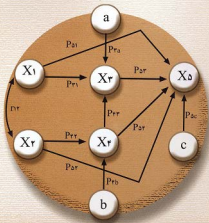


# پردازش و تحلیل داده ها در تحقیقات اجتماعی - اقتصادی

با استفاده از نرم افزار SPSS



دکتر خلیل کلانتری  
دانشیار دانشگاه تهران

# Data Processing and Analysis in Socio-Economic Research

BY:

Dr. Khalil Kalantari

Associate Prof. of Tehran University

شابک: ۹۶۴-۵۶۴۹-۲۲-۲



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

# پردازش و تحلیل داده‌ها در تحقیقات اجتماعی-اقتصادی

دکتر خلیل کلانتری  
دانشیار دانشگاه تهران

کلانتری، خلیل، ۱۳۴۳ -  
پردازش و تحلیل داده‌ها در تحقیقات اجتماعی و اقتصادی / خلیل  
کلانتری - تهران: شریف، ۱۳۸۲.  
۳۹۲ ص.: مصور، جدول، نمودار.

ISBN 964-5649-23-4

فهرست‌نویسی بر اساس اطلاعات فیپا.  
واژه‌نامه

کتابنامه: ص. ۳۶۱-۳۶۸؛ همچنین به صورت زیرنویس.  
۱. اس. پی. اس. اس. (برنامه کامپیوتر). ۲. اس. پی. اس. اس. تحت  
ویندوز (برنامه کامپیوتر). ۳. علوم اجتماعی -- روشهای آماری --  
برنامه‌های کامپیوتری. ۴. علوم اجتماعی -- تحقیق -- روشهای آماری.  
الف. عنوان.

۳۰۰ / ۲۸۵۵۳۶۹ الف ۵  
م ۸۲ - ۱۵۶۰۵

۸۸۳۲ / ک ۸ ر ۹  
کتابخانه ملی ایران



مهندسين مشاور طرح و منظر

پردازش و تحلیل داده‌ها در تحقیقات اجتماعی - اقتصادی

تألیف: دکتر خلیل کلانتری

ناشر: شریف

حروفچینی و صفحه‌آرایی: منیژه عیسایی

تعداد: ۲۵۰۰ نسخه

نوبت چاپ: دوم - ۱۳۸۵

لیتوگرافی: راستین

چاپ: شریف

قیمت: ۴۰۰۰۰ ریال ۱۵۵۰۰ تومان

شابک: ۹۶۴-۵۶۴۹-۲۳-۴ ISBN: 964-5649-23-4

تمامی حقوق برای مهندسين مشاور طرح و منظر محفوظ است.

تلفن پخش: ۶۶۴۹۳۸۵۳

E-mail: tarhomanzar@yahoo.com

انقلاب صنعتی در جهان غرب هم‌زادهای خود را به همراه آورد. تمرکز و تراکم جمعیت در نقاط شهری و افزایش میزان شهرنشینی منجر به بروز بحران‌های اجتماعی و اقتصادی گردید. افزایش ناهنجاری‌ها، وقوع جرم و جنایت، بیکاری و بسیاری مسائل دیگر که تا آن زمان سابقه نداشتند، جامعه‌شناسان را برآن داشت تا حل بحران‌های اجتماعی و رویارویی با پیامدهای اقتصادی و اجتماعی را هدف خود قرار دهند. گروهی از اندیشمندان، ضرورت شناخت مسائل اجتماعی را دنبال نمودند و چارچوب‌های تئوریک آن را فرموله کردند. دسته‌ای دیگر با استفاده از روش‌های علمی به کاربردی کردن نظریه‌ها پرداختند. اما با گذشت زمان روابط حاکم بر مسائل اجتماعی - اقتصادی از پیچیدگی بیشتری برخوردار شدند. بنابراین شناخت مسائل بغرنج با استفاده از روش‌های ساده مرسوم در گذشته امکان پذیر نبود. اقتضای روابط پیچیده آن بود که شناخت آنها با دقت علمی و به روش‌های نظام یافته کمی انجام شود. هر چه محاسبات و روش‌های کمی پیچیدگی بیشتری پیدا کردند، استفاده از ابزارها و تکنیک‌های جدیدتر الزامی‌تر شد. با گسترش رایانه و نرم‌افزارهای رایانه‌ای محاسبات پیچیده امکان پذیر و سرعت تحلیل افزایش یافت. با افزایش سهولت در استفاده از تکنیک‌ها و روش‌های آماری، احتمال بروز اشتباهات توسط کاربران نیز فزونی می‌یابد. برای گریز از این موضوع، استفاده درست از تکنیک‌ها و روش‌های تحلیل و پردازش اطلاعات ضرورت دارد. نوشتار حاضر برآن است تا این وظیفه مهم را با استفاده از نرم‌افزار SPSS که به صورت وسیعی توسط جامعه‌شناسان، برنامه‌ریزان شهری و منطقه‌ای، روانشناسان و سایر متخصصین به کار گرفته می‌شود به سرانجام رساند.

آقای دکتر خلیل کلانتری زمینه مطالعاتی اولیه خود را در حوزه علوم اجتماعی با تخصص مطالعات و برنامه‌ریزی منطقه‌ای تلفیق کرده و به تکنیک‌ها و روش‌های دو حوزه مطالعاتی به خوبی تسلط دارند. موفقیت ایشان در طول آموزش چندین دوره کارشناسی ارشد و دکترا و راهنمایی و مشاوره پایان‌نامه‌ها و رساله‌هایی با موضوعات اجتماعی و توسعه منطقه‌ای گواه این امر است. بخش تحقیق و توسعه مهندسی مشاور طرح و منظر امیدوار است انتشار این کتاب گامی هرچند کوچک در پاسخ به نیاز دانشجویان و کارشناسان توسعه، جامعه‌شناسی، اقتصاد، مدیریت، روانشناسی، برنامه‌ریزی منطقه‌ای، برنامه‌ریزان شهری و سایر متخصصین باشد.

مهندسین مشاور طرح و منظر



پیشگفتار: ..... بیست

### فصل اول: آشنایی با نرم افزار SPSS Win

۲۳	مقدمه: .....
۲۳	نحوه شروع بکار با برنامه SPSS: .....
۲۴	نحوه ورود به SPSS: .....
۲۶	پنجره Data Editor: .....
۲۸	اصلاح کردن اطلاعات: .....
۲۸	روش کپی کردن یا جابجا کردن اطلاعات در صفحه گستر: .....
۲۹	نحوه وارد کردن ستون یا سطر خالی در صفحه گستر: .....
۲۹	نحوه حذف کردن سطرها یا ستونها: .....
۲۹	نحوه حرکت در صفحه گستر: .....
۳۰	نحوه پیدا کردن یک متغیر: .....
۳۰	نحوه پیدا کردن یک سطر: .....
۳۰	نحوه پیدا کردن یک مقدار یا متغیر خاص یا یک پاسخگوی خاص: .....
۳۱	نحوه تعریف متغیرها و مقادیر: .....
۳۲	نحوه تعریف کردن متغیرها در SPSS: .....
۳۴	مقادیر نامعلوم (Missing Values): .....
۳۶	نحوه تعریف نوع متغیر: .....
۳۷	نحوه مرتب کردن ستونهای متغیرها: .....
۳۸	تعریف متغیرها با SPSS نسخه ۱۰ و بالاتر: .....
۴۰	نحوه استفاده از Templates برای تعریف کردن متغیرها: .....
۴۱	نحوه Template کردن در SPSS نسخه ۱۰ و بالاتر: .....
۴۲	نحوه ذخیره کردن اطلاعات در فایل: .....



## فهرست مطالب

صفحه	عنوان
۴۲	نحوه ذخیره کردن اطلاعات در دفعات بعدی:
۴۳	نحوه بازیابی یک فایل اطلاعاتی:
۴۳	نحوه بستن برنامه SPSS:
۴۴	پنجره Output:
۴۵	نحوه ذخیره یک فایل Out Put:
۴۶	نحوه بازیابی یک فایل Out Put از فلاپی:
۴۷	نحوه ویرایش محتوای پنجره Output:
۴۷	نحوه پرینت کردن:
۴۸	تسهیلات Help:

### فصل دوم: پردازش داده‌ها

۵۱	مقدمه:
۵۱	پردازش داده‌ها:
۵۳	۱- ویرایش داده‌ها:
۵۴	۲- کدگذاری:
۵۵	۳- استخراج و ورود داده‌ها به رایانه:
۵۶	۴- تعریف داده‌ها در رایانه:
۵۶	۵- کدبندی مجدد:
۶۲	کدبندی مجدد بوسیله نرم‌افزار SPSS:
۶۶	حساب کردن:
۶۸	نحوه انجام محاسبات با شرط:
۶۹	نحوه انتخاب زیر گروه‌ها برای تجزیه و تحلیل:
۶۹	الف) انتخاب زیرگروه دائمی:
۷۱	ب) انتخاب زیر گروه موقتی:

عنوان	صفحه
نحوه انتخاب نمونه تصادفی (Selecting Random Sample) : .....	۷۳
۶- آزمون برازش داده‌ها: .....	۷۳
<b>فصل سوم: توصیف داده‌ها</b>	
مقدمه: .....	۷۹
جداول توزیع فراوانی: .....	۸۰
شاخص‌های مرکزی: .....	۸۰
میان‌ه: .....	۸۲
نما یا مد: .....	۸۳
اندازه‌گیری پراکندگی: .....	۸۴
۱- دامنه: .....	۸۴
۲- پراکندگی دامنه دهک: .....	۸۵
۳- ضریب انحراف متوسط: .....	۸۵
۴- انحراف معیار: .....	۸۶
سنجش عدم تقارن: .....	۸۷
۱- چولگی: .....	۸۷
۲- درجه کشیدگی نمودار: .....	۸۹
آمار توصیفی و سطوح اندازه‌گیری: .....	۸۹
سطوح اندازه‌گیری: .....	۹۰
۱- سطح اسمی: .....	۹۰
۲- سطح ترتیبی: .....	۹۱
۳- سطح فاصله‌ای: .....	۹۲
۴- سطح نسبی: .....	۹۲
نحوه محاسبه جداول توزیع فراوانی، شاخص‌های مرکزی و پراکندگی با SPSS: .....	۹۳





**فصل چهارم: تحلیل روابط بین متغیرها**

مقدمه: ..... ۹۷

تحلیل روابط بین متغیرها: ..... ۹۷

ضرایب همبستگی مرتبط با متغیرهای اسمی: ..... ۹۹

کای اسکویر:..... ۹۹

شرایط استفاده از آزمون کای اسکویر:..... ۱۰۰

روش محاسبه جداول دوبعدی با نرم افزار SPSS..... ۱۰۱

ضریب همبستگی چوپروف: ..... ۱۰۴

ضریب همبستگی Phi:..... ۱۰۴

ضریب همبستگی توافق پیرسون: ..... ۱۰۵

ضریب کرامر: ..... ۱۰۵

ضرایب همبستگی مرتبط با متغیرهای ترتیبی:..... ۱۰۵

ضریب همبستگی رتبه‌ای کندال:..... ۱۰۶

ضریب گاما:..... ۱۰۶

ضریب همبستگی رتبه‌ای اسپرمن: ..... ۱۰۷

ضریب همبستگی پیرسون: ..... ۱۰۸

ضرایب همبستگی و سطوح اندازه گیری:..... ۱۰۸

نحوه محاسبه ضرایب همبستگی پیرسون اسپرمن و کندال با برنامه SPSS:..... ۱۰۸

ضریب همبستگی جزئی:..... ۱۱۲

نحوه محاسبه ضریب همبستگی جزئی با برنامه SPSS:..... ۱۱۲

**فصل پنجم: مقایسه میانگین‌ها**

مقدمه: ..... ۱۱۵

انتخاب آزمون مناسب برای مقایسه میانگین‌ها:..... ۱۱۶



## فهرست مطالب

صفحه	عنوان
۱۱۷	پیش فرضهای آزمونهای پارامتری
۱۱۸	روش محاسبه آزمون T با نرم افزار SPSS:
۱۱۹	آزمون t همبسته و نحوه محاسبه آن با SPSS:
۱۲۰	نحوه برخورد با مواردی که دارای Missing Data هستند:
۱۲۲	آزمون t مستقل و نحوه محاسبه آن با برنامه SPSS:
۱۲۳	تحلیل واریانس:
۱۲۵	اصول اساسی تحلیل واریانس:
۱۲۷	تحلیل واریانس یک طرفه با استفاده از برنامه SPSS:
۱۳۰	نحوه محاسبه میانگین زیرگروهها با نرم افزار SPSS:
۱۳۲	تحلیل واریانس دو طرفه با استفاده از برنامه SPSS:
۱۳۵	تحلیل واریانس برای گروههای همبسته:
<b>فصل ششم: آزمونهای غیر پارامتری</b>	
۱۳۹	مقدمه:
۱۴۰	انتخاب آزمون غیر پارامتری مناسب:
۱۴۱	آزمون نشانه:
۱۴۲	نحوه محاسبه آزمون نشانه با برنامه SPSS:
۱۴۴	آزمون مک نمار:
۱۴۴	نحوه محاسبه آزمون مک نمار با برنامه SPSS:
۱۴۷	آزمون ویلکاکسون:
۱۴۸	نحوه محاسبه آزمون ویلکاکسون با برنامه SPSS:
۱۵۰	آزمون فریدمن:
۱۵۰	روش محاسبه آزمون فریدمن با برنامه SPSS:
۱۵۲	آزمون کوکران:

## فهرست مطالب

صفحه	عنوان
۱۵۲	نحوه محاسبه آزمون کوکران با برنامه SPSS
۱۵۳	آزمون من و یتنی
۱۵۴	نحوه محاسبه آزمون من ویتنی با برنامه SPSS
۱۵۶	آزمون کولموگروف سیمرنف
۱۵۶	روش محاسبه آزمون کولموگروف - سیمرنف با برنامه SPSS
۱۵۷	آزمون کروسکال والیس
۱۵۷	روش محاسبه آزمون کروسکال والیس با برنامه SPSS
۱۶۰	آزمون میانه
۱۶۰	نحوه محاسبه آزمون میانه با برنامه SPSS
<b>فصل هفتم: طبقه‌بندی روشهای تحلیل چند متغیره</b>	
۱۶۳	مقدمه
۱۶۳	طبقه بندی روش های تحلیل چند متغیره
<b>فصل هشتم: رگرسیون چندگانه</b>	
۱۶۹	مقدمه
۱۶۹	تحلیل رگرسیون و ضریب همبستگی
۱۷۱	تحلیل رگرسیون
۱۷۵	داده‌های سازگار با رگرسیون
۱۷۷	روش محاسبه رگرسیون با برنامه SPSS
۱۸۰	روشهای رگرسیون چندگانه
۱۸۱	۱- روش رگرسیون چند گانه توأم (Enter)
۱۸۴	۲- روش گام به گام
۱۸۸	۳- روش حذف پس رو
۱۸۹	۴- روش پیشرو

**فصل نهم: رگرسیون لجستیک**

مقدمه: ..... ۱۹۱

روش محاسبه رگرسیون لجستیک بوسیله SPSS ..... ۱۹۲

**فصل دهم: تحلیل تشخیصی چندگانه**

مقدمه: ..... ۲۰۱

کاربرد تحلیل تشخیصی: ..... ۲۰۱

شباهتها و تفاوت‌های بین تحلیل تشخیصی با تحلیل رگرسیون و تحلیل واریانس: ..... ۲۰۴

مراحل انجام تحلیل تشخیصی: ..... ۲۰۷

۱- انتخاب متغیر: ..... ۲۰۷

۲- محاسبه تحلیل تشخیصی با برنامه SPSS: ..... ۲۰۸

۳- تفسیر نتایج در تحلیل تشخیصی: ..... ۲۱۲

**فصل یازدهم: تحلیل مسیر**

مقدمه: ..... ۲۲۳

اصول تحلیل مسیر: ..... ۲۲۴

تجزیه همبستگی: ..... ۲۲۹

محاسبه تحلیل مسیر با استفاده از برنامه SPSS: ..... ۲۳۰

**فصل دوازدهم: تحلیل واریانس چند متغیره**

مقدمه: ..... ۲۴۳

تحلیل واریانس چند متغیره چیست؟ ..... ۲۴۳

تفاوت بین دو گروه مستقل: ..... ۲۴۵

رهیافت یک متغیره: t تست: ..... ۲۴۶

رهیافت چند متغیره: T هتلینگ: ..... ۲۴۶

تفاوت بین K گروه مستقل: ..... ۲۴۸

صفحه	عنوان
۲۴۸	رهیافت یک متغیره ANOVA: .....
۲۴۹	رهیافت چند متغیره MANOVA: .....
۲۵۰	آزمون Post Hoc: .....
۲۵۱	مثالی از رهیافت چند متغیره: .....
۲۵۳	روش محاسبه MANOVA با برنامه SPSS: .....
۲۶۴	طرحهای عاملی: .....
۲۶۷	تحلیل کواریانس (ANCOVA) و تحلیل کواریانس چند متغیره (MANCOVA): .....
<b>فصل سیزدهم: تحلیل همبستگی کانونی</b>	
۲۶۹	مقدمه: .....
۲۷۰	اهداف همبستگی کانونی: .....
۲۷۱	نحوه استنتاج توابع کانونی: .....
۲۷۲	روش محاسبه تحلیل کانونی با برنامه SPSS: .....
۲۷۶	تفسیر نتایج بدست آمده از تحلیل کانونی: .....
<b>فصل چهاردهم: تحلیل عاملی</b>	
۲۸۱	مقدمه: .....
۲۸۱	بنیادهای اولیه تحلیل عاملی: .....
۲۸۲	تحلیل عاملی چیست؟ .....
۲۸۴	فرآیند تحلیل عاملی: .....
۲۸۵	نمودار تصمیم‌گیری در تحلیل عاملی: .....
۲۹۱	رهیافت‌های استخراج ماتریس همبستگی: .....
۲۹۳	تحلیل عاملی مشترک و تحلیل مؤلفه‌های اصلی: .....
۲۹۵	عاملهای چرخش یافته: .....
۳۰۳	معیارهای استخراج تعداد عاملها: .....



فهرست مطالب

صفحه	عنوان
۳۰۴	۱ - معیار مقدار ویژه: .....
۳۰۵	۲ - معیار پیشین: .....
۳۰۵	۳ - معیار درصد واریانس: .....
۳۰۶	۴ - معیار تست بریدگی: .....
۳۰۹	معیار تعیین سطح معنی داری بارهای عاملی: .....
۳۱۱	تفسیر ماتریس عاملی: .....
۳۱۳	روش انتخاب متغیرهای جایگزین برای تحلیل‌های بعدی: .....
۳۱۴	نحوه استفاده از مقادیر عاملی: .....
۳۱۵	محاسبه تحلیل عاملی با برنامه SPSS: .....
۳۱۵	روش محاسبه با برنامه SPSS: .....
۳۲۶	نامگذاری عاملها یا استنتاج اشتراک مفهومی: .....

فصل پانزدهم: تحلیل خوشه‌ای

۳۲۹	مقدمه: .....
۳۲۹	تحلیل خوشه‌ای چیست؟ .....
۳۳۰	بیان گرافیکی روش خوشه‌ای: .....
۳۳۲	فرآیند انجام تحلیل خوشه‌ای: .....
۳۳۲	نحوه اندازه‌گیری همگنی موارد یا موضوعات: .....
۳۳۵	الگوریتم خوشه‌ای کردن داده‌ها: .....
۳۳۷	روش خوشه‌ای سلسله‌مراتبی: .....
۳۳۹	پیوند تکی: .....
۳۴۰	پیوند کامل: .....
۳۴۰	پیوند متوسط: .....
۳۴۱	روش وارد: .....

فهرست مطالب

صفحه	عنوان
۳۴۲	روش مرکز ثقل:.....
۳۴۲	فرآیند خوشه‌ای غیرسلسله مراتبی:.....
۳۴۳	۱- روش آستانه متوالی.....
۳۴۳	۲- روش آستانه موازی.....
۳۴۳	۳- روش ایتیمم کردن.....
۳۴۴	الف) تعیین تعداد خوشه‌ها:.....
۳۴۴	ب) نام گذاری و توصیف خوشه‌ها:.....
۳۴۴	ج) تحلیل و تبیین خوشه‌ها:.....
۳۴۵	روش محاسبه تحلیل خوشه‌ای سلسله مراتبی با برنامه SPSS:.....
۳۵۴	روش خوشه‌ای K میانگین:.....
۳۵۵	روش محاسبه خوشه‌ای K میانگین با نرم افزار SPSS:.....
۳۶۱	منابع.....
۳۶۹	واژگان.....

## فهرست اشکال

صفحه	عنوان
۲۴	شکل (۱-۱) پنجره SPSS Data Editor
۳۳	شکل (۲-۱) پنجره Define Variable
۳۴	شکل (۳-۱) پنجره Define Labels
۳۵	شکل (۴-۱) پنجره Define Missing Values
۳۷	شکل (۵-۱) پنجره Define Variable Type
۳۸	شکل (۶-۱) پنجره Variable View
۴۱	شکل (۷-۱) پنجره Template
۴۴	شکل (۸-۱) پنجره Open File
۴۵	شکل (۹-۱) پنجره Output- SPSS Viewer
۴۸	شکل (۱۰-۱) پنجره مکالمه Print
۵۲	شکل (۱-۲) فرآیند پردازش داده‌ها
۶۳	شکل (۲-۲) پنجره Recode into Same Variables
۶۴	شکل (۳-۲) پنجره Old and New Values
۶۶	شکل (۴-۲) پنجره Compute Variable
۶۹	شکل (۵-۲) پنجره Compute Variable : If Cases
۷۰	شکل (۶-۲) پنجره Select Cases
۷۱	شکل (۷-۲) پنجره Select Cases:If
۷۲	شکل (۸-۲) پنجره Split File
۷۵	شکل (۹-۲) پنجره Reliability Analysis
۸۸	شکل (۱-۳) منحنی نرمال
۸۹	شکل (۲-۳) منحنی‌های دارای چولگی
۹۱	شکل (۳-۳) نوع داده و مقیاس اندازه‌گیری
۹۴	شکل (۴-۳) پنجره Frequency



## فهرست اشکال

صفحه	عنوان
۹۵	شکل (۳-۵) مشخص‌کننده‌های مرکزی و پراکندگی
۱۰۲	شکل (۴-۱) پنجره Crosstabs
۱۰۳	شکل (۴-۲) پنجره Crosstabs: Statistics
۱۱۰	شکل (۴-۳) پنجره Bivariate correlations
۱۱۳	شکل (۴-۴) پنجره Partial correlations
۱۱۷	شکل (۵-۱) درخت تصمیم‌گیری در مورد انتخاب آزمون مناسب برای بررسی سطح معنی‌داری
۱۱۹	شکل (۵-۲) پنجره Paired - Samples T Test
۱۲۲	شکل (۵-۳) پنجره Independent - Samples T Test
۱۲۹	شکل (۵-۴) پنجره One-Way ANOVA
۱۳۱	شکل (۵-۵) پنجره Means
۱۳۴	شکل (۵-۶) پنجره Univariate
۱۳۶	شکل (۵-۷) پنجره Repeated Measures Define Factor(s)
۱۳۷	شکل (۵-۸) پنجره تحلیل واریانس با اندازه‌گیری تکراری
۱۴۱	شکل (۶-۱) درخت تصمیم‌گیری برای انتخاب آزمون غیرپارامتری مناسب
۱۴۳	شکل (۶-۲) پنجره Two-Related Samples Tests
۱۵۱	شکل (۶-۳) پنجره Tests for several Related samples
۱۵۵	شکل (۶-۴) پنجره Two Independent samples Tests
۱۵۸	شکل (۶-۵) پنجره Tests for several Independent samples
۱۶۵	شکل (۷-۱) طبقه‌بندی روشهای تحلیل چند متغیره
۱۷۲	شکل (۸-۱) رابطه بین ساعات مطالعه و نمره دانشجو
۱۷۸	شکل (۸-۲) پنجره Linear Regression
۱۹۲	شکل (۹-۱) پنجره Logistic Regression

## فهرست اشکال

عنوان	صفحه
شکل (۲-۹) نمودار تقسیم‌بندی افراد بر اساس متغیرهای مستقل در رگرسیون لجستیک ...	۱۹۷
شکل (۱-۱۰) نمایش یک متغیره تشخیص مقدار Z (Z-Score) .....	۲۰۴
شکل (۲-۱۰) پنجره Discriminant Analysis .....	۲۰۹
شکل (۱-۱۱) نمودار تحلیل مسیر .....	۲۲۳
شکل (۲-۱۱) دیاگرام ساده‌ای از روش تحلیل مسیر .....	۲۲۷
شکل (۳-۱۱) نمودار تحلیل مسیر .....	۲۲۹
شکل (۴-۱۱) پنجره Linear Regression .....	۲۳۰
شکل (۶-۱۱) نمودار تفکیک شده مرحله سوم .....	۲۳۵
شکل (۷-۱۱) نمودار تفکیک شده مرحله چهارم .....	۲۳۷
شکل (۸-۱۱) نمودار نهایی تحلیل مسیر، همراه با ضرایب بتا .....	۲۳۷
شکل (۱-۱۲) پنجره Multivariate .....	۲۵۴
شکل (۲-۱۲) پنجره Multivariate Model .....	۲۵۴
شکل (۳-۱۲) پنجره Multivariate : Contrasts .....	۲۵۵
شکل (۴-۱۲) پنجره Post Hoc Multiple Comparisons for ... ..	۲۵۶
شکل (۵-۱۲) پنجره Multivariate : Options .....	۲۵۷
شکل (۱-۱۳) پنجره Multivariate .....	۲۷۵
شکل (۱-۱۴) دیاگرام تصمیم‌گیری تحلیل عاملی .....	۲۹۰
شکل (۲-۱۴) مقایسه نموداری تحلیل عامل نوع Q و تحلیل خوشه‌ای .....	۲۹۴
شکل (۳-۱۴) انواع واریانس وارد شده به ماتریس عاملی .....	۲۹۶
شکل (۴-۱۴) چرخش عاملی متعامد .....	۲۹۸
شکل (۵-۱۴) چرخش عاملی متمایل .....	۳۰۱
شکل (۶-۱۴) معیار تست بریدگی برای تعیین تعداد عاملها .....	۳۰۹
شکل (۷-۱۴) پنجره Factor Analysis .....	۳۱۹



## فهرست اشکال

صفحه	عنوان
۳۲۰	شکل (۸-۱۴) پنجره Factor Analysis: Extraction
۳۲۱	شکل (۹-۱۴) پنجره Factor Analysis: Rotation
۳۲۲	شکل (۱۰-۱۴) پنجره Factor Analysis: Options
۳۳۴	شکل (۱-۱۵) رابطه سن و میزان محافظه کاری
۳۳۷	شکل (۲-۱۵) نمودار شناسایی مقادیر دور افتاده
۳۳۷	شکل (۳-۱۵) فاصله اقلیدسی بین دو موضوع یا مورد
۳۳۹	شکل (۴-۱۵) نمودار خوشه‌ای، نشان دهنده واریانس درون و بین خوشه‌ای‌ها
۳۴۰	شکل (۵-۱۵) نمودار خوشه‌ای سلسله مراتبی
۳۴۲	شکل (۶-۱۵) در این نمودار نقاط A و B از همگنی بالایی برخوردار نیستند.
۳۴۴	شکل (۷-۱۵) نمودار مقایسه سنجش فاصله برای پیوند تکی و پیوند کامل
۳۴۸	شکل (۸-۱۵) پنجره Hierarchical Cluster Analysis
۳۴۹	شکل (۹-۱۵) پنجره Hierarchical Cluster Analysis: Method
۳۵۱	شکل (۱۰-۱۵) پنجره Hierarchical Cluster Analysis: Statistics
۳۵۲	شکل (۱۱-۱۵) پنجره Hierarchical Cluster Analysis: Plots
۳۵۸	شکل (۱۲-۱۵) پنجره K-Means Cluster Analysis
۳۶۰	شکل (۱۳-۱۵) پنجره K-Means Cluster Analysis

## فهرست جداول

صفحه	عنوان
۵۶	جدول (۱-۲): نحوه ورود داده‌ها به صفحه گستر SPSS
۶۰	جدول (۲-۲) کدگذاری مجدد برای هم جهت سازی گویه‌ها در زمینه سنجش گرایش روستائیان
۶۱	جدول (۳-۲) رابطه بین میزان درآمد و رضایت شغلی
۶۲	جدول (۴-۲) رابطه بین میزان درآمد و رضایت شغلی
۷۶	جدول (۵-۲) تغییر در مقدار آلفا در صورت حذف برخی از متغیرها
۹۳	جدول (۱-۳) آمار توصیفی و سطح اندازه‌گیری
۹۶	جدول (۲-۳) توزیع فراوانی پاسخگویان بر اساس جنسیت
۱۰۹	جدول (۱-۴) ضرایب همبستگی و سطح اندازه‌گیری
۱۱۱	جدول (۲-۴) نتایج محاسبات ضریب همبستگی پیرسون
۱۱۴	جدول (۳-۴) نتایج محاسبات ضریب همبستگی جزئی
۱۲۱	جدول (۱-۵) نتایج T تست همبسته
۱۲۴	جدول (۲-۵) مقایسه تفاوت بین میانگین درآمد زنان و مردان با استفاده از آزمون t مستقل
۱۳۰	جدول (۳-۵) مقایسه تفاوت میانگین نمرات اخذ شده بوسیله سه گروه از افراد آموزش دیده
۱۳۲	جدول (۴-۵) میانگین و انحراف معیار درآمد برای گروه‌های مختلف شغلی
۱۳۵	جدول (۵-۵) نتایج تحلیل واریانس دو طرفه
۱۴۳	جدول (۱-۶) نتایج آزمون نشانه
۱۴۵	جدول (۲-۶) نتایج آزمون مک نمار
۱۴۶	جدول (۳-۶) نظرات افراد در قبل و بعد از اجرای برنامه تبلیغات
۱۴۶	جدول (۴-۶) نحوه ورود داده‌ها به پنجره Data Editor
۱۴۷	جدول (۵-۶) نتایج آزمون مک نمار
۱۴۹	جدول (۶-۶) نتایج آزمون ویلکاکسون
۱۵۱	جدول (۷-۶) نتایج آزمون فریدمن
۱۵۳	جدول (۸-۶) نتایج آزمون کوکران



## فهرست جداول

عنوان	صفحه
جدول (۶-۹) نتایج آزمون من ویتنی.....	۱۵۵
جدول (۶-۱۰) نتایج آزمون کولموگروف - سیمرنف.....	۱۵۷
جدول (۶-۱۱) نتایج محاسبات آزمون کروسکال والیس.....	۱۵۹
جدول (۶-۱۲) نتایج آزمون مانه.....	۱۶۱
جدول (۸-۱۰) نحوه ورود داده‌های کیفی یا کدهای مجازی.....	۱۷۷
جدول (۸-۲) Model Summary.....	۱۷۹
جدول (۸-۳) ANOVA.....	۱۸۰
جدول (۸-۴) Coefficients.....	۱۸۰
جدول (۸-۵) Variables Entered/Removed.....	۱۸۳
جدول (۸-۱۶) Model summary.....	۱۸۳
جدول (۸-۷) ANOVA.....	۱۸۳
جدول (۸-۸) Coefficients.....	۱۸۴
جدول (۸-۹) Variables Entered/Removed.....	۱۸۶
جدول (۸-۱۰) Model Summary.....	۱۸۶
جدول (۸-۱۱) ANOVA.....	۱۸۷
جدول (۸-۱۲) Coefficients.....	۱۸۷
جدول (۸-۱۳) Excluded Variables.....	۱۸۷
جدول (۹-۱) نتایج حاصل از تحلیل رگرسیون لجستیک.....	۱۹۳
جدول (۱۰-۱) ارزیابی مشتریان از محصول جدید.....	۲۰۶
جدول (۱۰-۲) متغیرهای تحقیق قبل از ورود به تحلیل.....	۲۱۶
جدول (۱۰-۳) متغیرهایی که در گامهای مختلف وارد تحلیل تشخیصی شده‌اند.....	۲۱۷
جدول (۱۰-۴) متغیرهایی که در هر مرحله خارج از معادله باقی مانده‌اند.....	۲۱۸
جدول (۱۰-۵) Eigenvalues.....	۲۱۹

فهرست جداول

صفحه	عنوان
۲۱۹	جدول (۶-۱۰) Wilks' lambda
۲۲۰	جدول (۷-۱۰) Standardized canonical Discriminant Function coefficients
۲۲۰	جدول (۸-۱۰) Canonical Discriminant Function coefficients
۲۲۱	جدول (۹-۱۰) Structure Matrix
۲۲۱	جدول (۱۰-۱۰) Function at Group centroids
۲۳۱	جدول (۱-۱۱) Variables Entered / Removed
۲۳۱	جدول (۲-۱۱) Model Summary
۲۳۱	جدول (۳-۱۱) ANOVA
۲۳۲	جدول (۴-۱۱) Coefficientsa
۲۳۳	جدول (۵-۱۱) Variables Entered / Removed
۲۳۳	جدول (۶-۱۱) Model summary
۲۳۳	جدول (۷-۱۱) ANOVA
۲۳۳	جدول (۸-۱۱) Coefficients
۲۳۴	جدول (۹-۱۱) Variables Entered/Removed
۲۳۴	جدول (۱۰-۱۱) Model Summary
۲۳۵	جدول (۱۱-۱۱) ANOVA
۲۳۵	جدول (۱۲-۱۱) Coefficients
۲۳۶	جدول (۱۳-۱۱) Variables Entered/Removed
۲۳۶	جدول (۱۴-۱۱) Model Summary
۲۳۶	جدول (۱۵-۱۱) ANOVA
۲۳۶	جدول (۱۶-۱۱) Coefficients
۲۳۹	جدول (۱۷-۱۱) محاسبه اثرات مستقیم و غیرمستقیم سن (X1) بر رضایت شغلی (X5)
	جدول (۱۸-۱۱) محاسبه اثرات مستقیم و غیرمستقیم سابقه کار (X2) بر رضایت شغلی

## فهرست جداول

عنوان	صفحه
(X5) ..... ۲۳۹	
جدول (۱۱-۱۹) محاسبه اثرات مستقیم و غیرمستقیم مرتبه شغلی (X3) بر رضایت شغلی	
(X5) ..... ۲۴۰	
جدول (۱۱-۲۰) محاسبه اثرات مستقیم و غیرمستقیم درآمد (X4) بر رضایت شغلی (X5) .. ۲۴۰	
جدول (۱۱-۲۱) مجموع تاثیرات مستقیم و غیر مستقیم متغیرهای مستقل بر متغیر وابسته .. ۲۴۰	
جدول (۱۲-۱) نمرات اخذ شده توسط گروههای سه گانه .. ۲۵۲	
جدول (۱۲-۲) نتایج محاسبات تحلیل واریانس چند متغیره .. ۲۵۸	
جدول (۱۲-۳) تصویر شماتیک طرحهای عاملی .. ۲۶۵	
جدول (۱۳-۱) نتایج محاسبات تحلیل همبستگی کانونی .. ۲۷۶	
جدول (۱۴-۱) مقایسه تحلیل عاملی نوع Q با تحلیل خوشه‌ای سلسله مراتبی .. ۲۹۴	
جدول (۱۴-۲) مقایسه بارهای عاملی دوران یافته و دوران نیافته .. ۳۰۰	
جدول (۱۴-۳) متغیرهای انتخاب شده برای تحلیل عاملی .. ۳۱۸	
جدول (۱۴-۴) استخراج مجموعه اولیه عامل ها .. ۳۲۳	
جدول (۱۴-۵) ماتریس عاملی دوران نیافته .. ۳۲۶	
جدول (۱۴-۶) ضرایب عاملی دوران یافته .. ۳۲۷	
جدول (۱۴-۷) نمایی از عاملهای تحقیق و سهم هر یک از آنها .. ۳۲۸	
جدول (۱۵-۱) نتایج محاسبات تحلیل خوشه‌ای سلسله مراتبی .. ۳۵۴	
جدول (۱۵-۲) قندیل عمودی و نمودار درختی .. ۳۵۵	
جدول (۱۵-۳) نتایج محاسبات خوشه‌ای K میانگین .. ۳۶۱	

## پیشگفتار:

یک پژوهش علمی زمانی به نتایج مورد انتظار دست می‌یابد که داده‌های گردآوری شده، به روش علمی پردازش و تجزیه و تحلیل شوند. در این مرحله، آمار به عنوان ابزاری سودمند برای پردازش، توصیف و تحلیل داده‌ها و نتیجه‌گیری از یافته‌های تحقیق مورد استفاده قرار می‌گیرد. آماده‌سازی داده‌ها که شامل پالایش، کدگذاری، طبقه‌بندی و ورود داده‌ها به رایانه است خود مقدمه‌ای برای تحلیل داده‌ها محسوب می‌شود. در تحقیقات اجتماعی - اقتصادی، محققان علاقه‌مندند تا بررسی‌های گوناگون از جمله تعیین میزان تکرار و تواتر پدیده‌ها، گرایش‌های مرکزی، میزان تغییرپذیری و پراکندگی که در قالب آمار توصیفی مورد بررسی قرار می‌گیرند را از یک طرف، و تعیین روابط بین متغیرها، یا تفاوت‌های میان دو یا چند گروه و تبیین متغیر یا متغیرهای وابسته از طریق متغیرهای مستقل دیگر که عمدتاً از طریق آمار استنباطی بررسی می‌شوند را از سوی دیگر مورد تجزیه و تحلیل و تبیین قرار دهند.

امروزه کاربرد گسترده رایانه در تحقیقات امکان استفاده از روشهای آماری پارامتری و غیرپارامتری پیچیده و چند متغیره را آسانتر کرده است. بطوریکه در کمتر مطالعه‌ای است که از تحلیل‌های چند متغیره استفاده نشود. اگرچه همه این دستاوردهای بشری باعث شده‌اند تا دانشجویان و محققان از روشهای آماری در تحلیل و تبیین داده‌ها در سطح گسترده استفاده کنند، اما به دلیل عدم آشنایی با فلسفه و زمینه‌های بکارگیری هر یک از روشهای فوق، کاربردهای نابجای این تکنیکها در حوزه پژوهشهای اجتماعی - اقتصادی نیز افزایش یافته است. متأسفانه منابع موجود در زمینه روشهای تحقیق نیز کمتر به بخش تحلیل داده‌ها پرداخته‌اند و همین امر باعث گردیده است تا دانشجویان و محققان در مرحله تحلیل داده‌ها با مشکلاتی مواجه شوند و در مواردی نیز روشهای مختلف آماری را به صورت نادرست مورد استفاده قرار دهند و یا در تفسیر نتایج با مشکل مواجه گردند. به همین دلیل این کتاب درصدد است تا ضمن معرفی و تحلیل و تبیین نحوه محاسبه و تفسیر نتایج حاصل از روشهای مختلف آماری، خلاء مذکور را تا حد امکان پر کند.

در این کتاب طیف وسیعی از تکنیکهای آماری مورد بررسی قرار گرفته، و برای تحلیل و



تبیین هریک از آنها مثالهایی آورده شده است تا کاربرد هر روش بطور ملموس تر مشخص شود. برای انجام محاسبات سعی گردیده تا نحوه محاسبه هر روش نیز بوسیله نرم افزار SPSS تحت ویندوز توضیح داده شود و خروجی هر روش نیز در هر مرحله ارائه گردد.

این کتاب از ۱۵ فصل تشکیل شده است. از آنجا که آشنایی با نرم افزار SPSS تحت ویندوز برای پردازش و تحلیل داده‌ها ضروری می‌باشد، بنابراین فصل اول این کتاب به «آشنایی با برنامه SPSS» تحت ویندوز اختصاص یافته است. در این فصل خوانندگان محترم با اصول کار با برنامه SPSS آشنا می‌شوند تا قادر باشند در فصول مختلف عملیات آماری لازم را با استفاده از این نرم افزار انجام دهند. فصل دوم به «پردازش داده‌ها» اختصاص دارد. در این فصل سعی گردیده است تا نحوه آماده سازی داده‌ها و مراحل آن به تفصیل بیان گردد. از آنجا که توصیف داده در هر مطالعه‌ای مقدم بر تحلیل و تبیین آن می‌باشد بنابراین در فصل سوم روشهای «توصیف داده‌ها» و آماره‌های قابل استفاده در هر سطح توضیح داده شده است. نحوه محاسبه روابط بین متغیرها در بدو استفاده از آمار تحلیلی و دستیابی به شناخت لازم در خصوص روابط و همبستگی بین داده‌ها در تحقیقات اجتماعی - اقتصادی از جایگاه مهمی برخوردار می‌باشد به دلیل اهمیت این موضوع فصل چهارم به «تحلیل روابط بین متغیرها» اختصاص یافته است. از آنجا که مقایسه میانگین‌ها بویژه استفاده از آزمونهای «t» و «F» از روشهای آماری پر کاربرد در تحقیقات محسوب می‌گردند، بنابراین آزمونهای مربوط به «مقایسه میانگین‌ها» در فصل پنجم بطور مستقل مورد بحث و بررسی قرار گرفته است. آزمونهای غیرپارامتری نیز که طیف بسیار وسیعی را تشکیل می‌دهند در فصل ششم تحت عنوان «آزمونهای غیرپارامتری» بررسی شده‌اند. به دلیل اهمیت تصمیم‌گیری در انتخاب روشهای چند متغیره در تحلیل داده‌ها فصل هفتم به طبقه‌بندی این روشها اختصاص یافته و اصول حاکم بر بکارگیری هریک از آنها نیز مورد بررسی قرار گرفته است. از فصل هشتم تا فصل یازدهم به بررسی تکنیکهای مختلف تحلیل چند متغیره اختصاص یافته است، بطوریکه «رگرسیون چندگانه» در فصل هشتم، «رگرسیون لوجستیک» در فصل نهم، «تحلیل تشخیصی چندگانه» در فصل دهم، «تحلیل مسیر» در فصل یازدهم، «تحلیل واریانس چند متغیره» در فصل دوازدهم، «تحلیل همبستگی کانونی» در فصل

سیزدهم، «تحلیل عاملی» در فصل چهاردهم و «تحلیل خوشه‌ای» در فصل پانزدهم مورد بحث قرار گرفته است.

اگرچه در سراسر این کتاب سعی گردیده است تا از منابع مختلف داخلی و خارجی استفاده لازم بعمل آید، اما قطعاً کاستی‌هایی در آن وجود دارد که امید است اساتید، دانشجویان و محققان محترم خطاها و نارسائیهای موجود را در هر زمینه‌ای که در این کتاب وجود دارد به نگارنده گوشزد نمایند تا در چاپ‌های بعدی مورد توجه قرار گیرد.

در اینجا بر خود لازم می‌دانم تا از زحمات آقایان دکتر حسین شعبانعلی فمی، مهندس تقی شعبانعلی فمی و مهندس امیر حسین علی‌بیگی که متن اولیه این کتاب را مطالعه و نکات ارزنده‌ای را یاد آور شده‌اند صمیمانه تشکر کنم. همچنین از مهندسین مشاور طرح و منظر به دلیل تامین هزینه‌های انتشار این کتاب و از انتشارات شریف بدلیل ارائه خدمات چاپ صمیمانه تشکر و از خانم منیژه عیسائی نیز که با دقت تمام، مسئولیت حروفچینی این کتاب را تقبل کردند، قدردانی کنم. در آخر لازم است از همسر و فرزندم که تمام وقتی را که باید به آنها اختصاص می‌دادم صرف این کتاب کرده‌ام صمیمانه سپاسگزاری کنم.

خلیل کلانتری

پاییز ۱۳۸۲

## فصل اول

# آشنایی با نرم افزار SPSSWin

### مقدمه:

از آنجا که کاربرد رایانه در مطالعات اجتماعی - اقتصادی گسترش وسیعی یافته است و تحلیل داده‌ها در این کتاب نیز براساس استفاده از نرم افزارهای رایانه‌ی بنا نهاده شده است. لذا قبل از پرداختن به تکنیک های آماری، در این فصل به معرفی برنامه SPSS و نحوه کار با این نرم افزار پرداخته شده است. در فصول بعدی پس از تشریح و تبیین هر روش آماری، محاسبه آن با نرم افزار SPSS نیز توضیح داده شده است.

### نحوه شروع بکار با برنامه SPSS:

فرض کنید که برنامه SPSS که تحت ویندوز عمل می‌کند در رایانه شما کار گذاشته شده است. برای اینکه بتوانید از آن برنامه استفاده کنید موارد ذیل را باید بخوبی بدانید:

- ۱ - چگونه وارد برنامه SPSS شوید.
- ۲ - چگونه فایل ایجاد کنید و چطور اطلاعات خود را ذخیره نمایید.
- ۳ - چگونه به دستورات مختلف دست بیابید.

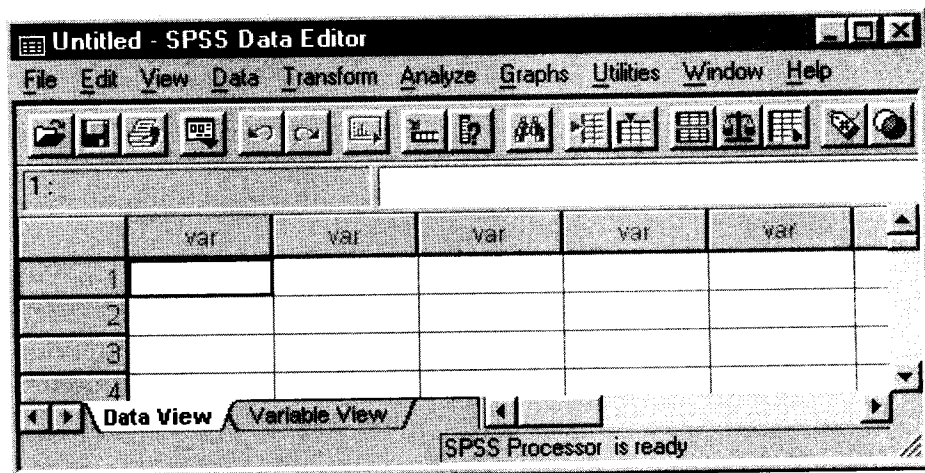
۴ - چگونه دستورات مختلف را در عمل بکار ببرید.

۵ - چگونه محتویات فایل Output را در فلاپی ذخیره کنید.

مباحث فوق سعی شده است در قسمتهای مختلف مورد بررسی و تجزیه و تحلیل قرار گیرند.

### نحوه ورود به SPSS:

چنانچه رایانه شما دارای ویندوز ۹۵، ۹۸ و یا ۲۰۰۰ باشد در این صورت با روشن کردن رایانه با صفحه‌ای مواجه می‌شوید که به آن میزکار گویند. در این صفحه مجموعه‌ای از آیکونها که یکی از آنها نیز ممکن است به SPSS تعلق داشته باشد وجود دارد. با دوبار کلیک کردن بر روی آیکون مربوط به SPSS برنامه فوق اجراء شده و شکل ذیل ظاهر می‌شود. این پنجره به پنجره Data Editor معروف است و تا زمانی که اطلاعاتی در آن وارد و ذخیره نشده باشد کلمه untitled در بالای آن ظاهر می‌شود و در صورت ذخیره کردن اطلاعات نام فایل در آنجا ثبت می‌گردد.



شکل (۱-۱) پنجره SPSS Data Editor

جعبه کنترل با علامت × در این پنجره در بالاترین خط و در سمت راست قرار دارد که با فشار

دهید و هم زمان با آن حرف H را فشار دهید).

آیتم‌های موجود در فهرست منوها نیز هریک دارای حرفی هستند که زیر آن خط تیره (-) وجود دارد. در اینجا بدون استفاده از کلید ALT و تنها با فشار دادن حروف فوق می‌توان به فعال کردن و اجرای منوها اقدام کرد.

خط پایین در پنجره Application که به نوار وضعیت (Status Bar) معروف است، پیغامی را صادر می‌کند که نشان می‌دهد برنامه SPSS در حال انجام چه کاری است.

### پنجره Data Editor:

این پنجره همانطوریکه در شکل (۱-۱) نشان داده شده است به صورت یک صفحه گستر، با اجرای SPSS در مانیتور ظاهر می‌شود و جایی است که اطلاعات در آن وارد می‌گردد. در ابتدا در بالای صفحه untitled نوشته شده است و این زمانی است که هیچ گونه اطلاعاتی در جدول ذخیره نشده است. در قسمت ستون، عناوین ستون‌ها تحت عنوان Var و در قسمت سطر، عناوین سطور با اعداد ۱، ۲، ۳ و ... مشخص شده است. بنابراین اگر اولین ستون را ستون A و اولین سطر را سطر ۱ بنامیم تقاطع این دو، مربعی را تشکیل می‌دهد که به سلول (Cell) معروف است. این سلول در مثال فوق با  $A_1$  مشخص می‌شود. به همین ترتیب سلول سمت راست به  $B_1$  و بعدی به  $C_1$  و الی آخر. ردیف زیر نیز  $A_2$ ،  $B_2$ ،  $C_2$  و غیره خواهد بود. برای وارد کردن اطلاعات در این سلولها کافی است که کرزر را در سلول مورد نظر قرار دهید و با تایپ کردن اعداد و فشار دادن دکمه‌های Enter و یا Tab یا کلیدهای  $\downarrow$  به سمت مورد نظر حرکت کنید. با فشار دادن دکمه Enter و یا کلید  $\downarrow$  عدد فوق در سلول مورد نظر ثبت و کرزر به سمت پایین حرکت می‌کند. در حالیکه با فشار دادن دکمه Tab و یا کلید  $\Rightarrow$  عدد تایپ شده در سلول مورد نظر وارد و کرزر به سمت راست حرکت می‌کند. این پنجره دارای Scroll Bar در سمت راست و پایین می‌باشد که با استفاده از دکمه‌های آن می‌توان قسمت‌های مختلف صفحه گستر را مشاهده کرد.

در SPSS تحت ویندوز، اطلاعات مستقیماً به صفحه گستر که در پنجره Data Editor قرار

دارد وارد می شود. وقتی که داده ها وارد SPSS می شوند، اطلاعات مربوط به یک فرد باید در یک ردیف وارد شود. در اینجا هر ردیف به Case معروف است.

هر ستون، داده های یک متغیر را شامل می شود. به عنوان مثال ستون دوم (Var00002)، ممکن است اطلاعات مربوط به جنسیت پاسخ دهنده را شامل شود. هریک از سلولها مقادیری را در خود جای می دهند که معمولاً به صورت اعداد می باشند. (اگر قرار باشد از حروف استفاده شود باید ابتدا موارد فوق تعریف شوند که نحوه اینگونه تغییرات بعداً بحث خواهد شد).

برای وارد کردن اطلاعات مراحل ذیل را دنبال کنید.

۱ - مطمئن باشید که پنجره Data Editor فعال است. اگر چنانچه این پنجره فعال نباشد

بر روی Title Bar پنجره فوق کلیک کنید.

۲ - مطمئن باشید که اولین سلول در گوشه سمت چپ بالای صفحه گستر فعال باشد

چنانچه این سلول فعال نباشد کلیدهای Ctrl + Home را هم زمان فشار دهید تا Cell Selector

در سلول فوق قرار گیرد.

۳ - مقدار مربوط به اولین متغیر اولین مورد را تایپ کنید. سپس دکمه Enter یا کلیدهای → یا

⌘ را فشار دهید.

در وارد کردن اطلاعات بهتر است همیشه اطلاعات مورد اول را بطور کامل وارد کرده و بعد به

ترتیب اطلاعات موارد بعدی را وارد کنید. بنابراین فشار دادن دکمه → راحت ترین راه برای دادن

اطلاعات به رایانه است.

وقتی که اولین عدد را وارد سلول بکنید عنوان سلول به صورت Var00001 در می آید.

زمانیکه اطلاعات برای ستون های بعدی نیز وارد می شود، عناوین ستون ها بترتیب Var00002،

Var00003 و غیره در می آیند. بدین ترتیب می توان کلیه اطلاعات مورد نظر را در ستونهای مختلف

وارد کرد.

**اصلاح کردن اطلاعات:**

اگر در اطلاعات وارد شده اشتباهی رخ داده باشد، برای اصلاح آن کرزر را در سلول مورد نظر قرار دهید سپس مقادیر درست را تایپ کرده و Enter را فشار دهید. اگر می‌خواهید یک عدد از یک دادهٔ چند رقمی بزرگ را اصلاح کنید ابتدا بر روی سلول مورد نظر کلیک کنید. مقادیر این سلول در زیر خط نوار آیکونها ظاهر می‌شود، سپس روی عددی که می‌خواهید اصلاح کنید کلیک کرده و سپس عدد مورد نظر را با استفاده از صفحه کلید تغییر دهید. برای وارد کردن اطلاعات در سلول مورد نظر کافی است تا Enter را فشار دهید.

**روش کپی کردن یا جابجا کردن اطلاعات در صفحه گستر:**

برای کپی کردن یا جابجا کردن اطلاعات در صفحه گستر مراحل ذیل را انجام دهید:

۱ - ابتدا، اطلاعات مورد نظر را انتخاب کرده و از طریق کلیک کردن و کشیدن موس محدوده مورد نظر را مشخص کنید. در صورتی که بخواهید کل ستون انتخاب شود، روی سر ستون مربوط به ستون مورد نظر کلیک کنید و اگر بخواهید کل سطر انتخاب شود، روی سر سطر مربوط به سطر مورد نظر کلیک کنید.

۲ - منوی Edit را انتخاب کنید.

۳ - اگر قصد کپی کردن اطلاعات فوق را دارید، Copy را از لیست منوی Edit انتخاب کنید.

اما اگر قصد جابجا کردن اطلاعات را دارید Cut را انتخاب کنید.

۴ - سپس کرزر را در جایی که می‌خواهید اطلاعات فوق کپی یا منتقل شود قرار داده و کلیک کنید تا مقصد مورد نظر مشخص شود.

۵ - با مشخص شدن مقصد، منوی Edit/Paste را انتخاب کرده تا اطلاعات در نقطه مورد نظر وارد شود. (شما می‌توانید اطلاعات کپی یا Cut شده را در هر نقطه از صفحه گستر وارد کنید. اما چنانچه جای خالی بین اطلاعات قبلی و اطلاعات Paste شدهٔ فعلی وجود داشته باشد ستون یا سطرهای فوق بطور اتوماتیک بوسیله System- missing value پر می‌شود).

**نحوه وارد کردن ستون یا سطر خالی در صفحه گستر:**

بعد از وارد کردن اطلاعات اگر ستون یا سطر فراموش شده باشد شما می توانید ستون یا سطر خالی به داخل اطلاعات وارد کنید و داده های خود را در آنها تایپ نمایید. برای وارد کردن سطر کرزر را در قسمت پایین جایی که می خواهید سطر جدید ایجاد کنید قرار دهید و سپس از منوی اصلی Data/Insert case را انتخاب کنید تا سطر جدید در صفحه گستر وارد شود.

برای وارد کردن یک ستون خالی کرزر را در سمت راست جایی که می خواهید ستون خالی ایجاد کنید قرار داده و منوی Data/Insert variable را انتخاب کنید تا ستون جدید وارد شود. برای این کار می توانید از آیکونهای بالای صفحه نیز استفاده کنید. این آیکونها با شکلهایی شبیه صفحه گستر همراه با یک فلش مشخص شده اند. در صورتی که کرزر را بر روی آنها قرار دهید کلمات Insert case یا Insert variable ظاهر می شوند، که با کلیک کردن دستور فوق اجراء می گردد.

**نحوه حذف کردن سطرها یا ستونها:**

برای این کار ابتدا بر روی شماره Case در سمت چپ صفحه گستر و یا بر روی نام متغیر در بالای صفحه کلیک کنید. بعد منوی Edit/Clear را انتخاب کنید و یا دکمه Delete را از صفحه کلید فشار دهید.

**نحوه حرکت در صفحه گستر:**

در صفحه گسترهای بزرگ لازم است که بتوان بر روی صفحه با سرعت بیشتری حرکت کرد. برای این کار کلید Home شما را به اولین سلول یک case و کلید End شما را به آخرین سلول یک Case می برد. برای حرکت به سمت بالا یا پایین با فشار دادن همزمان کلیدهای  $\uparrow$  + Ctrl شما را به اولین سطر یک ستون و کلیدهای  $\downarrow$  + Ctrl شما را به آخرین سطر یک ستون می برد. با هر بار فشار دادن دکمه Page up یا دکمه page down صفحه گستر به اندازه ارتفاع یک پنجره به سمت بالا یا پایین حرکت می کند. برای مرور قسمتهای مختلف می توان از Scroll Bar نیز



استفاده کرد.

#### نحوه پیدا کردن یک متغیر:

برای این کار ابتدا منوی Utilities/Variables را انتخاب کنید. سپس متغیر مورد نظر را در لیستی که در یک Box آمده است مشخص کرده و بر روی کلید Go To کلیک کنید تا کرزر بر روی متغیر فوق قرار گیرد. علاوه بر این با اجرای این دستور هر متغیری که انتخاب گردد مشخصات تعریف شده برای آن نیز در پنجره‌ای تحت عنوان Variable Information ظاهر می‌شود. برای اجرای این دستور می‌توانید از آیکونی که به شکل یک ستون همراه با یک علامت سوال است نیز استفاده کنید.

#### نحوه پیدا کردن یک سطر:

برای این منظور منوی Data/Go To را انتخاب کنید. سپس شماره Case مورد نظر را در Text box تایپ کرده و سپس دکمه OK را فشار دهید. برای اجرای این دستور می‌توانید از آیکونی که به شکل یک سطر همراه با یک فلش می‌باشد نیز استفاده کنید.

#### نحوه پیدا کردن یک مقدار یا متغیر خاص یا یک پاسخگوی خاص:

برای این کار توصیه می‌شود ابتدا کرزر را در اولین سلول قرار دهید (با استفاده از کلید  $\uparrow + \text{Ctrl}$ ) سپس از منوی اصلی Edit/Find... را انتخاب کنید. پس از آن شماره مورد نظر را در پنجره مکالمه تایپ کرده و دکمه Find Next را فشار دهید. با این دستور کرزر بر روی عدد فوق قرار می‌گیرد. با فشار دادن مجدد دکمه Find Next اگر این عدد مجدداً در همان ستون وجود داشته باشد پیدا می‌شود. در صورتی که عدد فوق پیدا نشود پیام Not found ظاهر می‌شود.

## نحوه تعریف متغیرها و مقادیر:

نام گذاری متغیرها:

نام متغیرها در بالای هر یک از ستون‌ها قرار می‌گیرند. با وارد کردن اطلاعات در صفحه گستر، برنامه SPSS بطور اتوماتیک به متغیرها از سمت چپ به راست به ترتیب اسامی Var00003, var00002, var00001 و غیره را می‌گذارد. اینگونه اسامی هیچ نوع اطلاعاتی را به استفاده کننده نمی‌دهد. بنابراین ضروری است تا اسامی معنی‌داری که معرف اطلاعات موجود در هر متغیر باشد را برای متغیرها تعریف کرد.

برای نام‌گذاری هر متغیر روش مناسب استفاده از چند حرف اول نام آن متغیر است که تعداد آن نباید بیشتر از هشت کاراکتر باشد<sup>(۱)</sup>. در بین حروف نباید فضای خالی یا نقطه در انتهای آن گذاشته شود. از سوی دیگر نباید از علائمی نظیر! یا \* استفاده کرد. از طرف دیگر هر نام صرفاً می‌تواند یکبار استفاده شود. به عنوان مثال اگر می‌خواهید نمرات افراد را برای دو زمان مختلف نام‌گذاری کنید می‌توانید از نام Mark1 و Mark2 استفاده کنید<sup>(۲)</sup>.

در همین جا لازم است تفاوت بین Variable Name و Variable Labels و Value Labels مشخص شود. در برنامه SPSS مفهوم Variable Name به منظور شناسایی متغیرها برای تجزیه و تحلیل بکار می‌رود. با مشخص کردن نام متغیر به هنگام دادن دستور به SPSS این برنامه می‌داند که چه متغیری را باید تجزیه و تحلیل کند. Variable Label به نتایج یا Output اضافه می‌شود تا مشخص شود که این متغیر چیست. Value Label هم به Output اضافه می‌گردد تا مشخص کند که منظور از یک مقدار یا مقوله خاص در یک متغیر چیست.

۱- برای نام گذاری متغیرها می‌توان از X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub> و غیره نیز استفاده کرد.

۲- اگر کامپیوتر شما دارای ویندوز فارسی باشد می‌توانید نام متغیرها را به فارسی نیز تایپ کنید.

**: Variable Labels**

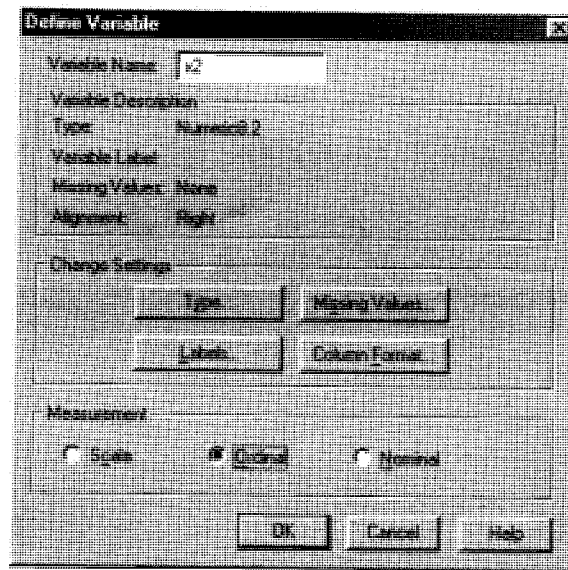
وقتی که شما به نتایج تجزیه و تحلیل نگاه می‌کنید ممکن است منظور از نام متغیر را فراموش کرده باشید. به عنوان مثال متوجه نشوید که منظور از Mark1 و Mark2 چیست؟ بنابراین بهتر است در نتایج تجزیه و تحلیل و یا در پرینت نهایی اطلاعات بیشتری وجود داشته باشد. برای این کار می‌توانید در موقع دادن نام به متغیرها، Variable Label هم به آنها اضافه کنید. Variable Label می‌تواند کاراکترهای بیشتری را بپذیرد و در این مورد می‌توان از فضای خالی یا فاصله نیز استفاده کرد. به عنوان مثال اگر شما Mark in first semester را تایپ کنید، دقیقاً به همین صورت به عنوان Label استفاده می‌شود. در صورتی که رایانه شما دارای ویندوز فارسی باشد می‌توانید Label فوق را به زبان فارسی نیز تایپ کنید.

**: Value Label**

Value Label به منظور مشخص شدن مقادیر، مورد استفاده قرار می‌گیرد که نتایج آن به Output اضافه می‌شود. به عنوان مثال اگر جنسیت پاسخگویان بوسیله کد ۱ برای مردان و کد ۲ برای زنان مشخص شده باشد، بهتر است در کنار کدهای فوق، Label توضیح دهنده نیز (Male و Female) در Output اضافه گردد.

**نحوه تعریف کردن متغیرها در SPSS:**

برای این منظور بر روی سرستون متغیر مورد نظر دوبار کلیک کنید. با این کار در SPSS نسخه‌های ۶ تا ۹ یک پنجره مکالمه در صفحه مانیتور باز می‌شود که کرزر بطور اتوماتیک در قسمت مربوط به Variable Name فعال است. برای اینکار می‌توانید از دستور Data/Define Variable نیز استفاده کنید (شکل ۱-۲).



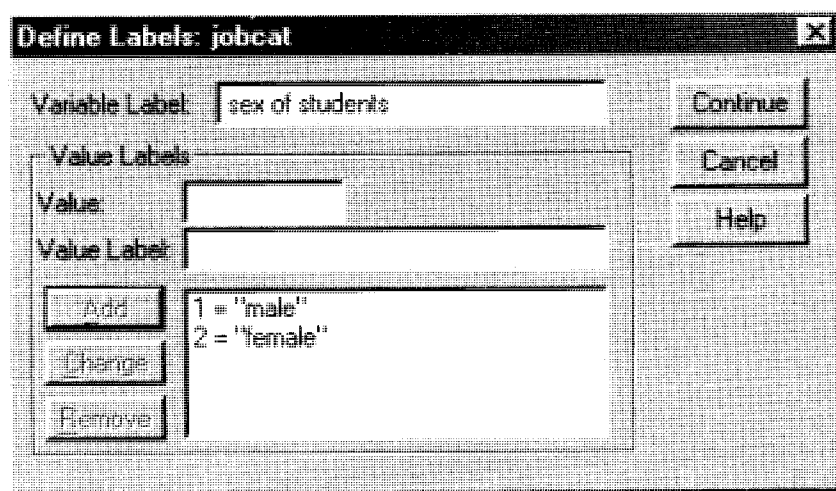
شکل (۱-۲) پنجره Define Variable

در این مرحله می‌توانید نام جدیدی برای متغیر فوق تایپ کنید. این نام می‌تواند حداکثر هشت کاراکتر داشته باشد. کاراکترهایی که برای نام فایل مجاز هستند در اینجا می‌توانند مورد استفاده قرار گیرند.

برای دادن Variable Label، ابتدا در همان پنجره مکالمه دکمه Labels را انتخاب کنید. با این کار پنجره دیگری باز می‌شود (شکل ۱-۳). بر چسب مورد نظر را در قسمت Variable Label تایپ کنید (آنچه که در این قسمت نوشته می‌شود بعد از انجام محاسبات در پنجره Output در عنوان جدول ظاهر می‌گردد).

به منظور دادن Label به مقادیر یا مقولات هر متغیر، ابتدا کرزر را در جعبه مقابل Value قرار داده و کد مورد نظر را تایپ کنید سپس Label مربوط به کد فوق را در جعبه مقابل Value Label تایپ کرده و بعد دکمه Add را کلیک کنید. به عنوان مثال، به منظور تعریف کد ۱ برای مردان در متغیر Sex، عدد ۱ را در قسمت Value و کلمه Male را در قسمت Value Label تایپ کرده و سپس دکمه

Add را فشار دهید. نتیجه به صورت "male"=1 در جعبه ظاهر می‌شود. به همین ترتیب ۲ را در قسمت Value و کلمه Female را در قسمت Value Label تایپ کرده و دکمه Add را فشار دهید (شکل ۱-۳). پس از تعریف مقادیر فوق دکمه Continue را کلیک کنید تا به پنجره Define variable بازگردید. با کلیک کردن دکمه OK تعاریف فوق اعمال می‌شوند.



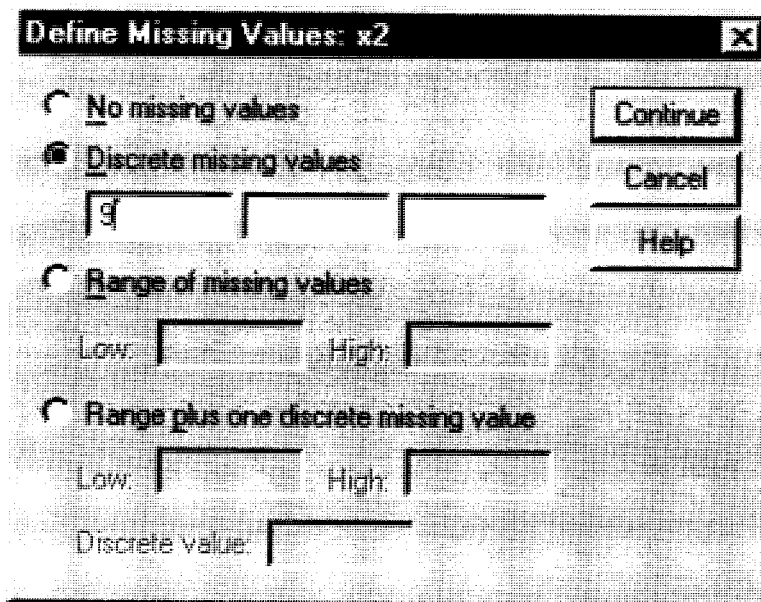
شکل (۱-۳) پنجره Define Labels

اگر قصد دارید که یکی از Labelها را تغییر دهید، ابتدا روی Label مورد نظر در جعبه لیست کلیک کنید، سپس Label جدید را در جعبه Value Label تایپ کرده و بعد دکمه Change را فشار دهید. اگر می‌خواهید که یکی از Labelها را از لیست حذف کنید برای این کار ابتدا روی Label مورد نظر در لیست کلیک کرده و سپس دکمه Remove را فشار دهید.

#### مقادیر نامعلوم (Missing Values):

برای آشنایی با نحوه برخورد با مقادیر نامعلوم، ابتدا باید تفاوت بین مفاهیم "System-Missing Values" با "User- Defined Missing Values" مشخص شود. در برنامه SPSS

تحت ویندوز هنگام وارد کردن اطلاعات به صفحه گستر اگر یکی از سلولها به صورت خالی رها شود و بقیه سلولها در ستون یا سطر فوق دارای اطلاعات باشد در این صورت SPSS بطور اتوماتیک در سلول خالی نقطه (.) گذاشته و آنرا به عنوان "System- Missing Values" لحاظ می کند. اما اگر پاسخگو در پرسشنامه ای به سؤالی پاسخ نداده باشد، نمی توان آن را به صورت سلول خالی رها کرد. در این صورت باید شماره یا کد خاصی را برای بدون پاسخ تعریف کرد. به عنوان مثال می توان کد ۱ برای مرد، کد ۲ برای زن و کد ۹ یا ۹۹ را برای بدون پاسخ در نظر گرفت. بنابراین در اینجا کد ۹ یا ۹۹ در متغیر Sex به عنوان "User- Defined Missing Values" خواهد بود. البته طبیعی است که باید برای SPSS مشخص کرد که کد فوق به معنای بدون پاسخ می باشد. معرفی Missing Values به SPSS از طریق قسمت Define Variable که در منوی Data/Define Variable می باشد انجام می گیرد. با دادن این دستور پنجره مکالمه تعریف متغیر در مانیتور ظاهر می شود. در این جعبه دکمه Missing Values را فشار دهید تا جعبه Define Missing Values باز شود (شکل ۱-۴).



شکل (۱-۴) پنجره Define Missing Values

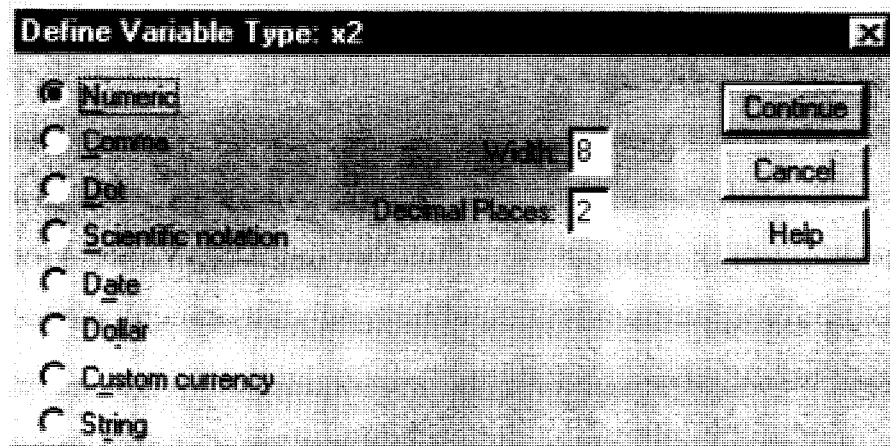
در اینجا شما می‌توانید تا سه سطح از Missing Value را وارد کنید. برای این منظور شما ممکن است یک کد برای بدون پاسخ (No- response)، کد دیگر برای نمی‌دانم (Dont know) و بالاخره یک سوم برای بی نظر (Undecided) انتخاب کنید. حتی شما می‌توانید مجموعه‌ای از اعداد را به عنوان Missing Value تعیین کرده و معرفی کنید.

برای تعریف کردن یک کد خاص به عنوان Missing Value آن را در قسمت Discrete Missing Value تایپ کرده و سپس دکمه Continue را کلیک کنید.

#### نحوه تعریف نوع متغیر:

وقتی که شما Data/Define Variable را از منو اصلی انتخاب می‌کنید، یا متغیر مورد نظر را دوبار کلیک کنید پنجره Define Variable در مانیتور ظاهر می‌شود (شکل ۱-۲). با فشار دادن دکمه Type اطلاعات بیشتری در اختیار شما قرار داده می‌شود (شکل ۱-۵). در این جعبه مشخص می‌شود که آیا متغیر فوق عدد (Numeric) است، یا متغیر غیر عددی و حروفی (String) است یا شکل دیگری از متغیرهاست. در صورتی که بخواهید در یکی از ستونها اسامی افراد یا موارد را به صورت حروفی بنویسید برای این کار باید ستون فوق را در پنجره type به string تبدیل کنید، تا امکان تایپ حروف در ستون مورد نظر در پنجره Data Editor فراهم گردد. زیرا در حالت عادی پنجره Data Editor تنها اعداد را می‌پذیرد. از طرف دیگر در همین جعبه عرض متغیر یا ستون در صفحه گستر مشخص است. به عنوان مثال، اگر در این جعبه در مقابل Width مقدار 8 و در قسمت Decimal Places عدد 2 باشد، نشان می‌دهد در این ستون تنها برای هشت عدد یا کاراکتر جا وجود دارد و این شامل اعداد اعشاری نیز می‌گردد و از طرف دیگر نشان می‌دهد که تنها تا دو رقم اعشار قابل وارد کردن است. بنابراین بزرگترین عددی که در این سلول جا می‌گیرد 99999.99 می‌باشد. تا زمانی که تغییری در این جعبه داده نشود، نوع متغیر قابل قبول عددی (Numeric) خواهد بود و عرض ستون نیز تنها هشت عدد را جا می‌دهد که شامل رقم اعشاری نیز می‌گردد. در صورتی که بخواهید می‌توانید اعشار ستون مورد نظر را از این قسمت حذف کنید.

اگر ترتیب پیش گزیده را عوض کنید، تا زمانی که با SPSS کار می کنید این تغییرات باقی می ماند اما چنانچه از برنامه خارج شوید ترتیب انجام شده از بین می رود زیرا این تغییرات همراه با ذخیره کردن فایل Save نمی شوند.



شکل (۵-۱) پنجره Define Variable Type

لازم است بدانید که مرتب کردن عرض ستون، تنها عرض ستون مشخص شده در صفحه را کنترل می کند نه شیوه وارد کردن متغیرها در برنامه را. بنابراین ممکن است شما با تغییر عرض ستون تعداد سه رقم اعشار مانند عدد 44.444 را در صفحه گستر وارد کنید، اما در مانیتور تنها 44.44 نشان داده می شود. در حالیکه وقتی محاسبات روی اعداد انجام می دهید رایانه اعداد اصلی را که با ۳ رقم اعشار وارد رایانه شده است لحاظ می کند.

#### نحوه مرتب کردن ستونهای متغیرها:

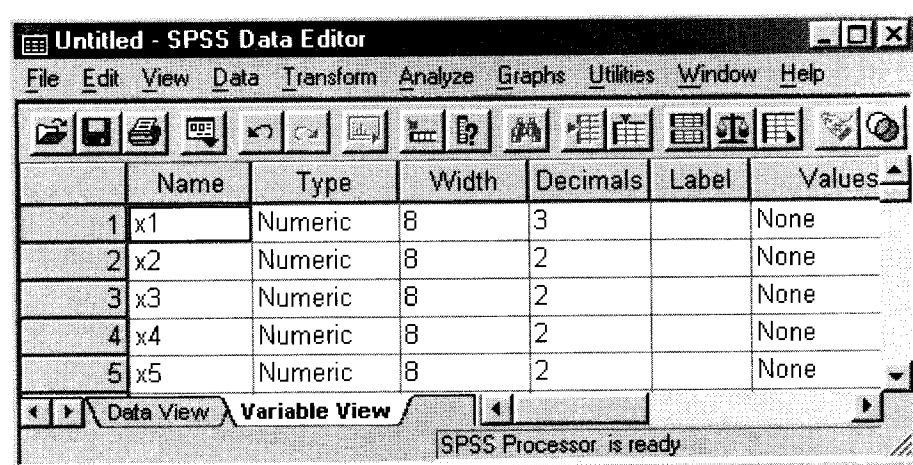
برای تغییر عرض ستونها و راست چین یا چپ چین کردن مقادیر متغیرها و یا قراردادن آنها در وسط ستون دکمه Column Format را از شکل ۱-۲ فشار دهید تا پنجره جدیدی باز شود. در این پنجره بطور پیش گزیده عرض ستون ۸ کارا کتر می باشد که می توان آن را تغییر داد. در این پنجره



از قسمت Alignment نیز می‌توان با انتخاب یکی از موارد Left، Center یا Right نسبت به تنظیم مقادیر در ستونهای پنجره Data Editor اقدام کرد.

### تعریف متغیرها با SPSS نسخه ۱۰ و بالاتر:

تعریف متغیرها در برنامه SPSS نسخه ۱۰ و بالاتر قدری متفاوت است. برای تعریف متغیرها چنانچه روی سر ستون هر متغیر دوبار کلیک کنید پنجره Variable View ظاهر می‌گردد. در این برنامه در واقع دو پنجره بر روی هم قرار گرفته‌اند یکی پنجره Data View و دیگری پنجره Variable View می‌باشد. پنجره Data View محل وارد کردن داده‌های اولیه و پنجره Variable View محل تعریف متغیرهاست (شکل ۱-۶). نام این دو پنجره در قسمت پایین سمت چپ نوشته شده است که با کلیک کردن موس بر روی هر کدام از آنها می‌توان پنجره مورد نظر را فعال نمود و به وارد کردن داده‌ها و یا تعریف متغیرها اقدام کرد.



شکل (۱-۶) پنجره Variable View

در پنجره Data View هر متغیر در یک ستون وارد می‌گردد، اما در پنجره Variable View مشخصات هر متغیر در یک سطر ثبت می‌شود. در اولین ستون پنجره Variable View باید نام متغیر نوشته شود. در این ستون تنها تا ۸ کاراکتر را می‌توان وارد نمود. هر نامی که در این قسمت ثبت گردد به عنوان نام متغیر در سر ستون پنجره Data Editor ظاهر می‌گردد. ستون دوم به تعریف نوع متغیر مربوط می‌باشد. متغیرها بطور پیش‌گزینه به صورت عددی (Numeric) تعریف شده‌اند اما اگر بخواهید در ستونهای Data View حروف تایپ کنید در این صورت باید از پنجره Variable View با فشار دادن سلول Type مربوط به متغیر مورد نظر دکمه کوچکی ظاهر می‌گردد که بر روی آن سه نقطه (...) وجود دارد با فشار دادن این دکمه پنجره Variable Type باز می‌شود که درست مانند پنجره‌ای است که در سایر نسخه‌های SPSS نیز وجود دارد. در این پنجره برای حروفی کردن ستون مورد نظر باید دکمه رادیویی String روشن و دکمه Ok کلیک شود. ستون سوم و چهارم در این پنجره تعداد کاراکترها و اعشار را نشان می‌دهند و ستون هشتم نیز عرض ستونها را تنظیم می‌کند. بطور پیش‌گزینه عرض ستونها به اندازه ۸ کاراکتر پیش‌بینی شده است. اما از این قسمت می‌توان عرض ستونها را کم یا زیاد کرد. برای این کار اگر در سلول مربوط به متغیر مورد نظر در ستون Columns کلیک شود، دو نوک فلش یکی به سمت بالا و دیگری به سمت پایین ظاهر می‌گردد که با فشار دادن فلش به سمت بالا عرض ستون افزایش و با فشار دادن فلش به سمت پایین عرض آن کاهش می‌یابد. ستون پنجم که به Label معروف است به منظور بیان توضیحات بیشتر در خصوص متغیر در نظر گرفته شده است. به عنوان مثال اگر در ستون اول نام متغیر جنسیت (Sex) تعریف شده باشد در این قسمت می‌توان توضیحات بیشتری راجع به آن متغیر بیان کرد. به عنوان مثال اگر در این ستون نوشته شود که «جنسیت دانشجویان دانشگاه تهران» در صورت محاسبه جداول توزیع فراوانی و سایر محاسبات آماری این جمله در عنوان جدول در پنجره Output ظاهر می‌شود. ستون Values که به عنوان ستون ششم در این قسمت پیش‌بینی شده است. به منظور تعریف مقادیر متغیر مورد استفاده قرار می‌گیرد. در این ستون اگر در سلول مربوط به متغیری که قرار است تعریف شود کلیک کنید دکمه کوچکی ظاهر می‌شود که بر روی آن سه نقطه (...) قرار دارد. با فشار دادن این دکمه، پنجره

Value Label باز می‌شود که در این پنجره به همان صورتی که در مورد تعریف برجسب‌ها در SPSS نسخه ۹ و پایین‌تر توضیح داده شد می‌توان به تعریف متغیرها پرداخت. سایر ویژگیهای متغیرها نیز در این پنجره با همین روش تعریف می‌شوند.

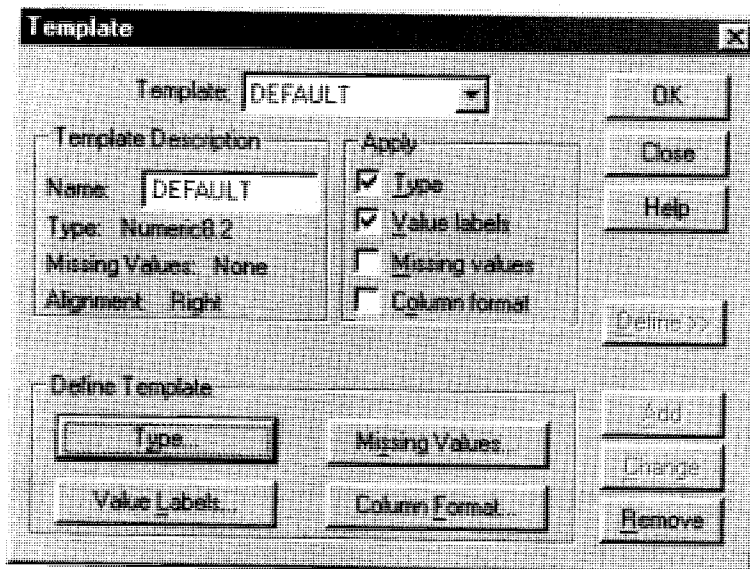
### نحوه استفاده از Templates برای تعریف کردن متغیرها:

ممکن است که تعداد زیادی متغیر به شرح ذیل داشته باشید و بخواهید برای همه آنها مقولات مشترک، عرض ستون یکسان و سایر شرایط مشترک تعریف کنید.

X <sub>۱</sub> - میزان علاقمندی و تمایل فراگیران به یادگیری و کسب دانش شغلی	(۱) بسیار کم	(۲) کم	(۳) متوسط	(۴) زیاد	(۵) بسیار زیاد
X <sub>۲</sub> - میزان هم سطحی بین فراگیران از نظر توانایی در یادگیری و درک مطالب آموزشی	(۱) بسیار کم	(۲) کم	(۳) متوسط	(۴) زیاد	(۵) بسیار زیاد
X <sub>۳</sub> - میزان ارتباط و هماهنگی مطالب و محتوای آموزش با اهداف دوره	(۱) بسیار کم	(۲) کم	(۳) متوسط	(۴) زیاد	(۵) بسیار زیاد
X <sub>۴</sub> - میزان کاربردی بودن مطالب و محتوای آموزش با نیاز فراگیران	(۱) بسیار کم	(۲) کم	(۳) متوسط	(۴) زیاد	(۵) بسیار زیاد
X <sub>۵</sub> - میزان تناسب بین مدت زمان دوره با حجم مطالب و محتوای آموزش	(۱) بسیار کم	(۲) کم	(۳) متوسط	(۴) زیاد	(۵) بسیار زیاد

برای اینکار ابتدا مقادیر هریک از سئوالات فوق را کدبندی کرده و هر سئوال (متغیر) را در یک ستون به Data Editor وارد کنید. پس از وارد کردن اطلاعات، برای اجرای دستور Template (در SPSS تحت ویندوز نسخه‌های ۹ و پایین‌تر) و تعریف همزمان چندین متغیر ابتدا در صفحه Data Editor با کشیدن موس بر روی عناوین، متغیرهای مورد نظر را مشخص کنید. سپس دستور

Data/Templates را اجرا کنید. با دادن این دستور پنجره Template در صفحه مانیتور ظاهر می شود. در این حالت دکمه Define را فشار دهید تا به دکمه های Define Template دست یابید (شکل ۷-۱). در این پنجره دکمه های Type، Value Labels، Missing Values و Column Format قرار دارند. با فشار دادن هر یک از دکمه های فوق می توان به تغییرات مورد نظر دست یافت و به همان ترتیبی که در مورد تعریف متغیرها در قسمت قبلی گفته شد به تعریف همزمان مجموعه ای از متغیرها اقدام کرد. اما پس از فشار دادن هر یک از دکمه های فوق و انجام تغییرات و کلیک کردن دکمه Continue، باید دکمه Change را در صفحه اصلی فشار دهید. و قبل از فشار دادن دکمه OK باید در قسمت Apply کلیه مواردی که در آن تغییرات ایجاد کرده اید را مشخص کنید تا تغییرات انجام گرفته اعمال شوند سپس دکمه OK را فشار دهید.



شکل (۷-۱) پنجره Template

نحوه Template کردن در SPSS نسخه ۱۰ و بالاتر:

اصول اجرای دستور Template در SPSS نسخه های ۱۰ و بالاتر قدری متفاوت است.

زمانیکه با این نسخه‌ها کار می‌کنید، می‌توانید از دستور Copy و Paste استفاده کنید. به عنوان مثال اگر در پنجره Variable View (شکل ۱-۶) متغیر  $X_3$  را با کدهای ۱ تا ۵ تحت عناوین بسیار کم، کم، متوسط، زیاد و بسیار زیاد تعریف کرده‌اید و قرار است متغیرهای  $X_4$  تا  $X_7$  نیز به همین شیوه تعریف شوند ابتدای روی سلول مربوط به متغیر  $X_3$  در ستون Values کلیک کنید، بعد دستور Edit/Copy را اجرا کرده و سپس سلولهای مربوط به متغیرهای  $X_4$  تا  $X_7$  را در همان ستون از طریق فشار دادن و کشیدن موس تیره کنید و بعد دستور Edit/Paste را اجرا نمایید. با این روش هر مشخصه دیگر را نیز می‌توان Template کرد.

### نحوه ذخیره کردن اطلاعات در فایل:

بعد از اینکه اطلاعات را وارد رایانه کردید، اولین کاری که باید انجام دهید ذخیره کردن اطلاعات است. در بسیاری از موارد ممکن است بخواهید اطلاعات فوق را در فلاپی ذخیره کنید. ابتدا مطمئن باشید که فلاپی فرمت شده در داخل drive A قرار گرفته باشد.

برای ذخیره کردن اطلاعات در فلاپی اقدامات زیر را انجام دهید:

۱ - مطمئن باشید که پنجره Data Editor فعال است.

۲ - منوی File/Save As را انتخاب کنید.

۳ - در این مرحله پنجره مکالمه Save Data As در صفحه مانیتور ظاهر می‌شود. نام فایل

را در قسمت File Name تایپ کنید (به عنوان مثال Mark)، بخاطر داشته باشید که از نقطه، جای

خالی و Comma استفاده نکنید. سپس دکمه Save را فشار دهید. با این دستور فایل شما در فلاپی

تحت عنوان Mark ذخیره خواهد شد. از آن پس نام فایل در قسمت Title bar در پنجره Data

Editor ظاهر می‌شود.

### نحوه ذخیره کردن اطلاعات در دفعات بعدی:

اگر اطلاعات وارد شده قبلاً یکبار ذخیره شده باشد و یا بر روی فایلی که قبلاً در دیسک وجود

داشته است کار می کنید، برای ذخیره کردن اطلاعات تحت نام قبلی، کافی است که منوی File/Save را انتخاب کنید. نسخه جدید این فایل بر روی فایل قبلی Over-Write می شود. اگر می خواهید که هم نسخه فایل قبلی و هم نسخه فایل جدید را داشته باشید در این صورت از منوی File/Save As (که قبلاً توضیح داده شد) استفاده کنید و نام دیگری به فایل بدهید. برای ذخیره کردن فایل می توانید از آیکون بالای پنجره که به شکل یک فلاپی می باشد نیز استفاده کنید.

### نحوه بازیابی یک فایل اطلاعاتی:

الف) بازیابی فایل اطلاعاتی با پسوند SAV

برای بازیابی یک فایل اطلاعاتی از فلاپی، دستور File/Open/Data را انتخاب کنید. و یا آیکون Open را فشار دهید. با این دستور پنجره Open File در صفحه مانیتور ظاهر می شود. (شکل ۸-۱).

فرض کنید که فایل اطلاعاتی مورد نظر شما در فلاپی است و قبلاً توسط برنامه SPSS ذخیره شده است در این صورت مورد Floopy را از قسمت Look in انتخاب کنید. کلبه فایل های ذخیره شده در فلاپی در Files Box لیست خواهند شد. روی فایل که می خواهید بازیابی کنید کلیک کرده تا به قسمت File Name منتقل شود برای باز کردن فایل دکمه Open را فشار دهید. بعد از باز شدن فایل در پنجره Data Editor نام فایل در قسمت بالای پنجره ظاهر می شود. همچنین از قسمت Type of Files می توانید فایل های دیگری از سایر برنامه ها نظیر Lotus 1-2-3, Excel و یا dBase و غیره وارد (Import) کنید.

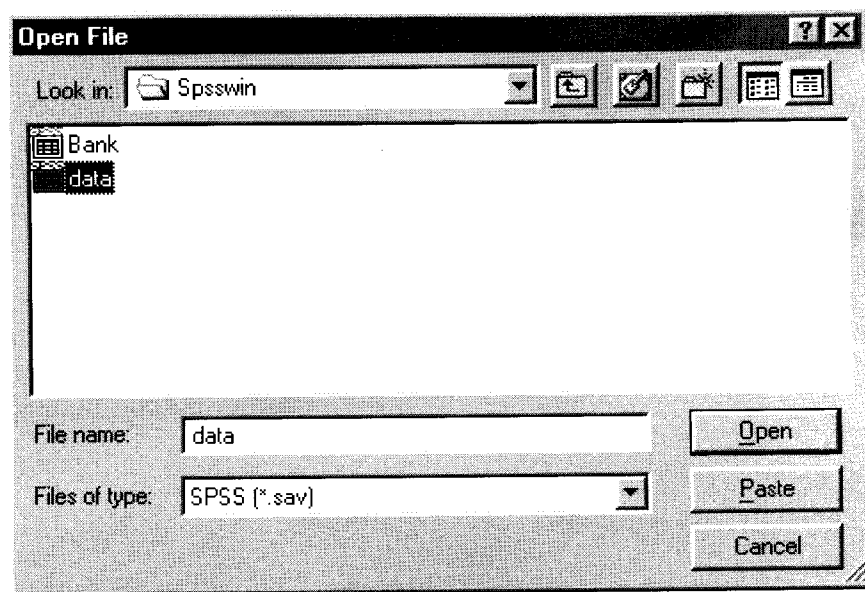
### نحوه بستن برنامه SPSS:

برای بستن برنامه SPSS به دو طریق می توانید اقدام کنید:

۱ - یا بر روی جعبه کنترل در گوشه راست (که به صورت علامت X می باشد) کلیک کنید.

Data Editor کلیک کنید و یا

۲ - منوی File Exit را از پنجره فوق انتخاب کنید.



شکل (۸-۱) پنجره Open File

در هر دو حالت چنانچه داده‌های وارد شده و یا محاسبات انجام گرفته ذخیره نشده باشند از شما خواهد خواست که آیا می‌خواهید داده‌های فوق ذخیره شوند یا خیر؟ در صورتی که دکمه Yes را بزنید در این صورت باید نام فایل را مشخص کرده و دکمه Save را کلیک کنید.

### پنجره Output :

دومین پنجره‌ای که در برنامه SPSS وجود دارد، پنجره Output نامیده می‌شود. نتایج محاسبات و تجزیه و تحلیل آماری در این پنجره نشان داده می‌شوند و مطالب ظاهر شده در این پنجره نیز قابل ویرایش و ذخیره کردن در فایل می‌باشند (شکل ۹-۱).

The screenshot shows the SPSS Output Viewer window. The main area displays a frequency table for variable X1. The table is titled 'X1' and has the following data:

	Frequency	Percent	Valid Percent
Valid male	14	46.7	46.7
Valid female	16	53.3	53.3
Total	30	100.0	100.0

شکل (۹-۱) پنجره Output- SPSS Viewer

وقتی که محاسبه در برنامه SPSS انجام گیرد پنجره Data Editor در پشت پنجره Out put قرار می گیرد. هر یک از پنجره های فوق دارای دکمه ای در پایین نوار وضعیت می باشند که با فشار دادن دکمه های فوق می توان بین پنجره های Data Editor و Output حرکت کرد. در این پنجره محاسبات در قالب جداول مختلف در سمت راست و فهرست Output در قسمت چپ لیست می شوند. با کلیک کردن بر روی هر یک از فهرستهای فوق جدول مربوط در سمت راست ظاهر می گردد. در صورتی که بخواهید هر یک از جداول فوق را حذف کنید می توانید جدول مورد نظر را از روی فهرست انتخاب کنید و سپس دکمه Delete را از جعبه کلید فشار دهید. در پنجره Output نیز مانند Data Editor منوها و آیکونهای مختلفی وجود دارد که برای کاربردهای مختلف مورد استفاده قرار می گیرند.

### نحوه ذخیره یک فایل Out Put :

برای ذخیره محاسبات:



- ۱ - مطمئن باشید که پنجره Output فعال است.
  - ۲ - منوی File/Save As... را انتخاب کنید.
- پنجره ذخیره Output که در مانیتور ظاهر می‌شود شبیه جعبه‌ای است که برای ذخیره کردن data file استفاده می‌شود. در این جعبه نام فایل را در قسمت File name بنویسید.
- ۳ - از قسمت Look in مورد Floppy را انتخاب کنید (در صورتی که بخواهید در فلاپی ذخیره کنید).
  - ۴ - دکمه Save را فشار دهید تا فایل شما در درایو A ذخیره شود.
- برای ذخیره کردن فایل Output می‌توانید از آیکونی که به شکل فلاپی در بالای صفحه قرار دارد نیز استفاده کنید.
- فراموش نکنید که هر نوع تغییر در فایل Output که بتازگی بازیابی شده است باید مجدداً ذخیره شود تا اصلاحات انجام شده از بین نرود. برای این کار کافی است تا در پنجره Output آیکون Save را که به شکل فلاپی می‌باشد فشار دهید.

#### نحوه بازیابی یک فایل Out Put از فلاپی :

- اگر شما محاسبات خود را در یک فایل ذخیره کرده باشید و بخواهید آن را پرینت یا ویرایش کنید، ابتدا باید فایل مورد نظر را در پنجره Output باز کنید. برای این منظور مراحل ذیل را انجام دهید:
- ۱ - از پنجره Output منوی File/Open/Output را انتخاب کنید تا پنجره مکالمه مربوط به Open File (شکل ۸-۱) در مانیتور ظاهر شود.
  - ۲ - اگر فایل مورد نظر در فلاپی باشد، از قسمت Look in مورد Floppy را انتخاب کنید. تا فایل‌های فلاپی در File box مانیتور ظاهر شوند. (در صورتی که نیاز باشد می‌توانید لیست را Scroll کنید).
  - ۳ - بر روی نام فایل که قصد بازیابی دارید کلیک کرده تا نام فایل فوق در قسمت File

Name ظاهر شود.

۴ - دکمه Open را فشار دهید تا محتوای فایل در پنجره Output ظاهر شود.

### نحوه ویرایش محتوای پنجره Output:

اگر محتوای پنجره Output متن باشد می توانید با دوبار کلیک کردن بر روی آن به اصلاح آن بپردازند. برای ویرایش اعداد داخل جداول نیز می توان به همین شیوه با دوبار کلیک کردن بر روی اعداد مورد نظر آنها را ویرایش کرد.

### نحوه پرینت کردن :

طبیعی است که بعد از انجام محاسبات آماری، شما نیاز دارید تا اطلاعات وارد شده در رایانه و اطلاعات Output را پرینت بگیرید. روش پرینت گرفتن برای همه این فایل ها، یکسان می باشد. اما قبل از اقدام به پرینت گرفتن باید مطمئن باشید که دستگاه پرینتر به رایانه شما وصل بوده و دارای کاغذ باشد و از طرفی پرینتر نیز On-line باشد.  
الف) برای پرینت کردن کل فایل، اقدامات ذیل را انجام دهید:

۱ - مطمئن باشید که پنجره ای که می خواهید از آنجا پرینت بگیرید فعال باشد.

۲ - منوی File/print... را انتخاب کنید یا آیکون پرینت که به شکل پرینتر می باشد را

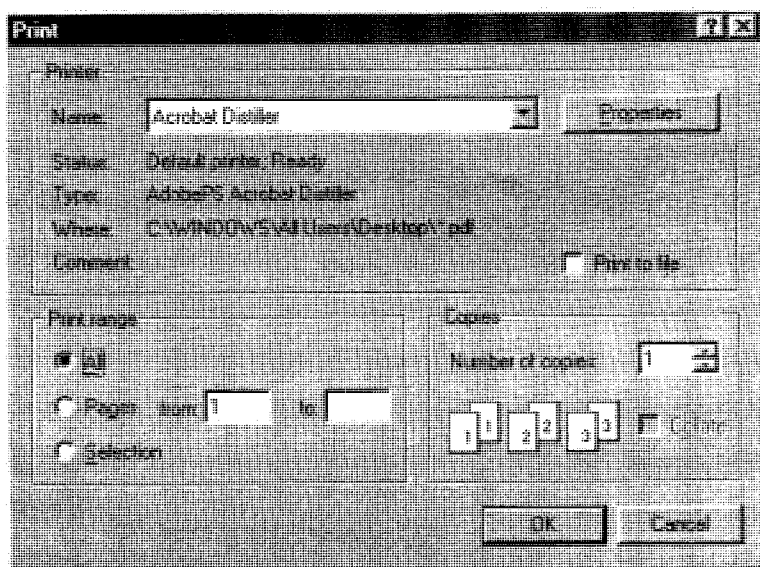
کلیک کنید.

۳ - در این مرحله پنجره مکالمه پرینت در مانیتور ظاهر می شود (شکل ۱-۱۰). بطور پیش

گزیده، فقط یک نسخه از فایل پرینت گرفته می شود. اما چنانچه نیاز به تعداد کپی بیشتری داشته

باشید در قسمت Copies تعداد مورد نظر را مشخص کنید.

۴ - دکمه OK را فشار دهید.



شکل (۱-۱) پنجره مکالمه Print

- ب) اگر بخواهید فقط بخشی از یک فایل را پرینت بگیرید به شیوه ذیل عمل کنید:
- ۱) ابتدا قسمتی که قصد دارید پرینت بگیرید را از طریق فهرست سمت چپ در پنجره Output مشخص کنید.
  - ۲) سپس منوی File/Print را انتخاب کنید یا آیکون پرینت را کلیک کنید.
  - ۳- وقتی که پنجره مکالمه پرینت در صفحه مانیتور ظاهر شد، دکمه Selection در آن بطور اتوماتیک انتخاب می‌گردد. دکمه OK را فشار دهید. تا مطالب انتخاب شده پرینت شود.
- اگر بخواهید تغییری در شکل صفحه پرینت شده ایجاد کنید باید دکمه Properties را از پنجره پرینت کلیک کنید و تغییرات مورد نظر را اعمال نمایید.

### تسهیلات Help :

برنامه SPSS تسهیلات Help وسیعی در اختیار استفاده کننده قرار می‌دهد. در این قسمت

نحوه استفاده از برنامه و حتی اصطلاحات آماری بکار رفته در برنامه توضیح داده شده است. این تسهیلات کمک زیادی به یادگیری برنامه SPSS می‌کند. با انتخاب منوی Help در پنجره اصلی، یا با کلیک کردن بر روی دکمه Help در هر قسمت که مشغول کار هستید می‌توانید پنجره Help را باز کنید.

## فصل دوم

# پردازش داده‌ها

### مقدمه:

در فصل قبل سعی گردید تا اصول کار با برنامه SPSS تشریح گردد و خواننده بتواند عملاً با این برنامه کار کند. در این فصل به تشریح نحوه آماده سازی داده‌ها برای ورود به رایانه و انجام محاسبات پرداخته شده است.

### پردازش داده‌ها:

داده‌های جمع آوری شده از طریق پرسشنامه یا سایر ابزارهای گردآوری اطلاعات قبل از هر نوع تجزیه و تحلیل باید آماده سازی شوند. بنابراین ویرایش<sup>(۱)</sup>، کدگذاری<sup>(۲)</sup>، ورود داده‌ها به رایانه، تعریف داده‌ها و طبقه‌بندی<sup>(۳)</sup> و کدگذاری مجدد<sup>(۴)</sup> آنها و همچنین حصول اطمینان از پایایی<sup>(۵)</sup> و

---

1- Editing

2- Coding

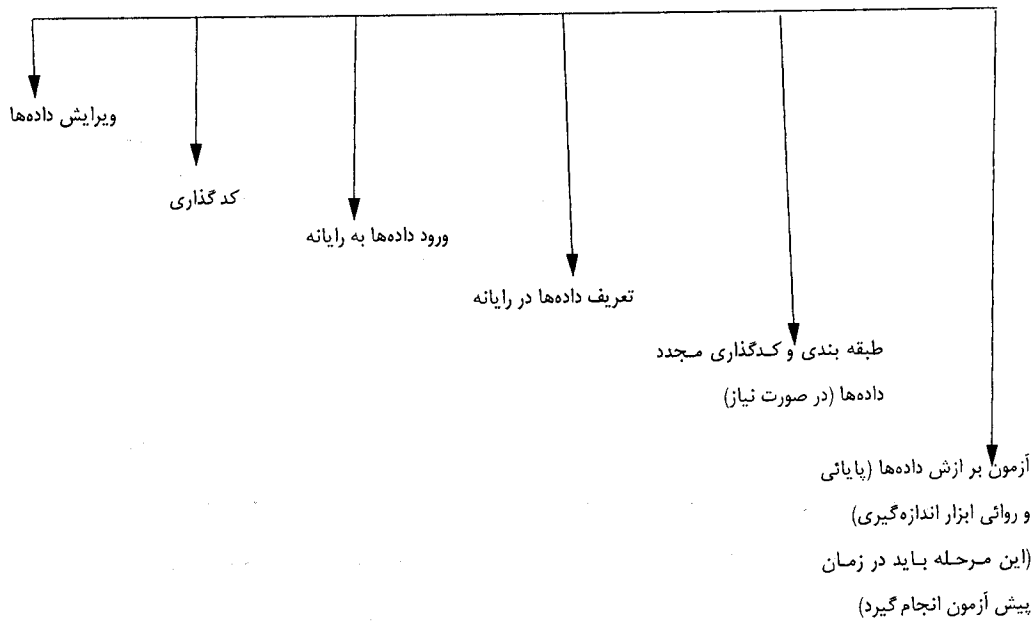
3- Classification

4- Recoding

5- Reliability

روایی<sup>(۱)</sup> ابزار اندازه‌گیری از جمله مراحل می‌باشند که باید قبل از تحلیل داده‌ها انجام گیرند. در تحقیقات اجتماعی - اقتصادی این مراحل را می‌توان تحت عنوان پردازش داده‌ها نامید. بنابراین همواره پردازش و آماده‌سازی داده‌ها مقدم بر تحلیل آنها بوده و هر نوع ساده‌انگاری در این مرحله می‌تواند نتایج و یافته‌های پژوهش را تحت تأثیر قرار دهد. لذا به دلیل اهمیت این موضوع فرآیند اساسی پردازش و تحلیل داده‌ها در شکل (۱-۲) ارائه و در ادامه، مراحل آن با تفصیل بیشتر مورد بحث قرار گرفته است.

شکل (۱-۲) فرآیند پردازش داده‌ها



## ۱- ویرایش داده‌ها:

ویرایش داده‌ها فرآیندی برای بررسی داده‌های خام گردآوری شده، رفع اشکالات، اصلاح و آماده سازی آنها می‌باشد. ویرایش داده‌ها به محقق کمک می‌کند تا از صحت و دقت داده‌ها قبل از تجزیه و تحلیل مطمئن شود. اینگونه ویرایش‌ها ممکن است در دو مرحله انجام گیرد: (۱) ویرایش در مرحله میدانی و زمان گردآوری داده‌ها و (۲) ویرایش در مرکز و پس از گردآوری آنها. ویرایش داده‌ها در میدان تحقیق و بلافاصله پس از گردآوری اطلاعات بویژه در روشهای مصاحبه و پرسشنامه‌های باز از اهمیت خاصی برخوردار است. زیرا محقق ممکن است در زمان مصاحبه و یا تکمیل پرسشنامه‌های باز به خلاصه نویسی بپردازد و یا از شرح و بسط موضوعات بپرهیزد. اما پس از مصاحبه و یا تکمیل پرسشنامه و قبل از فراموش کردن بحثهای مطرح شده در این مرحله پژوهشگر و یا مصاحبه کننده باید بلافاصله به ویرایش و اصلاح داده‌های جمع‌آوری شده بپردازد.

ویرایش در مرکز زمانی انجام می‌گیرد که کلیه فرمها و پرسشنامه‌ها جمع‌آوری شده‌اند. این ویرایش در مطالعات کوچک توسط یک نفر و در مطالعات وسیع‌تر توسط یک گروه انجام می‌گیرد. در این مرحله کلیه فرمها و پرسشنامه‌ها مورد بررسی قرار می‌گیرند و اشتباهات احتمالی آنها اصلاح می‌گردند. به عنوان مثال ممکن است محل ثبت پاسخها در برخی سؤالات اشتباه، یا زمان به جای هفته به ماه نوشته و یا مبلغ به جای ریال به تومان ثبت گردند. اینگونه اشتباهات عمومی در این مرحله اصلاح می‌گردند تا در زمان کدگذاری و استخراج وارد رایانه جهت تجزیه و تحلیل نشوند. در بعضی مواقع برخی پاسخهای غلط ممکن است از طریق سایر پاسخها قابل دستیابی باشند که پژوهشگر باید آنها را اصلاح کند. در مواقعی نیز ممکن است برای دستیابی به پاسخهای صحیح بویژه در سؤالات کلیدی و اصلی، محقق مجدداً به پاسخگو مراجعه نماید و یا در مواردی پاسخهای غلط را حذف و از وارد کردن آنها به مرحله تجزیه و تحلیل جلوگیری کند.

## ۲- کدگذاری:

کدگذاری عبارت از اختصاص یک عدد یا نماد خاص به هریک از پاسخها به منظور قابل طبقه‌بندی کردن داده‌های جمع‌آوری شده است. اینگونه کدگذاری به محقق کمک می‌کند تا در هر مرحله مجموعه یا طیفی از پاسخها را در یک طبقه خاص گروه‌بندی نماید تا فرآیند تحلیل‌های آماری را آسانتر کند. تصمیم‌گیری برای کدگذاری سوالات و پاسخها بهتر است در مرحله تدوین پرسشنامه انجام گیرد. این عمل باعث می‌شود تا محقق پس از تکمیل پرسشنامه، در موقع استخراج و وارد کردن داده‌ها به رایانه با مجموعه‌ای از داده‌های کدگذاری شده آماده، مواجه شود. ضمن اینکه این امر در مراحل مختلف از اشتباهات احتمالی نیز جلوگیری می‌کند. در تنظیم پرسشنامه و کدگذاری آن توصیه می‌شود که از کدبندی کردن متغیرهای کمی مانند سن، درآمد، نمره و غیره پرهیز گردد. حتی بهتر است ورود اینگونه داده‌ها به رایانه نیز بدون کدبندی انجام گیرد و مقادیر اصلی وارد رایانه شود. زیرا اینگونه متغیرها تنها با یک دستور در هر مرحله قابل کدبندی می‌باشند. اما اگر داده‌ها به صورت کدبندی شده وارد رایانه شود در این صورت امکان تحلیل اینگونه داده‌ها با استفاده از آزمونهای پارامتری از بین خواهد رفت و باب بکارگیری اینگونه آزمونها برای محقق بسته خواهد شد و او تنها می‌تواند از آزمونهای غیر پارامتری استفاده کند.

کدگذاری داده‌ها باید از یک سری اصول و استانداردهایی تبعیت کند. برای کدگذاری داده‌ها باید مراحل ذیل انجام گیرد.

۱- پس از تکمیل و جمع‌آوری کلیه پرسشنامه‌ها، ابتدا باید به هریک از آنها یک کد اختصاص یابد. (با توجه به اینکه در مرحله پرسشگری پرسیدن نام پاسخگو و درج آن در پرسشنامه ضرورت ندارد، بنابراین هریک از پرسشنامه‌ها با یک کد مانند 001, 002, 003 و الی آخر مشخص شوند. به عنوان مثال اگر تعداد پرسشنامه ۹۰ مورد باشد این کدها تا ۹۰ ادامه پیدا می‌کند. (در رایانه هریک از این کدها به یک Case معروف است).

۲- هریک از سوالات مطرح شده به ترتیب  $x_1$ ,  $x_2$ ,  $x_3$  و الی آخر تعریف گردد. اگر متغیر وابسته در پرسشنامه وجود داشته باشد متغیر فوق با علامت  $y$  نمایش داده شود.



۳ - مقولات مربوط به سؤالات باید با اعداد ۱، ۲، ۳ و الی آخر کدبندی شوند.

۴ - معمولاً کد ۸ برای مقوله بی‌مورد<sup>(۱)</sup> و کد ۹ برای مقوله بدون پاسخ اختصاص می‌یابد. چنانچه در پرسشنامه مقولات برخی از سؤالات بیش از ۹ مورد باشد در این صورت به منظور پرهیز از تداخل کدها باید کد ۸۸ به مقوله بی‌مورد و کد ۹۹ به مقوله بدون پاسخ اختصاص یابد. نمونه‌ای از کدنامه در ذیل ارائه شده است.

علامت	متغیر	مقولات
X <sub>1</sub>	جنسیت:	زن (۱) مرد (۲)
X <sub>2</sub>	وضعیت تأهل:	متاهل (۲) مجرد (۱)
X <sub>3</sub>	میزان تحصیلات:	بی‌سواد (۱) خواندن و نوشتن (۲) ابتدایی (۳) سیکل (۴)
		دیپلم (۵) لیسانس (۶) فوق لیسانس و بالاتر (۷)
Y	میزان رضایت از شغل	بسیار کم (۱) کم (۲) تا حدودی (۳) زیاد (۴) بسیار زیاد (۵)

### ۳- استخراج و ورود داده‌ها به رایانه :

استخراج داده‌ها می‌تواند به صورت دستی و در قالب یک جدول ماتریسی انجام گیرد و یا مستقیماً به صفحه گستر پنجره Data Editor در برنامه SPSS وارد شود. این صفحه درست شبیه جدول ماتریسی است که در ذیل آمده است. در روش استخراج دستی ابتدا داده‌های کدبندی شده به صورت یک جدول ماتریس استخراج می‌گردند. در این جدول کدهای اختصاص داده شده به هریک از پرسشنامه‌ها یا پاسخگویان در ستون اول و هریک از متغیرها یا سؤالات پرسشنامه در ستونهای بعدی وارد می‌شوند. در خانه محل برخورد کد پرسشنامه و سؤالات، کد مربوط به مقولات سؤالات

۱- منظور از "بی‌مورد" این است که ممکن است بدلیل شرایط خاص پاسخگو، بعضی از سؤالات برای او بی‌مورد باشد. به عنوان مثال اگر در سؤال قبلی وضعیت تأهل پرسیده شود و پاسخگو جواب دهد که مجرد است، اگر سؤال بعدی در مورد تعداد فرزندان باشد در این صورت سؤال فوق برای او بی‌مورد خواهد بود.

نوشته می‌شود (جدول ۱-۲).

جدول (۱-۲): نحوه ورود داده‌ها به صفحه گستر SPSS

Code	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	...	X <sub>45</sub>	Y
1	۱	۱	۱	۱	۱	۲	...	۲	۳
2	۲	۱	۳	۴	۵	۴	...	۲	۱
3	۱	۲	۴	۳	۵	۴	...	۳	۲
4	۲	۱	۲	۳	۲	۳	...	۱	۱
5	۱	۲	۱	۵	۳	۲	...	۲	۳
6	۲	۲	۵	۲	۴	۱	...	۳	۲
.	.	.	.	.	.	.	...	.	.
.	.	.	.	.	.	.	...	.	.
100	۲	۱	۳	۵	۳	۲	...	۱	۲

#### ۴- تعریف داده‌ها در رایانه:

در استخراج رایانه‌ای، اطلاعات جمع‌آوری شده به همان صورتی که در جدول (۱-۲) مشاهده می‌گردد مستقیماً در صفحه گستر رایانه وارد می‌شوند. پس از آن با استفاده از دستور Define Variable متغیرهای فوق به صورت جداگانه تعریف، و در صورت مشترک بودن مقولات متغیرها Template می‌شوند<sup>(۱)</sup>.

#### ۵- کدبندی مجدد:

اطلاعات جمع‌آوری شده ممکن است در اکثر مواقع اهداف تحقیق را تأمین نکند و طبقه

۱- نحوه تعریف متغیرها و یا Template کردن آنها در فصل ۱ توضیح داده شده است.

بندیها و کدگذاریهای اولیه نیاز به بازنگری مجدد داشته باشند. این فرآیند در پردازش داده‌ها به کدبندی مجدد<sup>(۱)</sup> معروف است. کدبندی مجدد به محقق امکان می‌دهد تا داده‌های گردآوری شده را به صورت هدفمند و متناسب با نیازهای تحقیق در هر مرحله از پژوهش بازسازی کند.

در این فرآیند کدهای اولیه اختصاص یافته به متغیرها تغییر یافته و نظم و ترتیب طبقات نیز تغییر می‌کند. این گونه کدبندیهای مجدد به اشکال مختلف انجام می‌گیرد که برخی از آنها در ذیل به تفصیل توضیح داده شده‌اند.

الف) در برخی مواقع ممکن است مقادیر مربوط به متغیرها به صورت کمی گردآوری و به رایانه وارد شده باشند، اما در مرحله‌ای خاص نیاز به طبقه‌بندی مجدد و تبدیل آن به مقادیر ترتیبی باشد. به عنوان مثال افرادی که درآمدشان کمتر از ۵۰ هزار تومان است در یک گروه و کسانی که درآمدشان بین ۵۱ تا ۱۰۰ هزار تومان است در گروه دیگر و الی آخر قرار گیرند. بدین ترتیب افراد فوق ممکن است در چند گروه درآمدی طبقه‌بندی شوند. اما مسئله اساسی این است که این طبقه‌بندی مجدد داده‌های کمی براساس چه معیار یا ملاکی انجام گیرد. روش معمول در اینگونه طبقه‌بندیها استفاده از حد بالا و پایین داده‌های کمی و محاسبه فاصله طبقاتی است. اما طبقه‌بندی داده‌ها براساس فاصله طبقاتی معمولاً با دو مشکل اساسی روبرو می‌باشد.

(۱) اولین مشکل این است که چند طبقه باید وجود داشته باشد و آیا فاصله طبقاتی برابر باشد

یا خیر؟

در مورد تعداد طبقات پاسخ مشخصی وجود ندارد و تصمیم‌گیری در مورد آن به تجربه محقق و شرایط تحقیق بستگی دارد. این تعداد باید به شیوه‌ای تعیین شوند که طبقه بندی را برای تجزیه و تحلیل معنادار کند و در این بین محقق ممکن است بین ۳ تا ۱۵ طبقه را تشخیص دهد. در مورد بخش دوم سؤال می‌توان گفت که فاصله طبقاتی ترجیحاً باید مساوی باشد، اما در برخی موارد ممکن است فاصله طبقاتی نابرابر مناسب‌تر بنظر رسد. بنابراین هدف تحقیق در این مورد تعیین‌کننده

اصلی است. ضرایب ۲، ۵ و ۱۰ به عنوان فاصله طبقاتی مناسب پیشنهاد می‌گردد. برخی از آماردانان فرمول پیشنهادی استورگس<sup>(۱)</sup> را برای تعیین فاصله طبقاتی مناسب می‌دانند. این فرمول بشرح ذیل است.

$$i = \frac{R}{(1 + 3.3 \log N)}$$

در این معادله:

$i$  = فاصله طبقاتی

$R$  = دامنه (که عبارتست از تفاوت بین بزرگترین مقدار و کوچکترین مقدار موجود در متغیری است که قرار است در سطوح مختلف طبقه بندی شود).

$N$  = تعداد مواردی که باید گروه بندی شوند.

گاهی ممکن است خود محقق راساً به تعیین تعداد طبقات بپردازد. در این صورت می‌توان از فرمول ذیل استفاده کرد.

$$i = \frac{R}{n}$$

$i$  = فاصله طبقاتی

$R$  = دامنه (که عبارتست از تفاوت بین بزرگترین مقدار و کوچکترین مقدار موجود در متغیری

است که قرار است در سطوح مختلف طبقه بندی شود).

$n$  = تعداد طبقات مورد نظر محقق

باید توجه داشت که چنانچه یک یا دو مورد از مقادیر مورد نظر بسیار زیاد و یا بسیار کم باشد

در این صورت می‌توان ابتدای طبقه اول و یا انتهای طبقه آخر را باز گذاشت. مانند در آمد کمتر از ۵۰ هزار تومان و یا بالاتر از ۲۰۰ هزار تومان. اگرچه اینگونه طبقه بندی چندان مناسب نمی‌باشد اما در برخی مواقع تنها راه حل بنظر می‌رسد.

(۲) مشکل انتخاب محدوده طبقات: در زمان تعیین محدوده طبقات، محقق باید ملاکی را

برای تعیین نقطه میانی فاصله طبقاتی در نظر بگیرد. نقاط میانی فاصله طبقاتی باید حتی الامکان به میانگین واقعی موارد قرار گرفته در هر طبقه نزدیک باشد. معمولاً در تحقیقات، محدوده طبقات ضربی از ۲، ۵، ۱۰، ۲۰، ۵۰ و ۱۰۰ انتخاب می‌گردد. محدوده طبقاتی ممکن است به یکی از شیوه‌های ذیل مشخص شود.

الف) فاصله طبقاتی غیرمشمول<sup>(۱)</sup>: این نوع طبقه بندی به صورت ذیل انجام می‌گیرد.

۱۰-۲۰

۲۰-۳۰

۳۰-۴۰

۴۰-۵۰

این فاصله طبقاتی به صورت ذیل خوانده می‌شود.

۱۰ تا کمتر از ۲۰

۲۰ تا کمتر از ۳۰

۳۰ تا کمتر از ۴۰

۴۰ تا کمتر از ۵۰

بنابراین در این نوع طبقه بندی، داده‌هایی که با حد بالای دامنه طبقاتی برابر هستند در طبقه بعدی جای می‌گیرند. به عنوان مثال موردی که مقدار آن دقیقاً ۳۰ می‌باشد در طبقه ۳۰-۴۰ قرار می‌گیرد نه در طبقه ۲۰-۳۰. بنابراین تحت قاعده طبقاتی غیرمشمول، محدوده بالای هر طبقه از طبقه فوق مستثناء شده و مشمول نمی‌گردد.

ب) فاصله طبقاتی شامل شده<sup>(۲)</sup>: در این نوع طبقه بندی محدوده طبقات به صورت ذیل

تعیین می‌شود.

۱۱-۲۰

1- Exclusive Type Class Intervals

2- Inclusive Type Class Intervals

۲۱-۳۰

۳۱-۴۰

۴۱-۵۰

براساس این قاعده مقدار حد بالای فاصله طبقاتی نیز در همان طبقه جای می‌گیرد. بنابراین اگر موردی دارای مقدار ۲۰ باشد در همان طبقه ۱۱-۲۰ طبقه بندی می‌شود. اگرچه حد بالای فاصله طبقاتی، طبقه ۱۱-۲۰ عدد ۲۰ می‌باشد، اما حد بالای واقعی ۲۰/۹۹۹۹۹ می‌باشد. براساس این منطق فاصله طبقاتی ۱۱-۲۰ به معنای ۱۱ تا کمتر از ۲۱ می‌باشد.

زمانیکه مقادیر موضوع مورد مطالعه به صورت اعداد صحیح (بدون اعشار) باشد بهتر است از قاعده فاصله طبقاتی مشمول یا شامل شده استفاده کرد. اما اگر مقادیر پدیده فوق به صورت ممتد (دارای اعشار) باشد، بهتر است از روش فاصله طبقاتی غیرمشمول استفاده شود.

ب) در مواقعی نیز ممکن است کلیه گویه‌های بکار گرفته شده در مقیاسها هم جهت نباشند بطوریکه برخی از گویه‌ها به صورت مثبت و برخی دیگر منفی باشند. در چنین مواقعی باید از طریق کدبندی مجدد به هم جهت سازی متغیرها و گویه‌ها اقدام کرد. نحوه هم جهت سازی در جدول (۲-۲) ارائه شده است.

جدول (۲-۲) کدگذاری مجدد برای هم جهت سازی گویه‌ها در زمینه سنجش گرایش روستائیان به

#### استفاده از کودشیمیایی

گویه یا معرفیها	کامل مخالف	مخالف	بیطرف	موافق	کامل موافق
۱- استفاده از کود موجب افزایش محصول می‌شود	۱	۲	۳	۴	۵
۲- استفاده از کود باعث آلودگی محصول و بروز بیماری می‌شود	۱	۲	۳	۴	۵
۳- کودشیمیایی کیفیت خاک را بهبود می‌بخشد	۱	۲	۳	۴	۵

همانطوریکه از جدول فوق برمی‌آید گویه دوم با دو گویه دیگر هم جهت نمی‌باشد. بنابراین کدهای اختصاص یافته به گویه فوق باید در جهت عکس مجدداً کدبندی شوند، تا بتوان از طریق جمع کردن کدها به سنجش میزان گرایش روستائیان نسبت به استفاده از کودشیمیایی پرداخت. (ج) در برخی مواقع ممکن است تعداد طبقات جداول دوبعدی به قدری زیاد باشد که امکان انجام عملیات آماری با نمونه‌های کوچک‌تر وجود نداشته باشد. به عنوان مثال برای محاسبه کای اسکویر برای جداول دوبعدی نباید بیش از ۲۰ درصد فراوانیهای مورد انتظار هر خانه جدول کمتر از ۵ باشد. بنابراین برای تحقق این شرط در صورتی که تعداد نمونه محدود باشد، ممکن است محقق به کدبندی مجدد به منظور تقلیل تعداد طبقات متغیرها اقدام کند. به عنوان مثال ممکن است محققى بخواهد رابطه بین میزان درآمد و رضایت شغلی افراد را در قالب یک جدول دوبعدی بشرح ذیل مورد تجزیه و تحلیل قرار دهد. همانطوریکه از جدول (۲-۳) برمی‌آید افراد نمونه که تعداد آنها ۱۵۰ نفر می‌باشند در یک جدول ۳×۵ که دارای ۱۵ خانه می‌باشد توزیع شده‌اند.

جدول (۲-۳) رابطه بین میزان درآمد و رضایت شغلی

رضایت شغلی	درآمد (هزار ریال)			جمع
	(۱) کمتر از ۵۰۰	(۲) ۵۰۱-۱۰۰۰	(۳) ۱۰۰۱-۱۵۰۰	
(۱) بسیار کم	۱۸	۸	۳	۲۹
(۲) کم	۵	۸	۵	۱۸
(۳) متوسط	۱۰	۹	۱۱	۳۰
(۴) زیاد	۵	۱۷	۱۸	۴۰
(۵) بسیار زیاد	۳	۱۱	۱۹	۳۳
جمع	۴۱	۵۳	۵۶	۱۵۰

در چنین مواقعی نیز ممکن است محقق در صدد ادغام برخی از طبقات جدول برآید. برای

این کار می‌توان از طریق کدبندی مجدد طبقات بسیار کم و کم و همچنین طبقات زیاد و بسیار زیاد را با یکدیگر ادغام نمود و رضایت شغلی را با سه سطح کم، متوسط و زیاد مورد تجزیه و تحلیل قرار داد. در صورت ادغام طبقات و کدبندی مجدد جدول (۲-۴) بدست می‌آید.

جدول (۲-۴) رابطه بین میزان درآمد و رضایت شغلی

رضایت شغلی	درآمد (هزار ریال)			جمع
	(۱) کمتر از ۵۰۰	(۲) ۵۰۱-۱۰۰۰	(۳) ۱۰۰۱-۱۵۰۰	
(۱) کم	۲۳	۱۶	۸	۴۷
(۲) متوسط	۱۰	۹	۱۱	۳۰
(۴) زیاد	۸	۲۸	۳۷	۷۳
جمع	۴۱	۵۳	۵۶	۱۵۰

همانطوریکه از جدول فوق برمی‌آید افراد نمونه بجای ۱۵ خانه در ۹ خانه جدول توزیع شده‌اند. این جدول امکان محاسبه آماره‌هایی نظیر کای اسکویر و تحلیل منطقی تر روابط بین متغیرها را فراهم می‌کند.

البته باید توجه کرد که ادغام طبقات متغیرها باید براساس یک منطق مناسب انجام گیرد. در غیر این صورت ممکن است گمراه کننده باشد و به تحریف نتایج منجر شود.

#### کدبندی مجدد بوسیله نرم افزار SPSS :

روش کدبندی مجدد قادر است کد اطلاعات قبلی که وارد فایل شده‌اند را تغییر دهد. به عنوان مثال اگر اطلاعات شما به صورت پنج سطحی بشرح ذیل در صفحه Data Editor وارد شده است.

(۱) خیلی کم (۲) کم (۳) متوسط (۴) زیاد (۵) خیلی زیاد

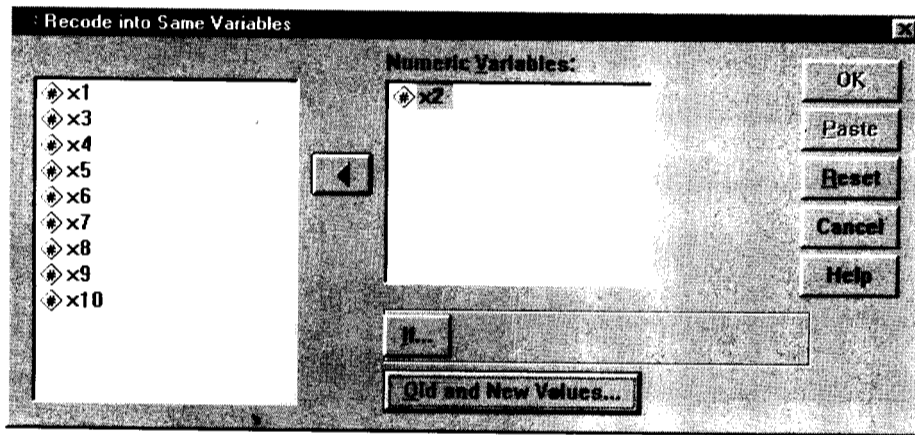
و اگر بخواهید در کدگذاری خود تجدیدنظر کرده و مثلاً کدهای ۱ و ۲ را با همدیگر و کدهای ۴

و ۵ را نیز با هم جمع کنید و در سه مقوله بشرح ذیل کدگذاری کنید،



(۱) کم (۲) متوسط (۳) زیاد

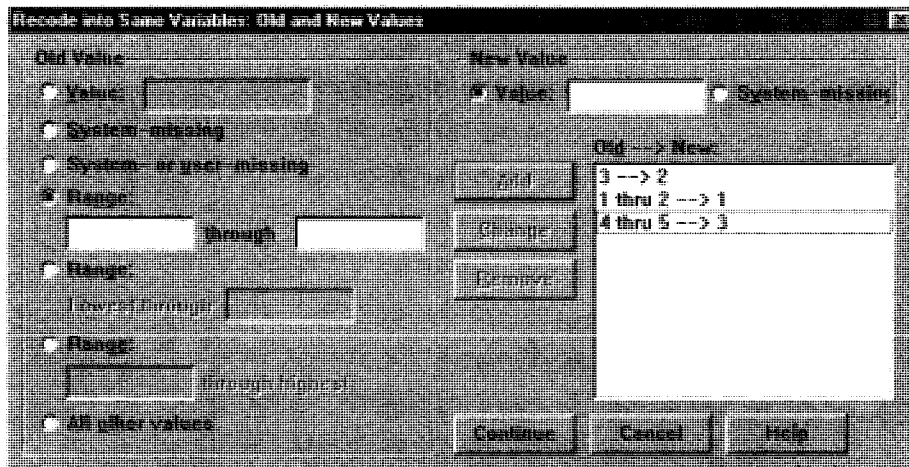
در این صورت می‌توانید از دستور Recode استفاده کنید. اجرای دستور کدبندی مجدد هم از طریق منوی Transform/Recode/Into Same Variables و هم از طریق منوی Transform/Recode/Into Different Variables امکان‌پذیر است. اگر شما دستور اول را اجرا کنید، داده‌های Recode شده جدید جای داده‌های قبلی را در Data Editor اشغال می‌کنند. اما اگر دستور دوم را انتخاب کنید، اطلاعات Recode شده متغیر جدیدی را در صفحه گستر می‌سازد که باید نام جدیدی برای این متغیر تعیین کنید. توصیه می‌شود از دستور دوم استفاده کنید. با اجرایی این دستور پنجره‌ای به صورت شکل (۲-۲) در مانیتور ظاهر می‌شود.



شکل (۲-۲) پنجره Recode into Same Variables

در این پنجره متغیری که باید Recode شود را مشخص کرده و وارد قسمت Numeric Variables کنید سپس دکمه Old and New Values را فشار دهید تا شکل (۳-۲) ظاهر شود.

بعد کدهای Old Value و New Values را مشخص کنید. به عنوان مثال اگر می‌خواهید که مقدار ۱ و ۲ به صورت کد ۱ تعریف شود در این صورت باید ابتدا Range را در قسمت Old Value انتخاب کنید و بعد عدد ۱ را در مستطیل اول و عدد ۲ را در مستطیل دوم نوشته و سپس در قسمت New Values عدد ۱ را در مستطیل مقابل Value وارد کنید و پس از آن دکمه Add را فشار دهید. سپس عدد ۳ را در قسمت Old Values در مقابل مستطیل Value وارد کرده و در قسمت New Value نیز عدد ۲ را در مقابل مستطیل Value تایپ و دکمه Add را فشار دهید. پس از آن در قسمت Old Value اعداد ۴ و ۵ را به ترتیب در مستطیل‌های مقابل Range وارد کرده و در قسمت New Value نیز عدد ۳ را تایپ و دکمه Add را فشار دهید.



شکل (۲-۳) پنجره Old and New Values

حتی شما می‌توانید تنها داده‌های بعضی از Case‌ها را Recode کنید. به عنوان مثال اگر بخواهید تنها مقادیر مربوط به مردان (Sex=1) را کدبندی مجدد کنید باید دکمه If... را از پنجره Recode in to same Variable (شکل ۲-۲) فشار دهید تا پنجره مورد نظر باز شود. (نحوه کار با

این پنجره بعداً توضیح داده خواهد شد.)

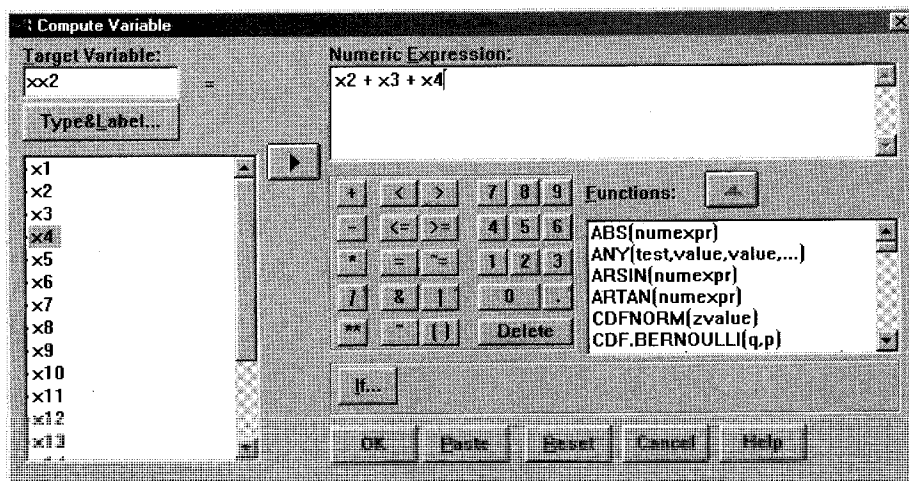
دستور Recode/Into Different Variables، درست مانند Recode/Into Same Variables عمل می‌کند. بجز اینکه دستور اولی نام جدیدی را برای متغیر جدید می‌طلبد. البته از نظر اینکه مطمئن باشید که اطلاعات قبلی از بین نمی‌رود روش Recode/Into Different Variables مناسب‌تر بنظر می‌رسد. روش اجرای آن نیز مانند Recode/Into Same Variables است. برای اینکار متغیری که باید Recode شود را مشخص کرده سپس وارد جعبه Numeric Variable → Output بکنید. در اینجا باید نامی برای متغیر Recode شده تعیین کرده و آن را در قسمت Output Variable Name تایپ کنید سپس دکمه Change را فشار دهید. با این دستور نام جدید در Numeric Variable → Output، در کنار نام قدیم قرار می‌گیرد. بعد از آن مقادیری که باید Recode شوند را به شیوه‌ای که قبلاً توضیح داده شد تعریف کنید.

وقتی مشخص کردید که مقادیر قبلی چگونه Recode شوند باید تصمیم بگیرید که با مقادیر مشخص نشده برای Recode کردن چگونه باید رفتار کرد. اگر در این مورد کاری انجام ندهید، رایانه به همه آنها System Missing Value خواهد داد و این چیزی است که نباید اتفاق بیافتد. برای اینکه مطمئن باشید که مقادیر قبلی که Recode نمی‌شوند، به همان صورت به متغیر جدید منتقل شوند در این صورت گزینه All Others Values را در پنجره Old and New Values انتخاب کنید و سپس مورد Copy Old Values را انتخاب کرده و دکمه Add را فشار دهید. در این صورت ELSE → Copy ظاهر می‌شود تا نشان دهد که کلیه مقادیری که کدبندی مجدد نشده‌اند به همان صورت به متغیر جدید منتقل خواهند شد. پس از اجرای دستور، نتیجه تحت عنوان یک متغیر جدید در صفحه گستر ظاهر می‌شود.

حساب کردن<sup>(۱)</sup>:

اغلب اتفاق می‌افتد که شما می‌خواهید مقادیر جدیدی را از روی داده‌های موجود در چند متغیر محاسبه کنید. برای این کار می‌توانید از دستور Compute استفاده کنید. این دستور در منوی Transform قرار دارد. برای اجرای آن منوی Transform/Compute را انتخاب کنید تا پنجره‌ای تحت عنوان Compute Variable در مانیتور ظاهر شود شکل (۴-۲).

نامی را برای متغیری که می‌خواهید محاسبه کنید تعیین کرده و در قسمت Target Variable تایپ کنید. شما می‌توانید یک Label نیز برای متغیر جدید تعیین کنید. برای اینکار دکمه Type & Label را فشار دهید و بر چسب مورد نظر را در آن وارد کنید.



شکل (۴-۲) پنجره Compute Variable

به عنوان مثال اگر می‌خواهید مجموع مقادیر متغیرهای  $X_2$ ,  $X_3$ ,  $X_4$  را بدست آورید برای اینکار فرمول را در قسمتی که تحت عنوان Numeric Expression قرار دارد وارد کنید. برای انتقال

1- Compute

متغیرها از فهرست، می‌توانید از دکمه فلش‌دار استفاده کنید. علامتهای چهار عمل اصلی و سایر علائم در قسمت پایین که شبیه به دکمه‌های ماشین حساب می‌باشند، قرار دارند که می‌توانید در نوشتن فرمولها از آنها استفاده کنید.

بعنوان مثال فرمول برای محاسبه مجموع مقادیر متغیرها عبارت خواهد بود از  $X_2 + X_3 + X_4$  که باید در قسمت Numeric Expression وارد شود. پس از وارد کردن فرمول دکمه OK را فشار دهید تا محاسبه انجام گیرد.

علاوه بر کلیدهای چهار عمل اصلی، دکمه‌های دیگر نیز نظیر > (به معنای بزرگتر)، < (به معنای کوچک‌تر) علامت & (به معنای و) I (به معنای یا) و علامت «-» (به معنای نه) وجود دارد. پس علامت = - به معنای مساوی نیست با می‌باشد.

علاوه بر اینها فهرستی در زیر عنوان Functions وجود دارد که امکان استفاده از عملیات مختلف را در موقع نوشتن فرمول برای شما فراهم می‌کند. به عنوان مثال اگر می‌خواهید لگاریتم مقادیر موجود در متغیر X را بدست آورید، ابتدا نام جدیدی را برای متغیر فوق تعیین کنید (مانند Log X). سپس فهرست Function را بچرخانید تا به گزینه LG10 برسید. آن را انتخاب کرده و سپس دکمه فلش‌دار بالای صفحه فهرست Function را فشار دهید تا گزینه فوق به قسمت Numeric Expression منتقل شود. این گزینه همراه با علامت سؤال مشخص می‌شود که نشان می‌دهد شما باید متغیر مورد نظر را در آنجا وارد کنید تا محاسبات مورد نظر (نظیر Log، میانگین، جذر و یا هر چیزی که از آن خواسته باشید) انجام گیرد. بنابراین روی علامت سؤال کلیک کرده سپس متغیر مورد نظر را از قسمت Variable List به آنجا منتقل کرده و سپس دکمه OK را فشار دهید، تا محاسبات تحت عنوان Logx در قسمت Data Editor ظاهر شود. فهرست Function این اجازه را به شما می‌دهد که گزینه‌های آماری دیگر را نیز انتخاب کنید. بعنوان مثال برای مجذور کردن مقادیر می‌توانید از SQRT استفاده کنید.

در اینجا گزینه‌های آماری و محاسباتی دیگری نیز وجود دارند که با استفاده از آنها می‌توان میانگین، مجموع، انحراف معیار، واریانس، مقادیر حداقل یا حداکثر را برای یک مجموعه از مقادیر

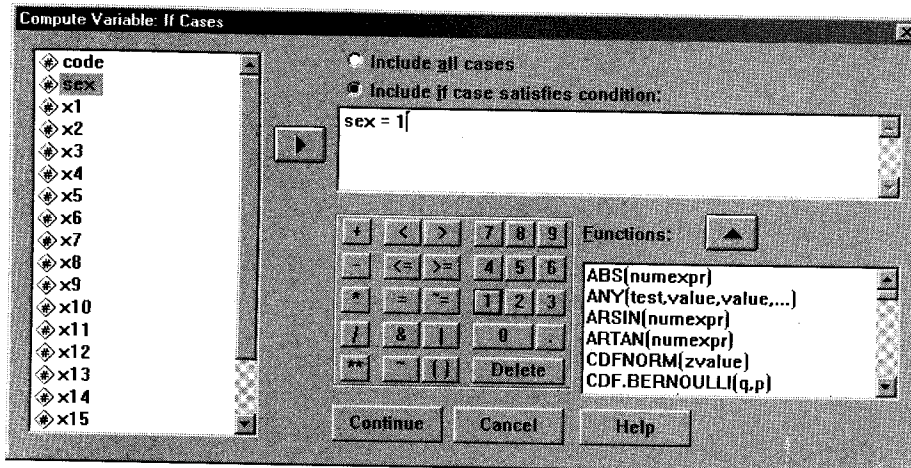
بدست آورد. به عنوان مثال می‌توان میانگین را برای  $X_3, X_2, X_1$  بدست آورد. برای اینکار نام متغیر جدید را در قسمت Target Variable بنویسید. سپس از فهرست Function گزینه MEAN (Numexpr, Numexpr...) را انتخاب کنید تا در پنجره Numeric Expression ظاهر شود. سپس متغیرهای مورد نظر را که می‌خواهید برای آنها میانگین محاسبه کنید انتخاب کرده و از طریق دکمه فلش‌دار آن را به قسمت Numeric Expression انتقال دهید. در این مثال متغیرهای  $X_3, X_2, X_1$  را انتخاب کنید. تا فرمول به صورت  $MEAN(X_3, X_2, X_1)$  ظاهر شود. سپس دکمه Ok را کلیک کنید تا میانگین متغیرهای فوق محاسبه و در ستون در نظر گرفته شده برای متغیر جدید ثبت شود.

فرمول جمع کردن مقادیر متغیرهای  $X_3, X_2, X_1$  که قبلاً توضیح داده شد می‌تواند با استفاده از گزینه Function به صورت  $Sum(X_3, X_2, X_1)$  نیز نوشته و محاسبه شود. برای گرد کردن ارقام نیز می‌توانید از گزینه RND و برای حذف اعشار می‌توانید از TRUNC استفاده کنید.

#### نحوه انجام محاسبات با شرط :

ممکن است بخواهید که صرفاً روی بعضی از موارد محاسبات خاصی انجام دهید. برای این کار شرایطی را برای متغیر فوق تعیین می‌کنید. به عنوان مثال ممکن است بخواهید تنها مواردی از  $X_3, X_2, X_1$  را با هم جمع کنید که جنسیت آنها زن باشد (یعنی  $Sex=1$  باشد). برای این کار دکمه If... را در پنجره مکالمه Compute Variable فشار دهید تا جعبه دیگری تحت عنوان Compute Variable: IF Cases باز شود شکل (۲-۵).

ساختار این جعبه شبیه به جعبه Compute Variable می‌باشد. اما در اینجا می‌توانید شرایط خود را قبل از اینکه دستور Compute اجراء شود اعمال کنید. به عنوان مثال برای جمع کردن موارد مربوط به زنان شما نیاز به Compute Variable: if Cases دارید تا شرط  $Sex=1$  را در آن وارد کنید. اجرای این مرحله نیز شبیه عمل Numeric Expression است. برای اینکار ابتدا دکمه رادیویی Included if case satisfies condition را روشن کنید و سپس متغیر مورد نظر را مشخص



شکل (۵-۲) پنجره Compute Variable : If Cases

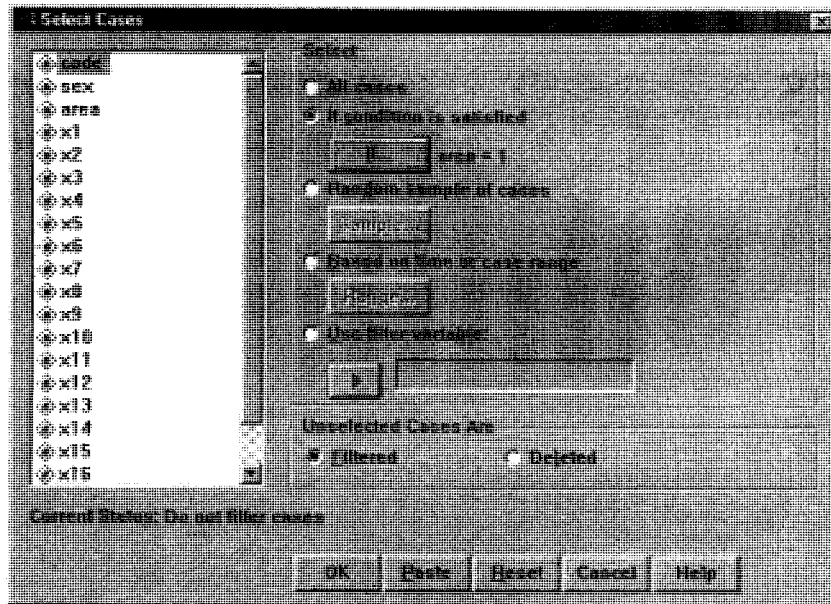
کرده و با کمک دکمه فلش دار به جعبه سمت راست منتقل کنید. سپس با استفاده از دکمه‌های ماشین حساب یا کلیدهای Functions شرط مورد نظر را وارد کنید. سپس دکمه Continue را کلیک کنید. در این مثال شرط  $Sex=1$  وارد شده است (شکل ۵-۲).

### نحوه انتخاب زیرگروه‌ها برای تجزیه و تحلیل:

#### الف) انتخاب زیرگروه دائمی:

ممکن است بخواهید زیرمجموعه‌ای از موارد را انتخاب کرده و تجزیه و تحلیل کنید. برای این کار منوی Data/Select Cases را انتخاب کنید. این دستور پنجره‌ای را باز می‌کند که در آن مسیرهای مختلفی وجود دارد شکل (۶-۲).

به عنوان مثال ممکن است بخواهید که میزان فروش مردان و زنان را با استفاده از آزمون  $t$  با هم دیگر در منطقه شمال کشور ( $Area=1$ ) مقایسه کنید. برای لحاظ کردن پاسخگویان

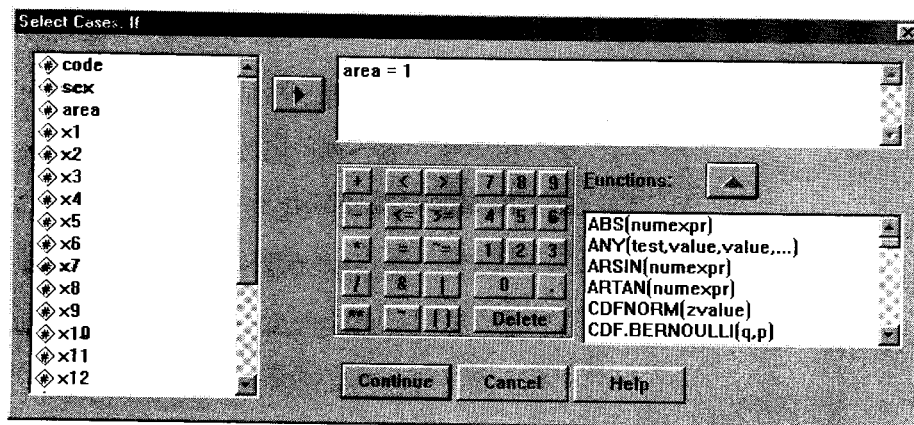


شکل (۶-۲) پنجره Select Cases

(یا فروشندگان منطقه شمالی) ابتدا دکمه رادیویی *If Condition is satisfied* را روشن کرده و سپس دکمه *IF...* را فشار دهید. تا پنجره‌ای تحت عنوان *Select Cases: IF* مانند شکل (۶-۲) باز شود.

در اینجا متغیر یا *Function* خاصی را مشخص کنید تا براساس آن موارد مورد نظر برای تجزیه و تحلیل انتخاب شوند. بنابراین در مثال اخیر باید  $area=1$  در فرمول لحاظ شود تا تنها افراد از منطقه شمال در تحلیل وارد شوند. سپس دکمه *Continue* و بعد دکمه *OK* را فشار دهید. توجه داشته باشید که دستور *Select Cases*، برای تحلیل‌های بعدی نیز از همین موارد استفاده می‌کند. به عنوان مثال: اگر ما صرفاً زنان را از طریق *Select Cases/If (Sex=2)* انتخاب کنیم و بعد روی آنها عملیات آماری انجام دهیم، بدنبال آن نمی‌توانیم دستور *Select Cases/If (Area=2)* را انتخاب کنیم. زیرا این دستور تمام مردان و زنان منطقه را شامل نمی‌شود. به عبارت دیگر اگر ما ابتدا دستور *Select Cases/If (Sex=2)* را انتخاب کنیم و بعد دستور





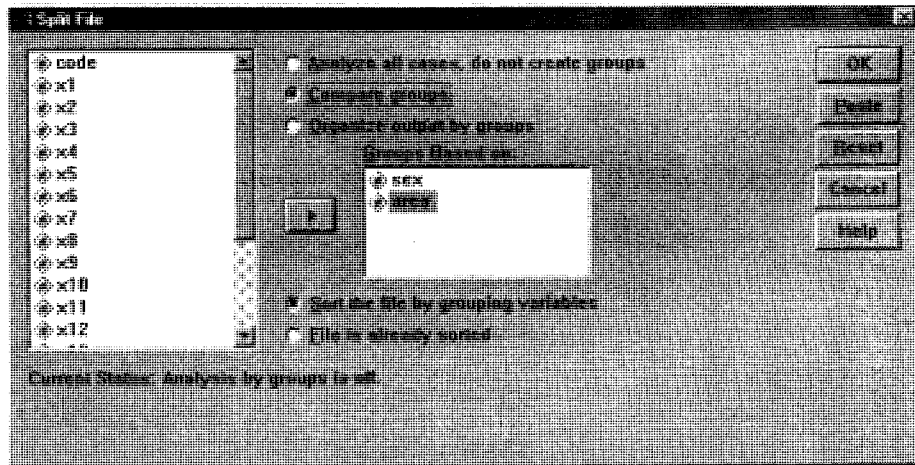
شکل (۷-۲) پنجره Select Cases:If

Select Cases/If (Area=2) را اجرا کنیم، گروه انتخابی ما تنها زنان منطقه ۲ خواهند بود. در حالیکه این موضوع مورد نظر ما نبوده است و ما می‌خواستیم ابتدا فقط زنان را انتخاب کنیم و تحلیل آماری انجام دهیم و بعد در مرحله بعدی می‌خواستیم منطقه جنوب (۲) را انتخاب کرده و تحلیل آماری را روی زنان و مردان آن منطقه انجام دهیم. برای رفع این مشکل باید از طریق اجرای دستور Data/Select Case مورد Select All Cases را فعال کنید.

#### ب) انتخاب زیرگروه موقتی :

اگر بخواهید تجزیه و تحلیل خود را بر روی یک سری از زیرگروهها انجام دهید باید از دستور Split File استفاده کنید. برای این کار منوی Data/Split File را فشار دهید تا شکل (۸-۲) ظاهر شود. در جعبه‌ای که باز می‌شود دکمه Compare Groups را انتخاب کنید. بعد متغیرهایی که باید استفاده شوند تا موارد را به زیرگروهها تقسیم کنند را مشخص کرده وارد Groups Based بکنید. در

مثال بالا متغیرهای Area و Sex باید وارد این قسمت شوند.



شکل (۲-۸) پنجره Split File

در این دستور چند نکته مهم وجود دارد. وقتی که دستور Split File اجرا می‌شود، این دستور تا زمانی که SPSS در حال کار است ادامه می‌یابد (از این نظر شبیه Select Cases عمل می‌کند). اما می‌توان آن را غیر فعال کرد. برای این کار در جعبه Split File دکمه Analyze all Cases را انتخاب کنید. بنابراین برای تحلیل آماری مردان و بعد زنان بطور جداگانه و بعد انجام تحلیل آماری بر روی هر دو گروه، شما باید از Split File برای تحلیل دو گروه بطور جداگانه استفاده کرده و بعد آن را بطوریکه در بالا گفته شد غیر فعال کنید تا بتوانید تحلیل آماری خود را بر روی هر دو گروه با هم انجام دهید.

وقتی که دستور اجراء می‌شود Split File اطلاعات را در فایل اطلاعاتی مرتب می‌کند. این کار به صورت اتوماتیک انجام می‌گیرد. اما اگر قبلاً اطلاعات به شیوه خاصی مرتب شده باشد می‌توانید با انتخاب دکمه File is already sorted مانع از مرتب کردن اتوماتیک شوید.

**نحوه انتخاب نمونه تصادفی (Selecting Random Sample) :**

اگر داده‌های بسیار وسیعی در فایل اطلاعاتی داشته باشید می‌توانید مواردی را به عنوان نمونه انتخاب کرده و مورد تجزیه و تحلیل قرار دهید. برای این کار منوی Data/Select Case را انتخاب کنید (شکل ۲-۶). سپس در پنجره Select Case، مورد Random Sample of Cases را انتخاب کنید و دکمه Sample را فشار دهید. در اینجا جعبه جدید دیگری باز می‌شود که می‌توانید درصد یا تعداد دقیق نمونه مورد نظران را تعیین نمایید.

**۶- آزمون برازش داده‌ها:**

آزمون برازش داده‌ها و تعیین میزان اعتماد یا پایایی<sup>(۱)</sup> و اعتبار یا روائی<sup>(۲)</sup> در تحقیقات اجتماعی و اقتصادی از اهمیت زیادی برخوردار است. زیرا اندازه‌گیری مطلوب آن است که نتایج حاصل از آن اولاً دقیق و صحیح باشد، یعنی بیانگر تفاوت‌های حقیقی در واقعیت یا پدیده مورد اندازه‌گیری باشد و ثانیاً مبین همان صفت یا صفاتی باشد که مورد نظر محقق بوده است، و درست همان چیزی باشد که در تعریف مسئله آمده است.

**۱- اعتماد (پایایی)**

در اینجا سعی می‌گردد مشخص شود که آیا نتایج حاصل از اندازه‌گیری توسط یک ابزار مشخص قابل اعتماد است یا خیر؟ یعنی ابزار اندازه‌گیری ما (پرسشنامه) با چه دقت یا صحتی موضوع مورد بررسی را قابل اندازه‌گیری کرده است و اگر اندازه‌گیری مربوطه تحت همان شرایط تکرار شود آیا همان نتایج قبلی بدست می‌آید یا خیر؟

بنابراین همانطوریکه مشخص است اعتماد یا پایایی مسئله‌ای کمی و تکنیکی است و بیشتر ناظر به این سوال است که ابزار اندازه‌گیری با چه دقت و صحتی پدیده یا صفت مورد نظر را

اندازه‌گیری می‌کند<sup>(۱)</sup>.

در تحقیقات اجتماعی محقق به کمک یک سری از روشها حدود اعتماد یا پایائی ابزارهای اندازه‌گیری را مشخص می‌کند. مهمترین این روشها عبارتند از: ۱- روش بازآزمایی<sup>(۲)</sup>، ۲- روش موازی یا آزمون هم‌تا<sup>(۳)</sup>، ۳- روش دو نیمه کردن<sup>(۴)</sup>، ۴- روش ضریب ثبات<sup>(۵)</sup>، ۵- روش کودر - ریچاردسون<sup>(۶)</sup> و ۶- روش آلفای کرونباخ<sup>(۷)</sup>.

بسیاری از این روشها قابل محاسبه با نرم‌افزار SPSS می‌باشند اما به دلیل اینکه استفاده از ضریب آلفای کرونباخ از عمومیت بیشتری برخوردار است لذا به تشریح نحوه محاسبه آن پرداخته می‌شود. اصولاً این روش برای محاسبه انسجام درونی ابزار اندازه‌گیری یا مقیاسها بکار می‌رود. زمانیکه مقیاسی از نوع لیکرت وجود داشته باشد و محقق بخواهد از طریق گویه‌های مختلف یک مفهوم پیچیده را اندازه‌گیری کند برای سنجش انسجام درونی مقیاس می‌تواند از آماره آلفای کرونباخ استفاده کند. دقت کنید که نباید سایر سوالات ناسازگار با گویه‌های این مقیاس وارد تحلیل شود. (این نکته از این نظر تاکید می‌شود که مانع از اشتباه عمومی ورود کل سوالات پرسشنامه در تحلیل نهایی شود).

این آماره برای محاسبه انسجام درونی پرسشنامه‌ها یا ابزارهایی که با مقیاس رتبه‌ای تنظیم شده‌اند مورد استفاده قرار می‌گیرد. برای محاسبه این آزمون از طریق برنامه SPSS ابتدا دستور Analyze/Scale/Reliability را اجراء کنید تا شکل (۲-۹) ظاهر شود. در این پنجره متغیرهای مورد نظر را به قسمت Items منتقل کنید. سپس از قسمت Method، آماره Alpha را انتخاب کنید (البته این آزمون به صورت پیش‌گزیده در SPSS بکار رفته است) پس از آن با فشار دادن دکمه

۱- عبداللهی، محمد. جزوه درسی روش تحقیق در علوم اجتماعی، دانشکده علوم اجتماعی دانشگاه علامه طباطبائی.

2- Test - Retest Method

3- Parallel or Equivalent Method

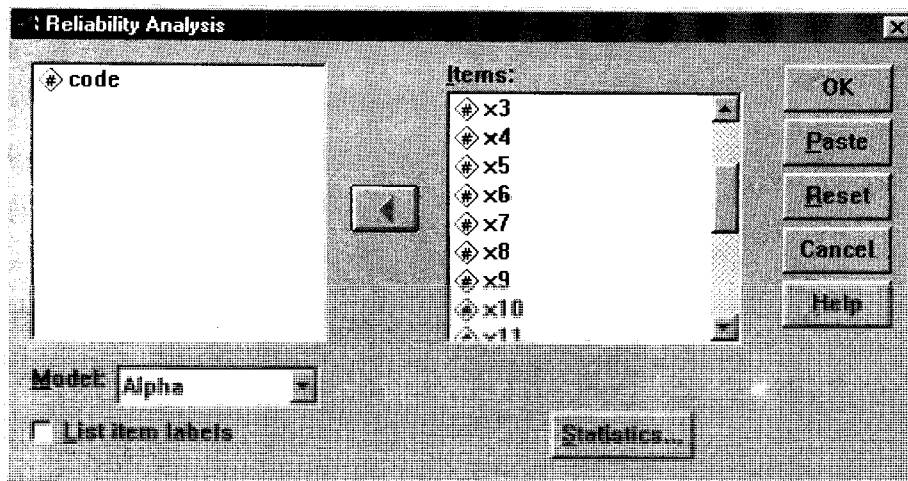
4- Split - Half Method

5- Stability coefficient Method

6- Kuder - Richardson Method

7- Cronbach's Alpha

از پنجره فوق مورد Scale if item deleted را انتخاب کرده و دکمه Continue را فشار داده و سپس دکمه Ok را کلیک کنید. مقدار Alpha در پنجره Output ظاهر می‌شود. اگر مقدار بدست آمده بیشتر از  $0/70$  باشد پایایی ابزار اندازه‌گیری قابل قبول تلقی می‌گردد. اما علاوه بر مقدار آلفای کرونباخ جدولی نیز محاسبه می‌گردد که در ستون آخر آن نشان می‌دهد که با حذف هر یک از گویه‌ها چه تغییری در مقدار آلفا ایجاد می‌گردد. از طریق بررسی مقادیر این ستون می‌توان تصمیم گرفت که در صورت حذف کدام گویه یا گویه‌ها مقدار آلفا افزایش می‌یابد. اینگونه گویه‌ها، مواردی هستند که با سایر گویه‌ها سازگار نبوده و لذا با حذف آنها دقت اندازه‌گیری مقیاس افزایش می‌یابد. اگر مقدار بدست آمده کمتر از  $0/70$  باشد باید عمل حذف گویه‌های با آلفای پایین تا جایی ادامه یابد که مقدار آلفای نهایی به  $0/7$  یا بیشتر از آن برسد.



شکل (۲-۹) پنجره Reliability Analysis

نتایج محاسبات برای یک مقیاس ۱۶ گویه‌ای در جدول (۲-۵) ارائه شده است. همانطوریکه از جدول فوق برمی‌آید مقدار آلفای بدست آمده برابر با  $0/768$  می‌باشد. اما در صورتی

که گویه  $X_5$  حذف گردد مقدار آلفا به  $0/77$  نیز قابل افزایش است. البته چنین کاری در این مثال ضرورت ندارد زیرا آلفای بدست آمده قابل قبول می‌باشد. در اینجا از نماد  $X_i$  برای نشان دادن گویه‌ها استفاده شده است.

جدول (۵-۲) تغییر در مقدار آلفا در صورت حذف برخی از متغیرها

	Scale Mean Deleted	Scale Variance Deleted	Corrected Item Correlation	Alpha if Item Deleted
X2	47.666	104.092	.220	.769
X3	47.333	97.747	.485	.746
X4	47.600	101.558	.333	.759
X5	47.766	106.047	.187	.770
X6	48.000	96.482	.493	.745
X7	47.500	96.327	.438	.749
X8	47.366	104.723	.259	.769
X9	47.733	102.271	.346	.758
X10	47.466	98.395	.424	.751
X11	47.400	99.006	.448	.749
X12	47.900	104.369	.267	.764
X13	47.333	96.436	.562	.740
X14	47.833	100.971	.363	.756
X15	48.133	101.912	.298	.762
X16	47.266	101.926	.263	.766
X17	47.700	100.217	.414	.752

Alpha =  $0/768$ 

۲- اعتبار (روایی):

بر خلاف اعتماد یا پایایی، که عمدتاً یک مسئله کمی و قابل اندازه‌گیری است، اعتبار یا روایی مسئله‌ای عمدتاً کیفی بوده و ارزیابی آن بسیار مشکل می‌باشد. در اینجا چگونگی حرکت محقق از تعریف نظری به تعریف عملی مورد توجه قرار می‌گیرد. برای سنجش اعتبار و روایی ابزار اندازه‌گیری

Reliability (۱)

Validity (۲)

روشهایی وجود دارد که مهمترین آنها عبارتست است: ۱- اعتبار محتوا<sup>(۱)</sup>، ۲- اعتبار سازه‌ای<sup>(۲)</sup> و ۳- اعتبار عاملی<sup>(۳)</sup>. اعتبار محتوا، روشی برای سنجش میزان اعتبار اجزای تشکیل دهنده یک ابزار اندازه‌گیری می‌باشد که معمولاً توسط افراد متخصص در موضوع مورد مطالعه انجام می‌گیرد. از این رو اعتبار محتوا به قضاوت داوران بستگی دارد. اعتبار سازه‌ای نیز نشان می‌دهد که ابزار اندازه‌گیری تا چه حد یک سازه یا خصیصه را که مبنای نظری دارد می‌سنجد، در اینجا رابطه متغیر وابسته با هر یک از معرفهای مورد نظر در مفهوم کلی بررسی می‌شود. اعتبار عاملی نیز نوعی اعتبار سازه‌ای است که از طریق تحلیل عاملی<sup>(۴)</sup> انجام می‌گیرد<sup>(۵)</sup>. روش تحلیل عاملی در فصل ۱۴ به تفصیل بحث شده است.

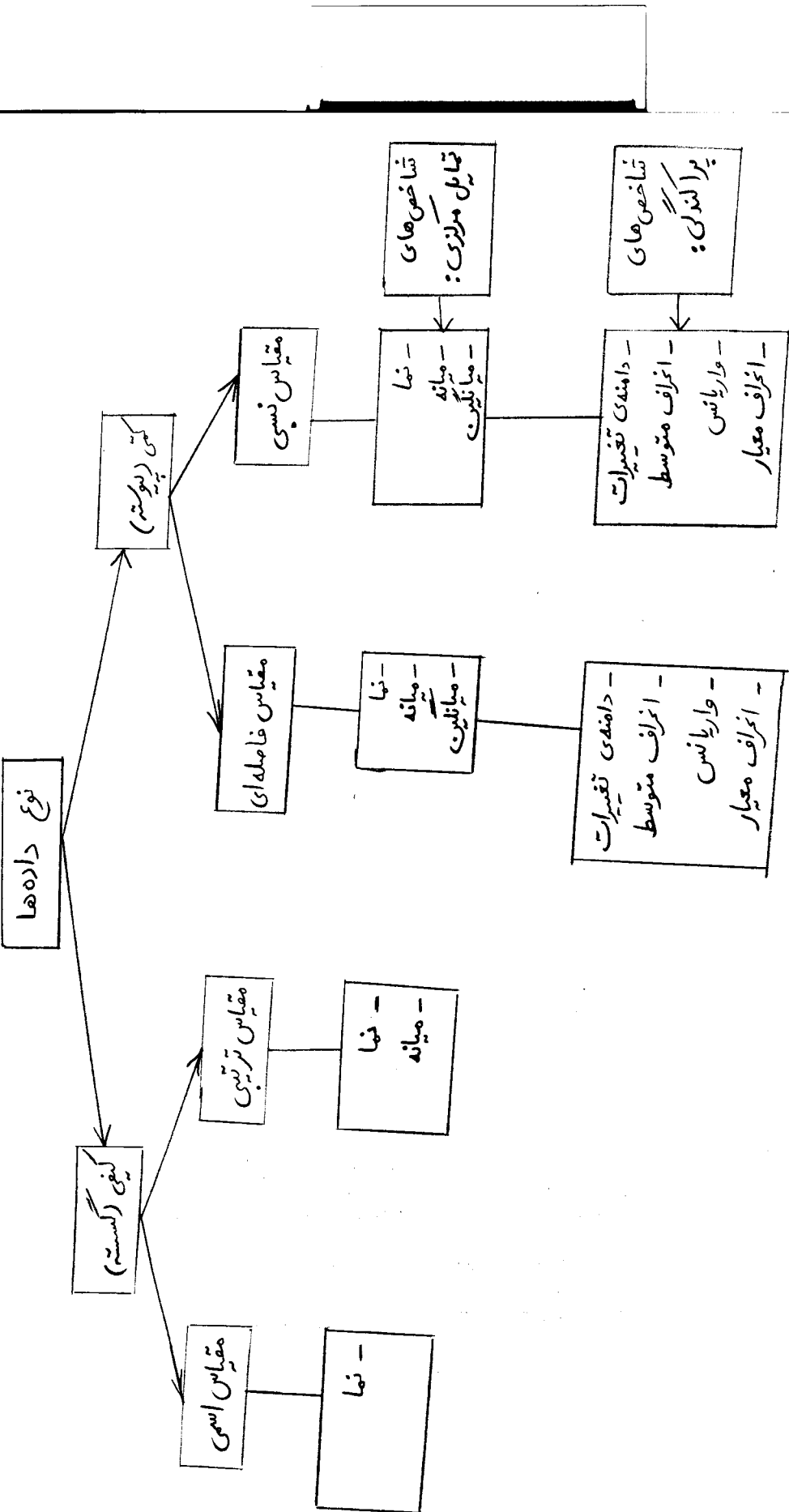
1- Content Validity

2- Construct validity

3- Factor Validity

4- Factor Analysis

۵- عبداللهی، محمد. «جزوه درسی روشهای تحقیق در علوم اجتماعی» دانشکده علوم اجتماعی دانشگاه علامه طباطبائی.



جدول نوع داده‌ها، سطح اندازه‌گیری و آمار توصیفی



## فصل سوم

# توصیف داده‌ها

### مقدمه:

پردازش داده‌ها در پژوهش‌های اجتماعی - اقتصادی معمولاً در دو سطح انجام می‌گیرد این دو سطح عبارتند از: ۱ - توصیف داده‌ها و ۲ - تجزیه تحلیل و تبیین داده‌ها. در مرحله توصیف داده‌ها جامعه مورد مطالعه با توجه به متغیرهای مورد نظر توصیف شده و تصویری از وضعیت موجود ارائه می‌گردد. برای تبیین این وضعیت از آمار توصیفی که طیف نسبتاً گسترده‌ای را تشکیل می‌دهد استفاده می‌گردد. در بخش آمار توصیفی محقق درصدد اثبات یا رد فرضیه خاصی نمی‌باشد. زیرا در این سطح از پردازش، در هر مرحله تنها یک متغیر مورد بررسی قرار می‌گیرد و روابط بین متغیرها مدنظر نمی‌باشد. مهمترین روش‌های آمار توصیفی و موارد کاربرد آنها در مطالعات اجتماعی - اقتصادی در این فصل به تفصیل مورد بررسی قرار گرفته و روش محاسبه آنها با استفاده از نرم‌افزار SPSS نیز توضیح داده شده است.

**جداول توزیع فراوانی:**

جداول توزیع فراوانی<sup>(۱)</sup> تعداد دفعات سطوح یا مقادیر یک متغیر را نشان می‌دهد. اینگونه جداول معمولاً برای نشان دادن فراوانی متغیرهای اسمی و ترتیبی بکار می‌روند. برای اینگونه جداول می‌توان درصد و درصد تجمعی را نیز محاسبه کرد تا تصویر روشنتری از متغیر مورد بررسی ارائه گردد. آنچه که باید در تهیه جداول توزیع فراوانی، مدنظر قرار گیرد، این است که برای متغیرهای اسمی و ترتیبی محاسبه شاخصهایی مانند میانگین و میانه امکانپذیر نمی‌باشد. در چنین مواردی می‌توان از نما یا مد به عنوان یک شاخص مرکزی استفاده نمود. نما یا مد در واقع سطح یا مقداری از متغیر است که در آن بیشترین فراوانی وجود دارد. در مواردی نیز ممکن است محقق با متغیر دو نمایی مواجه گردد، و این زمانی اتفاق می‌افتد که تعداد فراوانی دو سطح از یک متغیر با هم برابر باشند.

**شاخص‌های مرکزی:**

شاخصهای مرکزی<sup>(۲)</sup> مقادیری هستند که مرکزیت داده‌ها را مشخص می‌کنند. این مقادیر در واقع عدد کلیدی و قابل جانشین برای کل داده‌ها بحساب می‌آیند. مهمترین شاخص‌های مرکزی، میانگین<sup>(۳)</sup>، میانه<sup>(۴)</sup>، و نما<sup>(۵)</sup> می‌باشند. میانگین حسابی<sup>(۶)</sup> یکی از انواع میانگین‌هاست که بیشترین کاربرد را دارد و از طریق تقسیم مجموع مقادیر به تعداد کل آنها بدست می‌آید و این شاخص برای متغیرهای فاصله‌ای و نسبی قابل محاسبه می‌باشد که معادله آن به صورت زیر است:

$$\mu_x = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

1- Frequency Distribution Tables

2- Central Tendency

3- Mean

4- Median

5- Mode

6- Arithmetic Mean

در این معادله:

$$\mu_x = \text{عبارت تست از میانگین مقادیر متغیر } x$$

$$\Sigma = \text{علامت مجموع}$$

$$X_i = \text{مقادیر متغیر } X$$

$$n = \text{تعداد موارد}$$

اگر مقادیر یک متغیر تکرار شوند بجای محاسبه میانگین ساده که فرمول آن بیان شد، بهتر است از فرمول میانگین وزنی<sup>(۱)</sup> استفاده شوند که فرمول آن به صورت زیر است.

$$\mu_w = \frac{\sum W_i X_i}{\sum W_i}$$

در این معادله:

$$\mu_w = \text{میانگین وزنی}$$

$$W_i = f_i = \text{فراوانی مربوط به مقدار } X_i$$

$$X_i = \text{مقدار متغیر برای فرد یا مورد } i \text{ ام}$$

میانگین از ساده‌ترین و مهمترین شاخص‌های مرکزی است، بطوریکه محققان با استفاده از آن تعداد وسیعی از موارد یا متغیرها را با همدیگر مقایسه می‌کنند. اما علیرغم مفید و پرکاربرد بودن آن، این شاخص دارای نقاط ضعفی نیز می‌باشد که یکی از اساسی‌ترین آنها این است که به شدت تحت تأثیر مقادیر دور افتاده می‌باشد. بنابراین چنانچه مقادیر دور افتاده زیاد باشند و یا اختلاف آنها با میانگین بیشتر باشد، استفاده از میانگین به عنوان شاخص مرکزی صحیح نمی‌باشد. زیرا ممکن است محقق را به نتیجه‌گیری ناروا وادار کند.

سه شاخص مهم برای تعیین پراکندگی مقادیر یک متغیر عبارتند از: دامنه تغییرات، واریانس و انحراف معیار. با توجه به اینکه در محاسبه دامنه تغییرات فقط بزرگترین مقدار و کوچکترین مقدار

یک متغیر مورد استفاده قرار می‌گیرد و به سایر مقادیر توجه نمی‌شود، بنابراین تحت تأثیر دو مقدار اول و آخر قرار می‌گیرد و به همین دلیل اغلب از واریانس و انحراف معیار که در محاسبه آنها کلیه مقادیر دخالت دارند استفاده می‌شود. در توزیع نرمال انحراف معیار به عنوان ابزاری بسیار مفید به دلایل ذیل کاربرد دارد.

۱- بیش از ۹۹ درصد از مقادیر یک متغیر در فاصله سه انحراف معیار از میانگین قرار دارند.

$$(\bar{X} \pm 3sd = \%99)$$

۲- بیش از ۹۵ درصد از مقادیر یک متغیر در فاصله دو انحراف معیار از میانگین قرار دارند.

$$(\bar{X} \pm 2sd = \%95)$$

۳- بیش از ۶۸ درصد از مقادیر یک متغیر در فاصله یک انحراف معیار از میانگین قرار دارند.

$$(\bar{X} \pm 1sd = \%68)$$

بنابراین اگر داده‌های مورد بررسی با مقیاس فاصله‌ای یا نسبی اندازه‌گیری شده باشند در اغلب موارد از میانگین به عنوان یک شاخص مرکزی و از واریانس و انحراف معیار به عنوان شاخص‌های پراکندگی می‌توان استفاده کرد.

#### میانه:

میانه مقداری از متغیر است که نیمی از مقادیر کمتر از آن و نیمی دیگر بیشتر از آن می‌باشند.

برای محاسبه میانه به روش زیر عمل می‌کنیم.

۱- داده‌ها را به صورت نزولی و یا صعودی مرتب می‌کنیم.

۲- محل میانه را با استفاده از فرمول  $C_m = \frac{n+1}{2}$  مشخص می‌کنیم.

۳- مقداری از متغیر را که در محل محاسبه شده قرار می‌گیرد به عنوان میانه در نظر می‌گیریم.

میانه در واقع متوسط مکانی است که معمولاً برای پدیده‌های کمی محاسبه می‌گردد. اگر

تعداد داده‌های دور از میانگین زیاد باشد و یا اختلاف داده‌های دور از میانگین با میانگین بیشتر باشد،

در این صورت میانگین شاخص مرکزی مناسبی نخواهد بود. در چنین مواقعی بهتر است از میانه به

عنوان شاخص مرکزی استفاده شود. اگر جداول توزیع فراوانی دارای درصد تجمعی باشد در این صورت می‌توان به آسانی نسبت به تعیین میانه اقدام کرد. در ستون درصد تجمعی نقطه ۵۰ درصدی به عنوان میانه انتخاب می‌شود.

چارکها، دهکها و صدکها نیز برای تعیین و یا تقسیم بندی مقادیر یک متغیر استفاده می‌شوند. درست همان گونه که میانه کل مقادیر مشاهدات را به دو قسمت مساوی تقسیم می‌کند، چارکها نیز آن را به چهار قسمت مساوی و دهکها آنها را به ده قسمت مساوی و صدکها آن را به صد قسمت مساوی تقسیم می‌کنند. صدک برای داده‌های کلان، مانند نمره‌های مربوط به آزمونهای ورودی دانشگاهی و غیره مفید هستند<sup>(۱)</sup>.

#### نما یا مد :

مد<sup>(۲)</sup> سطحی از یک متغیر یا مقداری است که بیشترین فراوانی را دارد. مد نیز مانند میانه از جمله شاخص‌های مرکزی است که تحت تأثیر مقادیر دور افتاده قرار نمی‌گیرد. بنابراین در صورتی که بخواهیم تأثیر مقادیر دور افتاده (مقادیر حداکثر و یا حداقل) را از بین ببریم استفاده از این شاخص مطلوب بنظر می‌رسد. مد برای دستیابی به اندازه یا سطحی از متغیر که بیشترین فراوانی را دارد، مفید می‌باشد. به عنوان مثال اگر یک کارخانه تولیدکننده کفش بخواهد شماره کفشی که بیشترین مصرف کننده یا متقاضی را دارد بدست آورد، تعیین مد بهترین شیوه برای دستیابی به این هدف است. اما علیرغم این ویژگیها، شاخص مد نیز دارای نقاط ضعف خاصی است. مثلاً اگر بخواهیم اهمیت نسبی سطوح یا مقادیر یک متغیر را بررسی کنیم این شاخص از کاربرد لازم برخوردار نمی‌باشد. از سوی دیگر اگر در سطوح یا مقادیر یک متغیر، فراوانی مربوط به چند طبقه یا چند مقدار با همدیگر برابر

۱- سکاران اوما (۱۳۸۰)، روشهای تحقیق در مدیریت، ترجمه محمد صائبی و محمود شیرازی، انتشارات مرکز آموزش مدیریت

دولتی، ص ۳۴۸.

باشند، این شاخص مناسب نمی‌باشد.

علیرغم محدودیتهایی که ممکن است بر این شاخص مترتب باشد، شاخص فوق برای متغیرهای اسمی که محاسبه میانه و میانگین امکانپذیر نمی‌باشد، کاربرد زیادی دارد. برای تعیین اینکه مد بدست آمده تا چه حد می‌تواند نماینده گروه محسوب شود، می‌توان از آماره نسبت تغییرات<sup>(۱)</sup> استفاده کرد. نسبت تغییرات در واقع بیانگر درصد افرادی است که در طبقه مد قرار ندارند. هرچه این درصد بزرگتر باشد مد کمتر می‌تواند نماینده کل توزیع باشد.

#### اندازه‌گیری پراکندگی<sup>(۲)</sup>:

شاخص‌های مرکزی مانند میانگین یا میانه در بهترین حالت تنها می‌توانند وضعیت یک مجموعه را در قالب یک عدد که تا حدودی نماینده‌ای از آن مجموعه است نمایش دهند. اما این شاخصها قادر به بیان چگونگی توزیع مقادیر متغیر در حول میانگین نمی‌باشند. به عبارت دیگر نحوه پراکندگی مقادیر یک متغیر با استفاده از شاخص‌های مرکزی امکانپذیر نمی‌باشد. برای دستیابی به این هدف باید به محاسبه شاخص‌های پراکندگی اقدام کرد. مهمترین شاخص‌های موجود در این زمینه عبارتند از: دامنه<sup>(۳)</sup>، انحراف متوسط از میانگین<sup>(۴)</sup>، انحراف معیار<sup>(۵)</sup> و واریانس<sup>(۶)</sup>.

#### ۱- دامنه:

دامنه ساده‌ترین شاخص برای بیان میزان پراکندگی مقادیر یک متغیر است که از طریق تفاوت بین مقادیر حداکثر و حداقل بدست می‌آید. این شاخص را از طریق فرمول ذیل می‌توان محاسبه نمود.

1- Variation Ratio

2- Measure of Dispersion

3- Range

4- Mean Deviation

5- Standard Deviation

6- Variance

کمترین مقدار همان متغیر - بیشترین مقدار یک متغیر  $R =$

با توجه به اینکه دامنه تغییرات به آسانی محاسبه می‌شود، بنابراین وضعیت پراکندگی مقادیر متغیر را بطور سریع برای محقق روشن می‌کند. اما از آنجا که در محاسبه آن فقط بزرگترین و کوچکترین مقدار متغیر دخالت دارند، لذا از اهمیت چندانی برخوردار نمی‌باشد.

### ۲- پراکندگی دامنه دهک:

زمانیکه اکثر مقادیر در جدول توزیع فراوانی در طبقات نزدیک به طبقه میانه قرار گرفته باشند میانه می‌تواند نماینده خوبی برای کل توزیع باشد. اما اگر مقادیر فوق در طبقاتی قرار گیرند که فاصله زیادی با طبقه میانه داشته باشند، شاخص میانه نماینده مناسبی برای بیان کل جامعه یا متغیر نخواهد بود. بنابراین چنانچه دامنه<sup>(۱)</sup> توزیع از گستردگی بیشتری برخوردار باشد کارایی میانه کمتر خواهد شد. این مسئله بویژه زمانی اتفاق می‌افتد که مقادیر دورافتاده در ابتدا یا انتهای توزیع وجود داشته باشد. برای حذف یا به حداقل رساندن اثرات موردها یا مقادیر دورافتاده می‌توان ۱۰ درصد پایین (دهک اول) و ۱۰ درصد بالایی (دهک دهم) موردها یا مقادیر را حذف کرد، و با احتساب ۸۰ درصد موردها که بین دهک اول و دهک آخر قرار دارند با حذف تأثیرات اغراق‌آمیز معدودی از موارد اکثریت اعضای نمونه را بطور منطقی‌تر مورد تجزیه و تحلیل قرار داد<sup>(۲)</sup>.

### ۳- ضریب انحراف متوسط<sup>(۳)</sup>:

عبارتست از متوسط تفاوت مقادیر یک متغیر از یکی از مشخص‌کننده‌های مرکزی همان متغیر. در محاسبه این آماره علائم منفی لحاظ نمی‌گردد و قدر مطلق مقادیر مدنظر قرار می‌گیرد. این

1- Range

۲- دواس، دی، ای (۱۳۷۶) "پیمایش در تحقیقات اجتماعی" ترجمه هوشنگ ناییبی، نشر نی، ص ۱۴۸-۱۴۶.

3- Coefficient of Mean Deviation

شاخص را می‌توان از طریق فرمولهای ذیل محاسبه کرد.

$$\begin{aligned} \text{انحراف متوسط از میانگین } (\delta \bar{X}) &= \frac{\sum |X_i - \bar{X}|}{n} \\ \text{انحراف متوسط از میانه } (\delta m) &= \frac{\sum |X_i - md|}{n} \\ \text{انحراف متوسط از مد } (\delta mo) &= \frac{\sum |X_i - mo|}{n} \end{aligned}$$

در این فرمول :

$\delta$  = علامت انحراف متوسط است که به صورت دلتا تلفظ می‌شود

$X_i$  = مقدار  $i$ ام مربوط به متغیر  $X$

$n$  = تعداد یا مقادیر یک متغیر

$\bar{X}$  = میانگین

$md$  = میانه

$mo$  = مد

اگر مقادیر انحراف متوسط یک متغیر را به میانگین همان متغیر تقسیم کنیم ضریب انحراف متوسط بدست می‌آید. این شاخص بیانگر پراکندگی نسبی مقادیر بوده و از طریق آن می‌توان پراکندگی متغیرهای مختلف را با همدیگر مقایسه کرد. این شاخص برخلاف دامنه که تنها براساس مقادیر حداکثر و حداقل بدست می‌آید، کلیه مقادیر یک متغیر را شامل می‌گردد.

#### ۴ - انحراف معیار :

انحراف معیار از شاخص‌های مهم برای سنجش پراکندگی مقادیر یک متغیر است. با توجه به اینکه در محاسبه آن کلیه مقادیر یک متغیر نقش دارند و پراکندگی مقادیر را حول میانگین محاسبه می‌کند و میانگین از شاخص‌های مرکزی مهم می‌باشد بنابراین انحراف معیار نیز از اهمیت زیادی برخوردار است.

انحراف معیار از طریق فرمول ذیل بدست می‌آید.



$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

در صورتی که مقادیر متغیر دارای فراوانی باشند. فرمول به شرح ذیل خواهد بود:

$$\sigma = \sqrt{\frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i}}$$

اگر مقدار انحراف معیار یک متغیر را بر میانگین آن تقسیم کنیم ضریب تغییرات<sup>(۱)</sup> که یک شاخص نسبی است بدست می‌آید. این شاخص امکان مقایسه متغیرهای مختلف را فراهم می‌کند. هرچه مقدار ضریب تغییرات بیشتر باشد از پراکندگی بیشتر مقادیر یک متغیر حکایت می‌کند و هر قدر مقدار بدست آمده کوچک‌تر باشد متمرکز بودن مقادیر متغیر را حول میانگین نشان می‌دهد. اگر مقدار انحراف معیار به توان ۲ رسانده شود واریانس، بدست می‌آید.

در برخی مواقع ممکن است میانگین دو جامعه با همدیگر برابر باشند اما دارای انحراف معیار متفاوتی باشند. در این صورت میانگین شاخص مناسبی برای مقایسه دو جامعه نخواهد بود. ولی با محاسبه ضریب تغییرات برای هر یک از جامعه‌ها می‌توان توزیع مقادیر دو جامعه را با یکدیگر مقایسه کرد. جامعه با ضریب تغییرات بالاتر دارای پراکندگی بیشتر و جامعه با ضریب تغییرات پائین‌تر دارای پراکندگی کمتر می‌باشد.

### سنجش عدم تقارن:

مهمترین شاخص‌های اندازه‌گیری عدم تقارن<sup>(۲)</sup> متغیرها عبارتند از:

۱ - چولگی<sup>(۳)</sup>: اگر توزیع آیت‌های یک متغیر کاملاً به صورت متقارن باشند نمودار توزیع آن به صورت شکل (۱-۳) خواهد بود. اینگونه نمودارها به منحنی نرمال<sup>(۴)</sup> معروفند که بیانگر توزیع نرمال مقادیر یک صفت می‌باشد. زمانیکه توزیع یک صفت در جامعه مورد مطالعه کاملاً نرمال باشد، در

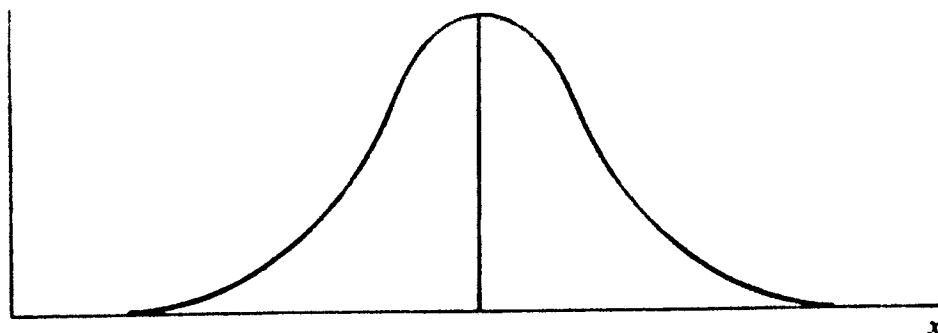
1- Coefficient of Variation

2- Asymmetry Measurement

3- Skewness

4- Normal Curve

این صورت مقادیر میانگین، میانه و مد با هم برابر خواهند بود. بنابراین، چنانچه این سه شاخص مرکزی در یک جامعه با هم برابر باشند می‌توان مدعی شد که جامعه فوق در خصوص صفت مورد نظر یک جامعه نرمال می‌باشد. اما چنانچه نمودار منحنی بسمت راست یا چپ خمیدگی داشته باشد منحنی فوق از چولگی برخوردار خواهد بود. اگر خمیدگی منحنی به سمت راست باشد چولگی مثبت و چنانچه خمیدگی به سمت چپ باشد چولگی منفی خواهد بود.



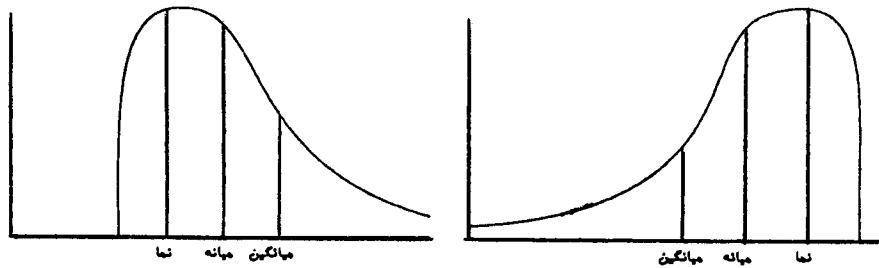
$$(\bar{X} = M = Z)$$

شکل (۱-۳) منحنی نرمال

همانطوریکه از اشکال (۱-۳) و (۲-۳) نیز برمی‌آید تفاوت‌های موجود بین میانگین، میانه و مد راحت‌ترین راه برای تشخیص چولگی در مقادیر یک صفت است. زمانی که مقادیر یک صفت دارای میانگین ( $\bar{X}$ )، میانه ( $M$ ) و مد ( $Z$ ) برابر باشند توزیع نرمال محسوب می‌شود. هنگامیکه  $Z < M < \bar{X}$  باشد، چولگی مثبت و چنانچه  $\bar{X} < M < Z$  باشد چولگی منفی خواهد بود.

۲- درجه کشیدگی<sup>(۱)</sup> نمودار:

این آماره برای اندازه‌گیری میزان پخی منحنی نرمال بکار می‌رود، چنانچه توزیع موردها در مرکز به صورت نرمال باشد منحنی پراکندگی به صورت منحنی نرمال ظاهر می‌شود. اما اگر پراکندگی به سمت اطراف میل کند منحنی به صورت پخ و اگر تراکم بیشتر در مرکز وجود داشته باشد منحنی نرمال از اوج بیشتری برخوردار خواهد شد.

چولگی مثبت  $Z < M < \bar{X}$ چولگی منفی  $\bar{X} < M < Z$ 

شکل (۲-۳) منحنی‌های دارای چولگی

## آمار توصیفی و سطوح اندازه‌گیری:

سطوح اندازه‌گیری عمده‌ترین معیار در استفاده از آزمونهای آماری است. آمار توصیفی نیز از

1- Kurtosis

این قاعده مستثنی نبوده و متناسب با طبیعت داده‌ها باید از آمار توصیفی سازگار با آن استفاده کرد. عدم توجه به سطوح اندازه‌گیری در بکارگیری روشهای مختلف آماری، اعم از آمار توصیفی یا استنباطی اشتباه عمومی است که در بسیاری از تحقیقات به چشم می‌خورد. به منظور انتخاب مناسب‌ترین روش آماری در توصیف داده‌ها اصول استفاده از آمار توصیفی بر مبنای مقیاسهای مختلف اندازه‌گیری در جدول (۳-۱) ارائه شده است.

### سطوح اندازه‌گیری:

بر خلاف علوم طبیعی فرآیند اندازه‌گیری در علوم اجتماعی و انسانی کار بسیار پیچیده‌ای بوده و علیرغم پیشرفتهای عمده‌ای که در این حوزه انجام گرفته است، هنوز سنجش و اندازه‌گیری در این بخش از دقت لازم برخوردار نمی‌باشد. از آنجا که کاربرد روشهای آماری مختلف در تحقیقات اجتماعی - اقتصادی تحت تأثیر نوع و طبیعت داده‌های موجود می‌باشد، بنابراین آشنایی با سطوح اندازه‌گیری، محقق را در انتخاب روشهای آماری مناسب برای توصیف و تحلیل داده‌ها کمک خواهد کرد. مفاهیم و متغیرهای مورد استفاده در علوم اجتماعی و انسانی با چهار مقیاس سنجیده می‌شوند که از نظر میزان دقت به سطوح اسمی<sup>(۱)</sup>، ترتیبی<sup>(۲)</sup>، فاصله‌ای<sup>(۳)</sup> و نسبی<sup>(۴)</sup> تقسیم می‌گردند. این مقیاسها در شکل (۳-۳) ارائه شده‌اند.

#### ۱- سطح اسمی:

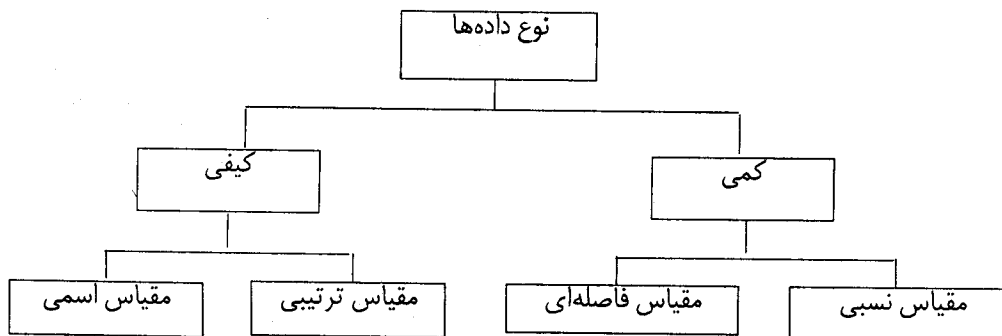
این سطح ابتدایی‌ترین سطح اندازه‌گیری است که در آن مقولات یک متغیر به صورت قراردادی در کنار یکدیگر قرار می‌گیرند تا چگونگی همسانی یا ناهمسانی آنها مشخص شود. مانند جنسیت: (۱) زن، (۲) مرد یا محل تولد: (۱) تهران، (۲) تبریز، (۳) شیراز، (۴) اصفهان.

1- Nominal Scale

2- Ordinal Scale

3- Interval Scale

4- Ratio Scale



شکل (۳-۳) نوع داده و مقیاس اندازه‌گیری

در این سطح از اندازه‌گیری فقط هر مقوله از مقوله دیگر متمایز می‌شود و هر فرد مورد بررسی فقط در یکی از این مقولات قرار می‌گیرد و اعدادی که به مقولات یک متغیر اختصاص داده می‌شود ارزش حقیقی عدد را ندارد. به عنوان مثال نمی‌توان گفت که مرد دو برابر زن است و یا شیراز سه برابر تهران است. زیرا ترتیب قرار گرفتن مقولات امری قراردادی است. در این سطح از اندازه‌گیری تنها می‌توان به محاسبه فراوانی، درصد و نما (یا مد) اکتفا کرد.

## ۲- سطح ترتیبی:

در سطح ترتیبی اندازه‌گیری، یک پله بالاتر می‌رود یعنی علاوه بر همسانی و ناهمسانی مقولات، ترتیب آنها هم مشخص می‌شود و این ترتیب هم واقعی و هم منطقی است. مثال، میزان علاقه به تحصیل: (۱) کم، (۲) متوسط، (۳) زیاد. یا طبقه اجتماعی: (۱) پایین، (۲) متوسط، (۳) بالا. در این سطح از اندازه‌گیری جهت تغییر از کم به زیاد یا از ضعیف به شدید است. بنابراین در این سطح، علاوه بر همسانی یا ناهمسانی مقولات، شدت و ضعف و ترتیب آنها نیز مشخص می‌شود، ولی چون فواصل بین مقولات مشخص نیست، اعداد اختصاص یافته به مقولات باز هم فاقد ارزش عددی است. یعنی نمی‌توانیم بگوییم که فلانی سه برابر دیگری به تحصیل علاقه دارد. بلکه تنها می‌توانیم او را از افراد کم علاقه متمایز کنیم و شدت علاقه او را مشخص کنیم. برای داده‌هایی که

در سطح ترتیبی جمع‌آوری شده‌اند تنها می‌توان به تعیین توزیع فراوانی، درصد، مد یا نما و ضریب همبستگی کندال اکتفا کرد و چنانچه داده‌ها به صورت رتبه‌ای باشند یعنی از سلسله اعداد طبیعی ۱، ۲، ۳، ... تبعیت کنند در این صورت می‌توان از آماره‌هایی نظیر میانه و ضریب همبستگی رتبه‌ای اسپیرمن نیز استفاده کرد.

### ۳- سطح فاصله‌ای:

در این سطح علاوه بر تعیین سطوح مختلف و ترتیب منطقی و واقعی بین مقادیر یک متغیر، فاصله بین مقادیر متغیر فوق نیز مشخص می‌شود. مانند: هزینه خانوار یا نمره دانشجویی. در این سطح می‌توانیم نمره یک دانشجوی را از دانشجوی دیگر متمایز کنیم. کمی یا زیادی آنرا مشخص کنیم و فاصله آنرا نیز از دیگر مقادیر متغیر تعیین کنیم. به عنوان مثال می‌توان گفت که دانشجویی که نمره ۲۰ دریافت کرده است، با دانشجویی که نمره ۱۰ گرفته است به اندازه ۱۰ نمره فاصله یا اختلاف دارد. اما در این مقیاس صفر مطلق وجود ندارد. به عنوان مثال اگر دانشجویی در درس آمار صفر گرفته باشد نباید فرض کرد که اصولاً فاقد دانش آمار است. برای داده‌های با مقیاس فاصله‌ای، شاخصهایی از قبیل میانگین، انحراف معیار، واریانس و ضریب همبستگی پیرسون، قابل محاسبه هستند و انجام آزمونهای T و F و غیره، نیز امکانپذیر است.

### ۴- سطح نسبی:

بالاترین و دقیق‌ترین سطح اندازه‌گیری است که علاوه بر تعیین سطوح و مقادیر یک متغیر و تعیین فاصله بین مقادیر یک متغیر، نسبت‌ها نیز قابل محاسبه و بیان هستند و مبنای اندازه‌گیری نیز صفر مطلق است. مانند: مقدار پول در حساب دانشجویی، یا میزان درآمد. در این متغیر اگر حساب دانشجوی صفر باشد، واقعاً هیچ پولی در حساب او نیست. زیرا این متغیر دارای صفر مطلق است و وجود صفر حقیقی به عنوان مبدا اندازه‌گیری محقق را قادر می‌سازد تا از روشهای مختلف آماری استفاده کند. برای داده‌های این مقیاس علاوه بر اینکه کلیه شاخص‌های ذکر شده در سطح فاصله‌ای قابل

محاسبه می‌باشند، بلکه شاخصهایی نظیر میانگین هندسی و ضریب پراکنش که به صفر حقیقی نیاز دارند نیز قابل محاسبه است.

در بحث اندازه‌گیری در تحقیقات اجتماعی و انسانی در مواقعی، با متغیرهایی سروکار داریم که نمی‌توان آنها را بالاتر از سطح اسمی اندازه‌گیری کرد، مانند محل تولد، مذهب، جنسیت و غیره. ولی اگر متغیری را در سطح نسبی بتوانیم اندازه‌گیری کنیم به نحو اولی می‌توان آنرا در سطح پایین‌تر نیز اندازه‌گیری کرد. مانند درآمد که هم می‌توان مبنا را صفر فرض کرد و هم می‌توان به صورت کم، متوسط و زیاد نیز مقوله‌بندی نمود<sup>(۱)</sup>. آمار توصیفی سازگار با هر یک از سطوح اندازه‌گیری در جدول (۱-۳) خلاصه شده است.

جدول (۱-۳) آمار توصیفی و سطح اندازه‌گیری

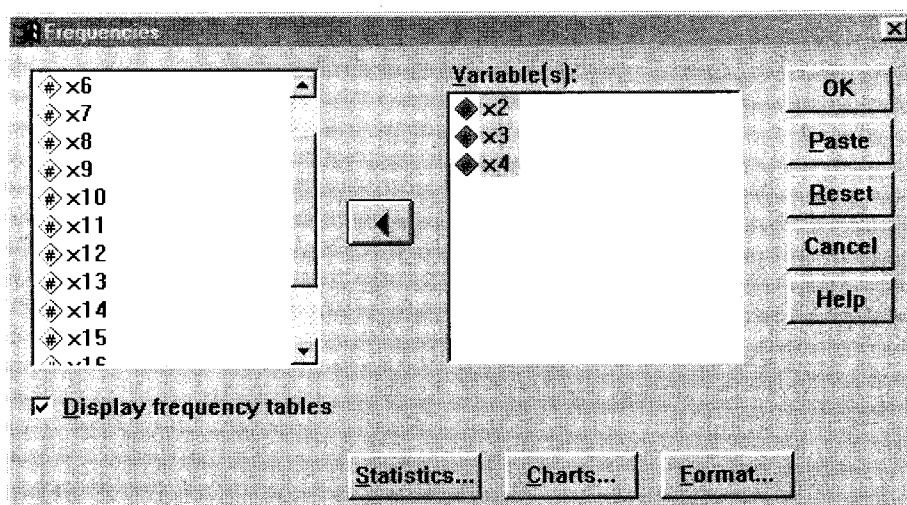
فاصله‌ای یا نسبی	ترتیبی	اسمی	سطح اندازه‌گیری آمار صیفی
مد، میانه، میانگین	مد، میانه	مد	شاخص‌های مرکزی
واریانس انحراف معیار دامنه تغییرات	چارک بندی	فراوانی نسبی طبقات	شاخص‌های پراکندگی
چولگی			قرینه (سطح توزیع مقارن)

نحوه محاسبه جداول توزیع فراوانی، شاخص‌های مرکزی و پراکندگی با SPSS:

برای محاسبه فراوانیها از طریق نرم‌افزار SPSS باید منوی Analyze/Descriptive Statistics/Frequencies... را انتخاب کنید تا پنجره‌ای مانند شکل (۳-۴) در مانیتور ظاهر شود.

۱- عبداللهی محمد «جزوه درسی روشهای تحقیق در علوم اجتماعی» دانشکده علوم اجتماعی دانشگاه علامه طباطبائی، تهران.

پنجره‌های مربوط به روشهای آماری دارای یک فرمت عمومی هستند که در آن در سمت چپ جعبه‌ای بدون عنوان وجود دارد که لیست متغیرها در آن ظاهر می‌شوند. این لیست به Source Variable List معروف است. از این پنجره متغیر مورد نظر را با استفاده از موس مشخص کرده و با کلیک کردن دکمه فلش دار آن را به جعبه Variable(s) منتقل کنید. اگر می‌خواهید بیش از یک متغیر را انتخاب کنید، روی اولین متغیر کلیک کرده و موس را روی متغیرها بکشید تا کلیه متغیرهای



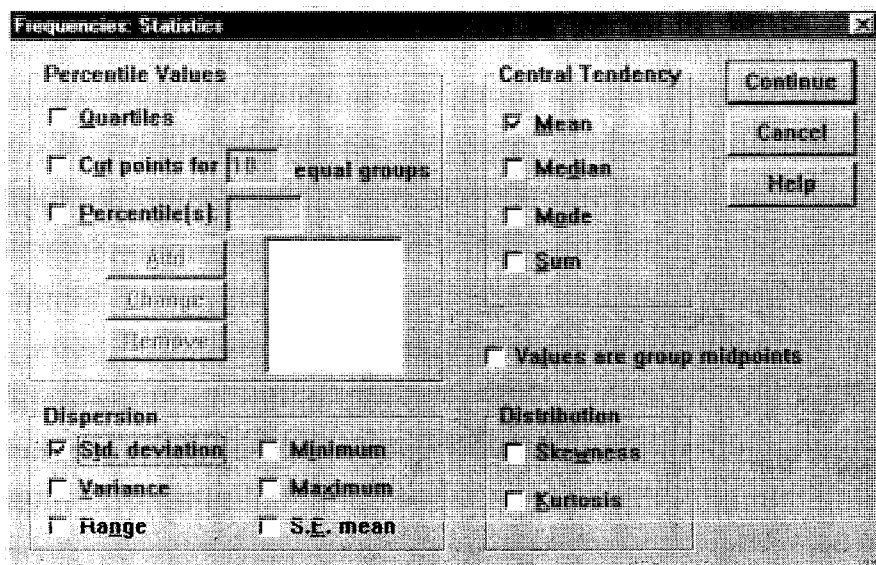
شکل (۳-۴) پنجره Frequency

مورد نظر تیره شوند، سپس دکمه فلش دار را فشار دهید تا کلیه متغیرها به قسمت Variable(s) منتقل شوند.

اگر اشتباهاً متغیری را از قسمت Source Variable List به جعبه سمت راست منتقل کردید و قصد تجزیه و تحلیل آن را ندارید، برای برگرداندن آن به جای اولیه خود، با استفاده از موس روی متغیر فوق کلیک کنید تا تیره شود. در این صورت دکمه فلش دار به سمت چپ تغییر جهت می‌دهد که با فشار دادن آن متغیر فوق از جعبه سمت راست به جعبه سمت چپ بازگردانده می‌شود.



برای محاسبه شاخص‌های مرکزی، پراکندگی و چولگی، در پنجره فراوانی، دکمه Statistics را فشار دهید تا پنجره مکالمه دیگری مانند شکل (۵-۳) در مانیتور ظاهر شود. عملیات آماری را که مورد نظر شماست باید از این پنجره انتخاب کنید. به عنوان مثال، برای محاسبه میانگین، کلمه Mean را در جایی که تحت عنوان Central Tendency قرار دارد انتخاب کنید و برای محاسبه انحراف معیار کلمه Std. deviation را در جایی که تحت عنوان Dispersion قرار دارد انتخاب کنید تا یک علامت  $\sqrt{\quad}$  در داخل مربع کوچک کنار کلمات فوق ظاهر



شکل (۵-۳) مشخص‌کننده‌های مرکزی و پراکندگی

شود. (اگر موردی را اشتبهاً انتخاب کردید، روی آن مجدداً کلیک کنید تا علامت داخل مربع از بین برود). پس از انتخاب آماره‌های مورد نظر دکمه Continue را فشار دهید تا به پنجره اصلی بازگردید. (اگر دکمه Cancel را فشار دهید تمام مواردی را که انتخاب کرده بودید از بین رفته و شما یک مرحله به عقب برمی‌گردید و به پنجره Frequencies می‌رسید.) در صورتی که بخواهید نمودارهای هیستوگرام و منحنی نرمال را نیز رسم کنید باید دکمه Charts را فشار داده و آنها را انتخاب کنید.

در صورت محاسبه جدول توزیع فراوانی از طریق نرم‌افزار SPSS، علاوه بر درصد و درصد تجمعی، ستون دیگری نیز تحت عنوان درصد معتبر<sup>(۱)</sup> محاسبه می‌گردد. زمانیکه داده‌های مورد استفاده، مقادیر گم شده نداشته باشند مقادیر ستون درصد معتبر با مقادیر ستون درصد برابر خواهند بود (جدول ۲-۳). اما چنانچه متغیر مورد بررسی دارای مقادیر گم شده باشد، در این صورت در ستون درصد معتبر، درصدهای مربوط به مقادیر گم شده لحاظ نمی‌گردد و درصدها صرفاً برای مقادیر موجود و معتبر محاسبه می‌شوند.

بخاطر داشته باشید که برای محاسبه میانگین از طریق زیرمنوی Analyze/ Compare Means / Means استفاده نکنید. زیرا زیرمنوی Means برای محاسبه میانگین زیرگروه‌های پاسخگویان مورد استفاده قرار می‌گیرد. به عنوان مثال اگر بخواهید میانگین را برای مردان و زنان بطور جداگانه محاسبه کنید می‌توانید از این دستور استفاده کنید.

جدول (۲-۳) توزیع فراوانی پاسخگویان بر اساس جنسیت

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	636	41.9	41.9	41.9
	Female	881	58.1	58.1	100.0
	Total	1517	100.0	100.0	

همانطوریکه از جدول (۲-۳) برمی‌آید از مجموع ۱۵۱۷ نفر در حدود ۴۲ درصد مرد و ۵۸ درصد زن هستند. با توجه به اینکه جدول فوق فاقد مقادیر گم شده می‌باشد، بنابراین مقادیر ستونهای درصد و درصد معتبر با هم برابر هستند.

1- Valid Percent

## فصل چهارم

# تحلیل روابط بین متغیرها

### مقدمه:

آمار توصیفی تنها می‌تواند تصویری از جامعه مورد مطالعه ارائه دهد، اما قادر به بیان روابط بین متغیرها و تبیین متغیر یا متغیرهای وابسته نیست. زیرا در این سطح محقق در هر مرحله تنها با یک متغیر سروکار دارد. نحوه توصیف داده‌ها و آماره‌های موجود و کاربرد هر یک از آن‌ها در فصل قبل مورد بررسی قرار گرفت. اما محقق در تحقیقات اجتماعی - اقتصادی عمدتاً بدنبال تحلیل و تبیین داده‌ها و متغیرهای مورد نظر می‌باشد. به همین دلیل سعی می‌کند با استفاده از آمار استنباطی به بررسی روابط بین متغیرها، تفاوت موجود در بین گروه‌های مورد مطالعه و یا تبیین متغیر وابسته از طریق متغیرهای مستقل و غیره بپردازد. در این سطح از تحلیل، محقق همواره با بیش از یک متغیر سروکار دارد که برای تحلیل و تبیین آنها از آزمونهای آماری مختلف که هر یک تحت شرایط خاصی به کار گرفته می‌شوند استفاده می‌گردد. اینگونه آزمونها در این فصل و فصول آتی مورد بررسی قرار می‌گیرند.

### تحلیل روابط بین متغیرها :

در اکثر مطالعات اجتماعی و انسانی محقق معمولاً با دو یا چند متغیر سروکار داشته و درصد بررسی روابط بین آنها می‌باشد. به عنوان مثال ممکن است محقق درصد رابطه بین نمرات

اخذ شده توسط دانشجویان با بهره هوشی، ساعات صرف شده برای مطالعه، جنسیت و غیره را مورد بررسی قرار دهد. در تحلیل اینگونه روابط دو سؤال اساسی مطرح می‌شود. این سؤالات عبارتند از:

۱- آیا ارتباطی بین دو یا چند متغیر وجود دارد؟ اگر جواب مثبت است این ارتباط در چه سطحی است؟

۲- آیا می‌توان تغییرات یک متغیر را از طریق متغیر یا متغیرهای دیگر پیش بینی و تبیین کرد؟ پاسخ سؤال اول از طریق تکنیک همبستگی و سؤال دوم از طریق روش تحلیل رگرسیون امکانپذیر است. هریک از تکنیک‌های فوق دارای روشهای متعددی هستند که عمده‌ترین روشهای همبستگی، در این قسمت مورد تجزیه و تحلیل قرار گرفته و تحلیل رگرسیون به فصل دیگری واگذار شده است.

در رابطه با تحلیل دو متغیره روابط بین متغیرها را می‌توان از طریق: (۱) جداول توافقی<sup>(۱)</sup>، (۲) ضریب همبستگی اسپیرمن<sup>(۲)</sup>، (۳) ضریب همبستگی پیرسون<sup>(۳)</sup> و غیره مورد بررسی قرار داد و برای پیشگویی تغییرات یک متغیر از طریق متغیر دیگر از رگرسیون ساده<sup>(۴)</sup> بهره جست.

در رابطه با جمعیت چند متغیره روابط بین متغیرها از طریق: (۱) ضریب همبستگی چندگانه<sup>(۵)</sup> و (۲) ضریب همبستگی جزئی<sup>(۶)</sup> و پیشگویی تغییرات یک متغیر از طریق متغیرهای دیگر با استفاده از رگرسیون چندگانه<sup>(۷)</sup> امکان پذیر است. در این بخش سعی می‌شود که هریک از روشهای همبستگی مورد بررسی قرار گیرند.

زمانیکه داده‌ها به صورت اسمی باشند جداول توافقی روش مناسبی برای بررسی روابط بین دو متغیر بحساب می‌آید. در این روش هر متغیر در دو یا چند طبقه تقسیم‌بندی شده و تعداد فراوانی

1- Cross Tabulation

2- Spearman's Coefficient of Correlation

3- Pearson's Coefficient of Correlation

4- Simple Regression

5- Coefficient of Multiple Correlation

6- Coefficient of Partial Correlation

7- Multiple Regression

مربوط به طبقات هر متغیر در جدول قرار می‌گیرند. از طریق اینگونه طبقه‌بندیها می‌توان روابط بین متغیرها را مشخص کرد. جداول توافقی عمدتاً به صورت جداول دو بعدی تشکیل می‌شوند تا وجود یا عدم وجود روابط بین متغیرها را بررسی کنند. در اینگونه جداول می‌توان از متغیر سومی نیز استفاده کرد و با فرض ثابت بودن اثر یک متغیر روابط بین متغیرهای دیگر را مورد بررسی قرار داد.

علیرغم کاربرد گسترده‌ای که جداول توافقی در تحلیل اولیه روابط بین متغیرها دارد، این روش، تکنیک بسیار قوی بنظر نمی‌رسد، بویژه اینکه قادر به تعیین سطح روابط بین متغیرها نمی‌باشد. به همین دلیل برای تحلیل‌های پیچیده‌تر بخصوص زمانی که طبیعت داده‌ها از نوع ترتیبی، فاصله‌ای و یا نسبی باشد ترجیحاً از سایر تکنیکها استفاده می‌شود.

#### ضرایب همبستگی مرتبط با متغیرهای اسمی:

روابط بین متغیرهای اسمی عمدتاً از طریق ضرایب کای اسکویر<sup>(۱)</sup> و ضرایب همبستگی چوپروف<sup>(۲)</sup>، فی<sup>(۳)</sup>، همبستگی توافق پیرسون<sup>(۴)</sup> و کرامر<sup>(۵)</sup> مورد بررسی قرار می‌گیرد. اگرچه کلیه این ضرایب به منظور بررسی روابط موجود بین متغیرهای اسمی مورد استفاده قرار می‌گیرند، اما علیرغم این نکته مشترک، هریک از این ضرایب تفاوت‌هایی با همدیگر دارند که تفسیر، تحلیل و کاربردهای خاصی را برای آنها تعیین کرده است. به منظور آشنایی با این ضرایب هرکدام از آنها بطور خلاصه در ذیل معرفی شده‌اند.

#### ۱- کای اسکویر:

آزمون کای اسکویر یکی از آزمونهای غیرپارامتری است که آن را با علامت  $X^2$  نمایش

1- Chi- square

2- Tschuprow's Coefficient of Correlation

3- Phi

4- Pearson Coefficient of Contingency

5- V Cramer

می‌دهند. این آزمون توسط فیشر<sup>(۱)</sup> ارائه شده و قادر است تا به سنجش آماری معنی داری تفاوت بین فراوانیهای مشاهده شده و فراوانیهای مورد انتظار بدست آمده از یک جامعه بپردازد. این آزمون نشان می‌دهد که آیا تفاوت موجود بین مقادیر فوق از نظر آماری معنی دار است یا این تفاوت عمدتاً بر اساس شانس می‌باشد. در محاسبه کای اسکویر فرض می‌شود که بین دو متغیر مورد بررسی (y , x) ارتباطی وجود ندارد. اما پس از محاسبه این آزمون از طریق نرم‌افزار SPSS چنانچه سطح معنی داری<sup>(۲)</sup> کوچک‌تر از ۰/۰۵ باشد می‌توان قضاوت کرد که به احتمال ۹۵ درصد بین دو متغیر رابطه وجود دارد. در این صورت فرض صفر که بر عدم وجود رابطه تأکید می‌کرد رد می‌شود. اما اگر سطح معنی داری محاسبه شده کوچک‌تر از ۰/۰۱ باشد در این صورت می‌توان ادعا کرد که به احتمال ۹۹ درصد بین دو متغیر رابطه وجود دارد. فرمول کای اسکویر بشرح ذیل است:

$$X^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

در این معادله:

$X^2$  = شاخص کای اسکویر

$F_o$  = فراوانی مشاهده شده

$F_e$  = فراوانی مورد انتظار می‌باشد.

#### شرایط استفاده از آزمون کای اسکویر:

برای استفاده از آزمون کای اسکویر شرایط ذیل باید صدق کند.

۱ - داده‌های مشاهده شده باید به صورت تصادفی گردآوری شده باشند.

۲ - کلیه موارد موجود در نمونه باید مستقل از هم باشند

۳ - حتی الامکان هیچ یک از خانه‌های جدول نباید کمتر از ۱۰ باشند. زمانیکه این شرط

تحقق پیدا نکند می‌توان از طریق ادغام منطقی برخی از گروهها و گروه‌بندی مجدد، تعداد فراوانی

1- Fisher

2- Significance Level

هریک از خانه‌ها را افزایش داد تا شرط فوق تحقق یابد. البته برخی از آماردانان تعداد ۵ مورد را برای هر یک از گروه‌ها یا خانه‌های جدول دو بعدی که قرار است برای آن کای اسکویر محاسبه شود کافی می‌دانند. اما آنچه که مسلم می‌باشد این است که نباید بیش از ۲۰ درصد فراوانی مورد انتظار خانه‌های اینگونه جداول کمتر از ۵ باشد. معمولاً زمانی که کای اسکویر از طریق برنامه SPSS محاسبه گردد در پنجره Output قید می‌شود که چند درصد از فراوانیهای مورد انتظار کمتر از ۵ می‌باشد.

۴ - تعداد نمونه باید به اندازه کافی بزرگ باشد. بهتر است حجم نمونه در اینگونه تجزیه و تحلیل‌ها حداقل ۵۰ مورد باشد.

آماره کای اسکویر می‌تواند برای یک متغیر اسمی و یک متغیر ترتیبی نیز بکار گرفته شود به عنوان مثال ممکن است محققى بخواهد تا عکس‌العمل چهارگروه از کارکنان یک مؤسسه را نسبت به یک سیاست مورد بررسی قرار دهد. در چنین مواردی او می‌تواند جدول تقاطعی ۱۶ خانه‌ای تشکیل دهد که در یک طرف جدول، گروه کارکنان واحدهای تولید، فروش، بازاریابی و بخش تحقیق و توسعه و در طرف دیگر بازتاب آنها نسبت به سیاست فوق در چهار گروه عدم علاقمندی، علاقمندی کم، علاقمندی متوسط و علاقمندی زیادی مقوله‌بندی کند<sup>(۱)</sup>.

در مواردی که فراوانیهای مورد انتظار کوچک باشند و جدول توافقی نیز  $2 \times 2$  باشد می‌توان از آزمونهای دیگری مانند آزمون احتمال فیشر<sup>(۲)</sup> استفاده کرد.

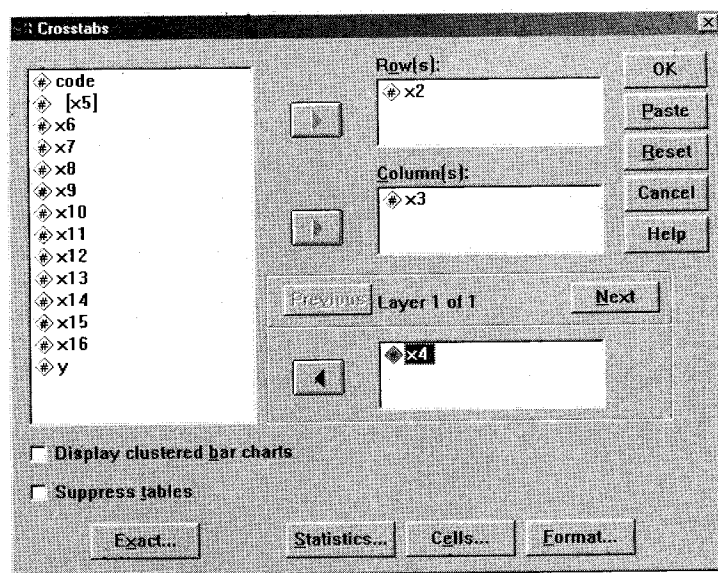
#### روش محاسبه جداول دو بعدی با نرم افزار SPSS

ممکن است بخواهید ارتباط دو متغیر مانند جنسیت و رضایت شغلی را با همدیگر بسنجید.

۱- سکاران، اوما (۱۳۸۰) روشهای تحقیق در مدیریت، ترجمه محمد صائبی و محمود شیرازی، انتشارات مرکز آموزش مدیریت دولتی

2- Fisher exact probability test

برای این کار باید جدول دو بعدی تهیه کنید. برای تهیه ی‌گونه جداول از Crosstabs استفاده می‌شود. این دستور از منوی Analyze/Descriptive Statistics/Crosstabs... بدست می‌آید. با انتخاب این منو شکل (۱-۴) در صفحه مانیتور ظاهر شد و از شما می‌خواهد که یک متغیر برای ردیف و یک متغیر برای ستون جدول انتخاب کنید. اگر شما بیش از یک متغیر در ردیف یا ستون وارد کنید مجموعه‌ای از جداول به دست می‌آید.

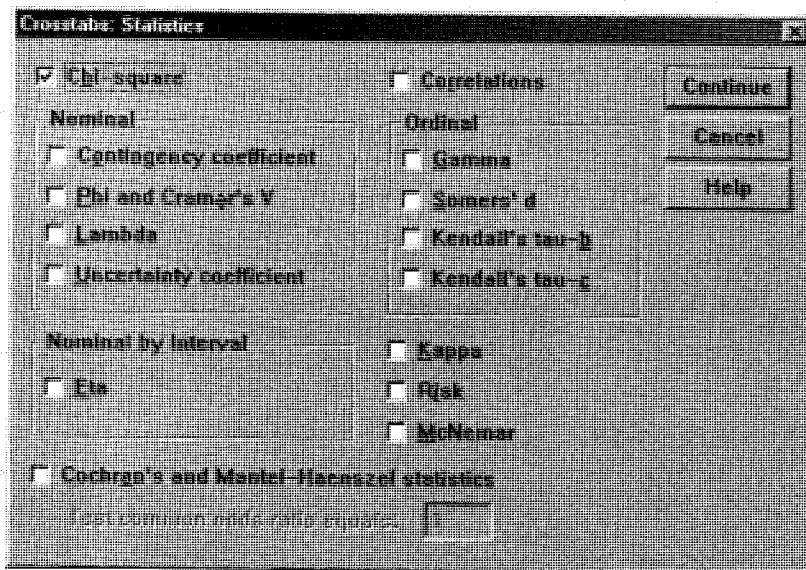


شکل (۱-۴) پنجره Crosstabs

ممکن است بخواهید اطلاعات خود را براساس یک متغیر سومی نیز طبقه‌بندی کنید. به عنوان مثال اگر بخواهید رابطه جنسیت و رضایت شغلی را بطور جداگانه برای دو منطقه (شهر و روستا) بدست آورید. در اینجا متغیر سوم (منطقه) باید به عنوان متغیر کنترل در جعبه ای که تحت عنوان Layers 1 of 1 می‌باشد وارد شود. شما می‌توانید متغیر کنترل بیشتری داشته باشید و این کار با فشار دادن دکمه Next امکانپذیر است. در پایین همین پنجره دکمه Cells را فشار دهید و گزینه‌های Observed Frequencies و Expected Frequencies را انتخاب کنید.



سپس دکمه Statistics را فشار دهید و آماره‌های اسکویئر را نیز انتخاب کنید (شکل ۴-۲). علاوه بر این، آماره‌هایی نظیر ضریب توافقی<sup>(۱)</sup>، فی و  $\chi^2$  کرامر<sup>(۲)</sup>، لامبدا<sup>(۳)</sup>، گاما<sup>(۴)</sup>، کندال تاو b<sup>(۵)</sup>، کندال تاو C<sup>(۶)</sup>، اتا<sup>(۷)</sup> و غیره در همین قسمت قرار دارند، که در صورت نیاز می‌توانید انتخاب کنید.



شکل (۴-۲) پنجره Crosstabs: Statistics

در صورتی که داده‌ها به صورت جدول دو بعدی ارائه شده باشند و بخواهید برای فراوانیهای خانه‌های جدول، آماره‌های اسکویئر محاسبه کنید برای این کار ابتدا متغیرهای مورد نظر را در جدول دوبعدی کدبندی نموده و در دو ستون در صفحه Data Editor وارد کرده سپس فراوانیهای جدول را

1- Contingency coefficient

2- Phi and Cramer's V

3- Lambda

4- Gamma

5- Kendall's tau b

6- Kendall's tau C

7- Eta

در ستون سوم ثبت کنید. به عنوان مثال (در صورتی که جدول  $3 \times 3$  باشد اطلاعات فوق باید در سه ستون و نه ردیف وارد شوند. در ستون اول کدهای یکی از متغیرها و در ستون دوم کدهای متغیر دوم و در ستون سوم تعداد فراوانیهای مرتبط با آنها ثبت می‌گردند). پس از آن با استفاده از دستور Data/Weight Cases ستون مربوط به فراوانی را تعریف کنید. سپس دستور Analyze/Descriptive Statistics/Crosstabs را اجرا کرده و یکی از متغیرهای جدول دوبعدی را در قسمت سطر و دیگری را در قسمت ستون وارد کنید. بعد از آن، از قسمت Statistics هر نوع محاسبات آماری را که مدنظر داشته باشید مشخص کرده و از قسمت Cells مورد Expected را انتخاب کنید. با فشار دادن دکمه OK نتایج محاسبات در قسمت Output ظاهر می‌شود.

#### ضریب همبستگی چوپروف :

ضریب همبستگی چوپروف که آن را با T نمایش می‌دهند به منظور تعیین شدت وابستگی بین متغیرهای مورد مطالعه بکار گرفته می‌شود و مقدار آن همواره بین صفر و یک در نوسان می‌باشد. این آزمون زمانیکه هر دو متغیر اسمی و یا یکی اسمی و دیگری ترتیبی باشد مورد استفاده قرار می‌گیرد، اما نباید تعداد سطر و ستون با هم برابر باشند. به عبارت دیگر این ضریب برای جداول توافقی مستطیلی بکار گرفته می‌شود. چنانچه جدول توافقی  $2 \times 2$  باشد نمی‌توان از این ضریب استفاده کرد در چنین مواردی باید از ضریب فی استفاده نمود.

#### ضریب همبستگی Phi:

همانطوری که قبلاً نیز بیان گردید ضریب همبستگی فی به منظور بررسی شدت همبستگی بین دو متغیر اسمی که به صورت دو وجهی و در قالب جدول توافقی  $2 \times 2$  می‌باشد مورد استفاده قرار می‌گیرد. به همین دلیل در اینگونه موارد بجای استفاده از کای اسکویر باید از ضریب همبستگی فی استفاده شود. تفاوت ضریب همبستگی فی با کای اسکویر در این است که کای اسکویر سطح معنی دار بودن همبستگی بین متغیرها را تعیین می‌کند، در حالی که ضریب فی شدت همبستگی کای

اسکویر را نشان می‌دهد<sup>(۱)</sup>. ضریب همبستگی فی نیز مانند کای اسکویر مورد تفسیر قرار می‌گیرد و مقدار آن نیز همواره بین صفر و یک در نوسان است<sup>(۲)</sup>.

#### ضریب همبستگی توافق پیرسون<sup>(۳)</sup>:

این ضریب که آن را با C نمایش می‌دهند، میزان همبستگی بین دو متغیر اسمی را که به صورت توافقی تنظیم شده‌اند محاسبه می‌کند و این شاخص زمانی بکار می‌رود که خانه‌های جدول توافقی بیشتر از ۲×۲ باشد. هرچه مقدار این ضریب بزرگتر باشد درجه همبستگی متغیرها با هم بیشتر خواهد بود و افزایش مقدار این ضریب بستگی به تعداد طبقات آن دارد<sup>(۴)</sup>.

#### ضریب کرامر:

این ضریب برای تعیین میزان شدت همبستگی بین دو متغیر اسمی مورد استفاده قرار می‌گیرد و آن را با علامت  $\chi^2$  نمایش می‌دهند که مقدار آن همواره بین صفر و یک در نوسان است. این ضریب در مقایسه با سایر ضرایب انعطاف بیشتری دارد. بطوریکه هم برای جداول توافقی بیشتر از ۲×۲ و هم برای جداول مستطیلی بکار می‌رود.

#### ضرایب همبستگی مرتبط با متغیرهای ترتیبی:

در صورتی که متغیرهای مورد بررسی از نوع ترتیبی باشند، برای بررسی روابط بین آنها باید

۱- دلاور علی (۱۳۶۹) "روشهای آماری در روانشناسی و علوم تربیتی" انتشارات دانشگاه پیام نور، تهران ص ۵۰۳

۲- ساعی علی (۱۳۷۷) "آمار در علوم اجتماعی با کاربرد نرم افزار SPSS/PC<sup>+</sup> در پژوهش‌های اجتماعی، مؤسسه نشر جهاد دانشگاهی ص ۱۴۱-۱۴۲.

3- Pearson's Coefficient of Contingency

۴- ساعی پیشین، ص ۱۴۲

از ضرایب کندال، گاما و اسپرمن استفاده کرد. برای آشنایی با اصول بکارگیری این شاخص‌ها و نحوه تفسیر آنها، هریک از ضرایب فوق بطور مختصر مورد بررسی قرار می‌گیرند.

#### ضریب همبستگی رتبه‌ای کندال<sup>(۱)</sup>:

شاخص کندال حالت تقارن دارد. به این معنا که، متغیرها قرینه بوده و برای محقق مهم نیست که کدامیک از متغیرهای مورد مطالعه وابسته و کدامیک مستقل می‌باشد. این شاخص مشخص می‌کند که تا چه میزان افزایش یا کاهش در یک متغیر با افزایش یا کاهش در متغیر دیگر همراه است. مقدار ضریب کندال همواره بین -۱ تا +۱ در نوسان می‌باشد. کندال به سه روش مختلف محاسبه می‌گردد. زمانی که جدول به صورت دوبعدی باشد و تعداد آزمودنی یا  $N$  نیز بیشتر باشد از کندال تائو  $a^{(۲)}$  استفاده می‌شود. کندال تائو  $b^{(۳)}$  نیز زمانی مورد استفاده قرار می‌گیرد که تعداد خانه‌های سطر و ستون جدول دوبعدی با هم برابر باشند. اما چنانچه تعداد ردیف‌ها و ستون‌ها با هم برابر نباشند و جدول به صورت مستطیل باشد باید از کندال تائو  $c^{(۴)}$  استفاده شود.

#### ضریب گاما:

ضریب گاما<sup>(۵)</sup> شاخصی است که از طریق آن می‌توان با آگاهی از پاسخهای یک متغیر پاسخها و نتایج متغیر دیگر را پیشگویی کرد. گاما در واقع میزان کاهش خطا را که در نتیجه آگاهی از پاسخ افراد در متغیر مستقل رخ می‌دهد، مشخص می‌کند. به عبارت دیگر محقق با آگاهی از وضعیت پاسخ افراد به متغیر مستقل ( $x$ ) با استفاده از گاما میزان کاهش خطا را در پیشگویی متغیر وابسته ( $y$ ) مورد آزمون قرار می‌دهد. مقادیر گاما بین +۱ تا -۱ در نوسان است. از آنجا که ضریب گاما همواره

1- Kendall's rank Correlation Coefficient

2- Kendall's tau a

3- Kendall's tau b

4- Kendall's tau c

5- Gamma Coefficient

روابط بین دو متغیر را خیلی قوی نشان می‌دهد بنابراین، این شاخص میل به بزرگ نشان دادن شدت ارتباط بین دو متغیر را دارد و به همین دلیل باید در بکارگیری این شاخص احتیاط کرد<sup>(۱)</sup>.

#### ضریب همبستگی رتبه‌ای اسپرمن:

ضریب همبستگی رتبه‌ای اسپرمن<sup>(۲)</sup> که در اوایل دهه ۱۹۰۰ توسط چارلز اسپرمن<sup>(۳)</sup> ابداع گردیده است زمانی مورد استفاده قرار می‌گیرد که داده‌ها به صورت رتبه‌ای متوالی ناپیوسته (۱، ۲، ۳، ...) باشد و یا اینکه مقادیر اصلی به رتبه تبدیل شوند. در صورتی که داده‌ها با مقیاس فاصله‌ای یا نسبی اندازه‌گیری شده باشند می‌توان آنها را به رتبه تبدیل کرد. بعد ضریب همبستگی رتبه‌ای اسپرمن را محاسبه نمود. برای اینکار رتبه ۱ به بیشترین مقدار، رتبه ۲ به مقدار بعدی و الی آخر داده می‌شود. در این رتبه‌بندی اگر در بین مقادیر اصلی دو یا چند مورد دارای مقادیر مساوی باشند در این صورت رتبه‌های مربوط به آنها با همدیگر جمع شده و بر تعداد آنها تقسیم می‌گردد و میانگین بدست آمده به عنوان رتبه برای مقادیر فوق در نظر گرفته می‌شود.

ضریب همبستگی اسپرمن که آن را با  $P$  یا  $r_s$  نمایش می‌دهند همواره بین  $+1$  و  $-1$  در نوسان است و از لحاظ سطح سنجش نیز ترتیبی و از نوع متقارن می‌باشد. به همین دلیل برای محقق مهم نیست که کدام متغیر مستقل و کدامیک وابسته باشد.

اگر چه ضرایب اسپرمن و کندال تائو هر دو از نوع رتبه‌ای بوده و در برنامه SPSS نیز در یک جا قرار دارند و از یک مسیر محاسبه می‌شوند، اما موارد کاربرد آنها تا حدودی متفاوت می‌باشد. چنانچه در داده‌های مربوط به متغیرها موارد هم رتبه زیاد وجود داشته باشد در این صورت بهتر است از روش کندال تائو استفاده کرد. اما اگر تعداد طبقات زیاد باشد و موارد هم رتبه نیز بسیار کم باشد و داده‌ها به صورت رتبه‌ای متوالی ناپیوسته باشد باید از ضریب همبستگی اسپرمن استفاده کرد.

۱- ساعی پیشین، ص ۱۵۱-۱۵۰

2- Spearman Rank Correlation Coefficient 3- Charles Spearman

**ضریب همبستگی پیرسون :**

ضریب همبستگی پیرسون<sup>(۱)</sup> از روشهای پرکاربرد جهت تعیین میزان رابطه بین دو متغیر محسوب گردیده و با علامت  $r$  نمایش داده می‌شود. این ضریب به منظور بررسی رابطه بین دو متغیر فاصله‌ای یا نسبی مورد استفاده قرار می‌گیرد و مقدار آن همواره بین  $+1$  و  $-1$  در نوسان است. چنانچه مقدار بدست آمده مثبت باشد به معنای این است که تغییرات در هر دو متغیر بطور هم جهت اتفاق می‌افتد. به عبارت دیگر با هرگونه افزایش در مقدار یک متغیر مقدار متغیر دیگر نیز افزایش می‌یابد و برعکس. اما چنانچه مقدار  $r$  منفی باشد بیانگر این نکته است که دو متغیر در جهت عکس همدیگر حرکت می‌کند. یعنی با افزایش مقدار یک متغیر، مقادیر متغیر دیگر کاهش می‌یابد و برعکس. اگر مقدار بدست آمده برای ضریب همبستگی صفر باشد به معنای این است که هیچگونه رابطه‌ای بین دو متغیر وجود ندارد. اگر مقدار  $r$  دقیقاً برابر با  $+1$  باشد بیانگر همبستگی مثبت کامل و اگر برابر با  $-1$  باشد نشان دهنده همبستگی کامل منفی بین دو متغیر است.

**ضرایب همبستگی و سطوح اندازه‌گیری :**

از آنجا که انتخاب ضریب همبستگی مناسب برای بررسی روابط بین متغیرها تحت تأثیر مقیاس اندازه‌گیری متغیرهای مورد بررسی قرار دارد، لذا تناسب بین سطوح اندازه‌گیری و ضریب همبستگی سازگار با آنها در جدول (۴-۱) ارائه شده است. این جدول تصمیم‌گیری برای انتخاب ضریب همبستگی را برای محققان تسهیل می‌کند.

**نحوه محاسبه ضرایب همبستگی پیرسون اسپرمن و کندال با برنامه SPSS:**

پس از اینکه داده‌های مربوط به متغیرهای مختلف در ستونهای مستقل در پنجره Data Editor وارد گردید، برای محاسبه ضرایب همبستگی فوق‌دستور Analyze/Correlate/Bivariate را اجرا کنید تا شکل (۴-۳) ظاهر شود.

1- Pearson correlation coefficient

جدول (۴-۱) ضرایب همبستگی و سطح اندازه‌گیری

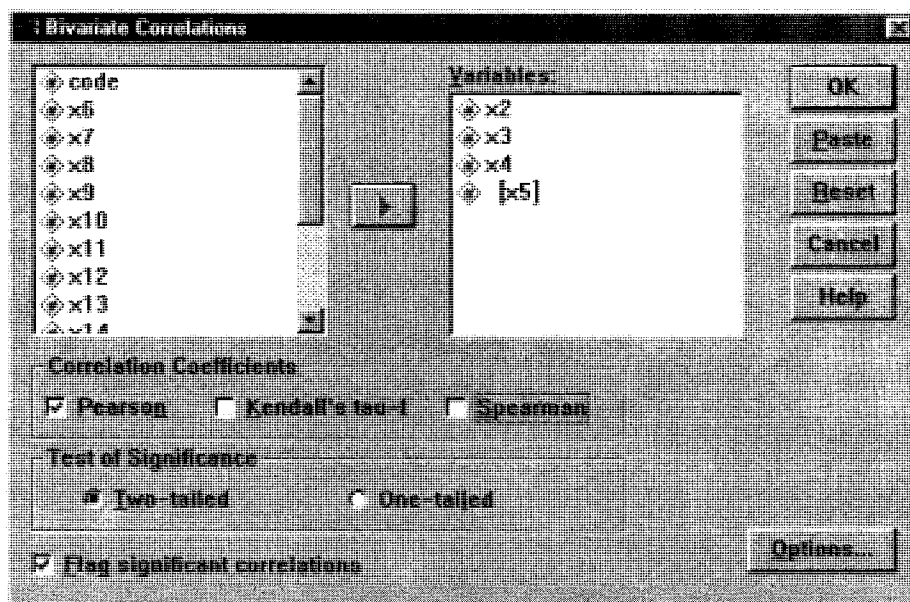
فاصله‌ای یا نسبی	ترتیبی	اسمی		متغیر Y	
		چند حالتی	دو حالتی	دو حالتی	اسمی
	رتبه‌ای	ضریب فی	ضریب چوپروف	دو حالتی	اسمی
				چند حالتی	اسمی
	رتبه‌ای	ضریب توافقی	ضریب چوپروف	دو حالتی	اسمی
				چند حالتی	اسمی
	رتبه‌ای	ضریب توافقی	ضریب چوپروف	دو حالتی	اسمی
				چند حالتی	اسمی
مقادیر فاصله‌ای به رتبه‌ای تبدیل و ضریب همبستگی	رتبه‌ای	ضریب توافقی	ضریب چوپروف	دو حالتی	اسمی
				چند حالتی	اسمی
مقادیر فاصله‌ای به رتبه‌ای تبدیل و ضریب همبستگی	رتبه‌ای	ضریب توافقی	ضریب چوپروف	دو حالتی	اسمی
				چند حالتی	اسمی

مأخذ:

Miller, L. E. (1998) "Appropriate Analysis" Journal of Agricultural Education, Vol. 39, No.2. pp. 1-10. (اصلاح شده بوسیله مؤلف)

در این پنجره متغیرهای مورد نظر را از جعبه سمت چپ به قسمت variables منتقل کنید و از قسمت Correlation Coefficients ضرایب مورد نظر را متناسب با طبیعت داده‌ها انتخاب کنید. برای اینکه سطح معنی داری روابط بین متغیرها را در سطح ۰/۰۵ و ۰/۰۱ مورد بررسی قرار دهید. مورد Flag significant correlations که در پایین پنجره فوق قرار دارد را روشن کنید. با کلیک کردن Ok و اجرای این دستور نتایج در پنجره Output ظاهر می‌شود. مقدار ضریب همبستگی

همواره بین ۱+ و ۱- در نوسان است. میزان همبستگی صفر به معنی آن است که رابطه‌ای بین دو متغیر وجود ندارد. در صورتی که مقادیر بدست آمده با یک ستاره (\*) مشخص شده باشد به معنای این است که رابطه در سطح ۹۵ درصد و اگر با دو ستاره (\*\*) مشخص شده باشند به معنای این است که رابطه در سطح ۹۹ درصد معنی‌دار است. نمونه‌ای از محاسبات ضریب همبستگی اسپرمن در جدول (۲-۴) ارائه شده است.



شکل (۴-۳) پنجره Bivariate correlations

دو نکته در رابطه با ضریب همبستگی وجود دارد که توجه به آنها بسیار ضروری است. نکته اول اینکه وجود همبستگی بین دو متغیر به معنای این نیست که آنها دارای اثرات مساوی هستند و یا یکی علت دیگری است. زیرا ممکن است هر دو تحت تأثیر عامل دیگری باشند. به عنوان مثال اگر بین میزان غیبت از کار و رضایت شغلی رابطه منفی وجود داشته باشد، نمی‌توان نتیجه گرفت که دلیل زیاد بودن غیبت، عدم رضایت شغلی است. زیرا ممکن است که غایب بودن، خود باعث احساس عدم



رضایت شغلی باشد و یا هر دو تحت تأثیر عامل دیگر قرار گرفته باشند. دومین نکته در مورد همبستگی ساده این است که این همبستگی نشان می‌دهد که چقدر ارتباط خطی بین دو متغیر وجود دارد. در یک همبستگی با ارتباط نموداری، مقدار پایین X در ارتباط با مقدار پایین Y و مقدار متوسط محور X با مقدار بالای محور Y و مقدار بالای محور X با مقدار پایین محور Y می‌تواند ارتباط داشته باشد. اما اینگونه روابط در ضریب همبستگی مشخص نمی‌شود. بنابراین همواره بهتر است که روابط دو متغیر در مواردی که به صورت خطی نیستند در قالب یک نمودار پراکنندگی<sup>(۱)</sup> رسم شود. برای اینکار می‌توان از دستور Graphs/Scatter/Simple استفاده کرد.

جدول (۲-۴) نتایج محاسبات ضریب همبستگی پیرسون

		X2	X3	X4	X6
X2	Pearson Correlation	1.000	.867**	.273	.382*
	Sig. (2-tailed)	.	.000	.144	.037
	N	30	30	30	30
X3	Pearson Correlation	.867**	1.000	.378*	.420*
	Sig. (2-tailed)	.000	.	.040	.021
	N	30	30	30	30
X4	Pearson Correlation	.273	.378*	1.000	-.013
	Sig. (2-tailed)	.144	.040	.	.944
	N	30	30	30	30
X6	Pearson Correlation	.382*	.420*	-.013	1.000
	Sig. (2-tailed)	.037	.021	.944	.
	N	30	30	30	30

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

## ضریب همبستگی جزئی:

همبستگی جزئی<sup>(۱)</sup> که به همبستگی خالص نیز معروف است، عکس همبستگی چندگانه است. این روش سعی می‌کند تا بطور جداگانه رابطه بین دو متغیر را از طریق حذف اثرات سایر متغیرهای مرتبط اندازه‌گیری کند. به عبارت دیگر، این ضریب با کنترل رابطه کاذب به محقق کمک می‌کند تا پس از تفکیک کردن اثرات متغیرهای کنترلی، همبستگی خالص بین دو متغیر مورد نظر سنجیده شود. هدف همبستگی جزئی نشان دادن اهمیت نسبی متغیرهای مستقل مختلف در تبیین واریانس متغیر وابسته است. معمولاً ضریب همبستگی جزئی به صورت  $r_{12.345...n}$  نشان داده می‌شود. این علامت همبستگی بین متغیر  $X_1$  و  $X_2$  را در حالی که متغیرهای  $X_3, X_4, X_5$  تا  $X_n$  ثابت نگهداشته شده‌اند، نشان می‌دهد. به عنوان مثال  $r_{12/3}$  به معنای رابطه بین متغیرهای  $X_1$  و  $X_2$  با ثابت نگهداشتن متغیر  $X_3$  می‌باشد، که فرمول آن بشرح ذیل می‌باشد.

$$r_{12/3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \times \sqrt{1 - r_{23}^2}} \quad \text{مثال:}$$

اگر ضریب همبستگی بین  $r_{12} = 0/8$ ،  $r_{13} = 0/65$  و  $r_{23} = 0/7$  باشد، ضریب همبستگی جزئی براساس فرمول بالا بشرح ذیل محاسبه می‌گردد.

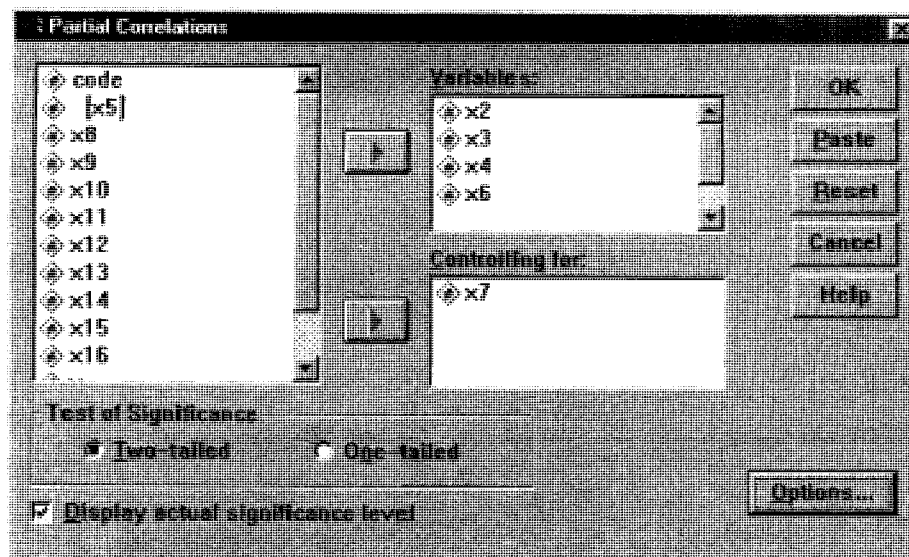
$$r_{12/3} = \frac{0/8 - 0/65 \times 0/7}{\sqrt{1 - 0/65^2} \times \sqrt{1 - 0/7^2}} = 0/635$$

نحوه محاسبه ضریب همبستگی جزئی با برنامه SPSS:

همانطوری که قبلاً نیز گفته شد همبستگی جزئی با حذف اثرات سایر متغیرها تنها روابط بین دو متغیر را مورد بررسی قرار می‌دهد. به عبارت دیگر در این آماره با ثابت نگهداشتن اثرات سایر متغیرها تنها رابطه یک متغیر وابسته با یک متغیر مستقل سنجیده می‌شود. بنابراین هر ضریب

همبستگی جزئی رابطه یک متغیر مستقل را با متغیر وابسته، مورد توجه قرار می‌دهد. اگر تنها یک متغیر ثابت نگهداشته شود در این صورت ضریب همبستگی جزئی، ضریب نوع اول<sup>(۱)</sup> نامیده می‌شود اما اگر دو متغیر ثابت نگهداشته شود ضریب نوع دوم<sup>(۲)</sup> نامیده می‌شود و الی آخر.

برای محاسبه ضریب همبستگی جزئی دستور Analyze/Correlate/Partial را اجزاء کنید تا شکل (۴-۴) ظاهر شود.



شکل (۴-۴) پنجره Partial correlations

در این پنجره متغیری که قرار است بر روی دو متغیر قبلی مورد کنترل قرار گیرد را در قسمت Controlling For قرار داده و سایر متغیرهای مورد نظر را به قسمت Variables منتقل کنید و دکمه ok را فشار دهید. با اجرای این دستور نتایج در پنجره Output ظاهر می‌شود. نمونه‌ای از نتایج محاسبات ضریب همبستگی جزئی در جدول (۴-۳) ارائه شده است. در این جدول همبستگی بین

1- First Order Coefficient

2- Second Order Coefficient

متغیرهای  $X_2, X_3, X_4, X_6$  و  $X_7$  با کنترل اثرات متغیر  $X_5$  محاسبه شده است.

جدول (۳-۴) نتایج محاسبات ضریب همبستگی جزئی

Controlling for.. X7

	X2	X3	X4	X6
X2	1.0000 ( 0) P= .	.8680 ( 27) P= .000	.2639 ( 27) P= .167	.3384 ( 27) P= .073
X3	.8680 ( 27) P= .000	1.0000 ( 0) P= .	.3740 ( 27) P= .046	.3310 ( 27) P= .079
X4	.2639 ( 27) P= .167	.3740 ( 27) P= .046	1.0000 ( 0) P= .	-.0494 ( 27) P= .799
X6	.3384 ( 27) P= .073	.3310 ( 27) P= .079	-.0494 ( 27) P= .799	1.0000 ( 0) P= .

(Coefficient / (D.F.) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

در جدول (۳-۴) علاوه بر مقادیر ضریب همبستگی سطح معنی‌داری واقعی همبستگی بین متغیرها نیز محاسبه شده است. براساس مقادیر  $P$ ، چنانچه این مقادیر کمتر از  $0.01$  باشد همبستگی بین متغیرها در سطح  $99\%$  درصد و اگر کمتر از  $0.05$  اما بزرگتر از  $0.01$  باشد، این ارتباط در سطح  $95\%$  درصد معنی‌دار خواهد بود.

## فصل پنجم

# مقایسه میانگین‌ها

### مقدمه:

چنانچه داده‌های مربوط به متغیر وابسته از نوع کمی با مقیاس (فاصله‌ای یا نسبی) و داده‌های متغیر مستقل یا گروه‌بندی از نوع کیفی با مقیاس (اسمی یا ترتیبی) باشد برای بررسی تفاوتها می‌توان به مقایسه میانگین‌ها پرداخت و معنی دار بودن تفاوت‌های موجود بین طبقات یا گروهها را مورد بررسی قرار داد. برای این منظور دو روش پرکاربرد تحت عنوان آزمون  $t$  و آزمون  $F$  وجود دارد. چنانچه متغیر مستقل یا متغیر گروه‌بندی تنها دو گروه داشته باشد (به عنوان مثال اگر بخواهیم درآمد زنان و مردان را با هم مقایسه کنیم) در این صورت باید از آزمون  $t$  استفاده کرد. اما اگر تعداد گروهها بیشتر از دو باشد در این صورت آزمون  $t$  کاربرد نخواهد داشت و باید از آزمون  $F$  که به تحلیل واریانس یا ANOVA نیز معروف است استفاده شود. (به عنوان مثال اگر بخواهیم میزان درآمد گروههای شغلی کارگر، کارمند و کشاورز را با همدیگر مقایسه کنیم باید از آزمون  $F$  استفاده کنیم). آنچه که باید مورد توجه قرار گیرد این است که آزمون  $F$  تنها معنی‌داری تفاوت بین میانگین گروهها را مورد بررسی قرار می‌دهد اما مشخص نمی‌کند که این تفاوتها در بین کدامیک از گروههای مورد بررسی وجود دارد. به همین دلیل برای اینکه بدانیم تفاوت‌های بدست آمده در بین کدامیک از

طبقات وجود دارد و از این طریق مقایسه‌ای بین گروهها انجام گیرد باید از آزمون شفیه<sup>(۱)</sup> یا LSD و یا از آماره‌هایی نظیر توکی<sup>(۲)</sup> یا دانکن<sup>(۳)</sup> استفاده شود. این آزمونها میانگین زوجها را با همدیگر به صورت دویبدو مقایسه کرده و وجود اختلاف معنادار بین آنها را نشان می‌دهد.

این فصل به تحلیل و تبیین آزمونهای پارامتری که به مقایسه میانگین گروهها می‌پردازد اختصاص یافته و علاوه بر تشریح زمینه‌های کاربرد هر یک از آنها، نحوه محاسبه آنها نیز با استفاده از نرم افزار SPSS توضیح داده شده است.

### انتخاب آزمون مناسب برای مقایسه میانگینها :

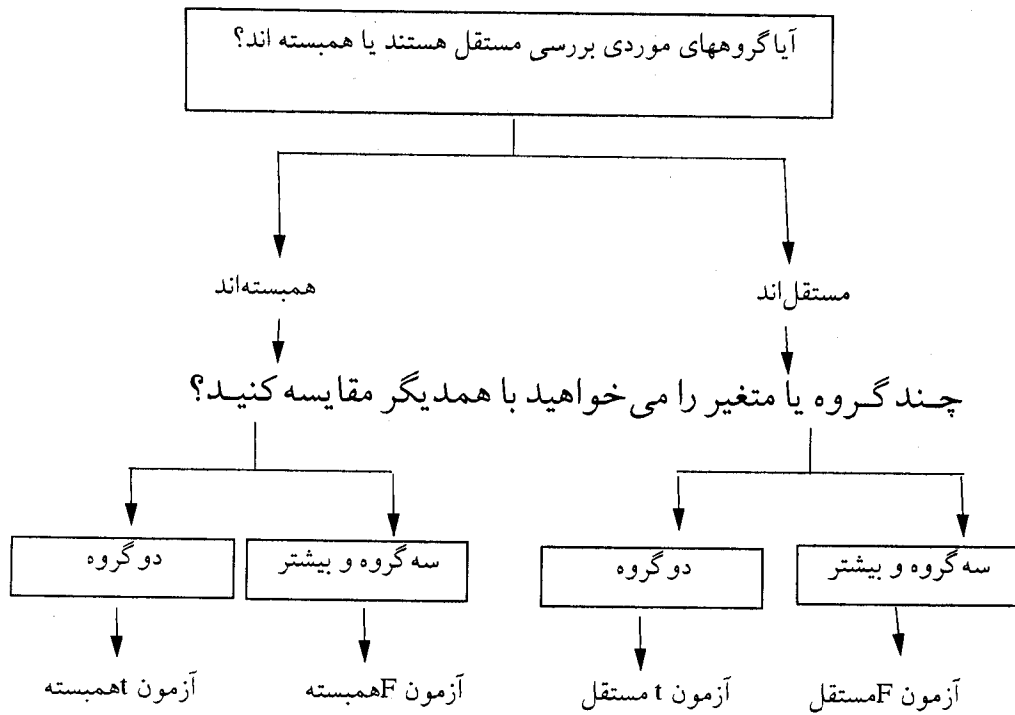
آزمونهای t و F عمده‌ترین آزمونهای آماری برای مقایسه میانگین گروهها می‌باشند. از آنجا که گروههای مورد بررسی ممکن است مستقل یا همبسته باشند، بنابراین هر یک از آزمونهای فوق به دو بخش مستقل و همبسته تقسیم می‌شوند. تصمیم‌گیری در خصوص اینکه در چه مواقعی باید از آزمونهای t یا F مستقل یا همبسته استفاده کرد مهمترین مسئله در تحلیل داده‌های کمی است. برای این منظور درخت تصمیم‌گیری در شکل (۱-۵) ارائه شده است که به محقق کمک می‌کند تا مناسب‌ترین آزمون آماری را انتخاب و بکار گیرد. براساس این نمودار اگر محقق تنها دو سؤال را مطرح و پاسخهای لازم را ارائه دهد قادر به انتخاب آزمون مناسب خواهد بود. اولین سوال این است که آیا گروههای مورد مطالعه مستقل هستند یا همبسته‌اند؟ و سوال دوم این است که چند گروه یا متغیر را می‌خواهید با همدیگر مقایسه کنید؟ پاسخ هر یک از سوالات فوق و آزمون مناسب برای آنها در شکل (۱-۵) آمده است.

1- Scheffe test

2- Tukey

3- Duncan

شکل (۵-۱) درخت تصمیم‌گیری در مورد انتخاب آزمون مناسب برای بررسی سطح معنی‌داری تفاوت‌های بین میانگین گروه‌ها (روش‌های پارامتری)



#### پیش فرض‌های آزمون‌های پارامتری

آزمون‌های پارامتری  $t$  و  $F$  را زمانی می‌توان مورد استفاده قرار داد که پیش فرض‌های ذیل در خصوص داده‌های مورد بررسی صادق باشد. این پیش فرضها عبارتند از:

- ۱ - مشاهدات از یک جامعه نرمال انتخاب شده باشند.
- ۲ - اطلاعاتی که با هم مقایسه می‌شوند باید تقریباً واریانس یکسانی داشته باشند. اگر گروه‌های مورد بررسی اندازه یکسانی داشته باشند در این صورت این فرض چندان مهم نیست. در جامعه نمونه بزرگ نیز حتی اگر واریانس یک گروه دو برابر دیگری باشد باز هم می‌توان از آزمون‌های

پارامتری استفاده کرد.

۳- داده‌های گردآوری شده دارای مقیاس فاصله‌ای یا نسبی باشند.

اگر اطلاعات جمع آوری شده این سه شرط را نداشته باشد می‌توان داده‌های فوق را به غیرپارامتری تبدیل کرد و از روشهای آماری غیرپارامتری استفاده نمود<sup>(۱)</sup>. روش عمده برای تبدیل داده‌های پارامتری به غیرپارامتری رتبه‌بندی کردن آنهاست.

### روش محاسبه آزمون T با نرم‌افزار SPSS:

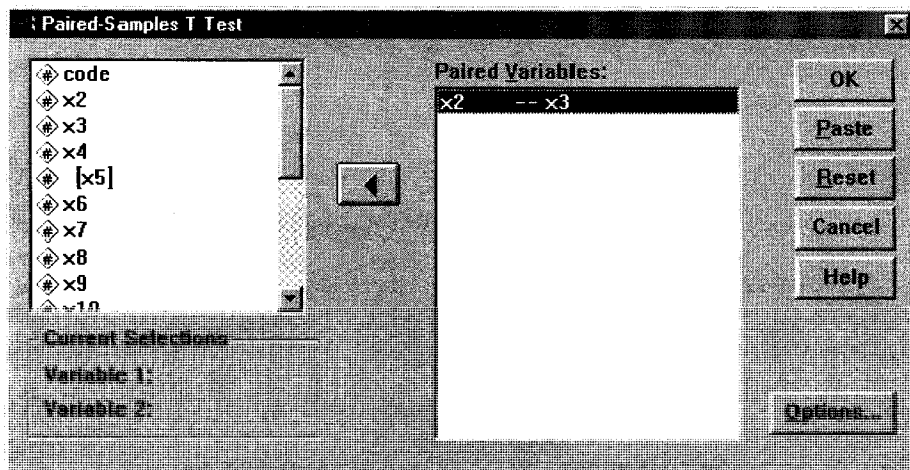
همانطوری که قبلاً نیز بیان گردید، آزمون t یکی از روشهای پارامتری می‌باشد که به منظور دستیابی به اینکه آیا تفاوت بین میانگین‌های دو گروه از نظر آماری معنی دار هست یا نه مورد استفاده قرار می‌گیرد. در به کارگیری این روش چنانچه گروهها همبسته باشند باید از آزمون t همبسته و در صورتی که گروهها مستقل باشند باید از آزمون t مستقل استفاده کرد. این دو روش بطور جداگانه در ذیل مورد بحث قرار گرفته است. نکته‌ای که باید در محاسبه آزمون t مدنظر قرار گیرد این است که چنانچه شما بیش از دو دسته اطلاعات یا متغیر داشته باشید نمی‌توانید روش t-test، Multiple را بکار ببرید. به عنوان مثال اگر از سه گروه دانشجوی (A, B, C) تست هوش گرفته شده باشد، نمی‌توانید با محاسبه آزمون t گروه A را با B و گروه B را با C و گروه C را با A مقایسه کنید. بنابراین در چنین مواقعی باید روشی را بکار ببرید که مناسب آن شرایط باشد. برای اینگونه اطلاعات پارامتری، تحلیل واریانس مناسب است. اما برای اطلاعات غیرپارامتری می‌توانید از آزمون فریدمن (در صورتی که گروهها همبسته باشند) یا تست کروسکال والیس (در صورتی که گروهها مستقل باشند) استفاده کنید.

۱- روشهای غیر پارامتری به تفصیل در فصل ششم مورد بررسی قرار گرفته‌اند.



## آزمون t همبسته و نحوه محاسبه آن با SPSS:

این روش زمانی مورد استفاده قرار می‌گیرد که دو موضوع یا دو متغیر در مورد یک گروه مورد بررسی قرار گیرد. فرض کنید که از تعدادی کودک خواسته شده است تا ابتدا با مداد رنگی نقاشی کنند و داوران به نقاشی آنها نمره دهند، سپس از همان کودکان خواسته شود تا با آب رنگ نقاشی کنند و مجدداً توسط داوران نمره داده شود. با استفاده از آزمون t همبسته می‌توان قضاوت کرد که آیا بین نمرات دو نوع نقاشی (نقاشی با مداد رنگی و نقاشی با آب رنگ) تفاوت معنی‌داری وجود دارد یا خیر. برای این کار نمرات کسب شده را در دو ستون در پنجره Data Editor وارد کرده و سپس منوی Analyze/Compare Means/ Paired-samples T-Test را انتخاب کنید تا پنجره آزمون t در مانیتور ظاهر شود (شکل ۲-۵).



شکل (۲-۵) پنجره Paired - Samples T Test

در اینجا شما باید دو متغیری که قرار است با همدیگر مقایسه شوند را مشخص کنید. با کلیک کردن بر روی متغیر اول این متغیر در جلو Variable 1 در قسمت Current selections

ظاهر می‌شود و با کلیک کردن بر روی متغیر دوم، این متغیر نیز به جلو 2 Variable منتقل می‌گردد. با این انتخاب دکمه فلش دار روشن می‌شود و در صورت فشار دادن آن هر دو متغیر به قسمت Paired variables در سمت راست، جهت محاسبه منتقل می‌گردند. سپس دکمه OK را فشار دهید تا محاسبات لازم انجام گرفته و مقادیر دو متغیر با همدیگر مقایسه شوند.

با نگاهی به نتایج محاسبات خواهید دید که در قسمت بالای صفحه Output، میانگین، انحراف معیار و خطای معیار میانگین<sup>(۱)</sup>، در قسمت میانی ضریب همبستگی بین دو متغیر و در قسمت پایین صفحه اختلاف بین میانگین‌ها<sup>(۲)</sup>، و انحراف معیار بین آنها و خطای معیار میانگین‌شان ثبت می‌گردد. پس از آنها مقدار t با درجه آزادی (df) و سطح احتمال یا سطح معنی داری قرار می‌گیرد. اگر سطح احتمال کمتر از ۰/۰۵ باشد از نظر آماری می‌توان نتیجه گرفت که بین دو مجموعه از متغیرها با احتمال ۹۵ درصد تفاوت معنی داری وجود دارد و اگر سطح معنی داری کمتر از ۰/۰۱ باشد می‌توان قضاوت کرد که تفاوت بین دو گروه در سطح ۹۹ درصد معنی دار است. نتایج محاسبات t تست همبسته بین نمرات دو گروه در جدول (۵-۱) ارائه شده است.

#### نحوه برخورد با مواردی که دارای Missing Data هستند:

روش Paired T-Test، مواردی را که دارای Missing Value در هریک از دو متغیری که قرار است با هم مقایسه شوند وجود داشته باشد حذف می‌کند. شما می‌توانید از طریق دستور Exculde Cases Listwise کلیه مواردی که دارای Missing Data هستند را حذف کنید. برای این کار دکمه Options را در جعبه Paired Samples T-Test انتخاب کنید و بعد موارد مورد نظر را از قسمت Missing Values انتخاب کنید. همچنین دکمه Options این امکان را برای شما فراهم می‌کند تا در صورت لزوم سطح اطمینان ۹۹٪ را تعیین کنید. برای این کار کافی است عدد ۹۹ را در محل مورد نظر تایپ کنید.

1- Standard Error of Mean

2- Means difference

## جدول (۱-۵) نتایج تست همبسته

## Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	X1	14.5815	48	2.0750	.2995
	X2	10.5900	48	2.4836	.3585

## Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	X1 & X2	48	.844	.000

## Paired Samples Test

		Paired Differences				t	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		
					Lower		Upper
Pair 1	X1 - X2	3.9915	1.3308	.1921	3.6050	4.3779	20.780

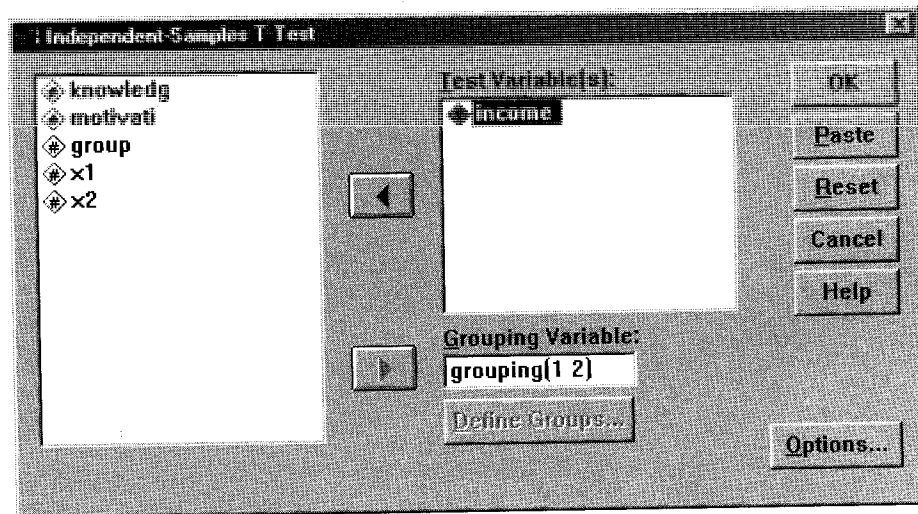
## Paired Samples Test

		df	Sig. (2-tailed)
Pair 1	X1 - X2	47	.000

### آزمون t مستقل و نحوه محاسبه آن با برنامه SPSS:

آزمون t مستقل زمانی مورد استفاده قرار می‌گیرد که بخواهیم میانگین بین دو گروه از افراد را که از همدیگر متفاوت هستند و در گروه‌های A و B سازماندهی شده‌اند مقایسه کنیم. در این روش سعی می‌گردد که تفاوت بین دو گروه مشخص شود.

به عنوان مثال اگر بخواهیم این فرضیه که بین درآمد زنان و مردان تفاوت معنی‌داری وجود دارد را آزمایش کنیم می‌توانیم از آزمون t مستقل استفاده کنیم. برای این کار باید درآمد افراد مورد بررسی در یک ستون به پنجره Data Editor وارد شود و سپس در ستون مقابل آن با کدهای ۱ و ۲ گروه‌ها از هم تفکیک گردند. به عنوان نمونه می‌توان در ستون گروه‌بندی زنان را با کد ۱ و مردان را با کد ۲ تعریف کرد. پس از وارد کردن داده‌ها به رایانه مانند شکل (۳-۵) در مانیتور ظاهر شود. متغیرهایی که قرار است با هم مقایسه شوند را در جعبه Test Variable (S) قرار دهید و متغیری که به عنوان متغیر گروه‌بندی در نظر گرفته شده است را در پنجره Grouping Variable وارد کنید.



(شکل ۳-۵) پنجره Independent - Samples T Test

سپس دکمه Define Groups را فشار دهید. با این دستور جعبه‌ای در مانیتور ظاهر می‌شود. روش معمول این است که مقادیری که در Grouping Variable برای تعریف کردن دو گروه استفاده می‌شود را در پنجره‌های Group 1 و Group 2 وارد کنید. سپس دکمه Continue و بعد دکمه OK را فشار دهید تا نتایج محاسبات در پنجره Output ظاهر شود. در محاسبات انجام شده، میانگین درآمد هریک از گروه‌ها همراه با انحراف معیار و خطای معیار میانگین مشخص می‌گردد. تفاوت بین میانگین‌ها و نتیجه تست لون<sup>(۱)</sup> نیز برای برابری واریانس مشخص شده و سپس مقدار  $t$ ، درجه آزادی، احتمال انحراف استاندارد  $t$  برای تفاوتها و فاصله اطمینان در سطح ۹۵ درصد نیز در آنجا محاسبه می‌شود. اگر مقدار تست لون نشان دهد که دو گروه دارای واریانس غیرمساوی هستند، (یعنی مقدار احتمال کمتر از ۰/۰۵ باشد) در این صورت باید مقدار  $t$  را برای واریانس‌های غیر برابر استفاده کرد. اما اگر سطح معنی‌داری تست لون بزرگتر از ۰/۰۵ بود، باید مقدار  $t$  را برای واریانس‌های برابر در نظر گرفت و پذیرفت که دو گروه واریانس برابر دارند.

نتایج محاسبات آزمون  $t$  مستقل برای مقایسه درآمد زنان و مردان در جدول (۵-۲) ارائه شده

است.

#### تحلیل واریانس:

فیشر<sup>(۲)</sup> اولین کسی بود که مفهوم «واریانس» را بکار برد و با ارائه تئوری در رابطه با تحلیل واریانس<sup>(۳)</sup> ارزش و اهمیت آن را در تحلیل پدیده‌های واقعی توجیه و تبیین کرد. پس از آن سندکور<sup>(۴)</sup> و دیگران تلاشهای زیادی را در بسط و توسعه این تکنیک انجام دادند. تحلیل واریانس که به آن ANOVA یا آزمون F نیز می‌گویند، یکی از تکنیکهای آماری مؤثر و پرکاربرد در تحقیقات اقتصادی، اجتماعی، علوم تربیتی، روانشناسی، مدیریت و حتی کشاورزی، بیولوژی و غیره است. همانطوری که قبلاً نیز بیان شد زمانی که بخواهیم میانگین‌های دو جامعه (یا نمونه) را با همدیگر

1- Levene's Test

2- R. A. Fisher

3- Analysis of variance

4- Sendecor

جدول (۲-۵) مقایسه تفاوت بین میانگین درآمد زنان و مردان با استفاده از آزمون t مستقل

Group Statistics

	GROUPING	N	Mean	Std. Deviation	Std. Error Mean
INCOME	male	23	196.7826	31.8561	6.6425
	female	25	134.9600	20.4073	4.0815

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
INCOME	Equal variances assumed	1.850	.180
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means			
		t	df	Sig. (2-tailed)	Mean Difference
INCOME	Equal variances assumed	8.072	46	.000	61.8226
	Equal variances not assumed	7.930	36.923	.000	61.8226

Independent Samples Test

		t-test for Equality of Means		
		Std. Error Difference	95% Confidence Interval of the Difference	
			Lower	Upper
INCOME	Equal variances assumed	7.6586	46.4066	77.2386
	Equal variances not assumed	7.7962	46.0249	77.6203

مقایسه کنیم و معنی‌داری تفاوت بین آنها را بررسی نماییم از آزمونهای  $z$  و  $t$  استفاده می‌گردد. اما زمانی که محقق بخواهد به بررسی تفاوت میانگین‌های بیش از دو جامعه (یا نمونه) بپردازد، بکارگیری آزمونهایی چون  $t$  امکانپذیر نخواهد بود. برای این منظور در اینگونه تحقیقات از روش تحلیل واریانس یا آزمون  $F$  استفاده می‌گردد. به عنوان مثال اگر بخواهیم تفاوت درآمد بین سه گروه شغلی کارمند، کشاورزی و کارگر را بررسی کنیم از آزمون  $F$  یا تحلیل واریانس استفاده می‌کنیم. این روش تفاوت معنی دار بودن درآمدهای گروههای شغلی سه گانه را از طریق مقایسه میانگین درآمدهای آنها بررسی می‌کند.

تحلیل واریانس در واقع روشی برای آزمایش تفاوت بین گروههای مختلف داده‌ها یا نمونه‌هاست. این روش کل واریانس موجود در یک مجموعه از داده‌ها را به دو بخش تقسیم می‌کند. بخشی از این واریانس ممکن است بخاطر شانس و تصادف حادث شده باشد و بخش دیگر ممکن است ناشی از دلایل یا عوامل خاصی باشد. از طرف دیگر واریانس موجود ممکن است ناشی از تفاوت بین گروههای مورد مطالعه و یا بخاطر تفاوت موجود در درون نمونه‌ها حادث شده باشد. بنابراین ANOVA به عنوان یک روش تحلیل با بررسی مجموع این تفاوتها به تبیین پدیده مورد نظر می‌پردازد. از طریق اینگونه تحلیل‌هاست که محقق می‌تواند بررسی کند که آیا بین درآمد گروههای مختلف شغلی تفاوت معنی‌داری وجود دارد یا نه؟

#### اصول اساسی تحلیل واریانس:

مهمترین اصل در تحلیل واریانس (ANOVA) آزمایش تفاوت‌های موجود در بین میانگین‌های جوامع یا گروههای مورد مطالعه از طریق بررسی میزان واریانس بین گروهها نسبت به واریانس درون گروههاست. در واریانس درون جامعه فرض بر این است که تفاوت بین مقدار  $X_{ij}$  نسبت به میانگین جامعه بخاطر شانس است. به عبارت دیگر تأثیرات بر روی  $X_{ij}$  غیرقابل توضیح و تبیین می‌باشد، در حالی که در بررسی تفاوت‌های بین جوامع و گروهها، فرض بر این است که تفاوت بین میانگین جامعه یا نمونه زام با میانگین کل به دلیل عوامل خاص می‌باشد. بنابراین زمانی که از

تحلیل واریانس استفاده می‌شود فرض می‌گردد که هریک از نمونه‌ها از یک جامعه نرمال انتخاب شده‌اند و هریک از این جوامع نیز واریانس برابری دارند همچنین فرض می‌شود کلیه عوامل، بجز عواملی که مورد مطالعه می‌باشند تحت کنترل هستند.

بطور خلاصه در تحلیل واریانس دو نوع واریانس باید برآورد گردد که یکی از آنها واریانس بین گروهها و دیگری واریانس درون گروههاست. واریانس بین گروهها عبارت از مجموع مجذورات انحراف میانگین هر گروه از میانگین کل و واریانس درون گروهها عبارت از مجموع مجذورات انحراف مقادیر هر فرد از میانگین گروه خود می‌باشد. بر مبنای تخمین این دو واریانس است که امکان محاسبه میزان  $F$  براساس فرمول ذیل فراهم می‌گردد.

$$F = \frac{\text{واریانس بین گروهها}}{\text{واریانس درون گروهها}}$$

با بدست آوردن میزان  $F$ ، در صورت عدم استفاده از رایانه، محقق باید میزان  $F$  بدست آمده را بر مبنای درجه آزادی مورد نظر با  $F$  جدول که معمولاً در پایان کتابهای آمار و روش تحقیق ارائه می‌شوند مقایسه کند. در صورتی که  $F$  بدست آمده بزرگتر از  $F$  جدول باشد می‌توان قضاوت کرد که بین میانگین گروههای مورد مطالعه تفاوت معنی‌داری وجود دارد. اما اگر محقق در محاسبات خود از رایانه استفاده کند و سطح معنی‌داری بدست آمده کوچک‌تر از  $0/05$  و بزرگتر  $0/01$  باشد می‌توان قضاوت کرد که بین میانگین گروههای مورد مطالعه تفاوت در سطح  $95$  درصد می‌باشد و در صورتی که عدد بدست آمده کوچک‌تر از  $0/01$  باشد می‌توان مدعی شد که این تفاوت در سطح  $99$  درصد معنی‌دار است.

نکته مهم در تحلیل واریانس (ANOVA) این است که، اگر در بین میانگین گروههای مختلف تفاوت معنی‌داری وجود داشته باشد، صرفاً از طریق نتایج ANOVA نمی‌توان محل این تفاوتها را بدست آورد. به عبارت دیگر اگر به مقایسه سه گروه  $A$ ،  $B$  و  $C$  بپردازیم و تفاوت معنی‌داری در بین آنها وجود داشته باشد، نمی‌توانیم قضاوت کنیم که آیا این تفاوت بین  $A$  و  $B$  است یا بین  $B$  و  $C$  و یا بین  $A$  و  $C$ . در چنین مواقعی نباید از طریق آزمون  $t$  به مقایسه دو بدو گروهها پرداخت، زیرا



هرقدر تعداد دفعاتی که آزمون  $t$  انجام می‌گیرد بیشتر باشد سطح اطمینان نتایج پایین می‌آید<sup>(۱)</sup>. به همین دلیل برای تعیین محل تفاوتها در بین گروههای سه گانه و بیشتر، باید از آزمونهایی مانند، آزمون شفیه<sup>(۲)</sup>، آزمون چند دامنه دانکن<sup>(۳)</sup>، آزمون توکی<sup>(۴)</sup> و آزمون استیودنت- نیومن- کیول<sup>(۵)</sup> برحسب ضرورت استفاده کرد.

#### تحلیل واریانس یک طرفه با استفاده از برنامه SPSS:

همانطوری که قبلاً نیز بیان گردید، در صورتی که سه یا بیش از سه دسته اطلاعات کمی داشته باشید و بخواهید فرضیات خود را آزمایش کنید برای اینکار نمی‌توانید از آزمون  $t$  استفاده کنید. زیرا روش فوق تنها می‌تواند دو دسته از اطلاعات را مقایسه کند. از طرف دیگر روش Multiple T-Test نیز برای این کار مناسب نمی‌باشد. به همین دلیل، روش تحلیل واریانس تکنیک مناسبی برای اینگونه بررسیهای آماری است. این روش نیز همان پیش فرضهای آزمون  $t$  را مبنا قرار می‌دهد. بنابراین تنها در صورتی که چنین پیش فرضی قابل تصور باشد می‌توان از این روش استفاده کرد.

فرض کنید سه دسته اطلاعات از سه گروه مختلف گردآوری شده است و می‌خواهیم بدانیم که آیا تفاوتی بین میانگین گروهها وجود دارد یا خیر؟. اگر تفاوتی بین گروهها وجود نداشته باشد به عبارت دیگر اگر فرض صفر<sup>(۶)</sup> درست باشد، اطلاعات آنها حتماً از یک جامعه آماری گردآوری شده است و هر سه دسته میانگین و واریانس برابری خواهند داشت و واریانس هرگروه تخمینی از واریانس جامعه آماری خواهد بود. (واریانس بدست آمده از طریق نوسانات تصادفی بین گروهها به واریانس

۱- به عنوان مثال انجام سه آزمون  $t$  بطور همزمان، سطح اطمینان را از ۹۵٪ به ۸۶٪ یعنی  $(0/۹۵)^۳$  کاهش می‌دهد.

2- Scheffe's Test

3- Duncan Multiple Range Test

4- Tukey's Test

5- Student- Newman- Keull's Test

6- Null Hypothesis

خطا<sup>(۱)</sup> معروف است زیرا این اختلاف به دلیل شانس انتخاب حادث می‌شود). بهترین تخمین واریانس جامعه آماری از طریق محاسبه میانگین واریانس‌های هر سه گروه بدست می‌آید. بنابراین از طریق نگاه به اختلاف در متوسط واریانس‌های سه گروه می‌توانیم تخمینی از خطاهای واریانس بدست آوریم.

هم چنین، اگر فرضیه صفر درست باشد میانگین سه گروه تقریباً با هم برابر بوده و واریانس میانگین‌ها نیز بسیار کوچک خواهد بود. اختلاف بین میانگین‌های سه گروه به واریانس تیمار<sup>(۲)</sup> معروف است. بنابراین اگر فرضیه صفر درست باشد و سه گروه با هم اختلافی نداشته باشند، واریانس بین میانگین‌ها برابر با خطای واریانس خواهد بود و اگر واریانس تیمار را به واریانس خطا تقسیم کنیم نتیجه برابر یک خواهد بود.

اگر فرضیه صفر درست نباشد، سه گروه با هم دیگر تفاوت خواهند داشت. در این صورت واریانس میانگین‌ها بیشتر از واریانس خطا خواهد بود و اگر واریانس تیمار را بر واریانس خطا تقسیم کنیم عددی که بدست می‌آید بیشتر از یک خواهد بود.

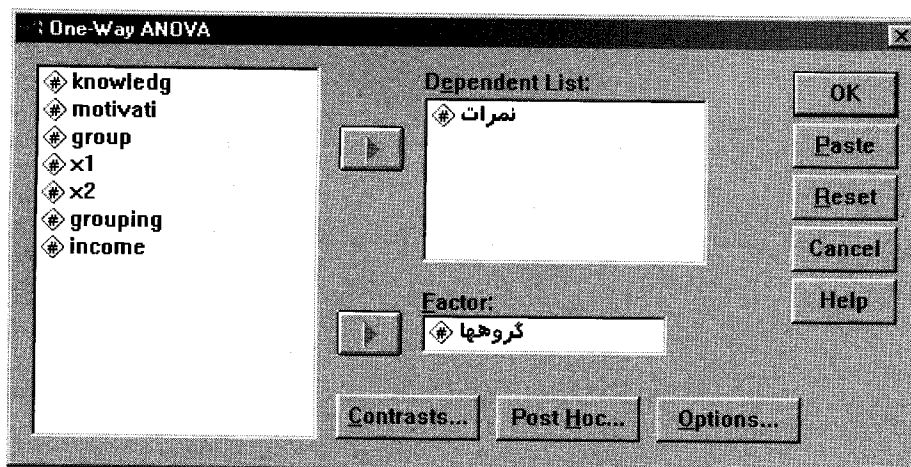
در تحلیل واریانس، ما در واقع واریانس تیمار را با واریانس خطا مقایسه می‌کنیم تا فرضیه را آزمایش کنیم و ببینیم که آیا تفاوت معنی‌داری بین میانگین گروه‌ها وجود دارد یا خیر. مسئله مهم در تحلیل واریانس یک طرفه مقدار  $F$  است، که در واقع بعد از محاسبه  $F$ ، اگر سطح معنی‌داری آن کمتر از  $0.05$  باشد می‌توان نتیجه گرفت که تفاوت معنی‌داری بین گروه‌ها وجود دارد.

به عنوان مثال اگر بخواهید اثر بخشی سه روس آموزش را با همدیگر مقایسه کنید و اگر سه گروه فراگیر داشته باشید، بطوریکه گروه اول به صورت انفرادی، گروه دوم بوسیله برنامه کلاسی و گروه سوم به صورت نیمه حضوری آموزش دیده باشند، برای محاسبه ANOVA و مقایسه سه گروه، نمرات آنها را در یک ستون در پنجره Data Editor وارد کنید. و در ستون دیگر با استفاده از کدهای

1- Error Variance

2- Treatment Variance

۱، ۲، ۳ گروه‌های فوق را تعریف نمایید. سپس برای محاسبه واریانس یک طرفه، منوی Analyze/Compare means/ One Way ANOVA را انتخاب کنید. در پنجره‌ای که در صفحه مانیتور ظاهر می‌شود (شکل ۵-۳)، متغیر وابسته نمرات را به جعبه Dependent list و متغیر گروه‌ها را به قسمت Factor وارد کنید.



شکل (۵-۴) پنجره One-Way ANOVA

نتایج تحلیل واریانس نشان می‌دهد که آیا تفاوتی بین میانگین‌های سه گروه یا بیشتر از پاسخگویان وجود دارد یا نه. اما این روش نمی‌تواند مشخص کند که این تفاوتها در کجا هستند به عبارت دیگر نمی‌تواند بگوید که آیا گروه ۱ با هر دو گروه ۲ و ۳ متفاوت است؟ یا گروه ۲ با هر دو گروه ۱ و ۳ متفاوت است؟ دستیابی به پاسخ اینگونه سوالات از طریق تست دانکن، تست توکی، تست شفه و غیره امکان‌پذیر می‌باشد. این محاسبات با فشار دادن دکمه Post Hoc... از طریق پنجره One Way ANOVA و انتخاب آزمونهای مورد نظر از لیست فوق انجام می‌گیرد. با انجام انتخابهای فوق دکمه Ok را فشار دهید تا محاسبات در پنجره Output ظاهر شود (جدول ۵-۳).

جدول (۵-۳) مقایسه تفاوت میانگین نمرات اخذ شده بوسیله سه گروه از افراد آموزش دیده با

استفاده از سه روش آموزشی مختلف

نمرات

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30995.792	2	15497.896	14.804	.000
Within Groups	47109.875	45	1046.886		
Total	78105.667	47			

## Post Hoc Tests

### Multiple Comparisons

Dependent Variable: نمرات

	گروهها (J)	گروهها (I)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	کلاسی	افراد	61.4375*	11.4394	.000	33.7126	89.1624
		نیمه حضوری	39.3750*	11.4394	.004	11.6501	67.0999
	افراد	کلاسی	-61.4375*	11.4394	.000	-89.1624	-33.7126
		نیمه حضوری	-22.0625	11.4394	.142	-49.7874	5.6624
	نیمه حضوری	افراد	-39.3750*	11.4394	.004	-67.0999	-11.6501
		کلاسی	22.0625	11.4394	.142	-5.6624	49.7874

\*. The mean difference is significant at the .05 level.

## Homogeneous Subsets

نمرات

گروهها	N	Subset for alpha = .05	
		1	2
Tukey HSD <sup>a</sup>	کلاسی	16	136.7500
	نیمه حضوری	16	158.8125
	افراد	16	198.1875
	Sig.		.142
Duncan <sup>a</sup>	کلاسی	16	136.7500
	نیمه حضوری	16	158.8125
	افراد	16	198.1875
	Sig.		.060

Means for groups in homogeneous subsets are displayed.

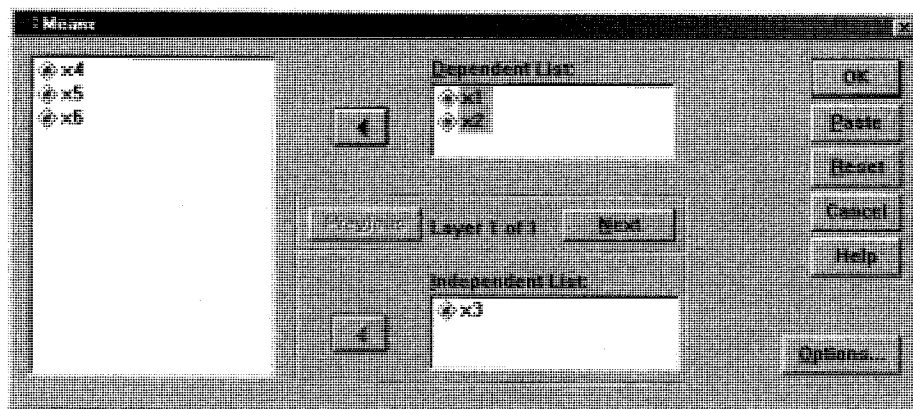
a. Uses Harmonic Mean Sample Size = 16.000.

نحوه محاسبه میانگین زیرگروهها با نرم افزار SPSS :

برای محاسبه میانگین زیرگروهها باید از قسمت Means اقدام کنید برای این کار منوی

Analyze/Compare Means/Means را انتخاب کنید تا شکل (۵-۵) ظاهر شود. سپس متغیر

وابسته درآمد (Income) را در جعبه Dependent List وارد کرده و متغیر گروه‌های شغلی (Group) را در جعبه Independent List قرار داده و سپس دکمه OK را فشار دهید تا نتیجه در Output ظاهر شود جدول (۴-۵).



شکل (۵-۵) پنجره Means

در جعبه Means قسمت دیگری وجود دارد که در آن Layer 1 of 1 نوشته شده است. این گزینه به شما این امکان را می‌دهد که پاسخگویان را یک مرحله دیگر به زیر گروه‌ها تقسیم کنید. فرضاً اگر می‌خواهید که میانگین درآمد پاسخگویان را به تفکیک گروه‌های مختلف شغلی برای دو منطقه مختلف شمال و جنوب محاسبه کنید. برای این کار متغیر درآمد را در Dependent list و متغیر گروه‌های شغلی را در Independent list و متغیر منطقه را با فشار دادن دکمه Next در قسمت Layer به جعبه Independent وارد کنید. با انتخاب Options می‌توانید واریانس میانگین‌ها را برای زیرگروه‌های وابسته بدست آورید. بنابراین اگر درآمد را به عنوان متغیر وابسته و مشاغل را به عنوان متغیر مستقل یا گروه‌بندی انتخاب کرده‌اید، جدول ANOVA و دکمه eta را از

قسمت Options انتخاب کنید تا واریانس یک طرفه درآمد زیرگروههای مشاغل مختلف شغلی را بدست آورید.

انتخاب گزینه eta زمانی مناسب است که متغیر وابسته در مقیاس فاصله‌ای و متغیر مستقل دارای مقیاس اسمی یا ترتیبی باشند.

#### جدول (۴-۵) میانگین و انحراف معیار درآمد برای گروههای مختلف شغلی

##### INCOME

GROUPS	Mean	N	Std. Deviation
کارمند	198.1875	16	37.0571
کشاورز	136.7500	16	21.1707
کارگر	158.8125	16	36.3212
Total	164.5833	48	40.7655

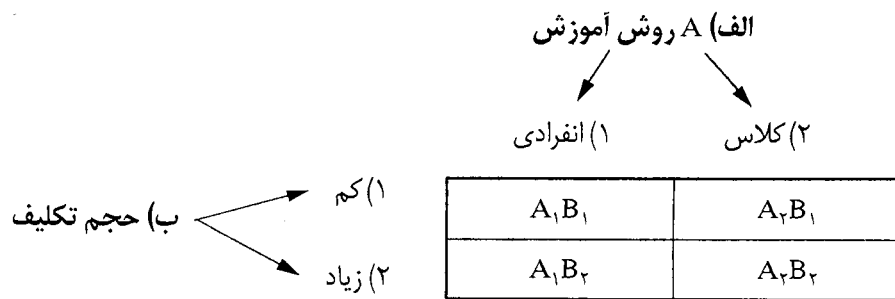
#### تحلیل واریانس دو طرفه با استفاده از برنامه SPSS:

همانطوری که قبلاً بیان شد اگر محقق تنها یک متغیر (مانند درآمد) را انتخاب کند و بخواهد تفاوت بین طبقات یا گروههای مختلف را بررسی کند در این صورت از تحلیل واریانس یک طرفه<sup>(۱)</sup> استفاده می‌کند. اما اگر بخواهد اثر دو عامل را بر روی یک متغیر وابسته بررسی کند باید از تحلیل واریانس دو طرفه<sup>(۲)</sup> استفاده کند. بنابراین در تحلیل واریانس دو طرفه داده‌ها براساس دو عامل تقسیم بندی می‌شوند. به عنوان مثال برداشت در واحد سطح یک محصول کشاورزی ممکن است براساس انواع مختلف بذر و انواع مختلف کود تقسیم‌بندی شوند. یا یک مؤسسه تجاری ممکن است میزان فروش خود را به تفکیک فروشندگان و به تفکیک مناطق مختلف تقسیم‌بندی کند و یا یک

1- One - way Analysis of Variance

2- Two - way Analysis of Variance

مؤسسه آموزشی میزان یادگیری فراگیران (سطح نمرات) را براساس روشهای مختلف آموزشی و حجم تکلیف تعیین شده برای فراگیران طبقه‌بندی کند. در این مثال اگر روشهای آموزش کلاسی و آموزش انفرادی به عنوان دوروش آموزشی و حجم تکلیف کم و تکلیف زیاد دو سطح مختلف از حجم تکلیف تعیین شده برای فراگیران باشد. در این صورت چهارگروه اطلاعات جداگانه بشرح ذیل بدست می‌آید.

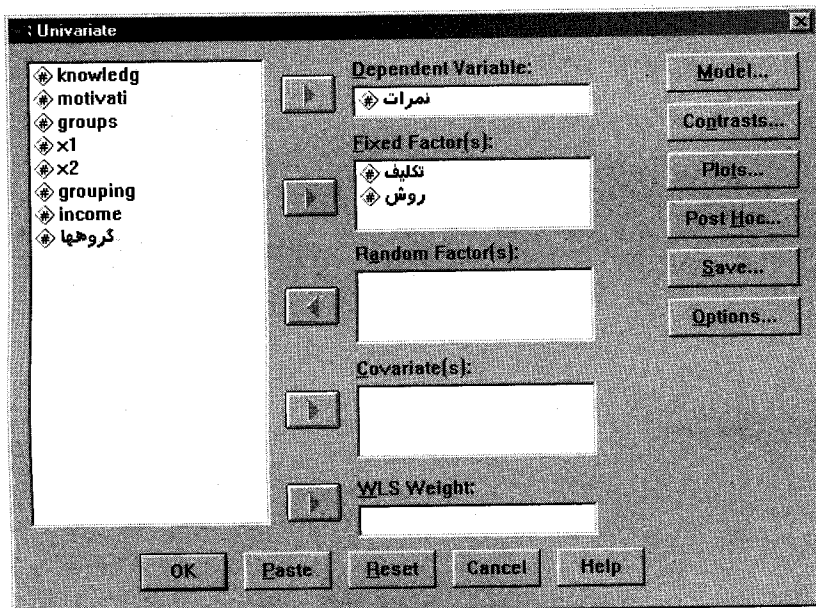


در اینجا ممکن است بخواهیم بدانیم که آیا میزان یادگیری فراگیران براساس روشهای مختلف آموزشی متفاوت است؟ همچنین آیا میزان یادگیری برحسب حجم تکلیف متفاوت می‌باشد؟ و آیا بین متغیرها اثر متقابل<sup>(۱)</sup> وجود دارد؟ (منظور از اثر متقابل این است که اثر یک متغیر تحت تأثیر متغیر دیگر قرار گرفته باشد.) به عنوان مثال ممکن است متوجه شوید که تفاوت در میزان یادگیری در روش آموزش کلاسی با آموزش انفرادی در حجم تکلیف کم به مراتب کمتر از حجم تکلیف زیاد است. اگر چنین یافته‌ای بدست آید تحلیل واریانس اثر متقابل معنی‌داری را نشان می‌دهد.

مثال فوق نمونه‌ای از تحلیل واریانس برای گروههای مستقل می‌باشد. در چنین مواقعی باید نمرات در یک ستون، روشهای آموزشی با کدهای ۱ و ۲ در ستون دیگر و حجم تکلیف با کدهای ۱ و ۲ در ستون سوم به قسمت Data Editor وارد شوند. سپس دستور:

1- Interaction

Analyze/General Linear Model/ Univariate... اجراء گردد<sup>(۱)</sup> تا شکل (۵-۶) ظاهر شود.



شکل (۵-۶) پنجره Univariate

در این پنجره متغیر وابسته (نمرات فراگیران) را به قسمت Dependent Variable و متغیرهای مربوط به روشهای آموزشی و حجم تکلیف را به قسمت Fixed Factor(s) وارد کرده و دکمه Ok را فشار دهید تا محاسبات در پنجره Output ظاهر شود، (جدول ۵-۵). در صورتی که بخواهید کواریانس نیز محاسبه شود می‌توانید متغیر یا متغیرهای مورد نظر را در قسمت Covariate وارد کنید. در صورت استفاده از کواریانس، اثر متغیرهای وارد شده به این قسمت کنترل شده و تحلیل واریانس صرفاً براساس مقادیر متغیرهای وارد شده به قسمت Factor(s) محاسبه می‌شود. همانطوری که از نتایج جدول فوق برمی‌آید اگر چه روشهای مختلف و حجم تکلیف هر دو

۱- دستور فوق در SPSS نسخه ۶ بصورت Statistics/ANOVA Models/Simple Factorial می‌باشد.



معنی‌دار می‌باشند اما، اثرات متقابل آنها معنی‌دار نمی‌باشد.

### جدول (۵-۵) نتایج تحلیل واریانس دو طرفه

#### Between-Subjects Factors

	Value Label	N
تکلیف 1	کم	24
تکلیف 2	زیاد	24
روش 1	انفرادی	24
روش 2	کلاسی	24

#### Tests of Between-Subjects Effects

Dependent Variable: نمرات

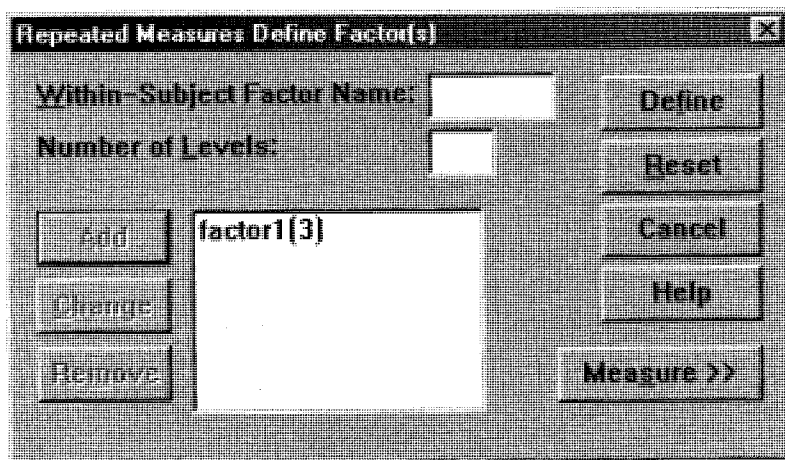
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	18015.000 <sup>a</sup>	3	6005.000	4.397	.009
Intercept	1300208.333	1	1300208.333	952.047	.000
تکلیف	7905.333	1	7905.333	5.788	.020
روش	7701.333	1	7701.333	5.639	.022
تکلیف * روش	2408.333	1	2408.333	1.763	.191
Error	60090.667	44	1365.697		
Total	1378314.000	48			
Corrected Total	78105.667	47			

a. R Squared = .231 (Adjusted R Squared = .178)

#### تحلیل واریانس برای گروههای همبسته:

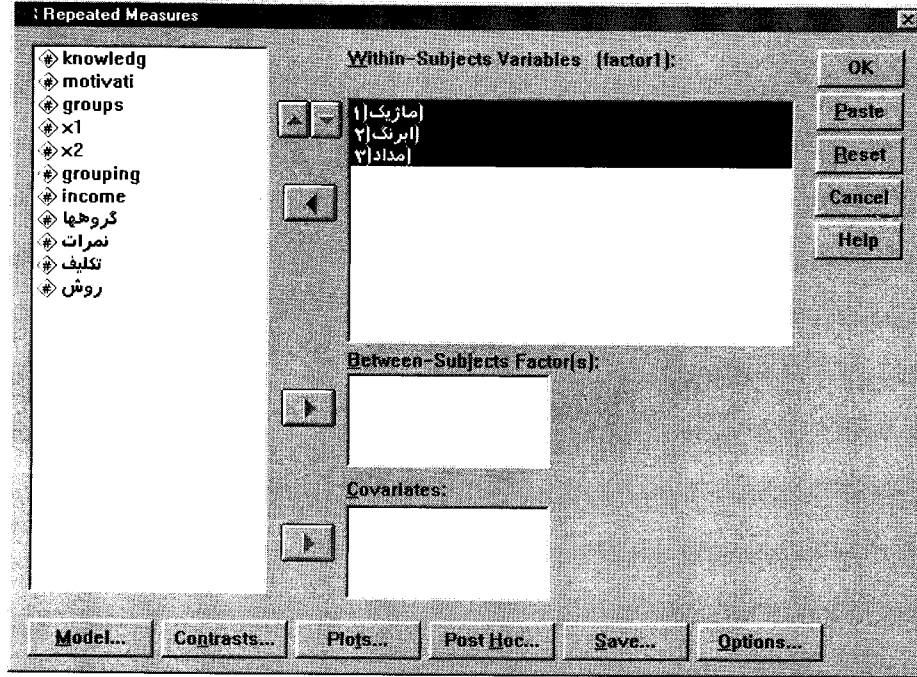
تحلیل واریانس برای گروههای همبسته، زمانی که در آن تنها یک گروه وجود دارد و چندین آزمون بر روی آنها اعمال می‌شود نیز بکار می‌رود. به عنوان مثال در این روش از یک گروه دانش‌آموز خواسته می‌شود تا با استفاده از سه ابزار (۱) ماژیک (۲) آب رنگ و (۳) مداد رنگی نقاشی کنند. برای

تحلیل واریانس، نمرات هر یک از نقاشی‌ها را در ستون جداگانه وارد کرده و دستور: Analyze/General Linear Model/Repeated Measures را اجراء کنید. با این دستور پنجره‌ای به شکل (۷-۵) در صفحه مانیتور ظاهر می‌شود.



شکل (۷-۵) پنجره Repeated Measures Define Factor(s)

در این پنجره تعداد سطوحی که باید با همدیگر مقایسه شوند را در مقابل مربع Number of levels مشخص کنید و سپس دکمه Add را کلیک کنید تا نام و تعداد سطوح آن به جعبه مربوط منتقل شود. سپس دکمه Define را فشار دهید، تا پنجره‌ای مانند شکل (۸-۵) در مانیتور ظاهر شود. در این پنجره به تعداد سطوح مشخص شده در پنجره قبلی، علامت سوال نمایش داده می‌شود که برای هر یک از آنها باید متغیر مورد نظر اختصاص یابد. برای این کار می‌توان با مشخص کردن هر متغیر و کلیک کردن دکمه فلش دار، متغیرهای مورد نظر را یکی پس از دیگری به قسمت مربوط منتقل کرد. با فشار دادن دکمه OK محاسبات در پنجره Output ظاهر می‌شود.



شکل (۸-۵) پنجره تحلیل واریانس با اندازه‌گیری تکراری

## فصل ششم

# آزمونهای غیر پارامتری

### مقدمه:

همانطوری که قبلاً نیز بحث گردید آزمونهای پارامتری علاوه بر اینکه نیاز به داده‌هایی از نوع فاصله‌ای دارند، باید از برخی پیش فرضهای اولیه نیز برخوردار باشند. نرمال بودن توزیع در جامعه و داشتن واریانس یکسان در مواردی که دو یا چند جامعه با هم مورد مقایسه قرار می‌گیرند از جمله این پیش فرضها محسوب می‌شوند. اما در آزمونهای غیر پارامتری چنین پیش فرضهایی مطرح نبوده و زمانی که داده‌ها در سطح اسمی و یا ترتیبی باشند و یا در صورتی که گروههای مورد مطالعه از واریانس نابرابر و یا از چولگی برخوردار باشند باید از آزمونهای غیر پارامتری استفاده کرد. این آزمونها از ویژگیهایی برخوردار هستند که آنها را از آزمونهای پارامتری متمایز کرده و به دلیل سادگی و کاربرد آنها در تحلیل متغیرهای کیفی جایگاه ویژه‌ای در تحقیقات اجتماعی یافته است. عمده‌ترین ویژگیهای این آزمونها را می‌توان بشرح ذیل خلاصه نمود.

۱ - این آزمونها هیچکدام از پیش فرضهای مطرح شده در آزمونهای پارامتری، نظیر نرمال بودن جامعه و یا برابر بودن واریانس گروهها، را مبنا قرار نمی‌دهد و حتی در صورت صادق نبودن مفروضات فوق در خصوص داده‌های فاصله‌ای به منظور استفاده از آزمونهای پارامتری امکان تبدیل

داده‌های فوق به داده‌های غیرپارامتری و رتبه‌ای و محاسبه آزمونهای ناپارامتری وجود دارد.

۲- از آنجا که در این آزمونها از مقادیر رتبه‌ای و حتی داده‌های اسمی استفاده می‌گردد بنابراین محاسبه آنها کار ساده‌ای است، در حالی که امکان استفاده از آزمونهای پارامتری برای تحلیل چنین داده‌هایی وجود ندارد.

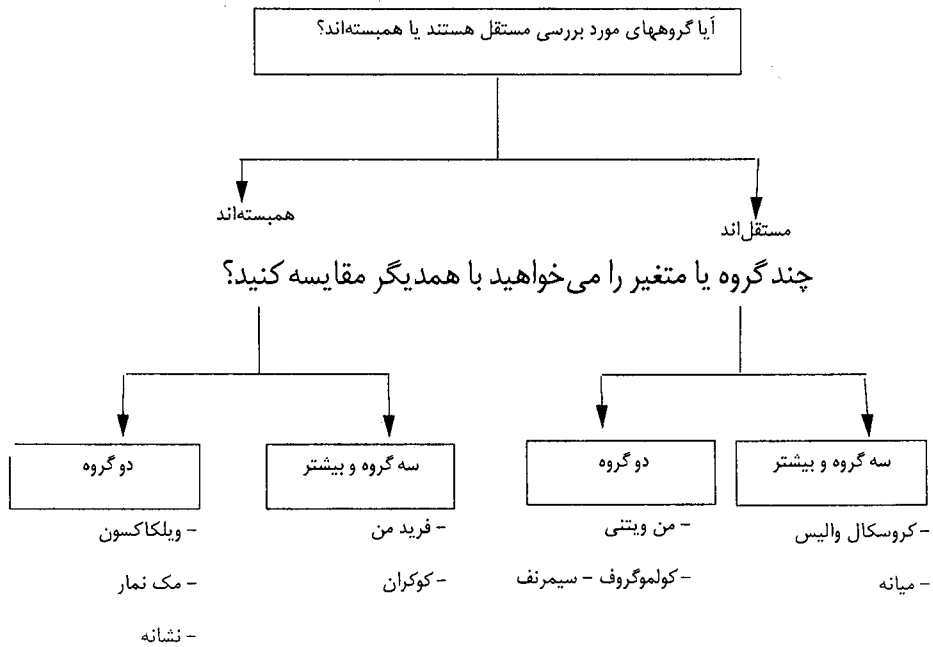
۳- این آزمونها در مقایسه با آزمونهای پارامتری از دقت بالایی برخوردار نمی‌باشند. دلیل آن نیز این است که با تبدیل داده‌های فاصله‌ای به مقادیر رتبه‌ای فواصل واقعی موجود در بین داده‌ها به فواصل یکسان بین رتبه‌ها تبدیل شده و در این فرآیند بخشی از اطلاعات ناپدید می‌گردند. به عبارت دیگر با تبدیل مقادیر اصلی و واقعی به مقادیر رتبه‌ای، بدلی از واقعیت ساخته می‌شود و این بدل بدست آمده به جای واقعیت مورد تجزیه و تحلیل قرار می‌گیرد.

این فصل به طبقه‌بندی، تحلیل و تفسیر آزمونهای غیرپارامتری پرداخته است و نحوه محاسبه هر یک از آنها نیز با استفاده از نرم‌افزار SPSS به تفصیل توضیح داده شده و در موارد ضروری مثالهایی نیز ارائه شده است.

#### انتخاب آزمون غیرپارامتری مناسب:

برای اینکه تصمیم بگیرید که چه آزمونی برای تحلیل داده‌های جمع آوری شده مناسب می‌باشد به درخت تصمیم‌گیری ذیل توجه کنید. در این درخت تصمیم‌گیری تنها با طرح دو سؤال و پاسخ دادن به آنها می‌توان مناسب‌ترین آزمون آماری را انتخاب کرد. سؤال اول این است که آیا گروههای مورد بررسی مستقل از هم هستند یا همبسته‌اند؟ سؤال دوم این است که چند گروه یا متغیر را می‌خواهید با همدیگر مقایسه کنید؟ پاسخ به این سوالات محقق را به انتخاب آزمون یا آزمونهای مناسب هدایت می‌کند. این فرآیند در شکل (۶-۱) نشان داده شده است.

شکل (۱-۶) درخت تصمیم‌گیری برای انتخاب آزمون غیر پارامتری مناسب



## ۱ - آزمون نشانه :

آزمون نشانه<sup>(۱)</sup> یکی از ساده‌ترین آزمونهای غیر پارامتری است. نام این آزمون از علائم مثبت و منفی مشاهدات در یک نمونه گرفته شده است. این آزمون به دو صورت یک نمونه‌ای و دو نمونه‌ای مورد استفاده قرار می‌گیرد. نظر به اینکه آزمون دو نمونه‌ای کاربرد زیادی در تحلیل داده‌ها دارد به همین دلیل تنها به توضیح و تبیین آن اکتفا می‌گردد. این آزمون زمانی که داده‌ها به صورت همبسته باشند مورد استفاده قرار می‌گیرد. در این روش، در صورتی که اولین مقدار مربوط به دومین نمونه ( $X_2$ ) بزرگتر از اولین مقدار مربوط به اولین نمونه ( $X_1$ ) باشد، با علامت مثبت (+) مشخص می‌شود. و اگر اولین مقدار مربوط به  $X_2$  کوچک‌تر از اولین مقدار مربوط به  $X_1$  باشد با علامت منفی

1- Sign Test

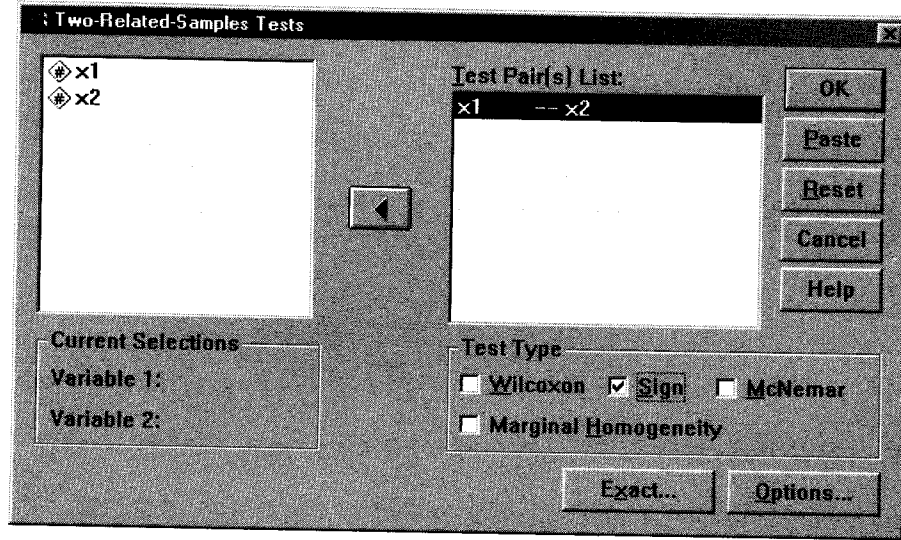
(-) مشخص می‌گردد. در صورتی که مقادیر دو نمونه با هم برابر باشند، جفت‌های فوق تحت عنوان Ties کنار گذاشته می‌شود<sup>(۱)</sup>.

### نحوه محاسبه آزمون نشان با برنامه SPSS:

فرض کنید اطلاعات مربوط به تعداد کالاهای مصنوعی کشف شده توسط دو باستان شناس در ۳۰ روز در یک منطقه گردآوری شده است. می‌خواهید با استفاده از آزمون نشانه، سطح معنی‌داری فرض صفر ( $H_0$ ) را که اکتشافات هر دو باستان شناس با هم برابرند را در مقابل فرضیه‌ای که یافته‌های باستان شناس اول ( $X_1$ ) بیشتر است، آزمایش کنید. برای این کار اکتشافات باستان شناس اول را در یک ستون تحت عنوان  $X_1$  و اکتشافات باستان شناس دوم را در ستون دیگر تحت عنوان  $X_2$  در پنجره Data Editor وارد کنید. سپس دستور: Analyze/Nonparametric Tests/2 Related Samples را اجرا کنید تا پنجره‌ای مانند شکل (۲-۶) ظاهر شود:

در این پنجره ابتدا بر روی متغیر  $X_1$  کلیک کنید تا به جلو: Variable 1 در قسمت Current Selections منتقل گردد، سپس بر روی متغیر  $X_2$  کلیک کنید تا در مقابل: Variable 2 قرار گیرد. با این دستور دکمه فلش دار روشن می‌شود. دکمه فوق را فشار دهید تا دو متغیر به قسمت Test Pair (s) List منتقل گردد. آزمون نشانه (Sign) را از قسمت Test Type انتخاب کرده و دکمه Ok را کلیک کنید تا نتایج در پنجره Output ظاهر شود، (جدول ۶-۱).

۱- در صورتی که تعداد موارد مربوط به نمونه‌های مورد مقایسه با هم برابر نباشند، پس از تعیین زوجهای دو نمونه بصورت تصادفی، مازاد مربوط به نمونه بزرگتر حذف می‌گردند.



شکل (۶-۲) پنجره Two-Related Samples Tests

جدول (۶-۱) نتایج آزمون نشانه

**Frequencies**

		N
X2 - X1	Negative Differences <sup>a</sup>	20
	Positive Differences <sup>b</sup>	5
	Ties <sup>c</sup>	5
	Total	30

- a.  $X2 < X1$
- b.  $X2 > X1$
- c.  $X1 = X2$

**Test Statistics<sup>b</sup>**

	X2 - X1
Exact Sig. (2-tailed)	.004 <sup>a</sup>

- a. Binomial distribution used.
- b. Sign Test



همانطوری که از جدول فوق بر می‌آید، سهم علامتهای منفی مشاهده شده برابر با ۲۰ و سهم علامتهای مثبت مشاهده شده برابر با ۵ می‌باشد. و آزمون نشانه نیز در سطح ۹۹ درصد تفاوت بین اکتشافات دو باستان شناس را تایید می‌کند. بنابراین، فرض صفر که فعالیت هر دو باستان شناس با هم برابرند در سطح ۱ درصد رد شده و فرض مقابل که باستان شناس اول ( $X_1$ ) بهتر عمل کرده است تایید می‌گردد.

### آزمون مک نمار :

آزمون مک نمار<sup>(۱)</sup> یکی از آزمونهای مهم غیرپارامتری است که اغلب در مواردی بکار برده می‌شود که داده‌ها به صورت اسمی و مربوط به دو نمونه مرتبط بهم<sup>(۲)</sup> یا همبسته باشند. این آزمون بویژه در مواردی بکار گرفته می‌شود که بخواهیم نظرات قبلی و بعدی افراد را مورد مقایسه و بررسی قرار دهیم. به عنوان مثال از این روش می‌توان برای سنجش اثر یک برنامه ترویجی و آموزشی بر روی نظرات فراگیران و میزان تغییر نگرش آنها استفاده کرد. این روش نشان می‌دهد که آیا فراگیران پس از اجرای برنامه ترویجی تغییر عقیده داده‌اند یا خیر. برای این منظور ابتدا تعدادی از افراد را به عنوان نمونه انتخاب کرده سپس نظر آنها را (به صورت موافق - مخالف) در رابطه با موضوع مورد بررسی جویا می‌شویم. پس از آن نسبت به اجرای برنامه (برگزاری دوره ترویجی یا جلسه توجیهی) اقدام کرده، سپس نظرات افراد را نسبت به موضوع فوق مجدداً سؤال می‌کنیم. در اینجا فرض صفر ( $H_0$ ) این است که تفاوتی بین نظرات افراد در قبل و پس از اجرای برنامه (دوره ترویجی) وجود ندارد.

### نحوه محاسبه آزمون مک نمار با برنامه SPSS:

فرض کنید نظرات ۳۰ نفر از کشاورزان در مورد استفاده از بذر خاص در قبل از برگزاری دوره ترویجی و پس از آن پرسیده شده است. برای وارد کردن اطلاعات به SPSS و محاسبه آزمون مک

1- Mc Nemar Test

2- Two Related Samples

نمار، نظرات افراد را باید به صورت مخالف یا موافق یا به صورت بلی و خیر در قالب کدهای ۱ و ۲ در دو ستون وارد کنید. سپس دستور Analyze/Non- Parametric Test/2 Related Sample را اجرا کرده تا شکل (۲-۶) در مانیتور ظاهر شود. در این پنجره، دو متغیری که قرار است بررسی شوند را مشخص کنید. متغیر اول در قسمت Current Selections در جلو Variable 1 و متغیر دوم را در جلو Variable 2 قرار دهید. برای وارد کردن این دو متغیر در قسمت Test pair (s) List دکمه فلش دار را فشار دهید و سپس آزمون مک نمار را از قسمت Test Type انتخاب کنید با فشار دادن دکمه OK نتیجه محاسبات در پنجره Output ظاهر می شود (جدول ۲-۶). اگر سطح معنی داری بدست آمده کوچک تر از  $0.05$  باشد، فرض صفر ( $H_0$ ) رد شده و نشان می دهد که برنامه ترویجی (یا جلسه توجیهی) اثر بخش بوده و تفاوت معنی داری بین نظرات افراد در قبل از برنامه ترویجی و پس از آن وجود دارد و این تفاوت در سطح  $0.95$  درصد تایید می گردد و اگر مقدار بدست آمده کوچک تر از  $0.01$

جدول (۲-۶) نتایج آزمون مک نمار

## Crosstabs

## X3 &amp; X4

X3	X4	
	1	2
1	8	2
2	12	8

Test Statistics<sup>b</sup>

	X3 & X4
N	30
Exact Sig. (2-tailed)	.013 <sup>a</sup>

a. Binomial distribution used.

b. McNemar Test

باشد همین تفاوت در سطح ۹۹ درصد معنی‌دار خواهد بود. لازم به توضیح است که روش مک‌نمار به تعداد افرادی که تغییر عقیده نداده‌اند توجهی ندارد و تنها افرادی که تغییر عقیده داده‌اند مد نظر قرار می‌گیرد.

در یک مثال دیگر، فرض کنید، نظرات ۱۰۰۰ نفر در رابطه با خرید و عدم خرید یک کالا قبل و بعد از برگزاری برنامه تبلیغاتی و معرفی کالا پرسیده شده است (جدول ۳-۶). با استفاده از آزمون مک‌نمار مشخص کنید که آیا برنامه تبلیغاتی در تغییر نگرش مشتریان مؤثر بوده است یا خیر.

جدول (۳-۶) نظرات افراد در قبل و بعد از اجرای برنامه تبلیغات

پس از تبلیغات	قبل از تبلیغات	
۱ - خواهم خرید	۳۰۰	۲ - نخواهم خرید
۲ - نخواهم خرید	۱۰۰	۲۰۰

برای محاسبه مک‌نمار از طریق برنامه SPSS زمانی که داده‌ها به صورت جدول بالا تنظیم شده باشد ابتدا باید مقادیر جدول را به عنوان فراوانی به برنامه SPSS معرفی کرد. برای اینکار باید داده‌های جدول فوق به صورت ذیل در سه ستون در پنجره Data Editor وارد گردد.

جدول (۴-۶) نحوه ورود داده‌ها به پنجره Data Editor

فراوانی	پس از تبلیغات	قبل از تبلیغات
۳۰۰	۱	۱
۲۰۰	۲	۱
۱۰۰	۱	۲
۴۰۰	۲	۲

پس از اینکه داده‌ها به صورت فوق در پنجره Data Editor وارد شد. دستور Data/Weight Cases را اجرا کنید و سپس متغیر فراوانی را به قسمت مربوط به Weight Cases by منتقل و دکمه OK را فشار دهید. سپس دستور Analyze/Non- Parametric Tests/2 Related Samples را اجرا کرده و متغیر «قبل» از تبلیغات را کلیک کنید تا در جلو Variable 1 قرار گیرد و سپس متغیر «پس از تبلیغات» را کلیک کنید تا در جلو Variable 2 قرار گیرد. پس از آن دکمه فلش دار را کلیک کنید و آنها را به قسمت Test Pair (s) List منتقل کرده و آماره مک نما را نیز انتخاب و دکمه OK را فشار دهید. تا نتایج در پنجره Output ظاهر شود جدول (۵-۶).

جدول (۵-۶) نتایج آزمون مک نما

### Crosstabs

قبل & بعد

قبل	بعد	
	1	2
1	300	200
2	100	400

### Test Statistics<sup>b</sup>

	قبل & بعد
N	1000
Chi-Square <sup>a</sup>	32.670
Asymp. Sig.	.000

a. Continuity Corrected

b. McNemar Test

آزمون ویلکاکسون:

در بسیاری از پژوهشهایی که نمونه‌ها به صورت جفت شده و همبسته هستند، ممکن است

محقق بخواهد تا هم، جهت تغییر و هم، میزان تغییر را مورد بررسی قرار دهد. برای این منظور آزمون ویلکاکسون<sup>(۱)</sup> تست مناسبی است. این آزمون برخلاف مک نماز علاوه بر تعیین جهت تغییرات میزان تغییرات، را نیز مشخص می‌کند. داده‌های مورد استفاده در این آزمون حداقل باید در سطح ترتیبی باشند. فرض کنید نظر تعدادی از مشتریان در رابطه با دو نوع کالای مشابه اما با مارکهای متفاوت از نقطه نظر کیفیت محصول سؤال شده است و امتیازات آنها در دو ستون وارد برنامه SPSS شده است. فرض صفر ( $H_0$ ) این است که تفاوت معنی داری از نقطه نظر کیفیت بین دو نوع کالا وجود ندارد. برای آزمون معنی داری می‌توان از آزمون ویلکاکسون استفاده کرد. این روش آماری تفاوت بین مقادیر و امتیازات دو نوع کالا را محاسبه کرده و این تفاوت‌های بدست آمده را با لحاظ کردن مقادیر قدر مطلق آنها رتبه بندی می‌کند. بطوریکه برای کوچکترین تفاوت رتبه ۱ به تفاوت بعدی رتبه ۲ و غیره می‌دهد. هر یک از مواردی که تفاوت بین پاسخها صفر باشد از تحلیل خارج می‌کند. سپس رتبه‌های دارای علائم + و - از هم تفکیک شده و با همدیگر جمع می‌شوند.

تست ویلکاکسون که آن را با علامت T نمایش می‌دهند، بر این استدلال استوار است که اگر تفاوتی بین دو مجموعه از مقادیر وجود نداشته باشد، تقریباً به همان میزان که تفاوت + کوچک وجود دارد، تفاوت - کوچک نیز وجود خواهد داشت. همچنین حدوداً به همان تعداد که تفاوت + بزرگ وجود داشته باشد به همان میزان نیز تفاوت - بزرگ وجود خواهد داشت. بنابراین مجموع رتبه‌ها برای تفاوت‌های + تقریباً برابر با مجموع رتبه‌ها برای تفاوت‌های - خواهد بود. اگر مجموع اختلافات + تفاوت زیادی با مجموع اختلافات - داشته باشد می‌توان قضاوت کرد که تفاوت معنی داری بین دو مجموعه مقادیر وجود دارد.

#### نحوه محاسبه آزمون ویلکاکسون با برنامه SPSS:

برای محاسبه آزمون ویلکاکسون براساس توضیحاتی که در بالا ارائه گردید، نتایج ارزیابی

افراد از کالای نوع (الف) را در یک ستون و نوع (ب) را در ستون دیگر (که در سطح ترتیبی ارزیابی شده‌اند) در پنجره Data Editor وارد کنید و سپس دستور: Analyze/Non- Parametric Test/2 Related Sample را اجرا کنید. با اجرای این دستور متغیرهای مورد نظر در جعبه سمت چپ پنجره لیست می‌شوند. روی متغیر اول کلیک کنید تا به جلو Variable 1 در قسمت Current Selections منتقل شود. سپس روی متغیر دوم نیز کلیک کنید تا به جلو Variable 2 منتقل شود. با فشار دادن دکمه فلش دار دو متغیر فوق به صورت جفت شده به جعبه سمت راست منتقل شده و آماده تجزیه و تحلیل می‌گردد. آماره ویلکاکسون را انتخاب و دکمه OK را فشار دهید تا نتایج در پنجره Output ظاهر شود (جدول ۶-۶).

جدول (۶-۶) نتایج آزمون ویلکاکسون

## Ranks

	N	Mean Rank	Sum of Ranks
X2 - X1 Negative Ranks	20 <sup>a</sup>	13.80	276.00
Positive Ranks	5 <sup>b</sup>	9.80	49.00
Ties	5 <sup>c</sup>		
Total	30		

a.  $X2 < X1$ b.  $X2 > X1$ c.  $X1 = X2$ Test Statistics<sup>b</sup>

	X2 - X1
Z	-3.112 <sup>a</sup>
Asymp. Sig. (2-tailed)	.002

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

نتایج بدست آمده از طریق نرم‌افزار SPSS نشان می‌دهد که تفاوت معنی‌داری بین کیفیت دو نوع محصول وجود دارد. بنابراین فرض  $(H_0)$  که معتقد است کیفیت دو نوع کالا با دو مارک متفاوت، یکسان می‌باشد رد می‌شود.

#### آزمون فریدمن:

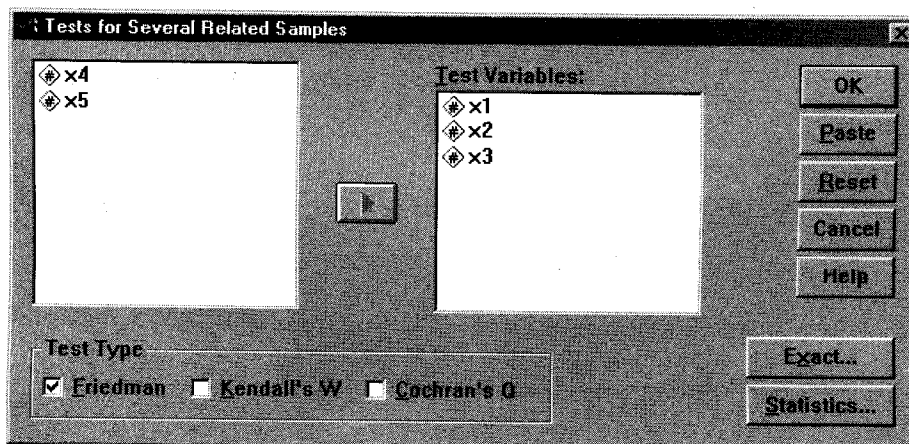
آزمون فریدمن<sup>(۱)</sup> یکی دیگر از آزمونهای غیر پارامتری است. این آزمون در واقع معادل آزمون  $F$  در روشهای پارامتری می‌باشد. اما در اینجا بر خلاف آزمون  $F$ ، فرض توزیع نرمال و برابر بودن واریانس ضرورتی ندارد.

این روش برای مقایسه سه گروه یا بیشتر از سه گروه همبسته بکار می‌رود. در این روش جدولی ایجاد می‌شود که در آن هر ستون اطلاعات یک موضوع را شامل شده و اطلاعات هر ستون رتبه بندی می‌شوند و بعد مجموع رتبه‌های هر ستون محاسبه می‌گردد. آزمون فریدمن در صدد است تا بداند آیا مجموع رتبه‌های هر ستون با چیزی که مورد انتظار می‌باشد بسیار متفاوت است یا خیر. اگر تفاوت زیادی بین شرایط مورد مطالعه وجود نداشته باشد در این صورت مجموع رتبه‌ها کم و بیش مثل هم خواهند بود.

#### روش محاسبه آزمون فریدمن با برنامه SPSS:

فرض کنید نظرات ۳۰ نفر از فراگیران را درخصوص سه روش مختلف تدریس جویا شده‌اید. پاسخها نیز از ۱ (بسیار نامناسب) تا ۵ (بسیار مناسب) امتیاز بندی شده‌اند. امتیازات هر روش تدریس را در یک ستون در پنجره Data Editor وارد کنید. برای محاسبه آزمون فریدمن منوی Analyze/Non Parametric Test/K Related Samples را انتخاب کنید. با این دستور پنجره‌ای مانند شکل (۳-۶) ظاهر می‌شود.

1- Friedman test



شکل (۶-۳) پنجره Tests for several Related samples

در این پنجره سه متغیر  $X_1$ ،  $X_2$  و  $X_3$  را انتخاب و به قسمت Test variables منتقل کنید. سپس آزمون فریدمن را نیز از قسمت Test Type انتخاب و دکمه Ok را فشار دهید تا نتایج در پنجره Output ظاهر شود، (جدول ۶-۷).

جدول (۶-۷) نتایج آزمون فریدمن

Ranks

	Mean Rank
X1	2.35
X2	1.62
X3	2.03

Test Statistics<sup>a</sup>

N	30
Chi-Square	9.740
df	2
Asymp. Sig.	.008

a. Friedman Test



همانطوری که از نتایج آزمون برمی‌آید، تفاوت معنی‌داری بین سه روش تدریس وجود دارد و بنابراین فرض صفر ( $H_0$ ) رد می‌شود.

### آزمون کوکران:

آزمون کوکران<sup>(۱)</sup> به عنوان یکی دیگر از روشهای غیر پارامتری، در واقع تعمیم یافته آزمون مک نمار است. با این تفاوت که این روش برای مواردی که تعداد گروهها یا تکرار سه یا بیشتر از سه باشد بکار می‌رود. داده‌های این آزمون به صورت اسمی می‌باشند و وجود تفاوت بین نظرات افراد را مورد بررسی قرار می‌دهد.

در این روش نیز مانند روش مک نمار نظرات افراد در زمانهای مختلف پرسیده می‌شود. فرض کنید قبل از برگزاری یک دوره آموزشی نظرات فراگیران در رابطه با موضوع مورد نظر بررسی می‌شود (و نوع موافقت یا مخالفت آنها سنجیده می‌شود)، در مرحله دوم پس از اینکه دوره آموزشی برگزار گردید، مجدداً نظرات افراد جمع‌آوری شده و در مرحله سوم پس از بکارگیری مباحث تدریس شده در عمل (مانند بکارگیری روش کشت جدید) برای بار سوم نظرات افراد در رابطه با موضوع فوق پرسیده می‌شود.

### نحوه محاسبه آزمون کوکران با برنامه SPSS:

زمانی که نظرات افراد براساس توضیحات بالا گردآوری شد در قالب کدهای ۱ و ۲ در سه ستون مستقل در پنجره Data Editor وارد می‌شود. سپس با استفاده از دستور Analyze/Non-Parametric Test/K Related Samples همان مراحلی که در تست فریدمن انجام گرفت در اینجا نیز دنبال می‌شود. با انجام مراحل فوق نتایج در پنجره Output ظاهر می‌گردد (جدول ۶-۸).

## جدول (۶-۸) نتایج آزمون کوکران

## Frequencies

	Value	
	1	2
X4	20	10
X5	10	20
X6	13	17

## Test Statistics

N	30
Cochran's Q	9.875 <sup>a</sup>
df	2
Asymp. Sig.	.007

a. 1 is treated as a success.

آزمون من و ويتنی<sup>(۱)</sup>:

این آزمون برای مقایسه میانگین‌های دو جامعه مستقل زمانی که داده‌ها به صورت رتبه‌ای، یا ترتیبی باشند مورد استفاده قرار می‌گیرد. آزمون من و ويتنی برای محاسبه تفاوت‌های موجود در بین دو گروه، مقادیر مربوط به هر دو نمونه را به صورت یکجا و بدون توجه به اینکه هر مقدار به کدام گروه تعلق دارد رتبه‌بندی می‌کند. در موقع رتبه‌بندی چنانچه مقادیر تکراری وجود داشته باشد رتبه‌های مربوط به آنها را با همدیگر جمع کرده و بر تعداد مقادیر مشترک تقسیم و رتبه مشترکی را برای همه آنها لحاظ می‌کند. به عنوان مثال اگر مقادیر مربوط به رتبه‌های ۵، ۶ و ۷ با هم برابر باشند در این صورت اعداد ۵، ۶ و ۷ را با هم جمع کرده و بر ۳ تقسیم می‌کند که حاصل آن عدد ۶ بدست می‌آید. در این حالت رتبه ۶ را برای هر سه مقدار فوق در نظر گرفته و رتبه عدد بعدی را از عدد ۸ آغاز می‌کند.

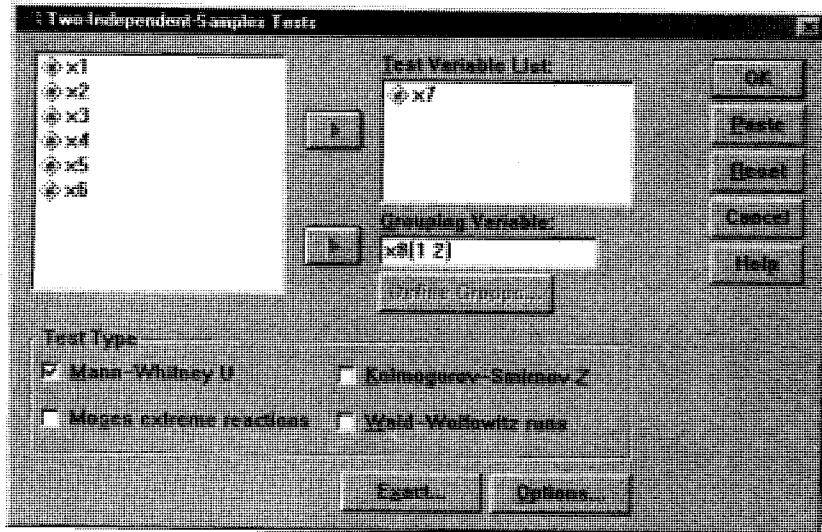
1- Mann-Whitney Test

پس از اینکه رتبه‌بندی به اتمام رسید، مجموع رتبه‌های هر دو گروه را بطور جداگانه محاسبه می‌کند.

#### نحوه محاسبه آزمون من ویتنی با برنامه SPSS:

فرض کنید دو گروه ۳۰ نفره از فراگیران با دو روش متفاوت آموزش دیده و نتیجه ارزیابی آنها از دوره‌های فوق در قالب امتیازات ۱ تا ۵ گردآوری شده است. از آنجا که در این آزمون گروهها مستقل می‌باشند بنابراین، برای وارد کردن داده‌ها به صفحه Data Editor در برنامه SPSS باید داده‌های مربوط به نتیجه ارزیابی هر دو گروه را در یک ستون وارد کرده و در ستون مقابل با دادن کدهای ۱ و ۲ دو گروه را از هم تفکیک کنید. پس از آن دستور: Analyze/Non- Parametric Test/2 Independent Sample را اجرا کنید، تا پنجره مربوط باز شود (شکل ۴-۶).

در این پنجره متغیر مربوط به داده‌های اصلی را در قسمت Test Variable List قرار داده و متغیر گروه‌بندی را در قسمت Grouping Variable قرار دهید. پس از آن با فشار دادن دکمه Define Groups کد گروهها را به صورت ۱ و ۲ تعریف کنید. سپس آزمون من ویتنی را انتخاب و دکمه OK را فشار دهید تا نتیجه محاسبات در پنجره Output ظاهر شود (جدول ۶-۹). اگر سطح معنی‌داری بدست آمده کوچک‌تر از  $0/05$  بود به احتمال ۹۵ درصد تفاوت معنی‌داری بین دو روش آموزشی وجود دارد و اگر کوچک‌تر از  $0/01$  بود به احتمال ۹۹ درصد این تفاوت معنی دار خواهد بود.



شکل (۶-۴) پنجره Two Independent samples Tests

جدول (۶-۹) نتایج آزمون من ویتنی

Ranks

X8	N	Mean Rank	Sum of Ranks
X7 آموزش با معلم	35	31.21	1092.50
X7 آموزش با تلویزیون	25	29.50	737.50
Total	60		

Test Statistics<sup>a</sup>

	X7
Mann-Whitney U	412.500
Wilcoxon W	737.500
Z	-.381
Asymp. Sig. (2-tailed)	.703

a. Grouping Variable: X8

همانطوری که از جدول فوق برمی‌آید تفاوت معنی‌داری بین دو روش آموزش وجود ندارد.

#### آزمون کولموگروف سیمرنف:

چنانچه در بحث کای اسکویر گفته شد اگر فراوانیهای مورد انتظار بیش از ۲۰ درصد خانه‌های جدول، کمتر از ۵ باشد، در این صورت نمی‌توان از فرمول کاسکویر استفاده کرد. این مشکل معمولاً زمانی پیش می‌آید که حجم نمونه کمتر از ۵۰ باشد و یا تعداد خانه‌های جدول بیشتر باشد. در چنین حالتی می‌توان از تست کولموگروف - سیمرنف<sup>(۱)</sup> استفاده کرد. اساس این آزمون مقایسه فراوانی تجمعی نسبی مشاهده شده با فراوانی تجمعی نسبی مورد انتظار می‌باشد.

#### روش محاسبه آزمون کولموگروف - سیمرنف با برنامه SPSS:

فرض کنید دو گروه ۳۰ نفره از فراگیران با دوروش متفاوت آموزش دیده و نتیجه ارزیابی آنها در قالب امتیازات ۱ تا ۵ گردآوری شده است. برای محاسبه این آزمون ابتدا مقادیر هر دو نمونه را در یک ستون وارد کرده سپس در ستون بعدی با استفاده از کدهای ۱ و ۲ دو گروه را مشخص کنید. پس از آن دستور Analyze/Non Parametric Tests/2Independent Samples را انتخاب کنید. تا شکل (۴-۶) در مانیتور ظاهر شود. سپس متغیر مورد مطالعه را در قسمت Test variable List و متغیری که برای گروه بندی مقادیر فوق ترتیب داده‌اید را در قسمت Grouping Variable وارد کرده و با فشار دادن دکمه Define Groups گروه‌های مورد نظر را تعریف کنید. پس از آن آماره کولموگروف - سیمرنف را انتخاب کرده و دکمه OK را فشار دهید. تا نتایج در پنجره Output ظاهر شود (جدول ۶-۱۰).

1- Kolmogrov-Smirnov test

جدول (۶-۱۰) نتایج آزمون کولموگروف - سیمرنف

## Frequencies

X8	N
X7 آموزش با معلم	32
آموزش با تلویزیون	28
Total	60

Test Statistics<sup>a</sup>

		X7
Most Extreme Differences	Absolute	.482
	Positive	.000
	Negative	-.482
Kolmogorov-Smirnov Z		1.863
Asymp. Sig. (2-tailed)		.002

a. Grouping Variable: X8

آزمون کروسکال والیس<sup>(۱)</sup>:

این آزمون در واقع معادل تحلیل واریانس یک طرفه می‌باشد، اما برخلاف آن نیازی به مفروضات آن نظیر اینکه نمونه‌ها از یک جامعه نرمال بدست آمده باشد و یا اینکه انحراف معیار یکسانی داشته باشند وجود ندارد. از سوی دیگر این آزمون یادبودی از روش من ویتنی می‌باشد که مقادیر نمونه‌ها را یکجا به صورت نزولی یا صعودی مرتب می‌کند. آزمون کروسکال والیس که آن را  $H_1$  نمایش می‌دهند زمانی مورد استفاده قرار می‌گیرد که تعداد نمونه‌ها بیش از دو گروه باشد.

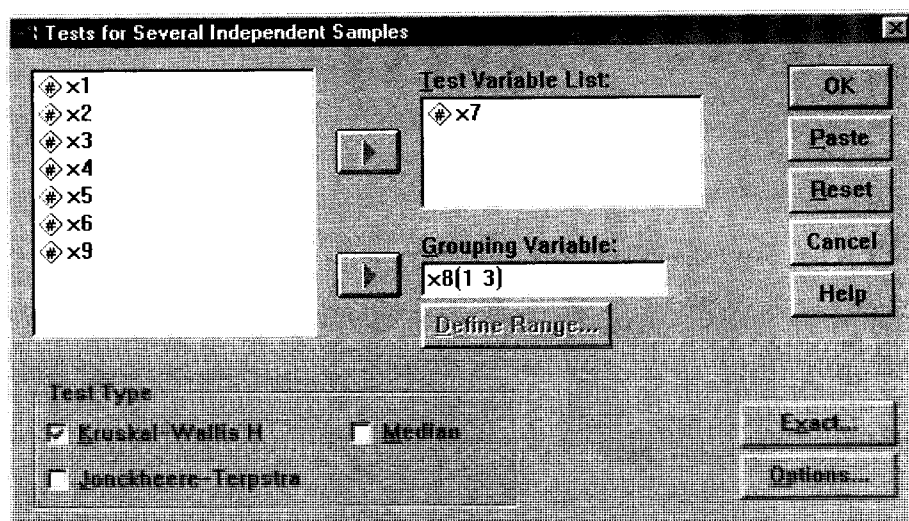
## روش محاسبه آزمون کروسکال والیس با برنامه SPSS:

فرض کنید از ۹۰ نفر از دانشجویان سه رشته مختلف خواسته شده است تا کیفیت برنامه‌های

1- Kruskal- Wallis Test

آموزشی دانشکده را ارزیابی کنند. امتیازات ارائه شده توسط افراد فوق از ۱ (بسیار ضعیف) تا ۵ (بسیار قوی) در نوسان بوده است. با استفاده از آزمون کروسکال والیس فرض صفر ( $H_0$ ) را که بین نظرات دانشجویان رشته‌های مختلف تفاوت معنی‌داری وجود ندارد را آزمون کنید.

باتوجه به اینکه گروهها از هم مستقل هستند، بنابراین برای وارد کردن داده‌ها در پنجره Data Editor باید امتیازات فوق در یک ستون وارد شده و در ستون دیگر رشته‌های مختلف بوسیله کدهای متفاوت از هم تفکیک شوند. (به عنوان مثال باید برای رشته A کد ۱، رشته B کد ۲ و رشته C کد ۳ اختصاص یابد). پس از وارد کردن داده‌ها دستور Analyze/Non- Parametric Test/K Independent Samples را اجرا کنید تا پنجره‌ای مانند شکل (۵-۶) ظاهر شود.



شکل (۵-۶) پنجره Tests for several Independent samples

در این پنجره متغیر مربوط به امتیازات را در قسمت Test Variable list و متغیری که در آن رشته‌های مختلف کدبندی شده‌اند را در قسمت Grouping Variable قرار داده سپس دکمه

Define Range را فشار داده و در قسمت حداقل عدد ۱ و در قسمت حداکثر عدد ۳ را وارد کنید و دکمه Continue را فشار دهید. سپس آزمون کروسکال والیس را انتخاب و دکمه OK را کلیک کنید تا نتایج در پنجره Output ظاهر شود (جدول ۶-۱۱). اگر سطح معنی داری کمتر از ۰/۰۵ باشد تفاوت بین ارزیابی دانشجویان از کیفیت برنامه‌های آموزشی در سطح ۹۵ درصد و اگر کوچک‌تر ۰/۰۱ باشد در سطح ۹۹ درصد به اثبات می‌رسد. در این صورت فرض صفر رد می‌شود. نتایج جدول فوق نشان می‌دهد که تفاوت ارزیابی در سطح ۹۹ درصد معنی دار است.

این آزمون اگرچه وجود تفاوت بین نظرات گروههای مختلف را نشان می‌دهد اما مشخص نمی‌کند که این تفاوت در بین کدام یک از گروهها وجود دارد.

جدول (۶-۱۱) نتایج محاسبات آزمون کروسکال والیس

## Ranks

X9	N	Mean Rank
X7 ترویج	32	57.64
زراعت	28	31.43
گیاهپزشکی	30	45.68
Total	90	

Test Statistics<sup>a,b</sup>

	X7
Chi-Square	15.548
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: X9



**آزمون میانه :**

آزمون میانه<sup>(۱)</sup> یکی دیگر از روشهای غیر پارامتری است که برای مقایسه سه یا بیشتر از سه گروه مورد استفاده قرار می‌گیرد. اطلاعات مورد نیاز در این روش باید در سطح رتبه‌ای بوده و حتی الامکان داده‌ها هم رتبه نباشند. زیرا اگر میانه مشترک بین گروهها جزو نمرات تکراری باشد در این صورت تشخیص تفاوت گروهها با مشکل مواجه می‌گردد<sup>(۲)</sup>.

روش میانه، مقادیر گروههای مورد بررسی را با هم ادغام کرده و یک میانه مشترک بین آنها تعیین و سپس تعداد مواردی که در هر گروه بطور جداگانه در بالا و پایین میانه مشترک قرارداد را بدست می‌آورد و پس از آن تفاوت بین گروهها را مشخص می‌کند. در این روش ضرورتی ندارد که تعداد گروهها با هم برابر باشند.

**نحوه محاسبه آزمون میانه با برنامه SPSS :**

فرض کنید می‌خواهیم بدانیم که آیا سرعت عمل کارگران سه شیفت مختلف یک کارگاه خیاطی با هم متفاوت است یا خیر؟ برای این کار اطلاعات مربوط به تعداد شلوار دوخته شده توسط ۴۰ کارگر (از سه شیفت مختلف) در یک هفته گردآوری شده است. از آنجا که آزمون میانه برای بررسی گروههای مستقل مورد استفاده قرار می‌گیرد بنابراین برای ورود داده‌ها به پنجره Data Editor، باید اطلاعات فوق در یک ستون ثبت گردند و در ستون دیگر بوسیله کدهای ۱، ۲ و ۳ شیفت‌های صبح، عصر و شب تعریف شوند. پس از تعریف داده‌ها باید دستور: Analyze/Nonparametric Tests/ K Independent Samples اجرا گردد. با این دستور شکل (۵-۶) در مانیتور ظاهر می‌شود. در این پنجره متغیری که قرار است مورد تجزیه و تحلیل قرار گیرد را در قسمت Test Variable list و متغیری که در آن گروهها تعریف شده‌اند را در قسمت

1- Median Test

۲- منصورفر، کریم. «جزوه درسی آمار ناپارامتری» دانشگاه تهران، ۱۳۷۹.

Grouping variable قرار دهید. سپس دکمه Define Range را فشار داده و مقادیر حداقل و حداکثر را با اعداد ۱ و ۳ تعریف کنید و دکمه Continue را کلیک کنید تا به پنجره اصلی باز گردید. در این پنجره آزمون Median را از قسمت Test Type مشخص کرده و دکمه Ok را فشار دهید تا محاسبات در پنجره Output ظاهر شود، (جدول ۶-۱۲).

جدول (۶-۱۲) نتایج آزمون میانه

## Frequencies

تولید	شیف	شیفت		
		صبح	عصر	شب
> Median	تولید	12	6	2
<= Median		1	8	11

Test Statistics<sup>b</sup>

	تولید
N	40
Median	37.50
Chi-Square	15.824 <sup>a</sup>
df	2
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.5.

b. Grouping Variable: شیفت

همانطوریکه از جدول (۶-۱۲) برمی آید کارگرانی که در سه شیفت مختلف کار می کنند از نظر

سرعت عمل و کارایی تفاوت معنی دار با همدیگر دارند. زیرا سطح معنی داری بدست آمده کمتر از

۰/۰۱ می باشد.

## فصل هفتم

# طبقه‌بندی روش‌های تحلیل چند متغیره

مقدمه :

انتخاب روش مناسب برای تجزیه و تحلیل داده‌ها مهمترین قدم در تحلیل داده‌های گردآوری شده محسوب می‌گردد. متناسفانه به دلیل عدم آشنایی با منطق حاکم بر روش‌های مختلف آماری چند متغیره، به اشتباه یکی بجای دیگری مورد استفاده قرار می‌گیرد و تکنیک‌های مختلف بطور نابجا بکار گرفته می‌شوند. این فصل در صدد است تا براساس منطق و اصول حاکم بر هر یک از روش‌های تحلیل چند متغیره، به طبقه‌بندی آنها بپردازد، تا محققان را در انتخاب روش یا روش‌های مناسب یاری دهد.

### طبقه‌بندی روش‌های تحلیل چند متغیره:

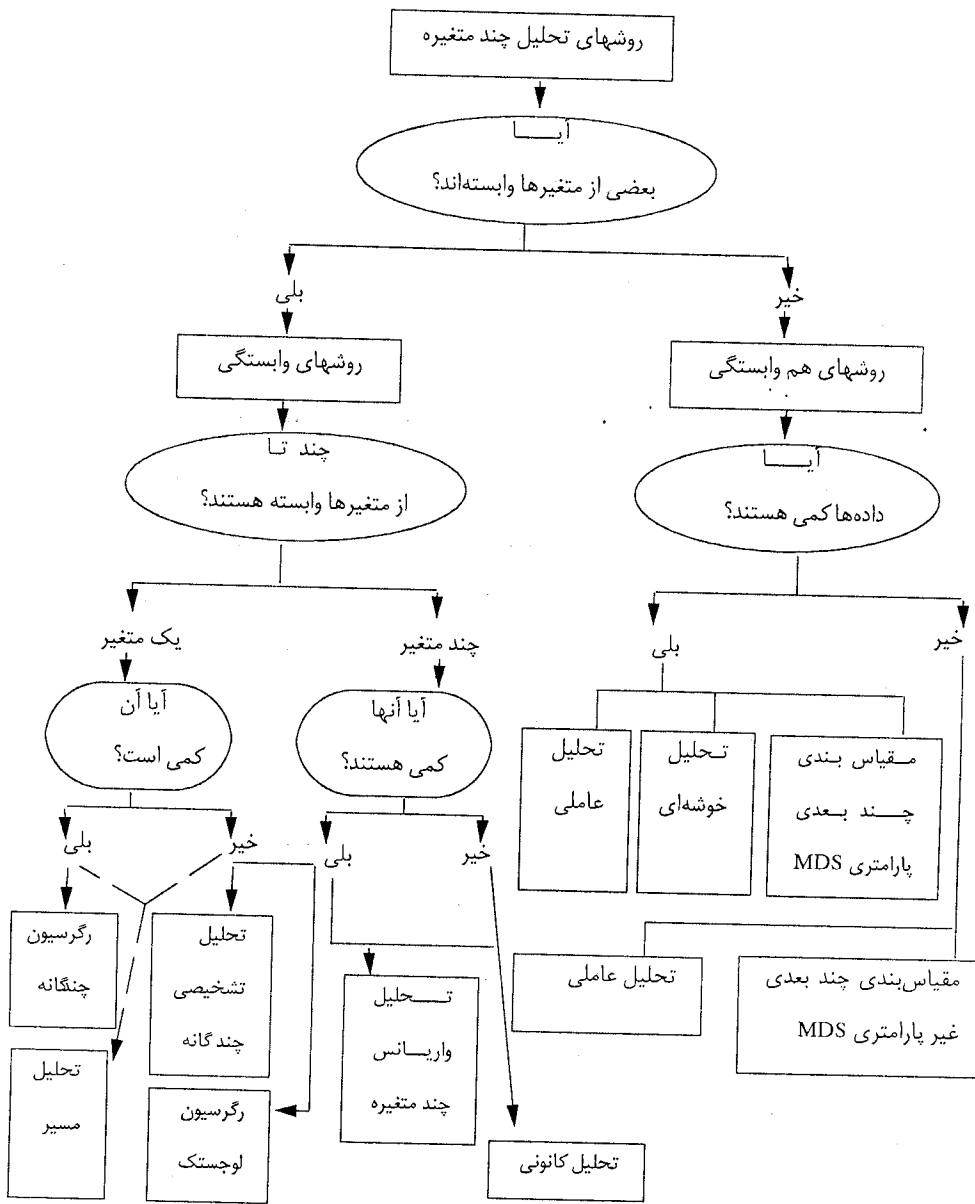
به منظور آشنایی با نحوه بکارگیری روش‌های چند متغیره در تجزیه و تحلیل داده‌ها، درخت تصمیم‌گیری برای استفاده اصولی و متناسب با نوع داده‌ها و هدف تجزیه و تحلیل در ذیل ارائه شده است. این طبقه‌بندی بر مبنای سؤالات و مفروضات خاصی انجام گرفته است که این مفروضات و قضاوت‌ها به شرح ذیل می‌باشند.

- ۱- آیا داده‌های مورد نظر قابل تقسیم به متغیرهای مستقل و متغیرهای وابسته هستند؟
- ۲- اگر چنین تقسیم‌بندی امکانپذیر است، چند متغیر از موارد فوق، متغیر وابسته می‌باشند؟
- ۳- متغیرهای موجود در تجزیه و تحلیل در چه سطحی اندازه‌گیری شده‌اند؟
- پاسخ این سه سؤال محقق را در تصمیم‌گیری برای انتخاب مناسب‌ترین تکنیک تجزیه و تحلیل چند متغیره کمک می‌کند. پاسخ سؤال اول مشخص می‌کند که آیا باید از تکنیکهای وابستگی<sup>(۱)</sup> استفاده کرد یا باید از تکنیکهای هم‌وابستگی<sup>(۲)</sup> بهره گرفت.
- روشهای وابستگی تکنیکهایی هستند که در آن یک یا چند متغیر به عنوان متغیرهای وابسته لحاظ شده و تغییرات آن براساس متغیرهای مستقل دیگر مورد بررسی و تبیین قرار می‌گیرد. مانند روش رگرسیون چندگانه یا تحلیل واریانس چند متغیره و غیره. روشهای هم‌وابستگی تکنیکهایی هستند که در آن متغیرهای مستقل و وابسته وجود ندارد، بلکه در این روشها مجموعه متغیرها بطور همزمان و با هم مورد تجزیه و تحلیل قرار می‌گیرند. مانند روش تحلیل عاملی.
- براساس سه سؤال فوق روشهای چند متغیره تحلیل داده‌ها را می‌توان براساس شکل (۷-۱) طبقه‌بندی کرد. برای فهم بهتر اصول طبقه‌بندی این تکنیک‌ها هر یک از آنها به صورت جداگانه مورد بررسی قرار می‌گیرند. برای این کار ابتدا از روشهای وابستگی آغاز می‌کنیم. تکنیکهای معمول در این روش را می‌توان بر مبنای دو ملاک: (۱) تعداد متغیرهای وابسته و (۲) سطح یا مقیاس اندازه‌گیری آنها طبقه‌بندی کرد. براساس تعداد متغیرهای وابسته تکنیکهای موجود در روش وابستگی را می‌توان به تکنیکهای یک متغیره وابسته و چند متغیره وابسته تقسیم کرد. هر یک از این روشها مجدداً می‌توانند براساس مقیاس اندازه‌گیری به دو بخش پارامتری و غیرپارامتری تقسیم شوند. اگر موضوع مورد تجزیه و تحلیل دارای یک متغیر وابسته از نوع پارامتری باشد روش مناسب برای تجزیه و تحلیل آن تحلیل رگرسیون چندگانه<sup>(۳)</sup> می‌باشد. چنانچه داده‌های فوق دارای یک متغیر وابسته باشد اما متغیر

1- Dependence Techniques

2- Interdependence Techniques

3- Multiple Regression Analysis



شکل (۱-۷) طبقه بندی روشهای تحلیل چند متغیره

فوق از نوع غیرپارامتری دو وجهی باشد روش مناسب رگرسیون لوجستیک و تحلیل تشخیصی چندگانه<sup>(۱)</sup> است. چنانچه متغیر وابسته غیر پارامتری چند وجهی باشد باید از تحلیل تشخیصی استفاده شود. در تحلیل تشخیصی متغیر وابسته از نوع غیرپارامتری و متغیرهای مستقل از نوع پارامتری هستند. اما چنانچه متغیرهای وابسته و مستقل هر دو غیرپارامتری باشند روش مناسب برای تجزیه و تحلیل داده‌ها رگرسیون لوجستیک<sup>(۲)</sup> می‌باشد. تحلیل مسیر<sup>(۳)</sup> روشی است که هم برای داده‌های پارامتری و هم غیر پارامتری سازگار می‌باشد.

چنانچه متغیرهای مربوط به موضوع مورد بررسی دارای چند متغیر وابسته باشد، در این صورت دو تکنیک دیگر برای تجزیه و تحلیل مناسب خواهند بود. اگر متغیرهای وابسته پارامتری باشند تحلیل واریانس چند متغیره<sup>(۴)</sup> یا تحلیل همبستگی کانونی<sup>(۵)</sup> مناسب خواهد بود. اما چنانچه متغیرهای وابسته مورد نظر غیرپارامتری باشند می‌توان از طریق تبدیل آنها به متغیرهای مجازی<sup>(۶)</sup> کدبندی شده به صورت (۰-۱) از روش تحلیل کانونی<sup>(۷)</sup> استفاده کرد.

در مقابل تکنیکهای وابستگی<sup>(۸)</sup>، روشهای هم وابستگی<sup>(۹)</sup> برمبنای وجود یا عدم وجود متغیرهای وابسته مورد تجزیه و تحلیل قرار نمی‌گیرند. زیرا در این روشها کلیه متغیرها بطور همزمان و با همدیگر بررسی می‌شوند. این روشها از نظر پارامتری یا غیرپارامتری بودن داده‌ها به دو گروه جداگانه تقسیم می‌شوند. در این روش، تکنیکهای تحلیل عاملی<sup>(۱۰)</sup> و تحلیل خوشه‌ای<sup>(۱۱)</sup> از جمله روشهای هم وابستگی پارامتری<sup>(۱۲)</sup> محسوب می‌شوند، در حالی که داده‌های غیرپارامتری از طریق کدبندی متغیر مجازی (به صورت ۰-۱) قابل بکارگیری در روشهای تحلیل عاملی و تحلیل خوشه‌ای

1- Multiple Discriminant Analysis

2- Logistic Regression

3- Path Analysis

4- Multivariate Analysis of Variance

5- Canonical Correlation Analysis

6- Dummy Variable

7- Canonical Analysis

8- Dependence Methods

9- Interdependence Methods

10- Factor Analysis

11- Cluster Analysis

12- Parametric Interdependence Techniques

هستند. از سوی دیگر هم داده‌های پارامتری و هم غیرپارامتری قابل استفاده در تکنیک مقیاس بندی چند بعدی<sup>(۱)</sup> می‌باشند.

روشهای وابستگی را می‌توان براساس نوع و طبیعت داده‌های مربوط به متغیرهای مستقل و وابسته نیز طبقه بندی کرد. این طبقه‌بندی که در جدول (۱-۷) ارائه شده است، انتخاب تکنیکهای تجزیه و تحلیل چندگانه را متناسب با نوع داده‌ها تسهیل می‌کند.

**جدول (۱-۷) تکنیکهای تجزیه و تحلیل چندگانه براساس طبیعت داده‌ها و نوع متغیر**

متغیر وابسته			
پارامتری	غیر پارامتری		
تحلیل رگرسیون چندگانه	تحلیل تشخیص چندگانه	پارامتری	متغیر مستقل
تحلیل همبستگی کانونی	همبستگی کانونی با متغیرهای مجازی	غیر پارامتری	
تحلیل واریانس چند متغیره	همبستگی کانونی با متغیرهای مجازی رگرسیون لوجستیک		

در انتخاب آزمون مناسب برای تجزیه و تحلیل از طریق جدول فوق باید دقت شود که در رگرسیون لوجستیک متغیر وابسته باید دو وجهی و در تحلیل تشخیصی دو یا چند وجهی باشد. و روشهای همبستگی کانونی و تحلیل واریانس چند متغیره نیز نیاز به چند متغیر وابسته دارند.

## فصل هشتم

# رگرسیون چندگانه

### مقدمه:

تحلیل رگرسیون یکی از روشهای پرکاربرد در مطالعات اجتماعی - اقتصادی است. این روش ارتباط تنگاتنگی با ضریب همبستگی داشته و عموماً بطور همزمان در مطالعات مورد استفاده قرار می‌گیرد. تحلیل رگرسیون این امکان را برای محقق فراهم می‌کند تا تغییرات متغیر وابسته را از طریق متغیرهای مستقل پیش بینی و سهم هر یک از متغیرهای مستقل را در تبیین متغیر وابسته تعیین کند. در این فصل به بررسی این روش آماری پرداخته شده و برای آشنایی با نحوه محاسبه آن با برنامه SPSS از یک مثال کمک گرفته شده است.

### تحلیل رگرسیون و ضریب همبستگی:

رگرسیون رابطه نزدیکی با ضریب همبستگی دارد. بدین معنا که برای انجام رگرسیون، باید ضریب همبستگی را محاسبه کرد. اگر میان متغیرهای مورد مطالعه، همبستگی وجود داشت، تنها در این صورت است که می‌توانیم از رگرسیون برای آزمون فرضیه‌های تحقیق استفاده نماییم. هرچه همبستگی بین متغیرها قویتر باشد، پیش بینی نیز دقیق‌تر خواهد بود. تفاوت



رگرسیون با ضریب همبستگی در این است که رگرسیون بدنبال پیش بینی است در حالی که ضریب همبستگی تنها میزان وابستگی دو متغیر را با هم بررسی می‌کند. علامت ضریب همبستگی در تحلیل رگرسیون  $R$  و مجذور آن  $R^2$  است. تفاوت  $R$  بزرگ با  $r$  کوچک در این است که اولاً  $R$  بیانگر رابطه میان بیش از دو متغیر است در حالی که  $r$  نشان دهنده رابطه بین دو متغیر  $x$  و  $y$  می‌باشد. به عبارت دیگر  $r$  ضریب همبستگی ساده<sup>(۱)</sup> است در حالی که  $R$  ضریب همبستگی چندگانه<sup>(۲)</sup> می‌باشد و براساس نمره‌های متغیر وابسته مشاهده شده ( $y$ ) و نمره‌های پیش بینی شده ( $\hat{y}$ ) محاسبه می‌شود.  $\lambda$  متغیر جدیدی است که از ترکیب خطی متغیرهای  $x_1, x_2, x_3$  و  $x_i$  ساخته می‌شود<sup>(۳)</sup>. ثانیاً دامنه تغییرات مقدار  $R$  بین صفر و یک است در حالی که مقدار  $r$  بین  $+1$  و  $-1$  در نوسان می‌باشد، و جهت تغییر را نیز نشان می‌دهد.

اگرچه تفاوت‌های اساسی بین تحلیل رگرسیون و ضریب همبستگی وجود دارد، اما در تحلیل داده‌ها به عنوان مکمل یکدیگر مورد استفاده قرار می‌گیرند. ضرایب همبستگی وجود و عدم وجود رابطه بین دو متغیر را نشان می‌دهد و همواره بین  $+1$  و  $-1$  در نوسان است. آماره ضریب همبستگی و خط رگرسیون را می‌توان بر روی یک نمودار پراکنندگی ترسیم و مورد تجزیه و تحلیل قرار داد. چنانچه نقاط مربوط به ضریب همبستگی کاملاً بر روی خط رگرسیون منطبق باشد در این صورت می‌توان تغییرات متغیر وابسته را از روی تغییرات متغیر مستقل بطور کامل پیش بینی کرد، و این حالتی است که در آن ضریب همبستگی برابر با ۱ می‌باشد. اما چنانچه این ضریب کمتر از یک باشد دقت پیش بینی رگرسیون نیز کاهش می‌یابد. بنابراین ضریب همبستگی مبنایی برای تعیین دقت برآورد رگرسیون می‌باشد. به همین دلیل است که در تحلیل داده‌ها این دو تکنیک باید با همدیگر مورد استفاده قرار گیرند.

## 1- Simple Correlation

## 2- Multiple Correlation

۳- چلبی مسعود (۱۳۷۲)، جزوه درس آمار در علوم اجتماعی، دانشگاه شهید بهشتی.

### تحلیل رگرسیون:

مفهوم رگرسیون برای اولین بار توسط فرانسیس گالتون<sup>(۱)</sup> در سال ۱۸۷۷ مورد استفاده قرار گرفت. او در مطالعه خود نشان داد که قد کودکان متولد شده از والدین بلند قامت گرایش به برگشت به متوسط قد افراد جامعه دارد. او فرآیند پیش بینی یک متغیر از طریق یک متغیر دیگر را با مفهوم «رگرسیون» بیان کرد. سپس رگرسیون چندگانه برای تخمین یا پیش بینی تغییرات یک متغیر بوسیله متغیرهای دیگر مورد استفاده قرار گرفت. از آن زمان به بعد تکنیک تحلیل رگرسیون برای تعیین رابطه آماری بین دو یا چند متغیر و پیش بینی تغییرات یک متغیر از روی متغیر یا متغیرهای دیگر در سطح وسیع توسط محققان رشته‌های مختلف بکار گرفته شد<sup>(۲)</sup>.

همانطوری که قبلاً نیز بیان گردید ضریب همبستگی که برای کشف وجود یا عدم وجود رابطه بین دو یا چند متغیر بکار می‌رود از جمله روشهای تحلیلی است که علاوه بر تعیین شدت و ضعف رابطه بین دو متغیر، جهت روابط را نیز نشان می‌دهد. اما این روش قادر به بیان روابط علی بین متغیرها نبوده و نمی‌تواند مدعی باشد که کدامیک از متغیرها علت و کدامیک معلول دیگری است. همچنین امکان پیش بینی یک متغیر از طریق متغیر یا متغیرهای دیگر وجود ندارد. به همین دلیل برای تحلیلهای پیشرفته‌تر و پیش بینی تغییر در متغیر وابسته در صورت تغییر در متغیر یا متغیرهای مستقل باید از روشهای دیگری نظیر تحلیل رگرسیون<sup>(۳)</sup> استفاده کرد. این روش ارتباط نزدیکی با ضریب همبستگی پیرسون و نمودار پراکنش داشته و خط رگرسیون در نمودار پراکنش<sup>(۴)</sup> مبنای تحلیل نقاط در این نمودار محسوب می‌شود و خط رگرسیون خطی است که مجموع مربعات انحراف نقاط روی نمودار با آن خط نسبت به هر خط دیگری حداقل مقدار را داشته و این خط مبنایی برای

1- Francis Galton

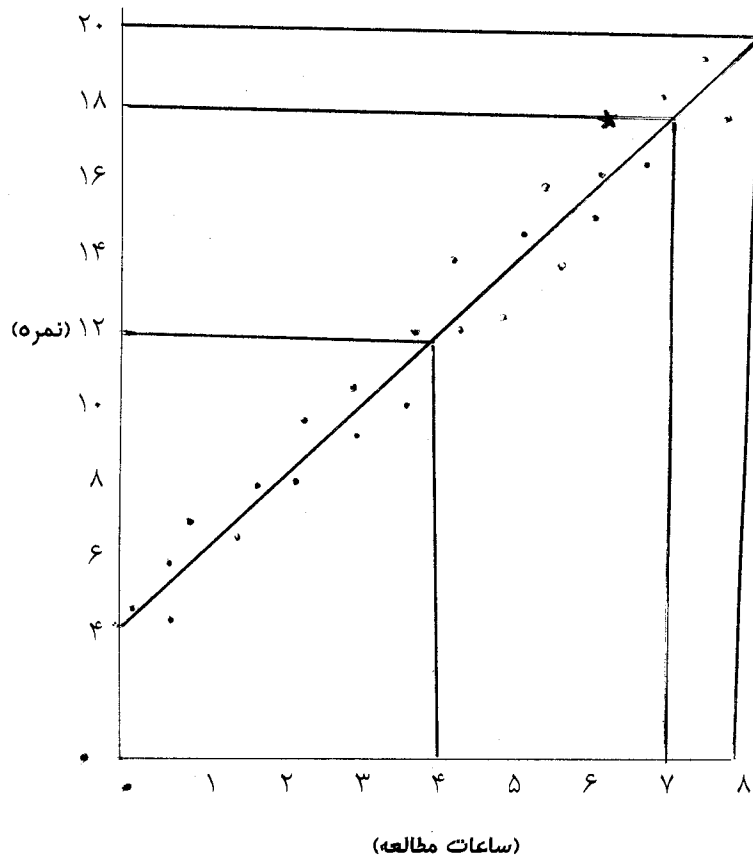
2- Kothari. C. R. (1992) "Quantitative Techniques" Vikas Publishing House New Delhi, PP. 206-207.

3- Regression Analysis

4- Scattergram

پیش بینی یک متغیر از طریق متغیر دیگر می‌باشد.

به عنوان مثال اگر میزان ساعات مطالعه به عنوان متغیر مستقل ( $X$ ) و نمره دانشجوی به عنوان متغیر وابسته ( $Y$ ) مدنظر باشد، می‌توان همبستگی موجود بین این دو متغیر را در قالب نمودار پراکنش ذیل ترسیم کرد (شکل ۸-۱).



شکل (۸-۱) رابطه بین ساعات مطالعه و نمره دانشجو

خط رگرسیون روابط بین متغیرهای  $X$  و  $Y$  را نشان می‌دهد. براساس این خط می‌توان نمره دانشجو ( $Y$ ) را از طریق میزان ساعات مطالعه ( $X$ ) پیش بینی کرد. به عنوان مثال دانشجویی که فقط

۴ ساعت مطالعه کرده باشد می تواند نمره ۱۲ بگیرد و یا دانشجویی که ۸ ساعت مطالعه کند قادر است نمره ۲۰ دریافت کند. اما سؤال اساسی این است که تا چه حد اینگونه پیش بینی که براساس خط رگرسیون انجام می گیرد از دقت لازم برخوردار می باشد. اگر به شکل (۸-۱) توجه کنید در این نمودار نمره واقعی فردی که تنها ۶ ساعت مطالعه کرده (و بوسیله ستاره مشخص شده است) ۱۸ می باشد. در حالی که این مقدار بر روی خط رگرسیون نمره ۱۶ را نشان می دهد. پس در اینجا خطای نمره پیش بینی شده با نمره واقعی ۲ نمره می باشد. مسئله اصلی این است که اگر این پیش بینی براساس میانگین نمرات انجام می گرفت، تا چه حد با نمرات واقعی افراد فاصله داشت. به عنوان مثال اگر این پیش بینی براساس میانگین نمرات انجام می گرفت، نمره ۱۴ (میانگین  $\bar{y}$ ) بدست می آمد که خطای پیش بینی را به ۴ افزایش می داد. در حالی که خطای پیش بینی انجام شده بر مبنای خط رگرسیون تنها ۲ نمره می باشد. در اینجا هرچه قدر ارتباط متغیرها با یکدیگر بیشتر باشد دقت پیش بینی را نیز افزایش می دهد. از آنجا که میزان خطا بر مبنای شاخص واریانس تعیین می گردد، بنابراین در تحلیل رگرسیون مقداری از واریانس متغیر  $Y$  را که بر مبنای متغیر  $X$  تبیین می گردد، مقدار واریانس تبیین شده<sup>(۱)</sup> می نامند که آنرا با  $R^2$  نمایش می دهند. اگر مقدار  $R^2$  از ۱ کم شود واریانس تبیین نشده بدست می آید. وجود واریانس تبیین نشده به این دلیل است که نمرات دانشجویان ( $Y$ ) فقط متأثر از ساعات مطالعه نمی باشد. واریانس ناشی از سایر متغیرها به واریانس تبیین نشده معروف است. چنانچه دلیل این تفاوت در نمرات دانشجویان، اختلاف در میزان ساعات مطالعه باشد می توان مدعی شد که تفاوت در نمرات از طریق تفاوت در ساعات مطالعه تبیین شده است. اما از آنجا که در برخی موارد علیرغم اینکه میزان ساعات مطالعه برابر می باشد اما نمرات دانشجویان با هم متفاوت است، بنابراین واریانس تبیین نشده نیز وجود خواهد داشت.

در تحلیل رگرسیون، واریانس تبیین شده را با  $R^2$  نمایش می دهند. باتوجه به اینکه دامنه  $R^2$  بین ۰ و ۱ می باشد بنابراین  $R^2$  برای بیان شدت ارتباط مناسب می باشد اما بیانگر جهت رابطه

نیست. به همین دلیل برای آگاهی از جهت رابطه، از ضریب همبستگی پیرسون استفاده می‌گردد<sup>(۱)</sup>.

### ضریب رگرسیون :

همانطوری که قبلاً نیز بیان شد خط رگرسیون در نمودار پراکنش به منظور پیش بینی متغیر وابسته (Y) براساس متغیر مستقل (X) مورد استفاده قرار می‌گیرد. اگر به نمودار قبلی برگردیم در این نمودار برای پیش بینی نمره کسی که ۸ ساعت مطالعه کرده است کافی است تا از نقطه ۸ در محور X خطی عمود بر محور X رسم کنیم تا خط رگرسیون را قطع نماید. سپس چنانچه از محل تقاطع خط رگرسیون خط دیگری عمود بر محور Y رسم کنیم میزان نمره پیش بینی شده بدست می‌آید، که مقدار آن ۲۰ می‌باشد. برای پیش بینی نمره دانشجویی که ۷ ساعت مطالعه کرده است باید از نقطه ۷ در محور X خطی عمود بر محور X رسم کرد تا خط رگرسیون را قطع و از همانجا خطی عمود بر محور Y رسم کرد تا نمره دانشجو بدست آید. اگر این خطوط برای سایر نقاط رسم شود خواهیم دید که به ازاء هر ساعت افزایش در میزان مطالعه، نمرات دانشجویان به اندازه ۲ نمره افزایش می‌یابد. به عبارت دیگر هر ساعت مطالعه اضافی به اندازه ۲ نمره بازده دارد که این تأثیر را ضریب رگرسیون یا شیب خط رگرسیون می‌نامند و با علامت b نمایش می‌دهند. ضریب b بدین معناست که با تغییر یک واحد در متغیر x، متغیر وابسته (y) نیز به اندازه b واحد تغییر می‌کند. ضریب b همواره برحسب مقیاس متغیر وابسته (y) بیان می‌شود. به عنوان مثال در اینجا مقدار b مساوی ۲ می‌باشد که منظور ۲ نمره است. نکته مهم دیگر در نمودار پراکنش، محل تقاطع خط رگرسیون با محور Y می‌باشد. این نقطه تقاطع را عرض از مبدأ Y یا مقدار ثابت می‌نامند و با علامت a نمایش می‌دهند. a در واقع نشان دهنده نمره کسانی است که مقدار متغیر X (ساعات مطالعه) آنها صفر می‌باشد. به عبارت دیگر نمره کسانی است که هیچگونه وقتی را صرف مطالعه نکرده‌اند.

علاوه بر نمودار پراکنش و خط رگرسیون، از طریق معادله جبری خط راست نیز می‌توان پیش

۱- دواس، دی، ای (۱۳۷۶) «پیمایش در تحقیقات اجتماعی» ترجمه هوشنگ نایبی، نشر نی، تهران.

بینی‌های لازم را انجام داد. این معادله بشرح ذیل است.

$$\hat{Y} = a + bx$$

در این معادله :

$\hat{Y}$  = مقدار پیش بینی شده متغیر  $y$

$a$  = مقدار ثابت یا عرض از مبدأ نقطه تقاطع خط رگرسیون با محور  $y$

$b$  = شیب خط یا ضریب رگرسیون

$x$  = مقادیر متغیر مستقل

براساس این معادله با بدست آوردن مقادیر  $a$  و  $b$  می‌توان مقدار  $\hat{Y}$  را برای هر مقدار  $x$  پیش

بینی کرد. در این مثال مقدار  $a = 4$  و  $b = 2$  می‌باشد. بنابراین می‌توان نمره فردی که ۶ ساعت مطالعه

کرده است را پیش بینی کرد.

$$\hat{Y} = 4 + 2 \times 6$$

$$\hat{Y} = 4 + 12$$

$$\hat{Y} = 16$$

همین مقدار از طریق نمودار پراکنش نیز قابل دستیابی است.

### داده‌های سازگار با رگرسیون :

مهمترین سؤال در بکارگیری روش تحلیل رگرسیون این است که برای استفاده از این

تکنیک چه نوع داده‌هایی باید بکار گرفته شوند؟ اگرچه ساختار این تکنیک عمدتاً براساس داده‌های

کمی می‌باشد، اما استفاده از داده‌های کیفی در متغیرهای مستقل شامل متغیرهای مستقل اسمی یا

ترتیبی نیز امکانپذیر می‌باشد. اما چنانچه داده‌ها از نوع اسمی (مانند جنسیت، رشته تحصیلی، گروه

خون، محل تولد و غیره) باشند و بخواهیم از تحلیل رگرسیون چندگانه استفاده کنیم در چنین مواردی

اینگونه داده‌ها باید به صورت متغیرهای مجازی<sup>(۱)</sup> که حالت دو وجهی داشته باشند تبدیل شوند. متغیری مانند جنسیت خود دو وجهی بوده و به صورت کدهای ۰ و ۱ قابل تعریف می‌باشند. (این کدها فاقد ارزش عددی بوده و صرفاً برای تعیین طبقات مورد استفاده قرار می‌گیرند.) اما متغیرهایی که چند وجهی هستند باید به چند متغیر دو وجهی تبدیل شوند. بطور کلی اگر یک متغیر دارای  $n$  وجه باشد آن متغیر قابل تبدیل به  $n-1$  متغیر خواهد بود که هر یک از متغیرها می‌تواند دو حالت (بله و خیر) داشته باشند. در چنین مواقعی هر یک از متغیرها باید به صورت مستقل در یک ستون در صفحه گستر SPSS وارد شود.

فرض کنید میزان رضایت شغلی ۲۰ نفر از شاغلین را اندازه‌گیری و در قالب متغیر  $y$  در رایانه وارد کرده‌اید. اگر بخواهید ارتباط رضایت شغلی با رشته تحصیلی (پزشکی، فنی، انسانی و هنر) را بررسی کنید در این صورت داده‌های مربوط به متغیر رشته تحصیلی که دارای چهار رشته مستقل است باید در قالب سه متغیر با کدهای ۰ و ۱ در رایانه وارد شود، بطوریکه در هر متغیر به رشته مورد نظر کد ۱ و به بقیه رشته‌ها کد ۰ اختصاص می‌یابد<sup>(۲)</sup>. اگر در داده‌های جمع‌آوری شده ۵ نفر اول دارای رشته پزشکی، ۵ نفر بعدی دارای رشته فنی و ۵ نفر بعدی دارای رشته انسانی و ۵ نفر آخر دارای رشته هنر باشند در این صورت داده‌های مورد نظر باید به صورت جدول (۸-۱) در پنجره Data Editor وارد شود.

محاسبه رگرسیون برای متغیرهای مجازی درست مانند محاسبه رگرسیون برای متغیرهای کمی است و معادله رگرسیون آن نیز مانند معادله رگرسیون چندگانه برای متغیرهای کمی می‌باشد. اما در این معادله مقادیر مربوط به متغیر مورد نظر (۱) و مقادیر مربوط به سایر متغیرها (۰) در نظر گرفته شده و در ضرایب بدست آمده ضرب می‌گردند.

#### 1- Dummy Variables

۲- برای اطلاع بیشتر به کتاب تحلیل داده‌های چند متغیری در پژوهش رفتاری، حیدرعلی هومن، نشر پارسا مراجعه کنید.

جدول (۸-۱) نحوه ورود داده‌های کیفی با کدهای مجازی

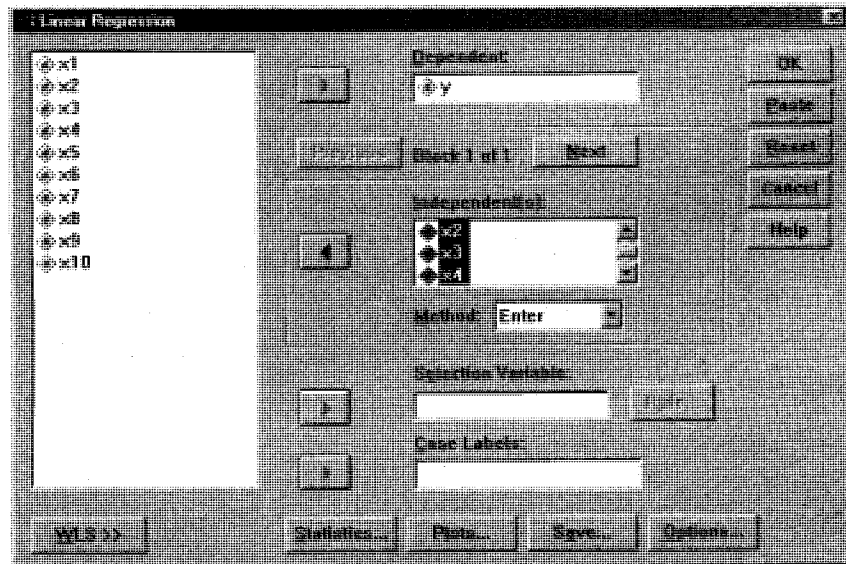
ردیف	y	X	X	X
۱	۱۸	۱	۰	۰
۲	۱۹	۱	۰	۰
۳	۱۶	۱	۰	۰
۴	۱۵	۱	۰	۰
۵	۱۸	۱	۰	۰
۶	۱۴	۰	۱	۰
۷	۱۵	۰	۱	۰
۸	۱۶	۰	۱	۰
۹	۱۷	۰	۱	۰
۱۰	۱۸	۰	۱	۰
۱۱	۱۵	۰	۰	۱
۱۲	۱۲	۰	۰	۱
۱۳	۱۰	۰	۰	۱
۱۴	۱۶	۰	۰	۱
۱۵	۱۱	۰	۰	۱
۱۶	۱۱	۰	۰	۰
۱۷	۱۲	۰	۰	۰
۱۸	۱۵	۰	۰	۰
۱۹	۱۳	۰	۰	۰
۲۰	۱۰	۰	۰	۰

## روش محاسبه رگرسیون با برنامه SPSS:

فرض کنید می‌خواهیم رابطه بین ساعات مطالعه (X) و نمره دانشجو (Y) را با استفاده از رگرسیون ساده مورد تجزیه و تحلیل قرار دهیم. برای این منظور داده‌های هر دو متغیر را در دو ستون مستقل در قسمت Data Editor وارد کرده و دستور Analyze/Regression/Linear را اجرا کنید تا شکل (۸-۲) ظاهر شود.







شکل (۲-۸) پنجره Linear Regression

در پنجره‌ای که ظاهر می‌شود متغیر نمره دانشجو که تحت عنوان Mark تعریف شده است در قسمت Dependent و ساعات مطالعه را که تحت عنوان h تعریف شده است در قسمت Independent وارد کنید. سپس دکمه OK را کلیک کنید تا نتایج در پنجره Output ظاهر شود. این نتایج در جداول (۱-۸) تا (۴-۸) آمده است همانطوری که از جدول (۱-۸) برمی آید، تنها متغیر ساعات مطالعه (H) به روش Enter وارد معادله شده است. با وارد شدن این متغیر ضریب تعیین  $R^2 = 0.711$  بدست آمده است (جدول ۲-۸). مقدار ضریب تعیین نشان می‌دهد که ۷۱/۱ درصد تغییرات در نمره دانشجو به ساعات مطالعه او مربوط می‌باشد و بقیه ۲۸/۹ درصد به عوامل دیگر بستگی دارد. البته اشکال وارد بر این ضریب این است که تعداد درجه آزادی را در نظر نمی‌گیرد. به همین دلیل برای رفع این مشکل معمولاً از ضریب تعیین تعدیل شده<sup>(۱)</sup> استفاده می‌گردد<sup>(۲)</sup>. مقدار

1- Adjusted R Square

این ضریب برابر با ۰/۶۹۹ می باشد (جدول ۸-۲). البته در رگرسیون چندگانه هرچه تعداد متغیرهای مستقل کمتر باشد فاصله ضریب تعیین ( $R^2$ ) با ضریب تعیین تعدیل شده ( $R^2Ad$ ) کمتر می شود. همچنین در این جدول، ضریب همبستگی چندگانه<sup>(۳)</sup> که با علامت R نشان داده شده است برابر با ۰/۸۴۳ می باشد. این ضریب شدت رابطه بین متغیر وابسته (Y) و متغیر یا متغیرهای مستقل را نشان می دهد که مقدار آن همواره بین ۰ و ۱ می باشد. معنی دار بودن رگرسیون نیز در جدول شماره (۳-۸) بوسیله F محاسبه شده است که در سطح ۹۹ درصد ( $sig=0.000$ ) معنی دار است. متغیر وارد شده در معادله رگرسیونی که هسته اصلی تحلیل رگرسیون می باشد در جدول (۴-۸) آمده است. معادله رگرسیون ساده را می توان با استفاده از ستون B به شرح ذیل محاسبه کرد.

$$\hat{Y} = a + bx$$

$$\hat{Y} = ۰/۳۸۰ + ۲/۲۸۵ (H)$$

آزمون t مربوط به ضریب رگرسیون نیز در این جدول نشان می دهد که این ضریب معنی دار

بوده و دربرآورد مقدار Y مؤثر می باشد.

جدول (۱-۸) Variables Entered / Removed

Model	Variables Entered	Variables Removed	Method
1	H <sup>a</sup>		Enter

a. All requested variables entered.

b. Dependent Variable: Mark

جدول (۲-۸) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.843 <sup>a</sup>	.711	.699	3.32

a. Predictors: (Constant), H.

۲- منصورفر «جزوه رگرسیون چند متغیره برای دانشجویان کارشناسی ارشد پژوهشگری» ص ۴۶

جدول (۸-۳) ANOVA

Model	Sum of squares	df	Mean Square	F	Sig.
1 Regression	623.308	1	623.308	55.697	.000 <sup>a</sup>
Residual	252.852	23	10.994		
Total	876.160	24			

a. Predictors: (Constant), H.

جدول (۸-۴) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.380	1.612		.236	8.16
	H	2.285	.303	.843	7.530	.000

a. Dependent Variable: Mark

## رگرسیون چندگانه :

رگرسیون چندگانه<sup>(۱)</sup>، برخلاف رگرسیون ساده که تنها با توجه به یک متغیر مستقل، تغییرات متغیر وابسته را پیش بینی می‌کند، با استفاده از ترکیب خطی چند متغیر مستقل به پیشگویی متغیر وابسته می‌پردازد. با این تفاوت، دو سؤال اساسی مطرح می‌شود. این سؤالات عبارتند از:

۱ - آیا اضافه کردن متغیرهای مستقل جدید (مانند میزان بهره هوشی، درآمد، سن و طبقه اجتماعی دانشجویان به متغیر مستقل قبلی (ساعات مطالعه (H)) میزان پیشگویی متغیر وابسته (نمره دانشجو) را دقیق‌تر خواهد کرد؟

۲- از میان این چند متغیر مستقل، کدامیک برای پیش بینی متغیر وابسته سودمندتر است. به عبارت دیگر نقش کدام متغیر مستقل بیشتر از سایر متغیرها می باشد؟

پاسخ قطعی این سؤالات را می توان از طریق تحلیل رگرسیون چند گانه بدست آورد. برای انجام این تمرین داده های مربوط به متغیرهای میزان بهره هوشی (IQ)، درآمد (Income)، سن (Age) و طبقه اجتماعی (Soclass) دانشجویان را در ستونهای مستقل در Data Editor فایل قبلی وارد کنید تا محاسبات بعدی براساس این داده ها انجام گیرد. از آنجا که رگرسیون چند گانه به چند روش محاسبه می گردد بنابراین مهمترین روشها در اینجا مورد بررسی قرار می گیرند.

### ۱- روش رگرسیون چند گانه توأم (Enter):

رگرسیون به شیوه Enter، روشی است که در آن کلیه متغیرهای مستقل همزمان وارد تحلیل شده و اثرات کلیه متغیرهای مستقل بر روی متغیر وابسته بررسی می گردد. برای اجرای این روش، ابتدا دستور: Analyze/Regression/Linear را اجرا کرده تا پنجره رگرسیون ظاهر شود (شکل ۲-۸). سپس متغیر نمره دانشجو (Mark) را در قسمت Dependent و متغیرهای مستقل را در قسمت Independent Variables وارد کرده و سپس روش Enter را از قسمت Method انتخاب کنید و دکمه OK را فشار دهید. با این دستور نتایج محاسبات در جداول مختلف در پنجره Output ظاهر می گردند. در این پنجره جدول اول روش مورد استفاده و متغیرهای مستقل و وابسته بکار گرفته شده در تحلیل رگرسیون را نشان می دهد (جدول ۵-۸). جدول بعدی که تحت عنوان Model Summary می باشد مقدار  $R^2$  را نشان می دهد که برابر با ۰/۹۱۱ می باشد (جدول ۶-۸). همانطوری که قبلاً نیز گفته شد در روش Enter کلیه متغیرها همزمان وارد معادله می شوند. بنابراین  $R^2$  بدست آمده نشان می دهد که ۹۱/۱ درصد تغییرات متغیر وابسته توسط این پنج متغیر مستقل تبیین می گردند. اگر  $R^2$  بدست آمده در این مرحله را با  $R^2$  رگرسیون ساده که برابر با ۰/۷۱۱ مقایسه کنیم پاسخ سؤال یک مشخص خواهد شد. بنابراین می توان نتیجه گرفت که با افزایش متغیرهای جدید دقت پیشگویی رگرسیون نیز افزایش می یابد.

جدول ANOVA نیز معنی‌دار بودن رگرسیون و رابطه خطی بین متغیرها را نشان می‌دهد و سطح معنی‌داری (sig=0.000) بدست آمده، معنی‌دار بودن آن را در سطح ۹۹ درصد تأیید می‌کند (جدول ۸-۷).

نتایج اصلی رگرسیون در جدول (۸-۸) آمده است. ستون B در این جدول به عنوان ضریب به منظور پیش بینی مقدار y در معادله رگرسیون مورد استفاده قرار می‌گیرد. فرمول عمومی معادله رگرسیون چندگانه عبارتست از:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

در این معادله:

$\hat{Y}$  = مقدار پیش بینی شده متغیر Y

a = مقدار ثابت یا عرض از مبدأ نقطه تقاطع خط رگرسیون با محور Y

b = ضریب رگرسیون یا شیب منحنی

x = مقادیر متغیرهای مستقل

براساس مقادیر ستون B معادله رگرسیون را می‌توان بشرح ذیل نوشت:

$$\hat{Y} = -29/196 + 0/634(H) + 0/316(IQ) - 0/0175(Income) - 0/0527(Age) + 0/535(Soclass)$$

در این جدول مقدار t تک تک ضرایب رگرسیون نیز محاسبه شده و سطح معنی‌داری آنها در آخرین ستون جدول آمده است. همانطوری که مقدار سطح معنی‌داری (sig) نشان می‌دهد، تنها اثرات دو متغیر H و IQ معنی‌دار شده است و اثرات سایر متغیرها معنی‌دار نمی‌باشد و آنها تأثیر بسیار ضعیفی در پیشگویی متغیر وابسته دارند.

اما در مورد اهمیت و نقش متغیرهای مستقل در پیشگویی معادله رگرسیون باید از مقادیر بتا استفاده کرد. از آنجا که مقادیر بتا استاندارد شده می‌باشند بنابراین از طریق آن می‌توان در مورد اهمیت نسبی متغیرها قضاوت کرد. بزرگ بودن مقدار بتا نشان دهنده اهمیت نسبی و نقش آن در

پیشگویی متغیر وابسته می‌باشد. بنابراین در اینجا می‌توان قضاوت کرد که متغیر IQ به مراتب سهم بیشتری در مقایسه با سایر متغیرها در پیشگویی متغیر وابسته دارد. زیرا یک واحد تغییر در انحراف معیار آن باعث می‌شود تا انحراف معیار متغیر وابسته (Mark) به اندازه  $0.773$  تغییر کند. در حالی که یک واحد تغییر در انحراف معیار H باعث می‌شود تا انحراف معیار متغیر وابسته به اندازه  $0.234$  تغییر نماید (جدول ۸-۸).

جدول (۸-۵) Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	SOCLASS AGE,IQ,H, INCOME <sup>a</sup>		Enter

a. All requested variables entered

b. Dependent Variable: Mark

جدول (۸-۶) Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.955 <sup>a</sup>	.911	.888	2.02

a. Predictors: (Constall). Soclass, Age, IQ, H, Income

جدول (۸-۷) ANOVA

Model		Sum of squares	df	Mean Square	F	Sig.
1	Regression	798.542	5	159.708	39.095	.000 <sup>a</sup>
	Residual	77.618	19	4.085		

Total		876.160	24			
جدول (۸-۸) Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficient	t	Sig.
		B	Std. Error	Beta		
		squares		Square		
1	(Constant)	-29.196	5.291		-5.518	.000
	H	.634	.365	.234	1.740	.048
	IQ	0.316	.061	.773	5.220	.000
	INCOME	-0.0175	.028	-.105	-.639	.530
	AGE	-0.0527	.042	-.100	-1.253	.225
	SOCLASS	.535	.664	.075	.805	.431

a. Dependent Variable: Mark

## ۲- روش گام به گام :

روش گام به گام<sup>(۱)</sup> روشی است که در آن قوی‌ترین متغیرها یک به یک وارد معادله می‌شوند و این کار تا زمانی ادامه می‌یابد که خطای آزمون معنی‌داری به ۵ درصد برسد. برای انجام رگرسیون گام به گام باید همان دستور قبلی Analyze/Regression/Linear را اجرا کرد و در پنجره ظاهر شده، متغیر وابسته را در قسمت Dependent و متغیرهای مستقل را در قسمت Independent وارد نمود، سپس از قسمت Method مورد Stepwise را انتخاب و بعد دکمه OK را فشار داد. با اجرای این دستورات نتایج رگرسیون در جداول مختلف در پنجره Output ظاهر می‌شود. در این مثال نمره دانشجو (mark) به عنوان متغیر وابسته و پنج متغیر بهره‌هوشی (IQ)، ساعات مطالعه (H)، درآمد (Income)، سن (Age) و طبقه اجتماعی (Soclass) به عنوان متغیرهای مستقل مورد تجزیه و

1- Stepwise Method

تحلیل قرار گرفته‌اند. همانطوری که از جداول (۸-۹ تا ۸-۱۳) برمی‌آید، تحلیل رگرسیون تنها تا دو گام پیش رفته است. در گام اول متغیر بهره‌هوشی (IQ) وارد معادله شده است که میزان ضریب همبستگی آن (R) با متغیر وابسته ۰/۹۳۴ بدست آمده است (جدول ۸-۱۰). در این مرحله میزان ضریب تعیین برابر با  $R^2 = 0/۸۷۲$  و ضریب تعیین تعدیل شده برابر با  $R^2_{Ad} = 0/۸۶۷$  بدست آمده است (جدول ۸-۱۰). در گام دوم با وارد شدن دومین متغیر یعنی ساعات مطالعه (H) ضریب همبستگی چندگانه به  $R = 0/۹۴۸$ ، ضریب تعیین به  $R^2 = 0/۹۰۰$  و ضریب تعیین تعدیل شده به  $R^2_{Ad} = 0/۸۹۰$  افزایش یافته است (جدول ۸-۱۰). به عبارت دیگر براساس ضریب تعیین تعدیل شده ۸۹ درصد تغییرات متغیر وابسته (نمره دانشجو) بوسیله دو متغیر بهره‌هوشی و ساعات مطالعه تبیین می‌شود. اگر ضریب تعیین بدست آمده توسط روش Enter با ضریب تعیین محاسبه شده توسط روش گام به گام مقایسه شود، خواهیم دید که سه متغیر سن، درآمد و طبقه اجتماعی نقش بسیار ناچیزی در تبیین واریانس متغیر وابسته داشته‌اند. این سهم در ذیل محاسبه شده است.

$$R^2 = 0/۹۱۱ - 0/۹۰۰ = 0/۰۱۱$$

مربوط به روش گام به گام -  $R^2$  مربوط به روش Enter

بنابراین، براساس ضریب تعیین ( $R^2$ ) تنها ۱/۱ درصد تغییرات متغیر وابسته توسط این سه متغیر تبیین شده‌اند. اما اگر این مقایسه توسط ضریب تعیین تعدیل شده انجام گیرد خواهیم دید که با خروج سه متغیر فوق از تحلیل رگرسیون حتی میزان ضریب تعیین تعدیل شده به روش Enter از ۰/۸۸۸ به ۰/۸۹۰ در روش گام به گام افزایش یافته است. این موضوع به این دلیل است که ضریب تعیین تعدیل شده براساس درجه آزادی تعدیل یافته است. بنابراین در روش Enter این تعدیل براساس پنج متغیر و در روش گام به گام براساس دو متغیر انجام گرفته است (جدول ۸-۱۰). میزان F نیز در جدول (۸-۱۱) در سطح ۹۹ درصد ( $sig=0.000$ ) معنی‌دار شده است که حاکی از معنی‌دار بودن رگرسیون می‌باشد.



جدول (۹-۸) Variables Entered/Removed

Model	Vaiables Entered	Variables Removed	Method
1	IQ		Stepwise (Criteria: Probability of-F-to-ent er<=.050, Probability- of-F-to-re move>= .100).
2	H		Stepwise (Criteria: Probability- of-F-to-ent er<=.050, Probability- of-f-to-re move>= .100).

a. Dependent variable: Mark

جدول (۱۰-۸) Model Summary

Model	R	R Square	Adjusted R Square	Std.Error of the Estimate
1	.934 <sup>a</sup>	.872	.867	2.21
2	.948 <sup>b</sup>	.900	.890	2.00

جدول (۱۱-۸) ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	764.110	1	764.110	156.846	.000 <sup>a</sup>
Residual	112.050	23	4.872		
Total	876.160	24			
2 Regression	788.123	2	394.061	98.476	.000 <sup>b</sup>
Residual	88.037	22	4.002		
Total	876.160	24			

جدول (۱۲-۸) Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-37.197	3.909		-9.517	.000
IQ	.382	.030	.934	12.524	.000
2 (Constant)	-29.450	4.749		-6.202	.000
IQ	.293	.046	.716	6.418	.000
H	.741	.302	.273	2.450	.023

جدول (۱۳-۸) Excluded Variables

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance
1 H	.273 <sup>a</sup>	2.450	.023	.463	.367
INCOME	.150 <sup>a</sup>	1.038	.311	.216	.265
AGE	-.135 <sup>a</sup>	-1.870	.075	-.370	.966
2 SOCLASS	.150 <sup>a</sup>	1.648	.113	.332	.623
INCOME	-.013 <sup>b</sup>	-.085	.933	-.019	.199
AGE	-.091 <sup>b</sup>	-1.276	.216	-.268	.872
SOCLASS	.08g <sup>b</sup>	.979	.339	.209	.551

جدول (۱۲-۸) نشان می‌دهد که متغیر بهره‌دهی (IQ) در سطح ۹۹ درصد (sig=0.000) و متغیر ساعات مطالعه (H) در سطح ۹۵ درصد (sig=۰/۰۲۳) معنی‌دار می‌باشند. این در حالی است که میزان خطای سایر متغیرها بیشتر از ۰/۰۵ بوده‌اند و به همین دلیل وارد معادله رگرسیون نشده‌اند.

براساس مقادیر B جدول (۸-۱۲)، معادله رگرسیون را می‌توان بشرح ذیل نوشت.

$$\hat{Y} = -29/450 + 0/293(IQ) + 0/741(H)$$

اما قضاوت در مورد سهم و نقش هریک از دو متغیر بهره هوشی (IQ) و ساعات مطالعه (H) در تبیین متغیر وابسته را باید به مقادیر بتا<sup>(۱)</sup> واگذار کرد. زیرا این مقادیر استاندارد شده بوده و امکان مقایسه و تعیین سهم نسبی هریک از متغیرها را فراهم می‌سازد.

اگرچه براساس مقدار B ضریب ساعات مطالعه ( $H=0/741$ ) به مراتب بیشتر از ضریب بهره هوشی ( $IQ=0/293$ ) می‌باشد، اما سهم و نقش بهره هوشی در تبیین متغیر وابسته بسیار بیشتر از ساعات مطالعه است. زیرا براساس بتای بدست آمده برای بهره هوشی، به ازاء یک واحد تغییر در انحراف معیار بهره هوشی به اندازه  $0/716$  در انحراف معیار متغیر وابسته (نمره دانشجوی) تغییر ایجاد می‌گردد. در حالی که به ازاء یک واحد تغییر در انحراف معیار ساعات مطالعه، تنها  $0/273$  در انحراف معیار متغیر وابسته تغییر ایجاد می‌شود.

### ۳- روش حذف پس رو<sup>(۲)</sup>:

در این روش، همانند روش Enter ابتدا کلیه متغیرهای مستقل وارد معادله شده و اثر کلیه متغیرها بر روی متغیر وابسته سنجیده می‌شود. اما در این روش به مرور متغیرهای ضعیف‌تر، یکی پس از دیگری از معادله خارج شده و نهایتاً این مراحل تا زمانی ادامه می‌یابد که خطای آزمون معنی‌داری به ده درصد برسد. در این روش با خروج متغیرها، بتدریج ممکن است مقدار ضریب تعیین تعدیل شده<sup>(۳)</sup> افزایش و سپس مجدداً کاهش یابد.

در روش حذف پس رو، متغیرهای ضعیف‌تر از معادله حذف و در جدولی تحت عنوان Excluded Variables قرار می‌گیرد، و متغیرهای باقیمانده بیشترین بخش از تغییرات متغیر وابسته

1- Beta

2- Backward Elimination Method

3- Adjusted R Square

را تبیین می‌کند. نتایج بدست آمده از این روش، معمولاً در مقایسه با نتایج روش گام به گام سهم بیشتری از تغییرات متغیر وابسته را تبیین می‌کند و این به این دلیل است که در این روش متغیرهای ضعیف‌تر تا زمانی حذف می‌گردند که خطای آزمون معنی‌داری به ده درصد برسد. در حالی که در روش گام به گام تا زمانی متغیرهای قوی‌تر وارد معادله می‌شوند که میزان خطای آزمون به پنج درصد برسد. به همین دلیل در روش حذف پس رو ممکن است متغیرهای بیشتری در مقایسه با روش گام به گام وارد معادله شوند. نحوه محاسبه این روش نیز در برنامه SPSS مانند روشهای قبلی می‌باشد، با این تفاوت که در این روش از قسمت Method باید گزینه Backward را انتخاب کرد. تحلیل نتایج در این روش نیز مانند روشهای قبلی می‌باشد.

#### ۴- روش پیشرو<sup>(۱)</sup>:

این روش نیز شبیه روش گام به گام است. در این روش نیز متغیرهای قوی‌تر انتخاب و وارد معادله می‌گردند. این عمل تا زمانی ادامه می‌یابد که خطای آزمون به ۵ درصد برسد. در این مرحله عملیات متوقف می‌شود. نتایج بدست آمده در این روش همانند روش گام به گام می‌باشد. روش محاسبه و تحلیل نتایج در این روش نیز مانند روش گام به گام انجام می‌گیرد، اما برای محاسبه این روش از پنجره اصلی رگرسیون از قسمت Method باید گزینه Forward را انتخاب کرد.

## فصل نهم

# رگرسیون لجستیک

### مقدمه:

همانطوریکه در فصل هشتم بیان گردید، شرط استفاده از رگرسیون چندگانه، کمی بودن متغیر وابسته است<sup>۱</sup>. در صورتی که متغیر وابسته از نوع اسمی دو وجهی (مانند بلی، خیر، عضو بودن، عضو نبودن و غیره) باشد در این صورت بجای رگرسیون چندگانه باید از رگرسیون لجستیک<sup>(۱)</sup> استفاده کرد. در این روش براساس متغیرهای مستقل مورد استفاده می توان احتمال هر یک از سطوح متغیر دو وجهی وابسته را محاسبه کرد. این فصل به نحوه محاسبه رگرسیون لجستیک و تفسیر نتایج آن اختصاص داده شده است.

د- در تحلیل رگرسیون، متغیرهای مستقل نیز باید عمدتاً کمی باشند، اگر چه ترکیبی از متغیرهای فاصله‌ای و ترتیبی نیز مجاز شمرده شده است. اما چنانچه برخی از متغیرها اسمی باشند باید اینگونه متغیرها بصورت متغیرهای مجازی (Dummy Variables) دو وجهی با کدهای ۰ و ۱ درآیند.

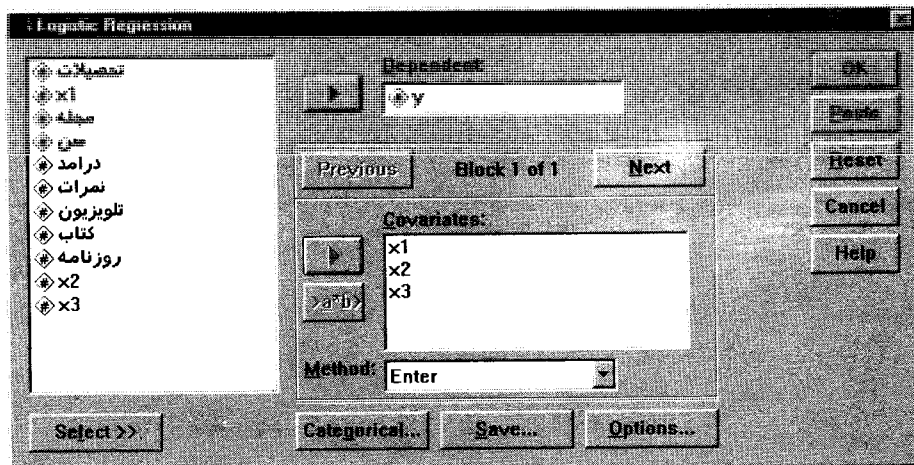
### 1- Logistic Regression

## روش محاسبه رگرسیون لجستیک بوسیله SPSS

برای محاسبه رگرسیون لجستیک ابتدا داده‌های مورد نظر را در پنجره Data Editor وارد کنید. فرض کنید تعداد ۵۰ نفر از شرکت کنندگان در کنکور را انتخاب کرده و می‌خواهیم بدانیم قبولی در کنکور تابع کدامیک از عوامل مدت مطالعه، سن دانشجو و معدل دیپلم می‌باشد. برای این منظور باید نتیجه آزمون کنکور (متغیر وابسته) بصورت قبول یا رد با کدهای 1 و 0 و سایر متغیرهای مستقل نیز در ستونهای جداگانه در پنجره Data Editor وارد شوند. پس از وارد کردن داده‌ها دستور:

Analyze / Regression / Binary Logistic

را اجرای کنید تا شکل (۹-۱) در مانیتور رایانه ظاهر شود. در این مرحله متغیر وابسته را به قسمت Dependent و متغیرهای مستقل را به قسمت Covariates وارد کنید و از قسمت Method گزینه Forward: Conditional را انتخاب نمایید.



شکل (۹-۱) پنجره Logistic Regression

در این پنجره با فشار دادن دکمه Save پنجره جدیدی باز می‌شود. در این پنجره عبارت‌های Probabilities و Group membership را علامت زده و دکمه Continue را فشار دهید. با انتخاب موارد فوق دو متغیر در پنجره Data Editor ایجاد می‌گردد که اولی احتمال تعلق فرد به سطح اول متغیر وابسته (که در این مثال احتمال قبولی در کنکور می‌باشد) را نشان می‌دهد. دومین متغیر نیز تعلق فرد به گروه‌های دوگانه را بر اساس متغیر قبلی پیش بینی می‌کند. چنانچه احتمال قبولی بیش از ۰/۵ باشد فرد قبول و اگر این احتمال کمتر از ۰/۵ باشد فرد مردود پیش بینی خواهد شد<sup>(۱)</sup>. سپس با کلیک کردن بر روی دکمه Options پنجره دیگری باز می‌شود که باید از پنجره فوق موارد Classification Plots و At last step را علامت بزینید با فشار دادن دکمه Continue و Ok نتایج در پنجره Output ظاهر می‌شود.

#### جدول (۹-۱) نتایج حاصل از تحلیل رگرسیون لجستیک

##### Case Processing Summary

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	50	100.0
	Missing Cases	0	.0
	Total	50	100.0
Unselected Cases		0	.0
	Total	50	100.0

a. If weight is in effect, see classification table for the total number of cases.

##### Dependent Variable Encoding

Original Value	Internal Value
fail	0
pass	1

۱- افشین نیا، فرساد (۱۳۷۸). «تحلیل کاربردی داده‌ها: راهنمای استفاده از SPSS» مدیریت مطالعات و توسعه آموزش پزشکی

دانشگاه علوم پزشکی اصفهان، ص ۱۶۱-۱۶۰.

**Block 0: Beginning Block**

Classification Table<sup>a,b</sup>

Observed		Predicted		
		کنکور		Percentage Correct
		fail	pass	
Step 0	کنکور	fail	pass	
		fail	pass	
	Overall Percentage			

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

Step	Variables	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.080	.283	.080	1	.777	.923

Variables not in the Equation

Step	Variables	Score	df	Sig.
Step 0	سن	2.901	1	.089
	معدل	27.338	1	.000
	مطالعه	27.862	1	.000
	Overall Statistics	31.428	3	.000

**Block 1: Method = Forward Stepwise (Conditional)**

Omnibus Tests of Model Coefficients

Step		Chi-square	df	Sig.
Step 1	Step	34.915	1	.000
	Block	34.915	1	.000
	Model	34.915	1	.000
Step 2	Step	4.774	1	.029
	Block	39.689	2	.000
	Model	39.689	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	34.319	.503	.670
2	29.546	.548	.731



Classification Table<sup>a</sup>

	Observed	Predicted			
		کنکور		Percentage Correct	
		fail	pass		
Step 1	کنکور	fail	23	3	88.5
		pass	2	22	91.7
	Overall Percentage				90.0
Step 2	کنکور	fail	25	1	96.2
		pass	1	23	95.8
	Overall Percentage				96.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	مطالعه	.907	.229	15.640	1	.000	2.477
	Constant	-5.371	1.448	13.749	1	.000	.005
Step 2	معدل	.585	.298	3.864	1	.049	1.795
	مطالعه	.552	.266	4.315	1	.038	1.737
	Constant	-12.274	4.111	8.915	1	.003	.000

a. Variable(s) entered on step 1: مطالعه

b. Variable(s) entered on step 2: معدل

Variables not in the Equation

		Score	df	Sig.
Step 1	Variables	سن	1	.293
		معدل	1	.035
	Overall Statistics		2	.103
Step 2	Variables	سن	1	.648
	Overall Statistics		1	.648

Step Summary<sup>a,b</sup>

Step	Improvement			Model		
	Chi-square	df	Sig.	Chi-square	df	Sig.
1	34.915	1	.000	34.915	1	.000
2	4.774	1	.029	39.689	2	.000

Step Summary<sup>a,b</sup>

Step	Correct Class %	Variable
1	90.0%	IN: مطالعه
2	96.0%	IN: معدل

a. No more variables can be deleted from or added to the current model.

b. End block: 1

همانطوریکه از نتایج فوق برمی‌آید تعداد افراد مورد بررسی ۵۰ نفر بوده و هیچگونه داده گم شده‌ای در آن وجود نداشته است و ۱۰۰ درصد افراد در تحلیل وارد شده‌اند. جدول بعدی نیز نشان می‌دهد که متغیر وابسته با کدهای ۱ برای قبول و ۰ برای مردود اختصاص یافته است. از آنجا که از روش پیشرو گام به گام استفاده شده است بنابراین متغیر مدت زمان مطالعه در گام اول و متغیر معدل دیپلم در گام دوم وارد معادله شده‌اند. همانطوریکه در تحلیل رگرسیون برای بررسی معنادار بودن رابطه از مشخصه‌های استاندارد  $F$  و  $t$  استفاده می‌شود، در رگرسیون لجستیک از مشخصه‌های آماری استفاده می‌شود که دارای توزیع کای اسکویر هستند. نتایج بدست آمده نشان می‌دهند که متغیر مدت زمان مطالعه با بهبود<sup>(۱)</sup> کای اسکویر برابر با  $۳۴/۹۱۵$  و سطح معنی‌داری  $P=۰/۰۰۰$  و متغیر معدل دیپلم در گام بعدی با بهبود کای اسکویر برابر با  $۴/۷۷۴$  و سطح معنی‌داری  $P=۰/۰۲۹$  وارد معادله شده‌اند. در نتایج بدست آمده،  $2 \log \text{Likelihood}$  که برای برازش مدل مورد استفاده قرار می‌گیرد، در دو مرحله ظاهر شده است. اولی مربوط به گام اول می‌باشد که در آن تنها یک متغیر مستقل وارد معادله شده است. دومین مقدار نیز نشان می‌دهد که با ورود متغیر مستقل بعدی در مدل، مقدار  $2 \log \text{Likelihood}$  اول نسبت به دومین مقدار آن یک بهبودی را نشان می‌دهد، بطوریکه مقدار آن از  $۳۴/۳۱۹$  به  $۲۹/۵۴۶$  کاهش یافته است. تفاسیل این دو، برابر با مقداری است که برای مقدار کای اسکویر بهبود یافته بدست می‌آید. مشخصه Improvement در این قسمت نشان می‌دهد که ورود متغیر جدید تا چه حد مقدار کای اسکویر را کاهش داده است و آیا این کاهش معنادار است یا خیر.

مقدار Model chi-square نشان می‌دهد که متغیر یا متغیرهای مستقل تا چه حد بر متغیر وابسته اثر گذار بوده‌اند. چنانچه مقدار  $p$  بدست آمد کوچکتر از  $۰/۰۵$  باشد تاثیر متغیر یا متغیرهای مستقل و ارتباط آن با متغیر وابسته در سطح  $۹۵$  درصد معنی‌دار تلقی شده و اگر مقدار آن کمتر از

1- Improvement

۰/۰۱ باشد سطح اطمینان به ۹۹ درصد افزایش می‌یابد<sup>(۱)</sup>.

در ادامه برونداد رایانه یک جدول ۲×۲ در قالب Classification table ابعاد دو وجهی متغیر وابسته را نشان می‌دهد. درصدهای این جدول حساسیت مدل در تعیین افراد مربوط به یک حالت را نشان می‌دهد. درصد کل در این جدول نشان می‌دهد که این مدل تا چند درصد کل افراد را به درستی از همدیگر تفکیک می‌کند. براساس نتایج این جدول حساسیت مدل در تعیین افراد مردود زمانیکه تنها متغیر میزان مطالعه وارد معادله شده است ۸۸/۵ درصد و حساسیت آن در تعیین افراد قبول ۹۱/۷ درصد بوده است. بطور کلی این مدل زمانیکه تنها یک متغیر مستقل وارد مطالعه شده است ۹۰ درصد کل افراد را به درستی تفکیک کرده است. اما با ورود دومین متغیر مستقل (معدل دیپلم) حساسیت مدل برای تعیین افراد مردود به ۹۶/۲ درصد و افراد قبول به ۹۵/۸ درصد افزایش یافته و بطور کلی ۹۶ درصد از افراد به درستی تفکیک شده‌اند. این موضوع از طریق نمودارهای گام اول و دوم Classification Plot نیز کاملاً قابل مشاهده است.

جدول Variables in the Equation مقدار ثابت، ضریب B، نسبت برتری<sup>(۲)</sup> (EXP(B))، آماره والد<sup>(۳)</sup> و غیره را نشان می‌دهد. در این قسمت ضریب مثبت متغیرهای مدت مطالعه و معدل دیپلم نشان می‌دهند که با افزایش مدت زمان مطالعه و معدل دیپلم احتمال قبولی در کنکور نیز افزایش می‌یابد. نسبت برتری بیانگر نسبت بین فراوانی تعلق به یک طبقه به فراوانی عدم تعلق به آن طبقه است. آماره والد نیز معنی داری متغیرهای وارد شده در معادله را نشان می‌دهد. بر مبنای مقدار ثابت و ضریب B می‌توان معادله بهینه رگرسیون لجستیک را بشرح ذیل نوشت.

(مدت مطالعه) + ۰/۵۵۲ + (معدل دیپلم) + ۰/۵۸۵ - ۱۲/۲۷۴ = برتری لگاریتمی

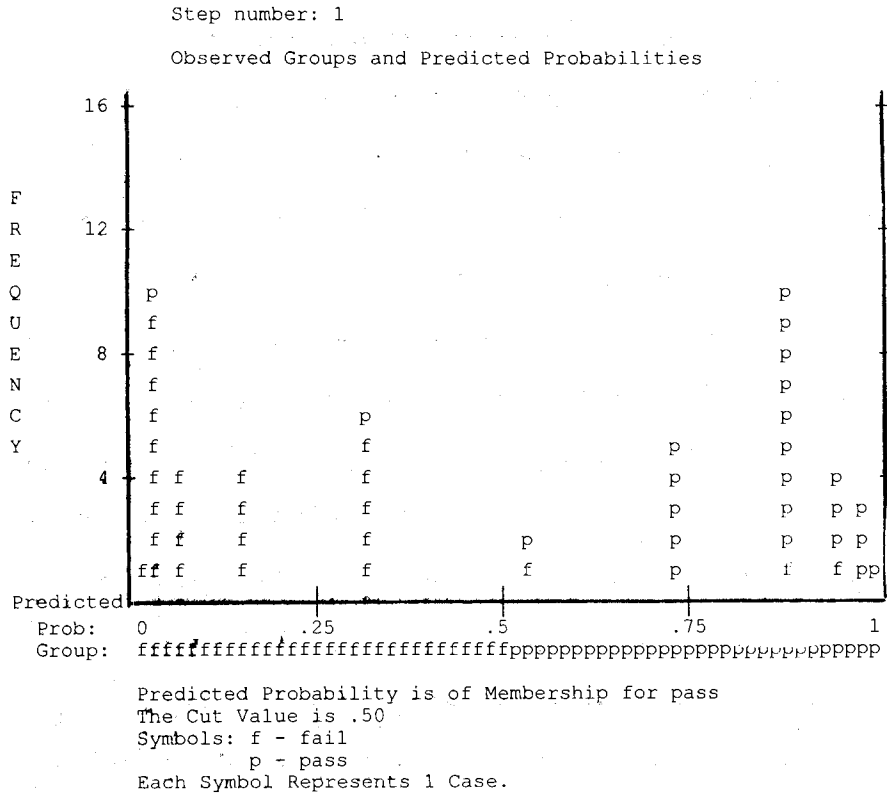
همانطوریکه قبلاً نیز بیان گردید، نتایج جدول توافقی که در قالب Classification Table بدست می‌آید، بصورت نمودار Classification Plot نیز ظاهر می‌گردد (شکل ۹-۲).

۱- هومن، حیدرعلی، (۱۳۸۰) «تحلیل داده‌های چند متغیری در پژوهش رفتاری» نشر پارسا، تهران ص ۱۷۶.

2- Odds ratio

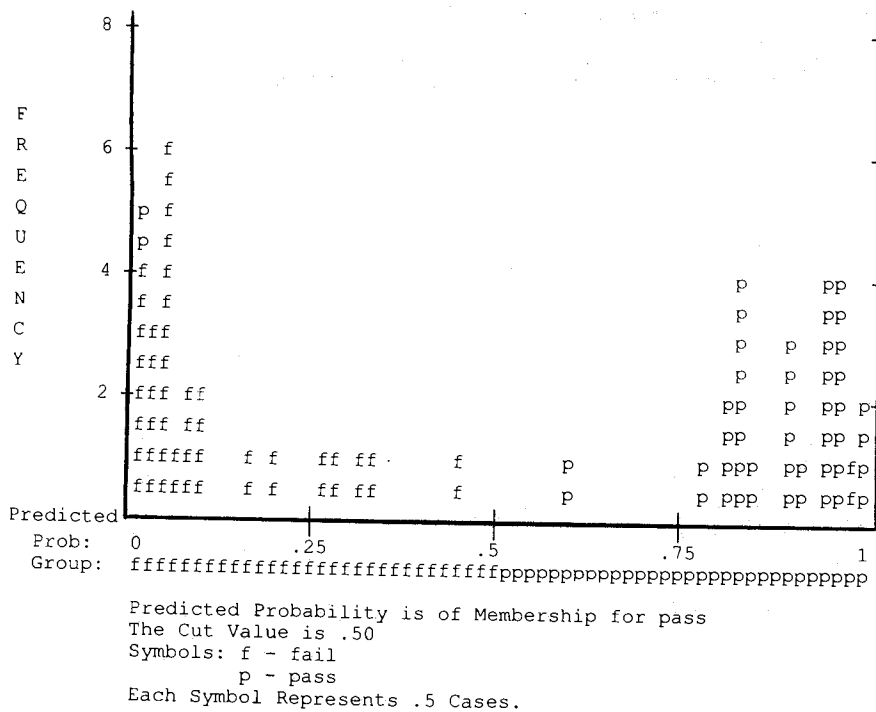
3- Wald test

شکل (۲-۹) نمودار تقسیم‌بندی افراد بر اساس متغیرهای مستقل در رگرسیون لجستیک



Step number: 2

Observed Groups and Predicted Probabilities



همانطوریکه از نمودارهای فوق برمی آید، بدلیل استفاده از روش پیشرو گام به گام نمودارها نیز در دو مرحله رسم شده اند. نمودار اول برای زمانی است که تنها یک متغیر مستقل وارد معادله شده است و نمودار دوم برای مرحله ای است که در آن دو متغیر مستقل وارد معادله گردیده و دقت پیش بینی و تقسیم بندی افراد را افزایش داده است. در این نمودارها در محور افقی احتمال قبولی یا تعلق فرد به گروه یک بین ۰ و ۱ قرار دارد. خط پایینی همان محور نشان می دهد که در صورتی که این احتمال بزرگتر از ۰/۵ باشد فرد قبول و در غیر این صورت فرد مردود پیش بینی خواهد شد. محتوای این نمودار نشان می دهد که چند نفر به درستی و چند نفر به اشتباه از یکدیگر

تفکیک شده‌اند. در این مثال نمودار مرحله اول نشان می‌دهد که با ورود تنها یک متغیر مستقل (مدت زمان مطالعه)، شش نفر به اشتباه تفکیک شده‌اند، در حالیکه براساس نمودار مرحله دوم، با ورود متغیر مستقل دوم (معدل دیپلم) دقت تفکیک افزایش یافته و تنها دو نفر به اشتباه تفکیک شده‌اند. نتایج جدول دو بعدی نیز نشان داد که بطور کلی ۹۶ درصد از افراد به درستی تفکیک شده‌اند.

## فصل دهم

# تحلیل تشخیصی چندگانه

مقدمه:

تحلیل تشخیصی<sup>(۱)</sup> از جمله روشهای تفکیکی است که تلاش می‌کند تا با بهره‌گیری از برخی متغیرهای مستقل، افراد گروهها را که داده‌های آنها بصورت اسمی یا ترتیبی است به بهترین وجه از هم تفکیک کند و نهایتاً متغیرهایی که بطور مناسب گروهها را از هم تفکیک می‌کنند مشخص کند. این روش شباهتها و تفاوتهایی از نظر کاربرد و تحلیل داده‌ها با سایر روشهای چند متغیره دارد که در این فصل مورد بحث قرار گرفته است. برای تحلیل و تبیین بهتر این روش از یک مثال فرضی که با برنامه SPSS محاسبه شده است نیز کمک گرفته شده است.

### کاربرد تحلیل تشخیصی:

تحلیل تشخیصی زمانی به عنوان یک روش مؤثر مورد استفاده قرار می‌گیرد که محقق بخواهد تفاوت‌های موجود بین گروهها را تشخیص دهد و یا اینکه درصد باشد افراد یا واحدهای مورد

مطالعه را به گروهها یا طبقات تقسیم کند. بنابراین، روش تحلیل تشخیصی زمانی مفید می‌باشد که یک متغیر گروه‌بندی شده (کیفی) و چندین متغیر مستقل کمی وجود داشته باشد. در چنین مواردی تحلیل رگرسیون مناسب بنظر نمی‌رسد. زیر در تحلیل رگرسیون، متغیر وابسته باید بصورت کمی باشد. بنابراین تحلیل تشخیصی چندگانه<sup>(۱)</sup> زمانی روش مناسب بحساب می‌آید که در آن متغیر گروه‌بندی در دو یا چند گروه طبقه‌بندی شده باشد و محقق بخواهد ارتباط آن را با تعدادی متغیر مستقل که بصورت کمی هستند بررسی کند. به عنوان مثال اگر محقق بخواهد ارتباط انتخاب یک اولویت مانند الف و ب را در رابطه با درآمد، سن و تحصیلات و غیره بررسی کند باید از تحلیل تشخیصی چندگانه استفاده کند. در بسیاری از مواقع اتفاق می‌افتد که متغیر وابسته در دو یا سه گروه مانند زن و مرد یا کم، متوسط و زیاد طبقه بندی می‌شوند. در هر دو این موارد می‌توان از تحلیل تشخیصی استفاده کرد. در صورتی که تنها دو گروه مورد بررسی قرار گیرند بطور ساده از مفهوم تحلیل تشخیصی و چنانچه سه گروه یا بیشتر مورد مطالعه قرار گیرند تحلیل تشخیصی چندگانه نامیده می‌شود. تحلیل تشخیصی ترکیب دو یا چند متغیر مستقل را که به بهترین وجه تفاوت بین دو گروه را تبیین می‌کند نشان می‌دهد. این موضوع از طریق حداکثر کردن واریانس بین گروهها نسبت به واریانس درون گروهها بر مبنای یک قاعده تصمیم‌گیری آماری انجام می‌گیرد، که بصورت نسبت واریانس بین گروهها به واریانس درون گروههاست.

ترکیب خطی برای تحلیل تشخیصی بر مبنای معادله ذیل انجام می‌گیرد.

$$Z = W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_nX_n$$

در این معادله :

$Z$  = میزان تشخیص (تفاوت)

$W$  = وزن تشخیص

$X$  = متغیرهای مستقل



زمانیکه فرضیه‌ای داشته باشیم مبنی بر اینکه میانگین‌های گروه‌های دو یا چندگانه با هم برابر هستند، برای آزمون آن می‌توان از تحلیل تشخیصی استفاده کرد. برای این کار تکنیک تحلیل تشخیصی، هر متغیر مستقل را در وزن خود ضرب کرده و آنها را با هم جمع می‌کند (براساس معادله بالا). نتیجه بدست آمده در واقع یک تفاوت (تشخیص) ترکیبی برای هر یک از افراد لحاظ شده در تجزیه و تحلیل است. با محاسبه میانگین تفاوت‌های کلیه افراد در یک گروه، میانگین گروه فوق بدست می‌آید. میانگین گروه در واقع به عنوان یک مرکز ثقل<sup>(۱)</sup> تلقی می‌گردد. اگر دو گروه وجود داشته باشد دو مرکز ثقل و اگر سه گروه وجود داشته باشد سه مرکز ثقل و الی آخر خواهد بود. مراکز ثقل، هر یک بیانگر موقعیت خاص هر فرد نسبت به گروه مورد نظر خواهد بود. مقایسه مرکز ثقل گروهها نیز نشان می‌دهند که هر یک در چه جایگاهی نسبت به ابعاد مورد بررسی قرار دارند.

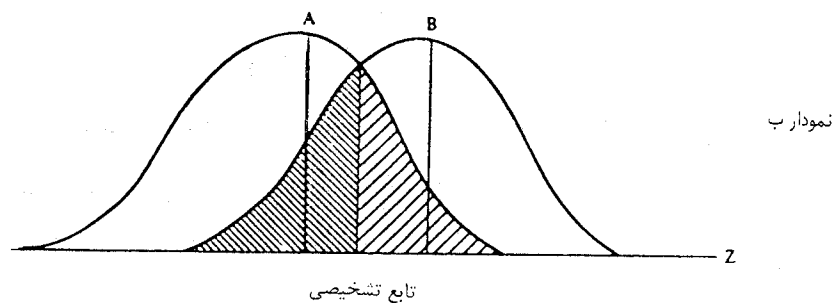
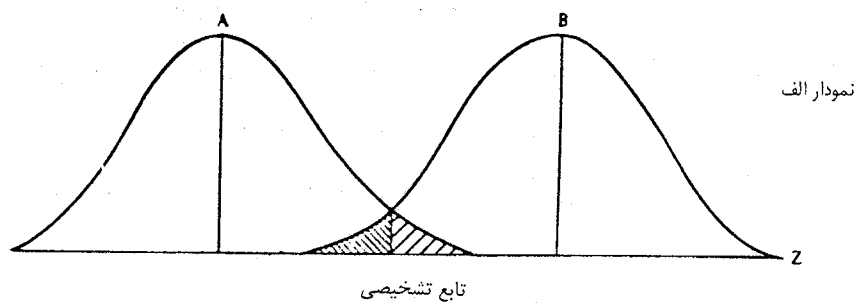
آزمون معنی‌دار بودن تشخیص (تفاوتها) از طریق سنجش فاصله بین مراکز ثقل گروههاست و این از طریق مقایسه تابع توزیع مقادیر تشخیص برای دو یا چند گروه انجام می‌گیرد. چنانچه روی هم افتادگی<sup>(۲)</sup> مقادیر توزیع کم باشد، تابع تشخیصی، گروهها را بطور روشن از هم تفکیک می‌کند. اما چنانچه روی هم افتادگی توزیع بیشتر باشد تابع تشخیصی تفاوت بین گروهها را با شفافیت کامل نشان نمی‌دهد. توزیع مقادیر تشخیص (تفاوتها) در نمودارهای ذیل ارائه شده‌اند.

همانطوریکه از شکل (۱۰-۱) برمی‌آید، نمودار الف وضعیت را نشان می‌دهد که تابع تشخیصی بطور روشنتر گروهها را از هم تفکیک کرده است. در حالیکه تابع نمودار ب در بیان تفاوت بین گروهها از شفافیت لازم برخوردار نمی‌باشد. (منطقه هاشور زده نشان دهنده طبقه‌بندی نادرست واحدهای آماری از A به B می‌باشد)<sup>(۳)</sup>.

1- Centriod

2- Overlap

3- Hair Joseph and et al (1990) "Multivariate Data Analysis" Macmillan Publishing Company. New York. P. 77.



شکل (۱۰-۱) نمایش یک متغیره تشخیصی مقدار  $Z$  (Z-Score)

شباهتها و تفاوت‌های بین تحلیل تشخیصی با تحلیل رگرسیون و تحلیل واریانس : کاربرد و تفسیر تحلیل تشخیصی تا حدود زیادی به تحلیل رگرسیون شباهت دارد، و آن ترکیب خطی دو یا چند متغیر مستقل کمی برای توضیح و تبیین یک متغیر گروه‌بندی وابسته است. تفاوت اصلی بین این دو روش در این نکته نهفته است که تحلیل تشخیصی زمانی مناسب می‌باشد که متغیر گروه‌بندی کیفی باشد در حالیکه در تحلیل رگرسیون متغیر وابسته از نوع کمی است. از سوی دیگر تحلیل تشخیصی قابل مقایسه با تحلیل واریانس نیز می‌باشد. در تحلیل تشخیصی متغیر گروه‌بندی کیفی و متغیرهای مستقل کمی یا پارامتری هستند. در حالیکه در تحلیل

واریانس وضعیت کاملاً عکس است و در آن یک متغیر وابسته یا متغیر اصلی پارامتری یا کمی و یک متغیر مستقل یا متغیر گروه‌بندی غیر پارامتری یا کیفی مورد استفاده قرار می‌گیرد.

مثال: فرض کنید که تولید کننده‌ای در صدد است تا بداند آیا محصول جدید تولید شده مانند جارو برقی نسبت به محصول قبلی، از طرف مصرف‌کنندگان مورد استقبال قرار می‌گیرد یا خیر. برای این کار تولید کننده باید افرادی که حاضر به خرید هستند و افرادی که حاضر به خرید محصول جدید نیستند را شناسایی کند. برای سنجش دلایل خرید یا عدم خرید، تولید کننده ممکن است سه مقیاس یا سه ملاک در سطح فاصله‌ای تعیین کند. این ملاکها در واقع معیارهایی هستند که خریدار بر مبنای آن محصول جدید را ارزیابی می‌کند تا نسبت به خرید یا عدم خرید آن تصمیم‌گیری کند. این سه معیار عبارتند از: ۱- دوام ۲- نحوه کار ۳- شکل ظاهری

در هر یک از موارد فوق پاسخگو می‌تواند امتیاز صفر تا ۱۰ را برای هر یک از آنها اختصاص دهد. ترکیب وزنی این معیارها نشان خواهد داد که تا چه حد احتمال خرید کالای فوق وجود دارد. علاوه بر آن، این روش نشان می‌دهد که کدام معیار بیشتر در تصمیم به خرید مؤثر است و کدامیک بهتر خریداران را از کسانی که حاضر به خرید نیستند تفکیک می‌کند. بعبارت دیگر کدامیک از معیارهای دوام، نحوه کار و شکل ظاهری بهتر خریداران را از کسانی که خریدار نیستند جدا می‌کند. به عنوان مثال اگر پاسخهای «خرید خواهم کرد» با پاسخهای «دوام بالا» همبستگی داشته باشند و پاسخهای «خرید نخواهم کرد» با پاسخهای «بی دوام» همبستگی داشته باشند می‌توان نتیجه گرفت که ملاک دوام بهتر خریداران را از کسانی که خریدار نیستند تفکیک می‌کند. در مقابل اگر به همان میزان که خریداران «شکل ظاهری» را بالا ارزیابی می‌کنند به همان نسبت نیز کسانی که خرید نمی‌کنند، شکل ظاهری را بهتر ارزیابی کنند. در این صورت ملاک وضع ظاهری، معیاری است که کمتر خریداران را از غیر خریداران متمایز می‌کند.

جدول ذیل پاسخهای ده نفر از زنان خانه‌دار را نسبت به محصول جاروبرقی بر مبنای یک قیمت خاص و براساس سه معیار دوام، نحوه کار و شکل ظاهری نشان می‌دهد. در هر مورد پاسخگو

محصول را با محصول قبلی مقایسه کرده و به آن نمره‌ای از ۰ تا ۱۰ را منظور می‌کند. پس از ارزیابی و پاسخ به این معیارها باید از پاسخگو خواسته شود تا نسبت به تکمیل ستون تصمیم به خرید اقدام کند. براساس این مقیاس ۵ پاسخگو در گروه خریدار و ۵ پاسخگو در گروه عدم خرید طبقه‌بندی شده‌اند. همانطوریکه از جدول فوق برمی‌آید، تفاوت بین میانگین پاسخهای کسانیکه حاضر به خرید محصول هستند و کسانیکه حاضر به خرید آن نیستند در رابطه با معیار دوام زیاد می‌باشد  $(۸ - ۳/۲ = ۴/۸)$ .

بنابراین «معیار دوام» ملاک خوبی برای تعیین تشخیص بین خریداران و کسانیکه حاضر به خرید نیستند بحساب می‌آید. از سوی دیگر تفاوت پاسخها در مورد معیار «شکل ظاهری»  $(۴ - ۳/۸ = ۰/۲)$  بسیار پایین می‌باشد. بنابراین معیار فوق دارای بار تشخیصی پایینی در مورد تصمیم به خرید یا عدم خرید دارد.

جدول (۱-۱۰) ارزیابی مشتریان از محصول جدید

تصمیم به خرید	شماره ردیف	نمره ارزیابی از ۰ تا ۱۰		
		دوام $X_1$	نحوه کار $X_2$	شکل ظاهری $X_3$
خواهم خرید	۱	۸	۹	۶
	۲	۶	۷	۵
	۳	۱۰	۶	۴
	۴	۹	۴	۳
	۵	۷	۸	۲
میانگین		۸	۶/۸	۴
نخواهم خرید	۶	۵	۴	۷
	۷	۳	۷	۲
	۸	۴	۵	۵
	۹	۲	۴	۳
	۱۰	۲	۲	۲
میانگین		۳/۲	۴/۴	۳/۸
فاصله بین میانگین‌ها		۴/۸	۲/۴	۰/۲

تکنیک تحلیل تشخیصی چندگانه فرآیندی شبیه به مثال فرضی قبلی را طی می‌کند این روش ابعاد و یا ویژگی‌هایی که در آن بیشترین تشخیص وجود دارد را در بین گروه‌ها شناسایی کرده و ضریب وزنی تشخیصی را برای هر متغیر به منظور انعکاس این تفاوتها بدست می‌آورد.

### مراحل انجام تحلیل تشخیصی:

برای انجام تحلیل تشخیصی باید مراحل ذیل انجام گیرد:

#### ۱ - انتخاب متغیر:

برای اجرای تحلیل تشخیصی محقق باید ابتدا متغیرهای مستقل و متغیر گروه‌بندی را تعیین کند، بطوری که متغیر گروه‌بندی وابسته از نوع غیر پارامتری و کیفی دو یا چند مقوله‌ای و متغیرهای مستقل از نوع پارامتری باشند. تعداد گروه‌های متغیر گروه‌بندی (تعداد طبقات) می‌تواند دو گروه (دو بخشی<sup>(۱)</sup>) باشد مانند زن و مرد و یا چند گروه (چند بخشی<sup>(۲)</sup>) باشد مانند نوع شغل: کشاورز، کارگر، کارمند، آزاد.

در مواردی ممکن است محقق متغیر گروه‌بندی را در سطح فاصله‌ای اندازه گرفته باشد، اما برای استفاده از روش تحلیل تشخیصی آن را به مقیاس ترتیبی تبدیل کند. (مانند طبقه‌بندی بر مبنای میزان درآمد: فقیر، متوسط، ثروتمند). در بعضی مواقع ممکن است تنها دو حداکثر (فقیر - ثروتمند) را مورد تجزیه و تحلیل قرار دهد، که آن را رهیافت دو قطبی<sup>(۳)</sup> می‌نامند. در این شیوه تنها دو گروه با هم مقایسه می‌شوند و گروه یا گروه‌های وسط حذف می‌گردند.

پس از تعیین متغیر وابسته، محقق باید نسبت به انتخاب متغیرهای مستقل اقدام کند، تا پس از جمع‌آوری داده‌ها، مورد تجزیه و تحلیل قرار گیرند. برای وارد کردن داده‌ها به SPSS داده‌های مربوط به هر متغیر باید در یک ستون وارد پنجره Data Editor شود. با توجه به اینکه متغیر

1- Dichotomous

2- Multichotomous

3- Polar- extremes approach

گروه‌بندی بصورت اسمی دو وجهی یا چند وجهی است، بنابراین کدهای اختصاص یافته به آنها باید تعریف شوند.

فرض کنید محقق با انتخاب تعدادی از کشاورزانی که عضو تعاونیهای کشاورزی هستند و تعدادی که عضو این تعاونیها نیستند می‌خواهد از طریق ترکیب خطی تعدادی از متغیرهای مستقل در یک تابع، تعیین کند که چگونه یک فرد در بین این دو گروه (عضو و غیرعضو) قابل تشخیص است. این محقق ممکن است تصور کند که متغیرهای ذیل می‌تواند به این تشخیص کمک کند.

$Y =$  عضویت کشاورزان در تعاونیهای کشاورزی

$X_1 =$  تأمین منابع مالی

$X_2 =$  تأمین خدمات و نهاده‌های کشاورزی

$X_3 =$  فروش و بازاریابی محصول

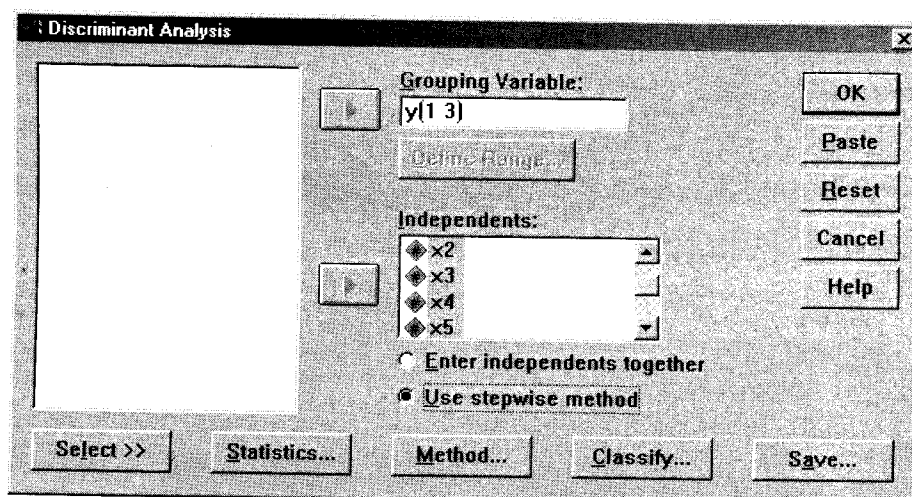
$X_4 =$  کشت یکپارچه

$X_5 =$  انسجام اجتماعی

او ممکن است از طریق گویه‌های مختلف و در قالب چند مقیاس اطلاعات مربوط را جمع‌آوری کند تا از طریق تحلیل تشخیصی این امکان را فراهم کند تا ضریب تشخیصی خطی را که شباهت زیادی به معادله رگرسیون خطی چندگانه دارد بدست آورد.

۲- محاسبه تحلیل تشخیصی با برنامه SPSS:

پس از وارد کردن داده‌ها، در پنجره Data Editor برای انجام محاسبات مربوط به تحلیل تشخیصی در برنامه SPSS باید دستور Analyze/Classify/Discriminant اجرا شود. با اجرای این دستور پنجره‌ای مانند شکل (۱۰-۲) باز می‌شود.



شکل (۲-۱۰) پنجره Discriminant Analysis

در این پنجره متغیرهای وارد شده در Data Editor در سمت چپ لیست می‌شوند در این مرحله باید متغیر گروه‌بندی را به قسمت Grouping Variable و متغیرهای مستقل را به قسمت Independents منتقل کرد. پس از منتقل کردن متغیر گروه‌بندی به قسمت Grouping Variable باید با فشار دادن دکمه پایین آن، مقادیر حداقل و حداکثر آن را مشخص کرد. تحلیل تشخیصی دارای دوروش کلی برای وارد کردن متغیرهای مستقل به تحلیل می‌باشد. این روشها عبارتند از:

۱ - روش وارد کردن کل متغیرهای مستقل بطور همزمان<sup>(۱)</sup>

1- Enter Independents Together

۲- روش گام به گام<sup>(۱)</sup>

در صورتی که از پنجره اصلی تحلیل تشخیصی، مورد گام به گام انتخاب شود در این صورت دکمه Methods روشن می‌شود. با فشار دادن این دکمه پنجره‌ای باز می‌شود که دارای سه قسمت مجزاست. در قسمت Method می‌توان روش مناسب و مورد نظر را برای ورود یا خروج متغیرها در گامهای مختلف انتخاب کرد. آلترناتیوهای مختلف عبارتند از:

Wilks' Lambda

Unexplained Variance

Mahalanobis Distance

Smallest F Ratio

Rao's V

روش Wilks' Lambda روش پیش‌گزیده می‌باشد. در صورت انتخاب Rao's V باید حداقل افزایش در V برای وارد شدن یک متغیر در تحلیل مشخص شود.

در قسمت Criteria نیز می‌توان ملاک خاصی را برای ورود یا خروج یک متغیر برای تحلیل تعیین کرد. در این قسمت می‌توان یکی از دو ملاک ذیل را انتخاب نمود و شرط ورود یا خروج را در هر مورد مشخص نمود. این معیارها عبارتند از:

- Use F Value

- Use Probability of F

قسمت Display نیز نتایج مورد نیاز را در پنجره Output مشخص می‌کند.

در استفاده از تحلیل تشخیصی محقق سعی می‌کند تا از اعتبار تابع تشخیصی<sup>(۲)</sup> اطمینان حاصل کند. روشهای مختلفی برای کسب این اطمینان وجود دارد، که پرکاربردترین و معمول‌ترین آن دو نیمه کردن نمونه مورد بررسی است. در این روش نمونه انتخاب شده به دو یا چند گروه تقسیم

1- Stepwise method

2- Validity of discriminant function



می‌شود که یکی از آنها نمونه مورد تجزیه و تحلیل<sup>(۱)</sup> می‌باشد که برای تهیه تابع تشخیصی مورد استفاده قرار می‌گیرد و دیگری نمونه بسط یافته<sup>(۲)</sup> یا حاکی است که برای آزمون تابع تشخیصی استفاده می‌شود. این فرآیند تعیین اعتبار تابع را دو نصف کردن نمونه<sup>(۳)</sup> و یا رهیافت اعتبار تقاطعی<sup>(۴)</sup> می‌گویند.

توجیه قطعی برای دو نیمه کردن نمونه وجود ندارد. اما بیشتر محققان معتقدند که این دو نیمه کردن باید به شیوه‌ای انجام گیرد که نیمی در «نمونه مورد تحلیل» و نیمی دیگر در «نمونه بسط یافته یا حاکی» قرار گیرد. برخی دیگر از محققان نیز نسبت‌های ۴۰-۶۰ یا حتی ۲۵-۷۵ را نیز تجویز کرده‌اند. در تقسیم نمونه به دو گروه اگرچه تعداد قطعی برای نمونه‌ها وجود ندارد، اما بهتر است حداقل حجم نمونه ۱۰۰ نفر باشد تا به دو گروه تقسیم شوند<sup>(۵)</sup>.

برای تقسیم نمونه از پنجره اصلی تحلیل تشخیصی، دکمه Classify را فشار دهید تا پنجره دیگری باز شود. در این پنجره، پنج قسمت عمده وجود دارد. قسمت Prior Probabilities برای طبقه بندی نمونه مورد استفاده قرار می‌گیرد. در اینجا می‌توان گروهها را برابر در نظر گرفت (و این پیش‌گزیده در برنامه SPSS است) و یا می‌توان تعیین تعداد اعضای گروه را به اندازه گروهها واگذار کرد. در قسمت Display نیز می‌توان نتایجی که مورد نیاز می‌باشد را مشخص کرد. در قسمت پایین پنجره فوق چنانچه مورد Replace missing values with mean انتخاب شود، در این صورت میانگین متغیر مستقل فقط در مرحله طبقه بندی جایگزین مقادیر گم شده<sup>(۶)</sup> می‌گردد. در صورت نیاز با فشار دادن دکمه Statistics از پنجره اصلی تحلیل تشخیصی، می‌توان آماره‌های مورد نیاز را انتخاب کرد تا محاسبات لازم انجام گیرد.

1- Analysis Sample

2- Holdout Sample

3- Split- sample

4- Cross- validation approach

5- Hair joseph and etal (1990) "Multivariate Datd Analysis" macmillan publishing company.

6- Missing values

بعد از انجام مراحل فوق و انجام انتخابهای لازم دکمه OK را فشار دهید تا نتایج در پنجره Output ظاهر شود. براساس مثال مورد نظر داده‌های مربوط به ۱۰۰ نفر از کشاورزان در مورد پنج متغیر مستقل و یک متغیر گروه‌بندی مورد تجزیه و تحلیل قرار گرفته است که نتایج آن در صفحات بعد ارائه شده است.

تفسیر نتایج در تحلیل تشخیصی:

اولین مرحله در تفسیر یافته‌ها بررسی تابع تشخیصی برای تعیین اهمیت هر متغیر مستقل در تمیز گروهها از همدیگر است. سه روش عمده برای تعیین اهمیت متغیرهای مستقل در تعیین تمایز بین گروهها وجود دارد.

۱ - وزنه‌های تشخیصی استاندارد شده<sup>(۱)</sup>

۲ - همبستگی‌های ساختار تشخیصی<sup>(۲)</sup>

۳ - مقدار F بخشی<sup>(۳)</sup>

روش سنتی در تفسیر تابع تشخیصی<sup>(۴)</sup> بررسی علامت و بزرگی مقدار وزن تشخیصی استاندارد شده مربوط به هر متغیر است که آنرا ضریب تشخیصی نیز می‌گویند. براساس این ضرایب، هر متغیری که مقدار وزن تشخیصی استاندارد شده بالایی داشته باشد، سهم بیشتری در تعیین توان تشخیصی بخود اختصاص خواهد داد. بدون در نظر گرفتن علائم این مقادیر، سهم و نقش هر متغیر در معادله تابع تشخیصی تعیین می‌گردد. علائم مثبت و منفی تنها جهت این نقش را تعیین می‌کنند و تفسیر این وزنها نیز درست مانند مقادیر بتا در تحلیل رگرسیون می‌باشد.

اخیراً ماتریس ساختار<sup>(۵)</sup> که آن را همبستگی ساختار<sup>(۶)</sup> نیز می‌گویند در تفسیر نتایج تحلیل

1- Standardized Discriminant Weights

2- Discriminant Structure Correlations

3- Partial F- Value

4- Discriminant Functions

5- Structure Matrix

تشخیصی بطور وسیع مورد استفاده محققان قرار می‌گیرد. مقادیر ماتریس ساختار در واقع نشان دهنده میزان همبستگی خطی<sup>(۷)</sup> بین هر متغیر مستقل و تابع تشخیصی است. بعبارت دیگر مقادیر ماتریس ساختار انعکاس دهنده مقدار واریانسی است که توسط هر متغیر مستقل در خصوص تابع تشخیصی تبیین می‌گردد.

زمانیکه تحلیل تشخیصی به شیوه گام به گام محاسبه گردد، مقدار F بخشی نیز می‌تواند وسیله دیگری برای تفسیر توان تشخیصی متغیرهای مستقل باشد. هر متغیری که دارای مقدار F بزرگتر باشد توان تشخیصی بیشتری خواهد داشت.

با مرور روشهای تفسیر داده‌ها در تحلیل تشخیصی، سؤال اساسی این است که کدامیک از موارد فوق باید مبنای تفسیر قرار گیرد؟ آنچه که مسلم می‌باشد این است که امروزه نرم افزارهای مختلف رایانه‌ای در تحلیل داده‌ها مورد استفاده قرار می‌گیرند. بنابراین هر روشی که در نرم افزارهای فوق وجود داشته باشد می‌تواند مبنای تفسیر قرار گیرد. اما ماتریس ساختار مبنای بهتری برای تفسیر نتایج می‌باشد.

پس از انجام محاسبات تحلیل تشخیصی، معنی داری تابع تشخیصی بوسیله کای اسکور مشخص می‌گردد (جدول ۱۰-۶). همبستگی کانونی<sup>(۸)</sup> نیز در جدولی تحت عنوان Eigenvalues (جدول ۱۰-۵) ارائه می‌شود. در این مثال مقدار همبستگی کانونی ۰/۹۱۶ بدست آمده است. با مجذور این عدد، مقدار ۰/۸۳۹ بدست می‌آید. از این عدد می‌توان نتیجه گرفت که ۸۳/۹ درصد از واریانس متغیر گروه‌بندی توسط این مدل که در آن پنج متغیر مستقل وارد شده است تبیین می‌شود. ماتریس ساختار<sup>(۹)</sup> بدون توجه به علامت مقادیر، به ترتیب از بیشترین تا کمترین در جدول مرتب می‌شوند (جدول ۱۰-۸). مرکز ثقل گروهها<sup>(۱۰)</sup> در جدول مستقلی ارائه می‌شود (جدول

6- Structure Correlations

7- Linear Correlation

8- Canonical Correlation

9- Structure Matrix

10- Group Centroids

۱۰-۱). این مقادیر در واقع میانگین مقادیر استاندارد شده (Z-Score) هر یک از گروههاست. اگر این مقادیر در تعداد افراد گروههای مربوط به خود ضرب و با همدیگر جمع شوند نتیجه صفر خواهد شد.

$$(-2/259 \times 50 + 2/259 \times 50 = 0)$$

در تحلیل تشخیصی به شیوه گام به گام روشهای مختلفی در برنامه SPSS وجود دارد. روش ماهالانویس<sup>(۱)</sup> بر مبنای مربع فاصله اقلیدسی<sup>(۲)</sup> انجام می‌گیرد. این روش داده‌ها را در فضای اصلی مورد تجزیه و تحلیل قرار داده و به همین دلیل امکان گم شدن اطلاعات در آن کمتر است، در حالیکه در سایر روشها مانند روش فیشر<sup>(۳)</sup> بدلیل کاهش در ابعاد مورد تجزیه و تحلیل میزان گم شدن اطلاعات بالاست. قاعده اصلی در روش ماهالانویس در رابطه با ورود متغیرهای مستقل به تحلیل، افزایش مربع فاصله ( $D^2$ ) بین گروهها است.

همانطوریکه قبلاً نیز بیان گردید، در تحلیل تشخیصی از دو روش می‌توان بهره جست، در روش اول می‌توان کلیه متغیرهای مستقل را یک جا در تحلیل وارد کرد و یا از روش گام به گام استفاده نمود. در روش گام به گام، در هر مرحله تنها یک متغیر مستقل براساس توان تشخیصی<sup>(۴)</sup> وارد تابع تشخیص می‌گردد. در این روش ابتدا بهترین متغیر انتخاب و وارد تحلیل می‌شود. سپس در مرحله بعدی، اولین متغیر انتخاب شده با هر یک از متغیرهای مستقل بعدی جفت شده و مناسب‌ترین متغیر دوم انتخاب و وارد تحلیل می‌گردد. دومین متغیر مناسب، متغیری خواهد بود که در ترکیب با متغیر اول، به بهترین وجه توان تشخیصی تابع را افزایش دهد. متغیرهای سوم و بعدی نیز درست به همین شیوه انتخاب و وارد تحلیل می‌گردند. با ورود متغیرهای جدید ممکن است برخی از متغیرهای قبلی از تحلیل خارج شوند. این در صورتی اتفاق می‌افتد که اطلاعات مربوط به متغیرهای قبلی از طریق ترکیب متغیرهای جدید حاصل شده باشد.

1- Mahalanobis Procedure

2- Squared Eudclidean Distance

3- Fisher's Method

4- Discriminant Power

باتوجه به مثال مورد نظر در این بخش، متغیر گروه‌بندی، وضعیت عضویت کشاورزان در تعاونیهای کشاورزی است که به دو گروه عضو و غیرعضو تقسیم شده است. از آنجا که متغیر گروه‌بندی از نوع مقیاس اسمی می‌باشد، بنابراین برای تعیین میزان تأثیر هرکدام از متغیرهای مستقل از روش تحلیل تشخیصی که شباهت بسیاری به تحلیل رگرسیون دارد استفاده شده است. در این روش مقیاس تمام متغیرهای مستقل از نوع پارامتری بوده و مقیاس متغیر گروه‌بندی از نوع اسمی می‌باشد. در این مثال نیز داده‌ها از طریق مقیاسهای لیکرتی جمع‌آوری و بصورت فاصله‌ای درآمده‌اند. همانطوریکه مشخص است در تحلیل تشخیصی انتخاب متغیرها برای ورود به تحلیل بر مبنای حداقل مقدار لامبدای ویلکز<sup>(۱)</sup> انجام می‌گیرد. لامبدا نسبت مجموع مجزورات درون گروهی به کل مجموع مجزورات است. این مقدار نسبتی از واریانس نمرات تشخیصی که نمی‌توان از طریق تفاوت گروهها تبیین کرد را نشان می‌دهد. پس هرگاه میانگین مشاهده شده گروهها با هم برابر باشند، لامبدا نیز برابر است و هرگاه نسبت مجموع مجزورات درون گروهی به کل مجموع مجزورات کوچک باشد لامبدا نیز کوچک بوده و معنای آن این است که میانگین گروهها متفاوت می‌باشد<sup>(۲)</sup>.

قدرت تحمل<sup>(۳)</sup> نسبتی از تغییرپذیری در یک متغیر مستقل است که توسط سایر متغیرهای مستقل تبیین نشده است. لذا برای یک متغیر خارج از معادله این امکان وجود دارد که قدرت تحمل قابل قبولی داشته باشد، اما وقتی آن متغیر وارد معادله شود باعث کاهش قدرت تحمل سایر متغیرها می‌گردد، و قدرت تحمل همه متغیرهای معادله مجدداً محاسبه می‌شود. اگر قدرت تحمل این متغیر یا متغیرهای دیگری که قبلاً وارد معادله شده‌اند کمتر از ۰/۰۱ باشد، این متغیرها وارد معادله نمی‌شوند، مگر اینکه قدرت تحمل پیش فرض عوض شود<sup>(۴)</sup>.

1- Wilk's Lambda

۲- سرمدی و دیگران (۱۳۷۶)

3- Tolerance

۴- پورکریمی جواد (۱۳۷۹) «بررسی عوامل مؤثر بر خوداشتغالی فارغ التحصیلان مقطع کارشناسی دامپروزی» پایان نامه

با این توضیحات مقدماتی نتایج حاصل از تحلیل تشخیصی بوسیله نرم‌افزار SPSS در جداول ذیل ارائه شده‌اند.

جدول (۱۰-۲) متغیرهای تحقیق را قبل از ورود به تحلیل تشخیصی نشان می‌دهد. همانطوریکه از جدول فوق برمی‌آید، با توجه به اینکه متغیر تأمین منابع مالی دارای کمترین مقدار لامبدا و پلکز می‌باشد، طبیعی است که این متغیر اولین متغیری خواهد بود که در تحلیل تشخیصی به شیوه گام به گام وارد معادله خواهد شد. اما با وارد شدن این متغیر به معادله ترکیب مقادیر لامبدا مجدداً تغییر خواهد کرد. در مرحله بعدی نیز مجدداً متغیری که دارای کمترین مقدار لامبدا باشد وارد تحلیل می‌گردد.

جدول (۱۰-۲) متغیرهای تحقیق قبل از ورود به تحلیل

Step	Tolerance	Min Tolerance	F to Enter	Wilks' Lambda
تأمین منابع مالی	1.000	1.000	266.511	0.269
تأمین خدمات و نهاده‌های کشاورزی	1.000	1.000	5.891	0.943
فروش و بازاریابی	1.000	1.000	47.349	0.674
کشت یکپارچه	1.000	1.000	87.820	0.527
انسجام اجتماعی	1.000	1.000	1.138	0.989

جدول (۱۰-۳) متغیرهایی که در مراحل مختلف وارد معادله شده‌اند را نشان می‌دهد. همانطوریکه از این جدول برمی‌آید، متغیرهای فوق در پنج گام یکی پس از دیگری براساس کمترین مقدار لامبدا وارد تحلیل شده‌اند. در این تحلیل متغیرهای تأمین منابع مالی، فروش و بازاریابی، تأمین خدمات و نهاده‌های کشاورزی، کشت یکپارچه و انسجام اجتماعی به ترتیب یکی پس از دیگری وارد معادله تحلیل تشخیصی شده‌اند.

جدول (۱۰-۳) متغیرهایی که در گامهای مختلف وارد تحلیل تشخیصی شده‌اند.

Step	Min Tolerance	F to Enter	Wilks' Lambda
1	1.000	266.511	
2	.986	198.388	.674
	.986	20.786	.269
3	.904	222.58	.629
	.964	23.2424	.236
	.905	15.905	.221
4	.894	119.323	.379
	.964	19.785	.203
	.883	18.413	.201
	.953	12.331	.190
5	.868	124.300	.374
	.941	21.712	.198
	.648	5.372	.170
	.940	13.426	.184
	.624	4.078	.168

جدول (۱۰-۴) متغیرهایی که در هر مرحله خارج از معادله باقی مانده‌اند را نشان می‌دهد. همانطوریکه مشخص است، در مرحله صفر هیچ یک از متغیرها وارد معادله نشده‌اند. اما کوچکترین لامبدا در این مرحله مربوط به متغیر تأمین منابع مالی می‌باشد که مقدار آن  $0/269$  است. طبیعی است که در اولین گام این متغیر وارد معادله خواهد شد. اما با ورود این متغیر به تحلیل، مقادیر لامبدا مجدداً محاسبه شده و در گام اول آمده است. مقادیر لامبدا برای این گام نشان می‌دهد که متغیر فروش و بازاریابی دومین متغیری خواهد بود که وارد معادله خواهد شد. به این ترتیب متغیر انسجام اجتماعی

بدلیل داشتن بالاترین مقدار لامبدا، آخرین متغیری است که وارد معادله تحلیل تشخیصی شده است.

جدول (۴-۱۰) متغیرهایی که در هر مرحله خارج از معادله باقی مانده‌اند.

Step	Tolerance	Min Tolerance	F to Enter	Wilks Lambd
0	1.000	1.000	266.511	.269
تامین منابع مالی	1.000	1.000	5.891	.943
تامین خدمات و نهاده‌های کشاورزی	1.000	1.000	47.39	.674
فروش و بازاریابی	1.000	1.000	87.820	.527
کشت یکپارچه	1.000	1.000	1.138	.989
انسجام اجتماعی	1.000	1.000		
1	.926	.926	13.538	.236
تامین خدمات و نهاده‌های کشاورزی	.986	.986	20.786	.221
فروش و بازاریابی	.977	.977	12.925	.237
کشت یکپارچه	.920	.920	9.333	.245
انسجام اجتماعی				
2	.905	.904	15.905	.190
تامین خدمات و نهاده‌های کشاورزی	.976	.963	9.904	.201
کشت یکپارچه	.880	.880	13.726	.194
انسجام اجتماعی				
3	.953	.883	12.331	.168
کشت یکپارچه	.633	.633	2.992	.184
انسجام اجتماعی				
4	.624	.624	4.078	.161
انسجام اجتماعی				

جدول (۵-۱۰) مقدار ویژه و جدول (۶-۱۰) مقدار لامبدا و ویلکز را که بیانگر معنی‌داری معادله متمایزکننده می‌باشد، نشان می‌دهند. چنانچه از این جداول برمی‌آید، معادله تشکیل شده برای تمایز دو گروه دارای مقدار کای اسکویر ۱۷۴/۳۴۱ می‌باشد که با درجه آزادی ۵ در سطح بالایی



معنی دار است. این آماره بر معنی داری و قدرت تمیز خوب تابع تشخیص دلالت دارد. عدد ۱ در ستون تعداد معادله، حاکی از تنها تابع تشخیصی است که با دو سطح از متغیر ملاک به دست آمده است و درجه آزادی، تعداد متغیرهای موجود در تابع تشخیصی را نشان می دهد. مقدار ویژه نسبت مجذورات بین گروهی به کل مجموع مجذورات درون گروهی را نشان می دهد و همبستگی کانونی میزان همبستگی بین نمرات تشخیص و سطوح گروه بندی وابسته است.

جدول (۵-۱۰) Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.206 <sup>a</sup>	100.0	100.0	.916

جدول (۶-۱۰) Wilks' lambda

Test of Function (s)	Wilk's Lambda	Chi-square	df	Sig.
1	.161	174.341	5	.000

ضرایب استاندارد شده و استاندارد نشده معادله متمایز کننده کانونی یا متعارف در جداول (۷-۱۰) و (۸-۱۰) ارائه شده است. ضرایب استاندارد شده بیانگر اهمیت نسبی هریک از متغیرها در تمایز بین گروههای مورد نظر در متغیر گروه بندی بوده و ضرایب استاندارد نشده، مقادیر ضرایب معادله تشخیصی یا متمایز کننده دو گروه عضو و غیر عضو تعاونیهای کشاورزی می باشند.

جدول (۷-۱۰) Standardized canonical Discriminant Function coefficients

متغیرها	Function
	1
تامین منابع مالی	.884
تامین خدمات و نهاده‌های کشاورزی	-.315
فروش و بازاریابی	.488
کشت یکپارچه	.398
انسجام اجتماعی	-.282

جدول (۸-۱۰) Canonical Discriminant Function coefficients

متغیرها	Function
	1
$(x_1)$ تامین منابع مالی	.483
$(x_3)$ تامین خدمات و نهاده‌های کشاورزی	-.082
$(x_2)$ فروش و بازاریابی	.045
$(x_5)$ کشت یکپارچه	.045
$(x_4)$ انسجام اجتماعی	-.073
(Constant)	-7.507

براساس مقادیر ضرایب استاندارد نشده، معادله رگرسیونی متمایز کننده دو گروه عضو و

غیرعضو را می‌توان بشرح ذیل نوشت:

$$y = -7/50.7 + 0/483(x_1) - 0/082(x_3) + 0/045(x_2) + 0/045(x_5) - 0/073(x_4)$$

جدول (۹-۱۰) Structure Matrix

متغیرها	Function
	1
تامین منابع مالی	.723
کشت یکپارچه	.415
فروش و بازاریابی	.305
تامین خدمات و نهاده‌های کشاورزی	-.107
انسجام اجتماعی	-0.47

جدول (۱۰-۱۰) Function at Group centroids

عضویت کشاورزان در تعاونیه‌های تولیدی	Function
	1
عضو	2.259
غیر عضو	-2.259

محاسبه تحلیل تشخیصی برای مواردی که در آن متغیر گروه‌بندی دارای سه طبقه می‌باشد نیز شبیه به فرآیندی است که برای دو گروه انجام گرفت. اما در اینجا زمانیکه متغیر گروه‌بندی در پنجره اصلی تحلیل تشخیصی در SPSS به قسمت مربوطه وارد گردید، باید با فشار دادن دکمه پایین آن، محدوده متغیر وابسته را با مقادیر ۱ و ۳ (به عنوان مقادیر حداقل و حداکثر) تعریف کرد. زمانیکه سه طبقه در متغیر گروه‌بندی وجود داشته باشد، در این صورت بجای یک تابع کانونی، دو تابع کانونی برای تشخیص گروه‌های سه گانه محاسبه می‌شود. تابع اول در واقع یکی از گروه‌ها را از دو گروه دیگر جدا می‌کند و تابع دوم نیز دو گروه باقیمانده را از هم تفکیک می‌کند. در تحلیل تشخیصی تعداد توابعی که می‌توان ساخت برابر با تعداد گروه‌ها منهای یک است. البته در تحقیقات اجتماعی اقتصادی معمولاً استفاده بیش از سه تابع مرسوم نمی‌باشد. با استفاده از مقادیر تشخیصی این توابع می‌توان نمودار دوبعدی بدست آورد و مقدار ثقل گروه‌ها و مقادیر تشخیص هر یک را مشخص کرد و مرزهای هر یک از گروه‌ها را تعیین نمود و رسم چنین نموداری از طریق SPSS امکان پذیر می‌باشد.

## فصل یازدهم

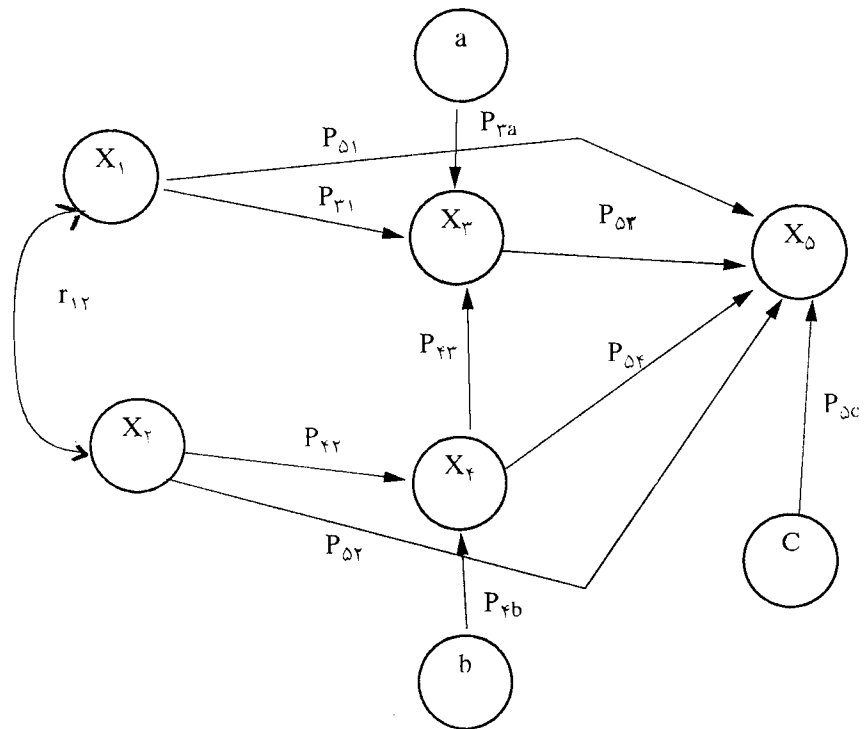
# تحلیل مسیر

### مقدمه:

در حوزه مطالعات اجتماعی و اقتصادی روشهای تحلیل چند متغیره زیادی وجود دارند که به بررسی اثرات و روابط بین متغیرهای مورد مطالعه می پردازند. این روشها عمدتاً اثرات مستقیم یک متغیر بر متغیر دیگر را مورد بررسی قرار می دهند. اما در این میان تحلیل مسیر از جمله تکنیکهای چند متغیره می باشد که علاوه بر بررسی اثرات مستقیم متغیرهای مستقل بر متغیر وابسته، اثرات غیر مستقیم این متغیرها را نیز مد نظر قرار می دهد و روابط بین متغیرها را مطابق با واقعیتهای موجود، در تحلیل وارد می کند. این فصل به بررسی این تکنیک آماری پرداخته و اصول کاربرد آنرا مورد بحث قرار داده است. برای آشنایی با نحوه محاسبه آن از یک مثال فرضی نیز کمک گرفته شده است.

## اصول تحلیل مسیر:

مفهوم تحلیل مسیر<sup>(۱)</sup> برای اولین بار توسط یک بیولوژیست بنام سوال راییت<sup>(۲)</sup> در سال ۱۹۳۴ در ارتباط با تجزیه کردن مجموع همبستگی بین دو متغیر در یک سیستم علی معرفی گردید. تکنیک تحلیل مسیر بر پایه مجموعه‌ای از تحلیل رگرسیون چندگانه و براساس فرض ارتباط بین متغیرهای مستقل و وابسته استوار است. این روش بر استفاده ابتکاری از نمودار تصویری<sup>(۳)</sup> که به دیاگرام مسیر<sup>(۴)</sup> معروف است تأکید خاص دارد.



شکل (۱-۱۱) نمودار تحلیل مسیر

1- Path Analysis

2- Sewall Wright

3- Visual Diagram

4- Path Diagram

دیاگرام مسیر به منظور بیان تصویری روابط بین مجموعه متغیرهای مورد نظر در تحلیل مسیر بکار می‌رود. این دیاگرام معمولاً ترکیبی از متغیرهای بیرونی و متغیرهای درونی است. منظور از متغیر بیرونی، متغیری است که تغییرات آن تحت تاثیر عواملی است که در خارج مدل قرار دارد. و متغیر درونی متغیری است که تغییرات آن به کمک متغیرهای بیرونی و درونی دیاگرام مسیر تبیین می‌شود. اگر چه در تحلیل مسیر این دو متغیر مورد توجه قرار می‌گیرد، اما برای ساده سازی دیاگرام مسیر ممکن است متغیرهای بیرونی از دیاگرام حذف گردند. برای آشنایی با متغیرهای بیرونی و درونی در تحلیل مسیر به شکل (۱-۱۱) توجه کنید.

همانطوریکه از دیاگرام فوق برمی‌آید، به جز متغیرهای  $X_1$ ،  $X_2$  بقیه متغیرها بوسیله یک فلش مستقیم یک طرفه بهم وصل شده‌اند که بیانگر وجود یک رابطه علت و معلولی است. در حالیکه متغیرهای  $X_1$ ،  $X_2$  بوسیله یک خط منحنی با فلش دو طرفه نشان داده شده‌اند. این خط منحنی بیانگر این است که رابطه بین این دو متغیر یک رابطه علت و معلولی نیست، بلکه این متغیرها متغیرهای بیرونی می‌باشند و رابطه بین آنها نیز بصورت تبیین نشده باقی می‌ماند. به همین دلیل است که روابط بین متغیرهای درونی از طریق ضرایب بتا بیان می‌شود، در حالیکه روابط بین متغیرهای بیرونی بوسیله ضرایب همبستگی نشان داده می‌شود. علاوه بر متغیرهای  $X_1$ ،  $X_2$  متغیرهای  $a$ ،  $b$  و  $c$  نیز در این مدل به عنوان متغیرهای بیرونی بحساب می‌آیند. از آنجا که تمام واریانس یک متغیر قابل تبیین نمی‌باشد، بنابراین بخش تبیین نشده به حساب متغیرهای باقیمانده بیرونی گذاشته می‌شود. آنچه که در مورد متغیرهای باقیمانده باید مد نظر قرار گیرد این است که، فرض می‌شود که بین یک متغیر باقیمانده با سایر متغیرهای باقیمانده و متغیرهایی که قبل از آن در مدل آمده‌اند همبستگی وجود ندارد<sup>(۱)</sup>. بعبارت دیگر بین متغیرهای  $a$  و  $b$ ،  $a$  و  $c$ ،  $b$  و  $c$ ،  $a$  و  $X_1$ ،  $X_2$  و غیره همبستگی وجود ندارد.

۱- کرلینجر و پدهازور (۱۳۶۶) «رگرسیون چند متغیری در پژوهش رفتاری» ترجمه حسن سرایی، انتشارات مرکز نشر دانشگاهی،

این نمودار بر مبنای روابط علی بین متغیرهای مستقل و متغیر وابسته که در آن روابط و تأثیرات مستقیم و غیرمستقیم متغیرها مشخص می‌گردد تنظیم می‌شود. در طراحی چنین مدلی باید نظم علی و تقدم و تأخر متغیرها در نظر گرفته شود و این روابط براساس استنباط و استنتاج محقق از چارچوب نظری تحقیق تدوین گردد. به همین دلیل پیش شرط استفاده از تحلیل مسیر برخورداری تحقیق از یک چارچوب تئوریک منسجم براساس عوامل تأثیرگذار و روابط منطقی بین این عوامل است. در تنظیم دیاگرام مسیر، متغیرهای تأثیرگذار مقدم باید در سمت چپ و متغیرهای تأثیرگذار متأخر در سمت راست نمودار قرار گیرند، بطوریکه نهایتاً به متغیر وابسته در منتهی‌الیه نمودار در سمت راست ختم شوند.

روابط بین متغیرها و تأثیرات آنها باید توسط فلش‌های جهت‌دار مشخص شوند. آنچه که باید مدنظر باشد این است که روابط علی باید بصورت یک طرفه در نظر گرفته شوند. اگرچه تحلیل مسیر با جهت علی و اثرات دو طرفه نیز وجود دارد. اما در این فصل، تحلیل مسیر با روابط علی یک طرفه مدنظر می‌باشد. تحلیل مسیر بر مبنای مفروضاتی استوار است که عمده‌ترین آنها بشرح ذیل‌اند:

۱- روابط بین متغیرهای موجود در مدل خطی، جمع‌پذیر و علی هستند و روابط انحنایی و تعاملی ملحوظ نمی‌گردند.

۲- متغیرهای باقیمانده با همدیگر و با متغیرهایی که قبل از آن در مدل قرار گرفته‌اند همبسته نیستند.

۳- جریان علیت در دستگاه یک طرفه می‌باشد و علیت متقابل بین متغیرها ملحوظ نمی‌شوند<sup>(۱)</sup>. پس از تنظیم دیاگرام مسیر باید روابط بین متغیرها بصورت اندیس‌هایی مشخص شوند. این اندیس‌ها بوسیله حرف P و شماره متغیرهایی که بوسیله فلش بهم وصل شده‌اند مشخص می‌شوند بطوریکه اولین شماره بیانگر شماره متغیر وابسته و دومین شماره بیانگر شماره متغیر مستقل است.

۱- کرلینجر و یدهازور (۱۳۶۶) «رگرسیون چند متغیری در پژوهش رفتاری» ترجمه حسن سرایی، انتشارات مرکز نشر دانشگاهی،

به عنوان مثال  $P_{51}$  در شکل قبل بیانگر اثر متغیر  $(X_1)$  بر متغیر  $(X_5)$  می‌باشد. این یک اثر مستقیم است اما همین متغیر  $(X_1)$  اثر غیرمستقیمی نیز از طریق  $(X_3)$  بر متغیر  $(X_5)$  می‌گذارد که این مسیر  $P_{31} \rightarrow P_{53}$  نامیده می‌شود.

در تحلیل مسیر از ضریب تعیین ( $R^2$ ) بدست آمده از طریق رگرسیون، میزان مناسب بودن مدل بدست می‌آید. این تکنیک از ضرایب رگرسیون استاندارد شده جزئی<sup>(۱)</sup> (که به وزن بتا معروف است) به عنوان ضرایب مسیر استفاده کرده و اثرات مستقیم هر متغیر را تعیین می‌کند. به عبارت دیگر این ضریب سهم یا وزن متغیر مستقل را در تبیین واریانس متغیر وابسته نشان می‌دهد. روش تحلیل مسیر به محقق کمک می‌کند تا اثرات مستقیم و غیرمستقیم هر متغیر را محاسبه و نقش هر یک از متغیرها را در دیاگرام مسیر مشخص کند.

در تحلیل مسیر اثرات افزایشی متغیرها با اضافه شدن هر متغیر مدنظر قرار می‌گیرد، بطوریکه این روابط از طریق مجموعه‌ای از معادلات بیان می‌شوند که نشان دهنده میزان وابستگی هر متغیر به متغیر قبلی است. اصل اساسی مورد نظر در تحلیل مسیر این است که هر ضریب همبستگی بین دو متغیر را می‌توان به مجموعه‌ای از مسیرها تجزیه کرد. مسیرهای مجزا از طریق ترتیب تقدم و تأخر متغیرهای واسطه، اثرات متغیرها و روابط آنها را نشان می‌دهند.

اهمیت و ارجحیت تحلیل مسیر در مقایسه با تحلیل همبستگی در این است که تحلیل مسیر امکان سنجش اثرات نسبی هر متغیر مقدم یا توضیحی بر متغیرهای بعدی یا وابسته را ابتدا از طریق مشخص کردن مفروضات مربوط به روابط علی و بعد از طریق تعیین اثرات غیرمستقیم متغیرهای مستقل یا توضیحی فراهم می‌کند.

استفاده از تکنیک تحلیل مسیر مستلزم قبول پیش فرضی است مبنی بر اینکه روابط بین متغیرهایی که حداقل در سطح مقیاس شبه فاصله‌ای<sup>(۲)</sup> هستند، بطور خطی با افزایش متغیرها اضافه

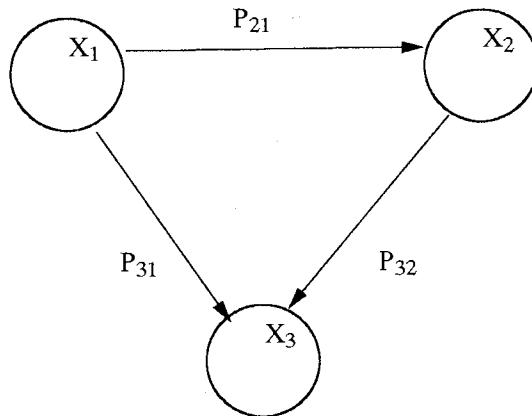
1- Standardized Partial Regression Coefficients

2- Quasi- Interval Scale



می‌گردد. هر متغیر وابسته براساس اثرات متغیرهای قبلی در دیاگرام مسیر و متغیر باقیمانده<sup>(۱)</sup> مورد تحلیل و تبیین قرار می‌گیرد.

متغیر باقیمانده، بخش واریانس تبیین نشده در متغیر وابسته است. متغیرهای مستقل تبیین کننده متغیر وابسته، متغیرهایی هستند که در داخل مدل تحلیل مسیر قرار می‌گیرند و متغیرهای خارجی، که آن را معمولاً با  $e$  نشان می‌دهند مقادیر باقیمانده را به خود اختصاص می‌دهند. شاید بهتر بتوان روش تحلیل مسیر و معادلات آن را با یک مثال ساده بیان کرد. فرض کنید که دیاگرام مسیر ذیل با سه متغیر وجود دارد.



شکل (۱۱-۲) دیاگرام ساده‌ای از روش تحلیل مسیر

معادله ساختاری برای دیاگرام فوق را می‌توان اینگونه نوشت :

$$\begin{array}{|c|} \hline X_1 \\ \hline X_2 \\ \hline X_3 \\ \hline \end{array} = \begin{array}{|c|} \hline e_1 \\ \hline P_{21}X_1 + e_2 \\ \hline P_{31}X_2 + P_{32}X_1 + e_3 \\ \hline \end{array} = P_X + e$$

1- Residual Variable

در این معادله متغیرهای  $X$  براساس انحراف از میانگین شان اندازه‌گیری شده‌اند و  $P_{21}$  نیز می‌تواند از طریق رگرسیون  $X_2$  بر  $X_1$  (بعبارت دیگر  $X_2 = b_{21}X_1$ ) و  $P_{31}$  و  $P_{32}$  نیز از طریق رگرسیون  $X_3$  بر  $X_2$  و  $X_1$  بصورت ذیل بدست آید:

$$X_3 = b_{31.2}X_1 + b_{21}X_2$$

در این معادله  $b_{31.2}$  به معنای ضریب رگرسیون جزئی استاندارد شده<sup>(۱)</sup> برای پیشگویی متغیر ۳ از طریق متغیر ۱ زمانیکه اثر متغیر ۲ ثابت نگهداشته شود، می‌باشد.

در تحلیل مسیر ضریب بتا بیانگر اثر مستقیم  $X_j$  ( $j = 1, 2, 3, \dots, P$ ) بر متغیر وابسته است. با مجذور کردن اثر مستقیم، سهم واریانس هر یک از متغیرهای مستقل ( $X_j = 1, 2, 3, \dots, P$ ) در متغیر وابسته ( $Y$ ) بدست می‌آید<sup>(۲)</sup>.

#### تجزیه همبستگی:

پس از محاسبه اثرات مستقیم یک متغیر بر متغیر دیگر، محقق ممکن است بخواهد مجموع اثرات غیرمستقیم متغیر فوق بر متغیر وابسته را از طریق کم کردن ضریب بتا ( $b_j$ ) از ضریب همبستگی ( $r_{yxj}$ ) بدست آورد. این عمل بر مبنای تجزیه همبستگی امکان‌پذیر است. همانطوریکه قبلاً نیز اشاره شد، تحلیل مسیر برای تبیین دقیق‌تر روابط علی متغیرها به تجزیه همبستگی بین متغیرها می‌پردازد. از طریق این تجزیه، اثرات مستقیم و غیرمستقیم یک متغیر، بر متغیر دیگر مشخص می‌گردد. آنچه که ضرایب همبستگی نشان می‌دهند اثرات مستقیم و غیرمستقیم متغیرهاست. بعبارت دیگر:

$$\text{اثرات غیر مستقیم} + \text{اثرات مستقیم} = \text{همبستگی}$$

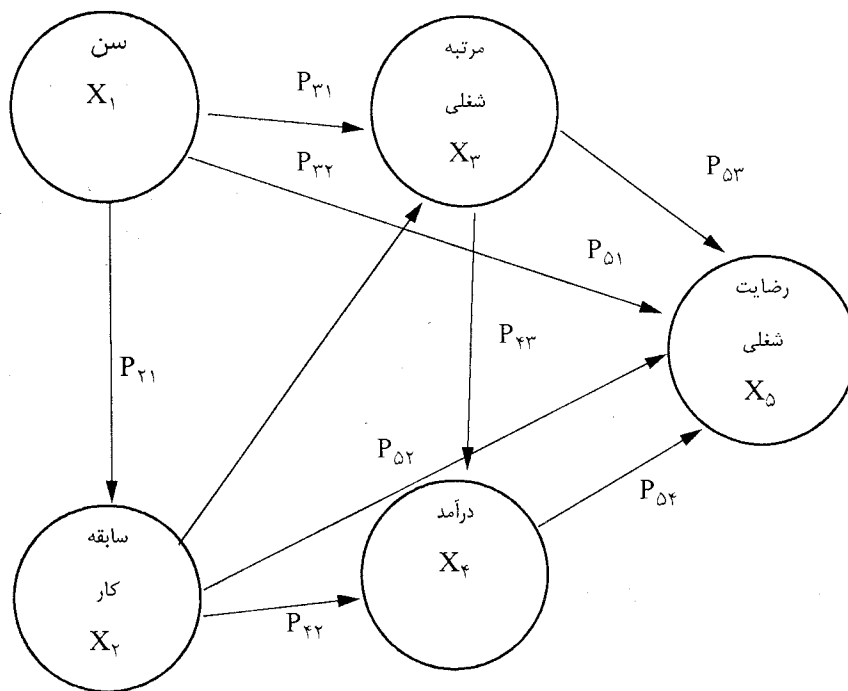
1- Standardized Partial Regression Coefficient

2- Kothari, C.R. (1994) "Research Methodology: Methods and Techniques" V.S. Johri, New Delhi. PP. 398-399.

بر همین استدلال چنانچه بخواهیم مجموع اثرات غیرمستقیم یک متغیر بر متغیر دیگر را بدست آوریم، کافی است تا اثرات مستقیم آن را از میزان ضریب همبستگی پیرسون دو متغیر فوق کم کنیم.

### محاسبه تحلیل مسیر با استفاده از برنامه SPSS:

فرض کنید که هدف، بررسی میزان تأثیر سن، سابقه کار، مرتبه شغلی و درآمد



شکل (۱۱-۳) نمودار تحلیل مسیر

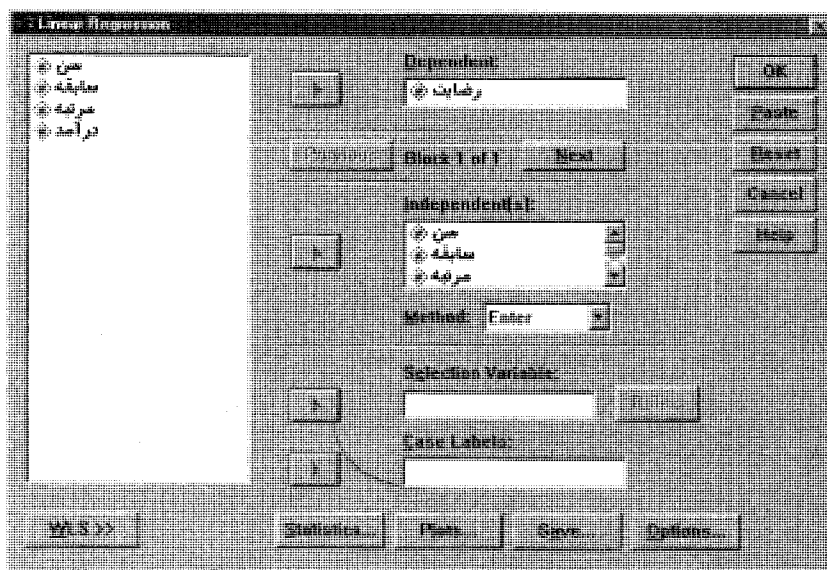
به عنوان متغیرهای مستقل بر رضایت شغلی به عنوان متغیر وابسته است. برای این کار تعداد ۳۰ نفر از کارکنان یک سازمان براساس نمونه‌گیری تصادفی انتخاب و اطلاعات مربوط به متغیرهای فوق جمع‌آوری و وارد برنامه SPSS شده است. برای انجام تجزیه و تحلیل به شیوه تحلیل

مسیر و محاسبه اثرات مستقیم و غیرمستقیم متغیرهای مستقل بر متغیر وابسته، ابتدا باید نمودار مسیر که بیانگر روابط و اثرات مستقیم و غیرمستقیم هر متغیر بر متغیر دیگر است رسم گردد (شکل ۳-۱۱).

برای تعیین ضرایب مسیر و محاسبه اثرات مستقیم و غیرمستقیم متغیرها با استفاده از تکنیک رگرسیون، ابتدا باید براساس نمودار ترسیمی، مسیرها را از هم تفکیک کرد. تفکیک مسیر براساس مبدأ فلشها که بیانگر متغیر مستقل و انتهای فلشها که نشان دهنده متغیر وابسته می باشد انجام می گیرد. در این فرآیند در هر مرحله یکی از متغیرها به عنوان متغیر وابسته و متغیرهای مرتبط با مبدأ فلشهای ختم شده به متغیر فوق به عنوان متغیرهای مستقل در تحلیل رگرسیون مورد استفاده قرار می گیرند تا از این طریق ضرایب بتاها که نشان دهنده اثرات مستقیم متغیرهای مستقل بر متغیر وابسته است بدست آید.

محاسبه رگرسیون در برنامه SPSS از طریق دستور Analyze/ Regression/ Linear

انجام می گیرد. با اجرای این دستور پنجره رگرسیون مانند شکل (۴-۱۱) در مانیتور ظاهر می شود.



شکل (۴-۱۱) پنجره Linear Regression

در این پنجره متغیر وابسته یعنی رضایت شغلی را در قسمت Dependent Variable و متغیرهای مستقل سن، سابقه کار، مرتبه شغلی و درآمد را در قسمت Independent Variables قرار داده و روش Enter را که بصورت پیش گزیده می‌باشد انتخاب و دکمه OK را فشار دهید، تا نتایج بشرح ذیل بدست آید.

جدول (۱-۱) Variables Entered / Removed

Model	Variables Entered	Variables Removed	Method
	سن سابقه کار مرتبه شغلی درآمد		Enter

a. All requested variables entered.

b. Dependent Variable: رضایت شغلی

جدول (۲-۱) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.984 <sup>a</sup>	.969	.964	.27

جدول (۳-۱) ANOVA

Model		Sum of squares	df	Mean Square	F	Sig.
1	Regression	55.568	4	13.892	193.109	.000 <sup>a</sup>
	Residual	1.798	25	7.19AE-02		
	Total	57.367	29			

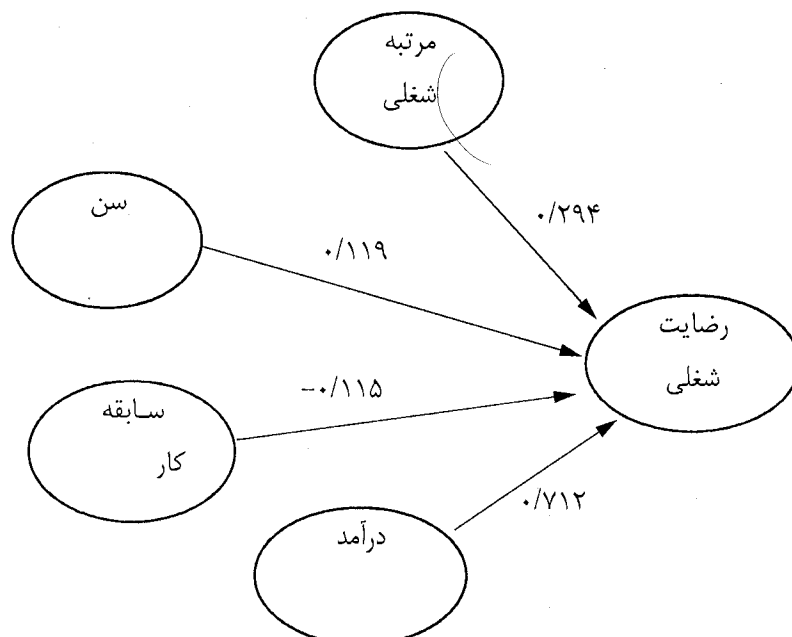
جدول (۴-۱۱) Coefficients<sup>a</sup>

Model		Unstandardized		Standardized	t	Sig.
		Coefficients		Coefficients		
		B	Std.Error	Beta		
1	(Constant)	-6.593E-02	.237	---	-.278	.783
	سن	8.670E-02	.068	.119	1.227	.213
	سابقه کار	-8.582E-02	.064	-.115	-1.333	.195
	مرتبه شغلی	.331	.108	.294	3.052	.005
	درآمد	.840	.086	.712	9.725	.000

a. Dependent Variable: رضایت شغلی

براساس مقادیر بدست آمده در محاسبات فوق می توان اثرات مستقیم متغیرهای مستقل بر

متغیر وابسته رضایت شغلی را در نمودار تفکیک شده شکل (۴-۱۱) مشخص کرد.



شکل (۴-۱۱) نمودار تفکیک شده مرحله اول

در مرحله دوم، مرتبه شغلی به عنوان متغیر وابسته و سن، سابقه کار به عنوان متغیرهای مستقل در نظر گرفته شده و ضرایب بتا با استفاده از روش رگرسیون به شیوه Enter محاسبه و نتایج در ذیل ارائه شده است. اثرات مستقیم این دو متغیر بر متغیر وابسته براساس ضرایب بتای بدست آمده در نمودار تفکیکی (۱۱-۵) ارائه شده است.

جدول (۱۱-۵) Variables Entered / Remored

Model	Variables Entered	Variables Removed	Method
1	سن سابقه کار		Enter

a. Dependent Variable: مرتبه شغلی

جدول (۱۱-۶) Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.905 <sup>a</sup>	.819	.806	.55

جدول (۱۱-۷) ANOVA

Model		Sum of squares	df	Mean Square	F	Sig.
1	Regression	37.256	2	18.629	61.266	.000 <sup>a</sup>
	Residual	8.210	27	.304		
	Total	45.467	29	7.19AE-02		

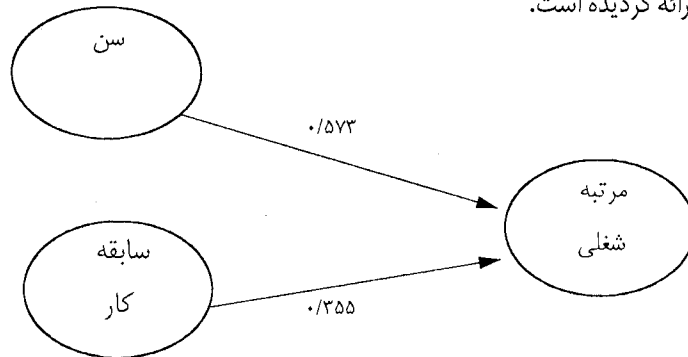
جدول (۱۱-۸) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.231	.228		9.805	.000
	سن	.370	.120	.573	3.091	.005
	سابقه کار	.235	.123	.355	1.918	.066

a. Dependent Variable: مرتبه شغلی

اثرات مستقیم این متغیرها بر متغیر وابسته بر مبنای میزان ضریب بتا در نمودار تفکیکی

(۵-۱۱) ارائه گردیده است.



شکل (۵-۱۱) نمودار تفکیک شده مرحله دوم

در مرحله سوم درآمد به عنوان متغیر وابسته و متغیرهای سابقه کار و مرتبه شغلی به عنوان متغیرهای مستقل در نظر گرفته شده و مقدار بتا با استفاده از رگرسیون به شیوه Enter محاسبه و نتایج آن در ذیل ارائه شده است.

جدول (۹-۱۱) Variables Entered/Removed

Model	Variables	Variables	Method
	Entered	Removed	
1	مرتبه شغلی سابقه کار		Enter

a. Dependent Variable: درآمد

جدول (۱۰-۱۱) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.875 <sup>a</sup>	.766	.748	.60



جدول (۱۱-۱۱) ANOVA

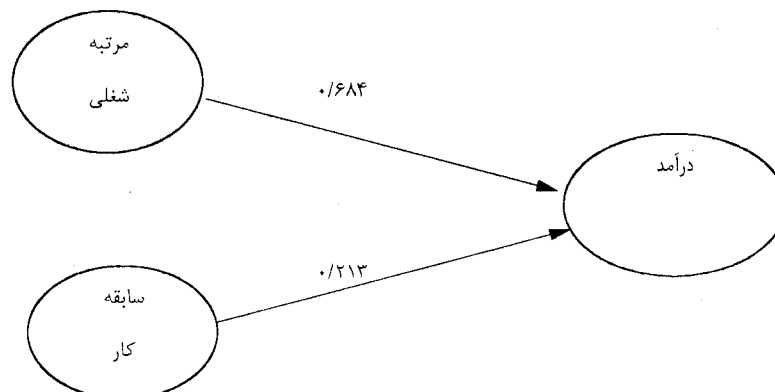
Model	Sum of squares	df	Mean Square	F	Sig.
1 Regression	31.550	2	15.775	44.138	.000 <sup>a</sup>
Residual	9.650	27	.357		
Total	41.200	29			

جدول (۱۲-۱۱) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
		1	(Constant)	.192		
	سابقه کار	.134	.119	.213	1.129	.269
	مرتبه شغلی	.651	.179	.684	3.630	.001

a. Dependent Variable: درآمد

اثرات مستقیم این متغیرها بر متغیر وابسته بر مبنای میزان ضریب بتا در شکل (۱۱-۶) ارائه شده است.



شکل (۱۱-۶) نمودار تفکیک شده مرحله سوم

در مرحله چهارم سابقه کار به عنوان متغیر وابسته و سن (که تنها فلش ختم شده به متغیر وابسته است) به عنوان متغیر مستقل در نظر گرفته شده و میزان بتا محاسبه گردیده است. نتایج این محاسبه در جداول ذیل ارائه شده است.

جدول (۱۱-۱۳) Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
	Fathers		Enter
1	سن		

a. Dependent Variable: سابقه کار

جدول (۱۱-۱۴) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897 <sup>a</sup>	.805	.798	.85

جدول (۱۱-۱۵) ANOVA

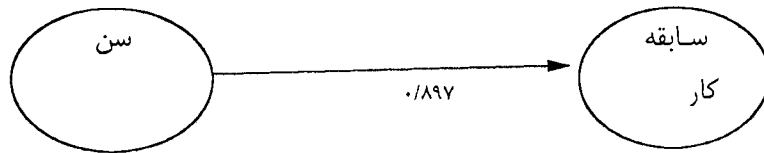
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	83.636	1	83.636	115.755	.000
	Residual	20.231	28	.723		
	Total	103.864	29			

جدول (۱۱-۱۶) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	.517	.337		1.533	.136
	سن	.877	.082	.897	10.759	.000

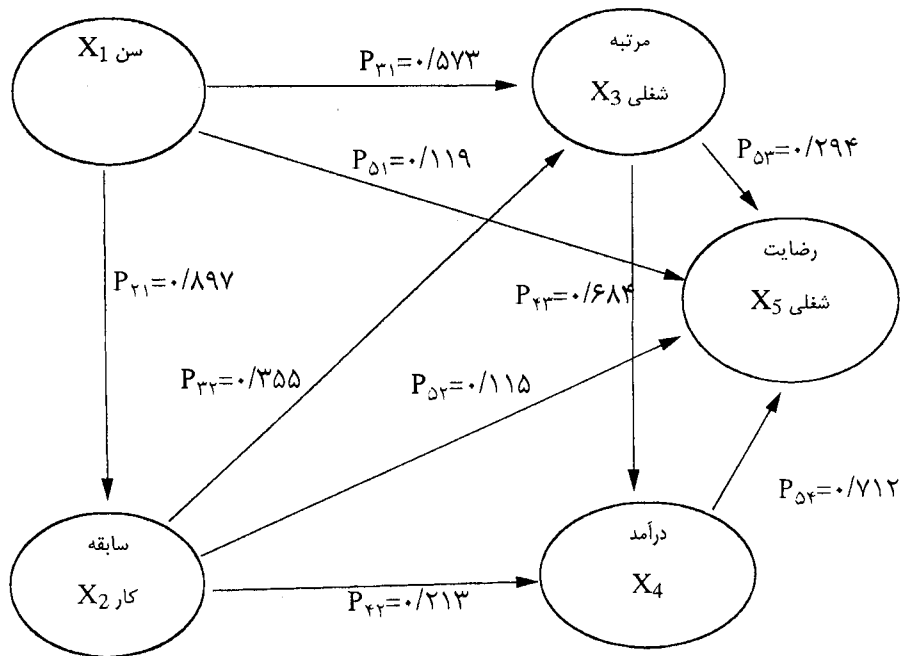
a. Dependent Variable: سابقه کار

براساس مقدار بدست آمده برای ضریب بتا، نمودار تفکیکی آن در ذیل ارائه شده است.



شکل (۷-۱۱) نمودار تفکیک شده مرحله چهارم

پس از آن که ضرایب مسیر برای کلیه مسیرهای تفکیکی بدست آمد، می‌توان از طریق ترکیب نمودارهای فوق اثرات مستقیم و غیرمستقیم متغیرهای مستقل را بر متغیر وابسته اصلی بدست آورد. ضرایب بدست آمده برای کلیه مسیرها در شکل (۸-۱۱) ارائه شده است.



شکل (۸-۱۱) نمودار نهایی تحلیل مسیر، همراه با ضرایب بتا

اینک پس از بدست آمدن ضرایب بتا، می‌توان تأثیرات مستقیم و غیرمستقیم هر یک از متغیرهای مستقل بر متغیر وابسته را محاسبه کرد. از آنجا که در دیاگرام تحلیل مسیر از ضرایب بتا

استفاده می‌شود و این ضرایب نیز بصورت استاندارد شده می‌باشند، بنابراین می‌توان اثرات متغیرهای مختلف را با همدیگر مقایسه کرد و مؤثرترین آنها را تعیین نمود. هر متغیری دارای دو اثر مستقیم و غیرمستقیم می‌باشد که از مجموع آنها اثر کلی متغیر بدست می‌آید. به عنوان مثال متغیر سن ( $X_1$ ) دارای اثر مستقیم بر متغیر رضایت شغلی ( $X_5$ ) است که میزان آن برابر با  $0/119$  می‌باشد. این متغیر علاوه بر اثر مستقیم، دارای اثرات غیرمستقیم دیگری است که از طریق مسیرهای مختلف بدست می‌آید. مسیر  $X_1 \rightarrow X_3 \rightarrow X_5$  یکی از اینگونه مسیرهاست که دارای اثرات غیرمستقیم می‌باشد. همچنین مسیر  $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$  نیز مسیر دیگری از اثرات غیرمستقیم در این نمودار می‌باشد.

برای فهم آسانتر موضوع، اثرات مستقیم و غیرمستقیم متغیر  $X_1$  بر متغیر  $X_5$  در جدول (۱۱-۱۷) ارائه شده است. برای محاسبه اثرات غیرمستقیم، ضرایب بتای هر مسیر را تا رسیدن به متغیر وابسته در هم ضرب می‌کنیم. از آنجا که هر متغیری ممکن است از طریق مسیرهای متعدد بر متغیر وابسته تأثیر بگذارد، بنابراین برای محاسبه کل اثرات غیرمستقیم یک متغیر بر متغیر دیگر، اثرات مسیرهای غیرمستقیم آن دو متغیر را با همدیگر جمع می‌کنیم. برای بدست آوردن اثرات علی کل، باید اثرات مستقیم و غیرمستقیم با همدیگر جمع شوند.

از طریق مقایسه ارقام ضریب بتا برای مسیرهای مختلف، محقق می‌تواند اهمیت هر مسیر را تعیین کند.

جدول (۱۷-۱۱) محاسبه اثرات مستقیم و غیرمستقیم سن ( $X_1$ ) بر رضایت شغلی ( $X_5$ )

نوع اثر	مسیر	میزان اثر بر اساس ضرایب بتا
اثر مستقیم	$X_1 \rightarrow X_5$	۰/۱۱۹
اثر غیر مستقیم	$X_1 \rightarrow X_3 \rightarrow X_5$	$(۰/۵۷۳) \times (۰/۲۹۴) = ۰/۱۶۸$
	$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_5$	$(۰/۸۹۷) \times (۰/۳۵۵) \times (۰/۷۱۲) = ۰/۲۲۶$
	$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$	$(۰/۸۹۷) \times (۰/۳۵۵) \times (۰/۶۸۴) \times (۰/۷۱۲) = ۰/۱۵۵$
	$X_1 \rightarrow X_2 \rightarrow X_5$	$(۰/۸۹۷) \times (-۰/۱۱۵) = -۰/۱۰۳$
	$X_1 \rightarrow X_2 \rightarrow X_4 \rightarrow X_5$	$(۰/۸۹۷) \times (۰/۲۱۳) \times (۰/۷۱۲) = ۰/۱۳۶$
کل اثرات غیر مستقیم		۰/۵۸۲
مجموع اثرات مستقیم و غیر مستقیم		۰/۷۰۱

جدول بالا تنها اثرات مستقیم و غیرمستقیم سن ( $X_1$ ) بر رضایت شغلی ( $X_5$ ) را نشان می‌دهد. اینگونه محاسبات باید برای مسیرهای مستقیم و غیرمستقیم کلیه متغیرها به تفکیک محاسبه و نتیجه نهایی در یک جدول آورده شود. این محاسبات در جداول بعدی آمده است.

جدول (۱۸-۱۱) محاسبه اثرات مستقیم و غیرمستقیم سابقه کار ( $X_2$ ) بر رضایت شغلی ( $X_5$ )

نوع اثر	مسیر	میزان اثر بر اساس ضرایب بتا
اثر مستقیم	$X_2 \rightarrow X_5$	-۰/۱۱۵
اثر غیر مستقیم	$X_2 \rightarrow X_3 \rightarrow X_5$	$(۰/۱۲۳) \times (۰/۲۹۴) = ۰/۰۳۶$
	$X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$	$(۰/۳۵۵) \times (۰/۶۸۴) \times (۰/۷۱۲) = ۰/۱۷۲$
	$X_2 \rightarrow X_4 \rightarrow X_5$	$(۰/۲۱۳) \times (۰/۷۱۲) = ۰/۱۵۲$
کل اثرات غیر مستقیم		۰/۳۶۰
مجموع اثرات مستقیم و غیر مستقیم		۰/۲۴۵

جدول (۱۱-۱۹) محاسبه اثرات مستقیم و غیرمستقیم مرتبه شغلی ( $X_3$ ) بر

رضایت شغلی ( $X_5$ )

نوع اثر	مسیر	میزان اثر بر اساس ضرایب بتا
اثر مستقیم	$X_3 \rightarrow X_5$	۰/۲۹۴
اثر غیر مستقیم	$X_3 \rightarrow X_4 \rightarrow X_5$	$(۰/۶۸۴) \times (۰/۷۱۲) = ۰/۴۸۷$
مجموع اثرات مستقیم و غیر مستقیم		۰/۷۸۱

جدول (۱۱-۲۰) محاسبه اثرات مستقیم و غیرمستقیم درآمد ( $X_4$ ) بر

رضایت شغلی ( $X_5$ )

نوع اثر	مسیر	میزان اثر بر اساس ضرایب بتا
اثر مستقیم	$X_4 \rightarrow X_5$	۰/۷۱۲

پس از محاسبه اثرات مستقیم و غیرمستقیم کلیه متغیرها، مجموع این اثرات در جدول (۱۱-۲۱) خلاصه شده است.

جدول (۱۱-۲۱) مجموع تاثیرات مستقیم و غیر مستقیم متغیرهای مستقل

بر متغیر وابسته

متغیر مستقل	اثرات غیرمستقیم	اثرات مستقیم	مجموع تاثیرات مستقیم و غیرمستقیم هر متغیر
$X_1$ - سن	۰/۵۸۲	۰/۱۱۹	۰/۷۰۱
$X_2$ - سابقه کار	۰/۳۶۰	-۰/۱۱۵	۰/۲۴۵
$X_3$ - مرتبه شغلی	۰/۴۸۷	۰/۲۹۴	۰/۷۸۱
$X_4$ - درآمد	-	۰/۷۱۲	۰/۷۱۲

در تحقیقات اجتماعی - اقتصادی معمولاً شناسایی کلیه عوامل تأثیرگذار بر متغیر وابسته برای محقق امکانپذیر نمی‌باشد. بنابراین متغیرهای تحلیل مسیر همواره می‌توانند تنها بخشی از واریانس متغیر وابسته را تبیین کنند. به همین دلیل در تحلیل مسیر آنچه که به عنوان اثر یا عوامل

ناشناخته باقی می‌ماند بوسیله «e» که به «کمیت خطا»<sup>(۱)</sup> معروف است، نشان داده می‌شود. مقدار e بیانگر میزان واریانس متغیری است که متغیرهای مستقل قبلی دیاگرام قادر به تبیین آن نبوده‌اند. با مجذور کردن e، واریانس تبیین نشده بدست می‌آید. چنانچه مقدار  $e^2$  از عدد ۱ کم شود مقدار واریانس تبیین شده (که آن را با  $R^2$  نشان می‌دهند) بدست می‌آید. بعبارت دیگر:

$$R^2 = 1 - e^2$$

حال در این مثال برای دستیابی به این موضوع که مدل علی ارائه شده در دیاگرام تحلیل مسیر چه میزان از واریانس متغیر وابسته را تبیین می‌کند، باید از ضریب تعیین ( $R^2$ ) کمک گرفت.  $R^2$  بدست آمده به روش Enter برای کلیه متغیرها در جدول (۲-۱۱) ارائه شده است. همانطوریکه از این جدول برمی‌آید مقدار ضریب تعیین<sup>(۲)</sup> برابر با ۰/۹۶۹ می‌باشد. یعنی ۹۶/۹ درصد از مجموع تغییرات متغیر وابسته توسط مدل تحلیلی فوق تبیین می‌شود. از طریق  $R^2$  می‌توان ضریب خطا یا مقدار e را نیز محاسبه کرد.

$$R^2 = 1 - e^2$$

$$۰/۹۶۹ = 1 - e^2$$

$$۱ - ۰/۹۶۹ = e^2 = ۰/۰۳۱$$

بنابراین می‌توان گفت که مدل علی بدست آمده تنها ۳/۱ درصد از واریانس متغیر وابسته را

تبیین نمی‌کند.

## فصل دوازدهم

# تحلیل واریانس چند متغیره

مقدمه :

در برخی از مطالعات اجتماعی - اقتصادی، ممکن است محقق با بیش از یک متغیر وابسته سروکار داشته باشد. در چنین مواقعی استفاده از تحلیل واریانس (ANOVA) برای مقایسه میانگین گروهها به تفکیک متغیرهای وابسته روش مناسبی نمی باشد. زیرا این روش در چنین مواقعی سطح خطا را افزایش می دهد. به همین دلیل روش تحلیل واریانس چند متغیره، تکنیک مناسبی برای اینگونه تجزیه و تحلیل ها به حساب می آید. این فصل به توضیح و تبیین این روش آماری اختصاص دارد.

تحلیل واریانس چند متغیره چیست؟

تحلیل واریانس چند متغیره<sup>(۱)</sup> (MANOVA) نیز مانند تحلیل واریانس (ANOVA) با بررسی تفاوت بین گروهها (یا تیمارهای تجربی) سروکار دارد. با این تفاوت که تحلیل واریانس یک



روش یک متغیره<sup>(۱)</sup> بوده و سعی می‌کند تا از این طریق به سنجش تفاوت گروهها براساس یک متغیر وابسته کمی بپردازد. در حالیکه تحلیل واریانس چند متغیره با بیش از یک متغیر وابسته سروکار داشته و سعی می‌کند تا به بررسی و سنجش تفاوت گروهها براساس چندین متغیر وابسته کمی بطور همزمان بپردازد. (بعبارت دیگر در این روش هر گروه تیمار براساس دو یا چند متغیر وابسته مورد بررسی قرار می‌گیرد).

از سوی دیگر، هر دو این روشها به منظور استنتاج معنی‌داری آماری تفاوت بین گروهها مورد استفاده قرار می‌گیرد. در تحلیل واریانس (ANOVA) فرضیه آماری<sup>(۲)</sup> برابر بودن میانگین بین گروههاست، در حالیکه چون در تحلیل واریانس چند متغیره با چندین متغیر وابسته مواجه هستیم بنابراین در اینجا بجای میانگین بین چندین گروه با بردار میانگین‌ها سر و کار داریم. به همین دلیل در این روش فرض آماری برابری بردارهای میانگین‌های چندین متغیر وابسته در بین گروهها مدنظر می‌باشد. تفاوت بین آزمایش فرضیات در ANOVA و MANOVA را می‌توان بصورت ذیل روشن کرد.

(۱) فرضیه آماری در ANOVA :

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

در اینجا فرض آماری این است که میانگین کلیه گروهها با هم برابر هستند و آنها همه از یک

جامعه گرفته شده‌اند.

(۲) فرضیه آماری در MANOVA:

$$H_0: \begin{vmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{vmatrix} = \begin{vmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{vmatrix} = \dots = \begin{vmatrix} \mu_{1k} \\ \mu_{2k} \\ \vdots \\ \mu_{pk} \end{vmatrix}$$

$\mu_{pk}$  = میانگین متغیر  $p$ ، گروه  $k$

در اینجا فرض آماری این است که کلیه بردارهای میانگین‌های گروهها با هم برابر بوده و از یک جامعه گرفته شده‌اند.

این دو روش بویژه در طرحهای عاملی تجربی که محقق از طریق دستکاری یک یا چند متغیر مستقل درصدد بررسی تأثیر آن بر روی یک متغیر وابسته (در روش ANOVA) و یا چند متغیر وابسته (در روش MANOVA) را مدنظر دارد بسیار مفید می‌باشند. این روشها این امکان را برای محقق فراهم می‌سازند تا قضاوت کند که آیا تغییرات مشاهده شده بر اثر تغییرات تیمار بوده و یا بر اثر شانس و تصادف در نمونه‌گیری بوجود آمده است. این روشها به شیوه‌های مختلف مورد استفاده قرار می‌گیرند که هر یک از آنها بطور مستقل در ذیل توضیح داده می‌شوند.

#### تفاوت بین دو گروه مستقل:

برای روشن شدن موضوع ابتدا به تحلیل تفاوت بین دو گروه مستقل می‌پردازیم. در گروههای مستقل هر یک از پاسخگویان بطور تصادفی تنها در یکی از گروههای (سطوح) مستقل قرار می‌گیرند. در تحلیل یک متغیره تنها یک متغیر وابسته کمی مورد بررسی قرار می‌گیرد و فرض صفر نیز برابر بودن میانگین گروهها می‌باشد. در تحلیل چند متغیره، حداقل دو متغیر وابسته کمی

مدنظر می‌باشد و فرض آماری نیز این است که بردارهای میانگین‌ها برای گروهها با هم برابر می‌باشند.

آزمون آماری مناسب برای تحلیل یک متغیر در دو گروه، آزمون  $t^{(۱)}$  و برای تحلیل چند متغیره، آماره  $T$  هتلینگ<sup>(۲)</sup> می‌باشد.

#### رهیافت یک متغیره: $t$ تست:

آزمون  $t$  به سنجش سطح معنی‌داری تفاوت بین میانگین‌های دو گروه مستقل می‌پردازد. به عنوان مثال ممکن است محقق دو گروه از افراد را در معرض دو پیام بازرگانی درخصوص یک معامله قرار دهد و به دنبال آن از افراد گروهها بخواهد تا میزان تصمیم به خرید خود را بر روی یک مقیاس ۱۰ نقطه‌ای مشخص کنند. برای مشخص کردن اینکه آیا دو پیام بازرگانی اثر متفاوتی بر تصمیم به خرید افراد گذاشته‌اند یا نه باید آزمون  $t$  محاسبه شود.

#### رهیافت چند متغیره: $T$ هتلینگ:

شاید غیرواقعی باشد که فرض شود تفاوت بین دو گروه (تیمار) تنها در یک متغیر وابسته اندازه‌گیری شده ظاهر شود. به عنوان مثال، دو پیام بازرگانی ممکن است تنها سطح تصمیم به خرید متفاوتی ایجاد نکند، بلکه ممکن است جنبه‌های دیگر (نظیر ارزیابی کلی از محصول، علاقه، انگیزش و غیره) را نیز تحت تأثیر قرار دهد. در چنین مواردی بسیاری از محققان هریک از این موارد را به عنوان یک متغیر وابسته بطور جداگانه و در قالب آزمون تک متغیره  $t$  مورد بررسی قرار می‌دهند. در این صورت ترکیب خطی این متغیرهای وابسته که ممکن است تفاوتیابی در بین گروهها ایجاد کنند مورد غفلت قرار می‌گیرند و آزمونهای مستقل، همبستگی بین متغیرهای وابسته را در نظر نگرفته و اطلاعات کمتری را در اندازه‌گیری تفاوت‌های کلی بین گروهها مورد استفاده قرار می‌دهد. به همین دلیل است که روش بررسی متغیرهای وابسته بطور جداگانه از ناکارآمدیهای قابل توجهی برخوردار

1-  $t$ - test2- Hotelling's  $T^2$

می‌باشد. زیرا در تحقیقات اجتماعی و اقتصادی سطح خطا ( $\alpha$ ) معمولاً  $0/05$  و  $0/01$  پذیرفته شده است و زمانیکه  $\alpha = 0/05$  باشد احتمال رد فرضیه صفر  $0/05$  و احتمال عدم رد آن  $0/95$  خواهد بود، که از طریق  $1 - \alpha = 0/95$  بدست می‌آید. اما چنانچه مقدار  $\alpha = 0/05$  باشد و در مدل تحلیلی دو متغیر وابسته وجود داشته باشد در این صورت مقدار  $\alpha$  به جای  $0/05$  در واقع برابر با  $0/0975 = (1 - 0/05)^2$  خواهد بود و اگر در مدل مورد نظر ۵ متغیر وابسته وجود داشته باشد میزان خطا برابر با  $0/23 = (1 - 0/05)^5$  خواهد بود<sup>(۱)</sup>. از سوی دیگر ممکن است هیچ یک از متغیرهایی که بصورت جداگانه مورد آزمون قرار گرفته‌اند معنی دار نباشند اما تحلیل یک جا و همزمان آنها با استفاده از تحلیل واریانس چند متغیره معنی دار گردد. زیرا ترکیب خطی متغیرها ممکن است اثرات و نتایجی را بدست آورند که بررسی جداگانه آنها فاقد آن باشند<sup>(۲)</sup>.

اما آماره T هتلینگ (در MANOVA) راه حل مناسبتری را برای این مسئله فراهم می‌کند. زیرا این آماره آزمون کلی را برای تشخیص تفاوت بین گروهها در رابطه با مجموعه متغیرهای وابسته بکار می‌گیرد و در این آزمون ترکیب خطی متغیرهای وابسته مورد استفاده قرار می‌گیرد. برای تحقق این هدف معادله‌ای بشرح ذیل برای ترکیب خطی متغیرهای وابسته به کار گرفته می‌شود.

$$C = W_1 Y_1 + W_2 Y_2 + \dots + W_n Y_n$$

که در این معادله:

C = مقدار ترکیبی پاسخها

W = وزن متغیرهای وابسته

۱- هومن، حیدرعلی (۱۳۸۰) «تحلیل داده‌های چند متغیری در پژوهش رفتاری» نشر پارسا، تهران، ص ۳۰۷.

2- Sharma Subhash (1996) "Applied Multivariate Techniques" John & sons, Inc. New York, PP.

$Y =$  متغیرهای وابسته

تحلیل واریانس چند متغیره براساس جدائی مراکز ثقل گروهها، به گونه‌ای وزنه‌ای مربوط به متغیرهای وابسته را اعمال می‌کند که واریانس بین گروهها را نسبت به واریانس درون گروهها به حداکثر برساند و این همان چیزی است که در تحلیل تشخیصی نیز انجام می‌گیرد.

### تفاوت بین K گروه مستقل :

طرح چند گروه مستقل، در واقع بسط یافته طرح دو گروهی است. بطور عمومی در طرح چند گروه مستقل، هر پاسخگو، بطور تصادفی در یکی از گروههای چندگانه قرار می‌گیرد. در مورد تحلیل تک متغیره تنها یک متغیر وابسته وجود داشت و فرض آماری نیز این بود که میانگین تمام گروهها با هم برابر هستند ( $\mu_1 = \mu_2 = \dots = \mu_k$ ). در مورد تحلیل چند متغیره، چندین متغیر وابسته کمی وجود دارد و فرض آماری نیز این است که تمام بردارهای مربوط به مقادیر میانگینها با هم برابرند. برای تحلیل تک متغیره، آزمون مناسب آماره  $F^{(1)}$  بدست آمده از ANOVA می‌باشد و آزمون مناسب برای تحلیل چند متغیره، اثر پیلای<sup>(۲)</sup>، لامبدای ویلکز<sup>(۳)</sup> (که به معیار نسبت حداکثر درستنمایی ویلکز<sup>(۴)</sup> یا آماره  $U^{(5)}$  معروف است)، اثر هتلینگ<sup>(۶)</sup> و معیار بزرگترین ریشه رُی<sup>(۷)</sup> بدست آمده از MANOVA می‌باشد.

### رهیافت یک متغیره ANOVA:

همانطوریکه در رهیافت یک متغیره، گفته شد ممکن است محقق دو گروه از افراد را در برابر

1- F- Statistics

2- Pillai's Trace

3- Wilk's lambda

4- Wilks' Likelihood ratio criterion

5- U- Statistic

6- Hotelling's trace

7- Roy's largest root

دو پیام بازرگانی در مورد معرفی کالا قرار دهد و سپس از آنها بخواهد تا تصمیم خود را نسبت به خرید کالای معرفی شده در قالب یک مقیاس ۱۰ نقطه‌ای تعیین کنند.

حالا فرض کنید که بجای دو پیام بازرگانی، در صد ارزیابی سه پیام بازرگانی هستیم.  $(K=3)$ . در این صورت پاسخگویان بطور تصادفی در سه گروه قرار گرفته و برای هر گروه یک میانگین محاسبه می‌شود. برای تجزیه و تحلیل تفاوت‌های موجود بین گروهها، یک نفر ممکن است به مقایسه دودو گروهها از طریق آزمون  $t$  پردازد و بدین ترتیب میانگین‌های جفت‌های  $X_1$  در مقابل  $X_2$ ،  $X_1$  در مقابل  $X_3$  و  $X_2$  در مقابل  $X_3$  را با هم مقایسه کند.

این نوع مقایسه نتایج درستی ارائه نمی‌دهد. برای مقایسه آنها باید از تحلیل واریانس (ANOVA) استفاده شود. در این صورت، تحلیل واریانس ضمن بررسی واریانس درون گروهها و بین گروهها از طریق محاسبه نسبت واریانس بین گروهها به واریانس درون گروهها به بررسی تفاوت‌های موجود می‌پردازد. این آزمون نشان می‌دهد که آیا تفاوتی بین میانگین‌های گروههای سه گانه یا بیشتر وجود دارد یا خیر. اما مشخص نمی‌کند که این تفاوتها در کجا قرار دارند. به همین دلیل برای دستیابی به اینگونه تفاوتها باید از آزمونهای پس تجربی<sup>(۱)</sup> استفاده کرد، که از جمله مهمترین این آزمونها می‌توان به آزمون دانکن<sup>(۲)</sup>، توکی<sup>(۳)</sup> و شفه<sup>(۴)</sup> اشاره کرد.

#### رهیافت چند متغیره MANOVA:

در برخی مواقع اتفاق می‌افتد که محققان بجای یک متغیر وابسته، با چندین متغیر وابسته مواجه می‌شوند. در چنین مواقعی برای آزمون تفاوت بین میانگین گروهها از روش  $F$  تست استفاده می‌کنند، بطوریکه با تکرار آزمون، در هر مرحله تنها یک متغیر وابسته را مورد بررسی قرار می‌دهند و این تکرار تا زمانی ادامه می‌یابد که کلیه متغیرهای وابسته بطور جداگانه مورد آزمون قرار گیرند.

1-Post Hoc

2-Duncan

3- Tukey

4-Scheffe

اشکال وارد بر روش t تست در خصوص مقایسه دوی دو میانگین گروهها در مواقعی که بیش از دو گروه وجود دارد، در این شیوه نیز قابل مشاهده است. زمانیکه بیش از یک متغیر وابسته وجود داشته باشد ممکن است همبستگی و ترکیب خطی متغیرهای فوق روابط و نتایج جدیدی را بوجود آورند که در صورت بررسی جداگانه آنها این نتایج مورد غفلت قرار می‌گیرد. به همین دلیل در چنین مواقعی استفاده از تحلیل واریانس یک متغیره (ANOVA) راه حل مناسبی بنظر نمی‌رسد. برای حل این مشکل باید از تحلیل واریانس چند متغیره (MANOVA) استفاده شود. تحلیل واریانس چند متغیره (MANOVA) در واقع حالت بسط یافته روش هتلینگ  $T^2$  می‌باشد. در این فرآیند می‌توان تمهیداتی در مورد تعیین وزن برای متغیر وابسته بدست آورد و مقدار ترکیب خطی خاصی برای هر پاسخگو محاسبه کرد. در واقع در تحلیل چند متغیره تلاش می‌شود تا مجموعه‌ای از وزنها برای حداکثر کردن مقدار F در مقدار ترکیبی، برای کل گروهها بدست آید. مجموعه‌ای از وزنها که مقدار F را به حداکثر می‌رساند اولین تابع تشخیصی<sup>(۱)</sup> نامیده می‌شود.

برای آزمون معنی‌داری تحلیل چند متغیره، برنامه‌های رایانه‌ای آزمون‌های متنوعی را محاسبه می‌کنند که مهمترین آنها لامبدای ویلکز<sup>(۲)</sup>، معیار پیلای<sup>(۳)</sup>، و اثر هتلینگ<sup>(۴)</sup> و معیار بزرگترین ریشه ری<sup>(۵)</sup> می‌باشد.

#### آزمون Post Hoc :

پس از اینکه، این نتیجه حاصل شد که بردارهای میانگین گروهها با هم برابر نیستند ممکن است محقق درصدد باشد تا مقایسه‌های دیگری را نیز در بین گروهها انجام دهد. به عنوان مثال ممکن است فردی بخواهد بداند که آیا تفاوتی بین گروهها در یک متغیر وابسته خاص یا متغیر

1- First Discriminant Function

2- Wilk's Lambda

3- Pillai's Criterion

4- Hotelling's Trace

5- Roy's Largest root

ترکیبی وجود دارد یا خیر؟ برای این کار آماره مقایسه مقید شفه بین گروهها<sup>(۱)</sup>، دانکن و توکی براساس هریک از متغیرهای وابسته از جمله مقایسه‌هایی است که انجام می‌گیرد.

بجز این تکنیک‌ها، روشهای دیگری نیز وجود دارند که یکی از آنها تحلیل گام به پس<sup>(۲)</sup> می‌باشد. این روش با حذف اثرات سایر متغیرهای وابسته، آماره F یک متغیره<sup>(۳)</sup> برای یک متغیر وابسته محاسبه می‌کند. فرآیند محاسبه آن شباهت زیادی به روش گام به گام در رگرسیون دارد. اما در این جا، این موضوع بررسی می‌شود که آیا یک متغیر وابسته خاص، بطور یکسان در تفاوت بین گروهها نقش ایفا می‌کند و این ارتباط با سایر متغیرها ناهمبسته<sup>(۴)</sup> است یا خیر. این فرآیند از طریق مقایسه مقید هلمرت<sup>(۵)</sup> نیز امکانپذیر است. روش دیگر در این زمینه توابع تشخیصی<sup>(۶)</sup>، بویژه اولین تابع تشخیصی<sup>(۷)</sup> است. این تکنیک مشخص می‌کند که کدامیک از متغیرها بهتر تفاوت بین گروهها را نشان می‌دهد.

#### مثالی از رهیافت چند متغیره:

فرض کنید مؤسسه‌ای درصدد است تا اثر بخشی دو دوره آموزشی سه روزه و پنج روزه‌ای که برای کارکنان بخش فروش برگزار شده است را مورد ارزیابی قرار دهد. هدف مؤسسه این است که بداند اولاً آیا دوره‌های برگزار شده بر میزان شناخت و آگاهی فروشندگان از اصول فروشندگی تاثیر داشته است و ثانیاً آیا میزان انگیزش افراد را برای دستیابی به اهداف فروش افزایش داده است یا خیر. همانطوریکه از این مثال برمی‌آید در اینجا دو متغیر وابسته (۱) افزایش شناخت و آگاهی و (۲).

- |   |                           |
|---|---------------------------|
| 1- Scheffe- type contrasts between groups | 2- Step- down analysis    |
| 3- Univariate F- statistic                | 4- Uncorrelated           |
| 5- Helmert contrasts                      | 6- Discriminant Functions |
| 7- First Discriminant Function            |                           |



جدول (۱۲-۱) نمرات اخذ شده توسط گروه‌های سه‌گانه در خصوص متغیرهای میزان آگاهی و

شناخت از اصول فروشندگی و انگیزش

تیمار ۱: شرکت کننده در دوره ۵ روزه			تیمار ۲: گروه کنترل			تیمار ۳: شرکت کننده در دوره ۳ روزه		
کد فروشنده	نمره شناخت (X <sub>۱</sub> )	نمره انگیزش (X <sub>۲</sub> )	کد فروشنده	نمره شناخت (X <sub>۱</sub> )	نمره انگیزش (X <sub>۲</sub> )	کد فروشنده	نمره شناخت (X <sub>۱</sub> )	نمره انگیزش (X <sub>۲</sub> )
۱	۱۳/۵۹	۱۰/۳۶	۱۷	۱۲/۹۵	۷/۴۷	۳۳	۱۱/۱۴	۷/۹۹
۲	۱۵/۲۲	۹/۸۳	۱۸	۱۵/۷۲	۱۰/۷۲	۳۴	۱۳/۲۹	۹/۹۹
۳	۱۴/۱۳	۱۱/۱۹	۱۹	۱۳/۹۱	۸/۸۲	۳۵	۱۶/۳۴	۱۲/۱۲
۴	۱۴/۷۷	۹/۸۶	۲۰	۱۲/۱۱	۷/۳۲	۳۶	۱۳/۱۲	۸/۹۵
۵	۱۶/۶۱	۱۳/۵۰	۲۱	۱۴/۶۵	۱۰/۳۷	۳۷	۱۵/۸۸	۹/۶۰
۶	۱۵/۵۵	۱۱/۷۹	۲۲	۱۳/۵۸	۶/۶۵	۳۸	۱۶/۹۲	۱۲/۱۲
۷	۱۲/۵۳	۹/۹۹	۲۳	۱۲/۶۸	۹/۴۳	۳۹	۱۵/۶۲	۷/۵۰
۸	۱۷/۹۱	۱۱/۹۱	۲۴	۱۳/۸۱	۸/۵۳	۴۰	۱۲/۷۲	۷/۳۲
۹	۱۴/۴۲	۹/۹۵	۲۵	۱۰/۸۶	۷/۸۱	۴۱	۱۷/۰۷	۱۰/۱۹
۱۰	۱۶/۸۷	۱۲/۰۳	۲۶	۱۳/۱۵	۸/۲۴	۴۲	۱۵/۸۶	۱۱/۸۴
۱۱	۱۵/۷۹	۱۱/۶۸	۲۷	۱۳/۰۳	۶/۸۲	۴۳	۱۴/۴۰	۱۰/۵۰
۱۲	۱۲/۷۷	۱۰/۹۶	۲۸	۱۱/۹۵	۸/۱۱	۴۴	۱۱/۳۶	۸/۵۰
۱۳	۱۵/۵۵	۹/۹۱	۲۹	۱۱/۳۶	۷/۷۵	۴۵	۱۵/۲۶	۱۱/۰۱
۱۴	۱۴/۱۵	۱۰/۱۶	۳۰	۱۲/۵۲	۷/۴۳	۴۶	۱۲/۴۹	۸/۸۵
۱۵	۱۳/۸۳	۹/۶۵	۳۱	۱۰/۵۸	۶/۶۱	۴۷	۱۴/۴۹	۸/۴۷
۱۶	۱۳/۲۱	۸/۱۷	۳۲	۱۲/۶۲	۸/۸۴	۴۸	۱۵/۵۶	۱۰/۰۱

افزایش انگیزش افراد مد نظر می‌باشد. برای این کار سه گروه ۱۶ نفره از کارکنان را انتخاب کرده و گروه اول در یک دوره آموزشی پنج روزه، گروه دوم به عنوان گروه کنترل و گروه سوم در یک دوره سه روزه شرکت کرده و سطح اطلاعات فروشندگی و میزان انگیزش آنها اندازه‌گیری شده است. نتایج این آزمون در جدول (۱-۱۲) ارائه شده است.

فرض صفر این است که بردارهای میانگین گروههای سه گانه با هم برابر هستند. با انجام تحلیل واریانس چند متغیره و استفاده از لامبدای ویلکز و سایر آماره‌ها می‌توان صحت و سقم این فرضیه را بررسی کرد. برای تحلیل‌های بیشتر می‌توان از آزمون‌های Post Hoc، نظیر مقایسه مقید شف‌ه<sup>(۱)</sup>، تحلیل گام به پس<sup>(۲)</sup> یا مقایسه مقید هلمرت<sup>(۳)</sup> و توابع تشخیصی<sup>(۴)</sup> و غیره نیز استفاده کرد.

#### روش محاسبه MANOVA با برنامه SPSS:

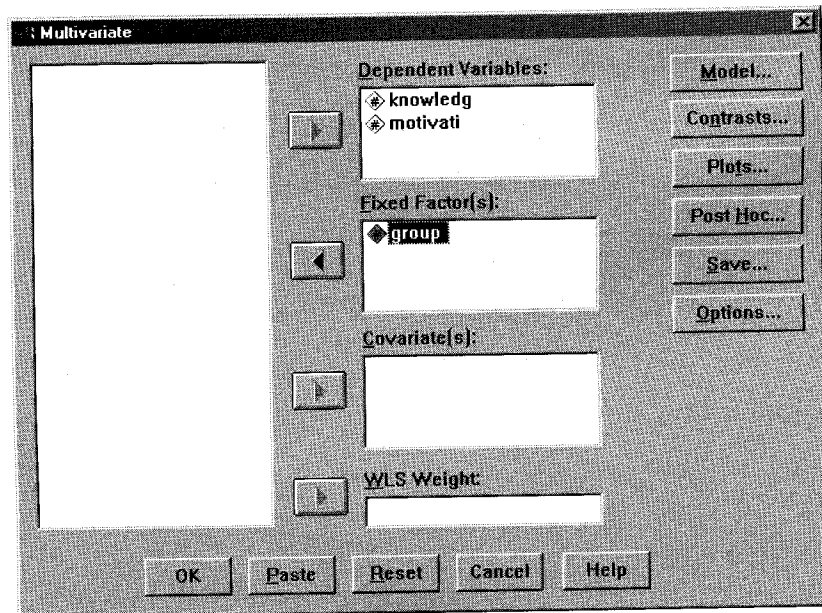
برای محاسبه تحلیل واریانس چند متغیره (MANOVA) با نرم‌افزار SPSS، برای گروههای مستقل، ابتدا مقادیر هر یک از متغیرهای وابسته را بطور جداگانه در یک ستون در پنجره Data Editor وارد کنید. به عنوان نمونه در این مثال، نمرات مربوط به میزان آشنایی با اصول فروشندگی را برای هر سه گروه در زیر هم در یک ستون وارد کرده و متغیر فوق را تحت عنوان knowledge تعریف کنید. سپس نمرات مربوط به انگیزش را برای هر سه گروه در ستون دوم در زیر یکدیگر وارد کرده و آنرا تحت عنوان Motivation تعریف نمایید. برای تشخیص سه گروه از یکدیگر در ستون سوم گروههای مورد نظر (تحت عنوان Group) را با استفاده از کدهای ۱، ۲ و ۳ تعریف نمایید. (به عنوان مثال گروهی که در دوره آموزشی ۵ روزه شرکت کرده است را با کد ۱، گروه کنترل را با کد ۲ و گروهی که در دوره آموزشی سه روزه شرکت کرده را با کد ۳ تعریف کنید). پس از وارد کردن داده‌ها دستور: Analyze/General Linear Model/ Multivariate را اجرا کنید تا شکل (۱-۱۲) ظاهر شود.

1- Scheffe's Contrast

2- Step- Down Analysis

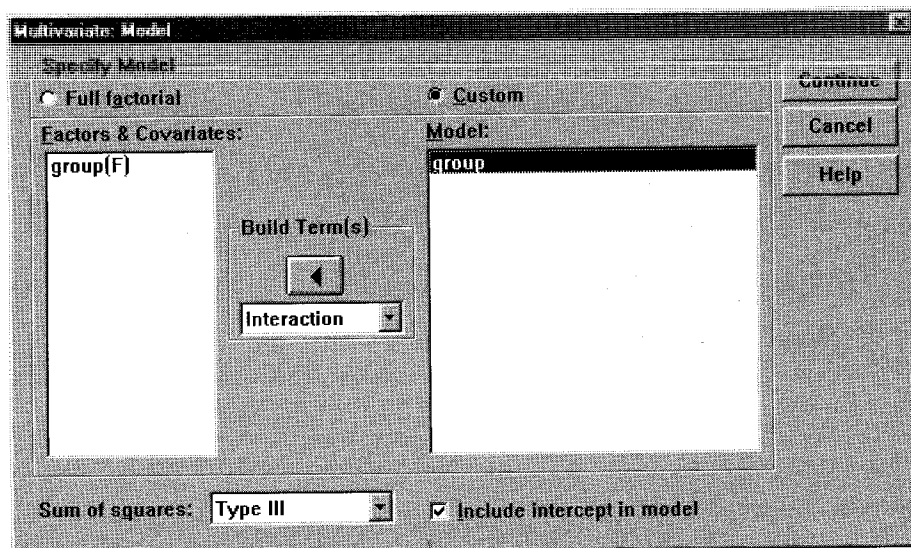
3- Helmert contrast

4- Discriminant Functions



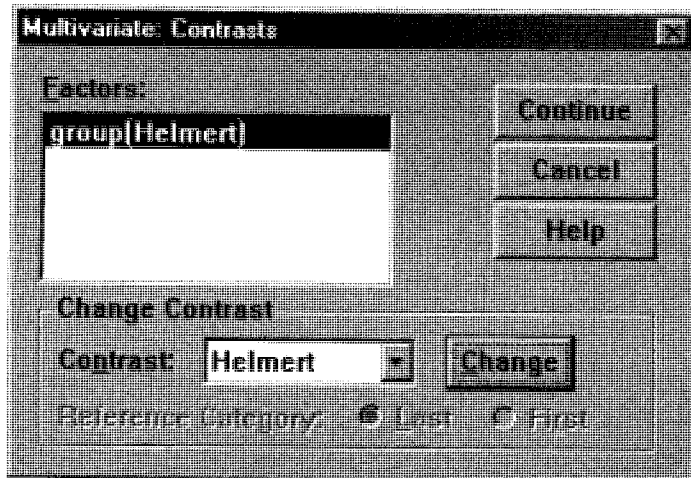
شکل (۱-۱۲) پنجره Multivariate

در این پنجره متغیرهای Knowledge و Motivation را به قسمت Dependent variables و متغیر Group را به قسمت Fixed Factor (S) وارد کنید. سپس دکمه Model... را از پنجره اصلی فشار دهید تا پنجره Multivariate model بشرح ذیل باز شود.



شکل (۲-۱۲) پنجره Multivariate Model

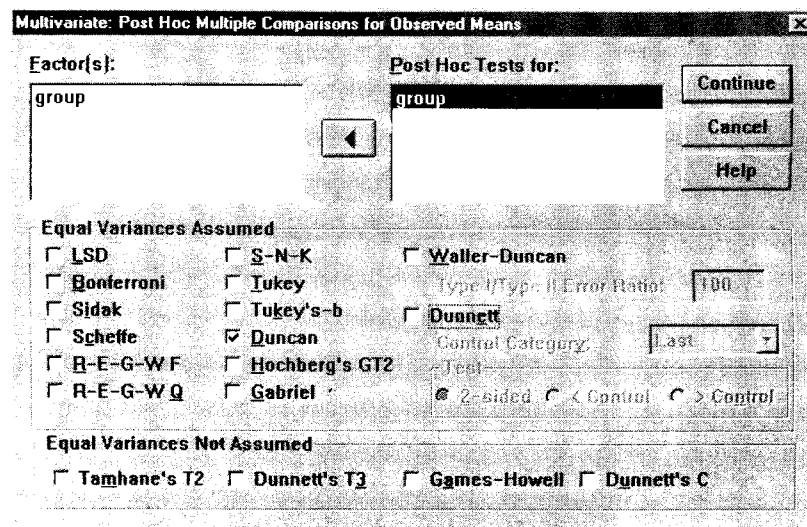
در این پنجره دکمه رادیویی Custom را روشن کرده و متغیر (F) group را به قسمت Model در سمت راست منتقل کنید و دکمه Continue را فشار دهید پس از آن دکمه Contrasts... را از پنجره اصلی فشار دهید تا شکل (۳-۱۲) ظاهر شود.



شکل (۳-۱۲) پنجره Multivariate: Contrasts

در این پنجره فلش مقابل Contrast را فشار دهید و از لیست کرکره‌ای ظاهر شده مورد Helmert را انتخاب نموده، سپس دکمه Change را که در مقابل آن قرار دارد فشار دهید تا آماره Helmert به جلو Group در پنجره Factors منتقل شود، سپس دکمه Continue را فشار دهید تا به پنجره اصلی باز گردید.

پس از آن دکمه Post Hoc... را از پنجره اصلی فشار دهید تا شکل (۴-۱۲) ظاهر شود.



شکل (۴-۱۲) پنجره Post Hoc Multiple Comparisons for ...

در این پنجره متغیر Group را از قسمت Factor(s) در سمت چپ به قسمت Post Hoc Tests for: در سمت راست منتقل کنید. در پایین این پنجره دو مجموعه آزمون وجود دارد. یک مجموعه از این آزمونها زمانی مورد استفاده قرار می‌گیرند که واریانس گروهها با هم برابر باشند، که این آزمونها در قسمتی تحت عنوان Equal variances Assumed قرار دارند (شکل ۴-۱۲) و مجموعه دیگر از آزمونها نیز در قسمتی تحت عنوان Equal variances Not Assumed قرار گرفته‌اند. برای اطلاع از اینکه آیا واریانسهای گروهها با هم برابر هستند یا نه می‌توان از آزمون لون<sup>(۱)</sup> استفاده کرد<sup>(۲)</sup>. در این قسمت در صورتی که واریانسهای گروهها با هم برابر فرض شوند از آزمونهایی مانند دانکن<sup>(۳)</sup>،

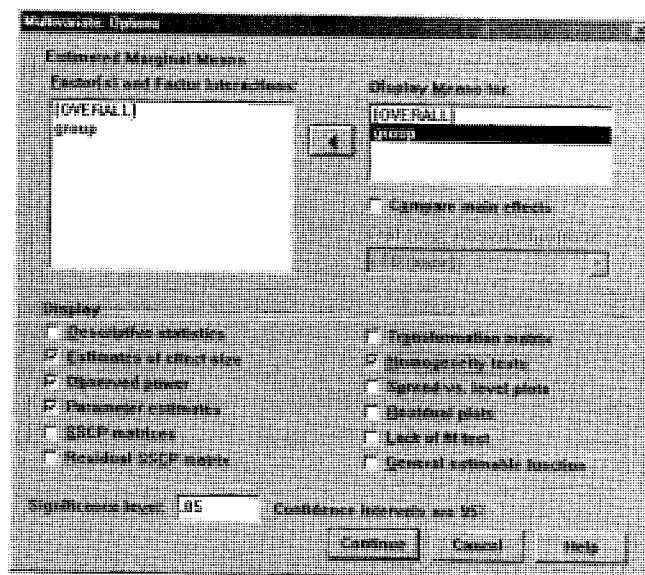
1- Leven Test

۲- برای آشنایی بیشتر با تست لون به فصل پنجم مربوط به آزمون F یا ANOVA مراجعه شود.

3- Duncan

شفه<sup>(۱)</sup>، توکی<sup>(۲)</sup> و غیره می‌توان استفاده کرد. برای انتخاب این آزمونها کافی است که در مربع‌های مقابل این آزمون‌ها کلیک کنید. پس از انتخاب آزمونهای مورد نظر دکمه Continue را فشار دهید تا به پنجره اصلی باز گردید.

از پنجره اصلی دکمه Options... را فشار دهید تا شکل (۱۲-۵) ظاهر شود. در این پنجره متغیر (OVERALL) و group را از قسمت Factor(s) در قسمت چپ به قسمت Display means for منتقل کنید. در این پنجره آماره‌های زیادی وجود دارند که بر حسب نیاز می‌توان به محاسبه آنها اقدام کرد. از آماره‌هایی که در قسمت Display قرار دارند موارد Estimates of effect size و observed power و Homogeneity tests را انتخاب کرده و دکمه Continue را فشار دهید تا به پنجره اصلی باز گردید. در این مرحله با کلیک کردن دکمه OK محاسبات در پنجره Output ظاهر می‌شود (جدول ۱۲-۲).



شکل (۱۲-۵) پنجره Options :Multivariate

1- Scheffe

2- Tukey

جدول (۲-۱۲) نتایج محاسبات تحلیل واریانس چند متغیره

**General Linear Model**

**Between-Subjects Factors**

GROUP	Value Label	N
1	5 days course	16
2	not attended	16
3	3 days course	16

**Box's Test of Equality of Covariance Matrices<sup>a</sup>**

Box's M	7.162
F	1.117
df1	6
df2	50469.231
Sig.	.049

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+GROUP

**Multivariate Tests<sup>d</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.988	1862.391 <sup>b</sup>	2.000	44.000	.000
	Wilks' Lambda	.012	1862.391 <sup>b</sup>	2.000	44.000	.000
	Hotelling's Trace	84.654	1862.391 <sup>b</sup>	2.000	44.000	.000
	Roy's Largest Root	84.654	1862.391 <sup>b</sup>	2.000	44.000	.000
GROUP	Pillai's Trace	.424	6.059	4.000	90.000	.000
	Wilks' Lambda	.591	6.618 <sup>b</sup>	4.000	88.000	.000
	Hotelling's Trace	.666	7.162	4.000	86.000	.000
	Roy's Largest Root	.625	14.060 <sup>c</sup>	2.000	45.000	.000

## Tests of Between-Subjects Effects

Source	Dependent Variable	Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	KNOWLEDG	.231	13.499	.899
	MOTIVATI	.382	27.874	.997
Intercept	KNOWLEDG	.988	3677.415	1.000
	MOTIVATI	.978	2031.302	1.000
GROUP	KNOWLEDG	.231	13.499	.899
	MOTIVATI	.382	27.874	.997
Error	KNOWLEDG			
	MOTIVATI			
Total	KNOWLEDG			
	MOTIVATI			
Corrected Total	KNOWLEDG			
	MOTIVATI			

- a. Computed using alpha = .05  
b. R Squared = .231 (Adjusted R Squared = .197)  
c. R Squared = .382 (Adjusted R Squared = .355)

## Custom Hypothesis Tests

## Contrast Results (K Matrix)

GROUP Helmert Contrast		Dependent Variable		
		KNOWLEDG	MOTIVATI	
Level 1 vs. Later	Contrast Estimate	1.125	1.937	
	Hypothesized Value	0	0	
	Difference (Estimate - Hypothesized)	1.125	1.937	
	Std. Error	.491	.442	
	Sig.	.027	.000	
	95% Confidence Interval for Difference	Lower Bound	.136	1.047
		Upper Bound	2.114	2.828
Level 2 vs. Level 3	Contrast Estimate	-1.627	-1.503	
	Hypothesized Value	0	0	
	Difference (Estimate - Hypothesized)	-1.627	-1.503	
	Std. Error	.567	.510	
	Sig.	.006	.005	
	95% Confidence Interval for Difference	Lower Bound	-2.769	-2.531
		Upper Bound	-.486	-.474

## Multivariate Test Results

	Value	F	Hypothesis df	Error df	Sig.	Eta Squared
Pillai's trace	.424	6.059	4.000	90.000	.000	.212
Wilks' lambda	.591	6.618 <sup>b</sup>	4.000	88.000	.000	.231
Hotelling's trace	.666	7.162	4.000	86.000	.000	.250
Roy's largest root	.625	14.060 <sup>c</sup>	2.000	45.000	.000	.385



**Multivariate Test Results**

	Noncent. Parameter	Observed Power <sup>a</sup>
Pillai's trace	24.234	.982
Wilks' lambda	26.471	.990
Hotelling's trace	28.649	.994
Roy's largest root	28.121	.998

- a. Computed using alpha = .05
- b. Exact statistic
- c. The statistic is an upper bound on F that yields a lower bound on the significance level.

**Univariate Test Results**

Source	Dependent Variable	Sum of Squares	df	Mean Square	F	Sig.
Contrast	KNOWLEDG	34.690	2	17.345	6.750	.003
	MOTIVATI	58.102	2	29.051	13.937	.000
Error	KNOWLEDG	115.639	45	2.570		
	MOTIVATI	93.800	45	2.084		

**Univariate Test Results**

Source	Dependent Variable	Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Contrast	KNOWLEDG	.231	13.499	.899
	MOTIVATI	.382	27.874	.997
Error	KNOWLEDG			
	MOTIVATI			

- a. Computed using alpha = .05

**Estimated Marginal Means**

**1. Grand Mean**

Dependent Variable	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
KNOWLEDG	14.031	.231	13.565	14.497
MOTIVATI	9.392	.208	8.972	9.812

## 2. GROUP

Dependent Variable	GROUP	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
KNOWLEDG	5 days course	14.781	.401	13.974	15.588
	not attended	12.843	.401	12.035	13.650
	3 days course	14.470	.401	13.663	15.277
MOTIVATI	5 days course	10.684	.361	9.957	11.411
	not attended	7.995	.361	7.268	8.722
	3 days course	9.498	.361	8.771	10.224

## Post Hoc Tests

## GROUP

## Homogeneous Subsets

## KNOWLEDG

Duncan<sup>a,b</sup>

GROUP	N	Subset	
		1	2
not attended	16	12.8425	
3 days course	16		14.4700
5 days course	16		14.7813
Sig.		1.000	.586

Means for groups in homogeneous subsets are displayed.  
Based on Type III Sum of Squares  
The error term is Mean Square(Error) = 2.570.

a. Uses Harmonic Mean Sample Size = 16.000.

b. Alpha = .05.

## MOTIVATION

Duncan<sup>a,b</sup>

GROUP	N	Subset		
		1	2	3
not attended	16	7.9950		
3 days course	16		9.4975	
5 days course	16			10.6838
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.  
Based on Type III Sum of Squares  
The error term is Mean Square(Error) = 2.084.

a. Uses Harmonic Mean Sample Size = 16.000.

b. Alpha = .05.

بر اساس دستوراتی که در تحلیل واریانس چند متغیره در برنامه SPSS صادر می‌شود، آماره‌های مختلفی محاسبه و در قالب جداول متفاوت در Output ظاهر می‌گردد. در اولین جدول معمولاً تعداد افراد هر یک از گروهها همراه با بر چسب<sup>(۱)</sup> آنها مشخص و در جدولی تحت عنوان Between-subject Factors ارائه می‌گردد. از آنجا که یکی از مفروضات MANOVA برابر بودن ماتریس‌های کواریانس برای گروههای مورد مطالعه می‌باشد (و این پیش فرضی است که در تحلیل تشخیصی نیز مد نظر است) بنابراین تست Box's M به همین منظور در جدول جداگانه‌ای محاسبه می‌گردد. اگر نتیجه این آزمون در سطح ۰/۰۵ یا ۰/۰۱ معنی‌دار باشد، نشان می‌دهد که ماتریس‌های کواریانس متغیرهای وابسته برای گروهها متفاوت است که در این صورت فرض صفر مبنی بر اینکه ماتریس کواریانس متغیرهای وابسته برای گروهها با هم برابر هستند رد می‌شود. بنابراین امکان انجام تحلیل واریانس چند متغیره وجود دارد.

آزمونهای معنی‌داری نظیر معیار پیلا<sup>(۲)</sup>، لامبدای ویلکس<sup>(۳)</sup>، معیار هتلینگ<sup>(۴)</sup> و بزرگترین ریشه ری<sup>(۵)</sup> که در جدول جداگانه تحت عنوان Multivariate Tests ظاهر می‌شوند، نشان می‌دهند که اثرات متغیرهای مورد نظر از لحاظ آماری معنی‌دار بوده است. به عبارت دیگر بردارهای میانگین‌های گسروهها با هم تفاوت معنی‌داری دارند. آماره‌های Noncentrality parameter و Observed power نیز که در ادامه همین جدول محاسبه می‌گردند، شدت رد فرض آماری را نشان می‌دهند. با افزایش مقدار آماره Noncentrality Parameter احتمال رد فرض آماری نیز افزایش می‌یابد. از سوی دیگر توان آماری<sup>(۶)</sup> نیز شانس درست رد کردن یک فرضیه نادرست را نشان می‌دهد. بالا بودن مقدار توان آزمونهای چند متغیره بدین معنی است که احتمال رد فرض صفر بسیار بالاست و چنانچه فرض آماری صحیح باشد

1- Lable

2- Pillai's Trace

3- Wilks' lambda

4- Hotelling Trace

5- Roy's Largest Root

6- Power of statistical test

مقدار آماره فوق نیز صفر خواهد بود.

نتایج بدست آمده از آزمونهای اثرات بین گروهها نیز نشان می دهند که آیا گروهها از همدیگر تفاوت معنی داری دارند یا خیر و آیا متغیرهای وارد شده اثر بخش بوده اند یا نه. اما مسئله اساسی این است که این تفاوتها در کجاست. برای تعیین اینگونه تفاوتها می توان از آماره مقایسه مقید هلمرت<sup>(۱)</sup> استفاده کرد. در این آزمون، اولین مقایسه مقید بین میانگین اولین گروه یا اولین سطح با متوسط میانگینهای بقیه گروهها یا سطوح انجام می گیرد و سطح معنی داری آن تعیین می گردد. دومین مقایسه مقید نیز بین میانگین دومین گروه یا دومین سطح با متوسط میانگینهای بقیه گروهها یا سطوح انجام می گیرد و این روند برای بقیه گروهها یا سطوح به همین ترتیب ادامه می یابد.

نتایج مربوط به آزمونهای تک متغیره نیز که متغیرهای وابسته را بطور جداگانه مورد بررسی قرار می دهد در جدولی تحت عنوان Univariate Test Results ارائه می گردد. در صورتی که این نتایج در سطح ۰/۰۵ یا ۰/۰۱ معنی دار باشد می توان قضاوت کرد که تفاوت معنی داری بین گروهها در سطح ۹۵ درصد یا ۹۹ درصد وجود دارد.

چنانچه دستور محاسبه میانگین کل و میانگین تک تک گروهها نیز صادر شده باشد، اینگونه محاسبات در جداول جداگانه ای تحت عنوان Grand Mean و Group ظاهر می شوند. حتی می توان مقایسه های دو به دو<sup>(۲)</sup> گروهها را نیز انجام داد که در این صورت نتیجه در جدول مستقلی ارائه می گردد. اما آزمونهای پس تجربی<sup>(۳)</sup> که از اهمیت زیادی نیز برخوردار هستند در جداول جداگانه ای محاسبه می گردند. از این گونه آزمونها می توان به آزمونهای دانکن، شفه، توکی و غیره اشاره کرد، که در این مثال از تست دانکن استفاده شده است. این آزمونها زیر مجموعه های همگن را مشخص می کنند. بطوریکه در مورد این مثال، گروه کنترل در یک طبقه و گروههایی که در دوره سه روزه و یا پنج روزه شرکت کرده اند. از نظر سطح شناخت (یعنی متغیر Knowledge) در یک طبقه قرار

1- Helmert Contrasts

2- Pairwise Comparisons

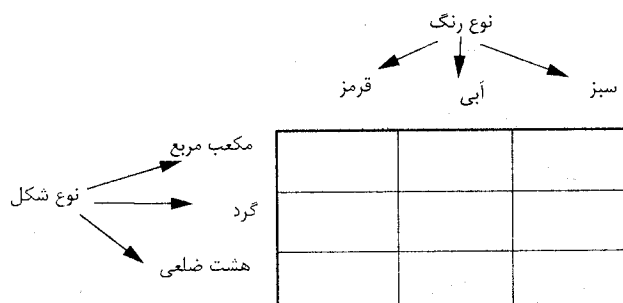
3- Post Hoc Tests

گرفته‌اند و تفاوت معنی‌داری از این حیث بین افراد شرکت کننده در دوره سه روزه و پنج روزه وجود ندارد، در حالیکه هر دو این گروهها تفاوت معنی‌داری با گروه کنترل (افرادى که در هیچ دوره‌ای شرکت نکرده‌اند) وجود دارد. به همین دلیل این افراد در طبقه جداگانه قرار گرفته‌اند. اما از نظر متغیر ایجاد انگیزش (Motivation) هر سه گروه از هم تفاوت معنی‌داری دارند و به همین دلیل آزمون دانکن، آنها را در زیر مجموعه‌های مختلف جای داده است.

### طرح‌های عاملی :

در تحلیل واریانس (ANOVA)، متغیرهای مستقل، عامل<sup>(۱)</sup> نامیده می‌شوند. یک عامل ممکن است دو یا چند سطح داشته باشد. یک طرح با بیش از دو عامل، طرح عاملی<sup>(۲)</sup> نام دارد. طرح عاملی روشی است که اثرات ساده و تعاملی دو یا چند متغیر مستقل را بر روی یک متغیر وابسته مورد بررسی قرار می‌دهد.

به عنوان مثال، فرض کنید که تولید کننده یک کالا درصدد است تا اثر سه رنگ مختلف (قرمز، آبی و سبز) و سه شکل متفاوت (مکعب مربع، گرد و هشت ضلعی) را بر ارزیابی مشتری از کالای فوق و تصمیم به خرید آن مورد بررسی قرار دهد. می‌توان اثر هر دو متغیر مستقل را بطور همزمان از طریق یک طرح عاملی  $3 \times 3$  بشرح ذیل مورد بررسی قرار داد.



جدول (۱۲-۳) تصویر شماتیک طرح‌های عاملی

1- Factor

2- Factorial design

همانطوریکه از جدول فوق برمی آید طرحهای عاملی شباهت زیادی به جداول متقاطع دو بعدی دارند، اما تفاوت آنها در این است که در خانه‌های جداول متقاطع، فراوانی یا درصد افراد نوشته می‌شود، در حالیکه در جدول طرحهای عاملی مقادیر متغیر وابسته نظیر نمره یا امتیاز ارزیابی مشتری از کالای فوق ثبت می‌گردد. (به عنوان مثال از پاسخگویان خواسته می‌شود تا ارزیابی خود را در قالب یک مقیاس ۱۰ نقطه‌ای بیان کنند). در این طرح، از طریق آزمون تحلیل واریانس (ANOVA) سه اثر مختلف مورد ارزیابی قرار می‌گیرد.

۱- اثر رنگ: آیا تفاوتی بین میانگین پاسخهای داده شده در خصوص رنگهای قرمز، آبی و سبز وجود دارد؟

۲- اثر شکل: آیا تفاوتی بین میانگین پاسخهای ارائه شده در مورد اشکال مختلف مکعب مربع، گرد و هشت ضلعی وجود دارد؟

۳- اثر متقابل رنگ و شکل: با نگاهی به تفاوت بین رنگها، آیا این تفاوتها در اشکال مختلف (مکعب مربع، گرد و هشت ضلعی) نیز به همان صورت همسان باقی مانده است. به عنوان مثال اگر رنگ قرمز بطور کلی به عنوان بهترین رنگ برای کالا ارزیابی شده باشد، اما این رنگ در کالایی که بصورت گرد تولید شده (در مقایسه با اشکال مکعب مربع و هشت ضلعی) در سطح پایین ارزیابی شود، این خود دلیلی بر اثر متقابل بین رنگ و شکل کالا خواهد بود و می‌توان اینگونه استنباط کرد که اثر رنگ بستگی به شکلی خواهد داشت که تعیین می‌گردد. همین ارزیابی را در مورد شکل کالا نیز می‌توان با بیان این سؤال که آیا اثر شکل بستگی به رنگی دارد که تعیین می‌گردد، مطرح ساخت.

در تحلیل واریانس (ANOVA) برای طرحهای عاملی، هر یک از این سه اثر از طریق آماره F مورد بررسی قرار می‌گیرد. جدول (۱۲-۴) اینگونه ارزیابی را برای یک مثال فرضی نشان می‌دهد.

جدول (۱۲-۴) نمونه‌ای از نتایج تحلیل واریانس برای یک طرح عاملی

منبع واریانس	مجموع میانگین	درجه آزادی	آماره F	Sig
رنگ	۴/۵	۲	۳/۰*	۰/۰۳۱
شکل	۱/۵	۲	۱/۰	۰/۱۵۸
شکل × رنگ	۱/۶	۴	۱/۱	۰/۲۱۰
	خطا (Error)	۱/۵	۶۳	

\*Significant;  $\alpha=0.05$ 

همانطوریکه از این مثال فرضی برمی آید، رنگ اثر معنی‌داری در ارزیابی کلی از محصول دارد در حالیکه شکل محصول تأثیر معنی‌داری ندارد. از طرف دیگر اثر رنگ بر روی هر یک از اشکال نیز از طریق اثر متقابل<sup>(۱)</sup> مورد ارزیابی قرار گرفته است که چنین اثری نیز معنی‌دار نمی‌باشد.

تحلیل واریانس چند متغیره (MANOVA) برای طرح‌های عاملی نیز از طریق بسط تحلیل واریانس (ANOVA) امکانپذیر می‌باشد. در تحلیل واریانس چند متغیره آماره لامبدای ویلکز بردار میانگین‌های متغیرهای وابسته را مورد ارزیابی قرار می‌دهد. همانطوریکه مشخص است انواع طرح‌های تحلیل واریانس (ANOVA)، نظیر طرح‌های عاملی<sup>(۲)</sup>، طرح‌های اندازه‌گیری مکرر<sup>(۳)</sup> و طرح‌های کرت‌های خرد شده<sup>(۴)</sup> وجود دارد که در منابع آماری مورد بحث قرار گرفته‌اند. برای هر یک از طرح‌های موجود در تحلیل واریانس (ANOVA)، همتایی در تحلیل واریانس چند متغیره (MANOVA) نیز وجود دارد.

1- Interaction

2- Factorial Designs

3- Repeated Measures Designs

4- Split- Plot Designs

### تحلیل کواریانس (ANCOVA)<sup>(۱)</sup> و تحلیل کواریانس چند متغیره (MANCOVA)<sup>(۲)</sup>:

در هر طرح تحلیل واریانس یک متغیره (ANOVA)، متغیرهای مستقل کمی، می‌توانند به عنوان متغیرهای کمکی<sup>(۳)</sup> در تحلیل در نظر گرفته شوند. در این صورت اینگونه طرحها، به عنوان طرح تحلیل کواریانس (ANCOVA) نامیده می‌شوند. تحلیل کواریانس زمانی قابل استفاده است که در آن متغیر وابسته کمی و چند متغیر مستقل کمی و کیفی وجود داشته باشد. در این گونه طرحها به متغیر مستقل کمی اصطلاح متغیر کمکی<sup>(۴)</sup> و به متغیر مستقل کیفی اصطلاح عامل<sup>(۵)</sup> اطلاق می‌گردد. متغیرهای کمکی پارامتری یا کمی، نوعاً در طرحهای تجربی<sup>(۶)</sup> و مطالعه پیمایشی<sup>(۷)</sup>، به منظور حذف و از بین بردن اثرات خارجی بر متغیر وابسته و افزایش دقت اندازه‌گیری مورد استفاده قرار می‌گیرد. در این نوع بررسیها، روشی شبیه رگرسیون خطی، برای رفع اختلاف در متغیر وابسته مرتبط با یک یا چند متغیر کمکی بکار گرفته می‌شود، و سپس یک تحلیل واریانس (ANOVA) مرسوم بر روی متغیر وابسته تعدیل شده انجام می‌گیرد.

یک متغیر کمکی مؤثر در تحلیل کواریانس (ANCOVA)، متغیری است که همبستگی بالایی با متغیر وابسته داشته، اما همبستگی با متغیرهای مستقل نداشته باشد. در اینجا واریانس متغیر وابسته، پایه‌ای را برای خطا در تحلیل واریانس تشکیل می‌دهد. اگر متغیر کمکی وجود داشته باشد که با متغیر وابسته همبستگی داشته باشد، می‌توان (از طریق تحلیل رگرسیون) بخشی از واریانس را تبیین کرد، و تنها واریانس باقیمانده، در متغیر وابسته باقی می‌ماند. این واریانس باقیمانده، تنها خطای کمی را برای آماره F باقی می‌گذارد<sup>(۸)</sup>.

1- Analysis of Covariance

2- Multivariate Analysis of Covariance

3- Covariates

4- Corariate

5- Factor

6- Exprimental Designs

7- Survey Research

8- Hair Joseph, etal (1999) "Multivariate Data Analysis" Macmillan pulishing Company, New York. PP.162-163.



استفاده از کواریانس (ANCOVA) شرایطی دارد که می‌توان آنرا بشرح ذیل خلاصه کرد:

۱- زمانیکه یک یا چند متغیر خارجی مزاحم وجود دارد که در متغیر وابسته اثر می‌گذارد.

۲- این متغیر مزاحم قابل اندازه‌گیری در مقیاس فاصله‌ای یا نسبی باشد.

۳- بین متغیر یا متغیرهای مزاحم و متغیر وابسته رابطه وجود داشته باشد.

۴- کنترل تجربی متغیرهای مزاحم خارجی امکان‌پذیر نباشد.

اگر این شرایط برقرار نباشند، تعدیلی که در متغیر وابسته از طریق کنترل عوامل مزاحم بدست می‌آید تورش دار خواهد بود، زیرا برخی اثرات قابل استناد، از متغیر وابسته حذف خواهد شد<sup>(۱)</sup>. تحلیل کواریانس نه تنها در چارچوب تحلیل واریانس (ANOVA) امکان‌پذیر است، بلکه بکارگیری آن در تحلیل رگرسیون نیز، هم ممکن و هم آسان می‌باشد. در تحلیل رگرسیون می‌توان به راحتی با کنترل برخی از متغیرها اثرات سایر متغیرهای مستقل را در تبیین متغیر وابسته بدست آورد. اثرات متغیرهای مزاحم از طریق مقایسه مقدار  $R^2$  کنترل شده و مقدار  $R^2$  کنترل نشده کاملاً قابل مشاهده می‌باشد.

تحلیل چند متغیره کوواریانس (MANCOVA)، بسط یافته اصول تحلیل کوواریانس (ANCOVA) به تحلیل چند متغیره است. بنابراین تحلیل چند متغیره کواریانس می‌تواند برای تبیین واریانس باقیمانده رگرسیون در متغیرهای وابسته که بوسیله متغیرهای کمکی تبیین نشده‌اند بکار گرفته شود.

۱- هومن، حیدرعلی (۱۳۸۰)، تحلیل داده‌های چند متغیری در پژوهش رفتاری. نشر پارسا، تهران، ص ۲۲۱-۲۲۲.

## فصل سیزدهم

# تحلیل همبستگی کانونی

مقدمه:

تحلیل همبستگی کانونی<sup>(۱)</sup>، از جمله روشهای تحلیل چند متغیره می باشد که ابتدا توسط هتلینگ<sup>(۲)</sup> در سال ۱۹۳۶ ارائه گردید. همانطوریکه در تحلیل رگرسیون چندگانه بیان شد، زمانی که محقق با یک متغیر وابسته کمی و مجموعه ای از متغیرهای مستقل سر و کار داشته باشد، می تواند برای پیش بینی تغییرات متغیر وابسته از تکنیک تحلیل رگرسیون چندگانه استفاده کند. اما در بعضی مواقع محقق ممکن است با مجموعه ای از متغیرهای وابسته و مجموعه ای از متغیرهای مستقل سروکار داشته باشد. در چنین مواقعی باید از روش تحلیل همبستگی کانونی استفاده کند. این فصل درصدد است تا ضمن بیان اصول حاکم بر این روش، نحوه محاسبه آن با استفاده از برنامه SPSS را توضیح دهد.

## اهداف همبستگی کانونی:

همبستگی کانونی در واقع همبستگی بین یک مجموعه متغیر وابسته و یک مجموعه متغیر مستقل می‌باشد. همانطوریکه رگرسیون چندگانه روابط بین مجموعه‌ای از متغیرهای مستقل را با یک متغیر وابسته مورد بررسی قرار می‌دهد، همبستگی کانونی نیز روابط بین تعدادی متغیر مستقل را با تعدادی متغیر وابسته بررسی می‌کند. مجذور همبستگی کانونی، درصد واریانس مجموعه متغیرهای وابسته را که بوسیله مجموعه متغیرهای مستقل تبیین می‌شود، نشان می‌دهد. این همبستگی، روابط بین دو مجموعه متغیرها را نشان می‌دهد نه همبستگی بین تک تک آنها را. بنابراین، اهداف این تکنیک عبارتست از:

- ۱- تعیین ارتباط بین مجموعه متغیرهای مستقل و مجموعه متغیرهای وابسته.
  - ۲- محاسبه مجموعه‌ای از وزنها برای هر یک از متغیرهای مستقل و وابسته، بطوریکه ترکیب خطی آنها حداکثر همبستگی را دارا باشد.
  - ۳- استخراج توابع خطی که باقیمانده همبستگی مستقل از مجموعه قبلی را به حداکثر برساند.
  - ۴- توضیح و تبیین هر نوع رابطه موجود بین مجموعه‌ای از متغیرهای مستقل و وابسته، از طریق تعیین سهم نسبی هر یک از متغیرها در تابع همبستگی کانونی استخراج شده.
- برای آشنایی بیشتر با تحلیل همبستگی کانونی، این روش آماری را می‌توان با برخی از تکنیکهای چند متغیره بشرح ذیل مقایسه کرد.

الف) تحلیل رگرسیون چندگانه:

$$X_1 + X_2 + X_3 + \dots + X_n = Y_1$$

(پارامتری)                      (پارامتری)

ب) تحلیل تشخیصی چندگانه:

$$X_1 + X_2 + X_3 + \dots + X_n = Y_1$$

(غیر پارامتری)                      (پارامتری)

ج) تحلیل واریانس چند متغیره:

$$X_1 + X_2 + X_3 + \dots + X_n = Y_1 + Y_2 + Y_3 + \dots + Y_n$$

(پارامتری) (غیرپارامتری)

(د) تحلیل همبستگی کانونی:

$$X_1 + X_2 + X_3 + \dots + X_n = Y_1 + Y_2 + Y_3 + \dots + Y_n$$

(پارامتری یا غیرپارامتری) (پارامتری یا غیرپارامتری)

### نحوه استنتاج توابع کانونی<sup>(۱)</sup>:

از آنجا که همبستگی کانونی جزو روشهای چندگانه خطی محسوب می‌گردد، بنابراین مفروضاتی که در اینگونه روشهای آماری نظیر، خطی بودن روابط، همواریانسی<sup>(۲)</sup> (یعنی اعتقاد به سطح روابط یکسان برای کل داده‌ها)، فاصله‌ای یا شبه فاصله‌ای بودن داده‌ها، و غیره وجود دارد، در مورد این روش نیز صادق می‌باشد.

همانطوریکه قبلاً نیز بیان گردید، محاسبه تحلیل همبستگی کانونی نیاز به دو دسته از متغیرها دارد. یک دسته متغیرهای مستقل و دسته دیگر متغیرهای وابسته. منطبق حاکم بر این روش بدست آوردن ترکیب خطی متغیرهای هر یک از مجموعه‌های فوق می‌باشد بطوریکه همبستگی بین دو ترکیب خطی محاسبه شده به حداکثر برسد. اما این روش تنها با محاسبه یک همبستگی کانونی بین مجموعه‌ای از متغیرها به اتمام نمی‌رسد. بلکه ممکن است با توجه به متغیرهای مورد استفاده چندین جفت ترکیب خطی بدست آید. تعداد متغیر کانونی<sup>(۳)</sup> بدست آمده، براساس تعداد متغیرهای مجموعه کوچک مورد نظر تعیین می‌شود و این مجموعه ممکن است مستقل یا وابسته باشد. به عنوان مثال اگر در مطالعه‌ای تعداد پنج متغیر مستقل و سه متغیر وابسته وجود داشته باشد. تعداد متغیر کانونی سه مورد خواهد بود.

1- Canonical Functions

2- Homoscedasticity

3- Canonical Variate

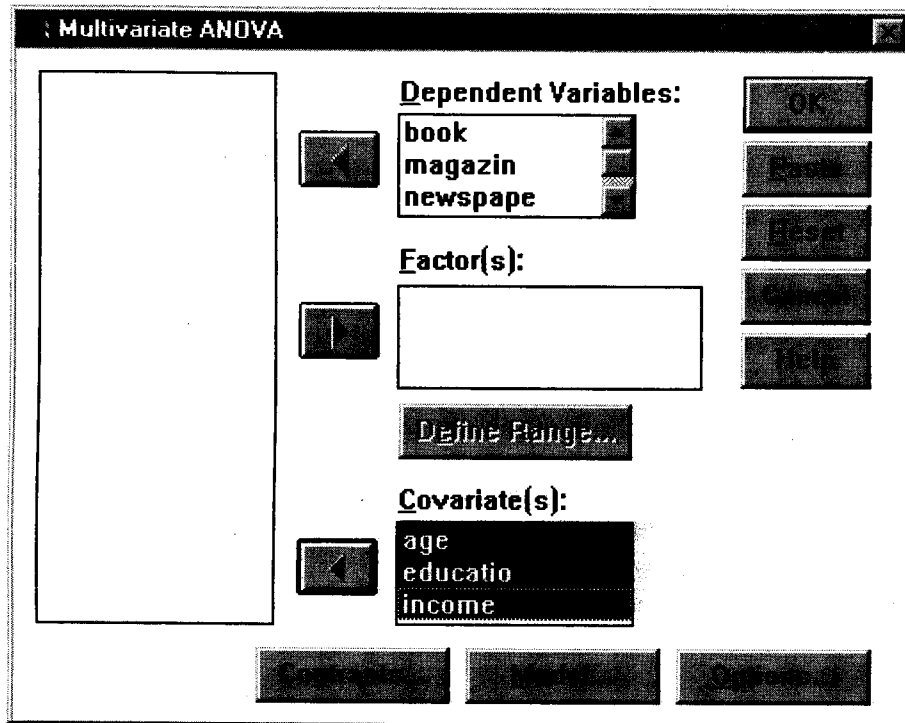
نحوه استنتاج متغیرهای کانونی شبیه فرآیند تحلیل عاملی چرخش نیافته است. بطوریکه در تحلیل عاملی، اولین عامل استخراج یافته بیشترین واریانس موجود در مجموعه داده‌ها را تبیین می‌کند. سپس دومین عامل حداکثر واریانس را که توسط اولین عامل قابل کنترل نبوده، تبیین می‌کند و به همین ترتیب عوامل بعدی استخراج می‌گردند. بنابراین عوامل متوالی، واریانس باقیمانده از عامل قبلی را تبیین می‌کند. تحلیل همبستگی کانونی نیز فرآیند مشابهی را طی می‌کند. اما حداکثر رابطه و همبستگی بین دو مجموعه از متغیرها را مدنظر دارد، نه روابط درون تنها یک مجموعه از متغیرها را. در اینجا نیز دومین متغیر کانونی بیشترین ارتباط و همبستگی بین دو مجموعه از متغیرها را که توسط متغیر کانونی اول کنترل نشده بود تبیین یا استنتاج می‌کند. بطور خلاصه، جفت‌های متوالی متغیرهای کانونی براساس واریانس باقیمانده از متغیرهای قبلی استنتاج می‌گردد و به همین دلیل همبستگی کانونی بدست آمده، بطور متوالی کوچک می‌شود.

#### روش محاسبه تحلیل کانونی با برنامه SPSS:

فرض کنید می‌خواهیم ارتباط بین ویژگیهای فردی، اجتماعی و اقتصادی افراد را با منابع و کانالهای ارتباطی که اطلاعات خود را بدست می‌آورند مورد بررسی قرار دهیم. در چنین مواقعی ممکن است مجموعه‌ای از متغیرهای مستقل نظیر میزان تحصیلات افراد، سن، درآمد و غیره در یک طرف و میزان استفاده از کانالها و منابع اطلاعاتی نظیر تلویزیون، کتاب، روزنامه و مجله به عنوان متغیرهای وابسته در طرف دیگر وجود داشته باشد. فرض اولیه این است که ویژگیهای فردی، اجتماعی و اقتصادی افراد بر استفاده از کانالهای اطلاعاتی مختلف تأثیر دارد. برای بررسی اینگونه روابط باید از تحلیل کانونی استفاده کرد. این روش بطور مستقل در برنامه SPSS وجود ندارد، اما برای محاسبه آن می‌توان از قسمت MANOVA استفاده کرد.

برای این کار در SPSS نسخه ۶ بساید دستور

Statistics/ANOVA Model/Multivariate را اجرا کرد تا شکل (۱-۱۳) ظاهر شود.



شکل (۱-۱۳) پنجره Multivariate

در این پنجره متغیرهای وابسته را به قسمت **Dependent Variables** و متغیرهای مستقل را به قسمت **Covariates** منتقل کنید. دکمه **Model** را فشار داده و مورد **Dimension Reduction** و **Eigenvalues, Multivariate Tests, Discriminant Analysis** را علامت زده و سپس دکمه **Continue** و بعد دکمه **Ok** را فشار دهید تا محاسبات در پنجره **Output** ظاهر شود (جدول ۱-۱۳).

جدول (۱۳-۱) نتایج محاسبات تحلیل همبستگی کانونی

\*\*\*\*\* Analysis of Variance -- design 1\*\*\*\*\*

EFFECT .. WITHIN+RESIDUAL Regression  
Multivariate Tests of Significance (S = 3, M = 0, N = 10 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	1.27544	4.62232	12.00	75.00	.000
Hotellings	9.68170	17.48085	12.00	65.00	.000
Wilks	.06338	9.35907	12.00	61.14	.000
Roys	.90153				

(الف)

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	9.155	94.561	94.561	.949	.902
2	.469	4.839	99.400	.565	.319
3	.058	.600	100.000	.234	.055

(ب)

Dimension Reduction Analysis

Roots	Wilks L.	F Hypoth. DF	Error DF	Sig. of F	
1 TO 3	.06338	9.35907	12.00	61.14	.000
2 TO 3	.64360	1.97202	6.00	48.00	.058
3 TO 3	.94513	.72570	2.00	25.00	.494

(ج)

Raw canonical coefficients for DEPENDENT variables  
Function No.

Variable	1	2	3
AGE	.029	.098	.184
EDUCATIO	.111	.158	-.562
INCOME	.007	-.030	.033

(د)

Standardized canonical coefficients for DEPENDENT variables  
Function No.

Variable	1	2	3
AGE	.224	.764	1.428
EDUCATIO	.523	.747	-2.654
INCOME	.333	-1.477	1.596

(ه)

Correlations between DEPENDENT and canonical variables  
Function No.

Variable	1	2	3
AGE	.807	.508	.302
EDUCATIO	.984	.035	-.173
INCOME	.915	-.397	.068

(9)

-----  
Variance in dependent variables explained by canonical variables

CAN. VAR.	Pct Var DE	Cum Pct DE	Pct Var CO	Cum Pct CO
1	81.908	81.908	73.842	73.842
2	13.898	95.806	4.434	78.276
3	4.194	100.000	.230	78.506

-----  
Raw canonical coefficients for COVARIATES  
Function No.

COVARIATE	1	2	3
BOOK	.091	-.020	-.261
MAGAZIN	.171	.100	-.482
NEWSPAPE	.124	.069	.513
TV	-.102	.543	-.086

-----  
Standardized canonical coefficients for COVARIATES  
CAN. VAR.

COVARIATE	1	2	3
BOOK	.284	-.061	-.813
MAGAZIN	.342	.200	-.961
NEWSPAPE	.410	.229	1.691
TV	-.180	.955	-.151

-----  
Correlations between COVARIATES and canonical variables  
CAN. VAR.

Covariate	1	2	3
BOOK	.930	-.044	-.145
MAGAZIN	.874	.270	-.307
NEWSPAPE	.913	.216	.344
TV	-.343	.936	-.028

\* \* \* \* \* A n a l y s i s o f V a r i a n c e -- design 1

Variance in covariates explained by canonical variables

CAN. VAR.	Pct Var DE	Cum Pct DE	Pct Var CO	Cum Pct CO
1	58.185	58.185	64.540	64.540
2	7.960	66.145	24.950	89.490
3	.322	66.467	5.871	95.361



VARIMAX rotated correlations between canonical variables and COVARIATES  
Can. Var.

DEP. VAR.	1	2	3
AGE	.625	.490	.507
EDUCATIO	.837	.424	.226
INCOME	.384	.897	.219

-----  
Transformation Matrix

	1	2	3
1	.584	.650	.487
2	.418	.273	-.866
3	-.696	.709	-.113

-----

### تفسیر نتایج بدست آمده از تحلیل کانونی:

همانند تحلیل عاملی، در همبستگی کانونی نیز ممکن است بیش از یک بعد معنی دار وجود داشته باشد. به عبارت دیگر، ممکن است بیش از یک همبستگی کانونی بدست آید، که در این صورت هریک از آنها الگوی روابط جداگانه متعامدی<sup>(۱)</sup>، بین مجموعه متغیرها را نشان می‌دهند. حداکثر تعداد همبستگی کانونی بین دو مجموعه متغیر، به تعداد متغیرهای مجموعه کوچک بستگی دارد و اولین همبستگی کانونی معمولاً بیشترین ارتباط را تبیین می‌کند.

همبستگی کانونی درست مانند همبستگی پیرسون تفسیر می‌گردد. اگر مقادیر آن مجذور شود، درصد واریانس تبیین شده یک مجموعه از متغیرها توسط مجموعه دیگر را نشان می‌دهد.

همبستگی کانونی ادغام شده<sup>(۲)</sup>، مجموع مجذورات کلیه ضرایب همبستگی کانونی است که نشان می‌دهد، یک مجموعه از متغیرها چه میزان از تغییرات مجموعه متغیرهای دیگر را تبیین یا پیشگویی می‌کند.

1- Orthogonal

2- Pooled Canonical Correlation(Pooled  $R^2_c$ )

جدول ANOVA آزمون‌های سطح معنی داری مختلفی را نشان می‌دهد. این آزمون‌ها عبارتند از: پیلائی<sup>(۱)</sup>، هتلینگ<sup>(۲)</sup>، ویلکز<sup>(۳)</sup> و ری<sup>(۴)</sup>. متداولترین آماره برای آزمون سطح معنی داری لامبدای ویلکز می‌باشد که سطح معنی داری اولین همبستگی کانونی را نشان می‌دهد. اگر مقدار P بدست آمده کوچکتر از ۰/۰۵ باشد می‌توان قضاوت کرد که دو مجموعه از متغیرها بطور معنی داری بوسیله همبستگی کانونی پیوند داشته‌اند. در قسمت الف جدول (۱-۱۳) نیز مشخص است که سطح معنی داری (sig = 0.000) با احتمال ۹۹ درصد وجود همبستگی کانونی بین دو مجموعه متغیر را تأیید می‌کند. تأکید می‌گردد که آزمون لامبدای ویلکز تنها به منظور بررسی سطح معنی داری اولین همبستگی کانونی مورد استفاده قرار می‌گیرد، نه ضرورتاً دومین همبستگی کانونی. برای بررسی سطح معنی داری دومین همبستگی کانونی می‌توان از آماره V بارتلت<sup>(۵)</sup> استفاده کرد، اما این آماره در برنامه SPSS محاسبه نمی‌گردد. اگر اولین همبستگی کانونی معنی دار نباشد، طبیعتاً همبستگی‌های کانونی بعدی نیز معنی دار نخواهند بود.

آزمون دیگری که برای بررسی سطح معنی داری مورد استفاده قرار می‌گیرد، آزمون نسبت درست‌نمایی<sup>(۶)</sup> می‌باشد، که در برنامه SAS قابل دستیابی است. این آزمون نیز برای بررسی معنی داری روابط خطی کلیه متغیرهای موجود در دو متغیر کانونی مورد استفاده قرار می‌گیرد و نه، تنها برای اولین همبستگی کانونی و یا هر نوع همبستگی کانونی مجزا<sup>(۷)</sup>.

زمانیکه بیش از یک همبستگی کانونی (که با برجسب Roots نیز نشان داده می‌شود) استخراج گردد، آماره مقدار ویژه<sup>(۸)</sup> نشان می‌دهد که چه سهمی از واریانس، بوسیله هر یک از

1- Pillai

2- Hotelling

3- Wilks

4- Roy

5- Bartlett's V

6- Likelihood Ratio Test

7- <http://www2.chass.ncsu.edu/garson/pa765/canonic.htm>, 2002, canonical correlation

8- Eigenvalue

همبستگی‌های کانونی مربوط به دو مجموعه متغیر تبیین می‌شود. هر متغیر کانونی دارای یک مقدار ویژه بوده و معمولاً اولین همبستگی کانونی مهمتر از دیگر همبستگی‌هاست. (قسمت ب جدول ۱۳-۱). در این مثال اولین همبستگی کانونی، در حدود ۹۰ درصد از واریانس متغیر وابسته را تبیین می‌کند. مقدار واریانس تبیین شده از طریق مجذور کردن مقدار همبستگی کانونی بدست می‌آید. به عنوان مثال اگر مقدار همبستگی کانونی مربوط به اولین همبستگی کانونی مجذور گردد، مقدار واریانس تبیین شده بدست می‌آید ( $0/949 \times 0/949 = 0/902$ ).

ضرایب همبستگی کانونی استاندارد شده، اهمیت نسبی هریک از متغیرهای اصلی را در محاسبه مقدار کانونی<sup>(۱)</sup> هریک از متغیرهای کانونی نشان می‌دهد. ضریب کانونی<sup>(۲)</sup> که به آن وزن کانونی<sup>(۳)</sup> نیز می‌گویند، مانند مقادیر بتا در تحلیل رگرسیون می‌باشد.

برای هریک از متغیرهای اصلی مربوط به هرکدام از مجموعه متغیرهای مستقل و وابسته یک ضریب کانونی بدست می‌آید. بنابراین، برای مجموعه متغیرهای وابسته، که در آن چهار متغیر وابسته و سه همبستگی کانونی (براساس کمترین تعداد در مجموعه متغیرهای مستقل و وابسته) وجود داشته باشد، مجموعاً ۱۲ ضریب کانونی بدست می‌آید که در قالب سه مجموعه، که هرکدام چهار ضریب را شامل می‌گردد، نشان داده می‌شود. (قسمت ه جدول ۱۳-۱). در این مثال آموزش و درآمد به ترتیب بیشترین نقش را ایجاد اولین همبستگی کانونی داشته‌اند.

ضرایب همبستگی ساختار<sup>(۴)</sup>، (قسمت و جدول ۱۳-۱) که به آن بارهای عاملی کانونی<sup>(۵)</sup> نیز می‌گویند، عبارتست از همبستگی کانونی یک متغیر با متغیر اصلی، و این شبیه بارهای عاملی است که در تحلیل عاملی نیز وجود دارد. این ماتریس نشان می‌دهد که چگونه هر متغیر اصلی، بر هر یک از متغیرهای کانونی سه گانه بار شده‌اند. همبستگی ساختار برای دو هدف بکار می‌رود. اولین کاربرد آن

1- Canonical score

2- Canonical coefficient

3- Canonical weight

4- Structure correlation coefficient

5- Canonical Factors Loadings

استفاده از آن برای تفسیر متغیرهای کانونی است. براساس یک قاعده‌ای کلی، متغیرهایی که همبستگی ساختار آنها بزرگتر از  $0/3$  باشد، به عنوان بخشی از متغیر کانونی<sup>(۱)</sup> محسوب می‌گردند و متغیرهایی که همبستگی ساختار آنها کمتر از مقدار فوق باشد، به عنوان بخشی از متغیر کانونی که بار معنی دار داشته باشد، لحاظ نمی‌گردد<sup>(۲)</sup>. دومین کاربرد همبستگی ساختار، محاسبه واریانس تبیین شده بوسیله متغیرهای اصلی<sup>(۳)</sup> است. چنانچه مقادیر ساختار مجذور شوند، درصد واریانس که بوسیله یک متغیر اصلی با متغیرهای کانونی تبیین می‌شود، بدست می‌آید. همین عملیات برای متغیرهای مستقل نیز محاسبه و در قالب متغیرهای کمکی<sup>(۴)</sup> در نتایج ظاهر می‌شود. علاوه بر آن در برنامه SPSS نتایج مربوط به Effect:Constant نیز محاسبه می‌گردد. اما بدلیل اینکه این محاسبات معمولاً در همبستگی کانونی مورد استفاده قرار نمی‌گیرد، بنابراین از جدول (۱-۱۳) حذف شده‌اند.

---

1- Cononical Variable

2- <http://www2.chass.ncsu.edu/garson/pa765/canonic.htm>, Year 2002, Canonical Correlation.

3- Original variables

4- Covariate variables

## فصل چهاردهم

# تحلیل عاملی

مقدمه:

تحلیل عاملی از جمله روشهای چند متغیره است که در آن متغیرهای مستقل و وابسته مطرح نیست. زیرا این روش جزو تکنیکهای هم وابسته محسوب می‌گردد و کلیه متغیرها نسبت به هم وابسته لحاظ گردیده و سعی می‌شود تا تعداد زیادی متغیر در چند عامل خلاصه شوند. تحلیل عاملی نوع Q و نوع R از مهمترین جنبه‌های این تکنیک آماری می‌باشد که بدلیل اهمیت موضوع هر دو این روشها بطور مشروح در این فصل مورد بحث و بررسی قرار گرفته‌اند. برای تبیین بیشتر موضوع از مثالهای عینی نیز کمک گرفته شده است.

بنیادهای اولیه تحلیل عاملی:

گالتون<sup>(۱)</sup> دانشمند اواخر قرن نوزده و آغاز قرن بیستم اولین کسی بود که بنیادهای اولیه تحلیل

عاملی<sup>(۱)</sup> را بنا نهاد. کارل پیرسون<sup>(۲)</sup> در اوایل قرن بیستم در مقاله‌ای معروف<sup>(۳)</sup> برای اولین بار روشی برای تحلیل عاملی براساس یک فضای چند بعدی هندسی ارائه داد. این روش در سال ۱۹۰۲ توسط مک‌دونال<sup>(۴)</sup> در زمینه شناسایی جرائم و ارتباط آن با ویژگیهای فیزیکی افراد نظیر شکل سر، بدن، اندازه لب و غیره عملاً مورد استفاده قرار گرفت. او با این روش ۳۰۰۰ نفر از مجرمین را با ۱۰۰۰ نفر از دانشجویان دانشگاه کمبریج مورد مقایسه قرار داد و به تحلیل مجرمین و ارتباط آن با وضع ظاهری و فیزیکی شان پرداخت.

اسپیرومن<sup>(۵)</sup> نیز در سال ۱۹۰۴ در گزارشی به معرفی مدل‌های ریاضی این روش همت گماشت. با این تلاشهای اولیه اصول و بنیانهای این روش، تدوین و مورد استفاده علوم مختلف قرار گرفت. امروزه روش تحلیل عاملی از جمله تکنیکهای تحلیل آماری است که در سطح وسیع توسط شعبات مختلف علوم نظیر روانشناسی، جامعه‌شناسی، اقتصاد، مدیریت، جغرافیا، برنامه ریزی شهری و منطقه‌ای، برنامه ریزی روستایی، و حتی رشته‌های مختلف کشاورزی، بیولوژی، زمین‌شناسی، پزشکی و غیره مورد استفاده قرار می‌گیرد.

#### تحلیل عاملی چیست؟

تحلیل عاملی نامی عمومی است برای برخی از روشهای آماری چند متغیره که هدف اصلی آن خلاصه کردن داده‌هاست. این روش به بررسی همبستگی درونی تعداد زیادی از متغیرها می‌پردازد و در نهایت آنها را در قالب عاملهای عمومی محدودی دسته بندی کرده و تبیین می‌کند.

1- Factor Analysis

2- Karl Pearson

3- Pearson, K. (1901) "On lines and plans of closest fit to systems of points in space". *Pill, Mag.*

4- Macdonnell, W. R. (1902) "On criminals anthropometry and the identification of criminals"

*Blometrika*, 1, 177-277.

5- Spearman

تحلیل عاملی (برخلاف رگرسیون چندگانه<sup>(۱)</sup>، تحلیل تشخیصی<sup>(۲)</sup> یا همبستگی کانونی<sup>(۳)</sup>) که در آن یک یا چند متغیر وابسته و تعداد زیادی متغیر مستقل وجود دارد) روشی هم وابسته<sup>(۴)</sup> بوده که در آن کلیه متغیرها بطور همزمان مدنظر قرار می‌گیرند. در این تکنیک، هر یک از متغیرها به عنوان یک متغیر وابسته لحاظ می‌گردد.

این روش در دهه‌های اخیر بویژه با عمومیت یافتن استفاده از رایانه در تحقیقات در سطح وسیع مورد استفاده محققان قرار گرفته است. قبل از پرداختن به این تکنیک آماری، لازم است برخی از مفاهیم کلیدی این روش معرفی شوند.

- اشتراک<sup>(۵)</sup>: اشتراک عبارتست از، میزان واریانس مشترک بین یک متغیر با سایر متغیرهای بکار گرفته شده در تحلیل.

- مقدار خاص<sup>(۶)</sup>: مقدار خاص، میزان واریانس تبیین شده بوسیله هر عامل را بیان می‌کند.  
- عامل<sup>(۷)</sup>: عبارتست از ترکیب خطی متغیرهای اصلی، که نشان دهنده جنبه‌های خلاصه شده‌ای از متغیرهای مشاهده شده است.

- بار عاملی<sup>(۸)</sup>: عبارتست از همبستگی بین متغیرهای اصلی و عوامل. اگر مقادیر بار عاملی مجذور شوند، نشان می‌دهند که چند درصد از واریانس در یک متغیر توسط آن عامل تبیین می‌شود. ماتریس عاملی<sup>(۹)</sup>: جدولی است که بارهای عاملی کلیه متغیرها را در هر عامل نشان می‌دهد.

- چرخش عاملی<sup>(۱۰)</sup>: فرآیندی است برای تعدیل محور عامل به منظور دستیابی به عاملهای

- |                          |                         |
|--------------------------|-------------------------|
| 1- Multiple Regression   | 2- Discriminat Analysis |
| 3- Canonical Correlation | 4- Interdependence      |
| 5- Communnality          | 6- Eigenvalue           |
| 7- Factor                | 8- Factor Loading       |
| 9- Factor Matrix         | 10- Factor Rotation     |

معنی‌دار و ساده. نمره عاملی<sup>(۱)</sup>: تحلیل عاملی متغیرهای اصلی را در تعداد محدودی از عاملها خلاصه می‌کند. وقتی این عاملهای محدود جدید در تحلیل‌های بعدی (مانند تحلیل تشخیصی یا رگرسیون) استفاده شود، بعضی از مقادیر باید برای استنتاج متغیرهای جدید بکار گرفته شوند. این مقادیر در واقع ترکیبی از کلیه متغیرهای اصلی است که در ساختن عامل جدید نقش عمده‌ای داشته‌اند. این ترکیب متغیرها، نمره عاملی نامیده می‌شوند.

#### فرآیند تحلیل عاملی:

هدف اصلی تحلیل عاملی تلخیص تعداد زیادی از متغیرها در تعداد محدودی از عاملها می‌باشد، بطوریکه در این فرآیند کمترین میزان گم شدن اطلاعات وجود داشته باشد. تحلیل عاملی کاربردهای مختلفی در تحلیل داده‌ها دارد که مهمترین آن را می‌توان بشرح ذیل خلاصه کرد.

۱ - دستیابی به ابعادی که بصورت پنهانی در مجموعه وسیعی از متغیرها وجود دارد ولی به آسانی قابل مشاهده نمی‌باشند. این نوع تحلیل عاملی به تحلیل عاملی نوع  $R^{(۲)}$  معروف است که در آن تعداد زیادی متغیر در تعداد محدودی از عاملها خلاصه می‌شوند.

۲ - ابداع روشی برای ترکیب و تلخیص تعداد زیادی از افراد در گروههای مختلف در درون یک جامعه بزرگ. این روش به تحلیل عاملی نوع  $Q^{(۳)}$  معروف است که در آن افراد یا موارد در تعدادی گروه طبقه بندی می‌شوند.

۳ - شناسایی متغیرهای مناسب از بین مجموعه وسیعی از متغیرها به منظور استفاده از آنها برای تحلیل‌های بعدی در رگرسیون چندگانه، یا تحلیل تشخیصی.

۴ - ایجاد مجموعه کوچک و کاملاً جدیدی از متغیرها که بطور کامل به جای متغیرهای

1- Factor Score

2- R-Type Factor Analysis

3- Q-Type Factor Analysis



اصلی در تحلیل‌های بعدی رگرسیون، یا تحلیل تشخیصی مورد استفاده قرار گیرد. رهیافتهای اول و دوم به منظور شناسایی ابعاد یا عاملها مورد استفاده قرار می‌گیرد. روش سوم با دستیابی به بارهای عاملی<sup>(۱)</sup>، آنها را به عنوان پایه‌ای برای شناسایی متغیرها برای تحلیل‌های بعدی بوسیله سایر تکنیکهای آماری قرار می‌دهد. روش چهارم، رهیافتی است که از طریق آن مقادیر عاملی<sup>(۲)</sup> محاسبه گردیده و سپس این مقادیر به عنوان متغیرهای مستقل در تحلیل رگرسیون یا تحلیل تشخیصی مورد استفاده قرار می‌گیرد.

#### نمودار تصمیم‌گیری در تحلیل عاملی:

شکل (۱۴-۱) مراحل عمومی کاربرد تکنیک عاملی را نشان می‌دهد. نقطه آغازین در تحلیل عاملی مانند سایر روشهای آماری بیان مسئله است. اگر هدف از تحقیق، کاهش داده‌ها و خلاصه کردن آنهاست، تحلیل عاملی تکنیک مناسبی برای آن می‌باشد. در این مرحله سوالاتی که محقق باید از خود بپرسد این است که چه نوع متغیرهایی باید در تحلیل بکار گرفته شود؟ چند متغیر در تحلیل باید مورد استفاده قرار گیرد؟ نوع داده‌ها از نظر سطوح اندازه‌گیری چگونه باید باشد؟ و حجم نمونه چقدر باید باشد؟ در پاسخ به این سوالات باید گفت که هر نوع متغیر مرتبط با مسئله تحقیق را می‌توان در تحلیل بکار گرفت. داده‌های خام برای تحلیل عاملی باید از نوع کمی باشند، اما در مواقعی می‌توان از متغیرهای مجازی<sup>(۳)</sup> (با کدهای ۰ و ۱) و غیرپارامتری یا کیفی نیز استفاده کرد.

یکی از روشهای انتخاب متغیرهای مناسب برای تحلیل عاملی استفاده از ماتریس همبستگی است. از آنجا که اساس روش تحلیل عاملی بر همبستگی بین متغیرها اما از نوع غیر علی استوار است، بنابراین در استفاده از این روش باید ماتریس همبستگی بین متغیرها نیز محاسبه گردد. معمولاً اینگونه ماتریسهای همبستگی وجود رابطه بین برخی متغیرها و عدم ارتباط آن با برخی دیگر

1- Factor Loadings

2- Factor Scores

3- Dummy Variables

را نشان می‌دهند. این الگو در تحلیل عاملی موجب شکل‌گیری خوشه‌هایی می‌گردد که متغیرهای درون خوشه با یکدیگر همبستگی داشته و با متغیرهای خوشه‌های دیگر همبستگی نداشته باشند. توصیه می‌شود متغیرهایی که با هیچ متغیری همبستگی لازم را نداشته باشند از تحلیل حذف گردند. البته آماره‌های دیگری نیز وجود دارند که محقق از طریق آنها نیز قادر به تعیین و تشخیص مناسب بودن داده‌ها برای تحلیل عاملی می‌باشد. از جمله این روشها، آزمون KMO<sup>(۱)</sup> می‌باشد که مقدار آن همواره بین ۰ و ۱ در نوسان است. در صورتی که مقدار KMO کمتر از ۰/۵۰ باشد داده‌ها برای تحلیل عاملی مناسب نخواهند بود و اگر مقدار آن بین ۰/۵۰ تا ۰/۶۹ باشد می‌توان با احتیاط بیشتر به تحلیل عاملی پرداخت. اما در صورتی که مقدار آن بزرگتر از ۰/۷۰ باشد همبستگی‌های موجود در بین داده‌ها برای تحلیل عاملی مناسب خواهند بود<sup>(۲)</sup>.

از سوی دیگر برای اطمینان از مناسب بودن داده‌ها برای تحلیل عاملی مبنی بر اینکه ماتریس همبستگی‌هایی که پایه تحلیل عاملی قرار می‌گیرد در جامعه برابر با صفر نیست، باید از آزمون بارتلت<sup>(۳)</sup> استفاده کرد. آزمون بارتلت این فرضیه را که ماتریس همبستگی‌های مشاهده شده، متعلق به جامعه‌ای با متغیرهای ناپسته است، می‌آزماید. برای آنکه یک مدل عاملی، مفید و دارای معنا باشد، لازم است متغیرها همبسته باشند، در غیر این صورت دلیلی برای تبیین مدل عاملی وجود ندارد. اگر این فرضیه که متغیرها با هم رابطه ندارند رد نشود، کاربرد تحلیل عاملی زیر سؤال خواهد رفت و بنابراین باید در آن تجدید نظر کرد<sup>(۴)</sup>. به همین دلیل است که قبل از تحلیل عاملی باید به تشکیل ماتریس همبستگی بین متغیرها اقدام کرد.

1- Kaiser-Meyer-Olkin

۲- دواس، دی، ای (۱۳۷۶) "پیمایش در تحقیقات اجتماعی" ترجمه هوشنگ ناییبی، نشر نی، ص ۲۵۶

3- Bartlett Test

۴- مرکز آموزش مدیریت دولتی و دانشگاه آزاد اسلامی (۱۳۷۹) "مقیاس سنجش سبک مدیریت" (M.S.S)، انتشارات مرکز

آموزش مدیریت دولتی، ص ۷۴-۷۵

در رابطه با حجم نمونه نیز باید تأکید کرد که تعداد نمونه نباید کمتر از ۵۰ مورد باشد و ترجیحاً باید حجم نمونه را به بیش از ۱۰۰ مورد افزایش داد. به عنوان یک قاعده کلی تعداد نمونه باید در حدود چهار یا پنج برابر تعداد متغیرهای مورد استفاده باشد. این نسبت تا حدودی محافظه کارانه است، در بسیاری از مواقع محقق مجبور است تا با نسبت ۲ به ۱ نیز به تحلیل عاملی بپردازد. اما زمانیکه این نسبت پایین باشد و حجم نمونه نیز کم باشد، تفسیر یافته‌ها باید با احتیاط بیشتری انجام گیرد.

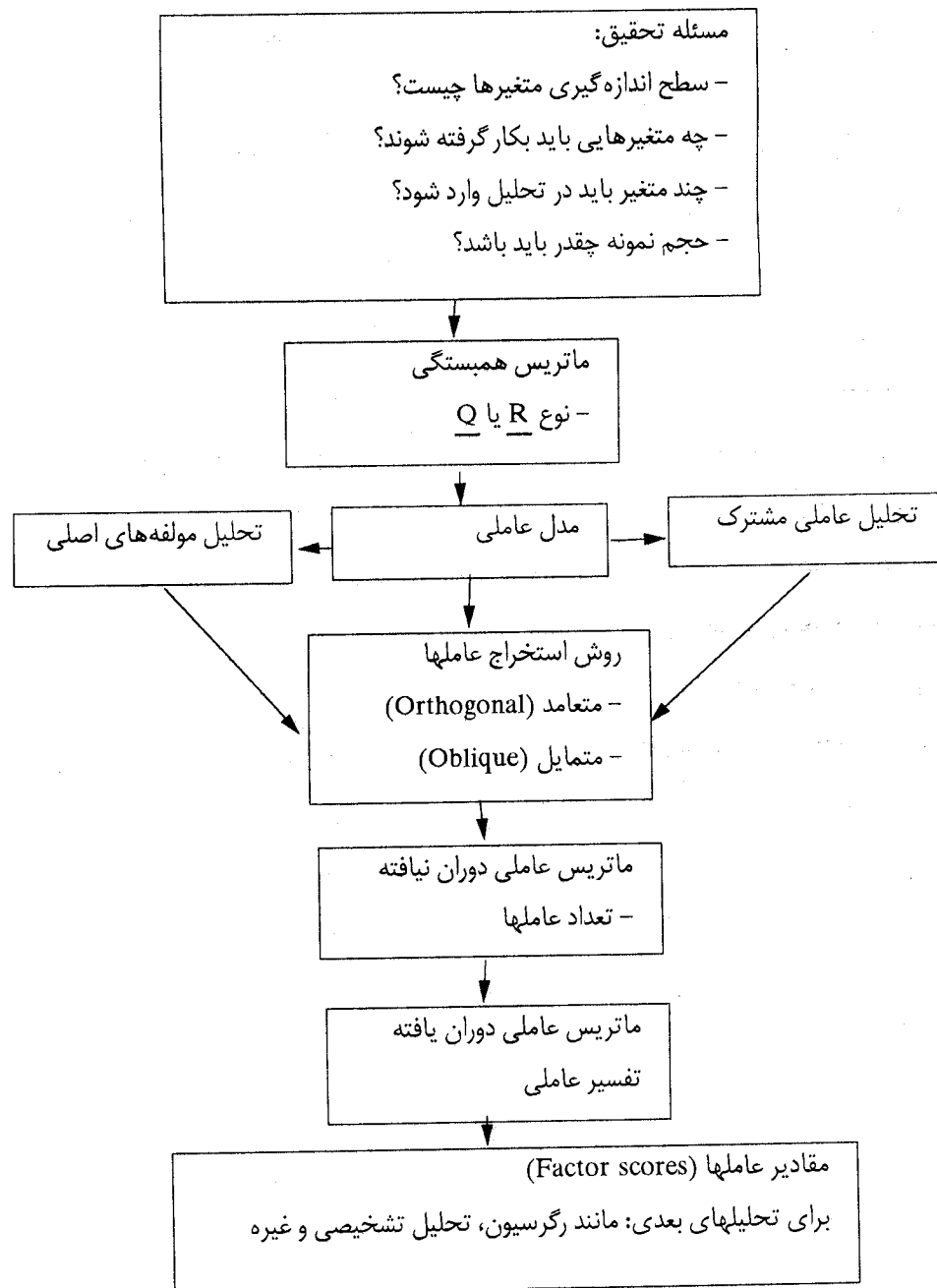
اولین تصمیم در بکارگیری تحلیل عاملی، محاسبه ماتریس همبستگی است. برای اینکار باید مشخص شود که آیا هدف، محاسبه همبستگی بین متغیرهاست یا بین پاسخگویان. به عنوان مثال ممکن است داده‌هایی از ۱۰۰ نفر از پاسخگویان در مورد ۱۰ متغیر گردآوری شده باشد. این امکان وجود دارد که همبستگی بین هریک از ۱۰ متغیر و یا بین هریک از پاسخگویان محاسبه شود. اگر هدف مطالعه، تلخیص متغیرها باشد، در این صورت باید همبستگی بین متغیرها محاسبه شود. این روش یکی از تکنیک‌های عمومی و پر کاربرد در مطالعات می‌باشد که به تحلیل عاملی نوع R<sup>(۱)</sup> معروف است. اما تحلیل عاملی ممکن است برای ماتریس همبستگی بین پاسخگویان نیز بکار گرفته است. این نوع تحلیل را تحلیل عاملی نوع Q<sup>(۲)</sup> می‌نامند. این نوع تحلیل عاملی شاید بدلیل مشکل بودن کمتر مورد استفاده محققان قرار گرفته است و بجای آن از روشهایی نظیر تحلیل خوشه‌ای<sup>(۳)</sup> یا گروه‌بندی سلسله‌مراتبی<sup>(۴)</sup> برای طبقه‌بندی پاسخگویان یا موارد استفاده می‌شود.

1- R-Type Factor Analysis

2- Q-Type Factor Analysis

3- Cluster Analysis

4- Hierarchical Grouping Technique



شکل (۱-۱۴) دیاگرام تصمیم‌گیری تحلیل عاملی

در تحلیل عاملی مدل‌های مختلفی وجود دارد که از میان آنها دو روش تحلیل مؤلفه‌های اصلی<sup>(۱)</sup> و تحلیل عاملی مشترک<sup>(۲)</sup> از پرکاربردترین این روشهاست<sup>(۳)</sup>. انتخاب هریک از این مدلها به هدف محقق بستگی دارد. مدل تحلیل مؤلفه‌های اصلی زمانی مورد استفاده قرار می‌گیرد که هدف محقق تلخیص متغیرها و دستیابی به تعداد محدودی عامل برای اهداف پیش بینی باشد. در مقابل، تحلیل عاملی مشترک زمانی بکار می‌رود که هدف شناسایی عاملها یا ابعادی باشد که به سادگی قابل شناسایی نیستند.

علاوه بر انتخاب مدل تحلیل، محقق باید مشخص کند که عاملها چگونه باید استخراج شوند. برای استخراج عاملها دو روش وجود دارد. عاملهای متعامد<sup>(۴)</sup> و عاملها متمایل<sup>(۵)</sup>. در روش متعامد، عاملها به شیوه‌ای استخراج می‌شوند که محورهای عاملی در حالت ۹۰ درجه قرار می‌گیرند، و این بدین معناست که هر عامل، مستقل از سایر عاملها می‌باشد. بنابراین، همبستگی بین عاملها، بطور قراردادی صفر تعیین می‌گردد. مدل عاملی متمایل پیچیده‌تر از مدل متعامد می‌باشد. در واقع در این روش، فرآیند تحلیلی کاملاً رضایت بخش بدست نمی‌آید. در این روش عاملهای استخراج شده دارای همبستگی می‌باشند. روش عاملی متمایل فرض می‌کند که متغیرهای اصلی دارای همبستگی بوده بنابراین عاملها نیز باید دارای حدی از همبستگی باشند. بطور خلاصه مدل عاملی متعامد از نظر ریاضی ساده بوده، در حالیکه مدل عاملی متمایل قابل تعدیل می‌باشد و بیشتر واقع‌گراست، زیرا از نظر تئوریک این فرض پذیرفته نیست که ابعاد و عاملهای مورد نظر با همدیگر مرتبط نباشند.

انتخاب اینکه چرخش عاملها بصورت متعامد یا متمایل باشد باید براساس نیازهای خاصی که در مسئله تحقیق وجود دارد انجام گیرد. اگر هدف تحقیق تلخیص تعداد متغیرهای اصلی بدون

1- Principal Components Analysis

2- Common Factor Analysis

۳- برای آشنایی بیشتر با روش تحلیل مؤلفه‌های اصلی به کتاب "برنامه‌ریزی و توسعه منطقه‌ای" (تئوریا و تکنیکها)، خلیل

کلانتری، انتشارات خوشبین، ۱۳۸۰ مراجعه کنید.

4- Orthogonal Factors

5- Oblique Factors

توجه به اینکه نتایج عاملهای استخراج شده تا چه حد معنی‌دار خواهد بود انجام گیرد، در این صورت روش متعامد انتخاب مناسبی خواهد بود. همچنین اگر محقق بخواهد از طریق تلخیص تعداد زیادی متغیر به مجموعه کوچکی از متغیرهای غیرهمبسته برای تحلیل‌های بعدی از طریق رگرسیون و سایر روشهای پیش‌بینی دست یابد، در این صورت نیز روش متعامد، تکنیک مناسبی خواهد بود. در حالیکه اگر هدف نهایی از تحلیل عاملی، دستیابی به عواملی که از نظر تئوریک معنی‌دار باشد مدنظر قرار گیرد، روش متمایل، مناسبترین تکنیک خواهد بود<sup>(۱)</sup>.

زمانیکه در خصوص ماتریس همبستگی، مدل عاملی و روش استخراج، تصمیم مناسب اتخاذ گردید، زمینه برای استخراج عاملهای اولیه چرخش نیافته فراهم می‌گردد. با بررسی ماتریس چرخش نیافته محقق می‌تواند به جستجوی روشهای تلخیص داده‌ها و تعیین تعداد عاملهای استخراجی بپردازد، اما تعیین نهایی تعداد عاملها پس از دستیابی به ماتریس عاملی چرخش یافته امکان‌پذیر می‌باشد.

بسته به دلایل استفاده از تکنیک تحلیل عاملی، محقق ممکن است با تفسیر عاملها کار خود را به اتمام برساند و یا ممکن است به محاسبه مقادیر عاملی<sup>(۲)</sup> بپردازد تا داده‌هایی برای تحلیل‌های بعدی بوسیله سایر تکنیکهای آماری بدست آید. اگر هدف، ترکیب منطقی متغیرها و یا پاسخگویان باشد در این صورت فرآیند تحلیل با تفسیر عاملها پایان می‌یابد. اگر هدف شناسایی متغیرهای مناسب برای کاربردهای بعدی توسط سایر تکنیکهای آماری باشد، محقق ماتریس عاملی<sup>(۳)</sup> را مورد بررسی قرار داده و متغیرهایی که بیشترین بار عاملی<sup>(۴)</sup> را داشته باشند به عنوان جانشین برای بعد خاصی از عامل انتخاب می‌کند و اگر هدف ایجاد مجموعه کوچک کاملاً جدیدی از متغیرها به منظور جایگزین کردن متغیرهای اصلی برای استفاده در سایر تکنیکهای آماری باشد، مقادیر عاملی

1- Hair, Jr. Joseph and et al (1990) "Multivariate Data Analysis" Macmillan publishing company.

New York.

2- Factor Scores

3- Factor matrix

4- Factor Loading

ترکیبی<sup>(۱)</sup> برای معرفی هریک از عاملها محاسبه می‌گردد. در این صورت مقادیر عاملی به عنوان داده‌های خام به جای متغیرهای مستقل در تحلیل رگرسیون، تحلیل تشخیصی و غیره استفاده می‌گردد.

### رهیافت‌های استخراج ماتریس همبستگی:

همانطوریکه قبلاً نیز بیان گردید، اولین تصمیم در بکارگیری تحلیل عاملی تعیین رهیافت محاسبه ماتریس همبستگی است. محقق می‌تواند ماتریس همبستگی را براساس همبستگی بین متغیرها محاسبه کند، که در این صورت تحلیل عاملی از نوع R بوده و الگوی بدست آمده روابط بین متغیرها را نشان می‌دهد. همچنین می‌تواند ماتریس همبستگی را براساس همبستگی بین افراد پاسخگو بدست آورد، که به تحلیل عاملی از نوع Q منجر می‌گردد. در این صورت ماتریس عاملی نشان دهنده افراد همگن خواهد بود. به عنوان مثال، اگر پاسخگویان بوسیله شماره مشخص شوند، نتایج تحلیل عاملی از نوع Q ممکن است نشان دهد که افراد شماره‌های ۱، ۵، ۷ و ۱۰ شباهت بیشتری نسبت بهم دارند. از آنجا که این افراد بار بیشتری بر یک عامل دارند بنابراین، افراد فوق می‌توانند در یک گروه قرار گیرند. همچنین ممکن است پاسخگویان شماره‌های ۲، ۳، ۴ و ۸ بار بیشتری بر روی یک عامل داشته باشند و در یک گروه طبقه بندی شوند. بنابراین از طریق تحلیل عاملی نوع Q می‌توان افراد را براساس الگوی پاسخ مشابه آنها به متغیرها طبقه بندی کرد.

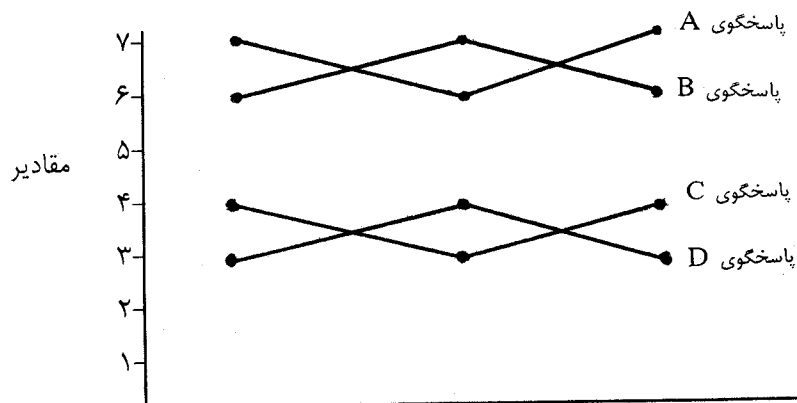
سؤال اساسی که در اینجا مطرح می‌شود این است که تفاوت بین تحلیل عاملی نوع Q با تحلیل خوشه‌ای چیست؟ پاسخ این است که هر دو این روشها پاسخگویان را براساس تعدادی از متغیرها مقایسه کرده و آنها را در گروه‌های مختلف طبقه بندی می‌کند. اما تفاوت این دو تکنیک در این است که گروه‌بندی افراد در تحلیل عاملی نوع Q براساس همبستگی درونی بین میانگین‌ها و انحراف معیارهای پاسخگویان انجام می‌شود. اما در تحلیل خوشه‌ای، گروه‌بندی براساس

1- Composite Factor Score

اندازه‌گیری فاصله (فاصله اقلیدسی و غیره) بین پاسخگویان در خصوص متغیرهای مورد استفاده در تحلیل انجام می‌گیرد. برای آشنایی با این تفاوتها به جدول (۱-۱۴) توجه کنید.

جدول (۱-۱۴) مقایسه تحلیل عاملی نوع Q با تحلیل خوشه‌ای سلسله مراتبی

پاسخگویان	متغیرها		
	الف	ب	ج
A	۷	۶	۷
B	۶	۷	۶
C	۴	۳	۴
D	۳	۴	۳



شکل (۲-۱۴) مقایسه نموداری تحلیل عاملی نوع Q و تحلیل خوشه‌ای

این جدول مقادیر چهار پاسخگو در سه متغیر را نشان می‌دهد. تحلیل عاملی نوع Q این چهار پاسخگو را در دو گروه با ساختارهای واریانس مشابه قرار داده است. این دو گروه، A و C در مقابل B و D خواهد بود. در مقابل تکنیک تحلیل خوشه‌ای بر فاصله بین مقادیر پاسخها حساسیت



می‌دهد و افراد را براساس نزدیک‌ترین پاسخها طبقه‌بندی می‌کند. بنابراین در این روش پاسخگویان A و B در یک گروه و C و D در گروه دیگر قرار می‌گیرند.

#### تحلیل عاملی مشترک و تحلیل مؤلفه‌های اصلی:

در تحلیل عاملی، دو مدل اساسی وجود دارد که محقق می‌تواند از آنها استفاده کند. این مدلها عبارتند از تحلیل عاملی مشترک و تحلیل مؤلفه‌های اصلی. برای انتخاب مناسبترین مدل، محقق باید یک نکته را در مورد انواع واریانس‌ها مورد توجه قرار دهد. با توجه به هدف تحلیل عاملی مجموعاً سه نوع واریانس وجود دارد.

۱- واریانس مشترک<sup>(۱)</sup>

۲- واریانس خاص<sup>(۲)</sup>

۳- واریانس خطا<sup>(۳)</sup>

انواع واریانس و رابطه آنها با یک مدل عاملی در ذیل با مثال روشن شده است. واریانس مشترک، آن بخش از واریانس است که با سایر متغیرهای لحاظ شده در تحلیل سهم می‌باشد. واریانس خاص واریانسی است که تنها به یک متغیر خاص مربوط می‌شود. واریانس خطا ناشی از بی‌اعتباری و ناپایداری در داده‌های جمع‌آوری شده و یا شانس و تصادف در اندازه‌گیری پدیده‌هاست. زمانیکه از تحلیل مؤلفه‌های اصلی استفاده می‌شود، واریانس کل مدنظر قرار می‌گیرد و عاملهای دو رگه‌ای استنتاج می‌شود که سهم کوچکی از واریانس خاص و واریانس خطا را شامل می‌شود، اما اینها در حدی نیستند که ساختار عاملی کل را تحت تأثیر قرار داده و آن را تحریف یا منحرف کند.

نکته دیگر اینکه، در تحلیل مؤلفه‌های اصلی، مقادیر قطرهای ماتریس همبستگی ۱

1- Common Variance

2- Specific Variance

3- Error Variance

می‌باشند. برعکس در تحلیل عاملی مشترک<sup>(۱)</sup>، مقادیر مشترکات<sup>(۲)</sup> در قطر ماتریس قرار می‌گیرد و عاملها تنها براساس واریانس مشترک<sup>(۳)</sup> استنتاج می‌گردند. از نقطه نظر واریانس، تفاوت زیادی بین قرار دادن مقادیر ۱ در قطر ماتریس و قرار دادن مقادیر مشترکات در آن وجود دارد. با قرار دادن مقدار ۱ در قطر ماتریس، همانطوریکه در شکل (۳-۱۴) نیز دیده می‌شود واریانس کل به ماتریس عاملی<sup>(۴)</sup> وارد می‌گردد، اما تحلیل عاملی مشترک مقادیر مشترک را در قطر ماتریس قرار می‌دهد و در نتیجه عوامل تنها براساس واریانس مشترک محاسبه می‌گردد.

واریانس استخراج شده

--

واریانس از دست رفته

--

۱ در قطر ماتریس

واریانس کل
------------

(تحلیل مؤلفه‌های اصلی)

مقادیر مشترکات در قطر ماتریس

واریانس مشترک	واریانس خاص و واریانس خطا
---------------	---------------------------

(تحلیل عاملی مشترک)

#### شکل (۳-۱۴) انواع واریانس وارد شده به ماتریس عاملی<sup>(۵)</sup>

هر دو این مدلها در سطح وسیع توسط محققان مورد استفاده قرار می‌گیرند. انتخاب هریک از

1- Common Factor Analysis

2- Communalities

3- Common Variance

4- Factor Matrix

5- Hair, Jr. Joseph and etal (1990) "Multivariate Data Analysis" Macmillan Publishing Company, New York.

این مدلها برای تحلیل داده‌ها به دو معیار مربوط می‌باشد.

۱- هدف محقق از استفاده از تحلیل عاملی

۲- میزان شناخت قبلی از واریانس موجود در متغیرها

زمانیکه، محقق درصدد پیش بینی و تعیین کمترین تعداد عاملهاست که قادر باشد بیشترین واریانس موجود در مقادیر اصلی را تبیین کند، و شناخت قبلی نیز وجود دارد که واریانس خاص و واریانس خطا سهم کمتری از کل واریانس را شامل می‌شود، در این صورت روش تحلیل مؤلفه‌های اصلی انتخاب مناسبی خواهد بود. در مقابل، اگر هدف اولیه، شناسایی ابعاد پنهان در متغیرهای اصلی باشد و محقق نیز شناخت کمتری از واریانس خطا و واریانس خاص دارد، علاقمند است که این نوع واریانس را حذف کند، مناسب‌ترین مدل در اینگونه مواقع تحلیل عاملی مشترک می‌باشد.

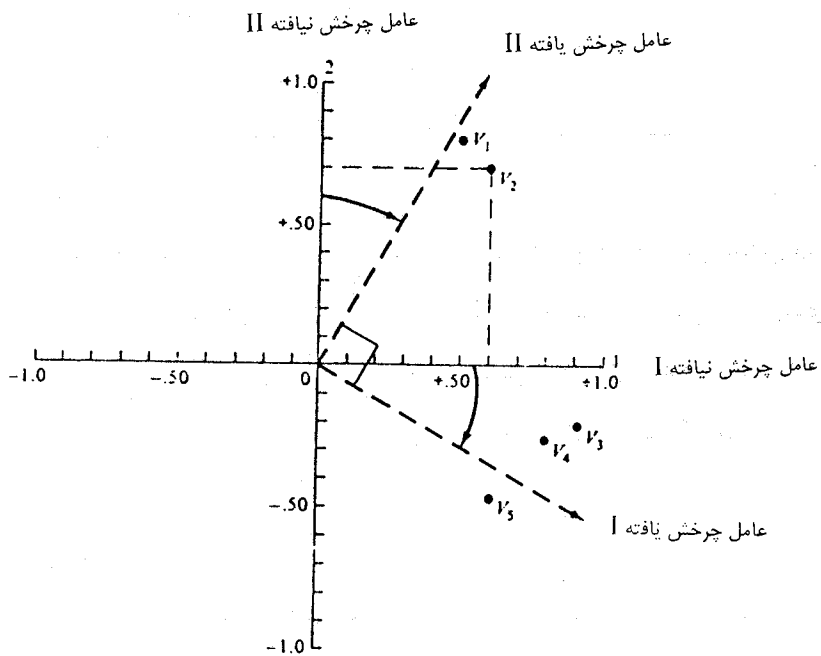
#### عاملهای چرخش یافته:

یکی از مفاهیم مهم در تحلیل عاملی، چرخش عاملهاست. این مفهوم دقیقاً به همان معنا دلالت دارد. در فرایند چرخش عاملی، محورهای مختصات عاملها به دور مبدأ، چرخش داده می‌شوند، تا اینکه موقعیت جدیدی کسب شود. در چرخش متعامد<sup>(۱)</sup> زاویه محورهای عمودی و افقی در حالت ۹۰ درجه باقی می‌ماند. اما این امکان نیز در چرخش عاملها وجود دارد که زاویه محورها در حالت ۹۰ درجه باقی نماند، که این نوع چرخش را چرخش متمایل<sup>(۲)</sup> می‌گویند. در اشکال (۴-۱۴) و (۵-۱۴) چرخش‌های عاملی متعامد و متمایل با استفاده از نمودار توضیح داده شده است.

همانطوریکه از شکل (۴-۱۴) برمی‌آید، برای استنتاج عاملهای نهایی، دو مرحله طی می‌شود. ابتدا ماتریس عاملی چرخش نیافته محاسبه می‌گردد تا نشانه مقدماتی برای استخراج تعداد عاملها بدست آید. در محاسبه ماتریس عاملی چرخش نیافته محقق درصدد است تا بهترین ترکیب خطی متغیرها بدست آید. منظور از بهترین ترکیب خطی، ترکیبی از متغیرهای اصلی است که

1- Orthogonal Rotation

2- Oblique Rotation



شکل (۴-۱۴) چرخش عاملی متعامد

بیشترین واریانس را در مجموعه داده‌ها، نسبت به هر نوع ترکیب خطی دیگر، تبیین کند. بنابراین ممکن است اولین عامل به عنوان تنها بهترین عاملی از همبستگی‌های خطی در بین داده‌ها باشد. دومین عامل، دومین بهترین ترکیب خطی از متغیرهاست که نسبت به عامل اول قرار گیرد، این عامل باید از بخش باقیمانده واریانس یعنی پس از استخراج عامل اول، استنتاج گردد. بنابراین عامل دوم ترکیبی خطی از متغیرهاست که در آن اثر اولین عامل حذف گردیده و به تبیین بخش عمده‌ای از باقیمانده واریانس می‌پردازد. به همین ترتیب سایر عاملها نیز استخراج شده تا اینکه کل واریانس داده‌ها تبیین شود.

ماتریس عاملی چرخش نیافته، هدف تلخیص داده‌ها را تحقق می‌بخشد. اما محقق باید

سؤال کند که آیا ماتریس عاملی چرخش نیافته اطلاعات لازم را برای تفسیر و تحلیل متغیرهای مورد مطالعه ارائه می‌دهد یا خیر؟ البته در بسیاری از موارد پاسخ این سؤال منفی است. بنابراین دلیل اصلی برای چرخش داده‌ها، دستیابی به ماتریس عاملی ساده و از نظر تئوریک معنی‌دارتر و قابل تفسیرتر است. در بسیاری از موارد چرخش عاملی از حضور برخی از متغیرها که همزمان در بسیاری از عاملها حضور می‌یابند جلوگیری می‌کند<sup>(۱)</sup>.

ماتریس عاملی چرخش نیافته در برخی موارد ممکن است الگوی معنی‌داری از متغیرها را ارائه دهد و در برخی مواقع نیز ممکن است چنین الگویی بدست نیاید. اگر عاملهای دوران نیافته الگوی قابل تفسیری از متغیرها ارائه دهد، ممکن است محقق تصمیم بگیرد که نیازی به چرخش عاملی وجود ندارد. اما از آنجا که تعیین این موضوع که آیا عاملهای چرخش نیافته معنی‌دارترند یا عاملهای چرخش یافته، مستلزم محاسبه عاملهای چرخش یافته می‌باشد، بنابراین، بطور عمومی لازم است تا در فرآیند تحلیل عاملی، ماتریس عاملی دوران یافته نیز محاسبه شود. زیرا این ماتریس، ساختار عاملی ساده‌تر و معنی‌دارتری برای تحلیل داده‌ها ارائه می‌دهد.

زمانیکه چرخش عاملی انجام می‌گیرد، مقادیر واریانس در بین عاملهای اول و عاملهای بعدی مجدداً توزیع می‌گردد. به همین دلیل اگرچه کل واریانس تبیین شده توسط عاملها برای ماتریس دوران نیافته و ماتریس دوران یافته ثابت باقی می‌ماند، اما واریانس تبیین شده توسط تک عاملها (یعنی مقادیر خاص) بدلیل توزیع مجدد واریانس در بین عاملها تغییر می‌یابد. این موضوع از طریق مقایسه مقادیر خاص هر یک از عاملها در ماتریس دوران نیافته با مقادیر خاص هر یک از عاملها در ماتریس دوران یافته کاملاً مشخص است.

همانطوریکه قبلاً نیز بیان گردید، ماتریس عاملی دوران نیافته، عاملها را به ترتیب اهمیت آنها استخراج می‌کند. بنابراین اولین عامل به یک عامل عمومی گرایش دارد که معمولاً اکثر بارهای

1- Dennis child (1973) "The Essentials of Factor Analysis" Holt Rinehart and Winston. New York. P. 47.

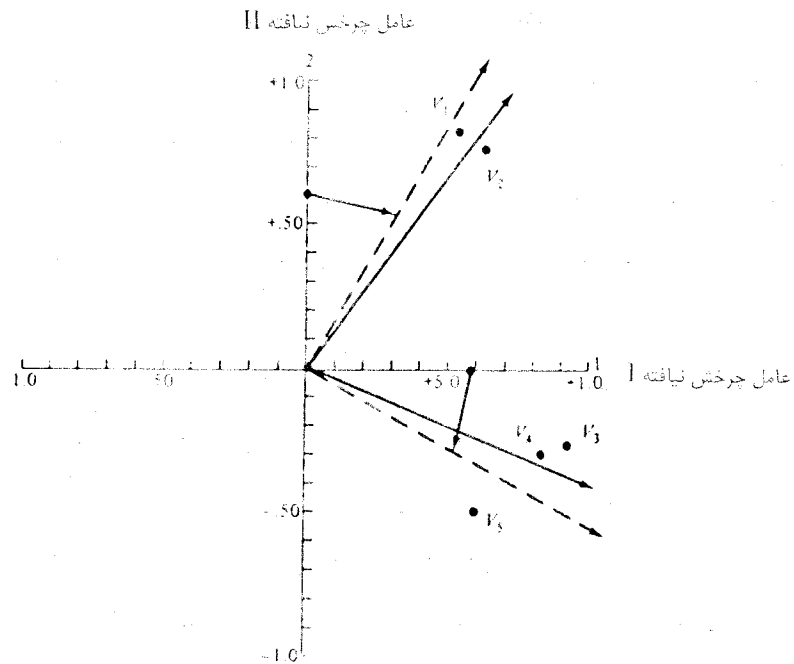
عاملی آن معنی‌دار هستند و این عامل بیشترین میزان واریانس را تبیین می‌کند. عامل دوم و عاملهای بعدی براساس مقدار واریانس باقیمانده انجام می‌گیرد، که هر یک از آنها سهم کمتری از واریانس را تبیین می‌کنند. اثر نهایی چرخش دادن ماتریس عاملی، توزیع مجدد واریانس از عاملهای اولی به عاملهای بعدی برای دستیابی به الگوی عاملی ساده‌تر و معنی‌دارتر است.

شکل (۱۴-۴) که در آن پنج متغیر در یک نمودار دوبعدی نمایش داده شده است، مفهوم چرخش عاملی را بطور واضح بیان می‌کند. محور عمودی، عامل چرخش نیافته II و محور افقی، عامل چرخش نیافته I را نشان می‌دهد. هر یک از این محورها از صفر تا +۱ یا -۱ مدرج شده‌اند. این مقادیر بیانگر بارهای عاملی<sup>(۱)</sup> می‌باشند. متغیرها نیز با علامت  $V_1$   $V_2$   $V_3$   $V_4$  و  $V_5$  نشان داده شده‌اند.

بار عاملی متغیر ۲ از طریق رسم یک خط نقطه چین از نقطه  $V_2$  به سمت محور عمودی عامل II بدست می‌آید. جایی که این خط محور عمودی متغیر ۲ چرخش نیافته را قطع می‌کند، بار عاملی متغیر ۲ می‌باشد. همچنین از طریق رسم یک خط نقطه چین از نقطه  $V_2$  به سمت محور افقی عامل I، بار عاملی متغیر ۱ بر روی عامل I بدست می‌آید. بارهای عاملی برای روشهای چرخش نیافته و چرخش یافته به منظور مقایسه نتایج این دو روش در جدول (۱۴-۲) ارائه شده است.

جدول (۱۴-۲) مقایسه بارهای عاملی دوران یافته و دوران نیافته

متغیر	بارهای عاملی دوران نیافته		بارهای عاملی دوران یافته	
	I	II	I	II
$V_1$	۰/۵۰	۰/۸۰	۰/۰۳	۰/۹۴
$V_2$	۰/۶۰	۰/۷۰	۰/۱۶	۰/۹۰
$V_3$	۰/۹۰	-۰/۲۵	۰/۹۵	۰/۲۴
$V_4$	۰/۸۰	-۰/۳۰	۰/۸۴	۰/۱۵
$V_5$	۰/۶۰	-۰/۵۰	۰/۷۶	-۰/۱۳



شکل (۵-۱۴) چرخش عاملی متمایل

همانطوریکه از جدول (۲-۱۴) برمی آید بار عاملی مربوط به عامل I در کلیه متغیرها مقدار بالایی را نشان می دهد. در عامل II دوران نیافته مقادیر بار عاملی برای متغیرهای ۱ و ۲ مثبت و بالا بوده، در حالیکه مقادیر بار عاملی برای متغیر ۳ در حد متوسط و برای متغیرهای ۲ و ۴ در حد پایین و بصورت منفی می باشد.

با نگاه به شکل (۵-۱۴) کاملاً مشخص است که در آن دو گروه مجزا از متغیرها وجود دارند، بطوریکه متغیرهای ۱ و ۲ در کنار یکدیگر در یک گروه و متغیرهای ۳ و ۴ و ۵ در گروه دیگر قرار دارند. در حالیکه این الگوی گروه بندی متغیرها از طریق مقادیر بارهای عاملی دوران نیافته (در جدول ۲-۱۴) به راحتی قابل استنتاج نیستند. همانطوریکه از شکل (۴-۱۴) نیز برمی آید با چرخش محورهای اصلی نمودار، در جهت عقربه های ساعت، الگوی بار عاملی کاملاً متفاوتی بدست می آید.

توجه کنید که در چرخش دادن عاملها، محورها باید همچنان در حالت ۹۰ درجه باقی بمانند. این بدین معناست که از نظر ریاضی عاملها مستقل از هم بوده و چرخش نیز بصورت قائم انجام می‌گیرد. پس از چرخش دادن محورها، متغیرهای ۳، ۴ و ۵ بار عاملی بالایی بر عامل I و متغیرهای ۱ و ۲ بار عاملی بالایی بر عامل II بدست می‌آورند. بنابراین طبقه‌بندی این متغیرها به دو گروه، پس از چرخش دادن بسیار شفاف و روشن انجام می‌گیرد، حتی پس از چرخش، موقعیت متغیر بدون تغییر باقی می‌ماند<sup>(۱)</sup>.

اصول حاکم بر روش چرخشی متعامد در مورد روش چرخش متمایل نیز صادق است. از آنجا که در روش چرخش متمایل ضرورتی ندارد که محورهای عاملها بصورت قائم باشند بنابراین، این روش از انعطاف بیشتری برخوردار است. همچنین این روش واقعی‌تر بنظر می‌آید، زیرا در این مدل فرض نمی‌شود که ابعاد مورد نظر از لحاظ تئوریک با یکدیگر همبستگی ندارند. در شکل (۱۴-۵) دو روش چرخش با همدیگر مقایسه شده‌اند. توجه کنید که چرخش عاملی متمایل، گروه‌بندی متغیرها را با دقت بیشتری نشان می‌دهد. این موضوع به این دلیل است که هر دو محور عاملی چرخش یافته به متغیرهای گروه مورد نظر نزدیک‌تر می‌باشند. همچنین روش چرخش متمایل اطلاعاتی در مورد اینکه عاملها تا چه حد بطور واقعی با یکدیگر همبستگی دارند ارائه می‌دهد.

بسیاری از آماردانان و محققان معتقدند که در بسیاری از موارد عاملهای دوران نیافته کافی نیستند. زیرا چرخش عاملها از ورود برخی از متغیرها در عاملهای مختلف جلوگیری کرده و ساختار عاملی مناسب و قابل تفسیرتری ارائه می‌دهد محصول نهایی هر دو روش چرخش متعامد و متمایل، دستیابی به یک ساختار عاملی ساده و از نظر تئوریک معنی‌دار است. اما رهیافت چرخش متعامد در سطح وسیع‌تر مورد استفاده محققان و آماردانان قرار می‌گیرد و شاید دلیل آن این است که این روش در کلیه برنامه‌های رایانه‌ای در نظر گرفته شده است. در حالیکه تنها تعداد کمی از نرم افزارهای

1- Hair Joseph and etal (1990) "Multivariate Data Analysis" Macmillan Publishing Company New York.



آماري روش متمایل را در خود جای داده‌اند. از سوی دیگر دلیل استفاده زیاد از روش چرخش متعامد این است که روشهای آنالیز برای بدست آوردن چرخش متمایل بطور مناسب توسعه داده نشده‌اند. همچنین اگر هدف محقق از تحلیل عاملی، بدست آوردن نتایج عاملها و استفاده از آن برای تحلیل آماری بعدی بوسیله سایر روشهای آماری باشد، در این صورت باید از تکنیک چرخش متعامد استفاده شود، و آن به این دلیل است که عاملها متعامد خواهند بود و بنابراین مسئله هم خطی<sup>(۱)</sup> از بین می‌رود. البته اگر هدف محقق بدست آوردن ساختار تئوریک معنی‌دار برای عاملها باشد روش چرخش متمایل مناسب‌تر خواهد بود، زیرا این روش هم از نظر تئوریک و هم از نظر تجربی به واقعیت نزدیک‌تر است<sup>(۲)</sup>.

آنچه که مسلم است، هدف همه روشهای چرخش عاملی ساده سازی سطرها یا ستونها و یا هر دو آنها در ماتریس عاملی برای تسهیل در تفسیر می‌باشد. هدف از ساده سازی سطرها، نزدیک کردن داده‌های هر سطر در حدامکان به صفر می‌باشد و هدف از ساده سازی ستونها نیز، نزدیک کردن مقادیر آن در حدامکان به صفر می‌باشد. سه رهیافت عمده متعامد برای تحقق این اهداف فرموله شده است. این روشها عبارتند از:

۱- کوار تیماکس<sup>(۳)</sup>

۲- واریماکس<sup>(۴)</sup>

۳- اکوایماکس<sup>(۵)</sup>

هدف نهایی چرخش کوار تیماکس ساده سازی سطرهای ماتریس عاملی است. بنابراین،

1- Collinearity

2- Fruchter, Benjamin (1964) "Introduction to factor Analysis" Princeton, Inc. New York. P. 83.

3- Quartimax

4- Varimax

5- Equimax

روش فوق به شیوه‌ای به چرخش ماتریس اولیه<sup>(۱)</sup> می‌پردازد که یک متغیر تاحد ممکن بیشترین بار را بر یک عامل داشته و کمترین بار را بر سایر عاملها داشته باشد. در مقابل، روش واریماکس سعی در ساده سازی ستونهای ماتریس عاملی دارد. توجه کنید که در رهیافت کوارتیماکس، بسیاری از متغیرها می‌توانند بار عاملی بالا یا نسبتاً بالایی بر یک عامل داشته باشند، زیرا این روش درصدد ساده سازی و مختصر کردن سطرهای ماتریس عاملی است. در روش واریماکس بیشترین سطح ساده سازی و مختصر کردن زمانی تحقق می‌یابد که در آن مقادیر یک ستون دارای کدهای ۰ و ۱ باشد. رهیافت اکوایماکس یک مصالحه و توافق بین معیارهای کوارتیماکس و واریماکس می‌باشد. در این روش بجای اینکه تنها سطرها و یا ستونها ساده سازی شوند سعی می‌گردد که بخشی از هر کدام از آنها انجام گیرد.

روش کوارتیماکس در ایجاد یک ساختار عاملی ساده‌تر موفق نمی‌باشد. مشکل این روش در آن است که سعی می‌کند در فرآیند چرخش، یک عامل عمومی بدست آورد. در حالیکه هدف اصلی چرخش ماتریس عاملی، دستیابی به یک عامل عمومی وسیع نمی‌باشد.

در مقابل روش واریماکس به دنبال مختصر کردن و ساده سازی ستونهای ماتریس عاملی است. به همین دلیل این روش درصدد حداکثر کردن مجموع واریانس بارهای ماتریس عاملی است. توجه کنید که همانطوریکه قبلاً نیز بیان شد در روش کوارتیماکس بسیاری از متغیرها می‌توانند بار بالا یا نسبتاً بالایی بر یک عامل داشته باشند زیرا این روش درصدد مختصر کردن سطرهاست. در رهیافت چرخش واریماکس، سعی می‌شود برخی از متغیرها بار عاملی بسیار بالا (یعنی نزدیک به +۱ یا -۱) و برخی متغیرها بار عاملی بسیار پایین (یعنی نزدیک به صفر) در هر ستون از ماتریس داشته باشند. منطق حاکم این است که زمانیکه همبستگی بین متغیر و عامل، نزدیک به +۱ یا -۱ باشد تفسیر به ساده‌ترین شکل انجام می‌گیرد و می‌توان قضاوت کرد که ارتباط روشن و قابل قبولی بین متغیر و عامل وجود دارد و اگر این همبستگی به صفر نزدیک باشد به راحتی می‌توان در خصوص عدم

وجود رابطه قضاوت نمود<sup>(۱)</sup>.

هرچند که راه حل کواریماکس از نظر محاسبات جبری ساده‌تر از راه حل واریماکس می‌باشد، اما واریماکس جداسازی عاملها را بطور شفاف‌تر و واضح‌تر نشان می‌دهد. بطور کلی کیسر<sup>(۲)</sup> تأکید می‌کند که تحلیل عاملی انجام یافته بوسیله چرخش واریماکس زمانیکه زیر مجموعه‌های مختلفی از متغیرها مورد تجزیه و تحلیل قرار می‌گیرند، پایدارتر و استوارتر از زمانی است که از روش کواریماکس استفاده می‌شود.

ثابت شده است که روش واریماکس به عنوان یک رهیافت تحلیلی در انجام چرخش عاملی متعادل موفق‌تر از سایر روشها بوده است. البته هنوز قاعده خاصی ارائه نشده است تا محققان را در انتخاب نوع تکنیک چرخش متعادل راهنمایی کند. در بسیاری از مواقع، محققان بدون دلیل خاصی از تکنیک‌های چرخشی که در برنامه‌های رایانه‌ای بطور استاندارد کار گذاشته شده‌اند استفاده می‌کنند. برخی از این نرم افزارها تنها قادر به انجام روش واریماکس هستند. به همین دلیل، این روش در سطح وسیع‌تر مورد استفاده قرار می‌گیرد. البته هنوز استدلال خاصی بر برتری یک روش نسبت به روشهای دیگر وجود ندارد. به همین دلیل انتخاب هر یک از این روشها باید متناسب با هدف تحقیق انجام گیرد.

#### معیارهای استخراج تعداد عاملها:

یکی از موارد مهم در تحلیل عاملی تعیین تعداد عاملهای قابل استخراج است. بطور معمول به تعداد متغیرهایی که به تحلیل وارد می‌شود، می‌توان عامل استخراج کرد. اما عاملهای آخری معمولاً سهم بسیار کمتری در تبیین موضوع دارند. بنابراین در این مرحله مهمترین مسئله تعیین معیار برای استخراج تعداد عاملهاست. اگر چه مبنای کمی دقیقی برای تصمیم‌گیری در مورد تعداد

1- Hair, Jr. Joseph and etal (1990) "Multivariate Data Analysis" Macmillan Publishing Company.

New York.

2- Kaiser

عاملهای استخراجی ارائه نشده است، اما ضوابطی وجود دارد که از آنها در تصمیم‌گیری برای تعیین تعداد عاملهای استخراجی استفاده می‌شود که این معیارها عبارتند از:

### ۱ - معیار مقدار ویژه<sup>(۱)</sup>:

یکی از ضوابط پرکاربرد در تعیین تعداد عاملها، مقدار ویژه است که آن را معیار ریشه را کد<sup>(۲)</sup> نیز می‌گویند. این معیار ساده‌ترین روش برای تعیین تعداد عاملهاست، اما بستگی به این دارد که کدامیک از مدل‌های تحلیل عاملی، یعنی تحلیل مؤلفه‌های اصلی<sup>(۳)</sup> و یا تحلیل عاملی مشترک<sup>(۴)</sup> مورد استفاده قرار گیرد. توجه کنید که در تحلیل مؤلفه‌های اصلی مقدار ۱ در قطر ماتریس همبستگی قرار می‌گیرد و واریانس کل در تحلیل لحاظ می‌گردد. در تحلیل مؤلفه‌های اصلی تنها عاملهایی که مقدار ویژه آنها بیشتر از ۱ باشد به عنوان عاملهای معنی‌دار در نظر گرفته می‌شود و کلیه عاملهایی که مقدار ویژه آنها کمتر از ۱ است به عنوان عاملهایی که از نظر آماری معنی‌دار نیستند از تحلیل کنار گذاشته می‌شوند.

بسیاری از محققان از جمله کیسر<sup>(۵)</sup> مقدار ویژه ۱ را مبنای تعیین تعداد عاملها قرار می‌دهند. اما زمانی که روش تحلیل عاملی مشترک، به عنوان مدل تحلیل انتخاب گردد، معیار مقدار ویژه ۱، باید قدری تعدیل شود، عاملهایی که مقدار ویژه آنها قدری کمتر از ۱ باشند نیز انتخاب می‌گردند. منطقی که در ورای معیار مقدار ویژه در تعیین تعداد عاملها وجود دارد این است که، اگر قرار است هر عاملی که برای تفسیر مورد استفاده قرار گیرد واریانس یک متغیر را در برگیرد، در این صورت رهیافت مقدار ویژه زمانی که تعداد متغیرها بین ۲۰ تا ۵۰ باشد قابل اعتماد بنظر می‌رسد، اما اگر تعداد متغیرها کمتر از ۲۰ باشد استفاده از این روش باید قدری با محافظه کاری انجام گیرد. چنانچه تعداد

1- Eigenvalue Criterion

2- Latent Root Criterion

3- Principal Component Analysis

4- Common Facto Analysis

5- Kaiser

متغیرها بیش از ۵۰ باشد در این صورت تعداد زیادی عامل استخراج می‌گردد که تاحدی غیر معمول خواهد بود.<sup>(۱)</sup>

### ۲- معیار پیشین:

معیار پیشین<sup>(۲)</sup> تحت شرایط خاصی مبنای قابل قبولی بنظر می‌رسد. براساس این معیار قبل از انجام تحلیل عاملی، محقق تاحدودی می‌داند که چند عامل ممکن است استخراج گردد. این رهیافت زمانی مفید است که محقق درصد آزمون تئوری یا فرضیه‌ای در مورد تعداد عاملهای قابل استخراج باشد. در استفاده از این معیار ممکن است محقق عواملی که در تحقیق مدنظر می‌باشد را در قالب گویه‌هایی اندازه‌گیری کرده و سپس با استفاده از تحلیل عاملی بخواهد که عوامل مورد نظر را که از قبل نیز تاحدودی برای محقق قابل پیش‌بینی بوده است استخراج نماید.

### ۳- معیار درصد واریانس:

معیار درصد واریانس تبیین شده توسط عاملها نیز ضابطه دیگری برای تعیین تعداد عاملهاست. در این رهیافت درصد تجمعی واریانس، مبنای تصمیم‌گیری قرار می‌گیرد. در تحقیقات غیرعلوم انسانی، فرآیند استخراج عوامل باید تا زمانی ادامه یابد که این عاملها بتوانند حداقل ۹۵ درصد واریانس داده‌ها را تبیین نمایند. اما در تحقیقات اجتماعی انسانی تبیین ۶۰ درصد واریانس نیز کفایت می‌کند.

1- Hair, Jr. Joseph and et al. (1990) "Multivariate Data Analysis" Macmillan Publishing Company, New York. p. 247.

2- A Prior Criterion

۴ - معیار تست بریدگی<sup>(۱)</sup>:

قبل از توضیح این معیار، لازم است تأکید شود که در روش تحلیل مؤلفه‌های اصلی، عاملهای استخراج شده بعدی، هم دارای واریانس مشترک<sup>(۲)</sup> و هم واریانس خاص<sup>(۳)</sup> است. هرچند که تمام عاملها حداقل دارای مقداری واریانس خاص هستند اما سهم آن در عاملهای بعدی بطور اسمی بیشتر از عاملهای قبلی است. با این مقدمه، معیار تست بریدگی، تعداد مناسب عاملها را بر مبنایی تعیین می‌کند که هنوز میزان واریانس خاص بر واریانس مشترک غلبه نکرده است. بنابراین تا زمانی که مقدار واریانس مشترک بیشتر از مقدار واریانس خاص باشد، عاملهای فوق به عنوان عاملهای معنی دار استخراج و مورد تجزیه و تحلیل قرار می‌گیرد<sup>(۴)</sup>.

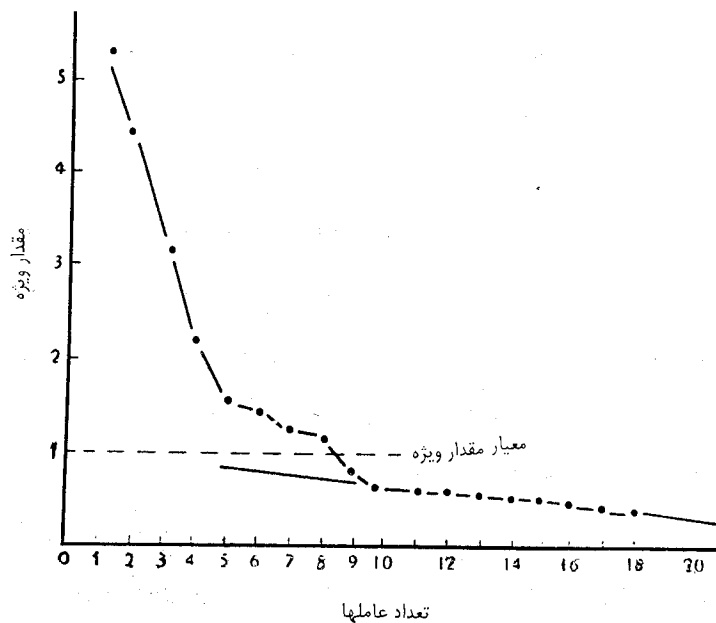
معیار بریدگی از طریق رسم نمودار مقدار ویژه در برابر تعداد عاملهایی که به ترتیب استخراج شده‌اند بدست می‌آید و از طریق حالت یا شکل نمودار، نقطه تعیین تعداد عاملها مشخص می‌شود. شکل (۱۴-۶) ۱۸ عامل استخراج شده در یک مطالعه را نشان می‌دهد.

1- Scree Test Criterion

2- Common Variance

3- Specific Variance

4- Cattell. R. B. (1966) "The Scree test for the number of factors"; *Multivariate Behavioral Research*, Vol. 1, April, pp 245-276.



شکل (۱۴-۶) معیار تست بریدگی برای تعیین تعداد عاملها

همانطوریکه از شکل (۱۴-۶) برمی آید با اولین عامل، نمودار با شیب نسبتاً تندی شروع شده و در نقطه‌ای تقریباً حالت افقی بخود گرفته است. براساس معیار بریدگی، این نقطه می‌تواند معیار استخراج حداکثر تعداد عاملها باشد. براساس این نمودار ۱۰ عامل اول از نظر آماری معنی‌دار بوده و می‌توانند برای تحلیل و تفسیر مورد استفاده قرار گیرند، اما از عامل ۱۰ به بعد مقدار واریانس خاص بر واریانس مشترک پیشی گرفته و این عاملها قابل قبول نمی‌باشند. توجه کنید که در این مثال اگر معیار مقدار ویژه مدنظر باشد تنها ۸ عامل اول باید استخراج گردد، اما براساس معیار بریدگی، دو عامل بعدی (یعنی عاملهای ۹ و ۱۰) نیز قابل استخراج می‌باشند. به عنوان یک قاعده کلی، معیار بریدگی معمولاً در مقایسه با معیار مقدار ویژه دو یا سه عامل بیشتر استخراج می‌کند<sup>(۱)</sup>.

1- Cattell, R. B. (1966) "The Scree test for the number of factor"; *Multivariate Behavioral Research*, Vol. 1, April, pp 245-276.

در عمل، بسیاری از محققان بندرت تنها یک معیار را مبنای تعیین تعداد عاملهای قابل استخراج قرار می‌دهند. در تحلیل عاملی، معیار مقدار ویژه به عنوان اولین راهنما برای تعیین عاملها مورد استفاده قرار می‌گیرد. سپس چند چرخش آزمایشی محاسبه شده و بر مبنای مقایسه آنها مناسب‌ترین معیار برای تعیین عاملها مشخص می‌شود.

بطور کلی نتایج تحلیل عاملی زمانی مناسب خواهد بود که عاملهای استخراج شده میزان قابل قبولی از مجموع واریانس را تبیین نماید. هرچه میزان کل واریانس تبیین شده بیشتر باشد نتایج بدست آمده مناسبتر خواهد بود. برای تحقق این هدف باید تعداد کافی از عاملها استخراج گردند تا واریانس کل قابل قبول بدست آید. این مقدار عمدتاً براساس مقدار ویژه که معمولاً باید بیشتر از یک باشد مشخص می‌شود.

در مواقعی نیز اتفاق می‌افتد که تعداد بیشتری از عاملها استخراج می‌گردند که مقدار ویژه همه آنها از یک بیشتر است. اما ممکن است محقق درصدد باشد تا با حذف برخی از آنها به یک مدل تحلیلی ساده‌تر دست یابد. راههای مختلفی برای حذف برخی از عوامل و خلاصه کردن متغیرها در تعداد محدودی از عاملها وجود دارد. برای این کار محقق ممکن است که در محاسبات خود در برنامه SPSS تعداد عاملهای استخراج شده را از قبل تعیین کند و از این طریق به استخراج تعدادی مشخص از عاملها و ساده سازی مدل بپردازد. از سوی دیگر با توجه به اینکه درصد تجمعی واریانس تبیین شده توسط هر عامل در جدول مقدار ویژه مشخص می‌شود و با افزایش هر عامل مقدار این درصد نیز افزایش می‌یابد، بنابراین محقق ممکن است با تعیین میزان واریانس کل تبیین شده که از نظر او قابل قبول می‌باشند بقیه عوامل را در مدل تحلیلی وارد نکند و از این طریق به ساده سازی مدل بپردازد. که البته هر دو این روشها با اشکالاتی مواجه می‌باشند که توصیه می‌شود، از این روشها برای تلخیص عاملها و ساده سازی مدل استفاده نگردد.

از آنجا که در تحقیقات، هدف دستیابی به عاملهایی است که بیشترین میزان واریانس را تبیین نماید، بنابراین باید سعی گردد به شیوه‌ای به تلخیص و محدود کردن عاملهای استخراج شده اقدام شود که کل واریانس تبیین شده نیز افزایش یابد. برای اینکار محقق باید سعی کند تا از طریق



حذف متغیرهای نامناسب که واریانس آنها توسط عاملهای اصلی تبیین نمی‌شوند، مقدار واریانس کل را افزایش دهد. برای تشخیص اینکه کدامیک از متغیرها در تحلیل مناسب نمی‌باشند و باید حذف گردند باید از آماره اشتراک<sup>(۱)</sup> استفاده کرد. مقدار اشتراک در واقع نسبت واریانس تبیین شده آن متغیر توسط ترکیبی از عاملهای استخراج شده است. برای بدست آوردن مقدار اشتراک هر متغیر باید مقدار همبستگی متغیر فوق با هر یک از عوامل را مجذور کرد و با همدیگر جمع نمود. مقدار اشتراک همواره بین صفر و یک در نوسان است. هرچه این مقدار به سمت یک میل کند، مجموعه عاملهای انتخاب شده به نحو بهتری واریانس آن متغیر را تبیین می‌کند و اگر مقدار اشتراک برای یک متغیر پایین باشد بیانگر این واقعیت است که واریانس آن متغیر توسط عاملهای انتخاب شده تبیین نشده است. بنابراین در صورتی که اینگونه متغیرها از تحلیل حذف گردند مقدار کل واریانس تبیین شده توسط عوامل فوق افزایش می‌یابند و حداقل عاملهای مقتضی نیز بدست می‌آید<sup>(۲)</sup>.

#### معیار تعیین سطح معنی‌داری بارهای عاملی:

برای تفسیر عاملها باید مشخص شود که کدامیک از بارهای عاملی باید به عنوان مقادیر معنی‌دار لحاظ گردند. برای اتخاذ این تصمیم معیارهای مختلفی وجود دارد که در ذیل به توضیح و تبیین آنها می‌پردازیم.

۱ - اولین معیاری که پیشنهاد شده است فاقد مبنای ریاضی است و عمدتاً براساس یک قاعده تجربی است که توسط محققان و آماردانانی که بطور مکرر از تحلیل عاملی استفاده کرده‌اند پیشنهاد گردیده است. براساس این معیار، بارهای عاملی که بزرگتر از  $0/30 \pm$  باشند معنی‌دار تلقی می‌شوند. بارهای عاملی که بزرگتر از  $0/40 \pm$  باشند دارای سطح معنی‌داری بالا و بارهایی که بزرگتر از  $0/50 \pm$  باشند بسیار معنی‌دار تلقی می‌شوند. بنابراین هرچه میزان بار عاملی بیشتر باشد سطح

#### 1- Commuality

۲- دواس، دی، ای (۱۳۷۶) "پیمایش در تحقیقات اجتماعی" ترجمه هوشنگ نایی، نشر نی، ص ۲۵۸-۲۵۹

معنی‌داری آنها نیز در تفسیر ماتریس عاملی افزایش می‌یابد. این معیار در مقایسه با سایر معیارها بیشتر مورد استفاده قرار می‌گیرد، اما زمانی استفاده از آن مناسب می‌باشد که تعداد نمونه بیشتر از ۵۰ باشد.

۲- همانطوریکه قبلاً نیز بیان شد، بار عاملی بیانگر همبستگی بین یک متغیر اصلی و عامل مربوط به آن می‌باشد. بنابراین، یکی از معیارهایی که می‌تواند برای معنی‌داری و تفسیر آن مینا قرار گیرد، همان معیاری است که برای تفسیر ضریب همبستگی مورد استفاده قرار می‌گیرد.

براساس این معیار، زمانیکه تعداد نمونه ۱۰۰ باشد بارهایی که حداقل  $\pm 0/19$  و  $\pm 0/26$  هستند به ترتیب در سطح ۵ و ۱ درصد معنی‌دار تلقی می‌شوند. زمانیکه حجم نمونه ۲۰۰ باشد، مقادیر  $\pm 0/14$  و  $\pm 0/18$  به ترتیب در سطح ۵ و ۱ درصد معنی‌دار لحاظ می‌گردند و بالاخره، اگر تعداد نمونه حداقل ۳۰۰ باشد، بارهایی که  $\pm 0/11$  و  $\pm 0/15$  هستند به ترتیب در سطح ۵ و ۱ درصد معنی‌دار در نظر گرفته می‌شوند. از آنجا که تعیین میزان خطا در تحلیل عاملی کار پیچیده‌ای است، بنابراین برای حصول اطمینان بیشتر در تفسیر عاملها، بهتر است سطح معنی‌داری ۱ درصد به عنوان مینا در نظر گرفته شود.

۳- یکی از اشکالاتی که در روشهای اول و دوم وجود دارد، این است که تعداد متغیرها و عاملی که مورد قضاوت و بررسی است مدنظر قرار نمی‌گیرد. ثابت شده است که با حرکت از یک عامل به عامل بعدی باید سطح پذیرش برای معنی‌دار بودن بارهای عاملی افزایش یابد. با قبول این واقعیت که واریانس خاص و واریانس خطا در عاملهای بعدی افزایش می‌یابند، بنابراین معیار سطح معنی‌داری در عاملهای بعدی در خصوص بارهای عاملی نیز باید افزایش یابد<sup>(۱)</sup>.

تعداد متغیرهایی که مورد تجزیه و تحلیل قرار می‌گیرند نیز در تعیین اینکه کدامیک از بارهای عاملی معنی‌دار هستند مهم می‌باشد. با افزایش تعداد متغیرها در تحلیل، سطح قابل قبول

1- Dillion. W. R. and M. Goldstein, (1984) "Multivariate Analysis: Methods and Applications", Wiley, New York, P. 256.

برای معنی‌داری بارها کاهش می‌یابد. تعدیل معیار بر اساس تعداد متغیر، بویژه زمانی صحیح است که از عامل اول به سمت عاملهای بعدی حرکت می‌کنیم. بطور خاص، زمانیکه حجم نمونه ۵۰ باشد و سطح معنی‌داری مناسب نیز ۰/۰۵ باشد اصول ذیل قابل اجراء می‌باشد: (۱) معیار معنی‌داری بار عاملی تا عامل پنجم با ۲۰ متغیر، ۰/۲۹۲± و با ۵۰ متغیر، ۰/۲۶۷± می‌باشد. (۲) معیار معنی‌داری بار عاملی برای عامل ششم تا دهم با ۲۰ متغیر، ۰/۳۵۳±، اما با ۵۰ متغیر فقط ۰/۲۷۴± خواهد بود. همین قاعده برای زمانیکه تعداد نمونه ۱۰۰ و سطح معنی‌داری قابل قبول ۰/۰۵ باشد بشرح ذیل خواهد بود: (۱) معیار معنی‌داری بار عاملی تا عامل پنجم با ۲۰ متغیر، ۰/۲۱۶± و با ۵۰ متغیر، ۰/۲۰۲± می‌باشد. (۲) معیار معنی‌داری بار عاملی برای عامل ششم تا دهم با ۲۰ متغیر، ۰/۲۶۱± و با ۵۰ متغیر، ۰/۲۱۴± خواهد بود<sup>(۱)</sup>.

بحث مربوط به معیار معنی‌داری بارهای عاملی را می‌توان بشرح ذیل جمع‌بندی کرد:

۱- هرچه حجم نمونه افزایش یابد، بارهای عاملی کوچکتر نیز به عنوان بارهای معنی‌دار لحاظ می‌گردند.

۲- هرچه تعداد متغیرهای مورد استفاده در تحلیل عاملی افزایش یابد، بارهای عاملی کوچکتر نیز معنی‌دار تلقی می‌گردند.

۳- هرچه تعداد عاملها بیشتر باشد، در عاملهای بعدی باید بارهای عاملی بزرگتر معنی‌دار تلقی شوند.

#### تفسیر ماتریس عاملی:

تفسیر روابط درونی ماتریس عاملی کار ساده‌ای نیست. اما از طریق دنبال کردن فرآیند ذیل می‌توان به ساده‌سازی و تفسیر روابط پرداخت.

1- Hair, Jr. Joseph and et al (1990) "Multivariate Data Analysis" Macmillam Publishing Company. New York. pp. 249-250.

۱- در ماتریس عاملی، هر ستون معرف یک عامل است. مقادیر موجود در هر ستون نشان دهنده بارهای عاملی هر متغیر بر یک عامل می‌باشد. در محاسبات رایانه‌ای عاملها به ترتیب از چپ به راست با شماره‌های ۱، ۲، ۳ و الی آخر قرار می‌گیرند. متغیرها نیز در ستون اول از بالا به پایین لیست می‌شوند. برای ساده‌سازی کار تحلیل ماتریس بهتر است در کنار شماره هر متغیر، نام متغیر نوشته شود.

۲- برای شروع تفسیر، محقق باید از اولین متغیر شروع کند و مقادیر مربوط به آن را در عاملهای مختلف بررسی نماید. هر جا که بیشترین مقدار مطلق بار عاملی وجود داشته باشد و از نظر آماری نیز معنی‌دار باشد زیر آن خط بکشد. به همین ترتیب مراحل باید برای متغیر دوم، سوم و الی آخر انجام گیرد. البته برای ساده کردن کار در این مرحله محقق می‌تواند با فشار دادن دکمه Option از پنجره اصلی تحلیل عاملی در SPSS سطح معنی‌داری بارهای عاملی را براساس یک منطق خاص تعیین نماید. در این صورت بارهای عاملی کمتر از عدد فوق از ماتریس عاملی حذف می‌گردند. به عنوان مثال اگر محقق مبنای معنی‌داری بار عاملی را ۰/۵۰ تعیین کند در این صورت تنها بارهای عاملی که بزرگتر از ۰/۵۰ باشند در ماتریس عاملی ظاهر می‌شوند. البته در برخی مواقع ممکن است یک متغیر بر بیش از یک عامل بار شده باشد، که این از موارد پیچیده و بغرنج در تحلیل عاملی است. اگرچه در بسیاری از موارد چرخش عاملها بخشی از اینگونه مشکلات را مرتفع می‌کند، اما در بعضی مواقع نیز اینگونه مشکلات، هنوز لاینحل باقی می‌مانند.

۳- زمانیکه با بررسی ماتریس عاملی، بارهای عاملی معنی‌دار مشخص شدند، باید متغیرهایی که بر روی هیچ یک از عاملها بار عاملی معنی‌دار ندارند نیز مشخص شوند. زمانیکه متغیرهایی وجود داشته باشند که بر روی هیچ عاملی بار نداشته باشند، محقق می‌تواند با اینگونه متغیرها به دو صورت بشرح ذیل برخورد کند. (۱) شیوه اول این است که اینگونه متغیرها را به فراموشی سپرده و تنها متغیرهای معنی‌دار را تفسیر کند و (۲) شیوه دوم این است که محقق با اینگونه متغیرها بطور نقادانه برخورد کند. در این روش می‌توان تأثیرات مجموعه متغیرها را بر روی نتایج

تحقیق و همچنین شاخص مشترک<sup>(۱)</sup> مدنظر قرار داد و بر این استدلال تأکید کرد که همه متغیرها سهمی حتی کوچک در اینگونه روابط و نتایج داشته‌اند. بنابراین برای رفع اثرات متغیرهایی که بار عاملی بالایی نداشته‌اند، آنها را از تحلیل حذف و مجدداً تحلیل عاملی را براساس متغیرهای معنی دار تکرار نمود و نتایج را مورد تجزیه و تحلیل قرار داد.

۴ - پس از اینکه کلیه متغیرهایی که در هر یک از عاملها دارای بار عاملی معنی‌دار هستند مشخص شد، باید با لحاظ کردن متغیرهای مربوط به هر عامل نام مناسبی برای عامل فوق تعیین کرد. انتساب نام برای عاملهای استخراج شده از طریق رایانه انجام نمی‌گیرد، بلکه این وظیفه‌ای است که بعهده محقق می‌باشد. برای تعیین نام برای عاملها، مقدار بار عاملی هر متغیر تعیین‌کننده می‌باشد، بدین معنی که نام عامل باید سنخیت بیشتری با متغیرهایی داشته باشد که از بار عاملی بالایی برخوردار هستند.

#### روش انتخاب متغیرهای جایگزین برای تحلیل‌های بعدی:

اگر هدف محقق انتخاب متغیرهای مناسب برای تحلیل‌های بعدی بوسیله سایر تکنیک‌های آماری باشد، می‌تواند از طریق بررسی ماتریس عاملی، متغیری که بیشترین بار عاملی را داراست به عنوان جایگزینی برای یک بعد عاملی خاص انتخاب کند. زمانیکه در هر عامل تنها یک متغیر وجود داشته باشد که بیشترین بار عاملی را داراست در این صورت انتخاب متغیر فوق به عنوان جانشین عامل مورد نظر، در تحلیل‌های بعدی کار ساده‌ای است. اما مشکل زمانی بروز می‌کند که در یک عامل بیش از یک متغیر با بار عاملی بالا و بسیار نزدیک بهم وجود داشته باشد. در این صورت محقق باید بارهای عاملی فوق را بطور نقادانه بررسی کند و نهایتاً یکی از آنها را به عنوان نماینده عامل فوق به منظور انعکاس بعد خاص انتخاب نماید. در این مرحله شناخت قبلی محقق می‌تواند مؤثر باشد. حتی ممکن است این شناخت محقق باعث شود که استدلال کند، مقادیر متغیری که بار

عاملی آن، قدری پایین‌تر است بهتر می‌تواند عامل فوق و بعد مورد نظر را نمایندگی کند و براساس همین منطق نیز متغیر فوق برای تحلیل بعدی انتخاب شود. توجه شود که متغیر جایگزین زمانی انتخاب می‌شود که چرخش متعامد<sup>(۱)</sup> انجام گرفته باشد. زیرا وقتی که محقق می‌خواهد متغیری را به عنوان جایگزین انتخاب کند، او فرض می‌کند که متغیرهای مستقل باید با یکدیگر همبسته نباشند و مستقل از هم عمل کنند و این زمانی تحقق می‌یابد که چرخش عاملی به روش متعامد انجام گرفته باشد.

### نحوه استفاده از مقادیر عاملی<sup>(۲)</sup>:

زمانیکه محقق بخواهد مجموعه‌ای جدید اما محدود از متغیرهای ترکیبی را انتخاب و جایگزین متغیرهای اصلی کند و در تحلیل‌های بعدی از آنها استفاده نماید، دست به محاسبه مقادیر عاملی می‌زند. مقادیر عاملی، سنجش ترکیبی است برای هر عامل، که نشان دهنده یک موضوع است. داده‌های خام اصلی برای محاسبه مقادیر عاملی برای هر فرد مورد استفاده قرار می‌گیرد. فردی که در برخی از متغیرهایی که بار عاملی بالایی در یک عامل دارد، نمره بالایی داشته باشد، در این صورت نمره عاملی بالایی در آن عامل بخود اختصاص خواهد داد. بسیاری از نرم‌افزارهای آماری، مقادیر عاملی هر فرد را برای استفاده در تجزیه و تحلیل‌های بعدی محاسبه می‌کنند، و استفاده کننده از رایانه تنها باید مورد مقادیر عاملی را انتخاب نماید تا محاسبات لازم انجام گیرد. محاسبه مقادیر عاملی خوشبختانه از طریق SPSS امکانپذیر است. بطور کلی محاسبه مقادیر عاملی کار پیچیده‌ای نیست، اما سؤال اساسی این است که آیا برای تحلیل‌های بعدی بهتر است محقق از مقادیر عاملی استفاده کند و یا از متغیرهای جایگزین؟ در پاسخ به این سؤال باید گفت که هر یک از آنها دارای مزایا و مضرات خاصی می‌باشند، و پاسخ قطعی وجود ندارد که کدامیک از آنها بر دیگری برتری دارد. مقادیر عاملی از این امتیاز برخوردار است که،

1- Orthogonal Rotation

2- Factor Scores

بیانگر ترکیبی از کلیه متغیرهایی است که بر یک عامل بار شده‌اند. در حالیکه متغیرهای جایگزین تنها نشان دهنده یک متغیر می‌باشند. اما از آنجا که مقادیر عاملی براساس همبستگیهای کلیه متغیرهایی که در یک عامل هستند محاسبه می‌گردد و این همبستگیها نیز احتمالاً بسیار کمتر از ۱ هستند بنابراین مقادیر عاملی تنها تخمینی از عاملها می‌باشند و به همین دلیل در واقع شاخصهایی خطا دار از عامل‌های مورد نظر هستند. بنابراین قاعده تصمیم‌گیری در مورد اینکه آیا باید مقادیر عاملی را برای تحلیل‌های آتی انتخاب کرد یا از متغیرهای جایگزین استفاده نمود، عمدتاً به وضعیت مقیاس مربوط می‌شود. اگر مقیاسهای طراحی شده مناسب باشند و ابزارای مطمئن و دارای روائی قابل قبول باشند، مقادیر عاملی می‌توانند بهترین آلترناتیو برای تحلیل‌های آتی باشند. اما اگر مقیاسهای ساخته شده، بدون پیش آزمون و با اعتبار و روائی غیرمطمئنی مورد استفاده قرار گرفته باشند در این صورت احتمالاً متغیرهای جایگزین مناسب‌ترین انتخاب در این زمینه خواهند بود.

#### محاسبه تحلیل عاملی با برنامه SPSS:

برای آشنایی با روش محاسبه تحلیل عاملی بهتر است از یک مثال واقعی استفاده گردد. فرض کنید درصد بررسی اثرات قطب صنعتی مبارکه اصفهان بر وضعیت کشاورزی این منطقه هستیم. برای این منظور تعداد ۳۰ متغیر انتخاب گردیدند که پس از بررسی آنها ۲۳ متغیر برای تحلیل عاملی مناسب تشخیص داده شد. این متغیرها در جدول (۱۴-۳) ارائه شده‌اند<sup>(۱)</sup>.

#### روش محاسبه با برنامه SPSS:

برای محاسبه تحلیل عاملی ابتدا متغیرهای مورد نظر را در ستونهای مختلف در صفحه گستر Data Editor وارد کرده و عناوین متغیرها را تعریف کنید. سپس دستور:

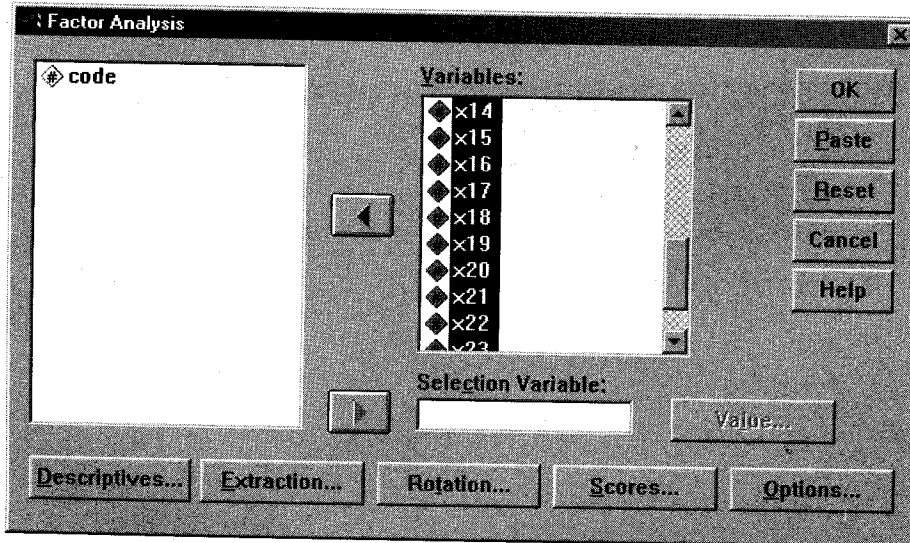
۱- این مثال از پایان نامه کارشناسی ارشد سرفراز حقیقی که در سال ۱۳۷۹ در دانشگاه تهران در رشته توسعه روستایی تحت راهنمایی نگارنده انجام گرفته اقتباس گردیده است.

Analyze/Data Reduction/Factor را اجراء نمائید تا شکل (۱۴-۷) ظاهر شود.

جدول (۱۴-۳) متغیرهای انتخاب شده برای تحلیل عاملی

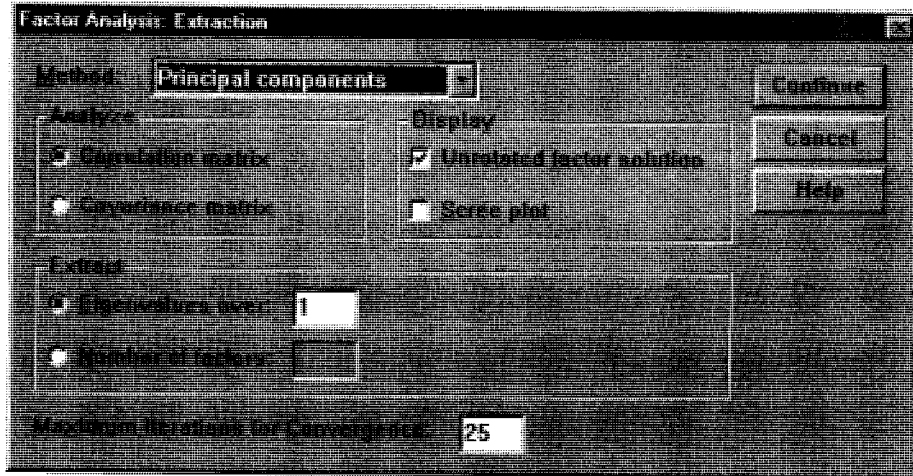
متغیر	کد متغیر
آلودگی آب و خاک کشاورزی	X <sub>1</sub>
تبدیل اراضی کشاورزی	X <sub>2</sub>
نابرابری توسعه بین روستاها	X <sub>3</sub>
کاهش سرمایه گذاری در بخش کشاورزی	X <sub>4</sub>
کاهش علاقه کشاورزان به کار کشاورزی	X <sub>5</sub>
کاهش سطح زیر کشت	X <sub>6</sub>
کاهش ارزش زراعی زمین‌های کشاورزی	X <sub>7</sub>
تشدید مهاجرت کشاورزان به شهر	X <sub>8</sub>
عدم رغبت جوانان روستایی به فعالیت کشاورزی	X <sub>9</sub>
تسهیل ارائه خدمات کشاورزی	X <sub>10</sub>
افزایش آگاهی کشاورزان	X <sub>11</sub>
ایجاد بازار مصرف مناسب	X <sub>12</sub>
افزایش تولیدات کشاورزی	X <sub>13</sub>
گسترش خدمات رفاهی و آموزشی	X <sub>14</sub>
افزایش فرصت‌های اشتغال	X <sub>15</sub>
افزایش قدرت خرید نهاده‌های کشاورزی	X <sub>16</sub>
استفاده بیشتر از نهاده‌های کشاورزی (کود، سم)	X <sub>17</sub>
افزایش درآمد روستائیان	X <sub>18</sub>
ایجاد شغل دوم	X <sub>19</sub>
واگذاری زمین کشاورزی	X <sub>20</sub>
فروش زمین کشاورزی	X <sub>21</sub>
تغییر الگوی کشت	X <sub>22</sub>
فاصله روستا تا صنایع منطقه	X <sub>23</sub>





شکل (۷-۱۴) پنجره Factor Analysis

در این پنجره متغیرهای مورد نظر را انتخاب کرده و به قسمت Variables منتقل نمائید. دکمه Descriptives را از پنجره اصلی فشار دهید و در پنجره مکالمه باز شده، مورد KMO and Bartlett's را انتخاب کنید. با فشار دادن دکمه Extraction پنجره دیگری مانند شکل (۸-۱۴) باز می‌شود.



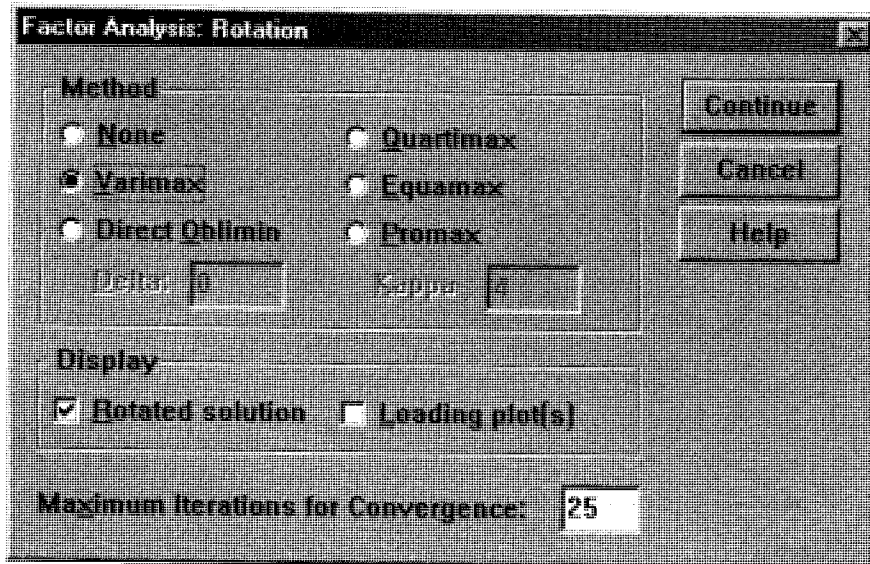
شکل (۸-۱۴) پنجره Factor Analysis: Extraction

در این پنجره از قسمت Method می‌توانید روشهای مختلف تحلیل عاملی را انتخاب کنید. روش تحلیل مولفه‌های اصلی به عنوان پرکاربردترین تکنیک بصورت پیش‌گزینه در این قسمت قرار دارد. در قسمت Extract بطور پیش‌گزینه مورد Eigenvalues over انتخاب شده و دکمه رادیویی آن روشن می‌باشد. در مقابل آن عدد ۱ نوشته شده است و آن بدین مفهوم است که به تعدادی عامل استخراج می‌گردد که مقدار ویژه<sup>(۱)</sup> آنها بزرگتر از ۱ می‌باشد. در صورتی که بخواهید تعداد عاملها را از قبل، خودتان انتخاب کنید، در این صورت می‌توانید دکمه رادیویی مورد Number of Factors را از همان قسمت انتخاب و در مقابل آن تعداد عاملهای مورد نظر را تایپ کرده و سپس دکمه Continue را فشار دهید تا به پنجره اصلی باز گردید.

برای تعیین نوع چرخش عاملی باید از قسمت Rotation استفاده کنید. با فشار دادن این

1- Eigenvalue

دکمه، پنجره‌ای مانند شکل (۹-۱۴) باز می‌شود.

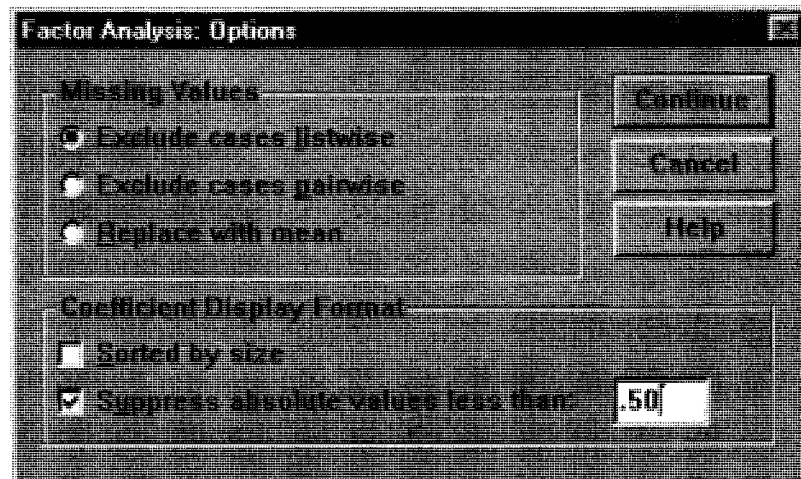


شکل (۹-۱۴) پنجره Factor Analysis: Rotation

در این پنجره امکان انتخاب روش‌های مختلف چرخش عاملی، مانند واریماکس، اکوایماکس، کواریماکس، و غیره وجود دارد. پس از انتخاب مورد مناسب دکمه Continue را فشار دهید. چنانچه بخواهید مقادیر عاملی<sup>(۱)</sup> را محاسبه و در تحلیل‌های بعدی نظیر رگرسیون یا تحلیل تشخیصی و غیره استفاده کنید، در این صورت دکمه Scores را از پنجره اصلی تحلیل عاملی فشار دهید و در پنجره باز شده مورد Save as variables را انتخاب کنید. توجه داشته باشد که نتایج این بخش در پنجره Data Editor ظاهر می‌شود، نه در پنجره Output. در صورتی که بخواهید بارهای عاملی استخراج کنید که مقدار آن بزرگتر از یک عدد خاص مانند ۰/۵۰ باشد، در این صورت دکمه Options را فشار دهید تا پنجره‌ای مانند شکل (۱۰-۱۴) در

1- Factor Scores

مانیتور ظاهر شود.



شکل (۱۴-۱۰) پنجره Factor Analysis: Options

در این پنجره بطور پیش‌گزیده مقدار  $0/10$  انتخاب شده است و بارهای عاملی بیشتر از آن استخراج می‌گردند. برای تغییر معیار فوق‌مورد: Suppress absolute values less than را انتخاب کرده و در مربع مقابل آن عدد  $0/50$  را تایپ و دکمه Continue را فشار دهید. با این انتخاب تنها بارهای عاملی بزرگتر از  $0/50$  در نتایج ظاهر می‌شوند و این کار تحلیل و تفسیر را برای محقق راحت‌تر می‌کند.

پس از انتخاب موارد فوق دکمه Ok را فشار دهید تا نتایج در پنجره Output ظاهر گردد. در صورتی که آزمون KMO و بارتلت نیز انتخاب شده باشند نتایج آن در اولین جدول در تحلیل عاملی ظاهر می‌شود. در این مثال، مقدار KMO برابر با  $0/80$  می‌باشد که بیانگر این واقعیت است که داده‌ها برای تحلیل عاملی مناسب بوده و نتیجه آزمون بارتلت نیز معنی‌دار می‌باشد.

جدول (۴-۱۴) استخراج مجموعه اولیه عامل‌ها

Variable	Communality	Factor	Eigenvalue	Pct of var	Cum Pct
X <sub>۱</sub>	1.000	1	6.205	24.8	24.8
X <sub>۲</sub>	1.000	2	3.390	13.6	38.4
X <sub>۱۴</sub>	1.000	3	1.808	7.2	45.6
X <sub>۴</sub>	1.000	4	1.520	6.1	51.7
X <sub>۵</sub>	1.000	5	1.380	5.5	57.2
X <sub>۶</sub>	1.000	6	1.096	4.4	61.6
X <sub>۷</sub>	1.000	7	1.043	4.2	65.9
X <sub>۳</sub>	1.000	8	1.000	4.2	69.8
X <sub>۱۳</sub>	1.000	9	.871	3.5	73.3
X <sub>۸</sub>	1.000	10	.789	3.2	76.4
X <sub>۱۱</sub>	1.000	11	.655	2.6	79.0
X <sub>۱۸</sub>	1.000	12	.612	2.5	81.5
X <sub>۱۱</sub>	1.000	13	.596	2.4	83.9
X <sub>۱۲</sub>	1.000	14	.542	2.2	86.0
X <sub>۱۵</sub>	1.000	15	.464	1.9	87.9
X <sub>۱۳</sub>	1.000	16	.451	1.8	89.7
X <sub>۱۶</sub>	1.000	17	.356	1.5	92.7
X <sub>۲۳</sub>	1.000	18	.341	1.4	94.0
X <sub>۲۱</sub>	1.000	19	.325	1.3	95.4
X <sub>۱۹</sub>	1.000	20	.305	1.2	96.6
X <sub>۲۰</sub>	1.000	21	.228	.9	98.5
X <sub>۱۷</sub>	1.000	22	.203	.9	99.3
X <sub>۱۹</sub>	1.000	23	.174	.7	100.0

جدول بعدی استخراج شده به روش تحلیل مؤلفه‌های اصلی، میزان اشتراک<sup>(۱)</sup>، را نشان می‌دهد. (این مقادیر در قطر ماتریس همبستگی قرار دارد و در روش تحلیل مؤلفه‌های اصلی برابر با یک می‌باشد)، مقادیر ویژه<sup>(۲)</sup>، درصد واریانس تبیین شده توسط هر عامل و درصد تراکمی نیز در جدول بعد ظاهر می‌شود، (جدول ۱۴-۴). در این جدول تعداد عاملها به اندازه تعداد متغیرهای استفاده شده می‌باشد که مقدار ویژه هر کدام نیز مشخص شده است<sup>(۳)</sup>. مجموع این عوامل کل واریانس متغیرها را تبیین می‌کند. اما عاملهای پایین تر معمولاً سهم ناچیزی در تبیین متغیرها دارند. به همین دلیل براساس ملاک کیسر تنها باید عاملهایی استخراج گردند که مقدار ویژه آنها بیشتر از یک باشد. اگرچه خود محقق نیز می‌تواند براساس چارچوب نظری و سهم واریانس تبیین شده، تعداد عامل‌های مورد نظر را تعیین کند.

جدول بعدی که معمولاً در تحلیل عاملی استخراج می‌گردد، ماتریس عاملی دوران نیافته است (جدول ۱۴-۵). در این جدول بار عاملی هر متغیر قبل دوران یافتن مشخص می‌شود. اما از آنجا که تمام این ضرایب در تعریف عاملها نقش ندارند، بنابراین ضرایب پایین تر باید از جدول حذف شوند. در مورد اینکه مبنای معنادار بودن این ضریب چیست نظرات متفاوتی وجود دارد. برای تفسیر این عاملها جونز<sup>(۴)</sup> (۱۹۵۴) کمترین مقدار این ضریب را برابر با ۰/۳۰، هومن<sup>(۵)</sup> (۱۹۸۲) برابر با ۰/۳۵ و رینولدز و همکاران<sup>(۶)</sup> (۱۹۸۸) برابر با ۰/۴۰ بکار برده‌اند<sup>(۷)</sup>. اما در این مثال مقدار ۰/۵۰ مبنای تعیین ضرایب در نظر گرفته شده است.

## 1- Communality

## 2- Eigenvalues

۳- این مثال بوسیله SPSS نسخه ۶ محاسبه شده است به همین دلیل جداول آن ممکن است از نظر شکل ظاهری یا تقدم و تاخر آن با نتایج نسخه‌های بالاتر قدری متفاوت باشد.

## 4- Jones

## 5- Hooman

## 6- Reynolds et. al

۷- مرکز آموزش مدیریت دولتی و دانشگاه آزاد اسلامی (۱۳۷۹) "مقیاس سنجش سبک مدیریت" (M.S.S) انتشارات مرکز آموزش مدیریت دولتی، ص ۷۹

مشخصه آمار نهایی در تحلیل عاملی در جدول دیگری استخراج می‌گردد. در این جدول مقدار اشتراک هر متغیر (که تنها واریانس مشترک بین متغیرها را تبیین می‌کند و واریانس اختصاصی و واریانس خطا را در نظر نمی‌گیرد که به همین دلیل کمتر از یک می‌باشد)، مقدار ویژه، درصد واریانس و درصد تراکمی استخراج می‌گردد. اما در این جدول چنانچه ملاک کیسر مدنظر باشد تنها عاملهایی استخراج می‌گردند که مقدار ویژه آن بزرگتر از یک باشد. چنانچه تعداد عاملها توسط محقق تعیین شده باشد، در این صورت به همان تعداد، عامل استخراج و مشخص‌های آماری آنها در این جدول ارائه می‌گردد.

بالاخره اینکه آخرین جدول در تحلیل عاملی به ماتریس عاملی دوران یافته اختصاص دارد که باید مبنای نامگذاری عوامل قرار گیرد (جدول ۱۴-۶). در این جدول هر متغیر معمولاً در یک عامل بار می‌شود. اما ممکن است که بعضی از متغیرها، بار عاملی در دو عامل داشته باشند و حتی ممکن است این بار عاملی بصورت دو قطبی نیز ظاهر شود، بطوریکه نیمی از بار عاملی منفی و نیمی دیگر مثبت باشد.

جدول (۱۴-۵) ماتریس عاملی دوران نیافته

Factor Matrix:	Factor 1	Factor 2	Factor 3	Factor 4
X <sub>۱</sub>	.496	-.053	.247	-.283
X <sub>۲</sub>	.662	-.370	-.069	.081
X <sub>۱۴</sub>	.497	-.545	.130	-.287
X <sub>۴</sub>	.578	.431	.001	.0733
X <sub>۵</sub>	.658	-.311	.078	.124
X <sub>۶</sub>	.647	-.467	-.005	-.089
X <sub>۷</sub>	.644	-.444	.050	.006
X <sub>۳</sub>	.687	-.344	-.222	.138
X <sub>۲۳</sub>	.087	-.068	-.308	.408
X <sub>۸</sub>	.640	-.170	.215	-.223
X <sub>۱۰</sub>	.402	.529	.158	-.096
X <sub>۱۸</sub>	.484	.244	-.249	.210
X <sub>۱۱</sub>	.452	.524	-.036	-.067
X <sub>۱۲</sub>	.493	.517	.036	-.156
X <sub>۱۵</sub>	.467	.265	-.250	.0900
X <sub>۱۳</sub>	.443	.541	-.155	.208
X <sub>۱۶</sub>	.561	.447	-.162	-.061
X <sub>۲۲</sub>	.428	.342	.125	-.244
X <sub>۲۱</sub>	-.149	-.056	.403	-.356
X <sub>۱۹</sub>	.026	.180	.613	.653
X <sub>۲۰</sub>	-.099	-.040	.399	-.334
X <sub>۱۷</sub>	.546	.466	-.025	.052
X <sub>۹</sub>	.669	-.424	.003	.083

Factor Matrix:	Factor 5	Factor 6	Factor 7	Factor 8
X <sub>۱</sub>	-.446	-.215	.155	.131
X <sub>۲</sub>	-.060	.120	.219	-.059
X <sub>۱۴</sub>	-.424	.034	-.131	.115
X <sub>۴</sub>	.202	-.000	.276	.013
X <sub>۵</sub>	.261	.100	-.195	-.011
X <sub>۶</sub>	.149	.070	-.101	-.037
X <sub>۷</sub>	-.093	-.152	.063	-.215
X <sub>۳</sub>	.222	.028	-.108	.202
X <sub>۲۳</sub>	-.092	.170	.420	.607
X <sub>۸</sub>	-.077	-.136	.277	-.084
X <sub>۱۰</sub>	.258	-.041	.312	-.250
X <sub>۱۸</sub>	.235	.273	.150	-.233
X <sub>۱۱</sub>	-.102	-.137	.002	.115
X <sub>۱۲</sub>	.303	-.217	-.138	-.013
X <sub>۱۵</sub>	-.405	.406	-.239	-.087
X <sub>۱۳</sub>	.319	-.101	-.013	.070
X <sub>۱۶</sub>	-.092	.251	-.204	-.069
X <sub>۲۲</sub>	-.006	-.134	-.017	.479
X <sub>۲۱</sub>	.449	.191	-.181	.301
X <sub>۱۹</sub>	-.120	-.021	-.036	.020
X <sub>۲۰</sub>	.042	.695	.170	.011
X <sub>۱۷</sub>	-.097	.076	-.268	.123
X <sub>۹</sub>	.001	-.131	-.285	-.017



جدول (۱۴-۶) ضرایب عاملی دوران یافته

Factor Matrix:	Factor 1	Factor 2	Factor 3	Factor 4
X <sub>۱</sub>				.7740
X <sub>۲</sub>	.7063			
X <sub>۱۴</sub>		.6512		
X <sub>۴</sub>	.7650			
X <sub>۵</sub>	.8040			
X <sub>۶</sub>	.7989			
X <sub>۷</sub>	.7230			
X <sub>۲</sub>	.7828			
X <sub>۲۳</sub>				
X <sub>۸</sub>	.5167			
X <sub>۱۰</sub>		.7996		
X <sub>۱۸</sub>				
X <sub>۱۱</sub>		.5830		
X <sub>۱۲</sub>		.7227		
X <sub>۱۵</sub>				
X <sub>۱۳</sub>		.6963		
X <sub>۱۶</sub>			.6352	
X <sub>۲۲</sub>				
X <sub>۲۱</sub>				
X <sub>۱۹</sub>				
X <sub>۲۰</sub>				
X <sub>۱۷</sub>			.5639	
X <sub>۹</sub>	.75994			

Factor Matrix:	Factor 5	Factor 6	Factor 7	Factor 8
X <sub>۱</sub>				
X <sub>۲</sub>				
X <sub>۱۴</sub>				
X <sub>۴</sub>				
X <sub>۵</sub>				
X <sub>۶</sub>				
X <sub>۷</sub>				
X <sub>۲</sub>				
X <sub>۲۳</sub>				
X <sub>۸</sub>				.9177
X <sub>۱۰</sub>				
X <sub>۱۸</sub>	.5229			
X <sub>۱۱</sub>				
X <sub>۱۲</sub>				
X <sub>۱۵</sub>	.5432			
X <sub>۱۳</sub>				
X <sub>۱۶</sub>				
X <sub>۲۲</sub>			.5881	
X <sub>۲۱</sub>		.5448		
X <sub>۱۹</sub>	.9136			
X <sub>۲۰</sub>		.8639		
X <sub>۱۷</sub>				
X <sub>۹</sub>				

## نامگذاری عاملها یا استنتاج اشتراک مفهومی:

پس از دوران عاملها، محقق باید از طریق انتساب متغیرها به عاملها به استنتاج مفهومی بپردازد. بدین معنی که باید برای هر یک از مجموعه متغیرهایی که بر یک عامل تعلق گرفته‌اند، مفهومی عام تعیین کرد تا بتوان به تفسیر آنها پرداخت. از طریق همین فرآیند است که محقق با استفاده از تحلیل عاملی، تعداد زیادی متغیر را در چند عامل خلاصه کرده و نام عمومی برای آن تعیین می‌کند تا بتواند به تلخیص داده‌ها که از اهداف اساسی این تکنیک آماری است دست یابد. نام گذاری عاملها برای این مثال در جدول (۱۴-۷) ارائه شده است.

جدول (۱۴-۷) نمایی از عاملهای تحقیق و سهم هر یک از آنها

شماره عامل	عامل	مقدار ویژه	درصد مقدار ویژه	درصد از کل عاملها
۱	اثر واپس روی	۶/۲	۲۴/۸	۳۵/۶
۲	اثر پخش	۳/۳	۱۳/۶	۱۹/۴
۳	اثر مکانیزاسیون	۱/۸	۷/۲	۱۰/۳۱
۴	اثر زیست محیطی	۱/۵	۶/۱	۸/۷
۵	اثر اقتصادی	۱/۴	۵/۵	۷/۸
۶	اثر تغییر در مالکیت	۱/۰۹	۴/۴	۶/۴
۷	اثر تجاری شدن	۱/۰۴	۴/۲	۶
۸	عامل مکانی	۱	۴	۵/۷
	جمع	--	۶۹/۸	۱۰۰

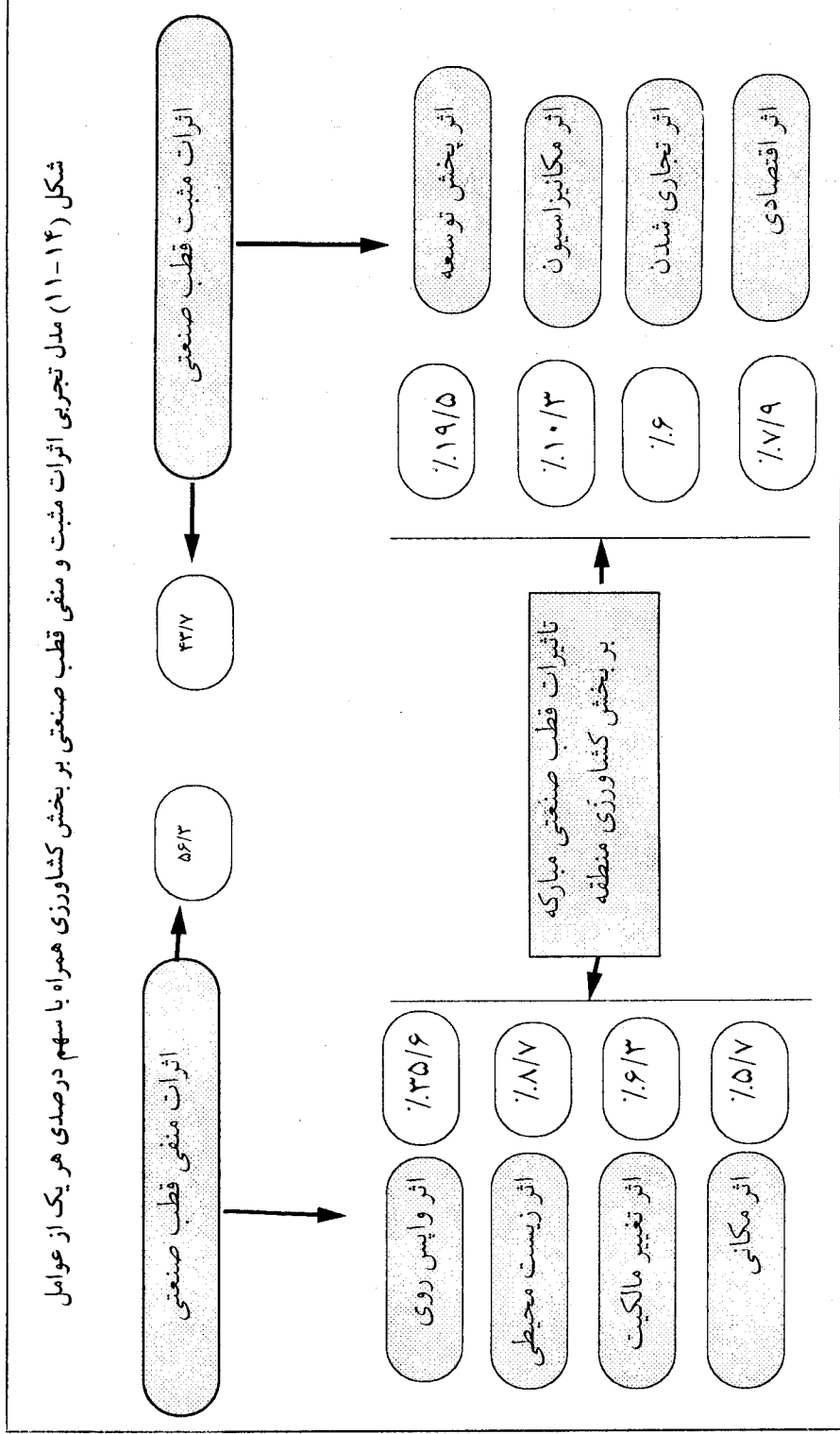
از طریق این عاملها می‌توان به یک مدل کلی‌تر برای تحلیل و تبیین پدیده مورد نظر دست یافت. از آنجا که برخی از این عاملها بیانگر اثرات مثبت قطب صنعتی بر وضعیت کشاورزی منطقه و برخی دیگر نشان دهنده اثرات منفی بوده‌اند بنابراین، اثرات مثبت و منفی عاملها در قالب شکل (۱۴-۱۱) ارائه شده و سهم درصدی هر یک از عاملها نیز محاسبه گردیده‌اند. لازم به

توضیح است که مجموع اثرات ۱۰۰ فرض گردیده و سهم درصدی هر عامل بر این مبنا محاسبه شده است. همانطوریکه شکل (۱۴-۱۱) نشان می‌دهد. میزان اثرات مثبت قطب صنعتی بر وضعیت کشاورزی منطقه ۴۳/۷ درصد بوده که در عملهای اثر پخش<sup>(۱)</sup> (۱۹/۵٪)، اثر مکانیزاسیون (۱۰/۳٪)، اثر تجاری شدن ۶٪ و اثر اقتصادی (۷/۹٪) نمود پیدا کرده است. در حالیکه اثرات منفی قطب صنعتی ۵۶/۳ درصد بوده است که در قالب عملهای اثر واپس‌روی (۳۵/۶٪) اثر زیست محیطی (۸/۷٪)، اثر تغییر مالکیت اراضی (۶/۳٪) و اثر مکانی (۵/۷٪) ظهور یافته است.

---

۱- اثر پخش و اثر واپس‌روی دو مفهومی است که توسط فرانسوا پرو در مورد قطب‌های صنعتی بکار برده شده است. منظور از اثر پخش، توزیع امکانات و فرآیندهای توسعه از قطب صنعتی به نواحی اطراف است و منظور از واپس‌روی جذب پتانسیلهای نواحی اطراف به سمت قطب صنعتی است که به عنوان اثرات منفی تلقی می‌گردد.

شکل (۱۴-۱۱) مدل تجربی اثرات مثبت و منفی قطب صنعتی بر بخش کشاورزی همراه با سهم درصدی هر یک از عوامل



## فصل پانزدهم

# تحلیل خوشه‌ای

### مقدمه:

تحلیل خوشه‌ای یکی از روشهای پر کاربرد در بسیاری از شاخه‌های علمی است. این روش به محقق امکان می‌دهد تا بر مبنای همگنی موجود در بین موارد یا موضوعات مورد مطالعه، آنها را به شیوه‌ای مناسب طبقه بندی کرده و سپس مورد تفسیر و تبیین قرار دهد. تحلیل خوشه‌ای به دو شیوه سلسله مراتبی و غیر سلسله مراتبی انجام می‌گیرد که هر دو آنها در این فصل به طور مجزا مورد بررسی قرار گرفته‌اند. برای تفسیر و توجیه بیشتر هر یک از روشهای فوق از یک مثال فرضی نیز کمک گرفته شده است.

### تحلیل خوشه‌ای چیست؟

تحلیل خوشه‌ای<sup>(۱)</sup> تکنیکی است برای گروه‌بندی افراد یا موضوعات بطوریکه در این گروه‌بندی موضوعات درون گروه شباهت زیادی با همدیگر داشته، اما تفاوت قابل توجهی با

گروه‌های دیگر دارند. بنابراین اگر گروه‌بندی فوق به شیوه صحیح انجام گیرد، در صورت رسم نمودار آنها افراد درون گروه در یک مجموعه نزدیک به هم قرار می‌گیرند در حالیکه فاصله زیادی با سایر گروه‌ها خواهند داشت.

تحلیل خوشه‌ای در رشته‌ها و موضوعات مختلف تحت عناوین متفاوتی نظیر، تحلیل Q<sup>(۱)</sup> گونه‌شناسی<sup>(۲)</sup>، تحلیل رده‌بندی<sup>(۳)</sup> و تکسونومی عددی<sup>(۴)</sup> بکار برده می‌شود<sup>(۵)</sup>. اختصاص نام‌های مختلف به تحلیل خوشه‌ای بدلیل کاربرد وسیع آن در رشته‌هایی نظیر روانشناسی، جامعه‌شناسی، بیولوژی، توسعه روستایی، برنامه‌ریزی منطقه‌ای و سایر رشته‌ها و شاخه‌های علمی است.

تحلیل خوشه‌ای در موقعیتها و موضوعات مختلف مورد استفاده قرار می‌گیرد. به عنوان مثال ممکن است محقق داده‌هایی را از طریق پرسشنامه از تعداد زیادی از افراد جمع آوری کند که بدون طبقه‌بندی آنها داده‌های فوق بی معنی باشند. در چنین مواردی تحلیل خوشه‌ای امکان تلخیص داده‌ها در تعدادی گروه یا طبقه را فراهم می‌کند تا محقق بتواند به تجزیه و تحلیل، استنتاج و نتیجه‌گیری مناسب دست یابد.

#### بیان گرافیکی روش خوشه‌ای:

طبیعت تحلیل خوشه‌ای را می‌توان با یک مثال دو متغیره نشان داد. فرض کنید که محقق می‌خواهد افراد یک جامعه را از نظر میزان محافظه‌کاری طبقه‌بندی کرده و آنها را مورد تجزیه و تحلیل قرار دهد. او ممکن است برای دو متغیر میزان سن و میزان محافظه‌کاری اطلاعات جمع آوری کند. اگر داده‌های فوق در قالب یک نمودار ترسیم شود ممکن است شکل (۱۵-۱) بدست آید.

1- Q-Analysis

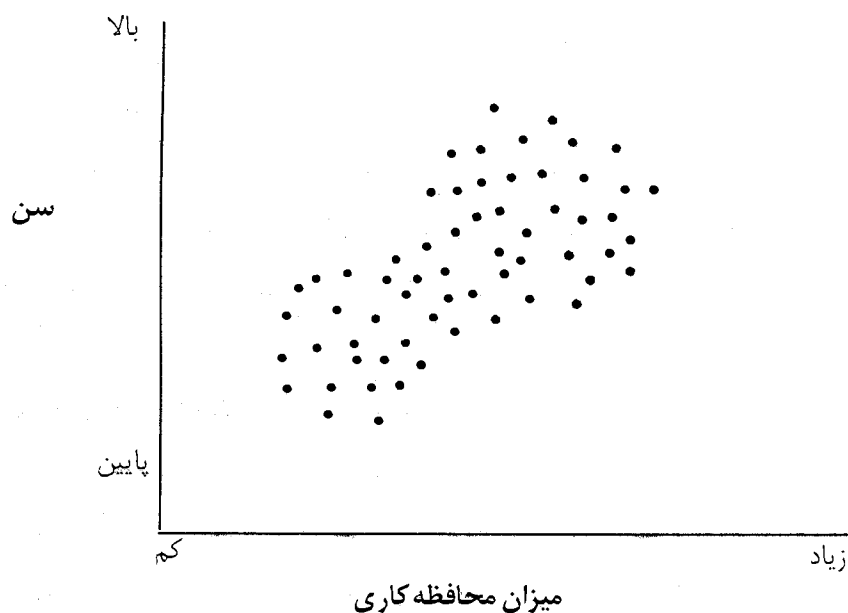
2- Typology

3- Classification Analysis

4- Numerical Taxonomy

۵- برای آشنایی بیشتر با تکنیک تکسونومی به کتاب "برنامه‌ریزی و توسعه منطقه‌ای (تئوریا و تکنیکها)" خلیل کلاتری، انتشارات خوشبین، سال ۱۳۸۰، مراجعه کنید.

در این نمودار محقق می‌تواند با رسم یک خط دو گروه از افراد جامعه را از هم تفکیک و در طبقات مختلف قرار دهد. ممکن است ضریب همبستگی در درون هر یک از گروه‌های فوق  $0/75$  - بدست آید، اما ضریب همبستگی بین دو گروه فوق معنی‌دار نباشد. این مسئله نشان دهنده این واقعیت خواهد بود که افراد درون گروه‌ها از همگنی بالایی برخوردار می‌باشند، در حالیکه گروه‌های فوق نسبت به هم ناهمگن هستند. اینگونه مقایسه‌ها در مواردی که محقق تنها با دو متغیر سر و کار داشته



شکل (۱-۱۵) رابطه سن و میزان محافظه کاری

باشد کار ساده‌ای است. اما چنانچه تعداد زیادی متغیر مدنظر باشد در این صورت مسئله با پیچیدگی زیادی مواجه می‌گردد. به عنوان مثال اگر تعدادی از متغیرها نظیر سن، درآمد، نمره و... بصورت کمی و تعدادی دیگر از آنها نظیر ملیت، جنسیت، مذهب و... بصورت متغیرهای کیفی باشند موضوع محاسبه و تحلیل خوشه‌ای از پیچیدگی بیشتری برخوردار خواهد شد.

### فرآیند انجام تحلیل خوشه‌ای:

تقسیم بندی<sup>(۱)</sup> موارد یا موضوعات براساس همگنی درونی آنها از اهم اهداف تکنیک تحلیل خوشه‌ای است. در اجرای این روش مراحل وجود دارد که بطور متوالی باید یکی پس از دیگری مدنظر قرار گیرد. در استفاده از این روش سه سوال اساسی بشرح ذیل مطرح می‌باشد:

۱ - چگونه باید همگنی موضوعات اندازه‌گیری شود؟

۲ - برای اینکه موضوعات یا موارد همگن در یک گروه قرار گیرند چه روشی باید بکار گرفته

شود؟

۳ - چند گروه یا طبقه باید تشکیل شود؟

پاسخهای مختلفی برای این سوالات ارائه شده‌اند اما هیچکدام از قطعیت لازم برخوردار نیستند و متأسفانه روشهای مختلف نیز نتایج متفاوتی برای داده‌های یکسان ارائه داده‌اند. بنابراین تحلیل خوشه‌ای همراه با تحلیل عاملی عمده‌تأ یک هنر است تا یک علم.

### نحوه اندازه‌گیری همگنی موارد یا موضوعات:

همگنی موضوعات و موارد می‌تواند به شیوه‌های مختلف انجام گیرد. یکی از این روشها مقایسه دودو موارد یا موضوعات و تعیین شباهت و همگنی هر جفت از موارد فوق است. راه دیگر تعیین فاصله یا میزان تفاوت هر جفت از موارد می‌باشد. تعیین فاصله و تفاوتها نیز می‌تواند محقق را در دستیابی به شباهتها و همگنی‌ها رهنمون سازد.

مهمترین روشی که برای تعیین قرابت بین موضوعات مطرح می‌شود تعیین ضریب همبستگی بین هر جفت از موضوعات یا متغیرهاست. اگر در اینجا بجای تعیین ارتباط بین متغیرها، ماتریس اولیه را معکوس کنیم بطوریکه ردیف‌ها نشان دهنده متغیرها و ستونها بیانگر موارد باشند، در این صورت محاسبه ضریب همبستگی بیانگر نزدیکی و قرابت موارد خواهند بود. هرچقدر میزان



همبستگی بدست آمده بیشتر باشد شباهت بین موارد نیز بیشتر خواهد بود و هرچه میزان همبستگی کمتر باشد شباهت موارد کمتر خواهد بود. این فرآیند معمولاً در روش تحلیل عاملی از نوع Q<sup>(۱)</sup> انجام می‌گیرد.

اندازه‌گیری فاصله<sup>(۲)</sup>، بویژه فاصله اقلیدسی<sup>(۳)</sup> نیز از روشهایی است که در سطح وسیع برای تعیین شباهتهای بین موضوعات یا موارد مورد استفاده قرار می‌گیرد. اما قبل از بحث در مورد این روشها، باید تأکید کرد که بسیاری از تکنیکهای تحلیلی مورد استفاده برای تعیین فواصل، نسبت به موارد دور افتاده<sup>(۴)</sup> حساسیت دارند. بنابراین قبل از انجام محاسبات تعیین فاصله، لازم است محقق از وجود اینگونه موارد دور افتاده اطلاع یابد. یکی از روشهای معمول در این زمینه ترسیم نمودار موارد یا موضوعات می‌باشد. برای اینکار محقق باید از طریق کم کردن میانگین از مقادیر صفت و تقسیم آن بر انحراف معیار مقادیر فوق را استاندارد نماید. پس از آن از طریق قرار دادن متغیرها در محور افقی و مقادیر استاندارد شده در محور عمودی نمودار مورد نظر بدست می‌آید. هر نقطه در این نمودار نشان دهنده مقدار متغیر فوق می‌باشد، که از طریق وصل این نقاط امکان تفسیر متغیرها تسهیل می‌گردد. با رسم این نمودارها مقادیر دور افتاده قابل شناسایی خواهند بود. بنابراین مقادیری که میزان انحراف معیار آنها از میانگین (مانند این مثال) بیشتر از ۲/۵ باشد باید مقادیر فوق به عنوان مقادیر دور افتاده، حذف گردد<sup>(۵)</sup>.

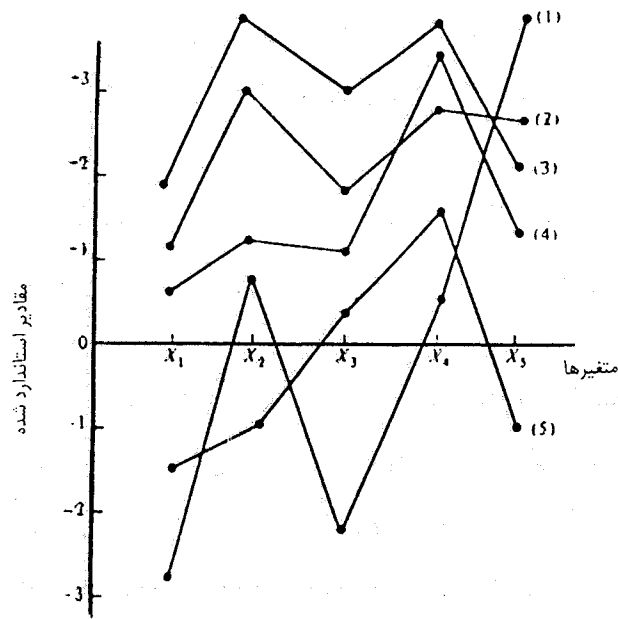
1- Q- type Factor Analysis

2- Distance Measures

3- Euclidean Distance

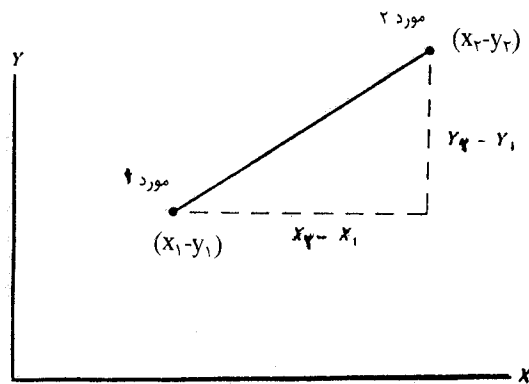
4- Outliers

5- Joseph, Hair (1990) "Multivariate Data Analysis" Macmillan publishing Company, New York, P. 298.



شکل (۲-۱۵) نمودار شناسایی مقادیر دور افتاده

از روشهای پرکاربرد برای اندازه‌گیری همگنی و شباهت، محاسبه فاصله اقلیدسی است. نمونه‌ای از محاسبه فاصله اقلیدسی در شکل (۳-۱۵) ارائه شده است.



شکل (۳-۱۵) فاصله اقلیدسی بین دو موضوع یا مورد

در این نمودار دو بردار  $(X_1, Y_1)$  و  $(X_2, Y_2)$  دو نقطه را نشان می‌دهند فاصله اقلیدسی این دو نقطه براساس فرمول ذیل محاسبه می‌گردد.

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

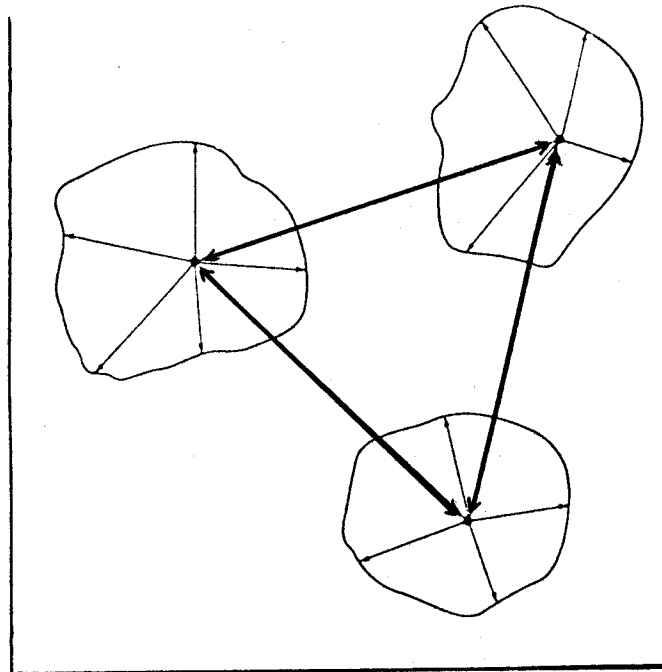
این نمونه تنها، مثالی از اندازه‌گیری فاصله اقلیدسی دو مورد یا دو موضوع می‌باشد، اما همین روند قابل بسط و استفاده برای متغیرهای دیگر نیز می‌باشد. علیرغم کاربرد وسیع فاصله اقلیدسی، این روش با دو محدودیت اساسی مواجه می‌باشد. نخست اینکه نتایج این روش به واحدهای اندازه‌گیری یا مقیاس متغیرها بستگی دارد، برای مثال، اگر یکی از ابعاد اندازه‌گیری شده، مانند طول بر حسب سانتیمتر باشد و آنرا به میلیمتر تبدیل کنید، فاصله اقلیدسی تا حد زیادی تحت تاثیر قرار می‌گیرد و در نتیجه، حاصل تحلیل خوشه‌ای ممکن است کاملاً متفاوت باشد. دوم اینکه، وقتی متغیرها بر پایه مقیاسهای مختلف (مانند قد و وزن) اندازه‌گیری می‌شوند، متغیرهائی که با اعداد بزرگتر اندازه‌گیری شده‌اند (مانند قد) نسبت به متغیرهای با اعداد کوچکتر (مانند وزن) سهم بیشتری در اندازه فاصله خواهند داشت<sup>(۱)</sup>. البته برای رفع این محدودیت، محقق می‌تواند داده‌های خام را با استفاده از روش Z- SCORE استاندارد کرده تا از این طریق به رفع اختلاف مقیاس متغیرها اقدام کند و سپس به محاسبه فاصله اقلیدسی بپردازد.

#### الگوریتم خوشه‌ای کردن داده‌ها:

دومین سؤال در مرحله طبقه بندی داده‌ها این است که چه فرآیند باید طی شود تا موارد یا موضوعات همگن در یک گروه قرار گیرند؟ بعبارت دیگر چه قاعده یا الگوریتمی برای این کار مناسب است؟ این یک سؤال ساده نیست و برنامه‌های رایانه‌ی دهها قاعده و الگوریتم در این زمینه بکار می‌گیرند. اما معیار اصلی برای همه آنها به حداکثر رساندن واریانس تفاوت‌های بین خوشه‌ها نسبت به

۱- هومن، حیدر علی (۱۳۸۰) «تحلیل داده‌های چند متغیری در پژوهش رفتاری»، نشر پارسا، تهران، ص ۴۵۱.

واریانس درون خوشه‌هاست (شکل ۱۵-۴).



← واریانس بین خوشه‌ها

→ واریانس درون خوشه‌ها

شکل (۴-۱۵) نمودار خوشه‌ای، نشان دهنده واریانس درون و بین خوشه‌ها

نسبت واریانس بین خوشه‌ای به واریانس درون خوشه‌ای با نسبت  $F^{(1)}$  در تحلیل واریانس قابل مقایسه است.

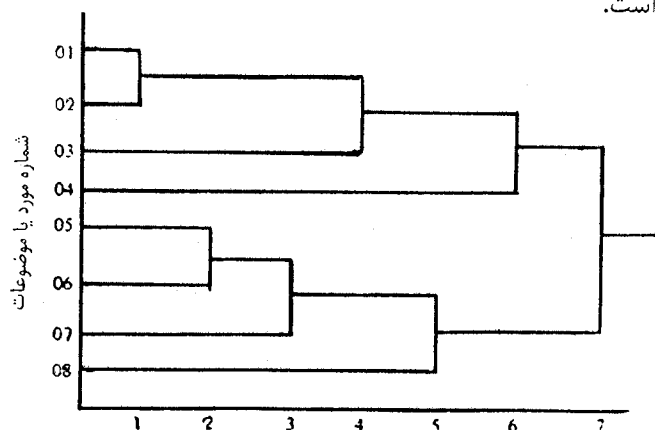
الگوریتم‌های عمومی مورد استفاده در روش خوشه‌ای را می‌توان در دو گروه طبقه بندی کرد.

۱ - روش خوشه‌ای سلسله مراتبی  $(2)$

۲- روش خوشه‌ای غیرسلسله مراتبی<sup>(۱)</sup>

## ۱- روش خوشه‌ای سلسله مراتبی:

این روش دارای یک ساختار درختی سلسله مراتبی است. روش خوشه‌ای سلسله مراتبی خود به دو شیوه مختلف انجام می‌گیرد. این شیوه‌ها عبارتند از: روش تراکمی<sup>(۲)</sup> و روش تفکیک‌پذیری<sup>(۳)</sup>. در روش تراکمی هر مورد یا موضوع با خوشه‌ای خاص خود آغاز می‌شود و سپس، دو مورد یا موضوع با هم ترکیب شده و خوشه‌ای تراکمی جدید می‌سازند. بنابراین در هر مرحله تعداد خوشه‌ها بصورت یک به یک کاهش می‌یابد. در بعضی موارد، مورد یا موضوع سوم با خوشه‌ای دو موردی ادغام شده و خوشه‌ای جدیدی را بوجود می‌آورند، و بعضی موارد نیز ممکن است دو گروه دو نفره یا دو موضوعی با هم ترکیب شوند و خوشه‌ای جدید را بوجود آورند. بدین ترتیب بتدریج تمام افراد یا موارد با همدیگر ادغام شده و نهایتاً یک خوشه بزرگی را پدید می‌آورند. این فرآیند در شکل (۱۵-۵) ارائه شده است.



شکل (۱۵-۵) نمودار خوشه‌ای سلسله مراتبی

1- Nonhierarchical Cluster Procedure

2- Agglomerative Method

3- Divisive Method

اگر فرآیند خوشه‌ای در جهت عکس روش سلسله مراتبی انجام گیرد، در واقع از روش تفکیک‌پذیری استفاده شده است. در روش تفکیکی فرآیند خوشه‌ای از یک خوشه بزرگ که تمام موارد یا موضوعات را در برمی‌گیرد آغاز می‌شود. در مراحل بعدی، مواردی که از همگنی کمتری برخوردار هستند از خوشه اولیه جدا شده و خوشه‌های کوچک و جدیدی تشکیل می‌شوند. این مراحل تا جایی ادامه پیدا می‌کند که دو فرد یا مورد به یک خوشه خاص خود منتهی می‌شوند. در شکل (۱۵-۵) روش خوشه‌ای تراکمی از سمت چپ به راست انجام می‌گیرد در حالیکه روش تفکیک‌پذیری از سمت راست به چپ حرکت می‌کند. اما بسیاری از برنامه‌های رایانه‌ای روش تراکمی را انجام می‌دهند.

پنج روش عمده برای تشکیل خوشه‌ای تراکمی مورد استفاده قرار می‌گیرند این تکنیکها عبارتند از:

۱- پیوند تکی<sup>(۱)</sup>

۲- پیوند کامل<sup>(۲)</sup>

۳- پیوند متوسط<sup>(۳)</sup>

۴- روش وارد<sup>(۴)</sup>

۵- روش مرکز ثقل<sup>(۵)</sup>

این روشها از نظر نحوه محاسبه فاصله بین خوشه‌ها از هم متفاوتند. به همین دلیل هر یک از روشهای فوق بطور جداگانه در ذیل توضیح داده شده‌اند.

1- Single Linkage

2- Complet Linkage

3- Average Linkage

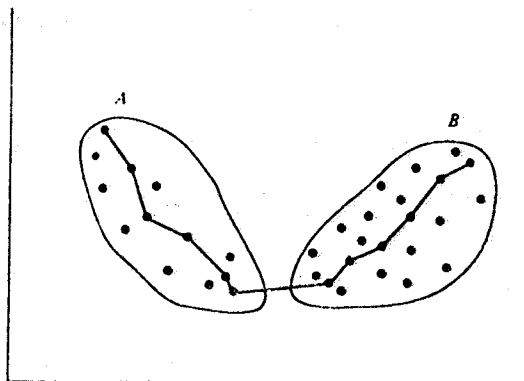
4- Ward's Method

5- Centroid Method

**پیوند تکی:**

روش پیوند تکی براساس حداقل فاصله بنا شده است. این روش دو مورد یا موضوعی که از نظر متغیر یا متغیرهای مورد نظر، کوتاهترین فاصله نسبت بهم دارند را پیدا کرده و در خوشه اول قرار می‌دهد. در مرحله دوم مجدداً کوتاهترین فاصله بعدی را پیدا کرده و از این طریق متغیر یا فرد سوم را با اولین خوشه ترکیب می‌کند و یا اینکه خوشه جدید دو موردی را تشکیل می‌دهد. این فرآیند تاجایی ادامه می‌یابد که همه موارد در یک خوشه بزرگ قرار گیرند. این فرآیند به روش نزدیکترین همسایه<sup>(۱)</sup> معروف است.

فاصله بین دو خوشه، کوتاهترین فاصله بین هر نقطه از یک خوشه به هر نقطه از خوشه دوم است. دو خوشه در هر مرحله بوسیله یک پیوند ضعیف یا قوی بین آنها به هم مرتبط می‌شوند. مشکل زمانی اتفاق می‌افتد که خوشه‌ها بطور بسیار ضعیفی نسبت بهم ارتباط داشته باشند. در چنین مواردی فرآیند پیوند تکی، زنجیره مارپیچی طولانی را تشکیل می‌دهد و در نهایت تمام افراد در یک زنجیره قرار می‌گیرند و افرادی که در انتهای این زنجیره واقع می‌شوند همگنی بسیار ضعیفی نسبت به بقیه خواهند داشت. نمونه‌ای از این مورد در شکل (۱۵-۶) ارائه شده است.



شکل (۱۵-۶) در این نمودار نقاط A و B از همگنی بالایی برخوردار نیستند.

**پیوند کامل :**

فرآیند پیوند کامل<sup>(۱)</sup> شبیه پیوند تکی است بجز اینکه در این روش معیار خوشه‌بندی براساس حداکثر فاصله می‌باشد. به همین دلیل از آن تحت عنوان رهیافت دورترین همسایه<sup>(۲)</sup> یاد می‌شود. در این روش حداکثر فاصله بین دو فرد در یک خوشه نشان دهنده کوچکترین (حداقل قطر دایره) فضا یا محدوده برای یک خوشه می‌باشد. از آنجا که در این روش تمامی موارد یا موضوعات براساس حداکثر فاصله یا حداقل همگنی به همدیگر پیوند می‌خورند، بنابراین از این روش تحت عنوان پیوند کامل نام برده می‌شود. می‌توان گفت که همگنی درون گروهی در این روش برابر با قطر دایره گروه می‌باشد. این روش مشکل ماریجی موجود در روش پیوند تکی را برطرف می‌سازد. هرچند که مشکل سنجش فاصله بین گروهها همچنان پابرجا می‌ماند. در تحلیل تشخیصی<sup>(۳)</sup> معمولاً از مراکز ثقل<sup>(۴)</sup> استفاده می‌گردد، اما ممکن است داده‌های تحلیل خوشه‌ای اجازه استفاده از میانگین‌ها<sup>(۵)</sup> را ندهد. شکل (۷-۱۵) نشان می‌دهد که چگونه ممکن است کوتاهترین و طولانی‌ترین فاصله‌ها همگنی واقعی بین گروهها را منعکس نکنند. استفاده از کوتاهترین فاصله بیانگر این واقعیت است که دو گروه نسبت به هم همگن تر هستند در حالی که بکارگیری طولانی‌ترین فاصله نشان دهنده این موضوع است که آنها بسیار غیرهمگن هستند<sup>(۶)</sup>.

**پیوند متوسط :**

روش پیوند متوسط<sup>(۷)</sup> نیز همانند پیوند تکی و پیوند کامل انجام می‌گیرد، اما معیار خوشه‌بندی در این روش، متوسط فاصله بین گروههاست. در این تکنیک، مقادیر حداکثر یا حداقل،

1- Complete Linkage

2- Furthest- Neighbor Approach

3- Discriminant Analysis

4- Centroids

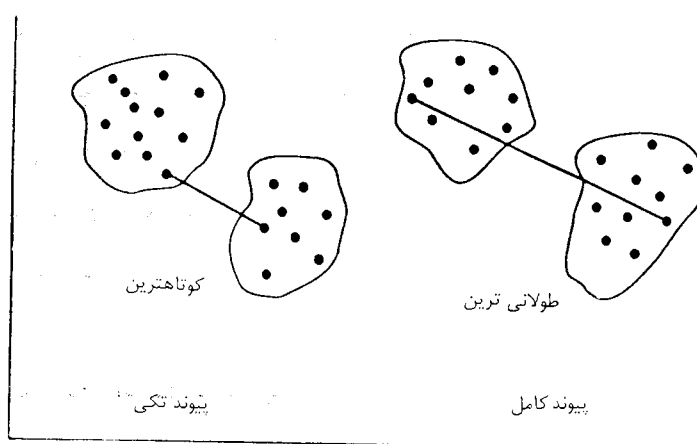
5- Means

6- Hair, Joseph and etal (1990) "Multivariate Data Analysis" Macmillan publishing company, New York, PP. 303-304.

7- Average Linkage



که در روشهای پیوند تکی و پیوند کامل وجود دارد مورد استفاده قرار نمی‌گیرد و گروه‌بندی یا طبقه‌بندی براساس مقادیر متوسط تمام اعضای خوشه انجام می‌گیرد. رهیافت پیوند متوسط، خوشه‌ها را براساس واریانسهای کوچک ترکیب می‌کند. این روش همچنین گرایش به ایجاد خوشه‌ها (بطور تقریبی) با واریانس یکسان دارد. (بعبارت دیگر واریانس خوشه‌ها به همدیگر نزدیک می‌باشد).



شکل (۷-۱۵) نمودار مقایسه سنجش فاصله برای پیوند تکی و پیوند کامل

### روش وارد:

در روش وارد<sup>(۱)</sup>، فاصله بین دو خوشه از طریق مجموع مجذورات بین دو خوشه بدست آمده از کل متغیرها تعیین می‌شود. در این روش در هر مرحله از خوشه‌بندی، خوشه‌های کوچک با هم ترکیب

1- Ward's Method

می‌شوند. در روش وارد تعداد افراد خوشه‌ها تقریباً با هم برابر هستند.

#### روش مرکز ثقل :

در روش مرکز ثقل<sup>(۱)</sup> فاصله بین دو خوشه عبارتست از فاصله اقلیدسی بین میانگین‌های آنها. در این روش در هر مرحله که افراد گروه‌بندی می‌شوند یک مرکز ثقل جدید<sup>(۲)</sup> محاسبه می‌گردد. با ادغام خوشه‌ها در هر مرحله، میانگین خوشه‌ها نیز تغییر می‌یابد. بعبارت دیگر با هر بار اضافه شدن فرد یا گروهی از افراد به خوشه قبلی، مرکز ثقل خوشه نیز تغییر می‌کند. البته ممکن است که در این روش سردرگمی بوجود آید، زیرا فاصله بین مراکز ثقل یک جفت ممکن است کوچکتر از فاصله بین مراکز ثقل جفت‌های قبلی که در خوشه ترکیب شده بودند باشد. اما از نکات قوت این روش، این است که کمتر تحت تأثیر مقادیر دور افتاده<sup>(۳)</sup> قرار می‌گیرد. در حالیکه سایر روشهای سلسله مراتبی تحت تأثیر این مقادیر قرار می‌گیرند<sup>(۴)</sup>.

بطور کلی روش مرکز ثقل نیاز به داده‌های کمی دارند، در حالیکه سایر روشهای بحث شده چنین شرطی ندارند.

#### فرآیند خوشه‌بندی غیرسلسله مراتبی :

در مقابل روشهای سلسله مراتبی، فرآیند خوشه‌بندی غیرسلسله مراتبی<sup>(۵)</sup> از یک ساختار درختی تبعیت نمی‌کند. بلکه در این روش اولین قدم انتخاب مرکز یا مبدأ و بذر<sup>(۶)</sup> اولیه است. با تعیین این بذر اولیه، کلیه افراد درون یک فاصله آستانه‌ای در آن خوشه قرار می‌گیرند. فرآیندهای خوشه‌ای

1- Centriod Method

2- New Centriod

3- outliers

4- Tryon, R.C. and Bailey, D.E. (1980) "Cluster Analysis" McGraw-Hill, New York. P. 66-67.

5- Nonhierarchical Clustering Procedure

6- Seed

غیرسلسله مراتبی بطور متوالی به خوشه‌بندی K میانگین<sup>(۱)</sup> ارجاع داده می‌شود. روشهای خوشه‌ای غیرسلسله مراتبی یکی از تکنیکها یا رویه‌های ذیل را مورد استفاده قرار می‌دهد.

#### ۱- روش آستانه متوالی<sup>(۲)</sup>

این روش با انتخاب یک مبنا یا بذر، تمام افراد درون فاصله آستانه‌ای از قبل تعیین شده را شامل می‌گردد. زمانیکه تمام افراد درون فاصله مورد نظر شامل شدند، دومین بذر خوشه انتخاب شده و تمام افراد درون فاصله انتخاب شده فوق را در برمی‌گیرد. سپس سومین بذر خوشه انتخاب و همان فرآیند ادامه می‌یابد وقتی یک موضوع یا فرد براساس یک مبنا خوشه‌بندی شد، در خوشه بعدی قرار نمی‌گیرد.

#### ۲- روش آستانه موازی<sup>(۳)</sup>

در این روش تعدادی مبنا یا بذر خوشه‌ای بطور همزمان در بدو شروع انتخاب می‌شود و سپس افراد درون فاصله آستانه‌ای به نزدیکترین بذر تعلق می‌گیرند. با آغاز فرآیند خوشه‌بندی، فاصله‌های آستانه‌ای می‌توانند جرح و تعدیل شوند تا تعدادی از افراد در هریک از خوشه‌ها طبقه‌بندی گردند. همچنین در این روش، ممکن است که برخی از افراد یا موضوعاتی که خارج از فاصله آستانه‌ای از قبل تعیین شده می‌باشند در هیچ خوشه‌ای قرار نگیرند.

#### ۳- روش ایتیم کردن<sup>(۴)</sup>

این روش نیز شبیه دو روش قبلی است اما با این تفاوت که در این روش امکان جابجایی از یک خوشه به خوشه‌ای دیگر برای افراد یا موضوعات وجود دارد.

1- K- means Clustering

2- The Sequential Threshold Procedure

3- The Parallel Threshold Procedure

4- Optimizing

در خوشه‌بندی غیر سلسله مراتبی باید مراحل ذیل دنبال شود:

#### الف) تعیین تعداد خوشه‌ها:

اصل مسلمی در زمینه تعیین تعداد خوشه‌ها وجود ندارد. اما توصیه می‌شود که چند خوشه (چهار یا پنج یا شش خوشه) استخراج گردد و سپس براساس موضوع تحقیق و تئوری مورد نظر از طریق مقایسه نتایج خوشه‌های فوق، خوشه مناسب نهایی انتخاب گردد.

#### ب) نام گذاری و توصیف خوشه‌ها:

در تفسیر خوشه‌ها مهمترین قدم تعیین نام یا برچسب مناسب برای خوشه‌هاست، بطوریکه این نام بیانگر طبیعت و سرشت خوشه فوق باشد. به عنوان مثال اگر گرایش یا نظر افراد بر مبنای چندگویه در یک طیف پنج یا هفت گانه سنجیده شود، پس از خوشه‌بندی آنها در سه گروه (خوشه) می‌توان بر حسب نظرات یا گرایش آنها، برای هر یک از گروه‌ها نام یا برچسب مناسبی را انتخاب کرد.

#### ج) تحلیل و تبیین خوشه‌ها:

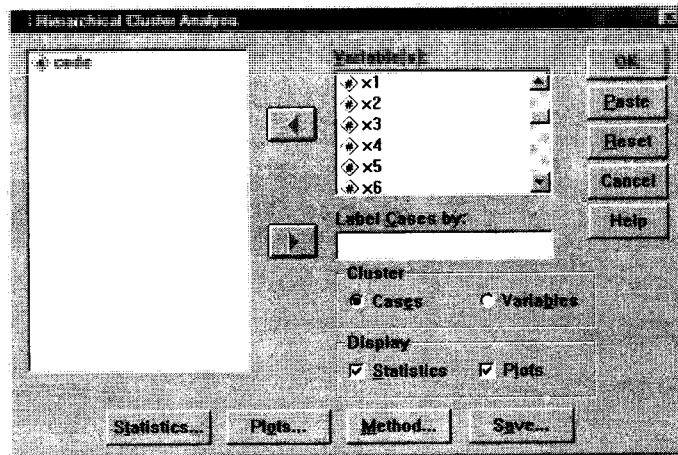
در این مرحله به توصیف و تحلیل ویژگیهای هر خوشه در جهت تبیین تفاوت‌های آنها پرداخته می‌شود، در این مرحله از تحلیل تشخیصی یا روشهای آماری مشابه استفاده می‌شود و محقق از داده‌ها و متغیرهایی استفاده می‌کند که در مرحله خوشه‌بندی مورد استفاده قرار نگرفته‌اند. اینگونه داده‌ها عمدتاً مربوط به متغیرهای جمعیتی، الگوی مصرف یا ویژگیهای فردی و شخصیتی است. با استفاده از روش تحلیل تشخیصی، محقق متوسط مقادیر خوشه‌ها را مقایسه می‌کند. در این تحلیل متغیر وابسته، خوشه‌های از قبل تعیین شده می‌باشد و متغیرهای مستقل ویژگیهای فردی، شخصیتی، روانی و غیره است. به عنوان مثال اگر گرایش افراد نسبت به یک موضوع خاص در سه خوشه طبقه‌بندی شده باشند در این مرحله می‌توان خوشه‌های فوق را در ارتباط با ویژگیهای فردی و شخصیتی، روانی، اقتصادی و غیره مورد بررسی قرار داد و مشخص کرد که گرایش‌ها مربوط به خوشه (الف) عمدتاً افراد جوان، تحصیل کرده و غیره هستند و به همین ترتیب ویژگیهای خوشه‌های بعدی

را می‌توان تعیین نمود.

### روش محاسبه تحلیل خوشه‌ای سلسله مراتبی با برنامه SPSS:

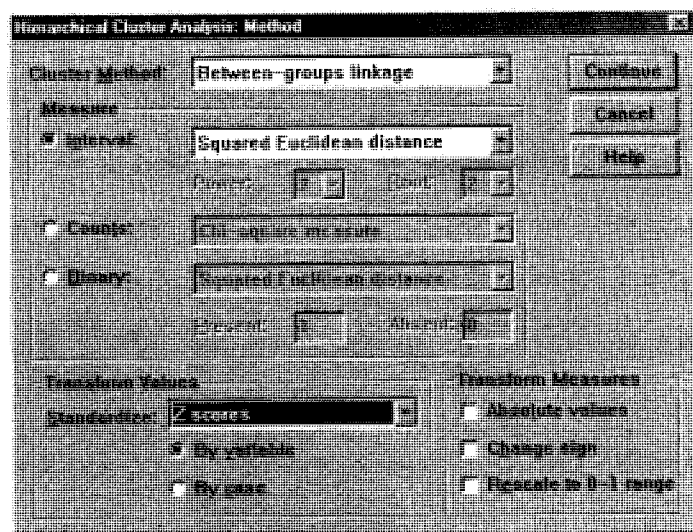
تکنیک تحلیل خوشه‌ای سلسله مراتبی سعی می‌کند تا گروه‌های نسبتاً همسان و مشابه را براساس ویژگی‌های انتخاب شده شناسایی نماید. برای این منظور الگوریتمی را انتخاب می‌کند که در آن هر مورد یا متغیر، با همدیگر ترکیب گردد تا جایی که با ترکیب خوشه‌های مختلف نهایتاً یک خوشه کلی تشکیل شود.

فرض کنید محقق درصدد است تا واحدهای جغرافیایی خاصی (نظیر استانها، شهرستانها، دهستانها یا روستاها) را براساس سطح توسعه یافتگی در گروه‌های مختلف طبقه‌بندی کند. برای اینکار تعدادی شاخص توسعه انتخاب و آنها را در ستونهای مختلف در Data Editor وارد می‌کند. برای خوشه‌بندی کردن آنها باید دستور: Analyze/Classify/Hierarchical Cluster اجرا گردد. با اجرای این دستور شکل (۸-۱۵) در مانیتور ظاهر می‌شود. در این قسمت شاخصهای توسعه مورد نظر باید در قسمت Variables و بر چسب مناطق یا روستاها در قسمت Label cases by وارد گردند.



شکل (۸-۱۵) پنجره Hierarchical Cluster Analysis

باتوجه به اینکه در تحلیل خوشه‌ای باید از داده‌های استاندارد شده استفاده گردد، لذا در صورتی که شاخصهای فوق دارای مقیاسهای متفاوت باشند با فشار دادن دکمه Method از پنجره مکالمه فوق، پنجره دیگری مانند شکل (۹-۱۵) باز می‌شود که در آنجا می‌توان از قسمت Transform Values مورد Z-Score را انتخاب کرد و داده‌ها یا شاخصهای فوق را استاندارد نمود.



شکل (۹-۱۵) پنجره Hierarchical Cluster Analysis: Method

در همین پنجره می‌توان روش خوشه‌ای کردن را از قسمت Cluster Method انتخاب کرد. روشهای مختلف در این زمینه عبارتند از:

- Between Group Linkage
- Within Group Linkage
- Nearest Neighbour
- Furthest Neighbour
- Centroid Clustering

- Median Clustering

- Ward's Method

در قسمت Measure در همین پنجره مکالمه این امکان وجود دارد تا محقق نحوه سنجش فاصله و یا همگنی که باید در فرآیند خوشه‌ای کردن مورد استفاده قرار گیرد را تعیین نماید. در این قسمت محقق باید نوع داده‌های خود را مشخص کرده و سپس برای هر مورد روش سنجش فاصله یا شباهت را تعیین کند. اگر داده‌ها از نوع فاصله‌ای باشد در این صورت روش‌های قابل انتخاب عبارتند از:

- Euclidean Distance

- Squared Euclidean Distance

- Cosine

- Pearson Correlation

- Chebychev

- Block

- Minkowski

- Customized

اگر داده‌ها از نوع شمارشی (Counts) باشند، روش سنجش قابل انتخاب عبارتند از:

- Chi- Square Measure

- Phi- Square Measure

و چنانچه نوع داده بصورت دوتایی (Binary) باشد در این صورت انتخاب‌های زیادی نظیر:

- Euclidean Distance

- Squared Euclidean Distance

- Size Difference

- Pattern Difference

- Variance

- Dispersion

و غیره وجود دارد.

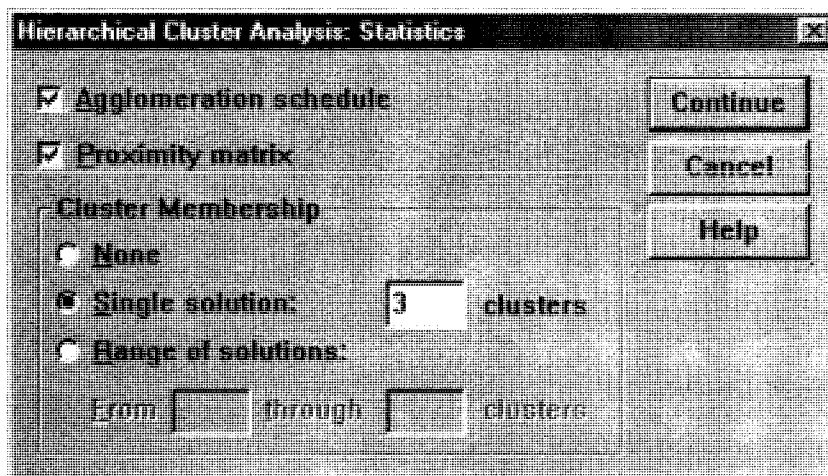
در همین پنجره، قسمت Transform Measures این امکان را فراهم می‌کند تا مقادیر بدست آمده از طریق سنجش فاصله به شیوه‌های مختلف تبدیل شوند. آلترناتیوهای موجود در این زمینه عبارتند از:

- Absolute Values
- Change Sign
- Rescale to 0-1 Range

این موارد در زمانیکه سنجش فاصله محاسبه گردید، انجام می‌گیرد.

با فشار دادن دکمه Statistics از پنجره اصلی تحلیل خوشه‌ای شکل (۱۰-۱۵) ظاهر

می‌شود.



شکل (۱۰-۱۵) پنجره Hierarchical Cluster Analysis: Statistics

در این قسمت می‌توان مواردی نظیر Agglomeration Schedule، proximity Matrix

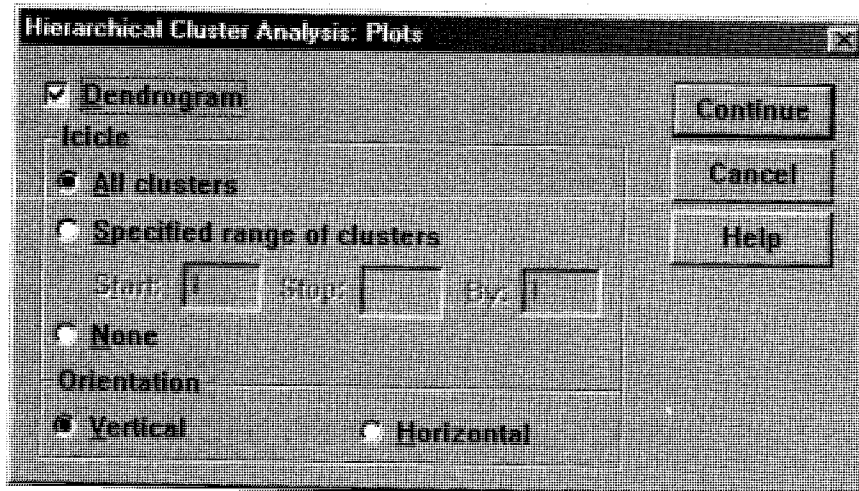


و Cluster Membership را انتخاب کرد.

Agglomeration Schedule فاصله بین موارد یا خوشه‌های ترکیب شده و آخرین سطحی که یک مورد (یا متغیر) به خوشه‌ای ملحق شده را نشان می‌دهد. Proximity Matrix فواصل یا همگنی بین آیتم‌ها را نشان می‌دهد و Cluster Membership نیز اعضای هر خوشه را مشخص می‌کند.

با فشار دادن دکمه Save از پنجره اصلی تحلیل خوشه‌ای سلسله مراتبی، پنجره دیگری باز می‌شود. در این پنجره، در صورت انتخاب Single Solution یا Range of Solutions اعضای هر خوشه مشخص شده و تحت عنوان متغیر جدید در پنجره Data Editor برای تجزیه و تحلیل و استفاده‌های بعدی ذخیره می‌گردد.

با فشار دادن دکمه Plots از پنجره اصلی شکل (۱۱-۱۵) ظاهر می‌شود.



شکل (۱۱-۱۵) پنجره Hierarchical Cluster Analysis: Plots

در این پنجره اولین گزینه، نمودار درختی<sup>(۱)</sup> می‌باشد. با انتخاب این مورد، نمودار درختی نحوه ترکیب موارد یا متغیرها، بطور سلسله مراتبی نمایش داده می‌شود. ادغام موارد یا خوشه‌ها تا جایی ادامه می‌یابد که نهایتاً همه آنها به یک خوشه بزرگ ختم می‌شوند. این نمودار همچنین ضرایب فاصله هر مرحله را نیز نشان می‌دهد. نمودار درختی فواصل اصلی را مجدداً مقیاس‌بندی کرده و اعداد بین ۰ تا ۲۵ برای آنها اختصاص می‌دهد که نشان دهنده نسبت فاصله‌های بین مراحل می‌باشد.

قسمت قندیل<sup>(۲)</sup> امکان رسم قندیل برای کل خوشه‌ها و یا برای محدوده خاصی از آنها را بصورت عمودی یا افقی فراهم می‌کند. این قندیل‌ها اطلاعاتی درباره اینکه چگونه موارد در هر تکرار (هر مرحله از تحلیل) در خوشه‌ها ترکیب شده‌اند را نشان می‌دهند. بعبارت دیگر این قندیل‌ها نشان می‌دهند که موارد چگونه در خوشه‌ها ترکیب و ادغام شده‌اند. در پایین این قندیل، (اگر بصورت افقی رسم شده باشد) هیچ یک از موارد با همدیگر ترکیب نمی‌شوند. اما اگر بطرف بالاتر حرکت کنیم (و یا در صورت عمودی بودن قندیل از راست به چپ حرکت کنیم) مواردی که ادغام و ترکیب شده‌اند با علامت X مشخص می‌شوند در حالیکه خوشه‌های مختلف بوسیله یک فضای خالی بین آنها نشان داده می‌شوند.

#### مثال:

فرض کنید که در صد هستیم تا با بررسی وضعیت توسعه ۸ دهستان، آنها را در سه گروه همگن از نظر توسعه یافتگی طبقه‌بندی کنیم. از سوی دیگر می‌خواهیم تا فاصله اقلیدسی دهستانها را لحاظ وضعیت توسعه نسبت به همدیگر تعیین نماییم. برای این کار تعدادی شاخص توسعه انتخاب و آنها را در پنجره Data Editor وارد کرده و دستور محاسبه تحلیل خوشه‌ای را به شرحی که در صفحات قبل گفته شد اجرا کرده‌ایم. نتایج در قالب جداول (۱-۱۵) و (۲-۱۵) بدست آمده‌اند.

## جدول (۱-۱۵) نتایج محاسبات تحلیل خوشه‌ای سلسله مراتبی

## Proximity Matrix

Case	Squared Euclidean Distance							
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H
1:A		15672835	1563930.6	12518578	11932646	3990639.3	1625089.9	2340939.8
2:B	15672835		7640814.5	8906370.0	16340506	18879066	7634861.5	15499059
3:C	1563930.6	7640814.5		9420849.0	11531003	6199174.5	294863.75	3814875.0
4:D	12518578	8906370.0	9420849.0		1381159.0	5846476.0	6677895.0	5656752.5
5:E	11932646	16340506	11531003	1381159.0		3110646.5	8171782.0	3966839.0
6:F	3990639.3	18879066	6199174.5	5846476.0	3110646.5		4109418.8	309670.66
7:G	1625089.9	7634861.5	294863.75	6677895.0	8171782.0	4109418.8		2261763.0
8:H	2340939.8	15499059	3814875.0	5656752.5	3966839.0	309670.66	2261763.0	

This is a dissimilarity matrix

## Average Linkage (Between Groups)

## Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	7	294863.750	0	0	4
2	6	8	309670.656	0	0	5
3	4	5	1381159.000	0	0	6
4	1	3	1594510.250	0	1	5
5	1	6	3786135.000	4	2	6
6	1	4	7883346.500	5	3	7
7	1	2	12939073.0	6	0	0

## Cluster Membership

Case	3 Clusters
1:A	1
2:B	2
3:C	1
4:D	3
5:E	3
6:F	1
7:G	1
8:H	1

جدول (۱۵-۲) قنذیل عمودی<sup>(۱)</sup> و نمودار درختی<sup>(۲)</sup>

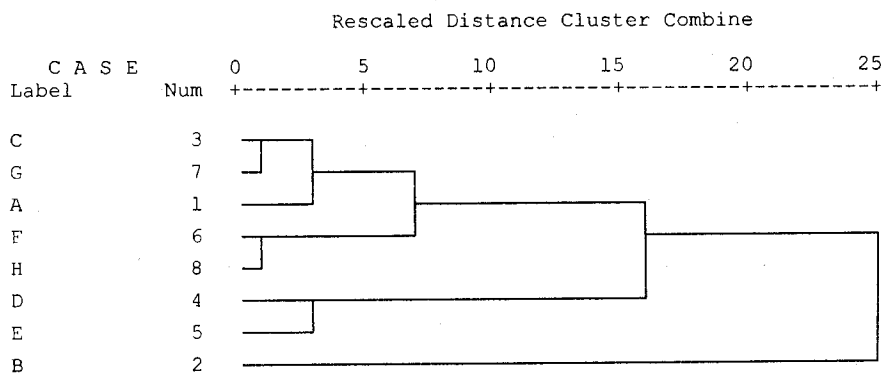
Vertical Icicle

Number of clusters	Case														
	2:B		5:E		4:D		8:H		6:F		7:G		3:C		1:A
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	X		X	X	X	X	X	X	X	X	X	X	X	X	X
3	X		X	X	X	X	X	X	X	X	X	X	X	X	X
4	X		X	X	X	X	X	X	X	X	X	X	X	X	X
5	X		X	X	X	X	X	X	X	X	X	X	X	X	X
6	X		X	X	X	X	X	X	X	X	X	X	X	X	X
7	X		X	X	X	X	X	X	X	X	X	X	X	X	X

Dendrogram

\*\*\*\*\* HIERARCHICAL CLUSTER ANALYSIS \*\*\*\*\*

Dendrogram using Average Linkage (Between Groups)



1- Vertical Icicle

2- Dendrogram

همانطوریکه از جدول (۱-۱۵) برمی‌آید، در قسمت اول معمولاً تعداد موارد یا دهستانها مشخص می‌گردد. قسمت دوم که به Proximity Matrix معروف است مربع فاصله اقلیدسی هر مورد یا دهستان نسبت به مورد یا دهستان بعدی را نشان می‌دهد. مقادیر کوچکتر در این ماتریس بیانگر میزان مشابهت یا همگنی دو مورد مربوط به آن عدد می‌باشد. با افزایش همگنی موارد یا دهستانها از همدیگر مقدار ماتریس شباهت مربوط به موارد فوق نیز افزایش می‌یابد.

قسمت سوم جدول که ترکیب تراکمی خوشه‌ها را تحت عنوان Agglomeration Schedule نشان می‌دهد، از طریق محاسبه پیوند متوسط<sup>(۱)</sup> انجام گرفته است، که این روش بصورت پیش‌گزیده در SPSS مورد استفاده قرار می‌گیرد. اما همانطوریکه در صفحات قبل در همین فصل نیز بیان گردید پنج روش اصلی برای تشکیل خوشه‌های تراکمی مانند پیوند تکی، پیوند کامل، پیوند متوسط، روش وارد و روش مرکز ثقل وجود دارد که در این مثال از پیوند متوسط استفاده شده است. معیار خوشه‌بندی در این روش فاصله بین گروهها یا موارد است. در این روش معیار طبقه‌بندی نه بر اساس مقادیر حداقل یا حداکثر بلکه بر اساس مقادیر متوسط تمام اعضای خوشه انجام می‌گیرد. همانطوریکه از جدول مربوط نیز برمی‌آید، فرآیند خوشه‌بندی در هفت مرحله انجام گرفته است، ستونهای دوم و سوم جدول نشان می‌دهد که در مرحله اول مورد ۳ و ۷، در مرحله دوم مورد ۶ و ۸، در مرحله سوم مورد ۴ و ۵ در مرحله چهارم مورد ۱ و ۳، در مرحله پنجم مورد ۱ و ۶، در مرحله ششم مورد ۱ و ۴ و بالاخره در مرحله هفتم مورد ۱ و ۲ خوشه‌بندی شده‌اند. این روند در نمودار درختی<sup>(۲)</sup> و قندیل عمودی<sup>(۳)</sup> رسم شده کاملاً مشخص است (جدول ۱-۲).

از آنجا که در اجرای دستور خوشه‌بندی خواسته شده بود که دهستانها در سه گروه همسان طبقه‌بندی شوند، بنابراین اعضای هر طبقه در جدول مستقلی تحت عنوان Cluster Membership مشخص شده است.

1- Areage linkage

2- Dendrogram

3- Vertical Icicle

روش خوشه‌ای K میانگین<sup>(۱)</sup>:

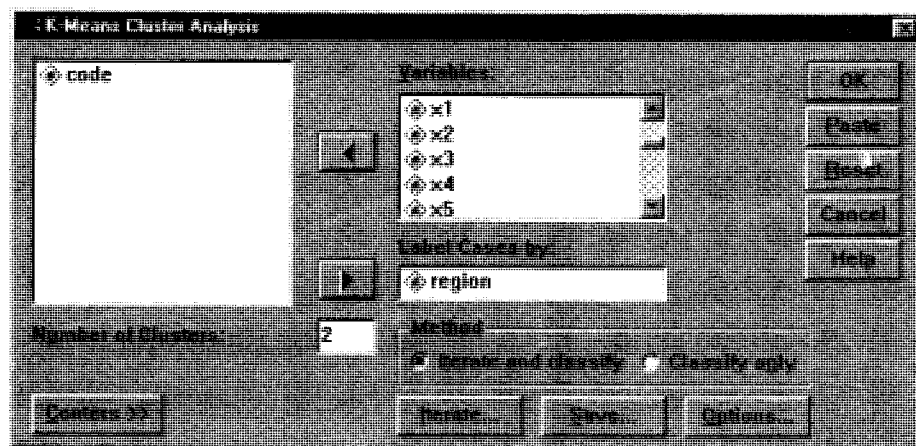
این روش سعی می‌کند تا گروه‌های همگنی از موارد یا موضوعات تحت مطالعه را براساس ویژگی‌های انتخاب شده شناسایی کند. این شناسایی براساس الگوریتمی انجام می‌شود که تعداد قابل توجهی از موارد را شامل می‌شود. بنابراین، این الگوریتم نیاز به تعیین تعداد خوشه‌ها دارد. داده‌هایی که در روش خوشه‌ای K میانگین مورد استفاده قرار می‌گیرند باید از نوع پارامتری یا کمی باشند. چنانچه داده‌ها غیرپارامتری یا کیفی باشند باید از تحلیل خوشه‌ای سلسله مراتبی استفاده کرد. در این روش فاصله‌ها براساس فاصله اقلیدسی تعیین می‌شود. اگر بخواهید از فاصله‌های اندازه‌گیری دیگر استفاده کنید باید از تحلیل خوشه‌ای سلسله مراتبی، که در آن علاوه بر فاصله اقلیدسی، انتخاب‌های دیگری نیز وجود دارد، استفاده کنید. نکته مهم در اینجا اختلاف مقیاس متغیرهاست. در صورتی که متغیرهای مورد استفاده دارای مقیاسهای متفاوتی باشند باید قبل از خوشه‌بندی، متغیرهای فوق استاندارد شوند.

انتخاب تعداد خوشه‌ها نیز از اهمیت بالایی برخوردار می‌باشد. اگر تعداد خوشه‌ها دقیق انتخاب نشوند و یا متغیرهای مهم حذف شوند ممکن است نتایج بدست آمده با واقعیتها مطابقت نداشته باشد.

روش خوشه‌ای در واقع تلاش می‌کند تا موارد یا موضوعات را در تعدادی گروه همگن طبقه‌بندی کند. روش خوشه‌ای K میانگین نسبت به تکنیک سلسله مراتبی، روش سریعتری است. اما روش سلسله مراتبی از انعطاف بیشتری برخوردار است و در آن روشهای سنجش فاصله دیگری غیر از ملاک فاصله اقلیدسی وجود دارد و می‌توان از داده‌های غیرپارامتری یا کیفی نیز استفاده کرد. علاوه بر آن در این روش ضرورتی ندارد تا محقق از قبل تعداد خوشه‌ها را تعیین کند.

## روش خوشه‌بندی K میانگین با نرم‌افزار SPSS:

برای انجام روش خوشه‌بندی K میانگین داده‌ها یا شاخصهای مورد نظر را در Data Editor وارد کرده و دستور Analyze/Classify/K- means Cluster را اجرا کنید تا شکل (۱۵-۲) در مانیتور ظاهر شود. در این قسمت متغیرها یا شاخصهای مورد نظر را در قسمت Variables و بر چسب مناطق یا روستاها و یا گروه‌ها را در قسمت Label Cases by قرار دهید.



شکل (۱۵-۱۲) پنجره K-Means Cluster Analysis

در این پنجره چندین دکمه وجود دارد که با فشار دادن هر یک از آنها می‌توان خروجیهای مورد نیاز را تعیین کرد. به عنوان مثال با فشار دادن دکمه Save پنجره‌ای باز می‌شود که در آن می‌توان موارد Cluster Membership و Distance from Cluster Center را انتخاب نمود. در صورت انتخاب مورد مربوط به اعضای خوشه، متغیر جدیدی در پنجره Data Editor ایجاد می‌شود که در آن اعضای نهایی هر خوشه را مشخص می‌کند. اعداد موجود در این متغیر از ۱ شروع شده و به تعداد خوشه‌هایی که تعیین شده است ختم می‌گردد.

در صورت انتخاب مورد مربوط به فاصله از مرکز خوشه، متغیر جدیدی ایجاد می‌شود که

فاصله اقلیدسی بین هر مورد و مرکز طبقه‌بندی آن را نشان می‌دهد. دکمه Iterate از پنجره اصلی زمانی روشن می‌شود که از قسمت Method در پنجره اصلی، مورد Iterate and Classify انتخاب شده باشد. با فشار دادن دکمه Iterate پنجره‌ای باز می‌شود که در آن می‌توان حداکثر تکرار<sup>(۱)</sup> را تعیین نمود. این مقدار تعداد تکرار را در الگوریتم K-means مشخص می‌کند و تکرار پس از این عدد متوقف می‌شود. تعداد تکرار می‌تواند بین ۱ تا ۹۹۹ در نوسان باشد.

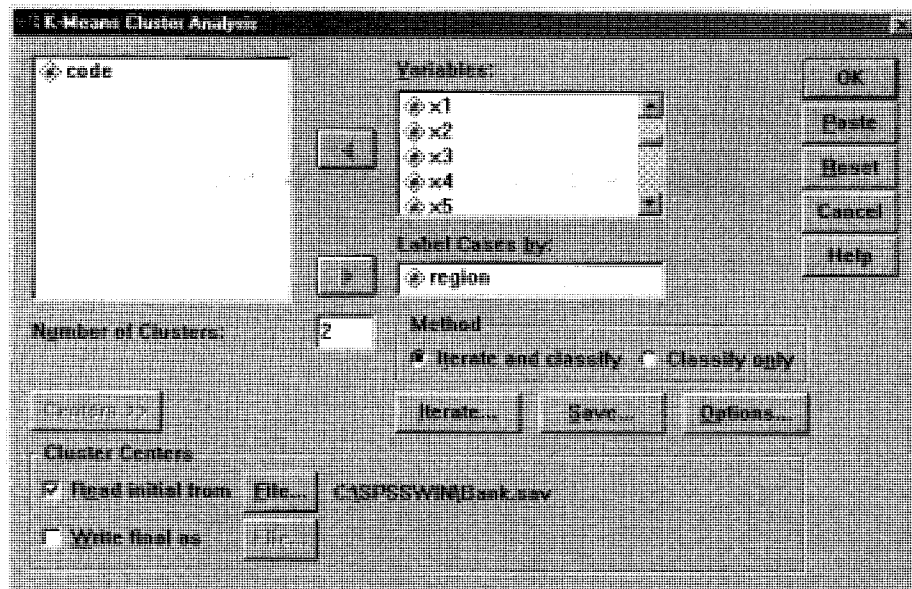
از سوی دیگر در همین پنجره می‌توان معیار همگرایی<sup>(۲)</sup> را نیز تعیین نمود. این انتخاب تعیین می‌کند که چه موقع موارد را تکرار کند. این مورد، نسبت حداقل فاصله بین مراکز خوشه اولیه را تعیین می‌کند. این مقدار باید بزرگتر از صفر و کوچکتر از یک باشد.

برای اجرای بهتر K-means Cluster ابتدا در پنجره اصلی متغیرهای مورد نظر را در قسمت Variables وارد کنید سپس از قسمت Method، مورد Iterate and Classify را انتخاب کنید، بعد دکمه Centers را فشار دهید تا شکل (۱۵-۱۳) ظاهر شود. در این پنجره مورد Write Final as File را علامت زده سپس دکمه File مقابل را فشار داده و نامی برای فایل فوق انتخاب کنید. سپس تعداد خوشه‌ها را نیز در قسمت Numbers of Clusters مشخص کنید. OK را اجرا کرده و به قسمت Output بروید. سپس به پنجره داده‌ها برگشته و بعد از پنجره اصلی K میانگین از قسمت Cluster Centers، مورد Read initial from File را انتخاب کرده و بعد دکمه File مقابل را فشار دهید و فایل قبلی را که نامگذاری کرده بودید باز کنید. سپس از پنجره اصلی مورد Classify only را انتخاب نمایید. تا محاسبات مورد نظر در پنجره Output ظاهر شوند.

1- Maximum Iterations

2- Convergence Criterion





شکل (۱۵-۱۳) پنجره K-Means Cluster Analysis

با فشار دادن دکمه Options می‌توانید آماره‌های

- Initial Cluster Centers
- ANOVA table
- Cluster information for each case

را نیز انتخاب کنید. با فشار دادن دکمه OK نتیجه محاسبات در پنجره Output ظاهر می‌شود.

مثال:

همان مثال قبلی که در روش خوشه‌ای سلسله مراتبی محاسبه گردید با روش خوشه‌ای K میانگین نیز محاسبه شده است. با این تفاوت که در مثال خوشه‌ای سلسله مراتبی دهستانها با استفاده از ۳۲ شاخص خوشه‌بندی شده‌اند در حالیکه در این مثال تنها از ۱۰ شاخص استفاده شده

است. تفاوت موجود در نتایج خوشه بندی دو روش، بدلیل تفاوت در تعداد متغیرهای مورد استفاده در تحلیل می‌باشد. نتایج بدست آمده در جدول (۳-۱۵) ارائه شده‌اند.

جدول (۳-۱۵) نتایج محاسبات روش خوشه‌ای K میانگین

Initial Cluster Centers

	Cluster		
	1	2	3
X1	105	107	38
X2	79.9	29.2	40.9
X3	65.9	50.7	43.9
X4	56.4	40.1	31.5
X5	146	142	124
X6	97	86	65
X7	7	4	4
X8	36.7	14.2	11.5
X9	10.2	2.4	2.6
X10	17	4	4

Input from FILE Subcommand

Iteration History<sup>a</sup>

Iteration	Change in Cluster Centers		
	1	2	3
1	.000	.000	.000

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 1. The minimum distance between initial centers is 62.875.

Cluster Membership

Case Number	REGION	Cluster	Distance
1	A	1	.000
2	B	2	.000
3	C	3	20.121
4	D	3	38.493
5	E	3	14.363
6	F	3	8.004
7	G	3	27.837
8	H	3	27.823

## Final Cluster Centers

	Cluster		
	1	2	3
X1	105	107	38
X2	79.9	29.2	40.9
X3	65.9	50.7	43.9
X4	56.4	40.1	31.5
X5	146	142	124
X6	97	86	65
X7	7	4	4
X8	36.7	14.2	11.5
X9	10.2	2.4	2.6
X10	17	4	4

## Distances between Final Cluster Centers

Cluster	1	2	3
1		62.875	97.606
2	62.875		76.304
3	97.606	76.304	

## ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
X1	3503.083	2	163.467	5	21.430	.004
X2	782.023	2	95.930	5	8.152	.027
X3	212.202	2	34.719	5	6.112	.045
X4	276.008	2	52.359	5	5.271	.059
X5	304.000	2	148.400	5	2.049	.224
X6	543.771	2	235.867	5	2.305	.195
X7	4.771	2	.667	5	7.156	.034
X8	272.863	2	8.030	5	33.982	.001
X9	25.293	2	.651	5	38.873	.001
X10	79.000	2	1.100	5	71.818	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

## Number of Cases in each Cluster

Cluster	1	1.000
	2	1.000
	3	6.000
Valid		8.000
Missing		.000

همانطوریکه از جداول فوق برمی آید. دهستانهای مورد مطالعه در سه گروه طبقه‌بندی شده‌اند. مرکز خوشه برای شاخصهای مختلف نیز محاسبه و اعضای هر خوشه نیز تعیین شده‌اند. بر اساس نتایج طبقه‌بندی، دهستانهای A و B هر یک بطور مستقل در یک گروه قرار گرفته‌اند در حالیکه، دهستانها، C، D، E، F، G و H بدلیل برخورداری از همگنی و سطح توسعه نسبتاً مشابه در یک خوشه طبقه‌بندی شده‌اند. فاصله مرکز هر خوشه با مرکز خوشه‌های دیگر نیز در جدول ماتریس دیگری تحت عنوان Distance between Final cluster centers ارائه شده است.

جدول تحلیل واریانس یا ANOVA در این محاسبات بصورت مستقل نمایش داده می‌شود. توجه کنید که از آزمون F در چنین مواردی نباید برای تست فرضیات استفاده شود. بلکه صرفاً برای استفاده‌های توصیفی توصیه می‌گردد. زیرا خوشه‌ها در این روش به شیوه‌ای انتخاب می‌شوند که تفاوت بین موارد یا موضوعات را در بین خوشه‌های مختلف به حداکثر برساند. به عبارت دیگر اعضای یک خوشه با همدیگر همگنی بسیار بالایی پیدا می‌کنند در حالیکه این همگنی در بین اعضای یک خوشه با اعضای خوشه دیگر به حداقل می‌رسد.

## منابع

### الف فارسی:

- ۱- افشین نیا فرساد (۱۳۷۸) «تحلیل کاربردی داده‌ها: راهنمای استفاده از نرم‌افزار SPSS» انتشارات دانشگاه اصفهان.
- ۲- پورکریمی جواد (۱۳۷۹) «بررسی عوامل مؤثر بر خوداشتغالی فارغ التحصیلان مقطع کارشناسی دامپروری» پایان نامه کارشناسی ارشد- دانشکده روانشناسی و علوم تربیتی دانشگاه تهران.
- ۳- رفیع پور فرامرزی (۱۳۶۰) «کندوکاوها و پنداشتها» شرکت سهامی انتشار، تهران.
- ۴- رونزر پیرت (۱۳۶۴) «میانی پژوهش در علوم اجتماعی» ترجمه محمد دادگران، نشر نوآور، تهران.
- ۵- چلبی مسعود (۱۳۷۲) «جزوه درسی آمار در علوم اجتماعی» دانشگاه شهید بهشتی، تهران.
- ۶- حقیقی سرفراز (۱۳۷۹) «بررسی اثرات قطب صنعتی مبارکه اصفهان بر وضعیت کشاورزی منطقه» پایان نامه کارشناسی ارشد توسعه روستایی، دانشکده کشاورزی دانشگاه تهران.
- ۷- دلاور علی (۱۳۷۶) «روشهای آماری در روانشناسی و علوم تربیتی» انتشارات دانشگاه پیام نور، تهران.
- ۸- دواس، دی، ای (۱۳۷۶) «پیمایش در تحقیقات اجتماعی» ترجمه هوشنگ نایبی، نشر نی، تهران.

تهران.

- ۹ - دوورژه موریس (۱۳۶۶) «روشهای علوم اجتماعی» ترجمه خسرو اسدی، مؤسسه انتشارات امیرکبیر، تهران.
- ۱۰ - ساروخانی باقر (۱۳۷۲) «روشهای تحقیق در علوم اجتماعی» انتشارات مؤسسه مطالعات و تحقیقات فرهنگی، تهران.
- ۱۱ - ساعی علی (۱۳۷۷) «آمار در علوم اجتماعی با کاربرد نرم افزار SPSS/PC<sup>+</sup> در پژوهش‌های اجتماعی»، مؤسسه نشر جهاد، تهران.
- ۱۲ - سرمدی زهره و دیگران (۱۳۷۸) «روشهای تحقیق در علوم رفتاری» دفتر نشر آگه، تهران.
- ۱۳ - سکاران اوما (۱۳۸۰) «روشهای تحقیق در مدیریت» ترجمه محمد صائبی و محمود شیرازی، انتشارات مرکز آموزش مدیریت دولتی، تهران.
- ۱۴ - سیف علی اکبر (۱۳۷۶) «اندازه‌گیری و سنجش در علوم تربیتی» انتشارات دانشگاه پیام نور، تهران.
- ۱۵ - طالقانی محمود (۱۳۷۰) «روش تحقیق نظری در علوم اجتماعی» انتشارات دانشگاه پیام نور، تهران.
- ۱۶ - عبداللهی محمد «جزوه درسی روشهای تحقیق در علوم اجتماعی» دانشکده علوم اجتماعی، دانشگاه علامه طباطبایی، تهران.
- ۱۷ - علیجانی بهلول (۱۳۷۴) «پژوهش میدانی در جغرافیا (روش تحقیق عملی)» انتشارات دانشگاه پیام نور، تهران.
- ۱۸ - قاضی طباطبایی محمود (۱۳۷۴) «تکنیکهای خاص تحقیق» انتشارات دانشگاه پیام نور، تهران.
- ۱۹ - کرلینجر و پدهازور (۱۳۶۶) «رگرسیون چند متغیری در پژوهش رفتاری» ترجمه حسن سرایی، انتشارات مرکز نشر دانشگاهی، تهران.
- ۲۰ - کورتز، نورمن آر (۱۳۷۴) «مقدمه‌ای بر آمار در علوم اجتماعی» ترجمه حبیب اله تیموری

نشرنی، تهران.

۲۱- کلانتری خلیل (۱۳۷۸) «جزوه درسی SPSS for Windows (برای دانشجویان کارشناسی ارشد ترویج و آموزش کشاورزی و توسعه روستایی)» دانشگاه تهران.

۲۲- کلانتری خلیل (۱۳۸۰) «برنامه ریزی و توسعه منطقه‌ای (تئوریها و تکنیکها)» انتشارات خوشبین، تهران.

۲۳- کلانتری خلیل (۱۳۸۰) «نقدی بر متدولوژی سنجش سطح توسعه انسانی UNDP» فصلنامه تحقیقات جغرافیایی، شماره ۶۱، سال ۱۳۸۰، مشهد.

۲۴- کیقبادی پرویز و حسن ستاری (۱۳۵۴) «آمار و کاربرد آن در مدیریت: علوم اداری، اقتصادی، بازرگانی و علوم اجتماعی» انتشارات دانشکده علوم اداری و مدیریت بازرگانی دانشگاه تهران.

۲۵- کینیریل و کلین گری (۱۳۷۸) «کتاب آموزشی SPSS در محیط ویندوز» ترجمه علیرضا منتظری، انتشارات آذرنگ، تهران.

۲۶- لومسبری، جان اف و فرانک تی، آلدريج (۱۳۷۱) «درآمدی بر روشها و فنون میدانی جغرافیا» انتشارات سمت، تهران.

۲۷- مرکز آمار ایران (۱۳۷۸) «واژه‌ها و اصطلاحات آماری» انتشارات مرکز آمار ایران، تهران.

۲۸- مرکز آموزش مدیریت دولتی و دانشگاه آزاد اسلامی (واحد ساوه) (۱۳۷۹) «مقیاس سنجش سبک مدیریت" (M.S.S)، انتشارات مرکز آموزش مدیریت دولتی، تهران.

۲۹- منصورفر کریم (۱۳۷۴) «آمار در علوم اجتماعی» انتشارات دانشگاه پیام نور، تهران.

۳۰- منصورفر کریم (۱۳۷۶) «روشهای آمار» انتشارات دانشگاه تهران.

۳۱- منصورفر کریم، «جزوه رگرسیون چند متغیره برای دانشجویان کارشناسی ارشد پژوهشگری» دانشگاه تهران

۳۲- موزر، س. ک و کالتون، ج (۱۳۶۷) «روش تحقیق» ترجمه کاظم ایزدی، انتشارات کیهان، تهران.

- ۳۳ - مهدوی مسعود (۱۳۷۳) «آمار و روشهای تجزیه و تحلیل داده‌ها در جغرافیا» نشر قومس، تهران.
- ۳۴ - نادری عزت الله و مریم سیف نراقی (۱۳۷۳) «روشهای تحقیق و چگونگی ارزشیابی آن در علوم انسانی» دفتر تحقیقات و انتشارات بدر، تهران.
- ۳۵ - نوروسیس ماریا (۱۳۷۸) «کتاب آموزشی آنالیز آماری داده‌ها با SPSS8» ترجمه اکبر فتوحی و فریبا اصغری، کانون نشر علوم، تهران.
- ۳۶ - هومن حیدعلی (۱۳۸۰) «تحلیل داده‌های چند متغیری در پژوهش رفتاری» نشر پارسا، تهران.

## (ب) لاتین:

- 37 - Alpert, Mark. I, and Robert A. Peterson, (1972) "On the Interpretation of Canonical Analysis" Journal of Marketing Research, vol. IX, May 1972, pp. 187-198
- 38 - Bailey, Kenneth D. (1988) "Methods of Social Research" New York.
- 39 - Boot, John C. G. and Cox, Edwin, B. (1979) "Statistical Analysis for Managerial Decisions, McGraw- Hill Publishing Co. Ltd. New Delhi.
- 40 - Cattell, R. B. (1978) "The Scientific Use of Factor Analysis in Behavioral Sciences", Plenum, New York.
- 41 - Cattell, R. B. (1966) "The Scree test for the Number of Factors" Multivariate Behavioral Research. vol.1, April, 1966.
- 42 - Cooley, W. W. and Lohnes, P. R. (1971) "Multivariate Data Analysis" John Wiley and Sons, New York.
- 43 - Crask, M. and W. Perreault (1977) "Validation of Discriminant Analysis in Marketing Research" Journal of Marketing Research, vol. XIV



- 
- 44 - Dennis Child (1973) "The Essentials of Factor Analysis" Holt Rinehart and Winston, New York.
- 45 - Dillion. W. R. and M. goldstein (1984) "Multivariate Analysis" Wiley, New York.
- 46 - Draper, Norman and Harry Smith, (1986) " Applied Regression Analysis" Wiley, Inc. New York.
- 47 - Dumteman, George H. (1989) " Principal Component Analysis" Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, No. 69, pp. 80-87.
- 48 - Ebdon David (1985)"Statistics in Geography" Billing and Sons Ltd, Oxford. UK.
- 49 - Gatty, R. (1966)" Multivariate Analysis for Marketing Research: An Evaluation" Applied Statistics, vol.15, November 1966, pp. 157-75.
- 50 - Ghosh, B. N. (1982)" Scientific Methods and Social Research" Sterling Publishers, Pvt. Ltd. New Delhi.
- 51 - Gopal, M. H. (1984)" An Introduction to Research Procedure in Social Sciences" Asia Publishing House, Bombay.
- 52 - Hair Jose P. F etal. (1990)" Multivariate Data Analysis" Macmillan Publishing Company, New York.
- 53 - Hammond, R. and McCullagh, D. S. (1974)" Quantitative Techniques in Geography: An Introduction, Oxford.
- 54 - Hotelling, H. (1935)"The Most Predicable Criterion" Journal of Educational Psychology, vol. 26: 139-142.
- 
-

- Educational Psychology, vol. 26: 139-142.
- 55 - [Http://www2.chass.ncsu.edu/grson/pa765/canonic.htm](http://www2.chass.ncsu.edu/grson/pa765/canonic.htm). Year 2002.
- 56 - Ferber Robert (1978) "Handbook of Marketing Research" McGraw-Hill, Inc. New York.
- 57 - Foster, J.J. (1993) "Starting SPSS/PC+ and SPSS for windows (a beginner's guide to data analysis), SIGMA Press, Wilmslow, U.K.
- 58 - Frank, R. E, W. F. Massey and D.G. Morrison (1985) "Bias in Multiple Discriminant Analysis" Journal of Marketing Research, vo 22. August 1985, pp. 250-58.
- 59 - Fruchter, Benjamin (1964) " Introduction to Factor Analysis" Princeton, Inc., New York.
- 60 - Jacques Tacq (1997)"Multivariate Analysis Techniques in Social Research, From Problem to Analysis" London.
- 61 - Kachigan, S.K. (1982)"Multivaiate Statistical Analysis" Radius Press, New York.
- 62 - Kothari, C.R. (1994)" Research Methodology: Methods and Techniques" V.S. Johri, New Delhi.
- 63 - Kothari, C. R (1992) "Quantitative Techniques" Pashupati Printers, Delhi.
- 64 - Krishan Gopal and Sundaresan, M. (1991) "Materials Management- An Integrated Approach "Variety Publications, Chandigarh, India.
- 65 - Labert, Z. and R. Durand (1975) "Some Precations in Using Canonical Analysis" Journal of Marketing Research, vol. XII, November, 1975. pp. 468-475.

- 
- 66 - Mahmood Aslam (1993) " Statistical Methods in Geographical Studies"  
Rajesh Publications, New Delhi.
- 67 - Marascuilo, L.A. and Mc Sweeney, M. (1977) "Nonparametric and  
Distribution- free Methods for the Social Sciences,"Monterey, Canada.
- 68 - Maxwell Albert, E. (1987)" Analyzing Qualitative Data" John Wiley and  
Sons' New York.
- 69 - McDonald, R. and H.W. Marsh (1990) "Chossing a Multivariate Model:  
Noncentrality and Goodness of Fit", Psychological Bulletin, 105,  
430-445.
- 70 - Mendenhall, W. (1989) "Statistics for Management and Economics"  
PWS-KENT Publishing Company. USA.
- 71 - Morrison, D. (1987) " Measurment Problems in Cluster Analysis"  
Mangment Science, vol. 4 August, 1987. pp. 39-48.
- 72 - Morrison Donald, G. (1969)" On the Interpretation of Discriminant  
Analysis" Journal of Marketing Research, vol. 6, May 1969, pp. 156-63.
- 73 - Morrison, D. F. (1967)" Multivariate Statistical Methods" McGraw- Hill  
inc. New York.
- 74 - Ott, R. L. (1993) "An Introduction to Statistical Methods and Data  
Analysis" Duxbury, California.
- 75 - Rao, C. R (1985) "Linear Statistical Inference and its Application" John  
Wiley, New York.
- 76 - Robinson, J. P. and P.R. Shaver, (1979) "Measures of Social Psychological  
Attitudes" Institutes of Social Research, The University of Michigan.
- 
-

- 77 - Sharma Subhash (1996) "Applied Multivariate Techniques" John and Sons, Inc. New York.
- 78 - Siegel, S. (1986) "Nonparametric Statistics for the Behavioral Sciences, New York.
- 79 - Spiegel Murray (1972) "Theory and Problems of Statistics" McGraw- Hill Book Company, U.K.
- 80 - Stevens, J. (1986) "Applied Multivariate for the Social Sciences" Hillsdale Publications, New York.
- 81 - Stevens, J.P. (1972) "Four Methods of Analyzing Between Variations of the K-Group MANOVA Problem" Multivariate Behavioral Research, vol. 7. October, 1972. pp. 442-454.
- 82 - Tabachinck, Barbara G. and Linda S. Fidell, (1996) "Using Multivariate Statistics" Harpercollins College Publishers, New. York.
- 83 - Takeuch, K. Yanai, H. and Mukherjee, B.N. (1992) "The Foundations of Multivariate Analysis" Wiley Eastern Ltd. New Delhi.
- 84 - Tryon, R. C. and Bailey, D. E. (1980)"Cluster Analysis" McGraw-Hill New York.

## واژگان

### A

Absolute	مطلق
Absolute Deviation	خطای مطلق
Absolute Frequency	فراوانی مطلق
Acceptance Region	ناحیه قابل قبول
Adjusted	تعدیل شده
Adjusted R Square	ضریب تعیین تعدیل شده
Aggregate	کل، مجموعه
Agglomerative Method	روش تراکمی
Algorithm	الگوریتم
Alternative Hypothesis	فرض مقابل
Analysis	تحلیل
Analysis of Covariance	تحلیل کواریانس
Analysis of variance	تحلیل واریانس
Application	کاربرد

Applied Statistics	آمار کاربردی
Approach	رهیافت
Approximate Eigenvalue	مقدار ویژه تقریبی
Approximate Value	مقدار تقریبی
Arithmetic Mean	میانگین حسابی
Arithmetic Progression	تصاعد حسابی
Assessment	ارزیابی
Association	پیوند
Asymmetric	نامتقارن
Asymmetric population	جامعه نامتقارن
Asymmetry	نامتقارنی
Attribute	صفت کیفی
Average	میانگین
Average Linkage	پیوند متوسط
Axioms	اصول موضوعه
Axis	محور
<b>B</b>	
Backward Elimination Method	روش حذف پس رو
Bar chart	نمودار میله‌ای
Bartlett Test	آزمون بارتلت
Beta coefficient	ضریب بتا
Beta Function	تابع بتا

Bias	اریب
Biserial	دو رشته‌ای
Biserial Correlation	همبستگی دو رشته‌ای
Bivariate Distribution	توزیع دو متغیره
Block Design	طرح بلوکی
Blocking	بلوک بندی
<b>C</b>	
Canonical Analysis	تحلیل کانونی
Canonical Correlation	همبستگی کانونی، همبستگی متعارف
Canonical Decomposition	تجزیه کانونی
Canonical Equation	معادله کانونی
Canonical Variate	متغیر کانونی
Canonical Function	تابع کانونی
Categorical Data	داده‌های طبقه بندی شده
Census	سرشماری
Central Statistic	آماره مرکز ثقل
Central Tendency	گرایش به مرکز
Centriod	مرکز ثقل
Chance	شانس
Chart	نمودار
Chi- square	کای اسکویر، خی دو
Chi- square Test	آزمون کای اسکویر

Chi- square Approximation	تقریب کای اسکویر
Circular	دایره‌ای
Class	رده، طبقه
Classification	رده‌بندی
Classified Data	داده‌های طبقه بندی شده
Cluster	خوشه
Cluster Analysis	تحلیل خوشه‌ای
Cluster Sampling	نمونه‌گیری خوشه‌ای
Clustering	خوشه‌بندی
Clustering Technique	روش خوشه بندی
Cochran's Test	آزمون کوکران
Cochran's Q- statistic	آماره Q کوکران
Code	کد
Coded Data	داده‌های کدبندی شده
Coding	کدگذاری
Coefficient	ضریب
Coefficient of Correlation	ضریب همبستگی
Coefficient of Determination	ضریب تعیین
Coefficient of Variation	ضریب تغییرات
Collection	گردآوری
Collinearity	هم خطی
Combination of Data	ترکیب داده‌ها
Common Difference	تفاضل مشترک



Commonality	مشترکات
Communality	همه داشت، مشترک
Comparison	مقایسه
Complet Linkage	پیوند کامل
Component Analysis	تحلیل مؤلفه‌ای
Composite	مرکب، ترکیبی
Condition	شرط
Confidence level	سطح اطمینان
Conjoint Analysis	تحلیل همپیوند
Constant	ثابت
Constrained	مقید
Contingency Table	جدول توافقی
Contrast	مقایسه مقید
Contrast Analysis	تحلیل مقایسه‌های مقید
Convergence	همگرایی
Correlation	همبستگی
Correspondence Analysis	تحلیل تناظر
Count	شمارش
Count Data	داده‌های شمارشی
Covariance	کواریانس
Covariate	متغیر کمکی
Cramer Statistic	آماره کرامر
Criterion	ملاک، معیار

Critical	بحرانی
Critical Path	مسیر بحرانی
Cronbach $\alpha$	آلفای کرونباخ
Cross- Validation Approach	رهیافت اعتبار تقاطعی
Cutting Point	نقطه برش، نقطه تفکیک
<b>D</b>	
Data	داده‌ها
Data Adjusting	تعدیل داده‌ها
Data Collection	گردآوری داده‌ها
Data Processing	داده‌پردازی
Data Reduction	تلخیص داده‌ها
Datum	داده
Decile	دهک
Decomposition	تجزیه
Deduction	استنتاج
Degree of Freedom	درجه آزادی
Dendrogram	نمودار درختی
Density	چگالی، تراکم
Dependence Techniques	روشهای وابستگی
Dependent Variable	متغیر وابسته
Description	توصیف
Descriptive Statistics	آمار توصیفی

Design	طرح
Design of Experiments	طرح آزمایشها
Determinant	دترمینان
Diagonal	قطر
Diagram	نمودار
Dichotomous Variables	متغیر دو حالتی
Dichotomy	دو حالتی
Discriminant Analysis	تحلیل تشخیصی
Discriminant Functions	تابع تشخیصی
Discriminant Power	توان تشخیصی
Discrimination	تشخیص
Dispersion Index	شاخص پراکندگی
Display	نمایش
Dissimilarity	عدم تشابه
Distribution	توزیع
Divisive method	روش تفکیک پذیری
Dot Diagram	نمودار نقطه‌ای
Dummy Variable	متغیر مجازی، متغیر ظاهری
Duncan's Test	آزمون دانکن
Durbin - Watson Test	آزمون دوربین واتسون

**E**

Editing Statistical Data	ویرایش داده‌های آماری
--------------------------	-----------------------

Effect	اثر
Eigenvalue	مقدار ویژه، مقدار خاص
Elimination Method	روش حذفی
Elimination of Variables	حذف متغیرها
Equalization	برابرسازی
Equation	معادله
Error Function	تابع خطا
Error of Measurement	خطای اندازه‌گیری
Error of Type II	خطای گونه ۲
Estimate	برآورد
Explained Variance	واریانس تبیین شده
Experimental Analysis	تحلیل آزمایشی
Experimental Design	طرح آزمایشی
<b>F</b>	
F- Test	آزمون F
Factor Analysis	تحلیل عاملی
Factorial	عاملی، فاکتوریل
Factorial Design	طرح عاملی
Fisher's Exact Test	آزمون دقیق فیشر
Fitness	برازندگی
Forecasting	پیش بینی
Forward Diffeence	تفاضل پیشرو

Forward Method	روش پیشرو
Frequency Distribution	توزیع فراوانی
Frequency Table	جدول فراوانی
Function	تابع

**G**

Geometric Mean	میانگین هندسی
Geometric Progression	تصاعد هندسی
Graphical Representation of Data	نمایش نموداری داده‌ها
Group Centroid	مرکز ثقل گروه‌ها
Grouping Variable	متغیر گروه‌بندی
Grouped Data	داده‌های گروه‌بندی شده
Guttman Scaling	مقیاس بندی گاتمن

**H**

Helmert Contrast	مقایسه مقید هلمرت
Hierarchical Classification	طبقه بندی سلسله مراتبی
Hierarchical Cluster Analysis	تحلیل خوشه‌ای سلسله مراتبی
Histogram	هیستوگرام
Holdout Sample	نمونه بسط یافته
Homogeneity	همگنی
Homoscedasticity	همواریانسی
Horizontal Axis	محور افقی

Hotelling's Trace	اثر هتلینگ
Hypothesis	فرض، فرضیه
Hypothesis Testing	آزمون فرضیه
<b>I</b>	
Icicle	قندیل
Ideal Point	نقطه ایده آل
Identification code	کد شناسایی
Independent Variable	متغیر مستقل
Indicator	شاخص
Inductive	استقرایی
Inequality	نابرابری، عدم تعادل
Inference	استنباط
Initial Data Analysis	تحلیل آغازین داده‌ها
Interaction	اثر متقابل
Interaction Model	مدل با اثر متقابل
Intercept	عرض از مبدا
Interdependence Methods	روشهای هم وابستگی
Intereffect	بین اثری
Interference	تداخل
Interpretation	تفسیر
Intersect	تقاطع
Interval	فاصله

Inverse Matrix	ماتریس وارون
Investigation	پژوهش
Iteration	تکرار
<b>J</b>	
Joint Distribution	توزیع توأم
Judgment	داوری
<b>K</b>	
Kendall's tau	کندال تائو
K-Means Clutstring	خوشه‌ای K میانگین
Kolmogorov - Smirnov Test	آزمون کولموگروف سیمرنف
Kruskal- Wallis Test	آزمون کروسکال والیس
Kurtosis tau	کشیدگی
<b>L</b>	
Latent	پنهان
Latent Root	ریشه راکد
Latent Structure Analysis	تحلیل ساختار پنهان
Least Powers	کمترین توان
Level of Significance	سطح معنی دار بودن
Levene's Test of Homogeneity of Variances	آزمون همگنی واریانس‌های لون
Likelihood	درست‌نمایی

Likelihood Ratio	نسبت درست‌نمایی
Limit	حد
Linear	خطی
Linear Combination	ترکیب خطی
Linear Dependence	وابستگی خطی
Linear Regression	رگرسیون خطی
Linkage	اتصال
Logistic Regression	رگرسیون لجستیک
<b>M</b>	
Main Effect	اثر اصلی
Mann-Whitney Test	آزمون من ویت‌نی
Match Statistics	آماره‌های جور شده
McNemar's Test	آزمون مک‌نمار
Mean	میانگین
Mean Difference	تفاضل میانگین
Measurement	اندازه‌گیری
Measurement Error	خطای اندازه‌گیری
Median	میانه
Missing Value	مقدار گم شده
Mode	مدء نما
Modeling	مدلسازی
Modification	تغییردهی



Modified	اصلاح شده
Modified Mean	میانگین تعدیل شده
Multidimensional Scaling	مقیاس بندی چند بعدی
Multichotomus	چند بخشی
Multiple Classification	طبقه بندی چندگانه
Multiple Correlation Coefficient	ضریب همبستگی چندگانه
Multiple Discriminant Analysis	تحلیل تشخیصی چندگانه
Multiple Regression	رگرسیون چندگانه
Multivariate	چند متغیره
Multivariate Analysis	تحلیل چند متغیره
Multivariate Analysis of Variance	تحلیل واریانس چند متغیره
Multivariate Distribution	توزیع چند متغیره
<b>N</b>	
Near- Neighbor Estimation	برآورد براساس همسایه نزدیک
Nearest- Neighbor Method	روش نزدیکترین همسایه
Negative Skewness	چولگی منفی
Network Analysis	تحلیل شبکه
Nominal Data	داده‌های اسمی
Nonhierarchical Cluster Procedure	روش خوشه‌ای غیرسلسله مراتبی
Nonparametric Data Analysis	تحلیل داده‌های ناپارامتری
Nonrespondents	پاسخ نداده‌ها
Nonresponse	بی پاسخی

Normal Distribution	توزیع نرمال
Null Hypothesis	فرض صفر
<b>O</b>	
Oblique Rotation	چرخش متمایل
Observation	مشاهده
Odds Ratio	نسبت برتری
One Sample Test	آزمون یک نمونه‌ای
One- Tailed Test	آزمون یک دمی
One- Way Analysis of Variance	تحلیل واریانس یک طرفه
Ordinal Data	داده‌های ترتیبی
Orthogonal Rotation	چرخش متعامد
Outliers	نقاط دور افتاده
Output	خروجی
Overlap	روی هم افتادگی
<b>P</b>	
Paired	جفت شده
Parameter	پارامتر
Parametric Interdependence Techniques	تکنیک‌های هم وابستگی پارامتری
Partial Association	پیوند جزئی
Partial Correlation	همبستگی جزئی
Path Analysis	تحلیل مسیر

Path Coefficients	ضرایب مسیر
Percentile	صدک
Phi- Coefficient	ضریب فی
Pillai's Trace	اثر پیلایی
Polar- extremes approach	رهیافت دو قطبی
Population Variance	واریانس جامعه
Positive Skewness	چولگی مثبت
Post Hoc	پس تجربی
Principal Components Analysis	تحلیل مؤلفه‌های اصلی
<b>Q</b>	
Qualitative Data	داده‌های کیفی
Quartile	چارک
Quasi- Experiment	شبه آزمایشی
Quasi- Factorial Design	طرح شبه عاملی
Questionnaire	پرسشنامه
Quota Sampling	نمونه‌گیری سهمیه‌ای
<b>R</b>	
Random Distribution	توزیع تصادفی
Random Error	خطای تصادفی
Random Number	عدد تصادفی
Random Sampling	نمونه‌گیری تصادفی

Range	دامنه تغییرات
RanK Analysis	تحلیل رتبه‌ای
RanK Test	آزمون رتبه‌ای
RanKing	رتبه بندی
Ratio Test	آزمون نسبت
Rectangular Distribution	توزیع مستطیلی
Rejection	رد
Reliability	اعتماد، پایداری
Research Hypothesis	فرضیه تحقیق
Residual	مانده
Rotation	چرخش
Rounding	گرد کردن
Row	ردیف، سطر
<b>S</b>	
Sample Design	طرح نمونه‌ای
Sample Size	اندازه نمونه، حجم نمونه
Sample Size Determination	تعیین اندازه نمونه
Sample Size Estimation	برآورد حجم نمونه
Sampling	نمونه گیری
Sampling Design	طرح نمونه‌ای
Sampling Method	روش نمونه‌ای
Scale	مقیاس

Scaling	مقیاس‌بندی
Scatter Diagram	نمودار پراکنش
Standard Deviation	انحراف معیار
Standardized Discriminant Weights	وزنهای تشخیصی استاندارد شده
Structure Correlation	همبستگی ساختاری
Structure Matrix	ماتریس ساختاری
Score	امتیاز، نمره
Scoring	امتیازبندی
Secondary Data	داده‌های ثانویه
Sector Diagram	نمودار قطاعی
Sign Test	آزمون نشانه
Significance Level	سطح معنی‌داری
Significant	معنی‌دار
Similarity	تشابه
Simple Random Sampling	نمونه‌گیری تصادفی ساده
Simultaneous	همزمان
Single Linkage	پیوند تکی
Size of a Population	اندازه جامعه، حجم جامعه
Skewed Distribution	توزیع چوله
Skewness	چولگی
Slope	شیب
Social Statistics	آمار اجتماعی
Spearman Rank Correlation Coefficient	ضریب همبستگی رتبه‌ای اسپیرمن

Split- sample	دو نیمه کردن نمونه
Squared Eudclidean Distance	مربع فاصله اقلیدسی
Statistical Software	نرم افزار آماری
Step- Down Method	روش گام به پس
Step- Up Method	روش گام به پیش
Stepwise Regression	روش گام به گام
Stratification	طبقه بندی
Stress	تنش
String	رشته
Structural Model	مدل ساختاری
Student Distribution	توزیع استیودنت
Student's t-Test	آزمون تی استیودنت
Sum	مجموع
Summary	خلاصه
Survey	بررسی، پیمایش
Survey Sampling	نمونه گیری پیمایشی
Survival Analysis	تحلیل بقاء
Symbol	نماد، سمبل
Symmetric	متقارن
Systematic Sampling	نمونه گیری سیستماتیک
<b>T</b>	
t-Statistics	آماره t

Table	جدول
Tabulation	جدول‌بندی
Target Population	جامعه هدف
Technique	فن، تکنیک
Test of Hypothesis	آزمون فرضیه
Test of Normality	آزمون نرمال بودن
Time- Series Analysis	تحلیل سریهای زمانی
Tolerance	قدرت تحمل
Transformation	تبدیل
Tukey's Test	آزمون توکی
Type I Error	خطای نوع ۱
Type II Error	خطای نوع ۲
<b>U</b>	
U- statistics	آماره U
Unequal	نا برابر
Uniformity	یکنواختی
Univariate	تک متغیره
Upper Bound	کران بالا
<b>V</b>	
V- statistics	آماره V
Validity	اعتبار

Value	مقدار
Variable	متغیر
Variance	واریانس
Variance Analysis	تحلیل واریانس
Variate	متغیر
Variation	تغییرات
Variation Ratio	نسبت تغییر
Vector	بردار
<b>W</b>	
Wald Statistic	آماره والد
Weight	وزن
Width	پهنای عرض
Wilk's Lambda	لامبدا ویلکز
Wilcoxon Test	آزمون ویلکاکسون
<b>Y</b>	
Yates' Correction	تصحیح یتس
<b>Z</b>	
Z- Score	نمره z (مقادیر استاندارد)
Z- Transformation	تبدیل به z





### معرفی کتاب:

کتابی که در اینجا معرفی خواهد شد، «پردازش و تحلیل داده ها در تحقیقات اجتماعی-اقتصادی: با استفاده از نرم افزار SPSS» نوشته‌ی «دکتر خلیل کلاحتری» دانشیار دانشگاه تهران است. این کتاب در تهران توسط انتشارات شریف چاپ شده است. اولین بار کتاب در سال ۱۳۸۲ منتشر شده است و کتابی حاضر مربوط به نوبت چاپ دوم در سال ۱۳۸۵ می‌باشد. کتاب داری ۳۸۸ صفحه است که صفحات پایانی آن به منابع و بخش واژگان اختصاص داده شده است. اشکال و جداول زیادی بنابر ماهیت کتاب که در ادامه می‌آید، در آن وجود دارد و بهای کتاب ۴۰۰۰۰ ریال می‌باشد.

همان‌طور که از نام کتاب برمی‌آید، محتوا بحث‌های آماری و تحلیل داده‌ها با آزمون‌های آماری از طریق نرم‌افزار SPSS می‌باشد. در آغاز، فهرست مطالب و پس از آن فهرست اشکال و جداول آورده شده است. کتاب شامل ۱۵ فصل می‌باشد که هر فصل، خود با عناوین مشخص قسمت‌بندی شده است.

قبل از آغاز مطالب اصلی، نویسنده به نوشتن پیش‌گفتاری مبادرت ورزیده است که در آن به طور خلاصه فصل‌های کتاب و محتوای آن را بیان نموده است و از اهمیت رایانه و امکانات آن برای محاسبات آماری در تحقیقات سخن گفته است.

فصل اول کتاب با عنوان «آشنایی با نرم افزار SPSS Win» مشخص شده است. لازم به ذکر است که در این فصل و دیگر فصل‌ها، نویسنده، ابتدا نحوه‌ی شروع به کار با برنامه‌ی SPSS، نحوه‌ی ورود به SPSS، پنجره‌ی Data Editor، اصلاح کردن اطلاعات، روش کپی کردن یا جابجا کردن اطلاعات در صفحه گستر، نحوه‌ی وارد کردن ستون یا سطر خالی در صفحه گستر، نحوه‌ی حذف کردن سطرها یا ستون‌ها، نحوه‌ی حرکت در صفحه گستر، نحوه‌ی پیدا کردن یک متغیر، نحوه‌ی پیدا کردن یک سطر، نحوه‌ی پیدا کردن یک مقدار یا متغیر خاص یا

## پردازش و تحلیل داده‌ها در تحقیقات اجتماعی - اقتصادی با استفاده از نرم افزار SPSS

نویسنده: دکتر خلیل کلاحتری  
ناشر: انتشارات شریف، ۳۸۸ صفحه  
قیمت: ۴۰۰۰ تومان  
فاطمه زنده بودی





به بحث انتخاب آزمون مناسب برای مقایسه‌ی میانگین‌ها می‌پردازد و پس از آن موارد استفاده از آزمون T برای گروه‌های مستقل و وابسته و همچنین آزمون F برای گروه‌های مستقل و وابسته را مشخص می‌سازد. در پایان نیز نحوه‌ی محاسبه‌ی آزمون T و تحلیل واریانس یک‌طرفه و دوطرفه را در نرم‌افزار SPSS توضیح می‌دهد.

فصل ششم با عنوان آزمون‌های غیرپارامتری، شرایط استفاده از این آزمون‌ها را بیان می‌کند. در این فصل آزمون‌های نشانه، مک‌نمار، ویلکسون، فریدمن، کوکران، من ویتنی، کولموگروف سیمونف، کروسکال والیس و آزمون میانه معرفی شده و نحوه‌ی محاسبه‌ی آنها با نرم افزار SPSS مشخص گردیده است.

طبقه‌بندی روش‌های تحلیل چندمتغیره در فصل هفتم در چند صفحه آورده شده که این فصل با همین توضیحات به پایان می‌رسد و نویسنده وارد فصل هشتم با عنوان رگرسیون چندگانه می‌شود. در این فصل تحلیل رگرسیون و ضریب همبستگی و موارد استفاده از آنها به رشته‌ی تحریر در آمده است. انواع روش‌های رگرسیون چندگانه یعنی روش رگرسیون چندگانه‌ی توأم، گام به گام، حذف پس رو و پیشرو و روش محاسبه‌ی آنها با نرم‌افزار SPSS توضیح داده شده است.

فصل نهم درباره‌ی رگرسیون لجستیک و طرز محاسبه‌ی آن به وسیله‌ی SPSS است.

در فصل دهم نیز که تحلیل تشخیصی چندگانه نامیده شده است، پس از مقدمه، نویسنده موارد کاربرد تحلیل تشخیصی و شباهت‌ها و

یک پاسخگوی خاص، نحوه‌ی تعریف کردن متغیرها و مقادیر، نحوه‌ی تعریف کردن متغیرها در SPSS، مقادیر نامعلوم (missing values)، نحوه‌ی تعریف نوع متغیر، نحوه‌ی مرتب کردن ستون‌های متغیرها، تعریف متغیرها با SPSS نسخه‌ی ۱۰ و بالاتر، نحوه‌ی استفاده از templates برای تعریف کردن متغیرها، نحوه‌ی template کردن در SPSS نسخه‌ی ۱۰ و بالاتر، نحوه‌ی ذخیره کردن اطلاعات در فایل، نحوه‌ی ذخیره کردن اطلاعات در دفعات بعدی، نحوه‌ی بازیابی یک فایل اطلاعاتی، نحوه‌ی بستن برنامه‌ی SPSS، پنجره‌ی output، نحوه‌ی ذخیره کردن یک فایل output، نحوه‌ی بازیابی یک فایل output از فلاپی، نحوه‌ی ویرایش محتوای پنجره‌ی output، نحوه‌ی پرینت کردن و تسهیلات help را بیان می‌کند. به طور کلی ابتدایی‌ترین و البته ضروری‌ترین مطالب درباره‌ی استفاده از این نرم‌افزار یعنی SPSS را بیان می‌کند.

فصل دوم مربوط به پردازش داده‌هاست؛ که پس از مقدمه‌ای کوتاه، در مورد پردازش داده‌ها که خود شامل ویرایش داده‌ها، کدگذاری، استخراج و ورود داده‌ها به رایانه، تعریف داده‌ها در رایانه و کدگذاری مجدد است، توضیح می‌دهد. سپس کدگذاری مجدد به وسیله‌ی نرم‌افزار SPSS، حساب کردن، نحوه‌ی انجام محاسبات با شرط، نحوه‌ی انتخاب زیرگروه‌ها برای تجزیه و تحلیل - الف - انتخاب زیرگروه دائمی، ب - انتخاب زیرگروه موقتی - نحوه‌ی انتخاب نمونه‌ی تصادفی (selecting random sample) و آزمون پردازش داده‌ها مورد بررسی قرار می‌گیرد.

عنوان فصل سوم توصیف داده‌هاست. پس از بیان مقدمه، نویسنده در مورد جداول توزیع فراوانی، توصیفی کوتاه می‌دهد و پس از آن، ابتدا شاخص‌های مرکزی چوی میانگین، میانه و مد را توضیح می‌دهد و سپس شاخص‌های پراکندگی نظیر دامنه، پراکندگی دهک، ضریب انحراف متوسط و انحراف معیار را مطرح می‌سازد. همچنین در مورد چولگی و درجه‌ی کشیدگی نمودار نیز تعاریفی ارائه می‌کند. در ادامه در ذیل عنوان آمار توصیفی، سطوح اندازه‌گیری اسمی، ترتیبی، فاصله‌ای و نسبی را معرفی می‌کند و در پایان نحوه‌ی محاسبه‌ی جداول توزیع فراوانی، شاخص‌های مرکزی و پراکندگی با نرم‌افزار SPSS را توضیح می‌دهد.

تحلیل روابط بین متغیرها عنوان فصل چهارم است. در این فصل مشخص می‌شود که با توجه به سطوح مختلف اندازه‌گیری در هر رابطه‌ای، چه آزمونی برای تبیین و تحلیل رابطه مناسب‌تر است. سپس به نحوه‌ی استفاده از هر یک از این آزمون‌ها از طریق نرم‌افزار SPSS می‌پردازد. به عبارت دیگر روش یافتن این آزمون‌ها و مسیر یافتن آنها در این نرم‌افزار و اعمال مقاصد مورد نظر بر روی داده‌ها را بیان می‌کند. عنوان فصل پنجم مقایسه میانگین‌هاست که در ابتدا به طور کوتاه

شده است. تیتز مطالب به صورت BOLD تایپ شده است که این خود در یافتن سریع‌تر مطالب راهگشاست. البته در کتاب به شیوه‌ی پاورقی ارجاع داده شده که بهتر می‌بود ارجاع، بلافاصله پس از پایان مطلب در یک پرانتز با نام نویسنده، سال انتشار و صفحه‌ی مربوط آورده شود و در پایان کتاب، نام کامل اثر در قسمت منابع ذکر شود.

از آنجا که کتاب مورد نظر، بیشتر با فرمول‌های آماری اثبات شده و نرم‌افزار SPSS سروکار دارد، از قضاوت‌های ارزشی به دور مانده است و سوگیری خاصی در آن مشاهده نمی‌شود. نحوه‌ی چینش فهرست، صحیح می‌باشد و مطالب در فهرست به زیرگروه‌های مختلف تقسیم شده است تا خواننده به راحتی به مطلب مورد نظر خود دست یابد. علاوه بر این فهرست، اشکال و جداول نیز به صورت جداگانه تهیه شده که خود از مزایای این اثر می‌باشد. از طرف دیگر ترتیب مطالب مناسب به نظر می‌رسد؛ در همه‌ی فصول نویسنده از توصیف به سمت تحلیل در حرکت است و ابتدا روش و آزمون مورد نظر را تعریف می‌کند و سپس به زمان استفاده از آن اشاره می‌کند. ترتیب مطالب به گونه‌ایست که خواننده‌ی مبتدی نیز می‌تواند با مطالب کتاب همگام شده و آنها را درک کند. نویسنده در هر فصل دقیقاً آنچه را که در عنوان به آن اشاره کرده است، تشریح می‌کند؛ در واقع تیتز مطلب به خوبی بیانگر محتوای آن مطلب است.

نویسنده برای توضیح مطالب مورد نظر بلافاصله پس از شرح موضوع، فرمول و جدول مربوطه را قرار داده است. همچنین در مورد محاسبه‌ی آزمون‌ها به وسیله‌ی نرم‌افزار SPSS نیز مسیر دقیق دست‌یابی به آزمون را توضیح داده و سپس تصویر صفحه‌ی مربوط به آن در نرم‌افزار SPSS را در همان صفحه یا صفحه‌ی بعد چاپ کرده است. حتی نتایج حاصل از آزمون‌ها یعنی "OUTPUT" ها نیز در کتاب چاپ شده است.

فرمول‌های به کار گرفته شده، تشریح شده است و مشخص گردیده که هر حرف لاتین یا ریاضی در فرمول نمایانگر چیست و چگونه محاسبه می‌شود.

از نظر محتوا نیز علاوه بر آنکه فرمول‌ها، جدول‌ها و تصویرهای مفید در کتاب به تفهیم مطالب کمک کرده است، مثال‌های مناسبی نیز در جهت تفهیم این فرمول‌ها و آزمون‌ها به کار گرفته شده و همانطور که از نام کتاب برمی‌آید این مثال‌ها در زمینه‌های اقتصادی و اجتماعی است که این نیز از نقاط قوت این کتاب به شمار می‌آید. در کل نویسنده به خوبی توانسته است در چپتی که نام کتاب می‌طلبد، در حرکت باشد. یعنی در پایان خواننده تا حدود زیادی شیوه‌ی توصیف، پردازش و تحلیل داده‌ها را فرامی‌گیرد. البته قابل انکار نیست که این کتاب نیز مانند هر کتاب دیگری نمی‌تواند خواننده را از خواندن مطالب کتاب‌های دیگر در این زمینه بی‌نیاز کند.

تفاوت‌های بین تحلیل تشخیصی و تحلیل رگرسیون و تحلیل واریانس را معرفی می‌کند. سپس مراحل انجام تحلیل تشخیصی، نحوه‌ی استفاده از این تحلیل در SPSS و تفسیر نتایج در این روش را تشریح می‌کند.

نویسنده تحلیل مسیر را در فصل یازدهم پی می‌گیرد و اصول تحلیل مسیر، تجزیه‌ی همبستگی و محاسبه‌ی تحلیل مسیر به وسیله‌ی SPSS را توضیح می‌دهد.

پس از آشنایی با تحلیل مسیر در فصل دوازدهم، خواننده تحلیل واریانس چندمتغیره را فرامی‌گیرد. این که اساساً تحلیل واریانس چندمتغیره چیست؟ تفاوت بین دو گروه مستقل و رهیافت‌های یک و چندمتغیره و آزمون‌های متناسب با آن چیست؟ در این فصل تحلیل کواریانس (ANOVA) و تحلیل کواریانس چندمتغیره (MANOVA) و محاسبه‌ی آنها با نرم‌افزار SPSS توضیح داده شده است.

در فصل سیزدهم با عنوان تحلیل همبستگی کانونی، اهداف این نوع همبستگی و نحوه‌ی استنتاج توابع کانونی مورد بحث قرار گرفته و در ادامه روش محاسبه‌ی تحلیل کانونی با برنامه‌ی SPSS و تفسیر نتایج به دست آمده از آن، پی گرفته می‌شود.

تحلیل عاملی عنوان آشنای فصل چهارم است که طبق معمول با مقدمه‌ای کوتاه وارد بحث می‌شود. نگارنده پس از بیان بنیان‌های اولیه‌ی تحلیل عاملی، تحلیل عاملی را تعریف کرده و فرایند آن را توضیح می‌دهد و این که نمودار تصمیم‌گیری در تحلیل عاملی چیست؟ رهیافت‌های استخراج ماتریس همبستگی، تحلیل عاملی مشترک و تحلیل مؤلفه‌های اصلی، عامل‌های چرخش یافته، معیارهای استخراج تعداد عامل‌ها و در کل روش انجام این نوع تحلیل و تفسیر نتایج آن و در پایان روش محاسبه‌ی آن با نرم‌افزار SPSS از دیگر مطالب این فصل است.

نویسنده در فصل پایانی، تحلیل خوشه‌ای را تعریف کرده، بیان گرافیکی و فرآیند انجام این تحلیل را مشخص می‌سازد. روش خوشه‌ای سلسله‌مراتبی و غیر سلسله‌مراتبی و فرآیند آن تشریح شده است. در ادامه نیز طرز محاسبه‌ی تحلیل خوشه‌ای سلسله‌مراتبی با برنامه‌ی SPSS بیان شده است.

صفحات پایانی کتاب نیز اختصاص به معرفی منابع و واژگان دارد.

#### نقد کتاب:

غلط‌های املائی متن کتاب بسیار کم بوده و آن تعداد معدود به تفهیم مطالب کتاب خدشه‌ای وارد نساخته است. فونت مورد استفاده در کتاب متداول و مناسب است. همچنین فاصله‌ی میان سطرها نیز به خوبی رعایت شده است و باعث کند شدن فرآیند خواندن برای خواننده نمی‌شود. دستور زبان فارسی در نگارش جملات کتاب به خوبی رعایت

