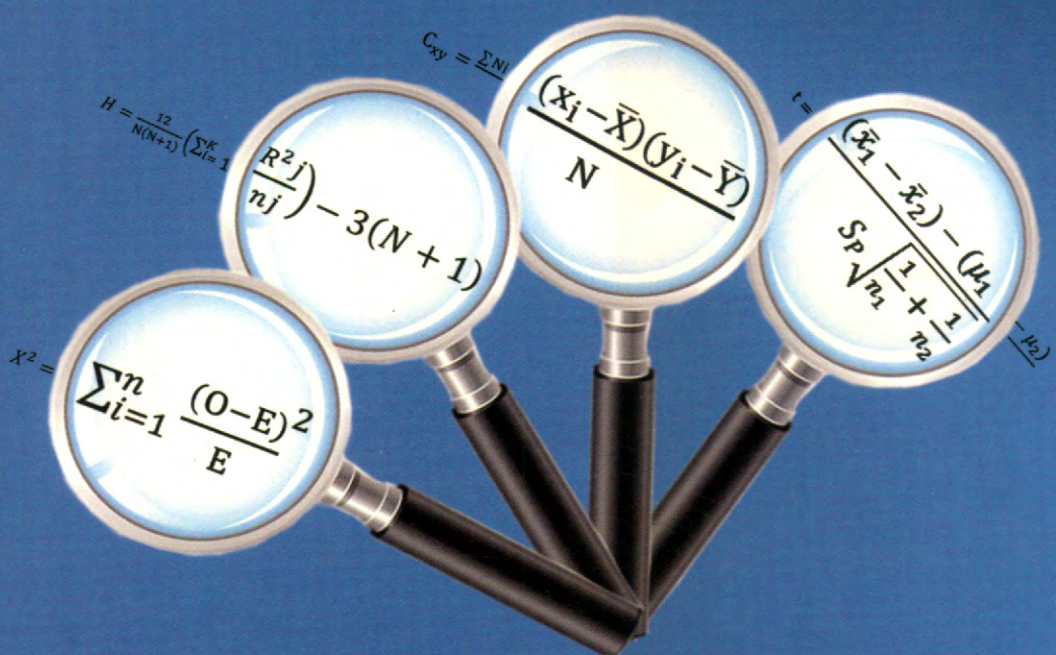
پژوهشگاه مهندسی
بحران های طبیعی و پدافند غیر عامل
شاخص پژوهش

آزمون فرضیه های پژوهش با استفاده از spss



دکتر مسعود تقوایی
عضو هیأت علمی دانشگاه اصفهان
محمدرضا بسحاق
اسماعیل سالاروند

پژوهشگاه در یک نگاه

۱- حضور ۵۰ عضو هیئت علمی و ۲۰۰ پژوهشگر جوان و خلاق در گروه های مختلف پژوهشکده

۲-دارای مجوز آموزش حین خدمت کارکنان دولت در مباحث پدافند غیرعامل و مدیریت بحران از وزارت علوم، تحقیقات و فن آوری

۳- پذیرش دانشجویان دکتری پژوهش محور در گرایش های عمران، مدیریت و جغرافیا با مجوز وزارت علوم ، تحقیقات و فن آوری

۴-برخورداری از آزمایشگاه های کاملاً تخصصی در زمینه های عمران، ژئوتکنیک و سازه

۵- چاپ و انتشار بیش از ۲۵ جلد کتاب تخصصی به زبان های فارسی و انگلیسی

۶- تهیه و تدوین دستورالعمل ها و استانداردهای تخصصی در حوزه های مدیریت بحران، پدافند غیرعامل و فن آوری های نوین برای سازمان ها و ادارات

۷-انتشار دو مجله علمی پژوهشی آمایش و پدافند غیرعامل و مدیریت ریسک و پدافند غیر عامل

۸- تاسیس انجمن آمایش سرزمین ایران

۹- برگزاری ۶ کنفرانس ملی و بین المللی

۱۰- چاپ و انتشار بیش از ۴۰ مقاله در مجلات معتبر و کنفرانس های داخلی و خارجی

۱۱-اجرای بیش از ۱۰۰ پروژه ی تحقیقاتی، مشاوره ای و آموزشی

۱۲-حمایت بیش از ۱۰۰ عنوان طرح تحقیقاتی و پایان نامه ی دانشجویی

۱۳-اجرای بیش از ۳۰ طرح پژوهشی در صنایع فولاد، حمل و نقل، نفت و گاز و برنامه ریزی شهری و منطقه ای

۱۴-اجرای بیش از ۳۰۰۰۰ نفر ساعت دوره های آموزشی تخصصی

۱۵- برگزاری بیش از ۲۰۰ کارگاه آموزشی در مباحث مدیریت بحران و پدافند غیرعامل در سراسر کشور

۱۶-طراحی ماشین آلات و تجهیزات مورد نیاز کشور در صنعت ساختمان همچون اولین دستگاه برش ملات در کشور برای آزمایش های مقاوم سازی ساختمان های بنایی

۱۷- دارای قراردادهای فعال علمی، آموزشی و مشاوره ای با ادارات و سازمان های مختلف

۱۸- خلق و توسعه ی فرمول ارزیابی سریع لرزه ای ساختمان ها

۱۹- طراحی و تدوین مدل های کاربردی مدیریت بحران

۲۰- حضور فعال در کنفرانس ها و نمایشگاه های بین المللی تخصصی در خارج از کشور

۲۱- عقد قراردادهای همکاری علمی و پژوهشی با دانشگاه های داخلی و خارجی

www.shakhespajouh.ir

Email:info@shakhespajouh.ir

پژوهشگاه معماری
و شهرسازی
شاخص پژوهه

تلفن: ۵-۲۶۱۶۲۴۳-۰۳۱۱

بسم الله الرحمن الرحیم

آزمون فرضیه‌های پژوهش
با استفاده از SPSS

دکتر مسعود تقوایی
محمدرضا بسحاق
اسماعیل سالاروند

سرشناسه	:	تقوایی، مسعود،
عنوان و نام پدیدآور	:	آزمون فرضیه‌های پژوهش با استفاده از SPSS مسعود تقوایی،محمدرضا بسحاق، اسماعیل سالاروند.
مشخصات نشر	:	اصفهان: انتشارات معظمی،۱۳۹۱.
مشخصات ظاهری	:	۱۶۸ص .: مصور، جدول.
شابک	:	۹۷۸-۶۰۰-۹۳۲۴۷-۶-۷
وضعیت فهرست نویسی	:	فیپا
یادداشت	:	کتابنامه.
موضوع	:	اس. پی. اس. اس تحت ویندوز
موضوع	:	آمار -- برنامه‌های کامپیوتری
شناسه افزوده	:	بسحاق، محمدرضا،
شناسه افزوده	:	سالاروند، اسماعیل،
رده بندی کنگره	:	HA ۲۹/۵/۲ت۶۶ ۱۳۹۱
رده بندی دیویی	:	۰۰۱/۴۲۲
شماره کتابشناسی ملی	:	۳۰۹۸۷۷۶

آزمون فرضیه های پژوهش با استفاده از SPSS

- نویسندگان: مسعود تقوایی، محمد رضا بسحاق، اسماعیل سالاروند
- ناشر : معظّمی
- مدیر وناظر فنی: علیرضا جنّت
- طراحی جلد: نگار اصفهان
- صفحه آرایی: مرضیه ضیائی
- چاپ : نگار اصفهان
- قطع اثر: وزیری
- نوبت چاپ : اول ۱۳۹۱
- تیراژ : ۱۰۰۰ جلد
- قیمت : ۸۵۰۰۰ ریال

شابک : ۹۷۸-۶۰۰-۹۳۲۴۷-۶-۷

کلیه حقوق نویسندگان اثر محفوظ می باشد.

ISFAHANNEGAR

E- mail: pubnegar@yahoo.com Cellphone: 09131703597

فهرست مطالب

عنوان **صفحه**

فصل اول: کلیات نرم افزاری

۱-۱. معرفی Spss ۹

۱-۲. معرفی منوها ۱۱

۱-۳. ورود داده ها ۱۲

۱-٤. ذخیره کردن فایلها در SPSS ۱٦

۱- ٥. باز کردن فایل ها در SPSS ۱۸

۱-٦. دستور compute ۱۹

فصل دوم: مفاهیم پایه ای آمار

۲-۱. تعریف علم آمار ۲۱

۲-۲. انواع آمار ۲۲

۲-۲-۱. آمار توصیفی ۲۲

۲-۲-۲. آمار استنباطی ۲۲

۲-۲-۳. آمار ناپارامتریک ۲۲

۲-۳. اصطلاحات آماری ۲۳

۲-۳-۱. جامعه ۲۳

۲-۳-۲. نمونه ۲۳

۲-۳-۳. متغیر ۲۳

۲-۳-٤. فرضیه ۲٦

۲-۳-٥. آزمون فرضیه ۲۷

۲-۳-٦. مراحل آزمون فرضیه ۲۸

۲-۳-٦-۱. تعیین فرضیه ها ۲۸

۲-۳-٦-۲. تعیین آماره آزمون مناسب ۲۸

۲-۳-٦-۳. تعیین ناحیه رد یا مرز اهمیت ۲۸

۲-۳-٦-٤. تعیین یک طرفه یا دو طرفه بودن آزمون ۲۹
۲-۳-٦-٥. محاسبه مقدار آماره آزمون ۳۰
۲-۳-٦-٦. نتیجه گیری ۳۰
۲-٤. شاخص های توصیف داده ها ۳۰
۲-٤-۱. شاخص های مرکزی ۳۰
۲-٤-۲. شاخص های پراکندگی ۳۲
۲-٤-۳. شاخص های شکل توزیع ۳۳
۲-٥. تعیین شاخص های توصیف داده ها به کمک SPSS ۳۷
۲-٥-۱. دستور Frequencies ۳۷
۲-٥-۲. دستور Descriptives ۳۹

فصل سوم: رسم نمودارها

۳-۱. نمودار ستونی (میله ای) ٤۳
۳-۲. نمودار خطی ٤٦
۳-۳. نمودار ستونی خطا ٤۸
۳-٤. نمودار پراکنش ٥۰
۳-٥. نمودار هیستوگرام (بافت نگار) ٥۲

فصل چهارم: نرمال بودن توزیع ها

٤-۱. مفهوم Significance ٥٥
٤-۲. نرمال بودن توزیع ها ٥٥
٤-۲-۱. آزمون تک متغیری مجذور کای(X^2) خی ۲ ..... ٥٦
٤-۲-۲. آزمون کولموگراف- اسمیرنوف(KS) ٦۱

فصل پنجم: آزمون های میانگین

٥-۱. آزمون تی تک نمونه ای ٦٥

۵-۲. آزمون تی مستقل(آزمون میانگین های دو نمونه مستقل) ٦۹
۵-۳. آزمون تی وابسته(آزمون میانگین های نمونه های جفتی) ۷۳
۵-٤. آزمون مقایسه میانگین چند جامعه(تحلیل واریانس) ۷۷

فصل ششم: آزمونهای ناپارامتری

٦-۱. آزمون علامت زوج نمونه ای ۸۳
٦-۲. آزمون مک نمار ۸٦
٦-۳. آزمون دو جمله ای ۹۰
٦-٤. آزمون من- ویتنی U ۹۳
٦-۵. آزمون ویلکاکسون ۹٦
٦-٦. آزمون کروسکال- والیس(آزمون H) ۱۰۰
٦-۷. آزمون فریدمن ۱۰۳

فصل هفتم: همبستگی

۷-۱. انواع همبستگی ۱۰۷
۷-۱-۱. همبستگی مستقیم کامل ۱۰۷
۷-۱-۲. همبستگی مستقیم ناقص ۱۰۷
۷-۱-۳. همبستگی معکوس کامل ۱۰۸
۷-۱-٤. همبستگی معکوس ناقص ۱۰۸
۷-۱-۵. عدم همبستگی ۱۰۸
۷-۲. ضریب همبستگی پیرسون ۱۰۹
۷-۳. ضریب همبستگی اسپرمن ۱۱۱
۷-٤. ضریب همبستگی تاو _ بی کندال ۱۱٤

فصل هشتم: رگرسیون

۸-۱. رگرسیون خطی ساده ۱۱۷

۸-۲. رگرسیون خطی چند گانه یا چند متغیری ۱۲۲
۸-۳. رگرسیون لجستیک ۱۲٦

فصل نهم: تحلیل عاملی

۹-۱. شاخص KMO و آزمون بارتلت ۱۳٤
۹-۲. اجرای تحلیل عاملی (اکتشافی) ۱۳٤

فصل دهم: تحلیل خوشه ای

۱۰-۱. الگوریتم های عمومی مورد استفاده در روش تحلیل خوشه ای ۱٤٥
۱۰-۲. اجرای تحلیل خوشه ای ۱٤٦

فصل یازدهم: روایی و پایایی

۱۱-۱. روایی ۱٥٥
۱۱-۲. روش های تعیین روایی ۱٥٥
۱۱-۳. پایایی ۱٥٦
۱۱-٤. روش های تعیین پایایی ۱٥٦
منابع و مآخذ ۱٦٤

پیشگفتار

با کاربرد گسترده رایانه، امروزه استفاده از روشها و تحلیل های پیچیده در تحقیقات بیش از پیش گردیده است ولی باید دانست زمانی می توان به نتایج مورد انتظار در یک تحقیق رسید که بدرستی فرضیه سازی کرده و پس از گردآوری اطلاعات با روش علمی داده ها را مورد پردازش و تجزیه و تحلیل قرار داد. بنابراین اگرچه کاربرد روشهای آماری در تحقیقات روبه گسترش نهاده اما بدلیل عدم آشنایی با فلسفه و زمینه های بکارگیری این روشها، کاربردهای نادرست این تکنیک ها نیز افزایش یافته است. به همین دلیل این کتاب درصدد است تا به معرفی و تبیین و نحوه محاسبه و تفسیر نتایج حاصل از روشهای مختلف آماری بپردازد.

یکی از ویژگی های کتاب حاضر سادگی آن است. در اغلب کتاب هایی که درباب نرم افزار SPSS نگاشته شده است، گستردگی و پراکندگی مطالب سبب گردیده که مطالب اصلی و مهم کمتر مورد توجه قرار گیرند. در این کتاب سعی شده از بیان مطالب غیر ضروری اجتناب گردد.

ویژگی دیگر کتاب حاضر طرح مثال هایی همراه با آزمون فرضیه است. نکته بارزی که در اغلب تحقیقات اجتماعی- اقتصادی چه در سطح مقالات و چه در سطح پایان نامه ها دیده می شود، نا آشنایی دانشجویان به ساختن فرضیه های آماری (H_0 و H_1) در تحقیقات است. در این کتاب سعی شده ضمن تشریح آزمون ها، نحوه ساختن فرضیه در هر آزمون با طرح مثال هایی ساده و کاربردی بیان گردند.

اگرچه در این اثر سعی شده تا با استفاده از منابع متعدد داخلی و خارجی مطالب کامل و بدون نقص باشد ولی بدون شک کاستی هایی در آن وجود دارد که امید است اساتید، دانشجویان و محققان محترم خطاها و نارسائیهای موجود در آن را به نگارندگان گوشزد نمایند تا در چاپ های بعدی مورد توجه قرار گیرد.

فصل اول

کلیات نرم افزاری

۱-۱. معرفی Spss

با پیشرفت علوم و گسترش تکنولوژی، اهمیت استفاده از روش های آماری در علوم مختلف بیش از پیش مورد توجه قرار گرفته است و آموختن آمار کاربردی در هر رشته جزء ملزومات گردیده است. فرآیند آنالیز آماری کمک می کند تا پژوهشگر بتواند از داده‌ها ی اولیه، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد.

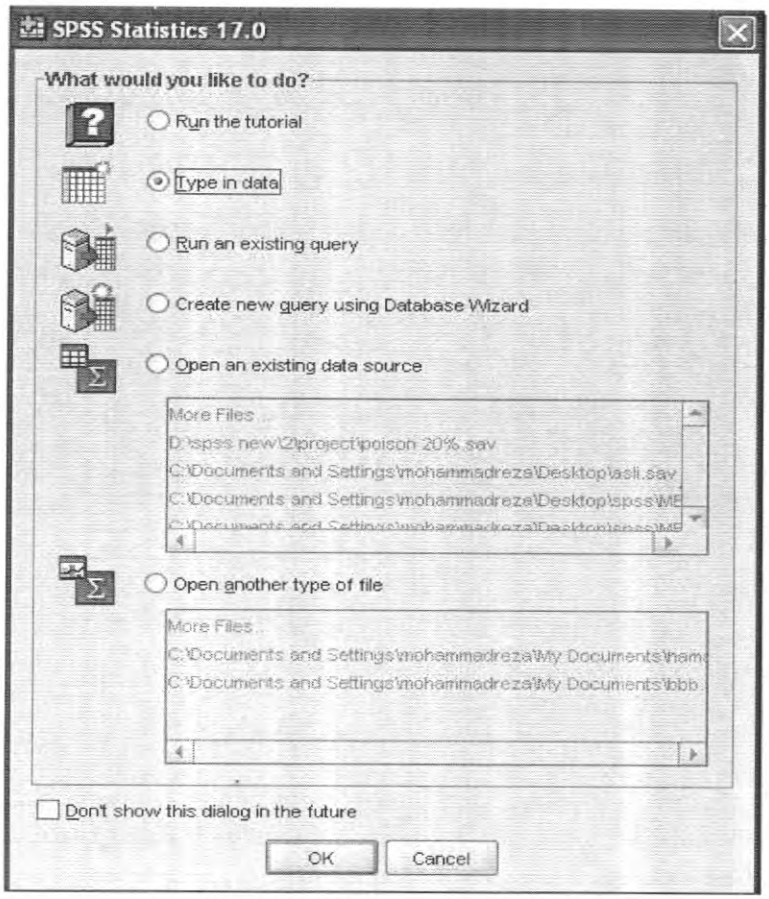
SPSS (مخفف Statistical Package for Social Sciences) یکی از تواناترین و جامع ترین نرم افزارهای آماری است که به دلیل سادگی و پوشش نسبتاً مطلوب روش-های آماری، به یکی از پرکاربردترین نرم افزارهای آماری در کشور تبدیل شده است. برای استفاده از این نرم افزار آشنایی کلی با محیط سیستم عامل ویندوز کافی است بنابراین ما تنها به نحوه اجرای SPSS در محیط ویندوز اشاره می کنیم.

برای ورود به محیط نرم افزار SPSS در صورت وجود Icon نرم افزار SPSS در Desktop روی آن دوبار کلیک می کنیم. یا از طریق مسیر زیر برنامه را فراخوانی می‌کنیم.

Start → All programs → SPSS Inc → Statistics 17.0 → SPSS Statistics 17.0

پس از اجرای SPSS پنجره ای مانند شکل(۱-۱) باز می شود و از شما می پرسد مایل به انجام چه کاری هستید؟ پس از انتخاب گزینه Type in data (ورود داده ها) OK را

کلیک کنید. اگر بر روی گزینه Cancel نیز کلیک کنید می توانید وارد صفحه ورود داده‌ها گردید.



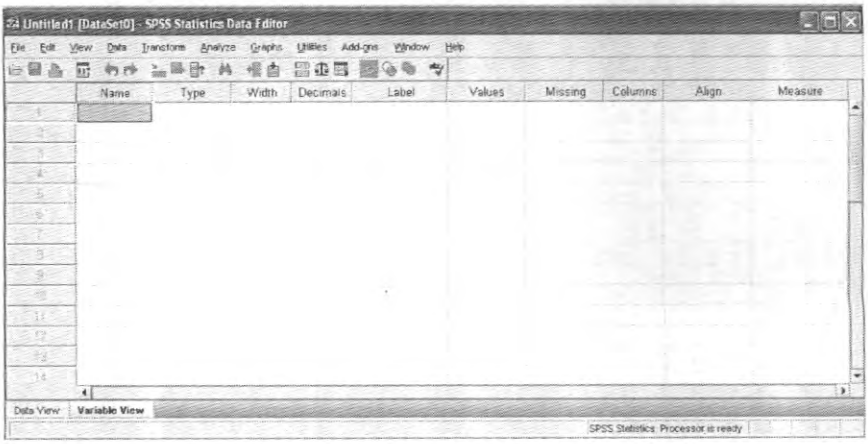
شکل(۱-۱) پنجره ورودی SPSS Statistics 17.0

سپس پنجره ای به نام Untitled-SPSS Editor باز می شود که شامل دو پنجره مختلف به نام های ذیل است:

۱) Data View (نمای داده): برای وارد کردن داده ها و در زیر ستون های معرفی شده استفاده می شود(شکل ۱-۲).

(شکل۱-۲) پنجره Data view

۲) **Variable view** (نمای متغیر): برای تعریف کردن متغیرها استفاده می‌شود (شکل۱-۳).



شکل(۱-۳) پنجره Variable view

۱-۲. معرفی منوها

منوی File: امکان گشودن، ذخیره کردن، پرینت گرفتن، بستن فایل و دسترسی به سایر فایل های موجود را فراهم می کند.

منوی Edit: امکان انجام کارهایی مثل کپی کردن، بریدن و چسباندن را می دهد.

منوی view: از این منو برای آرایش پنجره SPSS استفاده می شود. به عبارت دیگر با استفاده از گزینه های این منو می توان نوار ابزار، نوار وضعیت، خط تجهیزات و غیره را پنهان یا آشکار سازی کرد.

منوی Data: این منو برای مواردی همچون تعریف ویژگی متغیرها، مرتب سازی متغیرها، کار با الگوها، ترکیب و جمع کردن پرونده ها یا فایل های مختلف و وزن دهی به موردها استفاده می شود.

منوی Transform: به ما امکان می دهد متغیرهای جدیدی را محاسبه کنیم و تغییراتی را روی متغیرهای موجود اعمال کنیم.

منوی Analyze: کلیه پردازش های آماری از طریق این منو صورت می پذیرد. این منو ابزارهای مختلفی دارد که در مقوله های مختلفی طبقه بندی شده اند و در فصول آتی به تشریح آن خواهیم پرداخت.

منوی Graphs: این منو حاوی دستورهایی جهت ترسیم نمودارها و ویرایش آنها می‌باشد.

منوی Utilities: اطلاعاتی درباره متغیرها و فایل ها را در اختیار ما قرار می دهد.

منوی Add-ons: این منو اطلاعاتی را در مورد برنامه های دیگری از خانواده SPSS مانند Amos، Clementine و غیره به دست می دهد.

منوی Windows: از دستورهای این منو برای فعال کردن پنجره خاص و کوچک کردن صفحه کاربرگ استفاده می شود.

منوی Help: این منو امکان استفاده از راهنمای نرم افزار SPSS را فراهم می سازد.

۱-۳. ورود داده ها

در حالت کلی، روش معمول این است که ابتدا با استفاده از پنجره Variable view متغیرها را تعریف کرده و سپس با استفاده از پنجره Data view اعداد را به SPSS وارد کنیم.

بنابراین براساس این روش شروع به تعریف متغیرها در پنجره Variable view می‌نماییم. در هر سطر از این پنجره یک متغیر تعریف می شود و هر ستون آن بیانگر ویژگی خاصی در ارتباط با متغیر است (شکل ۱-۳) که در ادامه به توضیح هر یک می پردازیم.

ستون Name (وارد کردن نام متغیر): در این قسمت نام متغیر را وارد می کنیم .لازم به ذکر است که نام متغیر در نرم افزار SPSS مشمول محدودیت هایی است که در وارد کردن آن باید آنها را ذکر کرد:

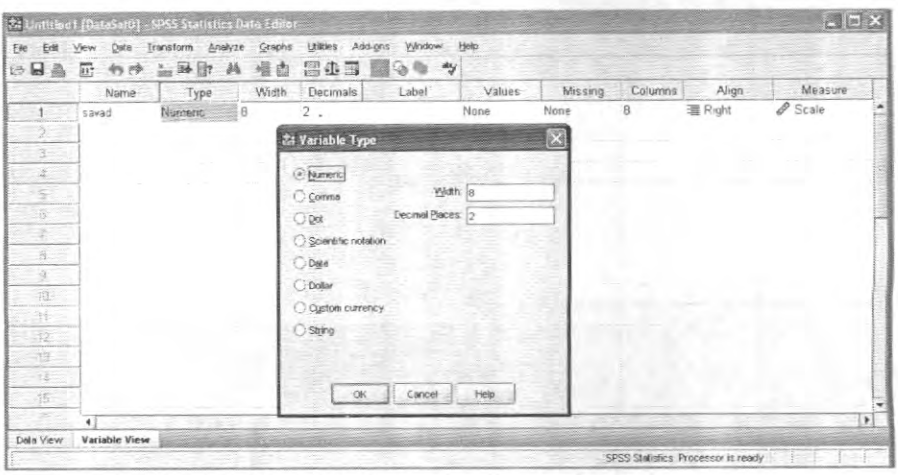
۱) باید با حرف شروع شود.

۲) نمی تواند شامل فاصله باشد.

۳) نباید با یک نقطه تمام شود.

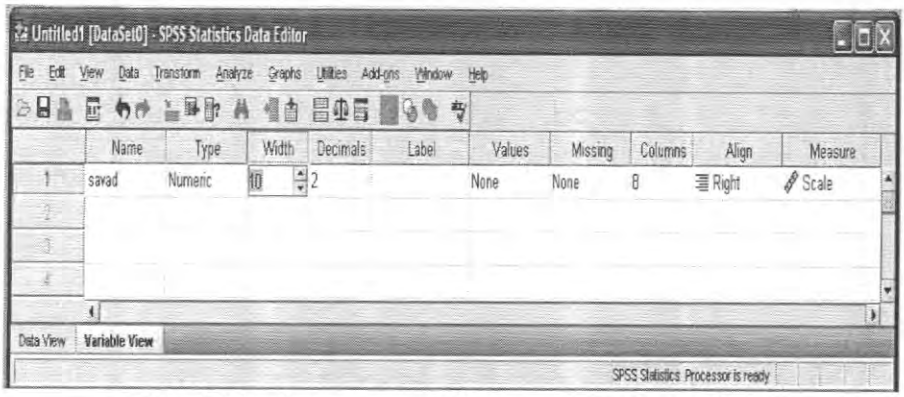
۴) کلمات کلیدی مانند: ALL ، AND ، BY و... که نرم افزار SPSS از آنها به عنوان عبارت های محاسباتی استفاده می کند، نمی تواند به تنهایی به عنوان اسم متغیر قرار گیرد.

ستون Type: این ستون به ما امکان می دهد نوع متغیر را تعریف کنیم. البته نرم افزار SPSS به صورت پیش فرض گزینه متغیر را "عددی" در نظر می گیرد (شکل ۱-۴) ولی انواع دیگر متغیرها نیز در این نرم افزار وجود دارد که بسته به نوع متغیر می توان از آن ها استفاده کرد. غیر از دو نوع Numeric (متغیرهای عددی مثل: قد، وزن، تعداد واحد و ...) و String (متغیرهای رشته ای مثل: اسم، گروه خونی، شغل و...) سایرانواع متغیرها کاربرد چندانی ندارند و در موارد بسیار خاصی به کار می روند.



شکل (۱-۴) پنجره Variable Type

ستون Width: این ستون برای تغییر دادن پهنای متغیر استفاده می شود. با کلیک روی ستون، دو پیکان کوچک بالا و پایین نشان داده می شود که می توانیم با بالا و پایین کردن، پهنای مورد نظر را تغییر دهیم (شکل ۱-۵).



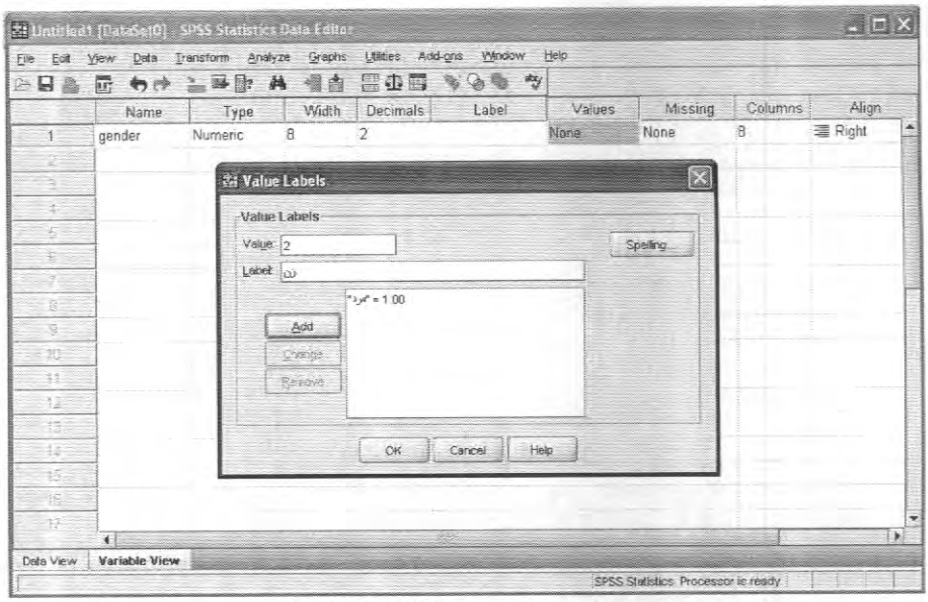
شکل (۱-۵) کادر ستون Width در پنجره Variable view

ستون Decimals: برای تعیین تعداد ارقام اعشار داده های مورد نظر استفاده می شود. مانند روش قبل روی ستون کلیک کرده و تعداد ارقام اعشار را تعیین می کنیم.

ستون Label: در این ستون می توان برای متغیر مورد نظر یک برچسب انتخاب کرد. که این برچسب می تواند ظاهر خروجی یک دستور را بهبود بخشد.

ستون Values: این ستون برای متغیرهای گروه بندی مورد استفاده قرار می گیرد. مانند متغیر جنسیت که می توان آن را صورت کدبندی بیان کرد.

بدین منظور با کلیک روی عبارت None صفحه ایی برای شما باز می شود تا بتوانید متغیرهای خود را کدبندی کنید، همانند شکل (۱-۶) در قسمت Value کد را وارد کنید سپس عنوان متغیر را در قسمت Label تایپ کنید پس از انجام این کار کلید Add برای شما فعال می گردد، روی آن کلیک کنید تا عنوان متغیر با کد مربوطه برای نرم افزار تعریف شود. در پایان کار نیز روی گزینه Ok کلیک کنید.



شکل (۱-۶) پنجره Value Labls

ستون Missing: مربوط به داده های گمشده می باشد و مقادیری هستند که محاسبات روی آنها انجام نمی شود. مقادیر از دست رفته می توانند ناشی از عواملی همچون وجود داده غیر منطقی یا عدم وجود داده (فردی به سئوال خاصی پاسخ نداده) و... باشد.

نکته: در این چنین مواردی می توان برای هر نوع داده از دست رفته یک کد تعریف کرد که این کار در ستون Missing تعریف می شود. مثلاً اگر داده ای وجود نداشته باشد کد (۱-) و برای داده های غیر منطقی کد (۲-).

ستون Columns: مربوط به تغییر دادن پهنای ستون در پنجره Data view می‌باشد که به مانند ستون Width و Decimals می توان تغییراتی در آن ایجاد کرد.

ستون Align: تراز کردن داده ها در این ستون قابل انجام شدن می باشد. داده ها به طور پیش فرض در پنجرهData view راست چین هستند ولی زمانی پیش می آید که می خواهیم داده ها چپ چین و یا وسط چین باشند. برای این کار ابتدا مکان نمای ماوس را روی ستون Align قرار داده و با دو بار کلیک کردن امکان مشاهده سایر گزینه ها را فراهم می کنیم و سپس گزینه مورد نظر را انتخاب می کنیم.

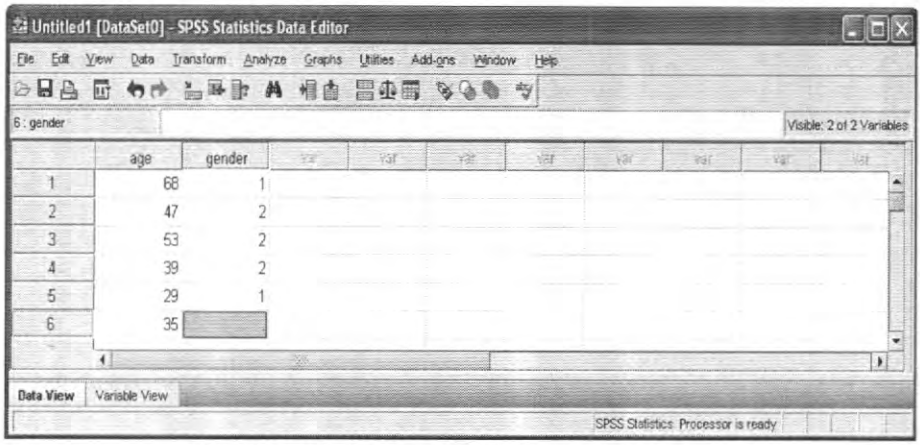
ستون Measure: این ستون مربوط به مقیاس اندازه گیری متغیرها است که شامل سه نوع زیر است:

Scale = مقیاس فاصله ای و نسبتی(نسبی)

Ordinal = مقیاس رتبه ای

Nominal = مقیاس اسمی

بعد از تعریف تمامی متغیرهای لازم همانند شکل (۱-۷) با کلیک روی عبارت Data view به محیط وارد کردن داده ها برگشته و داده ها را وارد کنید.

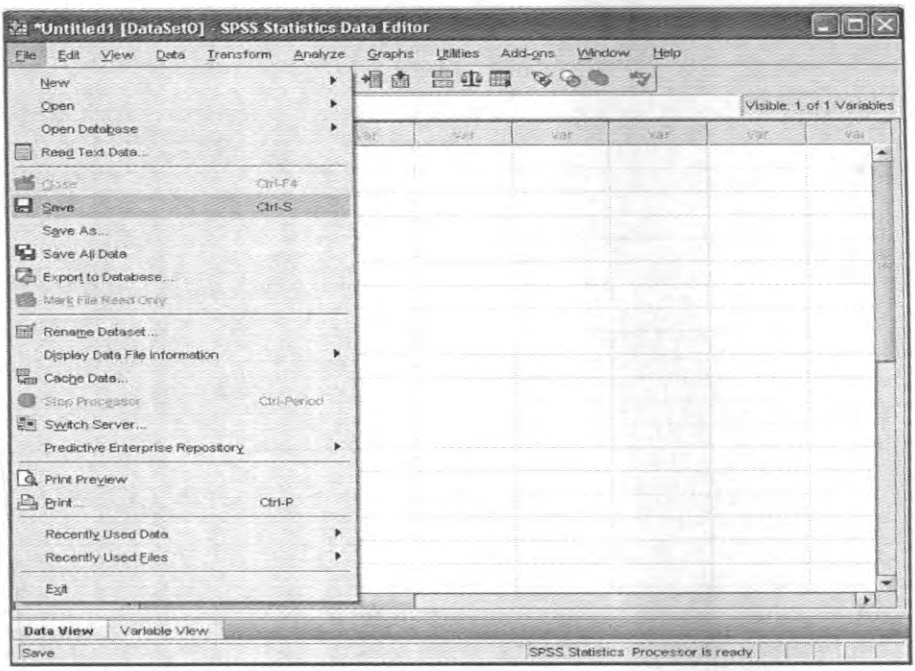


شکل (۱-۷) پنجره SPSS Statistics Data Editor

نکته: با توجه به این تفاسیر در پنجره Data View هر سطر یک آزمودنی و هر ستون یک متغیر است. پس برای ایجاد این ستون ها (متغیرها) باید در پنجره Variable View متغیرها را تعریف کرد.

۱-۴. ذخیره کردن فایلها در SPSS

برای ذخیره کردن فایل ها در SPSS همانند تمام برنامه ها از منوی File استفاده می‌شود(شکل ۱-۸).

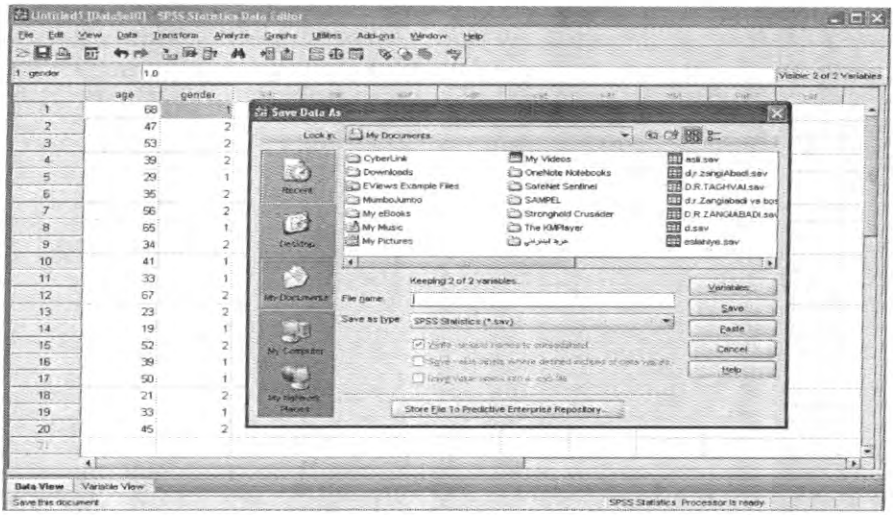


شکل(۱-۸) منوی File در پنجره SPSS Statistics Data Editor

به جای مسیر فوق می توانید از کلید میانبر (شکل۱-۹) استفاده کنید.

شکل (۱-۹) پنجره کلید میانبر Save کردن

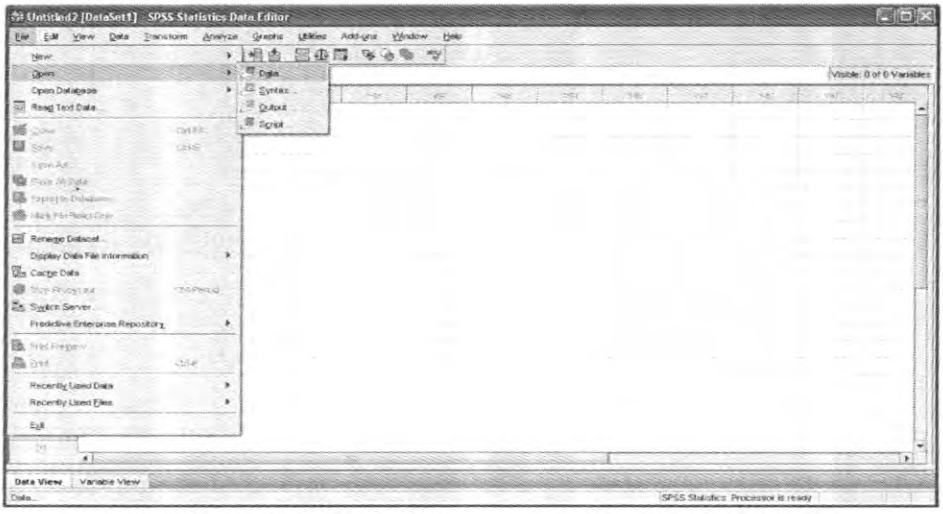
سپس پنجره ای باز می شود که شما می توانید با نام گذاری فایل خود آن را ذخیره کنید (شکل ۱-۱۰).



شکل (۱-۱۰) پنجره Save Data As

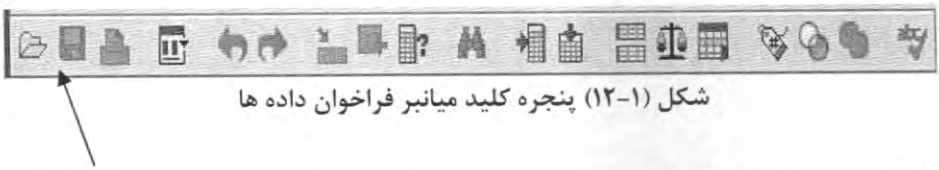
۱- ۵. باز کردن فایل ها در SPSS

برای باز کردن یک فایل حاوی داده های SPSS از مسیر شکل (۱-۱۱) استفاده کنید.



شکل (۱-۱۱) فراخوان داده ها در پنجره SPSS Statistics Data Editor

به جای مسیر فوق می توان از کلید میانبر (Ctrl + o) و یا مطابق شکل (۱-۱۲)، فایل داده های مورد نیاز خود را فراخوانی کنید.



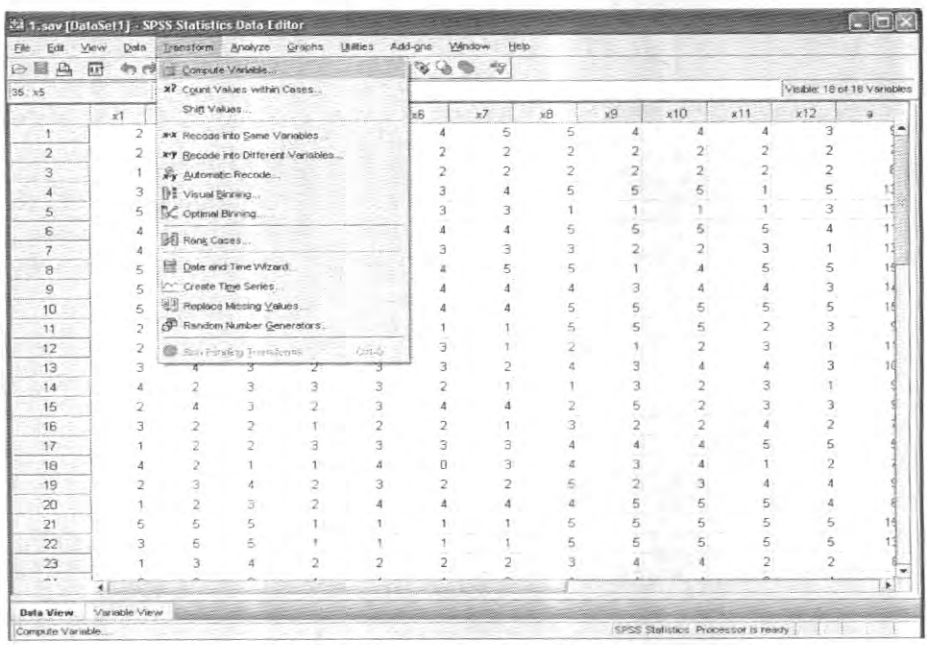
شکل (۱-۱۲) پنجره کلید میانبر فراخوان داده ها

۱-۶. دستور compute

از این دستور برای انجام محاسبات بر روی متغیرها استفاده می گردد. این دستور قابلیت ایجاد متغیر جدید، جایگزینی مقادیر یک متغیر معین، معدل گیری بین چند متغیر خاص و به کارگیری بیش از هفتاد تابع از پیش تعریف شده (شامل توابع ریاضی، توابع آماری، توابع توزیعی و توابع رشته ای) را دارد.

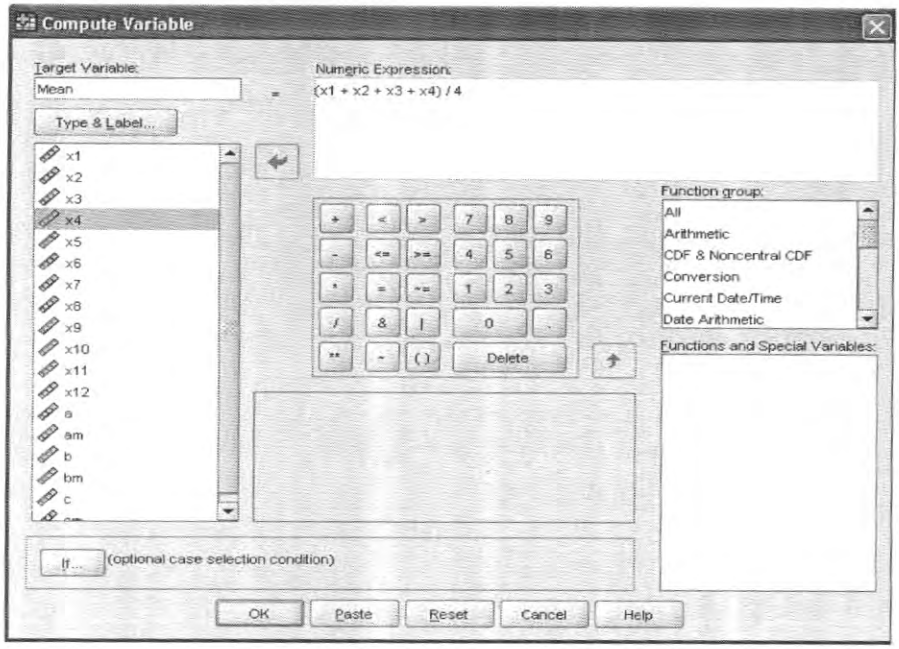
مثال: محققی برای سنجش پایداری اکولوژی کشاورزی فهرستی از چند متغیر را در غالب سئوالاتی با طیف پنج سطحی لیکرت(خیلی زیاد ۵، زیاد ۴، متوسط ۳، کم ۲ و بسیار کم ۱) مطرح کرده است. وی قصد دارد تا همه متغیرها را نهایتاً در یک متغیر ادغام کند. بنابراین برای این کار از دستور compute استفاده می کند که مراحل آن در شکل (۱-۱۳) تشریح می گردند.

Transform ⟶ compute Variable



شکل (۱-۱۳) پنجره 1.sav[DataSet1]- SPSS Statistics Data Editor

پس از اجرای این دستور پنجره compute Variable باز می گردد. همانند شکل (۱-۱۴) در گوشه سمت چپ جدول(قسمت Target Variable) نام متغیر جدیدی که می‌خواهید بوجود آید را بنویسید، سپس متغیرها را یکی یکی به قسمت Numeric Expression انتقال دهید. پس از انجام این مراحل گزینه Ok را کلیک کنید تا نتایج در متغیری جدید برای شما در صفحه اصلی SPSS ظاهر گردد.



شکل (۱-۱۴) پنجره Compute Variable

فصل دوم

مفاهیم پایه‌ای آمار

۲-۱. تعریف علم آمار

واژه آمار در فارسی به معنای شمارش و حساب است و معادل این واژه در زبان های اروپایی کلمه statistics می باشد. در اصطلاح عامیانه آمار به معنای ثبت و نمایش اطلاعات عددی در مورد یک موضوع مثلاً ثبت و نمایش تعداد بیکاران، میزان محصولات کشاورزی، میزان صدور نفت، جمعیت شهر تهران و غیره می باشد. ولی علم آمار امروزه دارای مفهومی بسیار وسیع تر از این کاربرد عامیانه است. از دیدگاه علمی، آمار به مجموعه روش هایی برای جمع آوری، تنظیم، تلخیص، تجزیه و تحلیل، انجام استنباط و نتیجه گیری مشاهدات اطلاق می شود. به بیان دیگر آمار عبارت است از هنر و علم جمع آوری، تعبیر، تجزیه و تحلیل داده ها و استخراج تعمیم های منطقی در مورد پدیده های تحت بررسی.

با توجه به تعاریف بالا می توان گفت یک فرآیند تحلیل آماری شامل دو بخش عمده است، اولین قدم نمایش دادن و خلاصه کردن داده ها می باشد تا توجه ما روی ویژگی های مهم داده ها متمرکز شود و جزئیات غیر ضروری کنار گذاشته شود. اما بخش دوم برای استخراج نکات کلی و استنباط هایی در مورد پدیده تحت مطالعه به کار می رود. بخش اول شامل روش های آمار توصیفی[1] و بخش دوم در برگیرنده روش های موسوم به آمار استنباطی[2] است. اما در کنار این دو نوع آمار باب تازه ای در علم آمار پس از جنگ

---

1. descriptive statistics
2. inferential statistics

دوم جهانی آغاز شد که بـه مبحـث روش هـای ناپارامتریـک[1] مشـهور اسـت. روش هـای ناپارامتریک به آزمون هایی اطلاق می گـردد کـه مشـروط بـه مفروضـات آمـار کلاسـیک نیستند. با توجه به مباحث مطرح شده ابتدا به بررسی اجمالی از انواع آمار می پردازیم.

۲-۲. انواع آمار

۲-۲-۱. آمار توصیفی

آمار توصیفی به آن دسته از روش های آماری گفته می شود که به پژوهشگر در طبقه بندی، خلاصه کردن، توصیف و تفسیر و برقراری ارتباط از طریق اطلاعات جمع آوری شده کمک می کند. نقش آمار توصیفی به عنوان اولین قدم در فرایند تحلیل آماری بسیار مهم است.

۲-۲-۲. آمار استنباطی

این نوع آمار از اوایل قرن بیستم مطرح شد. در این نوع آمار محقق با استفاده از مقادیر نمونه آماره ها را محاسبه می کند. سپس به کمک تخمین و آزمون فرض آماری، آماره[2] یا آماره ها به پارامتر[3] یا پارامترهای جامعه تعمیم داده می شود. به طورکلی در آمار استنباطی محقق یا برنامه ریز ورای مطالعات توصیفی سعی در ارائه تعمیم های آماری دارد. لذا آمار استنباطی به تجزیه و تحلیل مشاهداتی که از آمار توصیفی به دست می آید مربوط می شود.

۲-۲-۳. آمار ناپارامتریک۴

این نوع آمار در مقابل آمار پارامتریک[5] بیان می شود. فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از توزیع نرمال[6] است؛ در حالی که در فنون ناپارامتریک این فرض ضرورتی ندارد. بنابراین جوامع آماری که از توزیع نرمال برخوردار نیستند و داده های غیر کمی با نمونه های کوچک را می توان با این فنون بررسی کرد. این نوع آمار در بخش اعظم تحقیقات علوم انسانی و رفتاری که بیشتر متغیرهای آن با مقیاس های کیفی سنجیده می شود، کاربرد فراوانی دارد. ناگفته نماند که بحث های ناپارامتریک خود نوعی

---

1 . non-parametric methods
2 . statistic
3 . parameter
4 . non-parametric statistics
5 . parametric statistics
6 . normal distribution

از آمار استنباطی هستند اما چون از نظر بازه زمانی پس از بحث های پارامتریک مطرح شدند در اینجا بصورت جداگانه تشریح گردیده اند. به طورکلی برای انجام یک آزمون آماری با مفاهیم بسیاری روبه رو هستیم که در ابتدا باید با این مفاهیم در آمار توصیفی آشنایی داشت، که در ادامه به آن خواهیم پرداخت.

۲-۳. اصطلاحات آماری

۲-۳-۱. جامعه[1]

مجموعه افراد، اشیاء یا چیزهایی را که می خواهیم یک یا چند ویژگی درباره آنها مطالعه کنیم یک جامعه آماری می نامیم. به بیان دیگر جامعه آماری عبارت است از کلیه عناصر و افرادی که در یک مقیاس جغرافیایی مشخص (جهانی یا منطقه ای) دارای یک یا چند صفت مشترک باشند. جامعه آماری به دو نوع تقسیم می شود: محدود و نامحدود. اگر جامعه از تعداد محدود و ثابتی از مقادیر تشکیل شود و پایان پذیر باشد آن را محدود و در غیر این صورت، وقتی که جامعه از یک ردیف بی انتهای مقادیر تشکیل شده باشد آن را نامحدود گویند.

۲-۳-۲. نمونه[2]

تعداد محدودی از آحاد جامعه آماری که بیان کننده ویژگی های اصلی جامعه باشد را نمونه گویند. از آنجا که جوامع آماری معمولاً از حجم و وسعت جغرافیایی زیادی برخوردارند و محققان نمی توانند به تمام آنها مراجعه کنند، بنابراین ناگزیرند به انتخاب جمعی از آنها به عنوان نمونه و تعمیم نتایج آن به جامعه مورد مطالعه اکتفا کنند. توضیح این نکته ضرورت دارد که در آمار، مقادیر مربوط به ویژگی های یک نمونه (میانگین، انحراف معیار و غیره) را آماره و ویژگی های معادل آنها در جامعه اصلی را پارامتر می نامند.

۲-۳-۳. متغیر[3]

ویژگی یا خاصیت یک فرد، شئ و یا موقعیت است که می توان آنرا اندازه گرفت. به عبارت دقیق تر متغیر، خصوصیتی قابل اندازه گیری است که می تواند از فردی به فرد

---

1 . Population
2 . Sample
3 . Variable

دیگر و یا از مشاهده ای به مشاهده دیگر تغییر کند. دو نوع متغیر وجود دارد: متغیر کمی[1] و متغیر کیفی[2]

انواع متغیرها

۱) متغیرهای کمی

آن دسته از متغیرهایی هستند که در درجه تفاوت دارند تا نوع یا سطح. این متغیرها را می توان در مقیاس های فاصله ای و نسبی اندازه گیری کرد.

الف) مقیاس فاصله ای[3] : درجه و فاصله را از یک صفر قراردادی یا اختیاری در فواصل یکسان اندازه گیری واحد نشان می دهد. مثلاً درجه حرارت، نمره های امتحان و از این قبیل.

ب) مقیاس نسبی[4] : عالی ترین نوع مقیاس است که درجه و فاصله را از یک صفر طبیعی یا مطلق نشان می دهد. مثلاً ارتفاع، مصرف، بودجه سالیانه و از این قبیل.

۲) متغیرهای کیفی

متغیرهای کیفی آن دسته از متغیرهایی هستند که از نظر نوع تفاوت دارند تا از نظر درجه یا میزان. این متغیر را می توان با مقیاس های اسمی و یا ترتیبی اندازه گیری کرد.

الف) مقیاس های اسمی[5] : این مقیاس، ابتدایی ترین نوع مقیاس است که دسته بندی در گروه ها یا طبقات را نشان می دهند. برای مثال: جنسیت، مذهب، نژاد، رنگ، شغل و از این قبیل.

ب) مقیاس ترتیبی[6] : این مقیاس نیز همانند مقیاس اسمی به طبقه بندی داده ها می پردازد با این تفاوت که در مقیاس ترتیبی (همچنان که از نامش بر می آید) ترتیب مهم است. مثلاً مقیاس موافقت و مخالفت (عدد۱ مبین توافق بسیار زیاد و ۵ مبین مخالفت بسیار زیاد است) و درجه بندی های رضایت روستاییان (۱ مبین رضایت کامل و ۵ مبین نارضایتی کامل).

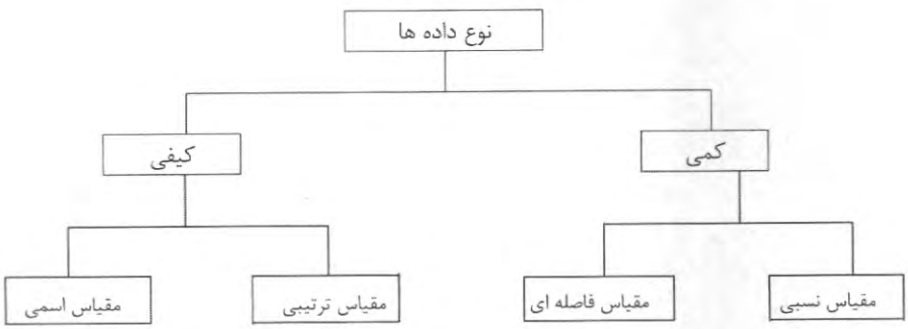
---

1 . quantitative Variable
2 . qualitative Variable
3 . interval scale
4 . ratio scale
5 . nominal scale
6 . rank scale

بطورکلی داده کیفی را می توان دارای دو بخش دانست که هر کدام تنها به دو طبقه (مثل جنسیت) یا چند طبقه(مثل نواحی جغرافیایی) اشاره می کنند.

نکته: SPSS هیچ تمایزی را بین داده های فاصله ایی و نسبی قائل نمی شود و آنها را تحت عنوان scale فهرست می کند.

در شکل(۲-۱) انواع داده ها به همراه مقیاس های آنها نمایش داده شده است. مقایسه مقیاس های چهارگانه نیز در جدول(۲-۱) انجام گرفته است.



شکل (۲-۱) نوع داده و مقیاس اندازه گیری

جدول (۲-۱) مقایسه مقیاس های اندازه گیری

مراتب / نوع مقیاس	ترتیب	فواصل	مبدأ صفر قراردادی	مبدأ صفر مطلق
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه ای	دارد	ندارد	ندارد	ندارد
فاصله ای	دارد	دارد	دارد	ندارد
نسبی	دارد	دارد	دارد	دارد

۲-۳-۴. فرضیه[1]

فرضیه یک ادعا درباره بعضی از ویژگی های یک جامعه است که باید براساس مدارک تجربی مورد حمایت قرار گیرد یا رد شود. فرضیه عبارت است از حدس و گمان اندیشمندانه دربارهٔ ماهیت و چگونگی روابط بین پدیده ها، اشیاء و متغیرها که محقق را در تشخیص نزدیکترین و محتمل ترین راه برای کشف مجهول کمک می نماید. اصولاً بدون داشتن فرضیه آماری امکان انجام یک آزمون دشوار است. مهمترین ویژگی های یک فرضیه خوب عبارتند از:

۱. فرضیه باید قدرت تبیین حقایق را داشته باشد.
۲. فرضیه باید بتواند پاسخ مسأله تحقیق را بدهد.
۳. فرضیه باید قابلیت حذف حقایق نامرتبط با مسأله تحقیق را داشته باشد.
۴. فرضیه باید شفاف، ساده و قابل فهم باشد.
۵. فرضیه باید قابلیت آزمون را داشته باشد.
۶. فرضیه نباید با حقایق و قوانین مسلم و اصول علمی تأیید شده و پذیرفته شده مغایرت داشته باشد.
۷. فرضیه نباید از واژه ها و مفاهیم ارزشی استفاده کند.
۸. فرضیه باید به مطالعه و پژوهش جهت بدهد و راهنمای فعالیت های محقق باشد.
۹. فرضیه باید به صورت جمله خبری باشد.

فرضیه آماری به دو دسته فرض صفر[2] (H_0) و فرض تحقیق[3] (H_1) بیان می‌شود.

فرض صفر (H_0): آن فرضیه ایی است که به عنوان یک فرضیه درست پذیرفته می شود مگر اینکه شواهد محکمی بر علیه آن وجود داشته باشد. یا آن دسته از فرضیاتی که به عدم وجود رابطه، اثر و تفاوت دلالت می کنند اصطلاحاً فرض صفر یا فرض آماری یا (H_0) اطلاق می شود.

---

1 . hypothesis
2 . null hypothesis
3 . research hypothesis

فرض تحقیق (H_1): آن دسته از فرضیاتی که به وجود رابطه، اثر و تفاوت دلالت می کنند اصطلاحاً فرض تحقیق یا (H_1) اطلاق می شود. این فرضیه زمانی پذیرفته می شود که (H_0) رد شده باشد.

فرضیه را در سه الگو یا غالب بکار می بریم:

۱- فرضیاتی که به بیان رابطه بین متغیرها می پردازند. مثلاً بین میزان تحصیلات و میزان رضایت از سکونت در روستا رابطه معنی داری وجود دارد.

۲- فرضیاتی که به بیان اثر یک متغیر بر روی متغیر دیگر دلالت می کنند. به این نوع فرضیات، فرضیات علی گفته می شود. مثلاً نوع شغل در نرخ مشارکت روستائیان در فعالیت های عمرانی مؤثر می باشد.

۳- فرضیاتی که به بیان جنبه تفاوت یا مقایسه می پردازند(فرضیه های تفاوتی یا مقایسه‌ای). مثلاً بین شیوه های سنتی و نوین آموزش در رابطه با افزایش سطح زیر کشت تفاوت وجود دارد.

گاهی اوقات در بیان فرضیات آنها را با جهت یا بدون جهت می سازیم. با جهت به این معنا که قید کوچکتر یا بزرگتر، یا بیشتر و کمتر را در ساخت فرضیه به کار می بریم. معمولاً فرضیه های تفاوتی، فرضیه هایی جهت دار هستند و باید سعی کرد تا آنها را جهت دار ساخت. مثلاً میزان مشارکت افراد تحصیل کرده در فعالیت های عمرانی بیشتر از بقیه افراد می باشد.

در فرضیه های بی جهت، جهت برای محقق مهم نیست و محقق بیشتر به دنبال بیان رابطه یا اثر می باشد. معمولاً فرضیه های رابطه ای را بدون جهت می گویند. مثلاً بین افزایش سطح زیر کشت و مهاجرت از روستا رابطه معنی دار وجود دارد.

۲-۳-۵. آزمون فرضیه[1]

ملاک اساسی برای با ارزش بودن فرضیه قابلیت آزمون آن است و آزمون فرضیه سنجش صحت یا درستی ادعای ارائه شده از دیدگاه محقق است. اگر چه یک فرضیه هرگز به طور قطعی ثابت نمی شود ولی آزمون فرضیه به دنبال این است که حدس محقق تا چه اندازه

---

1 . testing hypothesis

با واقعیت تطبیق دارد. در آزمون های آماری دو نوع خطای ممکن وجود دارد که به شرح آنها می پردازیم.

خطای نوع اول: این خطا زمانی روی می دهد که H_0 درست است اما به اشتباه رد می-گردد. این نوع تصمیم نادرست را خطای نوع اول می نامیم.

خطای نوع دوم: این خطا زمانی روی می دهد که فرض H_0 را بپذیریم در حالیکه باید H_0 رد گردد. این نوع تصمیم نادرست را خطای نوع دوم می نامیم.

۲-۳-۶. مراحل آزمون فرضیه:

۲-۳-۶-۱. تعیین فرضیه ها:

در آزمون فرضیه ابتدا لازم است فرضیه H_0 و فرضیه مخالف آن را صورت بندی نماییم. صورت های مختلف فرضیه به سه شکل زیر قابل طرح می باشند:

وقتی فرضیه جهت دار باشد از حالت اول و دوم و برای فرضیه های بدون جهت از حالت سوم استفاده می شود.

۲-۳-۶-۲. تعیین آماره آزمون مناسب:

تصمیم در مورد رد یا پذیرش H_0 بستگی به انتخاب آماره آزمون متناسب با متغیرها دارد. انتخاب یک آزمون آماری وابسته به موارد زیر است:

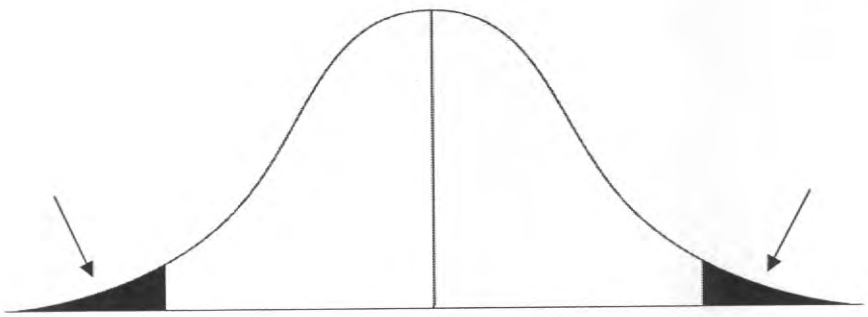
۱) سئوال پژوهش

۲) نقشه یا طرح پژوهش

۳) نوع داده های جمع آوری شده

۲-۳-۶-۳. تعیین ناحیه رد یا مرز اهمیت:

مرز اهمیت که با (α) نشان داده می شود نقطه ای است که در آن تفاوت آمار نمونه ای از مقدار مورد انتظار در جامعه آنقدر باشد که دیگر نتوان آن را نتیجه تغییر پذیری تصادفی دانست. در حقیقت تعیین مرز اهمیت همان تعیین حد فاصل میان تفاوت کم و زیاد می باشد. تفاوت اندک آنهایی هستند که احتمال وقوعشان زیاد و تفاوت زیاد آنهایی هستند که احتمال وقوعشان کم باشد. در شکل (۲-۲) نواحی بحرانی برای رد فرض H_0 با علامت نشان داده شده است.



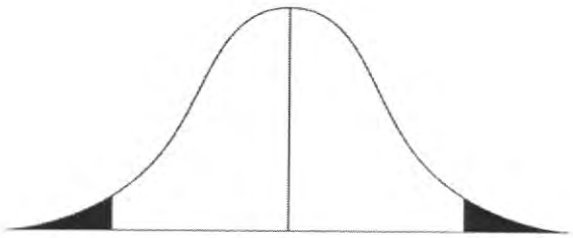
شکل (۲-۲) مقدار پارامتر مورد انتظار

نواحی بحرانی به این معنا هستند که اگر مقدار نمونه ای مشاهده شده در این نواحی باشند تفاوت آن با مقدار مورد انتظار در جامعه از لحاظ آماری برای رد فرض H_0 دارای اهمیت می باشد. بنابراین ورود به ناحیه بحرانی را می توان برابر با رد فرض H_0 دانست.

۲-۳-۶-۴. تعیین یک طرفه یا دو طرفه بودن آزمون:

در تعیین سطح معنی دار بودن باید مشخص کنیم که آیا منطقه رد فرضیه در یک طرف یا اینکه در هر دو طرف منحنی قرار می گیرد. اگر در آزمون فرضیه ای فقط از یک دنباله منحنی استفاده شود آن را آزمون یک طرفه یا یک دنباله می گویند ولی اگر هر دو طرف منحنی استفاده شود به آن آزمون دو طرفه یا دو دنباله می گویند. به طور معمول اگر جهت تفاوت در فرضیه H_0 دارای اهمیت باشد از آزمون یک طرفه شکل (۲-۳) و در صورتی که جهت مهم نباشد از آزمون دو طرفه شکل (۲-۴) استفاده می شود. آزمون‌های یک طرفه در اصل همان آزمون های جهت دار، و آزمون های دو طرفه آزمون های بی جهت می باشند.

شکل (۲-۳) آزمون های یک طرفه



شکل(۲-۴) آزمون دو طرفه

۲-۳-۶-۵. محاسبه مقدار آماره آزمون:

در این مرحله با توجه به داده های به دست آمده مقدار آماره آزمون را حساب می کنیم.

۲-۳-۶-۶. نتیجه گیری:

در این مرحله با توجه به این که مقدار آماره آزمون محاسبه شده در چه ناحیه ای قرار می گیرد می توان فرض H_0 را تأیید یا رد کرد.

۲-۴. شاخص های توصیف داده ها

شاخص های توصیف داده ها به سه گروه شاخص های مرکزی (مانند: میانگین، میانه...)، شاخص های پراکندگی (مانند: واریانس، انحراف معیار...)، شاخص های شکل توزیع (مانند: شاخص های چولگی و شاخص های کشیدی) تقسیم می شوند.

۲-۴-۱. شاخص های مرکزی[1]

- یکی از مهمترین موضوع ها در مطالعه هر جامعه آماری، تعیین مقدار مرکزی می-باشد، یعنی تعیین مقدار نماینده که مشاهدات در اطراف آن توزیع شده اند. میانگین، مد(نما)، میانه و چندک ها شاخص های گرایش مرکزی را تشکیل می دهند.

میانگین[2]: اصلی ترین و مورد استفاده تری شاخص مرکزی میانگین است. برای محاسبه میانگین، داده های یک متغیر را جمع کرده و بر تعداد مشاهدات تقسیم می‌کنیم. به عبارت دیگر میانگین، معدل مجموعه ای از مشاهدات به شمار می رود که معمولاً به

1 . central tendency
2 . mean

شکل ($\bar{x}$) برای نمونه و به شکل (μ) برای جامعه آماری نمایش داده می شود. میانگین به کمک رابطه ریاضی (۲-۱) محاسبه می شود.

$$\mu = \frac{\sum x}{n} \qquad (۲-۱)$$

مد[1] : مد یا نما در آمار مقداری است که بیشترین تکرار را در میان مشاهدات داشته باشد. مد از جمله شاخص های مرکزی است که تحت تأثیر مقادیر دور افتاده قرار نمی گیرد. بنابراین در صورتی که بخواهیم تأثیر مقادیر دورافتاده (مقادیر حداکثر و یا حداقل) را از بین ببریم، استفاده از این شاخص مطلوب به نظر می رسد.

میانه[2] : اگر داده های یک متغیر را به صورت صعودی یا نزولی مرتب کنیم، نمره وسط میانه است. بعبارت دیگر میانه مقداری است که ۵۰ درصد داده های جامعه پایین تر از آن و ۵۰ درصد بالاتر از آن قرار می گیرند. به طور کلی از میانه به عنوان اندازه تمایل به مرکز توزیع هایی که شکل آن ها نامتقارن است استفاده می شود. درباب میانه دو نکته قابل تأمل است:

۱. اگر جامعه ای که می خواهیم میانه آن را تعیین کنیم تعداد اعضای آن فرد باشد، عدد میانی میانه است.

۲. در صورتی که تعداد اعضای جامعه زوج باشد معدل دو عدد وسط میانه خواهد بود.

محل میانه را با استفاده از رابطه (۲-۲) محاسبه می کنند.

$$\mathrm{Md} = \frac{n+1}{2} \qquad (۲-۲)$$

چندک‌ها[3] : چندک ها برای تعیین و یا تقسیم بندی مقادیر یک متغیر به کار می روند. درست همان گونه که میانه کل مقادیر مشاهدات را به دو قسمت مساوی تقسیم می‌کند، چندک ها نیز آن را به چند قسمت برابر تقسیم می کند. اگر مقادیر یک متغیر را به چهار قسمت برابر تقسیم کنیم به چارک ها[4]، اگر به ده قسمت برابر تقسیم کنیم به دهک ها[5] و اگر به صد قسمت برابر تقسیم کنیم به صدک ها[6] دست پیدا کرده ایم.

---

1 . mode
2 . median
3 . quantiles
4 . quartiles
5 . decile
6 . percentile

۲-۴-۲. شاخص های پراکندگی[1]

برای این که داده ها به طور واقعی تری توصیف شوند و مجموعه های مشاهدات با هم قابل مقایسه باشند، باید میزان تفاوت های آن ها نیز از طریق معیارهای عددی سنجیده شود. سنجش این تفاوت ها از طریق شاخص های پراکندگی صورت می گیرد.

دامنه تغییرات[2] : دامنه تغییرات ساده ترین اندازه پراکندگی است که مبین تفاوت بین بزرگترین و کوچکترین مشاهده توزیع است و به صورت رابطه ریاضی (۲-۳) محاسبه می‌گردد.

$$R= Max(xi) - Min(xi) \qquad (۳-۲)$$

دامنه میان چارکی[3] : این شاخص دامنه تغییرات ۵۰ درصدی داده های وسط (بین چارک اول و سوم) را نشان می دهد. این شاخص (IQR) به صورت رابطه (۲-۴) محاسبه می شود.

$$IQR = Q_3 - Q_1 \qquad (۴-۲)$$

واریانس[4] : واریانس به میزان پراکندگی یک مجموعه از داده ها حول میانگین آن داده‌ها می پردازد. بیان علمی تر از این واژه، گستردگی توزیع احتمال و فاصله متغیر تصادفی از مقدار میانگین می باشد. بنابراین هر چه مقدار واریانس بیشتر باشد پراکندگی مقادیر صفت داخل جامعه نسبت به میانگین بیشتر خواهد بود و در نتیجه مقادیر صفات در جامعه از هم فاصله بیشتری دارند و برعکس هر چه مقدار واریانس کمتر باشد فاصله مقادیر صفات از یکدیگر کمتر بوده و نمره ها به هم نزدیک ترند. واریانس از رابطه (۲-۵) محاسبه می گردد.

$$S^2 = \frac{\sum(x_i - \mu_x)^2}{N} \qquad (۵-۲)$$

انحراف معیار[5] : اگر جذر واریانس را بگیریم به انحراف معیار خواهیم رسید. انحراف معیار از شاخص های مهم برای سنجش پراکندگی مقادیر یک متغیر است. با توجه به اینکه در محاسبه انحراف معیار کلیه مقادیر یک متغیر نقش دارند و پراکندگی مقادیر را

---

1 . dispersion indexes
2 . Range
3 . Interquartile Range
4 . variance
5 . Standard Deviation

حول میانگین محاسبه می کنند و میانگین نیز از مهم ترین شاخص های مرکزی می-باشد، بنابراین انحراف معیار نیز از اهمیت زیادی برخوردار است. انحراف معیار از طریق رابطه (۲-۶) به دست می آید.

$$S = \sqrt{S^2} \qquad (۲-۶)$$

ضریب پراکندگی[1] : یکی از معیارهای بسیار دقیق که کاربرد گسترده ای نیز در تحلیل پراکندگی مشاهدات دارد، ضریب تغییرات یا همان ضریب پراکندگی است. ضریب پراکندگی برای مقایسه پراکندگی دو متغیر زمانی که میانگین یکسانی ندارند به کار برده می شود. این معیار به واحد اندازه گیری متغیرها بستگی ندارد، بنابراین می توان از آن برای مقایسه پراکندگی داده ها با واحدهای اندازه گیری مختلف استفاده کرد. ضریب پراکندگی از طریق رابطه (۲-۷) به دست می آید.

$$C.V = \frac{S}{\mu_x} \qquad (۲-۷)$$

خطای معیار میانگین[2] : شاخص خطای معیار میانگین ($S_{\bar{x}}$)، انحراف معیار میانگین نمونه است. اگر توزیع داده ها نرمال باشد، با اطمینان ۹۵٪ ، میانگین جامعه آماری تقریباً در محدوده $\bar{x} \pm ۲S_{\bar{x}}$ قرار می‌گیرد. این شاخص از طریق رابطه (۲-۸) محاسبه می‌شود.

$$S_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (۲-۸)$$

۲-٤-۳. شاخص های شکل توزیع

این شاخص ها به بررسی شکل توزیع داده ها می پردازند و شامل چولگی[3] و کشیدگی[4] می باشند.

چولگی: چولگی میزان تقارن یک توزیع را اندازه گیری می کند که به سه شکل دیده می‌شود.

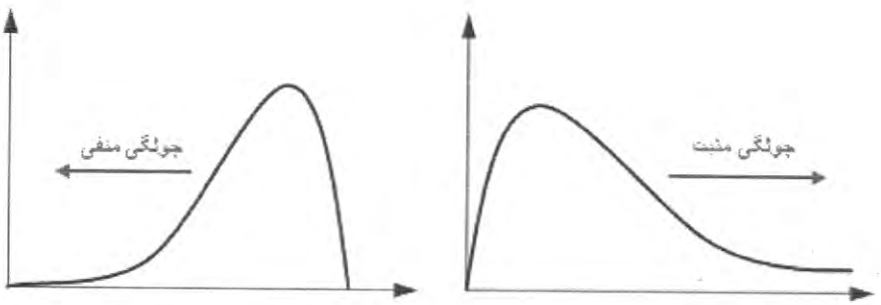
---

1 . coefficient of variation
2 . Standard Error of Mean
3 . Skewness
4 . Kurtosis

۱- چولگی مثبت است. زمانی که مد جامعه آماری کوچک تر از میانه باشد و میانه نیز از میانگین کمتر باشد، در این صورت چولگی مثبت است. به عبارت دیگر چنانچه $Mo < Md < \mu$ چولگی مثبت است.

۲- چولگی منفی است. هنگامی که مد جامعه آماری بزرگتر از میانه باشد و میانه نیز از میانگین بیشتر باشد. یعنی هنگامی که $\mu < Md < Mo$ چولگی منفی خواهد بود.

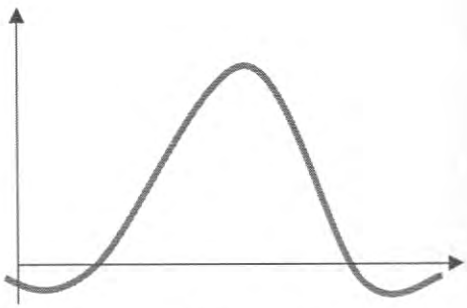
۳- چولگی نزدیک به صفر می باشد(توزیع نرمال است). هر چه قدر مقدار چولگی به صفر نزدیک باشد نشان از توزیع نرمال جامعه آماری است. به عبارت دیگر زمانی که سه شاخص گرایش مرکزی (میانگین، مد، میانه) در یک جامعه با هم برابر باشند، $\mu = Mo = Md$ در این صورت توزیع داده ها متقارن بوده و اصطلاحاً می گوییم توزیع داده ها نرمال می باشد. به صورت کلی اگر مقدار چولگی بین (۱+ و ۱-) باشد توزیع نرمال می‌باشد.



$Mo < Md < \mu$

$\mu < Md < Mo$

شکل(۲-۵) چولگی مثبت و منفی



$$\mu = \text{Mo} = \text{Md}$$

شکل (۲-۶) توزیع نرمال

مهمترین شاخص ضریب چولگی، ضریب چولگی گشتاوری است که از طریق رابطه (۲-۹) محاسبه می شود.

$$r_3 = \frac{\sum(x_i - \bar{x})^3}{n} \qquad (۹-۲)$$

و

$$SK = \frac{r_3}{s_x^3}$$

خطای استاندارد ضریب چولگی[1] : نسبت ضریب چولگی به خطای استاندارد آن را خطای استاندارد ضریب چولگی می گویند که از آن می توان برای سنجش نرمال بودن یک جامعهِ آماری استفاده کرد. در صورتی که خطای استاندارد ضریب چولگی کوچک تر از ۱- یا بزرگ تر از ۱+ باشد، نرمال بودن رد می شود.

کشیدگی: این آماره نیز همانند چولگی به بررسی میزان تقارن یک توزیع می پردازد. در حالت های خاص که تصمیم گیری به کمک پارامترهای مرکزی و چولگی امکان پذیر نیست، یکی از پارامترهای مناسب استفاده از مقایسه پراکندگی توزیع جامعه با توزیع نرمال است. شاخص سنجش پراکندگی جامعه نسبت به توزیع نرمال ضریب کشیدگی نام دارد که با (E) نشان داده می شود. با توجه به مقدار ضریب کشیدگی، درجات مختلفی برای کشیدگی پیدا خواهد شد که اندیشمندان آماری آنها را در سه گروه تقسیم کرده‌اند:

---

1 . Standard Error of Skewness

۱. آن دسته از توزیع ها که نسبت به توزیع نرمال از پراکندگی بیشتری برخوردارند، یعنی منحنی توزیع نسبت به توزیع نرمال کوتاه تر است. این دسته از توزیع ها دارای ضریب کشیدگی منفی خواهند بود.
۲. توزیع هایی هستند که از توزیع نرمال بلندترند، یعنی از اوج بیشتری برخوردارند. علت اوج گرفتن این توزیع ها آن است که داده ها حول میانگین متمرکز شده اند و پراکندگی اندکی دارند. ضریب کشیدگی این دسته از توزیع ها مثبت خواهد بود.
۳. توزیع هایی هستند که کشیدگی آنها با کشیدگی توزیع نرمال کاملاً برابر است. این توزیع ها، توزیع هایی با کشیدگی متوسط هستند. ضریب کشیدگی در این توزیع ها مساوی صفر است.

علاوه بر علامت ضریب کشیدگی، قدر مطلق آن $|E|$ نیز معنادار است. طبیعی است که هر چه قدر مطلق ضریب کشیدگی بزرگتر باشد تفاوت توزیع از نظر پراکندگی با توزیع نرمال بیشتر خواهد بود. در این زمینه استانداردهایی وجود دارد که عبارتند از:

۱. اگر $|E| \leq ۰/۱۰$ باشد، توزیع جامعه از نظر پراکندگی نرمال است.
۲. اگر $۰/۱۰ < |E| \leq ۰/۵۰$ باشد، توزیع از نظر کشیدگی دارای تفاوت اندکی با توزیع نرمال است.
۳. اگر $|E| > ۰/۵۰$ باشد، توزیع تفاوت زیادی با توزیع نرمال دارد.

ضریب کشیدگی از طریق رابطه (۲-۱۰) محاسبه می گردد.

$$r_4 = \frac{\sum(x_i - \bar{x})^4}{n} \qquad (۱۰-۲)$$

و

$$E = \frac{r_4}{s_x^4} - 3$$

خطای استاندارد ضریب کشیدگی[1]: نسبت ضریب کشیدگی به خطای استاندارد آن را خطای استاندارد ضریب کشیدگی می گویند. چنانچه خطای استاندارد ضریب کشیدگی کوچک تر از ۱- یا بزرگ تر از ۱+ باشد، نرمال بودن رد می شود.

---

1 . Standard Error of Kurtosis

۲-۵. تعیین شاخص های توصیف داده ها به کمک SPSS

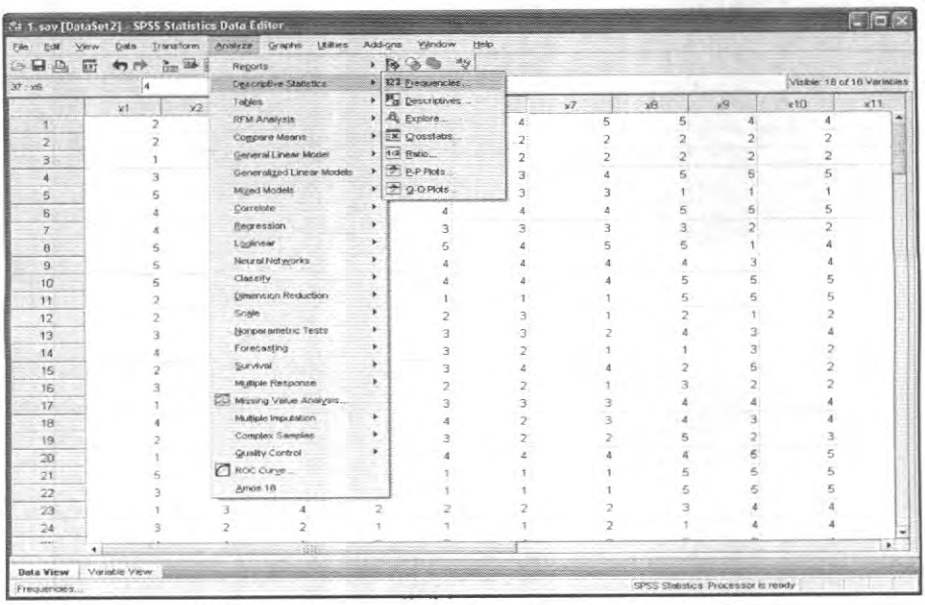
نرم افزار SPSS از دو دستور Frequencies و Descriptives قابلیت ارایه این شاخص ها را دارد که در ادامه شرح داده خواهند شد

۲-۵-۱. دستور Frequencies

این دستور دو حسن دارد ۱) برای انواع مختلف متغیرها قابل اجرا می باشد، ۲) قابلیت ارایه چندک ها (چارک ها، دهک ها و صدک ها)، میانه و مد را دارد.

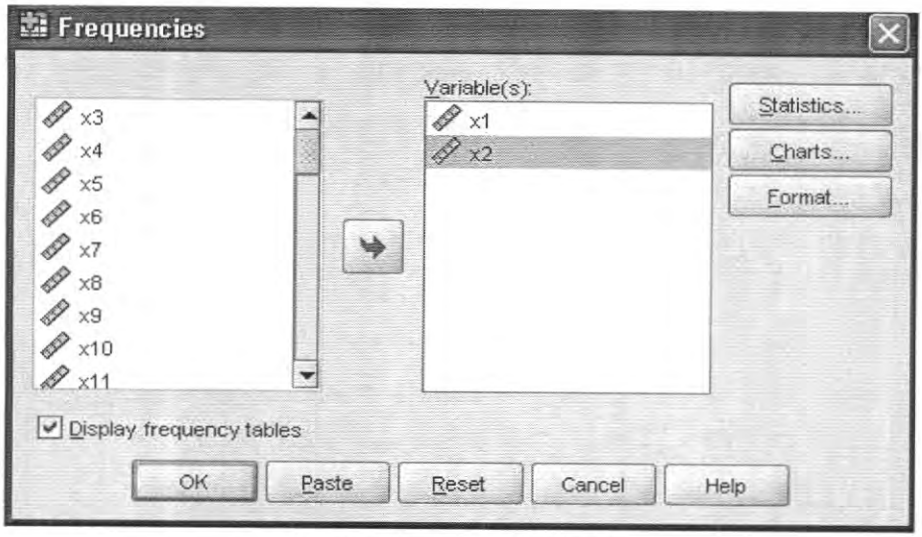
برای اجرای این دستور باید با توجه به شکل (۲-۷) به طریق زیر عمل کرد:

Analyze ⟶ Descriptive Statistics ⟶ Frequencies



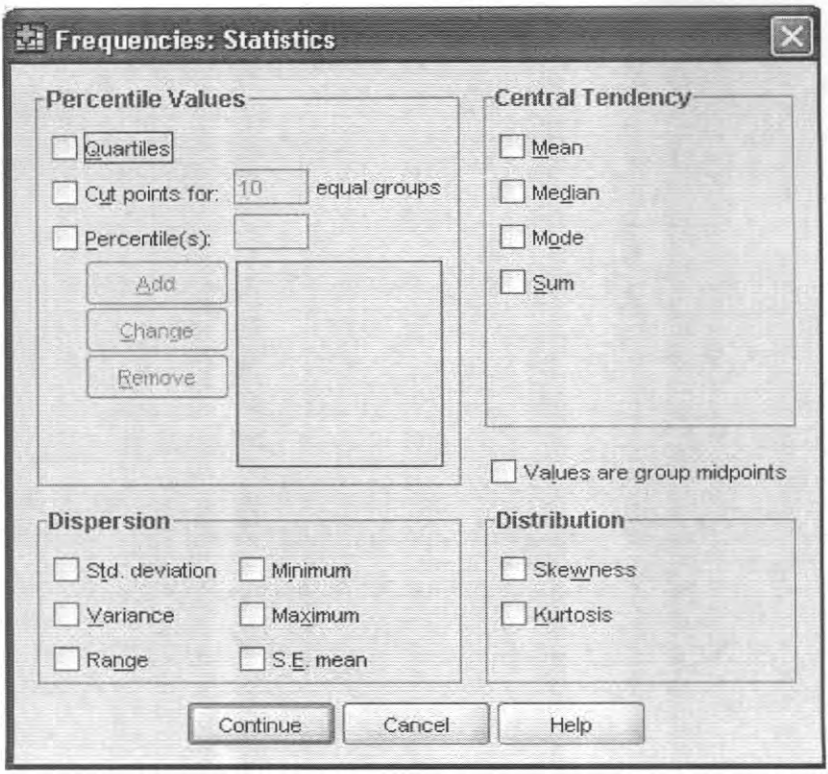
شکل (۲-۷) پنجره SPSS Statistics Data Editor 1

سپس پنجره ای تحت عنوان Frequencies باز می شود. همانند شکل (۲-۸) متغیرهای مورد نظر را به کادر Variable(s) منتقل کنید.



شکل (۲-۸) پنجره Frequencies

برای محاسبه شاخص های مرکزی، پراکندگی و چولگی، در پنجره فراوانی، دکمه Statistics را کلیک کنید تا پنجره ای همانند شکل (۲-۹) باز گردد. عملیات آماری را که مورد نظر شماست باید از این پنجره انتخاب کنید. به عنوان مثال، برای محاسبه میانگین، کلمه Mean را در کادری تحت عنوان Central Tendency فعال کنید. سایر شاخص های پراکندگی و چولگی نیز به ترتیب در کادرهای Dispersion و Distribution قرار دارد که در صورت نیاز می توان با فعال کردن هر کدام از آنها محاسبات مربوط هر داده را انجام داد. در کادر Percentiles Value این پنجره نیز می توانید با کلیک بر روی گزینه Quartiles چارک های ۱، ۲ و ۳ محاسبه گردند و به منظور ارائه صدک ها روی گزینه Percentile(s) کلیک کرده و مقدار مورد نظر را در کادر مربوطه وارد کنید. با کلیک روی دکمه Add این صدک انتخاب می شود.



شکل (۲-۹) پنجره Frequencies Statistics

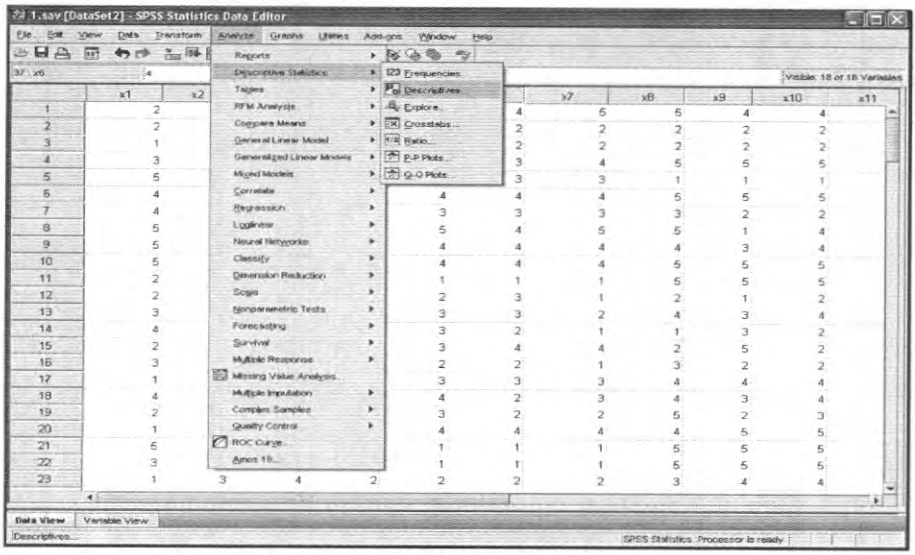
پس از این مرحله، Continue را کلیک کنید تا به پنجره اصلی بازگردید. در صورتی که بخواهید نمودارهای هیستوگرام و منحنی نرمال را رسم کنید، باید دکمه Charts را فشار دهید و آنها را انتخاب کنید. برای نمایش نتایج Ok را کلیک کنید تا نتایج برای شما ظاهر گردد.

۲-۵-۲. دستور Descriptives

این دستور نیز برای تعیین شاخص های توصیفی داده ها استفاده می شود با این تفاوت که دستور Descriptives توزیع فراوانی متغیرها را نشان نمی دهد، قابلیت ارایه چندک ها، میانه و مد را ندارد و برای متغیرهای رشته ای قابل استفاده نیست.

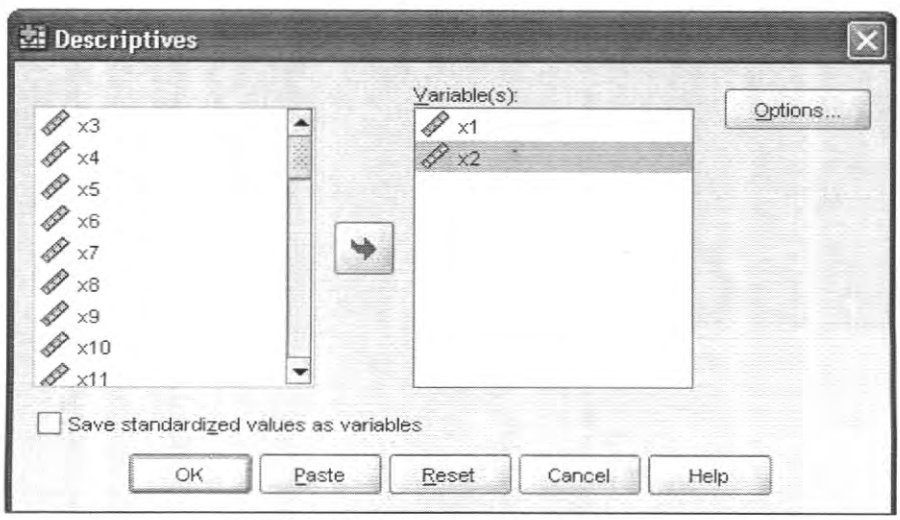
برای اجرای این دستور به طریق شکل (۲-۱۰) باید عمل کرد:

Analyze ⟶ Descriptive Statistics ⟶ Descriptives



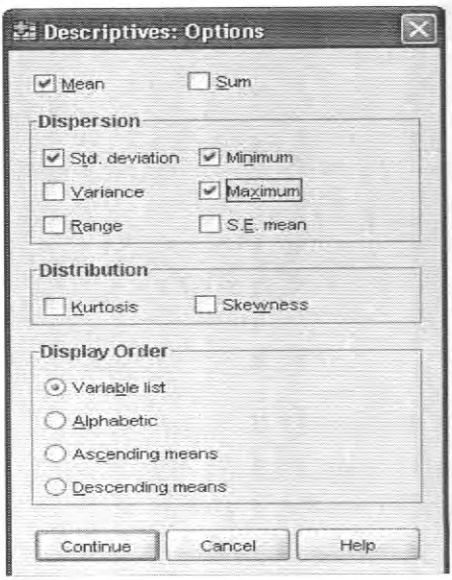
شکل (۲-۱۰) پنجره SPSS Statistics Data Editor 2

پس از اجرای این دستور پنجره ای تحت عنوان Descriptives باز می گردد همانند شکل (۲-۱۱) متغیرهای مورد نظر را به کادر Variable(s) منتقل کنید.



شکل(۲-۱۱) پنجره Descriptives

تنها دکمه موجود در این پنجره گزینه Options می باشد بر روی آن کلیک کنید و عملیات آماری را که مورد نظر شماست فعال کنید. (البته همان طوری که از شکل (۲-۱۲) پیداست نرم افزار به صورت پیش فرض گزینه های Mean، Minimum، Maximum و Std. deviation را فعال کرده است) پس از این مرحله، Continue را کلیک کنید تا به پنجره اصلی بازگردید و سپس برای نمایش نتایج Ok را کلیک کنید تا همانند جدول (۲-۲) برای شما ظاهر گردد.



شکل(۲-۱۲) پنجره Descriptives Options

جدول (۲-۲) نتایج توصیفی داده ها

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
x1	200	1	5	2.71	1.024
x2	200	1	5	3.43	.766
Valid N (listwise)	200				

فصل سوم

رسم نمودارها

در اغلب تحقیقات برای نمایش توزیع داده ها از نمودارها استفاده می گردد. به عبارت دیگر نمودارها امکان نمایش توزیع های فراوانی را به شیوه واضح و بیانی قابل فهم و روان امکان پذیر می سازند. در نرم افزار SPSS قابلیت ارائه نمودارهای فراوانی وجود دارد که هر کدام از آنها دارای کارکردهای مخصوص خود می باشند و در شرایط خاصی به کار می‌روند.

مهمترین و کاربردی ترین نمودارها در SPSS عبارتند از: نمودار ستونی یا میله ای (Bar Chart)، نمودار خطی (Line Chart)، نمودار ستونی خطا (Error bar Chart)، نمودار پراکنش (Scatter plot) و نمودار هیستوگرام (Histogram Chart). در راستای استفاده از نمودارها در نرم افزار SPSS معمولاً از منوی Graphs استفاده می گردد. در عین حال می توان از گزینه های دیگری نیز جهت ارائه نمودارها بهره گرفت.

نکته: تمامی نمودارها در نرم افزار SPSS قابل ویرایش هستند.

۳-۱. نمودار ستونی (میله ای)

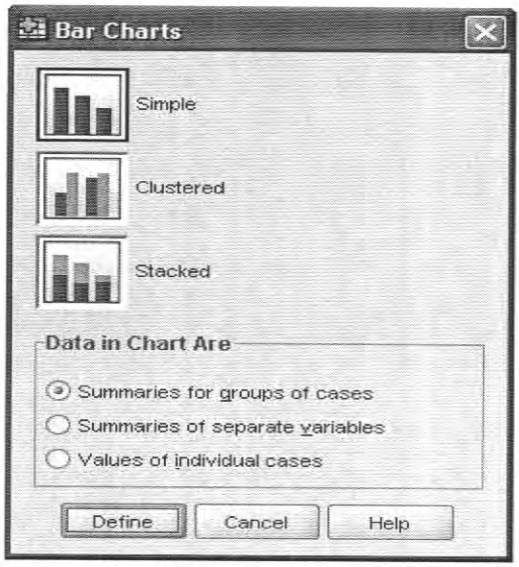
در این نمودار محور افقی نشان دهنده کیفیت خاصی در مشاهدات و محور عمودی آن نشان دهنده فراوانی هر گروه از مشاهدات می باشد. بنابراین باید زمانی از این نمودار استفاده کرد که متغیر ما به صورت طبقه بندی شده بوده و سطوح محدودی داشته باشد.

مثال: محققی درآمد سالیانه حاصل از کشاورزی را در یک روستا، به سه دسته کم درآمد (کمتر از ۳ میلیون تومان)، درآمد متوسط (۳ تا ۶ میلیون)، و درآمد بالا (بیشتر از ۶ میلیون) تقسیم بندی کرده است. وی سطوح کم درآمد را با کد ۱، درآمد متوسط کد۲ و

درآمد بالا را با کد ۳ وارد نرم افزار SPSS می کند. حال برای نمایش وضعیت درآمدی حاصل از کشاورزی از نمودار ستونی استفاده کرده است که در ادامه مراحل ترسیم نمودار آن ارائه می گردد.

مسیر زیر را انتخاب کنید تا کادر Bar Chart همانند شکل (۳-۱) باز گردد.

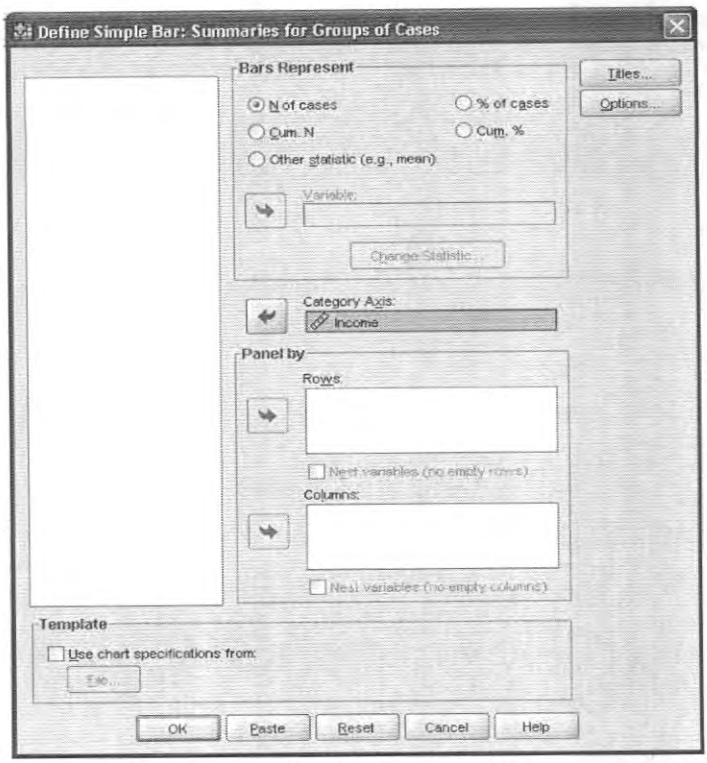
Graphs ⟶ Legacy Dialogs ⟶Bar



شکل(۳-۱) پنجره Bar Charts

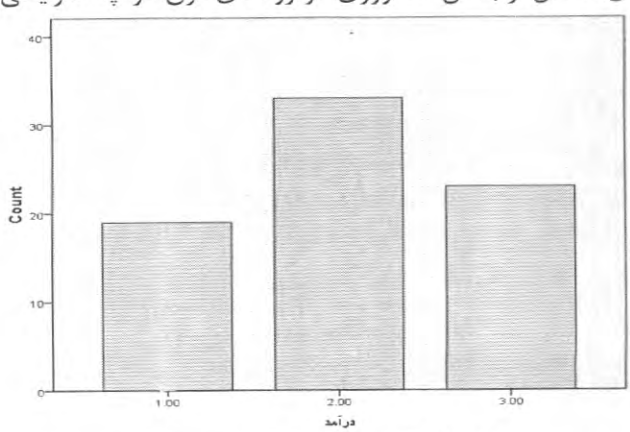
کادر Bar Chart دارای قابلیت های ارائه نمودارهای ستونی خوشه ای (Clustered) و ستونی انباشته (Stacked) نیز می باشد. اما ما در این مثال تنها از نمودار ستونی ساده (Simple) استفاده می کنیم بنابراین پیش فرض نرم افزار را تغییر ندهید.

گزینه Define را کلیک کنید تا پنجره مربوطه باز گردد. سپس همانند شکل (۳-۲) متغیر Income را به کادر Category Axis منتقل کنید و برای ترسیم نمودار ستونی Ok را کلیک کنید.



شکل(۳-۲) پنجره Difine Simple Bar Summaries for Group of Cases

خروجی حاصل از این دستور (شکل ۳-۳) نموداری است که به خوبی نشان می دهد وضعیت درآمدی حاصل از بخش کشاورزی در روستای فوق در چه شرایطی قرار دارد.



شکل (۳-۳) نمودار میله ای متغیر درآمد

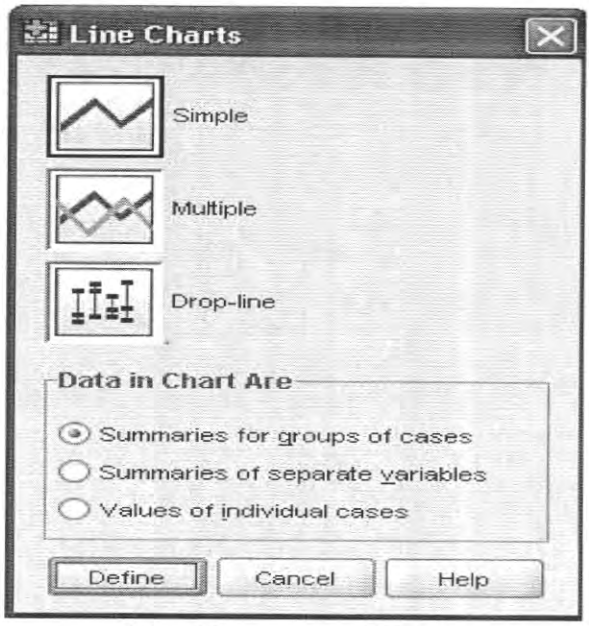
۳-۲. نمودار خطی

در این نمودار فراوانی هر طبقه به وسیله نقطه ای نمایش داده می شود که این نقطه به وسیله خطی مستقیم به نقطه کناری وصل می گردد و همینطور تمامی متغیرها زیر پوشش این خطوط قرار می گیرند. به کمک این نمودار به سادگی می توان تغییرات حاصل از یک متغیر کمی با سطوح نسبتاً محدود و یا طبقه بندی شده را به نمایش گذاشت.

مثال: محققی برای نمایش وضعیت سواد در یک منطقه روستایی می خواهد از نمودار خطی استفاده کند که در ادامه مراحل آن توضیح داده می شود.

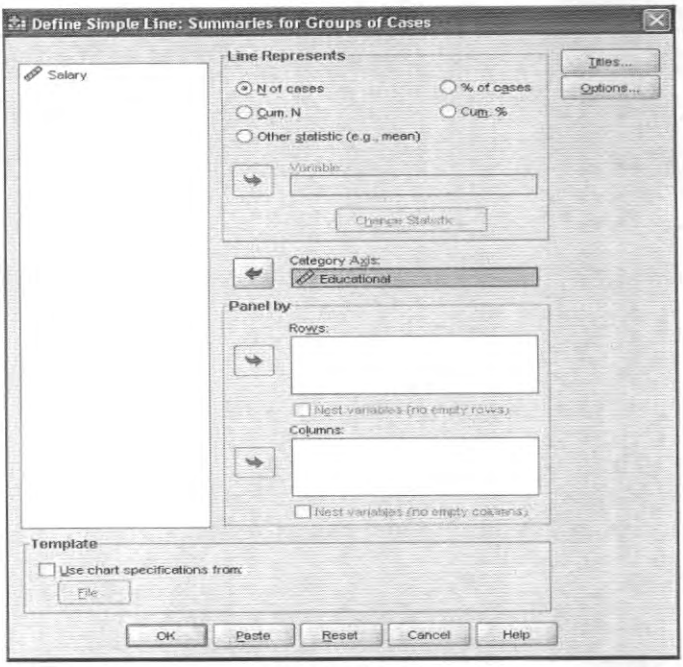
مسیر زیر را انتخاب کنید تا کادر Line Chart باز گردد (شکل ۳-۴).

Graphs ⟶ Legacy Dialogs ⟶ Line



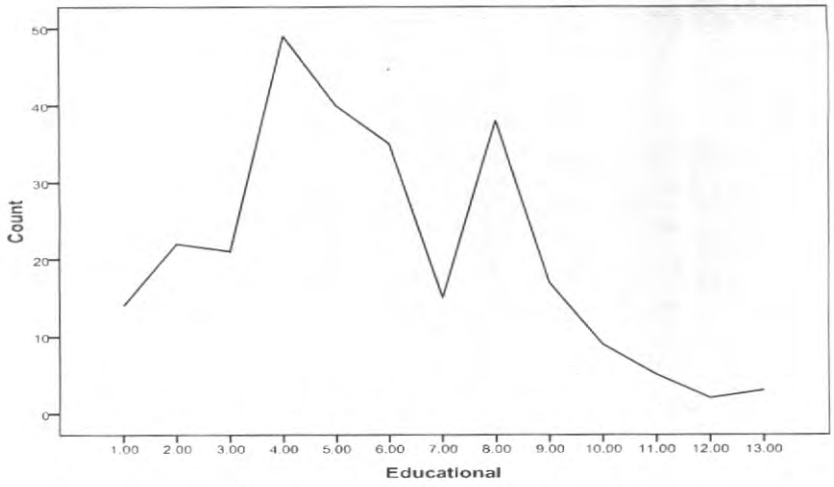
شکل(۳-۴) پنجره Line Chart

سپس مطابق شکل (۳-۵) گزینه Educational را به کادر Category Axis منتقل کنید و برای ترسیم نمودار ستونی Ok را کلیک کنید.



شکل (۳-۵) پنجره Difine Simple Line Summaries for Group of Cases

نتایج حاصل از این خروجی شکل (۳-۶) نموداری است که به خوبی نشان می دهد که اکثریت نمونه های مورد بررسی در این منطقه روستایی دارای چه تحصیلاتی هستند و عده ای قلیل تحصیلات بالایی داشته اند.



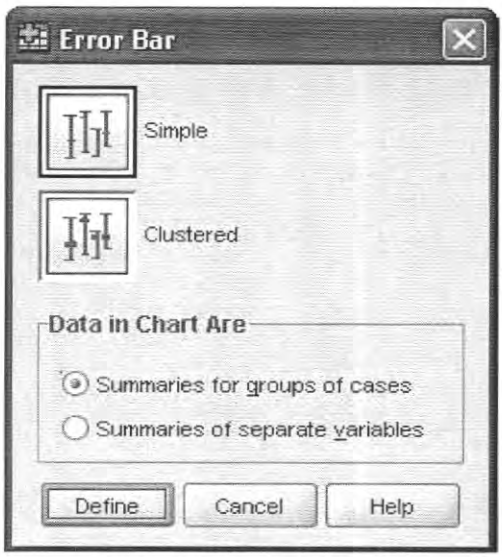
شکل (۳-۶) نمودار خطی متغیر تحصیلات

۳-۳. نمودار ستونی خطا

در این نمودار، میانگین مقادیر هر طبقه توسط یک نقطه نمایش داده می شود و پراکندگی طبقه به کمک خط عمودی که از درون این نقطه می گذرد مشخص می گردد. مثال: میانگین درآمد سالیانه سه گروه کارمندان(۱)، مغازه داران(۲) و تاکسی داران(۳) را در یک شهر مورد پرسش قرار داده ایم. حال برای اینکه تصویری روشن از وضعیت درآمدی هر کدام از این گروه ها ارائه دهیم از نمودار ستونی خطا استفاده کرده ایم که مراحل آن شرح داده می شود.

مسیر زیر را انتخاب کنید تا کادر Error bar همانند شکل (۳-۷) باز گردد.

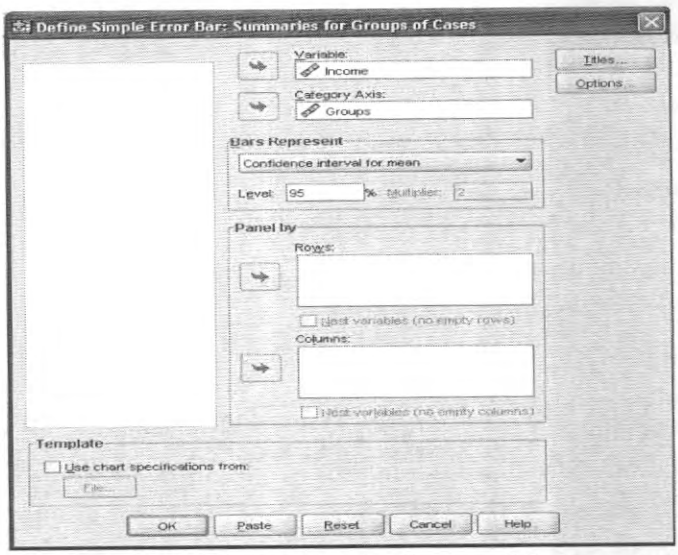
Graphs ⟶ Legacy Dialogs ⟶ Error bar



شکل (۳-۷) پنجره Error bar

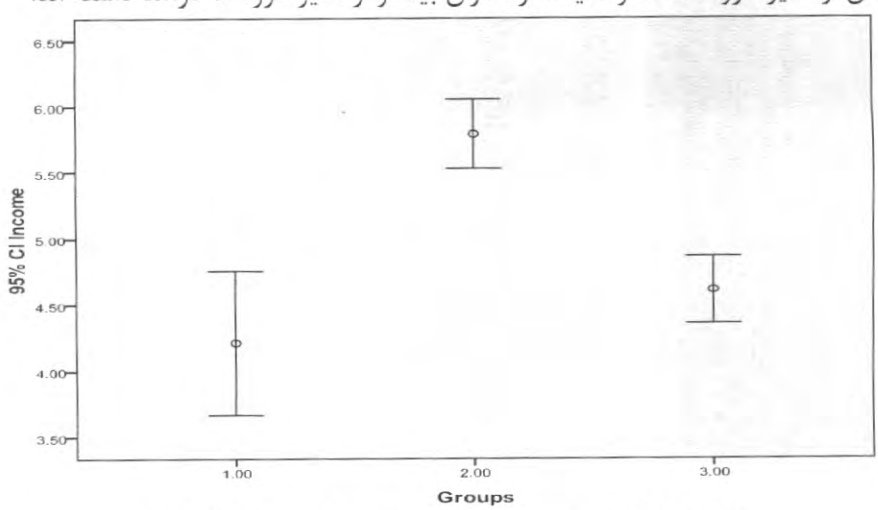
در کادر Error bar می توانید نوع نمودار ستونی خطا (ساده و خوشه ای) و نحوه نمایش داده ها را در خروجی تعیین کنید. نرم افزار بطور پیش فرض نمودار ستونی خطای ساده را فعال کرده است، آن را تغییر ندهید. روی دکمه Define کلیک کنید تا پنجره مربوطه باز گردد. سپس همانند شکل (۳-۸) متغیر Income را به کادر

Variable و متغیر Groups را به کادر Category Axis منتقل کنید. پس از انجام این مراحل گزینه Ok را کلیک کنید.



شکل (۳-۸) پنجره Difine Simple Error Bar Summaries for Group of Cases

خروجی (شکل ۳-۹) نموداری است که نشان می دهد، اولاً پراکندگی درآمد کارمندان بیش از سایر گروه هاست و ثانیاً مغازه داران بیشتر از سایر گروه ها درآمد داشته اند.



شکل (۳-۹) نمودار ستونی خطای متغیر درآمد گروههای شغلی

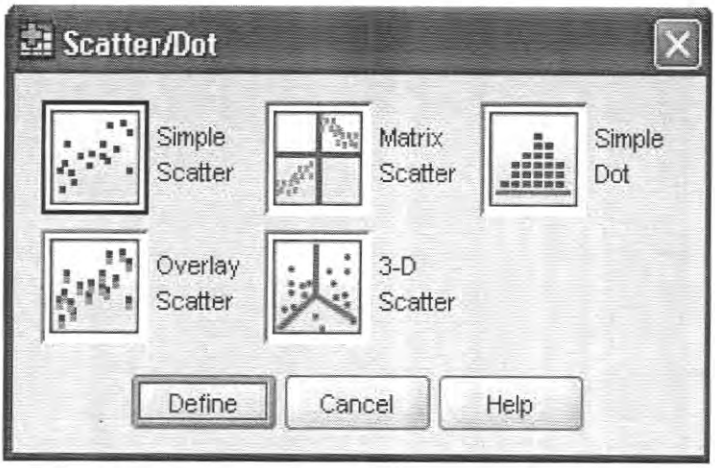
۳-۴. نمودار پراکنش

از نمودار پراکنش معمولاً قبل از محاسبه همبستگی و انجام محاسبات مربوط به رگرسیون استفاده می گردد. این نمودار قابلیت توزیع همزمان دو متغیر کمی را دارا می باشد.

مثال: محققی برای تهیه نمودار پراکنش متغیر سطح زیر کشت کشاورزی در ارتباط با وضعیت جمعیتی در یک منطقه روستایی از نمودار پراکنش استفاده می کند که در ادامه مراحل آن تشریح می گردد.

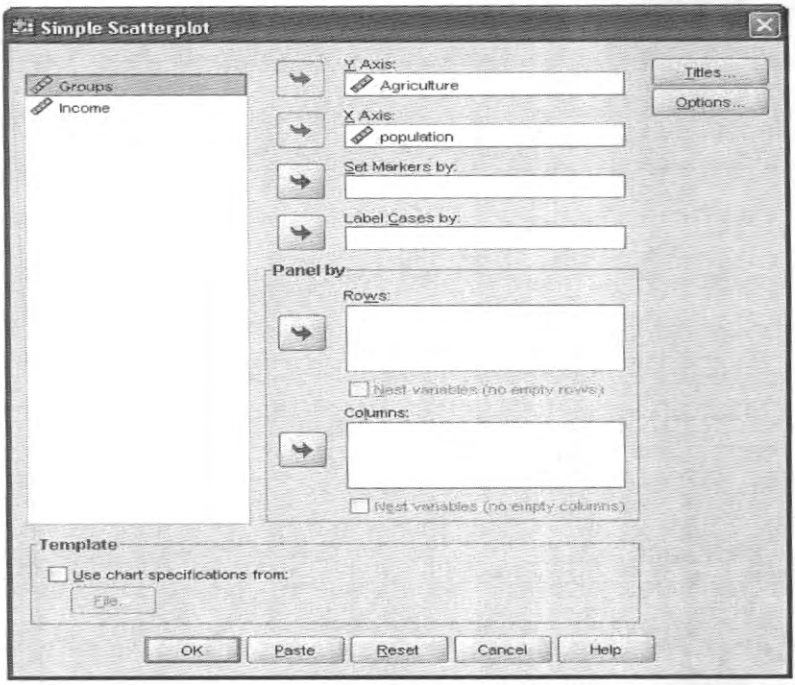
مسیر زیر را انتخاب کنید تا کادر Scatter/Dot همانند (شکل ۳-۱۰) باز گردد.

Graphs ⟶ Legacy Dialogs ⟶ Scatter/Dot



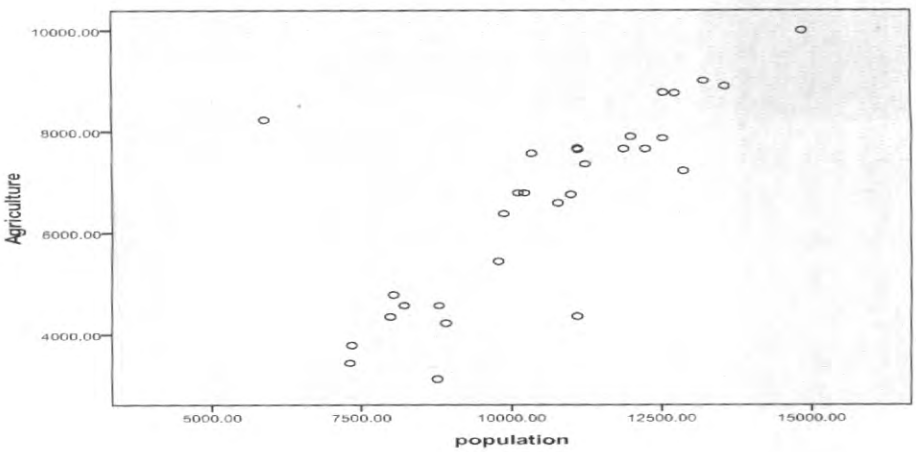
شکل (۳-۱۰) پنجره Scatter/Dot

در کادر Scatter/Dot می توانید نوع نمودار پراکنش را تعیین کنید. در این جا پیش گزیده نرم افزار را تغییر ندهید و روی دکمه Define کلیک کنید تا کادر Simple Scatterplot باز گردد. همانند (شکل ۳-۱۱) متغیر Agriculture را به کادر Y Axis و متغیر population را به کادر X Axis منتقل کنید. سپس جهت اجرای دستور روی Ok کلیک کنید.



شکل (۳-۱۱) پنجره Simple Scatterplot

خروجی (شکل ۳-۱۲) نمودار است که نشان می دهد این دو متغیر با یکدیگر همبستگی قابل ملاحظه ای دارند.



(شکل ۳-۱۲) نمودار پراکنش متغیر سطح زیر کشت کشاورزی در ارتباط با وضعیت جمعیتی

۳-۵. نمودار هیستوگرام (بافت نگار)

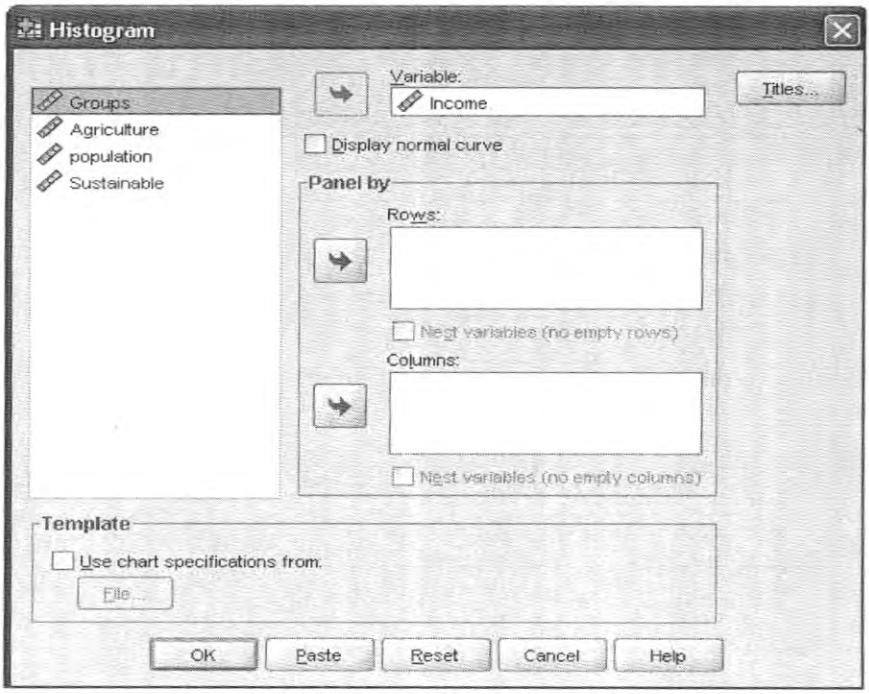
این نمودار نمایش مناسبی از الگوی توزیع داده ها را نشان می دهد. در نمودار هیستوگرام از فراوانی و حدود طبقات استفاده می گردد. به عبارت دیگر در این نمودار محور افقی حدود طبقات و محور عمودی فراوانی را ارائه می دهد.

مثال: محققی به منظور ارائه تصویری روشن از میانگین درآمد ماهیانه کارگران شهرداری از نمودار هیستوگرام استفاده می کند که در ادامه مراحل آن توضیح داده می شود.

مسیر زیر را انتخاب کنید تا کادر Histogram باز گردد.

Graphs ⟶ Legacy Dialogs ⟶ Histogram

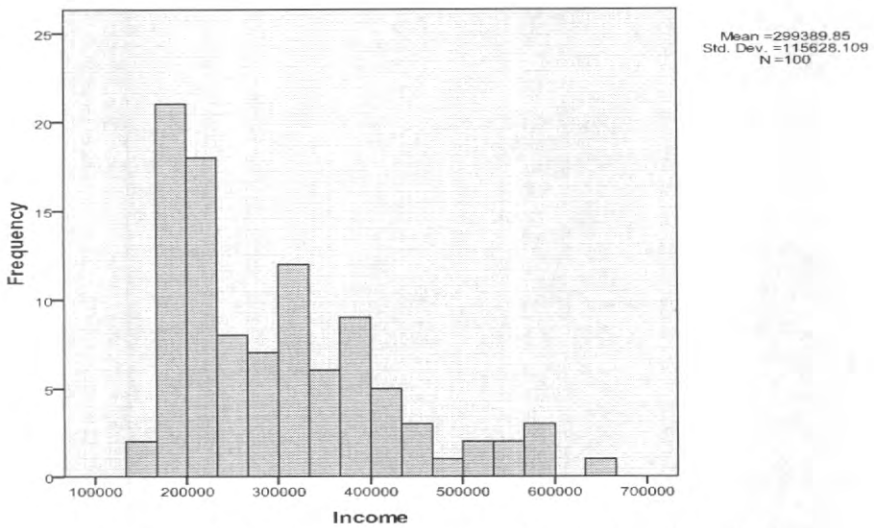
سپس گزینه Income را همانند شکل (۳-۱۳) به کادر Variable منتقل کنید و برای ارائه نتایج Ok را کلیک کنید.



شکل (۳-۱۳) پنجره Histogram

نتیجه (شکل ۳-۱۴) نموداری است که نشان می دهد از ۱۰۰ نفر نمونه مورد بررسی، میانگین درآمدی عده قلیلی بیش از ۶۰۰۰۰۰ تومان می باشد. همچنین در خروجی

نمودار هیستوگرام؛ میانگین، انحراف معیار و تعداد نمونه های مورد بررسی نیز در کنار نمودار ارائه می شوند.



شکل (۳-۱۴) نمودار هیستوگرام وضعیت درآمد ماهیانه کارگران شهرداری

فصل چهارم

نرمال بودن توزیع‌ها

پیش از شرح این بحث به بررسی مفهوم Significance می‌پردازیم.

۴-۱. مفهوم Significance:

معنی داری(Significance) که به اختصار آن را با Sig نشان می دهند، میزان خطایی است که در رد فرضیه H_0 مرتکب می شویم.

Sig به p-value نیز معروف است. اگر میزان Sig کمتر از ۵ درصد باشد فرض H_0 رد می شود و اگر این مقدار از ۵ درصد بیشتر باشد فرض H_0 پذیرفته می شود.

نکته: ممکن است یک فرض صفر در یک سطح خطا ، مثلاً ۱ درصد، معنی دار نباشد ولی در سطح دیگری، مثلاً ۵ درصد، معنی دار باشد. بنابراین انتخاب سطح معنی داری در رد یا تأیید H_0 عامل بسیار مهمی می باشد.

۴-۲. نرمال بودن توزیع ها

در بسیاری از آزمون های آماری نیاز به این داریم که توزیع داده ها نرمال باشد. گاهی آزمون نرمال بودن برای توزیع داده هایی به کار می رود که دارای مقیاس فاصله ای–نسبی (پارامتریک) و گاهی هم برای داده های ترتیبی و رتبه ای (ناپارامتریک) استفاده می گردد. به سه شکل می توان فهمید که داده ها دارای توزیع نرمال هستند یا خیر؟

۱- استفاده از یک مقدار عددی برای تشخیص نرمال بودن یا نبودن

۲- سنجش نرمال بودن از طریق بصری به واسطه شکل و نمودار بدست آمده

۳- استفاده از آزمون های آماری مناسب

یکی از روش های بررسی نرمال بودن استفاده از ضریب چولگی (Skewness) می باشد که در فصل ۲ تشریح گردید. اما اگر بخواهیم از آزمون های آماری مناسب برای بررسی توزیع داده ها (نیکوئی برازش) استفاده کنیم، دو آزمون کای ۲ (Chi square) و کولموگراف- اسمیرنوف(KS) را می توان مورد استفاده قرار داد.

فرض های آماری در این آزمون ها به این شرح می باشند:

H_0: فراوانی های تجربی و نظری یکسان هستند

H_1: فراوانی های تجربی و نظری یکسان نیستند

۴-۲-۱. آزمون تک متغیری مجذور کای(X^2) خی ۲:

این آزمون بیانگر شرایطی است که در آن یک متغیر طبقه ای واحد با بیش از دو ارزش وجود داشته باشد. این آزمون برای بررسی نوع توزیع داده ها در متغیر های کیفی یا متغیر های کمی دسته بندی شده مورد استفاده قرار می گیرد. آزمون تک متغیری مجذور کای این امکان را فراهم می کند که معلوم کنیم آیا تفاوتی میان فراوانی های مورد انتظار و فراوانی های مورد مشاهده در یک گروه نمونه نظام دار است یا حاصل شانس می باشد.

شرایط انجام این آزمون:

۱- هر یک از مشاهدات در هر طبقه باید کاملاً تصادفی و مستقل از مشاهده های دیگر نمونه برداری شده باشد.

۲- هر یک از مشاهدات باید در یک طبقه از متغیر مورد نظر قرار گیرد.

۳- آزمون کای دو هنگامی دقیق خواهد بود که فراوانی های مورد انتظار در هیچ یک از مقوله ها کمتر از ۵ نباشد و یا فراوانی های مورد انتظار در بیش از ۲۰ درصد مقوله ها کمتر از ۵ و در هیچ یک از آنها کمتر از یک نباشد.

مجذور کای دو روشی آماری برای بیان اختلاف بین تعداد موارد مشاهده شده در سه طبقه یا بیشتر می باشد و از طریق رابطه (۴-۱) محاسبه می گردد.

$$X^2 = \sum_{i=1}^{n} \frac{(O-E)^2}{E} \qquad (۴-۱)$$

در این رابطه O فراوانی مشاهده شده و E فراوانی مورد انتظار می باشد.

مثال: انگیزه داشتن شغل دوم از بدون انگیزه تا انگیزه خیلی زیاد(۰= بدون انگیزه، ۱=خیلی کم، ۲=کم، ۳=تاحدی، ۴= زیاد، ۵= خیلی زیاد) را برای ۱۵۰ کارمند استانداری بدست آورده ایم. می خواهیم بدانیم که آیا انگیزه داشتن شغل دوم دارای توزیع یکنواختی در میان کارمندان استانداری است یا خیر؟ در واقع می خواهیم بدانیم انگیزه داشتن شغل دوم دارای توزیع نرمالی است یا خیر؟

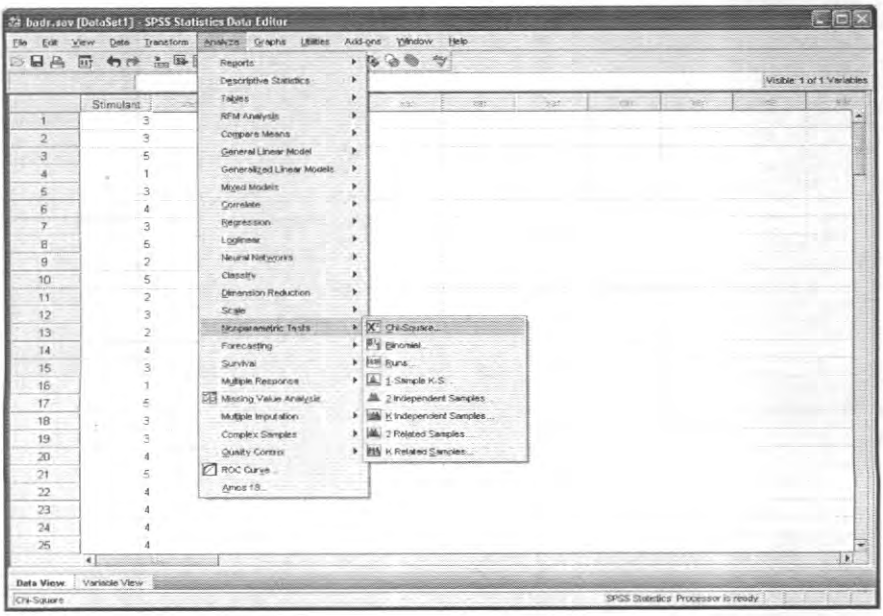
فرض آماری برای این سئوال بدین شکل مطرح می گردد:

H_0: انگیزه داشتن شغل دوم دارای توزیع یکنواختی بین کارمندان است

H_1: انگیزه داشتن شغل دوم دارای توزیع یکنواختی بین کارمندان نیست

پس از وارد کردن داده ها همانند شکل (۴-۱) از آزمون chi square استفاده می کنیم.

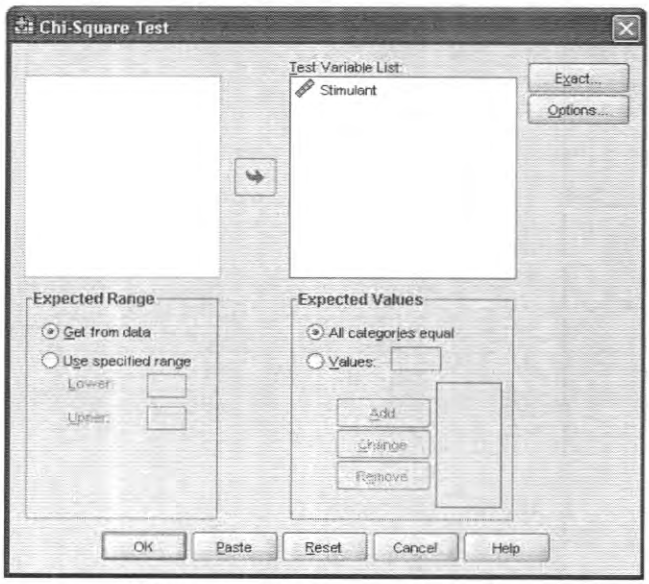
Analyze ⟶ Nonparametric Tests ⟶ Chi square



شکل (۴-۱) پنجره SPSS Statistics Data Editor 1

پس از انتخاب این منو پنجره ای تحت عنوان Chi-square Test باز می گردد. همانند شکل (۴-۲) متغیر Stimulant را به کادر Test Variable List منتقل کنید. در کادر Expected Values گزینه All categories equal به صورت پیش فرض نرم

افزار وجود دارد و بدین معنی است که H_0 براساس توزیع داده ها در طبقات مختلف (یکنواختی توزیع داده ها) قرار دارد. در صورتی که توزیع دیگری مد نظر باشد، می توان توزیع مورد نظر را به کمک گزینه Values وارد کرد که این مسأله در ادامه با یک مثال تشریح می گردد. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۴-۲) پنجره Chi-square Test1

نتیجه آزمون شامل دو خروجی است. خروجی اول جدول(۴-۱) به ترتیب فراوانی مشاهده شده، فراوانی مورد انتظار و مانده تفاوت آنها را برای طبقات مختلف نشان می دهد.

جدول(۴-۱) نتایج توصیفی آزمون خی ۲

Stimulant

	Observed N	Expected N	Residual
No option	3	25.0	-22.0
very low	7	25.0	-18.0
low	24	25.0	-1.0
somewhat	51	25.0	26.0
high	48	25.0	23.0
very high	17	25.0	-8.0
Total	150		

در خروجی دوم (جدول۴-۲) مقدار آماره کای دو، درجه آزادی و sig ارایه می شود. از آنجایی که sig آزمون کمتر از ۵ درصد است H_0 رد می شود و ادعای یکنواخت نبودن انگیزه داشتن شغل دوم در بین کارمندان پذیرفته می شود.

جدول(۴-۲) خروجی نهایی آزمون خی ۲

Test Statistics

	Stimulant
Chi-Square	83.120[a]
df	5
Asymp. Sig.	.000

مثال

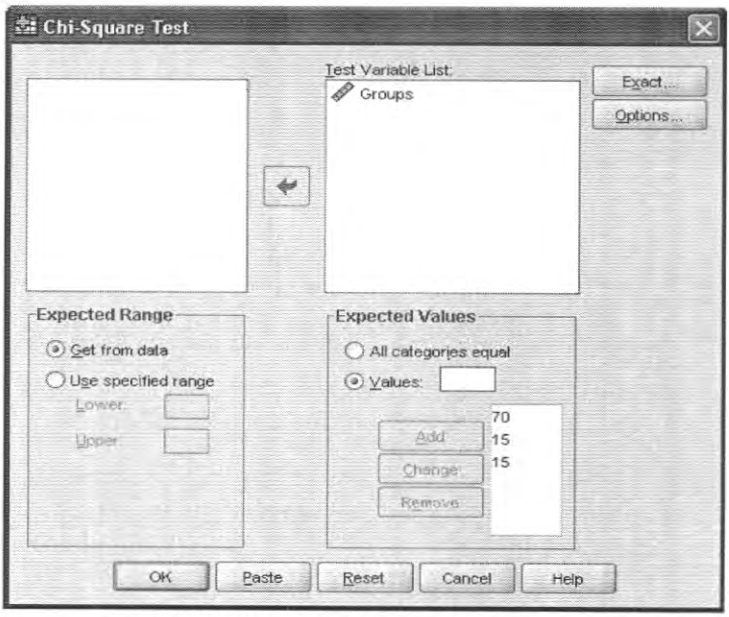
محققی بعد از بازدیدی ظاهری از یک روستا ادعا می کند که توزیع جمعیت روستا براساس ۷۰ درصد کشاورز، ۱۵ درصد دامدار و ۱۵ درصد باغدار است.

فرض آماری برای این سئوال بدین شکل مطرح می گردد:

H_0: توزیع جمعیت روستا براساس ۷۰ درصد کشاورز، ۱۵ درصد دامدار و ۱۵ درصد باغدار است

H_1: دست کم یکی از نسبت ها صحیح نمی باشد

همانند روش شرح داده شده در مثال قبل پنجره Chi-square Test را باز کنید و متغیر Groups را به کادر Test Variable List منتقل کنید. در کادر Expected Values گزینه Values را انتخاب کنید. از آنجا که گروهها به صورت کشاورز(۱)، دامدار(۲) و باغدار(۳) تعریف شده است، باید ابتدا عدد ۷۰ را در کادر Values وارد کرده و سپس روی Add کلیک کنید تا عدد ۷۰ به کادر زیرین آن منتقل شود. این کار را برای اعداد ۱۵ دوبار دیگر تکرار کنید شکل (۴-۳). سپس جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۴-۳) پنجره Chi-square Test2

نتیجه آزمون شامل دو خروجی است. خروجی اول (جدول۴-۳) به ترتیب فراوانی مشاهده شده، فراوانی مورد انتظار براساس توزیع تعریف شده و مانده تفاوت آنها را برای طبقات مختلف نشان می دهد.

جدول(۴-۳) فراوانی مشاهده شده، فراوانی مورد انتظار براساس توزیع تعریف شده و مانده تفاوت آنها

Groups

	Observed N	Expected N	Residual
Agri	29	105.0	-76.0
Anim	54	22.5	31.5
Garden	67	22.5	44.5
Total	150		

در خروجی دوم (جدول۴-۴) مقدار آماره کای دو، درجه آزادی و sig ارایه می شود. از آنجایی که sig آزمون کمتر از ۵ درصد است H_0 رد می شود و ادعای اینکه دست کم یکی از نسبت ها صحیح نمی باشد پذیرفته می شود.

جدول (۴-۴) نتایج نهایی آزمون خی ۲ براساس توزیع تعریف شده

Test Statistics

	Groups
Chi-Square	187.121[a]
df	2
Asymp. Sig.	.000

۴-۲-۲. آزمون کولموگراف- اسمیرنوف(KS)

آزمون **KS** نظیر آزمون نیکوئی برازش X^2 می باشد و معمولاً زمانی از این آزمون استفاده می شود که حجم نمونه کمتر از ۵۰ باشد. در نرم افزار SPSS می توان انواع توزیع های نرمال[1]، یکنواخت[2]، پواسون[3] و نمایی[4] را مورد بررسی قرار داد. توزیع نرمال برای متغیرهای کمی پیوسته، توزیع یکنواخت برای متغیرهای کمی گسسته و پیوسته، توزیع پواسون برای متغیرهای کمی گسسته و اعداد صحیح و توزیع نمایی برای متغیرهای کمی مورد استفاده قرار می گیرد. آماره آزمون کولموگراف- اسمیرنوف از طریق رابطه (۴-۲) محاسبه می گردد.

$$D = SUP_x \left|F_o(x) - F_e(x)\right| \qquad (۴-۲)$$

در این رابطه $F_o(x)$ فراوانی تجمعی نسبی یک نمونه تصادفی و $F_e(x)$ فراوانی تجمعی نظری مفروض می باشد.

مثال: محققی می خواهد بداند که آیا توزیع درآمد کشاورزان در یک ناحیه روستایی نرمال است یا خیر؟

---

1 . normal distribution
2 . uniform distribution
3 . poisson distribution
4 . exponential distribution

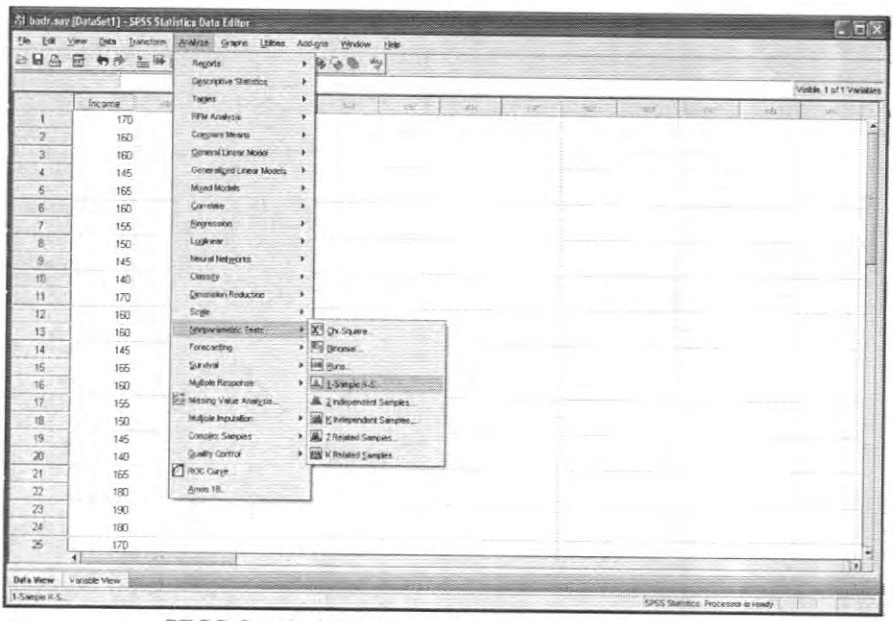
فرض آماری برای این سئوال بدین شکل مطرح می گردد:

H_0: توزیع درآمد کشاورزان نرمال است

H_1: توزیع درآمد کشاورزان نرمال نیست

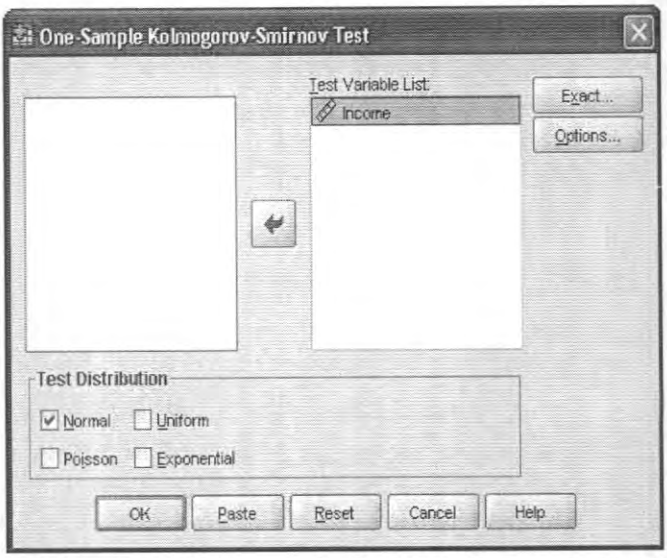
پس از وارد کردن داده ها همانند شکل (۴-۴) از آزمون KS استفاده می کنیم.

Analyze ⟶ Nonparametric Tests ⟶ 1-Sample K-S



شکل (۴-۴) پنجره SPSS Statistics Data Editor 2

پس از انتخاب این منو پنجره ای تحت عنوان One-Sample Kolmogorov-Smirnov Test باز می گردد. همانند شکل (۴-۵) متغیر Income را به کادر Test Variable List منتقل کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۴-۵) پنجره One-Sample Kolmogorov-Smirnov Test

نتیجه آزمون شامل یک خروجی جدول(۴-۵) است که به ترتیب تعداد داده ها، پارامترهای مورد نظر مانند میانگین و انحراف معیار در توزیع نرمال، قدر مطلق مقدار بیشترین انحراف، بیشترین انحراف مثبت، بیشترین انحراف منفی، مقدار آماره Z و مقدار sig را ارایه می کند. با توجه به اینکه sig کمتر از ۵ درصد است، H_0 رد شده و ادعای نرمال بودن توزیع درآمد کشاورزان پذیرفته نمی شود.

جدول(۴-۵) نتایج آزمون کولموگراف - اسمیرنوف

One-Sample Kolmogorov-Smirnov Test

		Income
N		38
Normal Parameters[a,,b]	Mean	159.47
	Std. Deviation	10.767
Most Extreme Differences	Absolute	.230
	Positive	.191
	Negative	-.230
Kolmogorov-Smirnov Z		1.418
Asymp. Sig. (2-tailed)		.036

فصل پنجم

آزمون‌های میانگین

این آزمونها برای آزمون معنی داری اختلاف بین میانگین های دو یا چند گروه به کار می‌رود. در حالت کلی بسته به توزیع جامعه مادر می توان این آزمون ها را به آزمونهای پارامتری که توزیع جامعه در آن مشخص است و آزمونهای ناپارامتری که توزیع جامعه در آنها مشخص نیست تقسیم کرد.

آزمون تی، یک آزمون پارامتری است که به منظور تعیین معنی داری تفاوت بین دو میانگین به کار می رود. سه نوع آزمون تی وجود دارد:

۱- آزمون تی تک نمونه ای

۲- آزمون تی مستقل

۳-آزمون تی وابسته یا تی جفت شده

۵-۱. آزمون تی تک نمونه ای :

آزمون تی تک نمونه ای (آزمون مقایسه میانگین یک جامعه با یک عدد) از آزمون هایی است که در اغلب تحقیق های کاربردی مورد استفاده قرار می گیرد. این آزمون تعیین می‌کند که آیا میانگین مشاهده شده در مقایسه با یک مقدار تعیین شده یا یک مقدار استاندارد تفاوت معنی داری دارد یا نه؟ این آزمون ممکن است به این شرح انجام شود که ابتدا عددی را برای میانگین جامعه فرض می کنیم سپس از طریق آزمون تعیین می‌کنیم که این فرض صحیح است یا نه. از این آزمون می توان برای مقایسه میانگین یک نمونه با یک عدد ثابت نیز استفاده کرد.

فرض های آماری در این صورت عبارتند از:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

آماره آزمون T تک نمونه ای از طریق رابطه (۵-۱) بدست می آید.

(۵-۱)

$$t = \frac{\bar{x} - \mu_0}{S_{\bar{x}}}$$

همچنین برآورد $S_{\bar{x}}$ طبق رابطه (۵-۲) محاسبه می گردد.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

(۵-۲)

آزمون تی تک نمونه ای را آزمون تی استودنت نیز می نامند. با توجه به تفاسیر فوق برای استفاده از این آزمون شرایط زیر باید وجود داشته باشد:

۱- داده ها باید دارای مقیاس فاصله ای – نسبی باشند

۲- جامعه مادر نمونه ها دارای توزیع نرمال باشد

۳- واریانس جامعه های مورد بررسی برابر باشند

مثال: شرط انجام کشت زعفران در یک شهرستان برخورداری منطقه از دمایی برابر با ۱۰ درجه سانتیگراد است برای این منظور ۱۵ ایستگاه انتخاب شده اند که دمای متوسط آنها به شرح زیر ثبت شده است:

۱۳ ۷ ۸ ۱۵ ۱۲ ۷ ۱۰ ۱۳ ۹ ۱۴ ۱۱ ۹ ۸ ۱۰ ۱۲

حال با توجه به اعداد به دست آمده می توان گفت که میانگین کل ایستگاه های هواشناسی این شهرستان برابر با ۱۰ است یا خیر؟

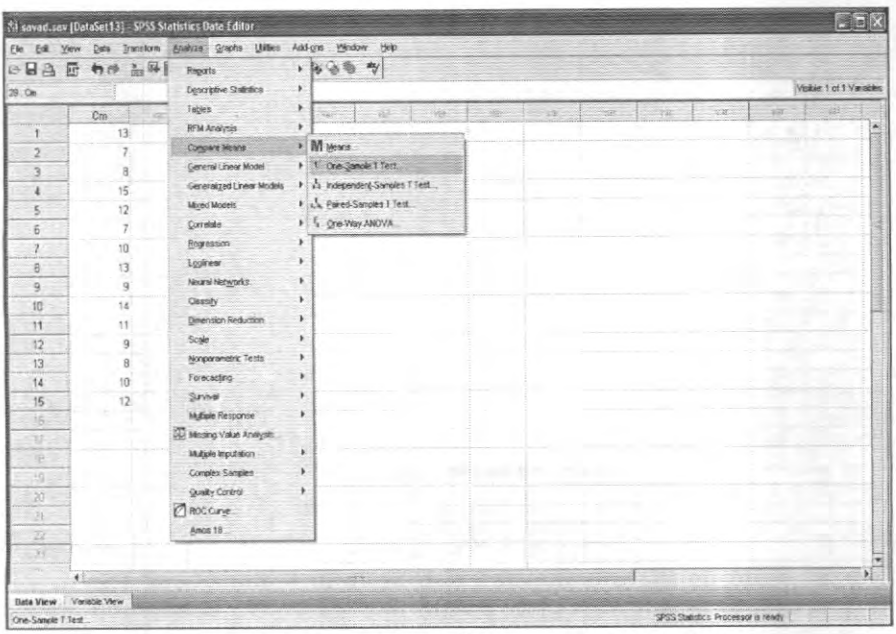
فرض آماری برای این سئوال بدین شکل مطرح می گردد:

$H_0 : \mu = ۱۰$

$H_1 : \mu \neq ۱۰$

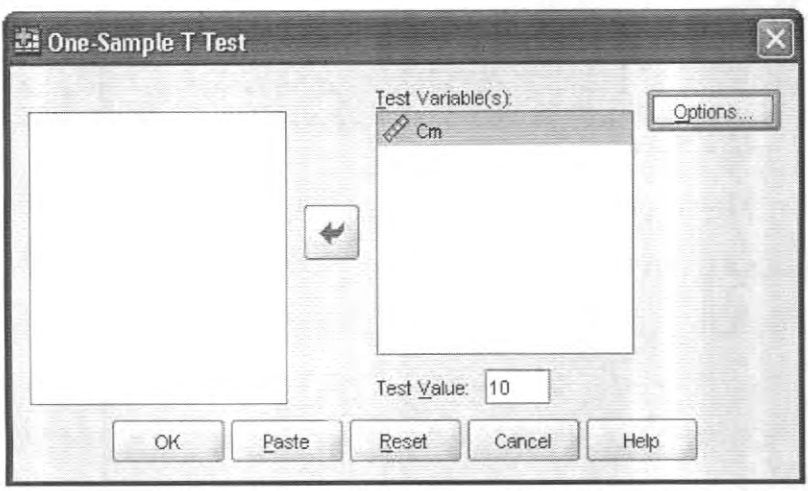
پس از وارد کردن داده ها همانند شکل (۵-۱) از آزمون T تک نمونه ای استفاده می کنیم.

Analyze ⟶ Compare Means ⟶ One-Sample T Test



شکل (۵-۱) پنجره SPSS Statistics Data Editor 1

پس از انتخاب این منو پنجره One-Sample T Test باز می گردد. همانند شکل (۵-۲) متغیر Cm را به کادر Test Variable(s) منتقل کنید. پس از این کار عدد ۱۰ را در قسمت Test Value وارد کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۵-۲) پنجره One-Sample T Test

نتیجه آزمون شامل دو خروجی است. خروجی اول (جدول ۵-۱) به ترتیب تعداد مشاهدات، میانگین، انحراف معیار و خطای معیار میانگین را نشان می دهد.

جدول(۵-۱) نتایج توصیفی آزمون تی تک نمونه ای

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Cm	15	10.53	2.560	.661

خروجی دوم (جدول۵-۲) نتایج آزمون را براساس آمار استنباطی بیان می کند. در این خروجی مقدار t، درجه آزادی، Sig، اختلاف میانگین و مقادیر حد پایین و بالا ارایه می-گردند. با توجه به اینکه Sig بزرگ تر از ۵ درصد است، H_0 رد نمی شود و ادعای برخورداری منطقه از دمایی برابر با ۱۰ درجه سانتیگراد را نمی توان رد کرد.

جدول(۵-۲) نتایج نهایی آزمون تی تک نمونه ای

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Cm	.807	14	.433	.533	-.88	1.95

نکته: این آزمون سه سطح را ایجاد می کند یا $\mu = \mu_0$ و شاخص X در شرایط متوسطی قرار دارد و یا $\mu \neq \mu_0$ یعنی میانگین مشاهده شده شاخص X در حد متوسط قرار ندارد در این صورت باید از مقادیر حد بالا و حد پایین استفاده کرد که:

۱- هرگاه حد بالا و پایین مثبت باشد، میانگین از مقدار مشاهده شده بزرگتر است.

۲- هرگاه حد بالا و پایین منفی باشد، میانگین از مقدار مشاهده شده کوچکتر است.

۵-۲. آزمون تی مستقل(آزمون میانگین های دو نمونه مستقل)

گاهی هدف بررسی اختلاف میانگین میان دو جامعه متفاوت است و می خواهیم بر اساس نمونه های گرفته شده از این دو جامعه، اختلاف معنی دار بین دو جامعه را با هم مقایسه می کنیم. یا به عبارت دیگر محقق بر آن است تا آزمون خود را برای هر داده از یک نمونه با داده های مخصوص از نمونه دیگر مقایسه کند. در بررسی های آماری سؤالاتی به این صورت مطرح می شود:

آیا میزان درآمد کارمندان و مغازه داران در یک شهر برابر است یا خیر؟

آیا احساس تعلق مکانی در بین زنان و مردان یک روستا یکسان است یا خیر؟

آیا میزان برداشت کشت گندم و برنج در یک شهرستان برابر است یا خیر؟

آیا سطح رضایت از سکونت گاه در بین زنان و مردان یک روستا یکسان است یا خیر؟

در سوالات بالا به دنبال این هستیم که آیا میانگین دو جامعه مستقل، برابر است یا خیر؟

فرض های آماری در این آزمون عبارتند از:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

شرایط انجام این آزمون:

۱- متغیر وابسته باید دارای توزیع نرمال و مقیاس فاصله ای- نسبی باشد.

۲- متغیر مستقل باید در سطح کیفی و به صورت مقیاس اسمی باشد.

۳- واریانس داده ها در دو جامعه باید یکسان باشد که برابری یا نابرابری واریانس ها از طریق آزمون leven's test انجام می شود که spss این کار را با استفاده از آماره F فیشر (رابطه ۵-۳) انجام می دهد و نتیجه را در خروجی نشان می دهد.

$$F = \frac{S_1^2}{S_2^2} \quad (۵-۳)$$

آماره t جهت آزمون تساوی میانگین دو جامعه در دو حالت تساوی و عدم تساوی واریانس دو جامعه مورد نظر را حساب می کند که در حالت تساوی واریانس ها از رابطه

(۵-۴) و در حالت عدم تساوی از رابطه (۵-۵) استفاده می گردد. لازم به ذکر است که در حالت عدم تساوی درجه آزادی از رابطه (۵-۶) محاسبه می شود.

$$S_P = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \qquad (۴-۵)$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (۵-۵)$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad (۶-۵)$$

$$df = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[S_1^2/n_1\right]^2}{n_1 - 1} + \frac{\left[S_2^2/n_2\right]^2}{n_2 - 1}}$$

مثال: انگیزه مهاجرت را در یک شهرستان بین دو گروه سنی ۴۰-۲۰ سال و ۶۰-۴۰ سال در نمونه ای به تعداد ۳۰ نفر برای هر گروه بر اساس طیف لیکرت(از ۰ تا ۱۰۰) که ۰ به معنای کمترین انگیزه و ۱۰۰ به معنای بالاترین انگیزه است حساب کرده ایم. می خواهیم بدانیم که انگیزه مهاجرت در بین دو گروه یکسان است؟

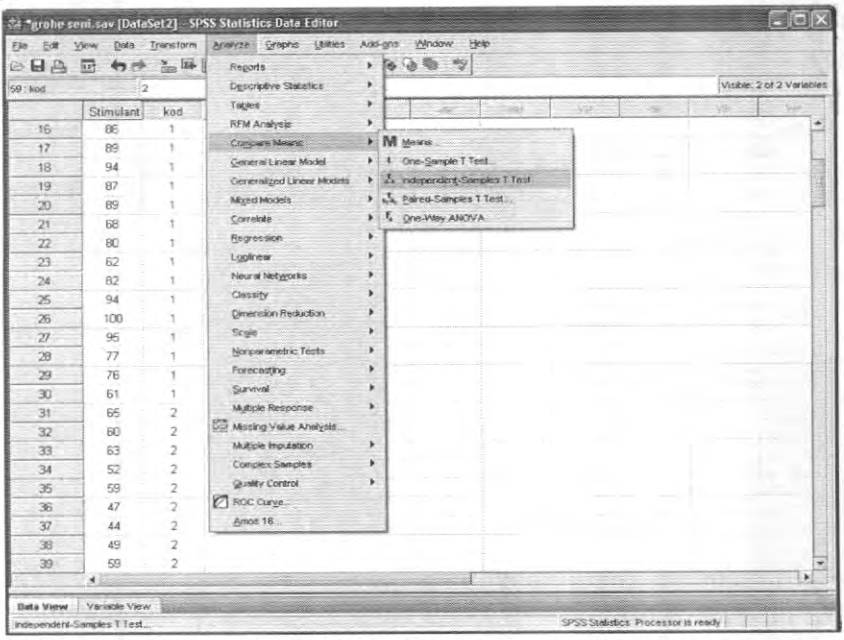
فرض های آماری در این سئوال عبارتند از:

H_0: انگیزه مهاجرت در دو گروه یکسان است

H_1: انگیزه مهاجرت در دو گروه یکسان نیست

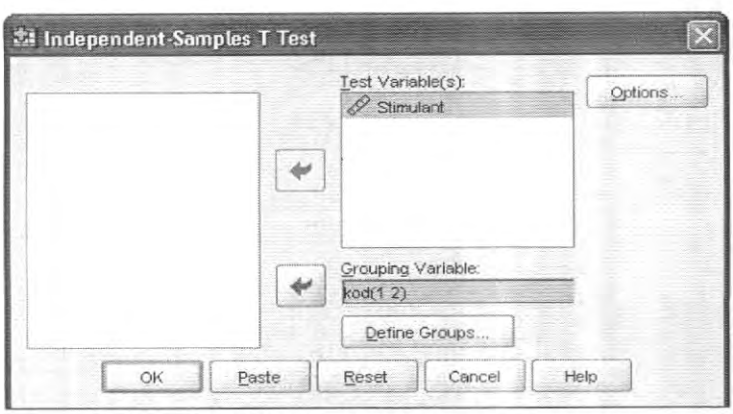
پس از وارد کردن داده ها همانند شکل (۵-۳) از آزمون T مستقل استفاده می کنیم.

Analyze ⟶ Compare Means ⟶ Independent-Sample T Test



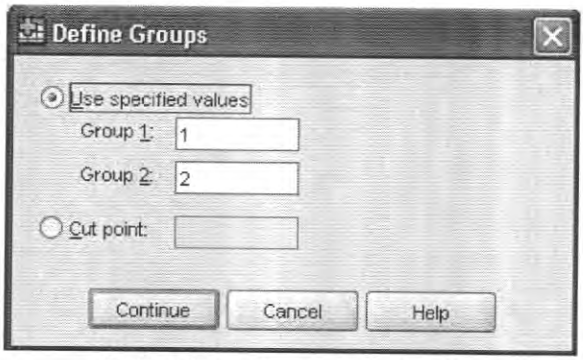
شکل (۵-۳) پنجره SPSS Statistics Data Editor 2

پس از انتخاب این منو پنجره Independent-Sample T Test باز می گردد. همانند شکل (۵-۴) متغیر Stimulant را به کادر Test Variable(s) منتقل کنید. سپس متغیر Kod را به کادر Grouping Variable انتقال دهید.



شکل (۵-۴) پنجره Independent-Sample T Test

پس از این کار بر روی دکمه Define Groups کلیک کنید تا پنجره ای با همین عنوان باز گردد. همانند شکل(۵-۵) در کادر Group 1 عدد ۱ و در کادر Group 2 عدد ۲ را وارد کرده و Continue را کلیک کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۵-۵) پنجره Define Groups

نتیجه آزمون شامل دو خروجی است. خروجی اول (جدول ۵-۳) آمار توصیفی مربوط به دو گروه را نشان می دهد.

جدول(۵-۳) آمار توصیفی آزمون تی مستقل

Group Statistics

	kod	N	Mean	Std. Deviation	Std. Error Mean
Stimulant	20-40	30	80.17	11.396	2.081
	40-60	30	57.43	9.807	1.791

خروجی دوم (۵-۴) نتایج آزمون را براساس آمار استنباطی بیان می کند. در این خروجی ابتدا به بررسی تساوی واریانس دو گروه براساس آزمون leven's پرداخته می شود، سپس نتایج آزمون تساوی میانگین دو گروه برای هر دو حالت تساوی و عدم تساوی واریانس ارایه می گردند.

از آنجا که sig آزمون لون بزرگ تر از ۵ درصد است لذا فرض برابری واریانس ها را می-پذیریم. بنابراین اطلاعات سطر اول را برای نتیجه گیری در مورد این مسأله مورد بررسی

قرار می دهیم. با توجه به اینکه sig آزمون کمتر از ۵ درصد است، پس فرض H_0 رد می‌شود و ادعای یکسان بودن انگیزه مهاجرت در دو گروه سنی پذیرفته نمی شود.

جدول(۵-۴) نتایج نهایی آزمون تی مستقل

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Stimulant	Equal variances assumed	.785	.379	8.282	58	.000	22.733	2.745	17.239	28.228
	Equal variances not assumed			8.282	56.740	.000	22.733	2.745	17.236	28.231

نکته:

۱. هرگاه حد پایین و بالا مثبت باشد تفاوت میانگین دو جامعه بزرگ تر از صفر است و میانگین جامعه اول از جامعه دوم بزرگ تر می باشد.

۲. هرگاه حد پایین و بالا منفی باشد تفاوت میانگین دو جامعه کمتر از صفر است و میانگین جامعه اول از جامعه دوم کوچک تر می باشد.

۳. هرگاه حد بالا مثبت و حد پایین منفی باشد تفاوت میانگین دو جامعه معنی دار نبوده و تساوی میانگین دو جامعه رد نمی شود.

۵-۳. آزمون تی وابسته(آزمون میانگین های نمونه های جفتی)

در بسیاری از بررسی های آماری اتفاق می افتد که با داده هایی مواجه هستیم که برای هر عنصر نمونه یک جفت متغیر اندازه گیری شده است یا به عبارتی مربوط بر طرح های ماقبل و مابعد می باشد. از این آزمون برای نشان دادن تأثیر یک نوع مداخله استفاده می-شود. به عبارت دیگر زمانی از این آزمون استفاده می کنیم که بخواهیم روی یک نمونه دو بار مختلف (قبل و بعد) مشاهداتی را صورت بدهیم. به عنوان مثال:

آیا انگیزه مهاجرت قبل و بعد از دادن امکانات در یک شهرستان یکسان است یا خیر؟

آیا میزان برداشت محصول قبل و بعد از دادن کود در زمینهای یک روستا یکسان است یا خیر؟

آیا میزان مشارکت روستاییان در طرح های عمرانی قبل و بعد از دادن آموزش یکسان است یا خیر؟

فرض های آماری در این آزمون عبارتند از:

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

علاوه بر شرایطی که برای آزمون های t گفته شد شرط دیگر برای این آزمون این است هر زوج مشاهده باید مستقل از زوج های دیگر باشند. بدیهی است در این ساختار (X1 و Y1)، (Xn و Yn) ... مستقل از یکدیگر هستند و انحراف استاندارد مجموع تفاوت در داده های نمونه از طریق رابطه (۵-۷) به دست می آید.

$$Sd = \sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N-1}} \qquad (۵-۷)$$

آزمون t برای دو گروه نمونه مانند آزمون t تک نمونه ای است با این تفاوت که در اینجا در مورد توزیع خاص نمره های تفاوت به کار برده می شود و از رابطه (۵-۸) بدست می‌آید. آزمون t وابسته دارای درجه آزادی n-1 است.

$$t = \frac{\overline{X}D - \mu_0 D}{\sqrt{n}} \quad \text{یا} \quad t = \frac{\overline{D} - d_0}{\sqrt{n}} \qquad (۵-۸)$$

در این رابطه $\overline{X}D$ و $\overline{D}$ میانگین مجموع تفاوت های دو گروه نمونه، Sd انحراف استاندارد در داده های نمونه و n تعداد زوج هاست.

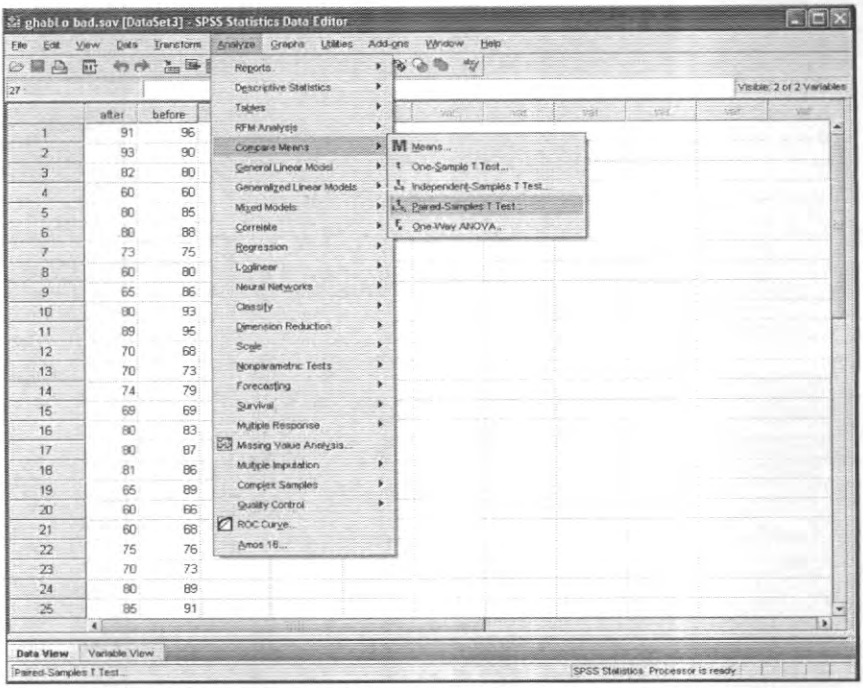
مثال: انگیزه مهاجرت در یک روستا را در قبل و بعد از دادن امکانات رفاهی بر اساس طیف لیکرت (از ۰ تا ۱۰۰) برای نمونه ای ۲۵ نفری بدست آورده ایم. حال می خواهیم بدانیم که با توجه به داده های بدست آمده، آیا دادن امکانات انگیزه مهاجرت را کاهش داده است؟

H_0: انگیزه مهاجرت در قبل و بعد از دادن امکانات رفاهی تغییری نکرده است

H_1: انگیزه مهاجرت در قبل و بعد از دادن امکانات رفاهی تغییر کرده است

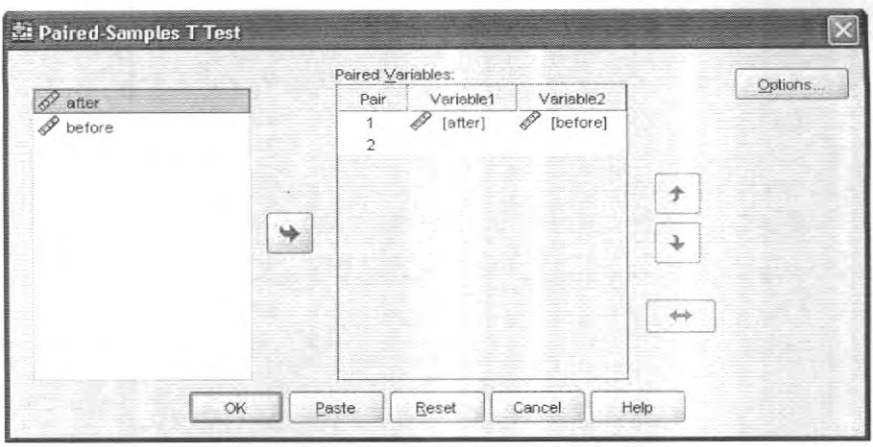
پس از وارد کردن داده ها مسیر شکل (۵-۶) را برای اجرای آزمون T وابسته استفاده می‌کنیم.

Analyze ⟶ Compare Means ⟶ Paired-Sample T Test



شکل (۵-۶) پنجره SPSS Statistics Data Editor 3

پس از انتخاب این منو پنجره Paired-Sample T Test باز می گردد. همانند شکل (۵-۷) متغیرهای after و before را به کادر Paired Variables منتقل کنید. سپس برای اجرای آزمون ok را کلیک کنید.



شکل (۵-۷) پنجره Paired-Sample T Test

نتیجه شامل سه خروجی است. خروجی اول (جدول ۵-۵) آمار توصیفی دو جامعه را نشان می دهد.

(جدول ۵-۵) آمار توصیفی آزمون تی جفت نمونه ای

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	after	74.88	25	9.846	1.969
	before	81.00	25	9.941	1.988

در خروجی دوم (جدول ۵-۶) به ترتیب تعداد مشاهدات، ضریب همبستگی بین دو متغیر و sig ضریب همبستگی ارایه می گردند.

جدول(۵-۶) نتایج ضریب همبستگی دو متغیر

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	after & before	25	.752	.000

در خروجی سوم (جدول ۵-۷) نتایج آزمون t نشان داده می شود. با توجه به اینکه sig آزمون کوچک تر از ۵ درصد است، بنابراین H_0 رد می گردد. یعنی ادعای عدم تغییر انگیزه مهاجرت در قبل و بعد از دادن امکانات رفاهی به روستا رد می شود. با توجه به اینکه حد پایین و بالا منفی هستند بنابراین میانگین انگیزه مهاجرت بعد از دادن امکانات رفاهی به روستا کمتر شده است.

جدول (۵-۷) نتایج نهایی آزمون تی جفت نمونه‌ای

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	after - before	-6.120	6.966	1.393	-8.995	-3.245	-4.393	24	.000

۵-۴. آزمون مقایسه میانگین چند جامعه (تحلیل واریانس)

تحلیل واریانس(ANOVA) روش آماری بسیار مفیدی است که در اکثر پژوهش‌های علمی به طور گسترده بکار می رود. با استفاده از این روش می توان فرضیه هایی مربوط به وجود تفاوت بین میانگین های بیش از دو جامعه را مورد بررسی قرار داد. هدف از تحلیل واریانس مقایسه بین میانگین های چند جامعه و بررسی معنی داری تفاوت آنها و مشخص کردن عوامل مؤثر در این تفاوتها است. به این روش که میانگین های دو یا بیش از دو سطح یک یا چند متغیر مستقل را مقایسه می کند تحلیل واریانس گفته می شود. اساس آزمون تحلیل واریانس بر اساس تحلیل واریانس درون گروه ها و بین گروه ها است. برای مثال می توان جواب سؤالات زیر را با روش تحلیل واریانس پاسخ داد:

- آیا میزان برداشت محصولات به نوع کود بستگی دارد یا نه؟
- آیا نوع شغل در میزان درآمد تأثیر دارد یا نه؟
- آیا تأثیر گذاری صنایع مختلف در تحقق شاخص های توسعه یکسان است یا خیر؟

در سؤالات فوق می بینیم که متغیرهای مستقل باعث تقسیم جامعه به چند زیر گروه یا زیر جامعه می شوند و هدف آزمون این مسأله است که آیا میانگین متغیر وابسته در زیر گروه ها یکسان است یا خیر؟ به عبارتی فرض صفر در تحلیل واریانس، یکسان بودن میانگین تمام زیر گروه هاست و فرض مخالف هم این است که حداقل یکی از میانگین زیر گروه ها با بقیه متفاوت است.

فرض های آماری در این آزمون عبارتند از:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_K$

H_1: حداقل دوتا از μ ها برابر نیستند

این آماره آزمون شکل گسترش یافته آزمون T می باشد. در مقایسه های چند گانه، تحلیل واریانس یک روش فراگیرتر از آزمون تی می باشد زیرا آزمون تی به منظور تعیین معنی داری تفاوت بین دو میانگین بکار می رود ولی در شرایطی که دارای بیش از دو شرایط آزمایشی یا دو متغیر باشیم در این موقعیت ها به جای استفاده از چندین آزمون تی برای مقایسه تمام تفاوت های ممکن، از تحلیل واریانس استفاده می کنیم. همچنین تحلیل واریانس این اجازه را به ما می دهد تا اثرات بیش از یک متغیر مستقل را بررسی کنیم در حالی که آزمون تی شامل بررسی اثر فقط یک متغیر مستقل بر یک متغیر وابسته می باشد.

شرایط استفاده از تحلیل واریانس:

۱- متغیر وابسته دارای مقیاس فاصله ای- نسبی باشد.
۲- توزیع داده ها برای متغیر وابسته نرمال باشد.
۳- متغیر مستقل اسمی یا رتبه ای باشد.
۴- نمونه ها به صورت مستقل و تصادفی از جامعه انتخاب شوند.

تحلیل واریانس یک طرفه را F فیشر هم می گویند. در واقع تحلیل ANOVA با توزیع F منطبق است و این شاخص آماری را به کار می گیرد.

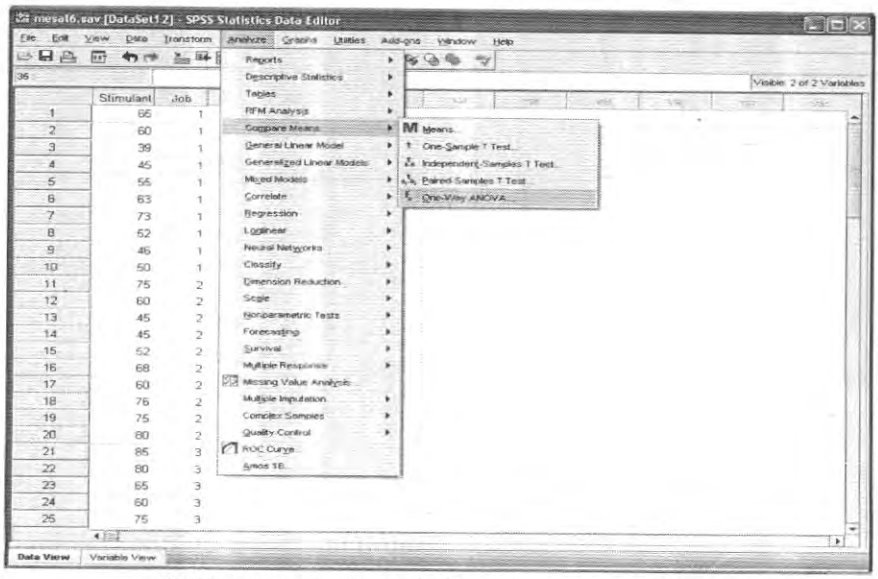
مثال: انگیزه داشتن شغل دوم در گروهای شغلی زیر را بر حسب لیکرت از نمونه‌ای به حجم ۱۰ نفر برای هر گروه شغلی در یک دهستان به صورت زیر بدست آورده ایم:

۱- کشاورزان: ۶۵- ۶۰- ۳۹- ۴۵- ۵۵- ۶۳- ۷۳- ۵۲- ۴۶- ۵۰
۲- دامداران: ۷۵- ۶۰- ۴۵- ۴۵- ۵۲- ۶۸- ۶۰- ۷۶- ۷۵- ۸۰
۳- باغداران: ۸۵- ۸۰- ۶۵- ۶۰- ۷۵- ۵۵- ۵۰- ۸۲- ۹۰- ۷۵

حال با توجه به این اعداد به دست آمده آیا انگیزه داشتن شغل دوم در این سه گروه متفاوت است یا خیر؟

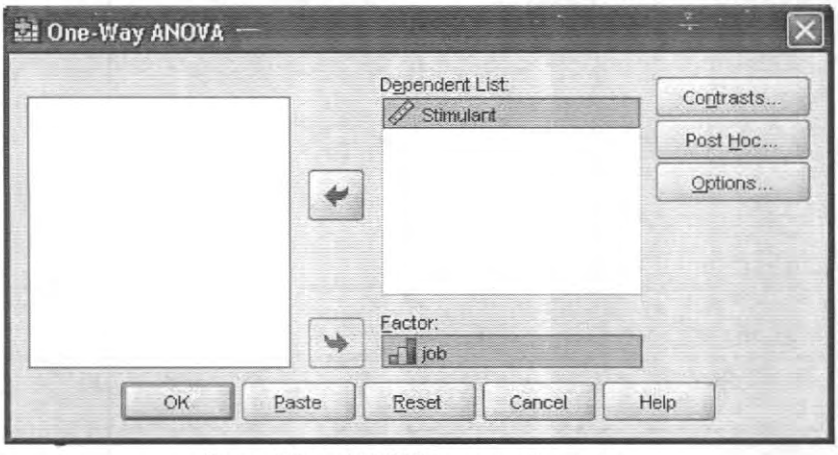
پس از وارد کردن داده ها مسیر شکل (۵-۸) را برای اجرای آزمون تحلیل واریانس استفاده می کنیم.

Analyze ⟶ Compare Means ⟶ One-way ANOVA



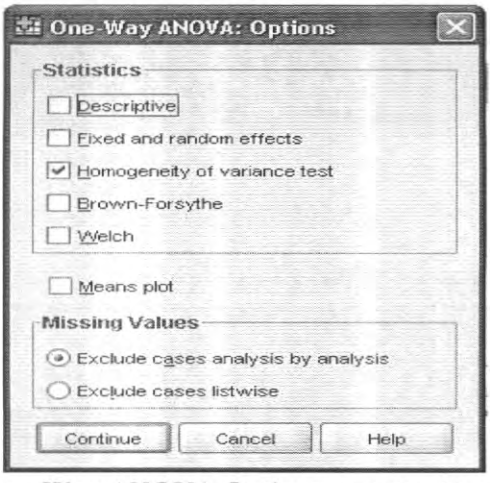
شکل (۵-۸) پنجره SPSS Statistics Data Editor 4

پس از انتخاب این منو پنجره One –Way ANOVA باز می گردد. همانند شکل (۵-۹) متغیرهای Stimulant را به کادر Dependent List و متغیر Job را به کادر Factor منتقل کنید. سپس بر روی دکمه Options کلیک کنید.



شکل (۵-۹) پنجره One –Way ANOVA

بعد از کلیک روی دکمه Options کادری همانند شکل (۵-۱۰) باز می گردد. جهت بررسی برابری واریانس ها همانند شکل گزینه Homogeneity of variance test را فعال کنید. بر روی دکمه Continue و سپس گزینه ok کلیک کنید تا نتایج برای شما ظاهر گردند.



شکل (۵-۱۰) پنجره One –Way ANOVA Options

نتیجه شامل دو خروجی است. در خروجی اول (جدول ۵-۸) نتایج آزمون لون برای برابری واریانس ها نشان داده شده است. نتیجه این خروجی در ادامه جهت انتخاب آزمون های پس از تجربه بسیار کاربرد دارد. با توجه به اینکه sig این خروجی بیشتر از ۵ درصد است، لذا فرض برابری واریانس ها رد نمی شود.

جدول (۵-۸) نتایج آزمون لون برای برابری واریانس ها

Test of Homogeneity of Variances

Stimulant

Levene Statistic	df1	df2	Sig.
.807	2	27	.457

خروجی دوم (جدول ۵-۹) نتایج حاصل از آزمون F را نشان می دهد. با توجه به اینکه sig آزمون کوچکتر از ۵ درصد است، می توان H_0 را رد کرد به این معنی که حداقل یک گروه شغلی با بقیه گروهها از لحاظ انگیزه داشتن شغل دوم متفاوت می باشد.

جدول (۵-۹) نتایج حاصل از آزمون F

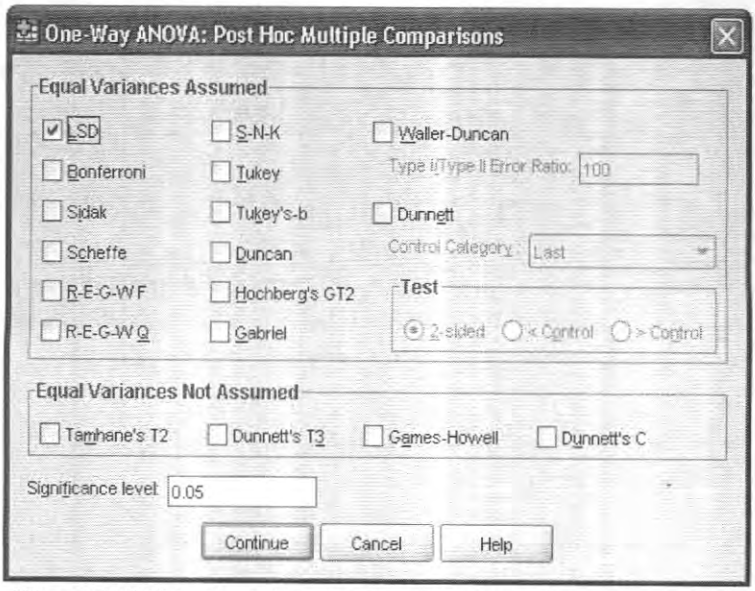
ANOVA

Stimulant

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1428.867	2	714.433	4.617	.019
Within Groups	4178.100	27	154.744		
Total	5606.967	29			

حال که دانستیم تفاوت وجود دارد می خواهیم بدانیم که این تفاوت بین کدام گروهها است. در اینجا نتایج آزمون لون که در جدول ۵-۸ انجام گرفت کاربرد دارد. با توجه به اینکه واریانس جوامع برابرند لذا باید از آزمون های پس از تجربه برابری واریانس ها (Equal Variances Assumed) استفاده کرد. بنابراین دوباره مسیر فوق را برای

تحلیل واریانس انتخاب کنید. در پنجره One –Way ANOVA بر روی دکمه Post Hoc کلیک کنید تا پنجره ای تحت عنوان One –Way ANOVA: Post Hoc Multiple Comparisons باز گردد. در این پنجره دو کادر وجود دارد. قسمت بالا مربوط به آزمونهای مورد استفاده در حالتی است که واریانس جوامع برابر باشند(Equal Variances Assumed) و قسمت پایین مربوط به آزمونهای مورد استفاده در حالتی است که واریانس جوامع متفاوت باشند(Equal Variances Not Assumed). با توجه به برابری واریانس ها همانند شکل(۵-۱۱) گزینه LSD را فعال کنید. بر روی دکمه Continue و سپس گزینه ok کلیک کنید تا نتایج برای شما ظاهر گردند.



شکل (۵-۱۱) پنجره One –Way ANOVA: Post Hoc Multiple Comparisons

نتیجه این بار هم شامل دو خروجی است. در خروجی اول همان نتایج آزمون F که در جدول (۵-۹) نشان داده شده است، ارائه می گردد که تشریح شد. اما در خروجی دوم (جدول ۵-۱۰) نتایج مربوط به آزمون LSD می باشد که از سه طریق می توان فهمید که کدام گروهها با هم متفاوت اند. روش اول از طریق ستاره ای که بر روی اختلاف میانگین هر گروه با گروه دیگر زده شده است. اگر ستاره روی آن باشد نشان از تفاوت آن

دو گروه می باشد. روش دوم از طریق sig می باشد که اگر کمتر از ۵ درصد باشد نشان از تفاوت بین گروه‌ها دارد. روش سوم از طریق Lower و Upper می باشد اگر صفر بین آنها قرار گیرد عدم تفاوت و اگر قرار نگیرد نشان از تفاوت است.

حال با توجه به توضیحات داده شده مشخص می شود که گروه ۱ (کشاورزان) با گروه ۳ (باغداران) از لحاظ انگیزه داشتن شغل دوم تفاوت وجود دارد.

جدول (۵-۱۰) نتایج مربوط به آزمون LSD

Multiple Comparisons

Stimulant
LSD

(I) job	(J) job	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-8.800	5.563	.125	-20.21	2.61
	3	-16.900*	5.563	.005	-28.31	-5.49
2	1	8.800	5.563	.125	-2.61	20.21
	3	-8.100	5.563	.157	-19.51	3.31
3	1	16.900*	5.563	.005	5.49	28.31
	2	8.100	5.563	.157	-3.31	19.51

فصل ششم

آزمون‌های ناپارامتری

آزمونهای ناپارامتریک برای سنجش تفاوت بین نمونه هایی که دارای جامعه مادر بدون توزیع مشخص یا اصطلاحاً آزاد توزیع می باشند به کار می روند. این آزمون ها به فرض های مقیاس های فاصله ای- نسبی، نرمال بودن و یا واریانس های برابر نیازی ندارد.

در شرایط زیر استفاده از آزمون های ناپارامتریک مؤثر خواهد بود:

۱- هنگامی که مقیاس داده ها ترتیبی باشد.

۲- هنگامی که داده ها فاصله ای یا نسبی اما دارای توزیع غیر نرمال هستند.

همچنین در مواردی نظیر نامساوی بودن حجم نمونه ها و واریانس ها که امکان برقراری فرض های لازم جهت آزمون های پارامتریک وجود ندارد استفاده از آزمونهای ناپارامتریک مشابه می تواند مؤثر باشد.

۶-۱. آزمون علامت[1] زوج نمونه ای

آزمون علامت(نشانه) یکی از ساده ترین آزمون های ناپارامتری است. این آزمون زمانی که داده ها به صورت همبسته باشند مورد استفاده قرار می گیرد. در این آزمون دو زوج مرتب (x_i , y_i) نشان دهنده صفت متغیر برای هر فرد نمونه قبل و بعد از انجام عمل مورد نظر می باشد. در این آزمون اگر x_i از y_i بزرگ تر باشد از علامت + و اگر x_i از y_i کوچک تر باشد از علامت - و اگر x_i و y_i مساوی باشند از واژه Ties استفاده می گردد.

مثال: محققی در راستای این مسأله که آیا بعد از دادن امکانات رفاهی انگیزه مهاجرت در روستا نسبت به قبل تغییر یافته است یا خیر، با استفاده از طیف لیکرت پنج سطحی(۱ تا ۵)

1 . Sign Test

نمونه ای ۲۵ نفری را برای بررسی این مسأله انتخاب کرده است.

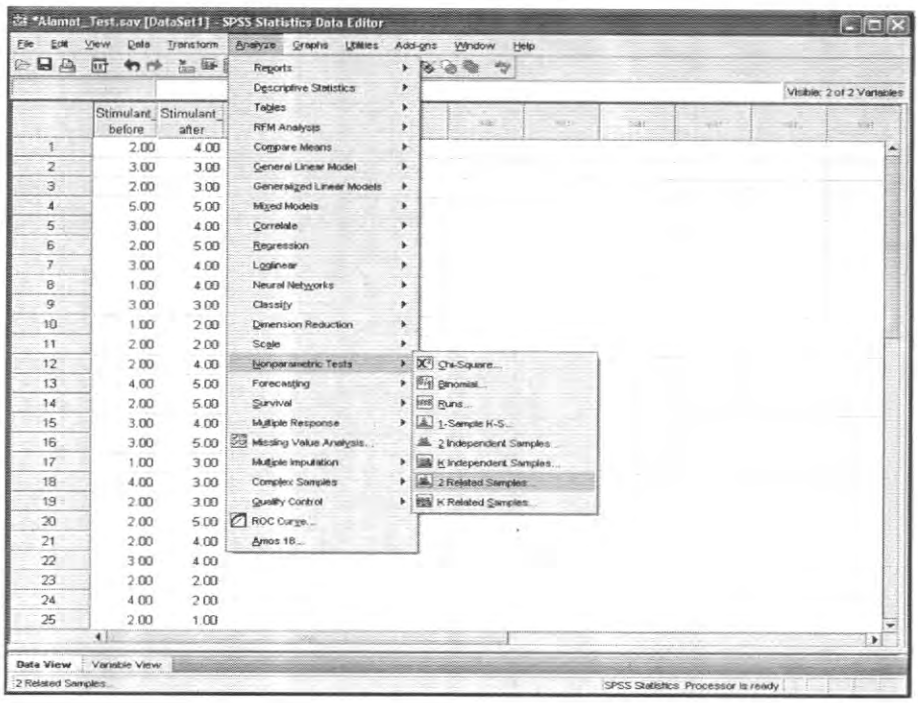
فرض های آماری در این آزمون عبارتند از:

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

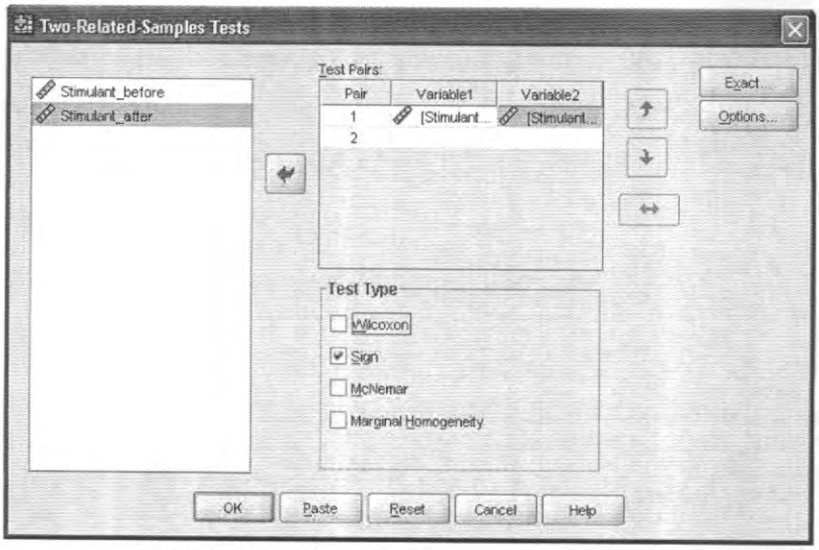
پس از وارد کردن داده ها مسیر شکل (۶-۱) را برای اجرای آزمون علامت زوجی استفاده می کنیم.

Analyze ⟶ Nonparametric Test ⟶ 2 Related Samples



شکل (۶-۱) پنجره SPSS Statistics Data Editor 1

پس از انتخاب این منو پنجره Two-Related-Samples Tests باز می گردد. همانند شکل(۶-۲) متغیرهای Stimulant_before و Stimulant_after را به کادر Test Pairs منتقل کنید. سپس در کادر Test Type گزینه Sign را انتخاب کنید. جهت اجرای آزمون ok را کلیک کنید.



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

نتیجه آزمون شامل دو خروجی است. خروجی اول (جدول ۶-۱) به ترتیب تعداد حالاتی که انگیزه بعد از دادن امکانات کاهش یافته است، تعداد حالاتی که انگیزه بعد از دادن امکانات افزایش یافته است، تعداد حالاتی که انگیزه قبل و بعد از دادن امکانات تفاوتی نکرده است و مجموع کل مشاهدات را نشان می دهد.

جدول (۶-۱) نتایج توصیفی آزمون علامت

Frequencies

		N
Stimulant_after - Stimulant_before	Negative Differences[a]	3
	Positive Differences[b]	17
	Ties[c]	5
	Total	25

a. Stimulant_after < Stimulant_before

b. Stimulant_after > Stimulant_before

c. Stimulant_after = Stimulant_before

خروجی دوم (جدول ۶-۲) مقدار آماره z و sig آزمون را ارایه می کند. با توجه به اینکه sig کمتر از ۵ درصد است، H_0 رد می شود. بنابراین تفاوت معنی داری بین انگیزه مهاجرت قبل و بعد از دادن امکانات به روستا وجود دارد.

جدول (۶-۲) نتایج نهایی آزمون علامت

Test Statistics[b]

	Stimulant_after – Stimulant_before
kExact Sig. (2-tailed)	.003[a]

a. Binomial distribution used.

b. Sign Test

۶-۲. آزمون مک نمار:

این آزمون در مواردی بکار برده می شود که دو مجموعه از اندازه های وابسته در یک متغیر دو ارزشی اندازه گیری شده باشند. متغیر مورد مطالعه می تواند بر حسب ماهیت خود در مقیاس های طبقه ای، ترتیبی و یا فاصله ای قابل اندازه گیری باشد. وضعیت ویژه کاربرد این آزمون در طرح ماقبل و مابعد است این آزمون محدود به یک جدول ۲×۲ است که هر خانه آن معرف عددی است که آنها را با حرف a, b, c, d نشان می دهند.

a	b
c	d

آزمون مک نمار همواره بر مقادیر (b وc) توجه داشته که در معرض تغییر قرار گرفته اند. بنابراین فرضیه H_0 در آزمون مک نمار این است که نسبت افراد واقع در خانه b با نسبت افراد واقع در خانه c برابر است یعنی هر یک از این نسبت ها برابر با ۰/۵ است.

$$\begin{cases} H_0: P = 0/5 \\ H_1: P \neq 0/5 \end{cases}$$

آزمون مک نمار چون تغییر از خانه های b و c را مستقیماً ارزیابی می کند صریحاً آزمون یک طرفه است. چنانچه مقادیر یا تعداد افراد در معرض تغییر کمتر یا مساوی ۲۵ باشد معنا دار بودن آزمون را می توان با مراجعه به جداول توزیع های دو جمله ای ارزشیابی کرد. اما اگر تعداد آزمودنی ها از عده گروه های نمونه در جداول دو جمله ای تجاوز کند برای آزمون تفاوت دو نسبت وابسته می توانیم از تقریب نرمال دو جمله ای مطابق رابطه (۶-۱) استفاده کنیم.

$$z_0 b = \frac{|b-c|-1}{\sqrt{b+c}} \qquad (۶-۱)$$

چنانچه n نسبتاً بزرگ باشد آماره آزمون با توزیع کای دو منطبق شده و مطابق رابطه (۶-۲) محاسبه می شود.

$$x^2 = \frac{(|b-c|-1)^2}{b+c} \qquad (۶-۲)$$

این مشخصهٔ آماری با یک درجه آزادی توزیع شده و زمانی به کار می رود که فراوانی مورد انتظار در هیچ یک از خانه ها کمتر از ۵ نباشد.

مثال: می خواهیم تأثیر آموزش (قبل و بعد از آموزش) را در افزایش محصول در یک دهستان بسنجیم برای این کار نمونه ای به حجم ۱۰۰ نفر را انتخاب کرده ایم که نتایج زیر بدست آمده است. حال با توجه به نتایج بدست آمده (جدول ۶-۳) آیا آموزش موجب افزایش محصول شده است یا نه؟

جدول (۶-۳) توزیع فراوانی نظرات پاسخگویان

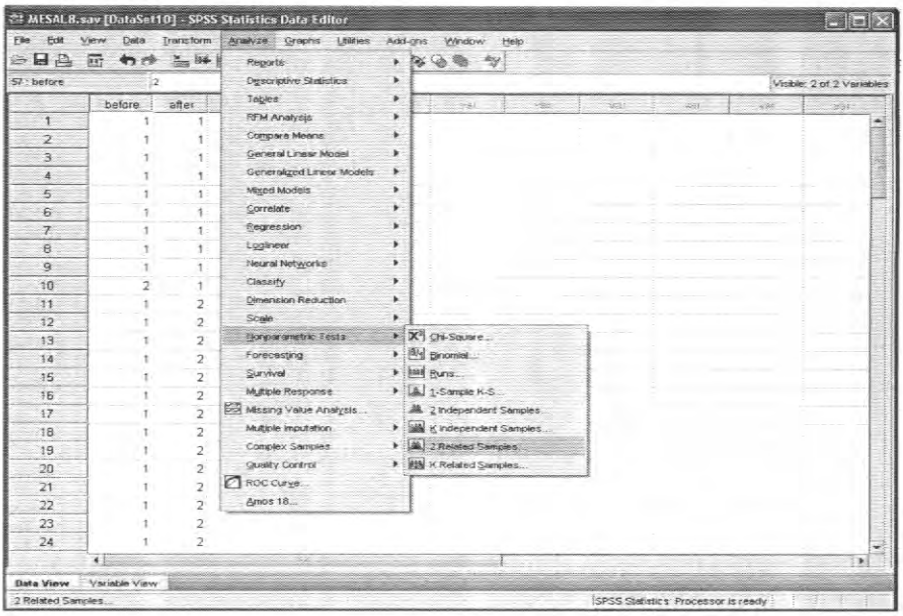
	بعد از دادن آموزش	
قبل از دادن آموزش	منفی(۱)	مثبت(۲)
منفی(۱)	۱۰	۴۵
مثبت(۲)	۱۵	۳۰

- ۱۰ نفر قبل و بعد از آموزش دیدگاهی منفی در مورد آموزش داشته اند.
- ۴۵ نفر قبل از آموزش دیدگاهی منفی داشته اند ولی بعد از آموزش دیدگاه آنها مثبت شده است.
- ۳۰ نفر نفر قبل و بعد از آموزش دیدگاهی مثبت در مورد آموزش داشته اند.

- ۱۵ نفر قبل از دادن آموزش دیدگاه شان مثبت بوده ولی بعد از آموزش دیدگاه شان منفی شده است.

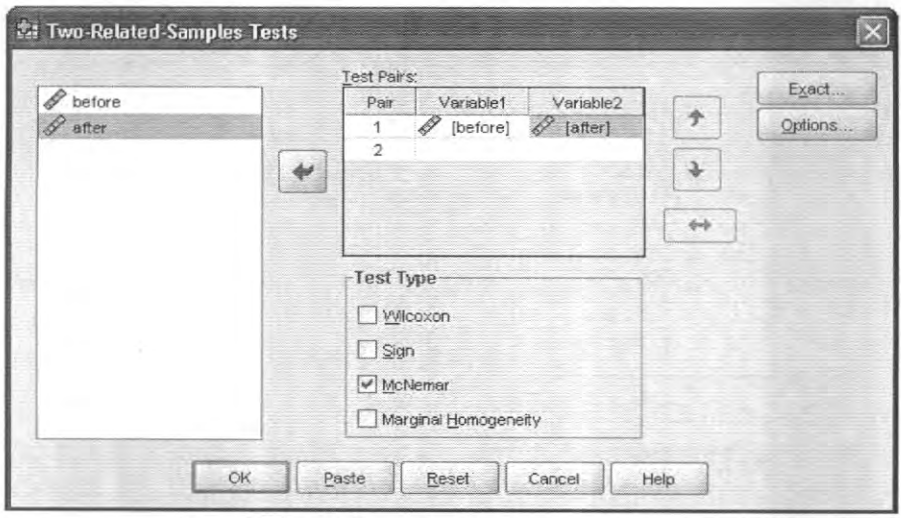
برای پاسخ به این سئوال از آزمون مک نمار استفاده می کنیم. بنابراین مسیر شکل (۶-۳) را برای اجرای این آزمون بکار می گیریم.

Analyze ⟶ Nonparametric Test ⟶ 2 Related Samples



شکل (۶-۳) پنجره SPSS Statistics Data Editor 2

پس از انتخاب این منو پنجره Two-Related-Samples Tests باز می گردد. همانند شکل(۶-۴) متغیرهای before و after را به کادر Test Pairs منتقل کنید. سپس در کادر Test Type گزینه McNemar را انتخاب کنید. جهت اجرای آزمون ok را کلیک کنید.



شکل (۶-۴) پنجره Two-Related-Samples Tests

نتیجه آزمون شامل دو خروجی می باشد. خروجی اول (جدول ۶-۴) فراوانی ترکیبی دو متغیر را نشان می دهد.

جدول (۶-۴) نتایج توصیفی آزمون مک نمار

before & after

before	after	
	1	2
1	10	45
2	15	30

خروجی دوم (جدول ۶-۵) به ترتیب تعداد موارد، مقدار آماره X^2 و sig آزمون را ارایه می کند. با توجه به مقدار sig که کمتر از ۵ درصد است می توان H_0 را رد کرد. بنابراین تفاوت معنی داری بین افزایش محصول در دهستان قبل و بعد از آموزش وجود دارد. به عبارت دیگر آموزش باعث افزایش محصول گردیده است.

جدول (۶-۵) نتایج نهایی آزمون مک نمار

Test Statistics[b]

	before & after
N	100
Chi-Square[a]	14.017
Asymp. Sig.	.000

a. Continuity Corrected

b. McNemar Test

۶-۳. آزمون دو جمله ای

این آزمون (که گاهی آزمون نسبت نامیده می شود) زمانی به کار می رود که بخواهیم نسبت خاصی را در جامعه بررسی کنیم یا زمانی که محقق بر آن است تا با طرح متغیرهای طبقه ای که تنها دارای دو ارزش می باشد (مثل جنسیت که یا مرد است یا زن) به دنبال آزمون فرضیه خود باشد. آزمون دوجمله ای معادل ناپارامتری آزمون T تک نمونه ای است. این آزمون عموماً برای بررسی فرضیه هایی با متغیرهای کیفی مورد استفاده قرار می گیرد. بعبارت دقیق تر اگر متغیرهای مورد بررسی از توزیع نرمال یا واریانس های مساوی برخوردار نباشند از آزمون دو جمله ای استفاده می گردد.

نکته: گاهی از این آزمون همانند آزمون میانگین یک جامعه (T تک نمونه ایی)، برای تشخیص تأثیر یا عدم تأثیر یک متغیر در پدیده ای معین استفاده می شود. در طراحی فرض های این آزمون H_0 بیانگر عدم تأثیر متغیر و H_1 بیانگر تأثیر متغیر می باشد. فرمول این آزمون از رابطه (۶-۳) به دست می آید.

$$Z = \frac{r_0 b - np}{\sqrt{npq}} \qquad (۶-۳)$$

در این فرمول $r_0 b$ تعداد واقعی موفقیت های مشاهده شده در گروه نمونه، n تعداد نمونه‌ها، p نسبت موفقیت در جامعه مورد نظر و $\sqrt{npq}$ نیز انحراف استاندارد توزیع در شرایط فرضیه صفر می باشد.

مثال:

بر اساس تجربیات قبلی فرض کرده ایم که ۶۵ درصد از افراد یک روستا باسواد هستند. یک نمونه ۷۵ نفری از جامعه انتخاب کرده‌ایم که ۴۱ نفر آنها باسواد بوده اند. حال سئوال این است که این ۴۱ نفر می توانند این فرض را که ۶۵ درصد افراد روستا باسواد هستند را تأیید یا رد نماید؟ برای این کار از آزمون دو جمله ای استفاده می کنیم.

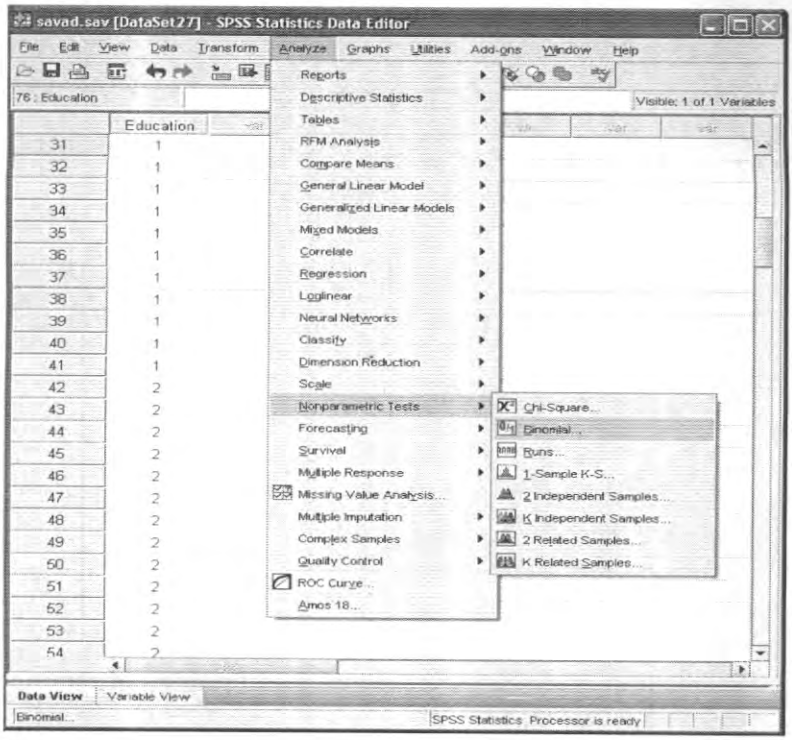
فرض های آماری در این آزمون عبارتند از:

$H_0 : p \leq 0/65$

$H_1 : p > 0/65$

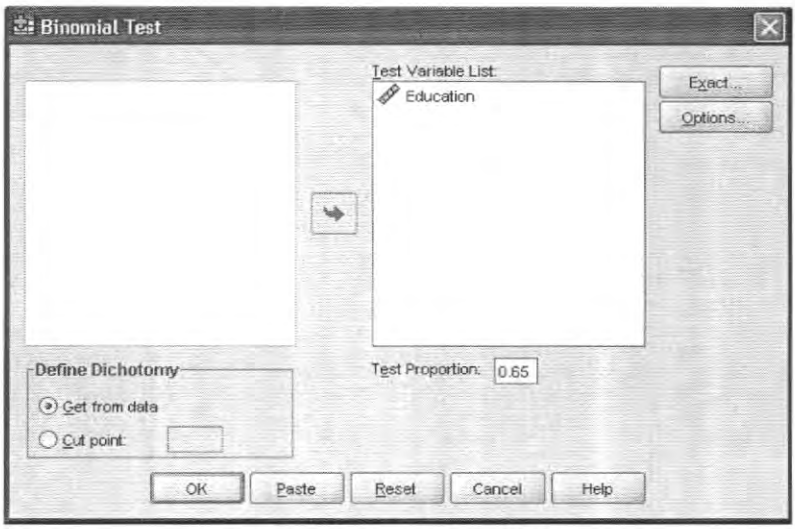
پس از وارد کردن داده ها مسیر شکل (۶-۵) را برای اجرای آزمون Binomial استفاده می کنیم.

Analyze ⟶ Nonparametric Test ⟶ Binomial



شکل (۶-۵) پنجره SPSS Statistics Data Editor 3

پس از انتخاب این منو پنجره Binomial Test باز می گردد. همانند شکل(۶-۶) متغیر Education را به کادرTest Variable List منتقل کنید. سپس در کادر روبروی Test Proporation پیش فرض نرم افزار را از ۰/۵۰ به ۰/۶۵ تغییر دهید. جهت اجرای آزمون ok را کلیک کنید.



شکل (۶-۶) پنجره Binomial Test

نتیجه شامل یک خروجی می باشد. در این خروجی (جدول ۶-۶) گروه ها، نحوه تعریف آنها، تعداد مشاهدات هر گروه، نسبت مشاهده شده، نسبت آزمون و sig را نشان می‌دهد. از آنجایی که sig کمتر از ۵ درصد می باشد، H_0 رد می شود و می توانیم بپذیریم که ۶۵ درصد افراد روستا باسواد هستند.

جدول (۶-۶) نتایج آزمون دوجمله ای

Binomial Test

	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Education Group 1	educated	41	.55	.65	.041[a,b]
Group 2	uneducated	34	.45		
Total		75	1.00		

a. Alternative hypothesis states that the proportion of cases in the first group < .65.

b. Based on Z Approximation.

۶-۴. آزمون من- ویتنی U

این آزمون معادل ناپارامتریک آزمون تی مستقل است و برای دو جامعه مستقل با مقیاس ترتیبی استفاده می شود. گاهی اوقات پژوهشگر مایل است آزمون فرضیه عدم تفاوت بین گروهها را براساس رتبه های واقعی مشاهده شده برای دو توزیع جامعه که دارای توزیع پیوسته هستند بکار گیرد که در این مورد آزمون یومن ویتنی از کاربرد مناسبی برخوردار است. اساس مقایسه آزمون من- ویتنی بر پایه مقایسه میانه است. در این آزمون نیازی به نرمال بودن توزیع و برابری واریانسها نمی باشد. شیوه عمل آزمون من- ویتنی به این طریق است که ابتدا کلیه مقادیر موجود را به صورت یکجا و بدون توجه به جامعه ای که مقادیر به آن تعلق دارد رتبه بندی می کند یعنی نمونه ها را به هم ریخته و رتبه بندی می کند، زمانی که در دو جامعه هم رتبه وجود داشته باشد رتبه ها را جمع کرده و بر تعداد افراد هم رتبه تقسیم می کند. فرضیه صفر مورد آزمایش در این آزمون همانند آزمون تی مستقل بدین شکل است:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

مثال: میزان رضایت دو گروه کشاورز و دامدار از میزان درآمد شغلی را در نمونه ای ۲۵ نفری بدست آورده ایم. حال با توجه به این داده ها آیا میزان رضایت از درآمد در دو گروه متفاوت است یا خیر؟ نمره های خام بر اساس پرسشنامه بدست آمده اند(طیف از ۱ تا ۱۰۰) که ما آنها را رتبه بندی کرده ایم(جدول ۶-۷).

جدول (۶-۷) توزیع فراوانی نمره پاسخگویان

کشاورز(۲)		دامدار(۱)	
نمره خام	رتبه	نمره خام	رتبه
۶۰	۱	۶۰	۱
۷۰	۲	۷۵	۳
۷۵	۳	۷۶	۴
۷۸	۶	۷۷	۵
۷۹	۸	۷۸	۷
۸۰	۱۰	۸۱	۱۱
۸۲	۱۲	۸۳	۱۳
۸۵	۱۴	۸۶	۱۵
۸۷	۱۶	۹۱	۱۹
۸۹	۱۷	۹۴	۲۲
۹۰	۱۸	۹۶	۲۳
۹۲	۲۰	۹۸	۲۴
۹۳	۲۱		

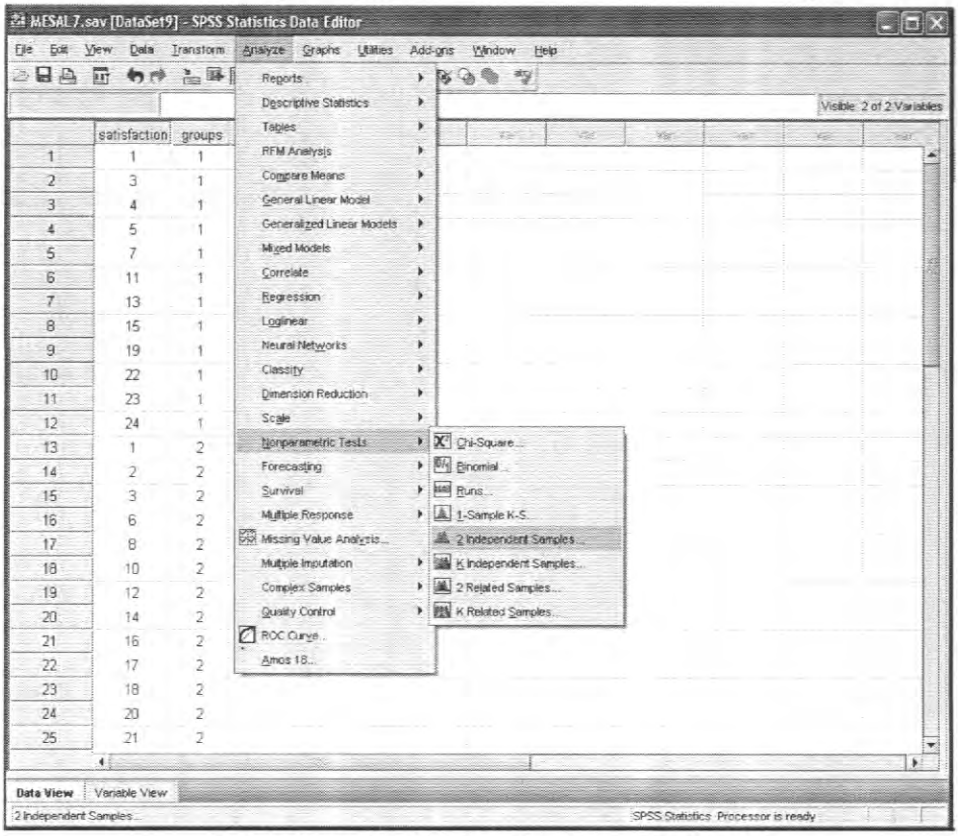
فرض های آماری در این آزمون عبارتند از:

H_0: دو گروه در میزان رضایت از درآمد شغلی با هم تفاوت ندارند

H_1: دو گروه در میزان رضایت از درآمد شغلی با هم تفاوت دارند

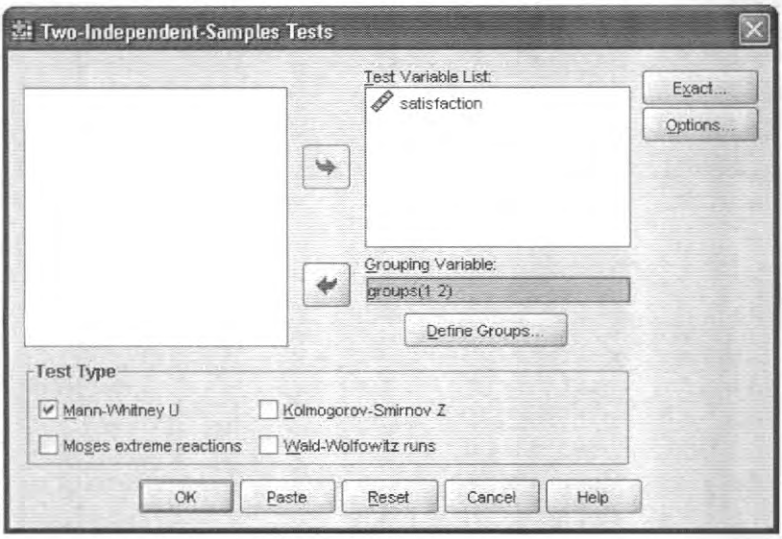
پس از وارد کردن داده ها مسیر شکل (۶-۷) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Nonparametric Test ⟶ 2 Independent Samples



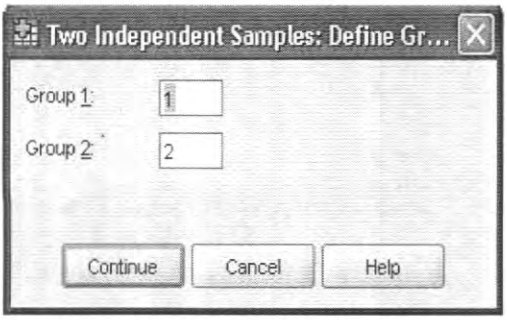
شکل (۶-۷) پنجره SPSS Statistics Data Editor 4

پس از انتخاب این منو پنجره Two-Independent-Samples Tests باز می گردد. همانند شکل(۶-۸) متغیر satisfaction را به کادر Test Variable List و متغیر groups را به کادر Grouping Variable منتقل کنید.



شکل (۶-۸) پنجره Two Independent Samples Tests

سپس بر روی دکمه Define Groups کلیک کنید. همانند شکل(۶-۹) در کادر Group 1 عدد ۱ و در کادر Group 2 عدد ۲ را وارد کرده و Continue را کلیک کنید. نهایتاً جهت اجرای آزمون ok را کلیک کنید.



شکل (۶-۹) پنجره Two Independent Samples Define Group

نتیجه شامل دو خروجی است. خروجی اول (جدول ۶-۸) آمار توصیفی مربوط به تعداد، میانگین رتبه ای و مجموع رتبه ها را برای هر دو گروه نشان می دهد.

جدول (۶-۸) نتایج توصیفی آزمون من ویتنی

Ranks

	groups	N	Mean Rank	Sum of Ranks
satisfaction	1	12	13.50	162.00
	2	13	12.54	163.00
	Total	25		

خروجی دوم (جدول ۶-۹) به ترتیب آماره U من- ویتنی، آماره W ویلکاکسون، آماره Z و مقدار sig را ارایه می کند. با توجه به اینکه sig بزرگ تر از ۵ درصد است، H_0 رد نمی شود. به این معنا که این دو گروه در میزان رضایت از درآمد شغلی با هم تفاوت ندارند و به یک اندازه از درآمد شغلی رضایت دارند.

جدول(۶-۹) نتایج نهایی آزمون من ویتنی

Test Statistics[b]

	satisfaction
Mann-Whitney U	72.000
Wilcoxon W	163.000
Z	-.326
Asymp. Sig. (2-tailed)	.744
Exact Sig. [2*(1-tailed Sig.)]	.769[a]

a. Not corrected for ties.

b. Grouping Variable: groups

۶-۵. آزمون ویلکاکسون

آزمون ویلکاکسون معادل ناپارامتریک آزمون t وابسته است. این آزمون مناسب طرح های ماقبل و مابعد با مقیاس رتبه ای می باشد که اندازه تفاوت میان رتبه ها را مورد نظر قرار می دهد. بر مبنای این آزمون قادر خواهیم بود نه تنها جهت نمونه ها را بلکه اندازه تفاوت بین دو گروه را نیز اندازه گیری نمائیم و مشخص کنیم که کدام گروه شرایط مناسب تر یا بهتری را نسبت به گروه هم جفت خود دارد. بر مبنای آزمون ویلکاکسون در شرایط فرض صفر (H_0) یعنی یکسان بودن توزیع دو جامعه، انتظار می رود که حدود نیمی از

تفاوت‌های زوجی مثبت و نیمی دیگر منفی باشد یعنی اندازه تفاوت های منفـی و مثبـت تا اندازه‌ای یکسان باشد. در این آزمون سه حالت اتفاق می افتد:

۱- یا اینکه تأثیر دوره بعد نسبت به دوره قبل بیشتر است.

۲- یا اینکه تأثیر دوره قبل نسبت به دوره بعد بیشتر است.

۳- یا اینکه تأثیر دوره قبل و بعد تفاوتی پیدا نکرده است.

مثال: میزان مشارکت در طرحهای عمرانی را در یک روستا برای نمونه ای به حجم ۲۰ نفر در قبل و بعد از دادن آموزش بصورت زیر بدست آورده ایم. حال با توجه به اعداد بدست آمده می توان گفت که آیا آموزش توانسته است باعث افزایش مشارکت در طرحهای عمرانی شده باشد یا خیر؟

کد ۱ مشارکت خیلی کم، کد ۲ مشارکت کم، کد۳ مشارکت متوسط، کد۴ مشارکت زیاد، کد۵ مشارکت خیلی زیاد.

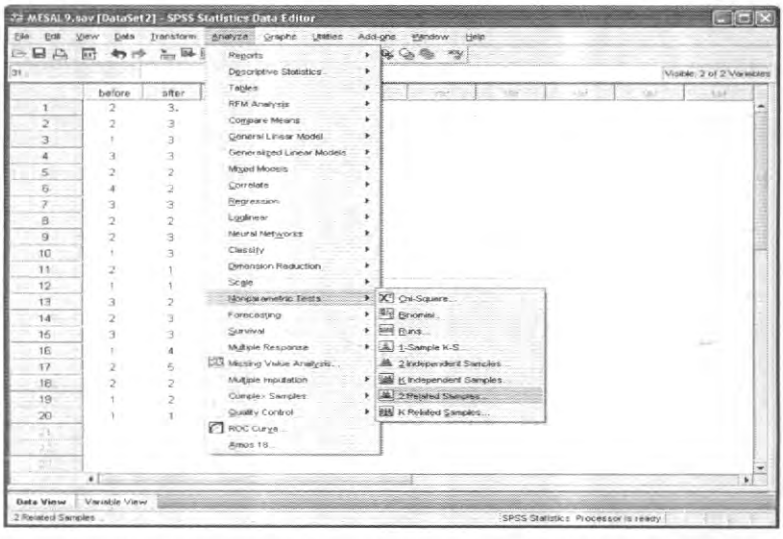
فرض های آماری در این آزمون عبارتند از:

H_0: میزان مشارکت قبل و بعد از دادن آموزش تفاوت نکرده است

H_1: میزان مشارکت در قبل و بعد از دادن آموزش متفاوت است

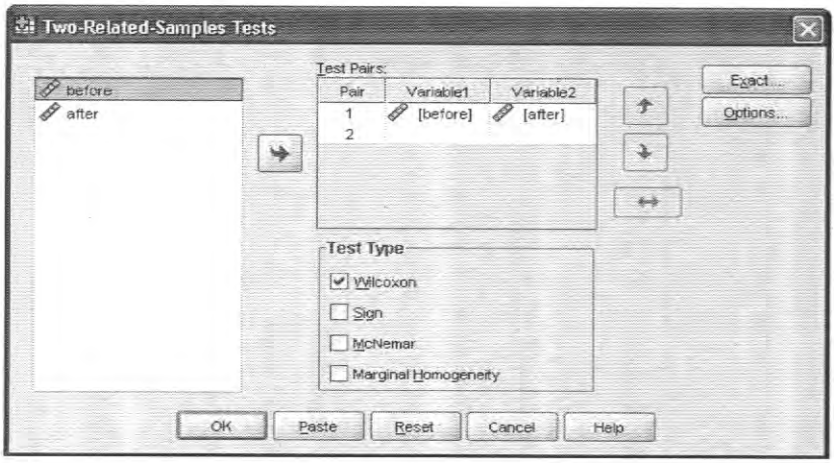
پس از وارد کردن داده ها مسیر شکل (۶-۱۰) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Nonparametric Test ⟶ 2 Related Samples



شکل (۶-۱۰) پنجره SPSS Statistics Data Editor 5

پس از انتخاب این منو پنجره Two-Related-Samples Tests باز می گردد. همانند شکل (۶-۱۱) متغیرهای after و before را به کادر Test Pairs منتقل کنید. سپس برای اجرای آزمون ok را کلیک کنید.



شکل (۶-۱۱) پنجره Two Related Samples Tests

نتیجه شامل دو خروجی است. خروجی اول (جدول ۶-۱۰) آمار توصیفی مربوط به تعداد مشاهدات، میانگین رتبه ای و مجموع رتبه ها را برای هر دو گروه نشان می دهد. در این خروجی سه شرایط نشان داده شده است:

a = تعداد افرادی که معتقدند بعد از آموزش مشارکت کمتر شده است. این داده ها تحت عنوان رتبه های منفی فهرست می شوند.

b = تعداد افرادی که معتقدند بعد از آموزش مشارکت بیشتر شده است. این داده ها تحت عنوان رتبه های مثبت فهرست می شوند.

c = تعداد افرادی که معتقدند قبل و بعد از آموزش مشارکت تغییری نکرده است. این داده ها تحت عنوان رتبه های برابر (Ties) فهرست می شوند.

جدول (۶-۱۰) آمار توصیفی آزمون ویلکاکسون

Ranks

		N	Mean Rank	Sum of Ranks
after - before	Negative Ranks	3[a]	5.67	17.00
	Positive Ranks	9[b]	6.78	61.00
	Ties	8[c]		
	Total	20		

a. after < before

b. after > before

c. after = before

خروجی دوم (جدول ۶-۱۱) مقدار آماره Z و Sig را نشان می دهد. با توجه به اینکه مقدار Sig از ۵ درصد بزرگ تر است، H_0 رد نمی شود. یعنی اگرچه تعداد افرادی که معتقدند بعد از آموزش مشارکت بیشتر شده است حدود ۳ برابر تعداد افرادی است که معتقدند بعد از آموزش مشارکت کمتر شده است اما تفاوت آن در کل اینقدر نبوده است که آزمون معنی دار گردد. بنابراین میزان مشارکت در قبل و بعد از دادن آموزش تفاوتی پیدا نکرده است.

جدول (۶-۱۱) نتایج نهایی آزمون ویلکاکسون

Test Statistics[b]

	after - before
Z	-1.768[a]
Asymp. Sig. (2-tailed)	.077

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

۶-۶. آزمون کروسکال- والیس(آزمون H)

آزمون کروسکال- والیس برای بیش از دو گروه مستقل با مقیاس رتبه ای بکار می رود. اگر جمعیت های مورد بررسی از توزیع نرمال یا واریانس های مساوی برخوردار نباشند از آزمون های ناپارامتریک که مشابه آزمون های پارامتریک است استفاده می شود. در آزمون کروسکال والیس متفاوت بودن یا نبودن سه یا بیش از سه گروه مستقل با توجه به رتبه های به دست آمده مورد ارزیابی قرار می گیرند. این آزمون جهت را نشان نمی دهد بلکه تأکید آن بر روی مقایسه میانگین ها و در خصوص تفاوت بین گروهها است. از مهمترین ویژگی های این آزمون تحلیل توزیع رتبه های متغیر وابسته در میان گروه متغیر مستقل بوده و نیز اینکه لازم نیست که حجم نمونه ها برابر باشد.

با توجه به این تفاسیر فرض های آماری در این آزمون عبارتند از:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_K$

H_1: حداقل دوتا از μ ها برابر نیستند

شرایط استفاده از این آزمون:

۱- یک متغیر مستقل با سه یا بیش از سه سطح

۲- حداقل رتبه ای بودن مقیاس داده ها

۳- زیاد نبودن تعداد هم رتبه ها

در مواردی که هیچ مشاهده تکراری در داده ها وجود نداشته باشد مشخصه آماری مشاهده شده در تحلیل واریانس که با H نمایش داده می شود از طریق رابطه (۶-۴) بدست می آید.

$$H = \frac{12}{\mathrm{N(N+1)}}\left(\sum_{i=1}^{K}\frac{R^2 j}{nj}\right) - 3(N+1) \qquad (۶-۴)$$

در این فرمول N معادل کل مشاهدات، nj تعداد مشاهدات در مورد گروههای مشخصی که با مقوله های مستقل تعریف می شوند و $R^2 j$ مربع جمع رتبه های هر گروه است.

در مواردی که هم رتبه در داده ها وجود داشته باشد ضمن استفاده از میانگین رتبه ای برای مشاهده های هم رتبه از رابطه (۶-۵) بهره گرفته می شود.

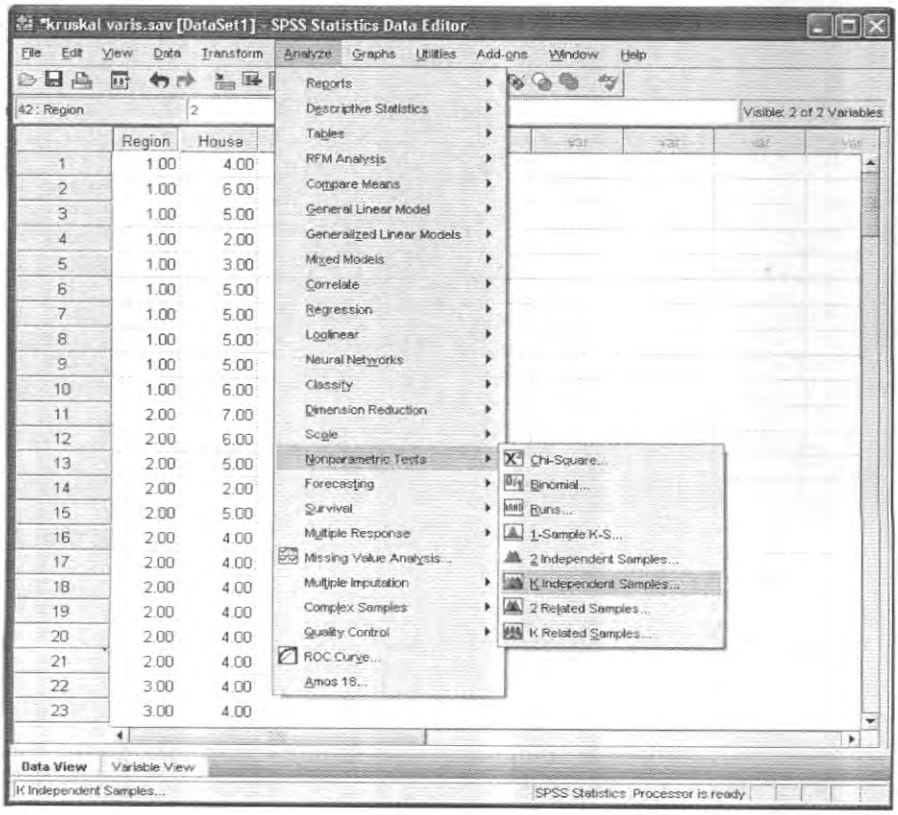
$$\mathrm{t} = 1 - \frac{\sum \mathrm{T_K}}{(\mathrm{N^3 - N})} \qquad (۶-۵)$$

در این رابطه $\mathrm{T_K} = \mathrm{t^3\text{-}k}$ است. میزان t نیز معرف تعداد مشاهده های تکراری برای هر رتبه و k معرف تعداد مجموعه های هم رتبه است.

مثال: کیفیت مسکن را در ۳ منطقه شهری تهران(منطقه ۱، منطقه ۲ و منطقه ۳) براساس طیف لیکرت ۵ سطحی سنجیده ایم حال آیا می توان گفت که کیفیت مسکن در این ۳ منطقه متفاوت است یا خیر؟

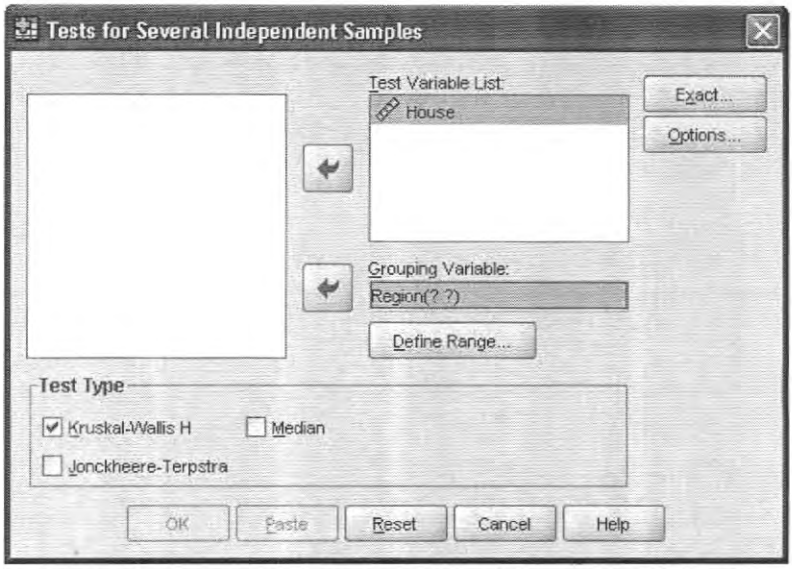
پس از وارد کردن داده ها مسیر شکل (۶-۱۲) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Nonparametric Test ⟶ K Independent Samples



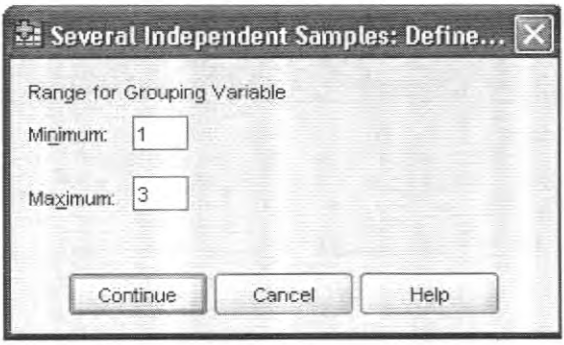
شکل (۶-۱۲) پنجره SPSS Statistics Data Editor 6

پس از انتخاب این منو پنجره Tests for Several Independent Samples باز می گردد. همانند شکل (۶-۱۳) متغیر House را به کادر Test Variable List و متغیر Region را به کادر Grouping Variable منتقل کنید.



شکل (۶-۱۳) پنجره Tests for Several Independent Samples

سپس در همان شکل(۶-۱۳) بر روی دکمه Define Range کلیک کنید. با توجه به اینکه ۳ منطقه داریم و می خواهیم تفاوت را بین هر ۳ منطقه بسنجیم همانند شکل (۶-۱۴) عمل می کنیم، پس عدد ۱ را در کادر Minimum و عدد ۳ را در کادر Maximum بنویسید. سپس Continue را کلیک کرده و جهت اجرای آزمون ok را کلیک کنید.



شکل (۶-۱۴) پنجره Tests for Several Independent Samples Define Group

نتیجه شامل دو خروجی است. خروجی اول (جدول ۶-۱۲) تعداد مشاهدات و میانگین رتبه ای را برای مناطق ۱، ۲ و ۳ نشان می دهد.

جدول (۶-۱۲) نتایج توصیفی آزمون کروسکال والیس

Ranks

	Village	N	Mean Rank
House	Region 1	15	21.03
	Region 2	11	18.73
	Region 3	14	21.32
	Total	40	

خروجی دوم (جدول ۶-۱۳) به ترتیب مقدار آماره کای دو، درجه آزادی و sig را نشان می دهد. با توجه به اینکه sig آزمون بزرگ تر از ۵ درصد است لذا فرض H_0 رد نمی شود و می توانیم بگوییم که بین کیفیت مسکن در سه منطقه مورد مطالعه تفاوت معنی داری وجود ندارد.

جدول (۶-۱۳) نتایج نهایی آزمون کروسکال والیس

Test Statistics[a,b]

	House
Chi-Square	.403
df	2
Asymp. Sig.	.818

a. Kruskal Wallis Test

b. Grouping Variable: Village

۶-۷. آزمون فریدمن[1]

از آزمون فریدمن برای مقایسه میانگین های چند جامعه وابسته (داشتن سه یا بیش از سه جامعه) استفاده می گردد. این آزمون معادل ناپارامتری تحلیل واریانس یک طرفه (F) درون آزمودنی ها است. استفاده از این آزمون مشروط به وابسته بودن و رتبه‌ای بودن متغیرهای مورد مقایسه است. به عبارت دیگر آزمون فریدمن را می توان جهت بررسی یکسان بودن اولویت بندی(رتبه بندی) به کار گرفت.

مبنای این آزمون ، توزیع مربع کای و آماره آزمون نیز بر اساس رابطه (۶-۶) می باشد.

$$k^2 = [\frac{12}{NK(K+1)} + \sum_{J-1}^{K}(R_J)2] - 3N(K+1) \qquad (۶-۶)$$

1 . Friedman Test

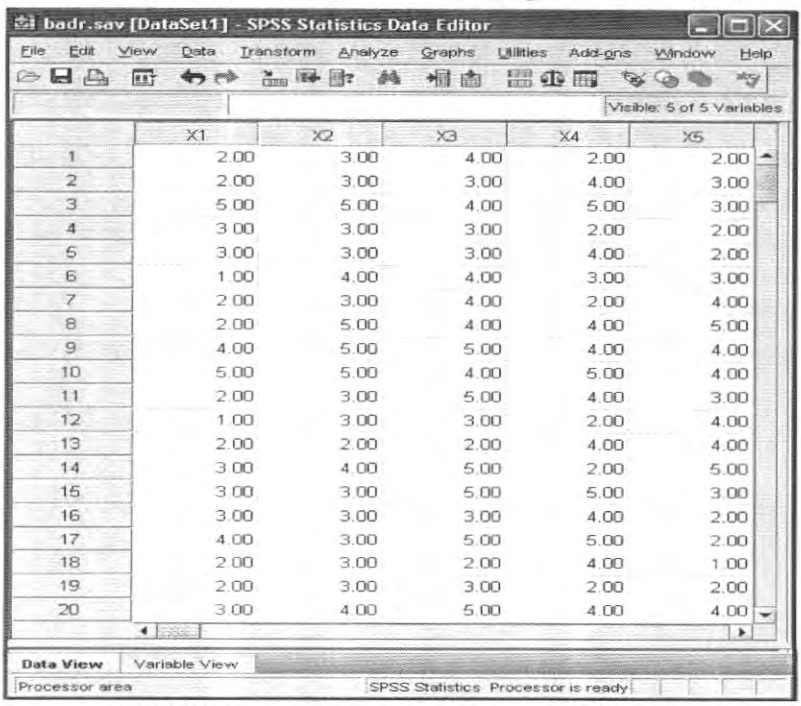
در این رابطه K تعداد مقوله های متغیر وابسته، N تعداد گروه های وابسته، $(R_J)^2$ مربع جمع رتبه های هر گروه است.

مثال: محققی پس از شناسایی عوامل مؤثر در مهاجرت های روستایی، می خواهد نگرش مهاجران را نسبت به اهمیت این عوامل مؤثر بدست آورد. وی می خواهد بداند که آیا همه این عواملی که وی شناسایی کرده است به یک اندازه در مهاجرت روستاییان تأثیر داشته است یا خیر؟ این عوامل عبارت اند از: ضعف بنیان های کشاورزی و عدم درآمد کافی حاصل از آن(X1)، ضعف خدمات و امکانات زیربنایی(X2)، عدم وضعیت اشتغال(X3)، جاذبه های زندگی شهری(X4) و سایر عوامل(تحصیل، ازدواج،...) (X5).

H_0: تأثیر همه عوامل یکسان است

H_1: دست کم یک عامل با سایرین فرق دارد

دقت داشته باشید که داده ها می باید همانند شکل (۶-۱۵) وارد محیط SPSS شوند.


badr.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Visible: 5 of 5 Variables

	X1	X2	X3	X4	X5
1	2.00	3.00	4.00	2.00	2.00
2	2.00	3.00	3.00	4.00	3.00
3	5.00	5.00	4.00	5.00	3.00
4	3.00	3.00	3.00	2.00	2.00
5	3.00	3.00	3.00	4.00	2.00
6	1.00	4.00	4.00	3.00	3.00
7	2.00	3.00	4.00	2.00	4.00
8	2.00	5.00	4.00	4.00	5.00
9	4.00	5.00	5.00	4.00	4.00
10	5.00	5.00	4.00	5.00	4.00
11	2.00	3.00	5.00	4.00	3.00
12	1.00	3.00	3.00	2.00	4.00
13	2.00	2.00	2.00	4.00	4.00
14	3.00	4.00	5.00	2.00	5.00
15	3.00	3.00	5.00	5.00	3.00
16	3.00	3.00	3.00	4.00	2.00
17	4.00	3.00	5.00	5.00	2.00
18	2.00	3.00	2.00	4.00	1.00
19	2.00	3.00	3.00	2.00	2.00
20	3.00	4.00	5.00	4.00	4.00

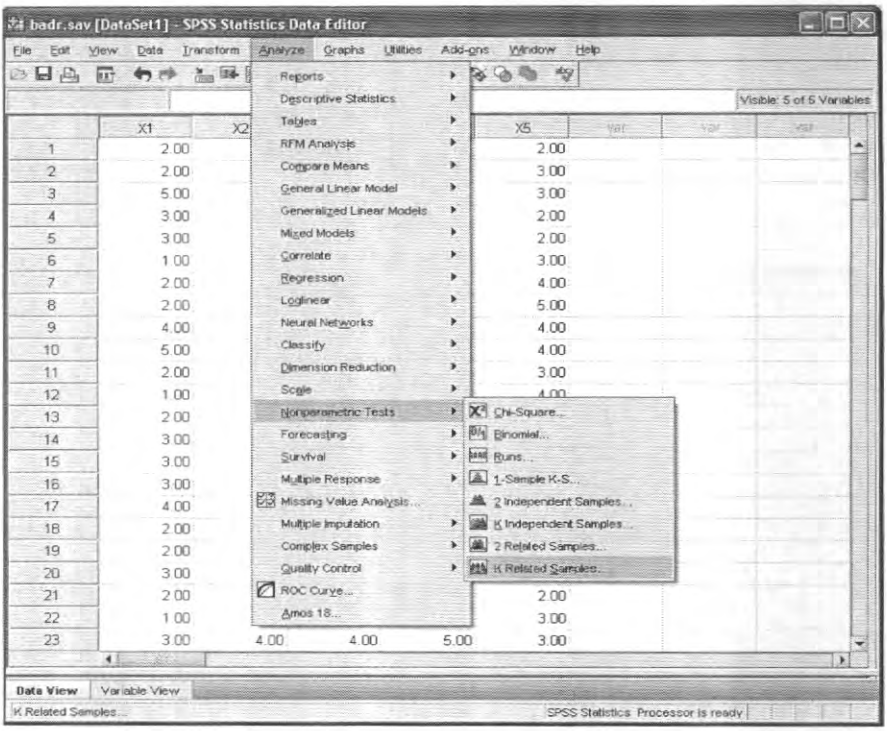
Data View | Variable View

Processor area | SPSS Statistics Processor is ready

شکل (۶-۱۵) پنجره SPSS Statistics Data Editor 7

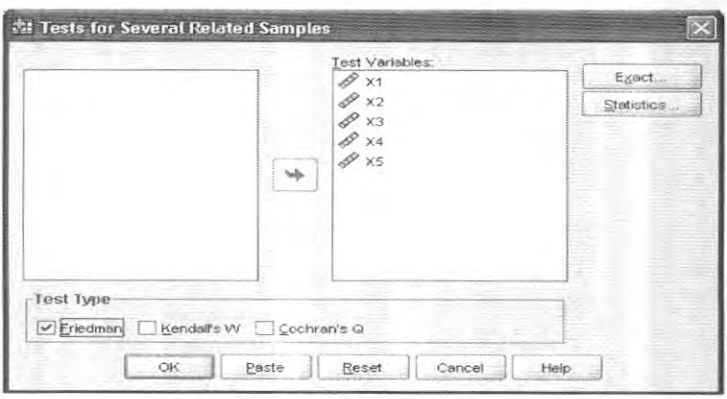
پس از وارد کردن داده ها مسیر شکل (۶-۱۶) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Nonparametric Tests ⟶ K Related Samples



شکل (۶-۱۶) پنجره SPSS Statistics Data Editor 8

پس از انتخاب این منو پنجره Tests for Several Related Samples باز می‌گردد. روی متغیرهای X1 ... X5 کلیک کرده و آنها را همانند شکل (۶-۱۷) به کادر Test Variables منتقل کنید. جهت اجرای آزمون فریدمن گزینه Ok را کلیک کنید.



شکل (۶-۱۷) پنجره Tests for Several Related Samples

نتیجه شامل دو خروجی می باشد. خروجی اول (جدول ۶-۱۴) میانگین رتبه ایی هر متغیر را نشان می دهد. هر چقدر میانگین رتبه هر متغیر بیشتر باشد، اهمیت آن نیز بیشتر است.

جدول (۶-۱۴) میانگین رتبه ای متغیرها

Ranks

	Mean Rank
X1	4.58
X2	3.00
X3	2.54
X4	2.22
X5	2.66

خروجی دوم (جدول ۶-۱۵) به ترتیب تعداد داده های هر متغیر، مقدار آماره کای دو (χ^2)، درجه آزادی و sig را ارائه می دهد. با توجه به این که sig کمتر از ۵ درصد است، H_0 رد شده و ادعای یکسان بودن نقش عوامل مؤثر در مهاجرت های روستایی پذیرفته نمی‌شود.

جدول (۶-۱۵) نتایج نهایی آزمون فریدمن

Test Statistics[a]

N	25
Chi-Square	38.304
df	4
Asymp. Sig.	.000

a. Friedman Test

فصل هفتم

همبستگی

مهمترین آزمون هایی که برای تبیین روابط بین متغیرها استفاده می شوند مربوط به آزمون های ضریب همبستگی می باشند. همبستگی وقتی حاصل می شود که ارتباط بین دو جامعه مورد بررسی قرار گیرد و این دو جامعه که با یکدیگر ارتباط دارند باید افرادشان متناظر و یا به عبارت دیگر جفت باشند، در غیر این صورت همبستگی مفهوم ندارد. مثلاً بین میزان تصادفات در فصل زمستان و تعداد تصادفات می توان یک همبستگی را نشان داد وقتی می خواهیم همبستگی بین متغیرها را بررسی کنیم باید انواع همبستگی را شناسایی کنیم. ۵ نوع ضریب همبستگی داریم:

۷-۱. انواع همبستگی

۷-۱-۱. همبستگی مستقیم کامل

ضریب همبستگی می باشد که اگر افراد جامعه اول با یک شکل منطقی و ثابت افزایش یا کاهش پیدا کنند، افراد جامعه دوم نیز با یک نظم ثابت و منطقی افزایش یا کاهش می یابند.

جامعه X : ۱ ۲ ۳ ۴ ۵

جامعه Y : ۲ ۴ ۶ ۸ ۱۰

۷-۱-۲. همبستگی مستقیم ناقص

در این همبستگی ارتباط مستقیم بین دو جامعه وجود دارد ولی کامل نیست به این معنی که افراد هر دو جامعه در یک جهت تغییر می کنند ولی به صورت ثابت و منطقی این افزایش یا کاهش قابل مشاهده نیست.

X جامعه : ۱ ۲ ۳ ۴ ۵ ۶

Y جامعه : ۱ ۲ ۴ ۷ ۱۰ ۱۱

۷-۱-۳. همبستگی معکوس کامل

هنگامی به وجود می آید که افزایش افراد یک جامعه همراه با کاهش افراد متناظر در جامعه دیگر به یک نسبت ثابت باشد.

X جامعه : ۱ ۲ ۳ ۴ ۵ ۶

Y جامعه : ۷ ۶ ۵ ۴ ۳ ۲

۷-۱-۴. همبستگی معکوس ناقص

در این نوع همبستگی اندازه واحدهای یکی از دو جامعه زیاد می شود و در مقابل اندازه واحدهای جامعه دیگر کم می گردد، ولی بین مقادیر افزایش یا کاهش روند ثابتی وجود نداشته باشد.

X جامعه : ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸

Y جامعه : ۱۳ ۱۰ ۹ ۷ ۶ ۳ ۲ ۱

۷-۱-۵. عدم همبستگی

این نوع همبستگی زمانی است که بین دو جامعه هیچ نوع همبستگی ای وجود ندارد و روندی در متغیرهای دو جامعه مشاهده نمی شود.

X جامعه : ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸

Y جامعه : ۱۰ ۳ ۴ ۳ ۸ ۲ ۱۰ ۹

برای اینکه بتوانیم به یک نُرم عددی در خصوص ۵ حالت گفته شده دسترسی پیدا کرد آماردانان به محاسبه ضرایبی به نام ضریب همبستگی پرداخته اند که مقدار این ضریب از ۱+ تا ۱- در نوسان است. معمولاً برای همبستگی مستقیم کامل مقدار این ضریب برابر با ۱+ و برای همبستگی معکوس کامل مقدار این ضریب ۱- می باشد. وقتی ضریب همبستگی بین صفر و یک ($0 < r < +1$) باشد همبستگی مستقیم ناقص و اگر ضریب همبستگی بین صفر و منفی یک ($0 > r > -1$) باشد همبستگی ناقص معکوس می باشد. وقتی که ضریب همبستگی صفر باشد نشان از عدم همبستگی دارد.

ضریب همبستگی سه چیز را مشخص می کند:

۱- اول اینکه آیا بین دو متغیر رابطه وجود دارد یا خیر؟

۲- دوم اینکه اگر رابطه بین متغیر ها وجود دارد شدت آن چقدر است؟

۳- سوم اینکه اگر رابطه وجود دارد جهت آن کدام است؟

هر چه شدت همبستگی به ۱+ و ۱- نزدیک تر باشد نشان از رابطه قوی بین دو متغیر است. اگر مقدار ضریب همبستگی بین ۲۵٪ تا ۳۵٪ مثبت یا منفی باشد همبستگی نسبتاً ضعیف، از ۳۵٪ تا ۶۵٪ تا حدودی قوی، از ۶۵٪ تا ۸۵٪ همبستگی قوی و از ۸۵٪ $\pm$ همبستگی بسیار قوی است.

سه نوع ضریب همبستگی بر اساس متغیرها داریم:

۷-۲. ضریب همبستگی پیرسون

زمانی از ضریب همبستگی پیرسون استفاده می شود که متغیر ها کمی باشند یعنی هر دو متغیر دارای مقیاس فاصله ای- نسبی باشند. فرمول پیرسون از طریق رابطه (۷-۱) به دست می آید.

$$r = \frac{C_{xy}}{\delta_n \delta_y} \qquad (۱-۷)$$

در این فرمول r مقدار ضریب همبستگی، C_{xy} کوواریانس دو جامعه و δ_y و δ_n ریشه دوم واریانس های دو جامعه x و y است. کوواریانس عبارت است از میانگین حاصلضرب انحراف از میانگین دو جامعه که از طریق رابطه (۷-۲) محاسبه می گردد.

$$C_{xy} = \frac{\sum Ni(x_i - \bar{X})(y_i - \bar{Y})}{N} \qquad (۲-۷)$$

مثال: بنا به تجربیات قبلی اعتقاد داریم که بین جمعیت(به نفر) و سطح زیر کشت(بر حسب هکتار) در روستا یک رابطه منطقی وجود دارد. برای این کار میزان جمعیت را با توجه به سطح زیر کشت آنها را در یک دوره زمانی ۱۰ ساله بدست آورده ایم. حال با توجه به اعداد بدست آمده آیا بین جمعیت و سطح زیر کشت رابطه وجود دارد یا خیر؟

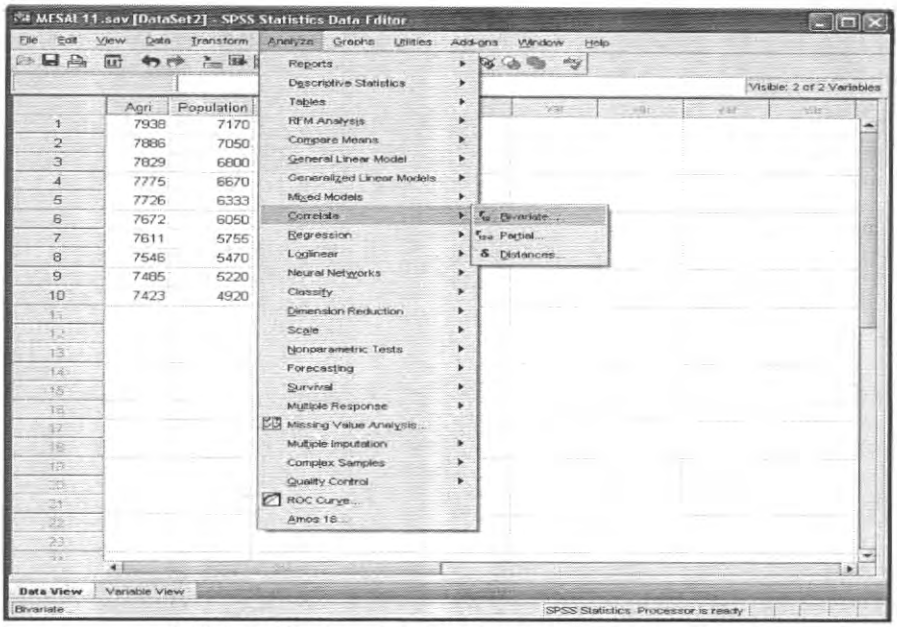
فرض های آماری در این آزمون عبارتند از:

H_0 : بین جمعیت و سطح زیر کشت رابطه وجود ندارد

H_1 : بین جمعیت و سطح زیر کشت رابطه وجود دارد

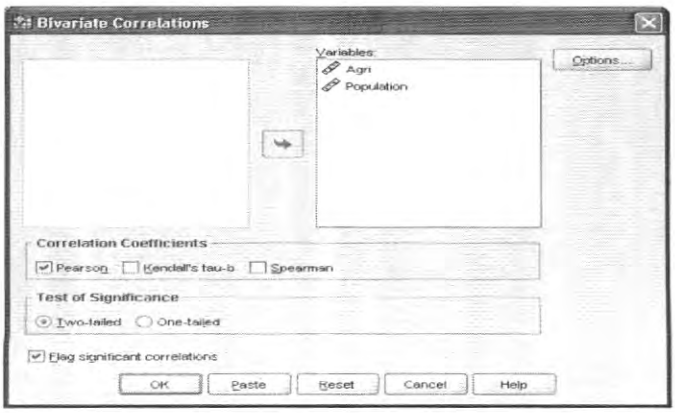
پس از وارد کردن داده ها مسیر شکل (۷-۱) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Correlate ⟶ Bivariate



شکل (۷-۱) پنجره SPSS Statistics Data Editor 1

پس از انتخاب این منو پنجره Bivariate Correlate باز می گردد. روی متغیرهای Agriو Population کلیک کرده و آنها را همانند شکل (۷-۲) به کادر Variables منتقل کنید. نوع ضریب همبستگی را در کادرCorrelate Coefficients می توانید مشخص کنید. با توجه به اینکه پیش فرض نرم افزار در این کادر Pearson است آن را تغییر ندهید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۷-۲) پنجره Bivariate Correlate 1

نتیجه شامل یک خروجی می باشد. این خروجی (جدول ۷-۱) ضریب همبستگی پیرسون، sig و تعداد مشاهدات را نشان می دهد. با توجه به ضریب همبستگی پیرسون(۰/۹۹۷) مشخص شده است که بین این دو متغیر یک ارتباط بسیار قوی وجود دارد. به عبارت دقیق تر می توان گفت که بین جمعیت و سطح زیر کشت در روستا رابطه مثبت و معنی داری در سطح اطمینان ۹۹ درصد وجود دارد. این بدان معناست که با کاهش جمعیت و مهاجرت در روستا سطح زیر کشت نیز کاهش می یابد و برعکس. لازم به ذکر است وجود دو ستاره بالای عدد ضریب همبستگی نشان دهنده معنی داری در سطح ۹۹ درصد، وجود یک ستاره نشان دهنده معنی داری در سطح ۹۵ درصد و عدم درج ستاره بالای این عدد حاکی از عدم وجود رابطه معنی دار بین دو متغیر است.

(جدول ۷-۱) نتایج ضریب همبستگی پیرسون

Correlations

		Agri	Population
Agri	Pearson Correlation	1	.997**
	Sig. (2-tailed)		.000
	N	10	10
Population	Pearson Correlation	.997**	1
	Sig. (2-tailed)	.000	
	N	10	10

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

۷-۳. ضریب همبستگی اسپرمن

اگر متغیرها کیفی باشند در این صورت باید از روشهای رتبه‌ای استفاده کرد. در این روش افراد جوامع به جای اینکه برحسب مقادیر صفتشان گروهبندی شوند بر اساس ترتیب و مرتبه آنها در جامعه گروهبندی می شوند. بدیهی است که مرتبه افراد در رابطه با ارزش گذاری هایی است که برای یک صفت بارز آنها به عمل می آید. یک از این روش ها روش همبستگی اسپرمن می باشد و معمولاً زمانی استفاده می شود که تعداد نمونه ها بزرگ (بیشتر از ۲۰) باشد. ضریب همبستگی رتبه ای اسپیرمن را با حروف یونانی ρ_S (رواندیس ρ) نشان می دهند و مقدار آن از طریق رابطه (۷-۳) به دست می آید:

$$\rho_S = 1 - \frac{6\sum d^2}{N^3 - N} \qquad (۳-۷)$$

در این فرمول ρ_S ضریب همبستگی، d تفاضل هر جفت رتبه متناظر و N تعداد افراد جامعه است.

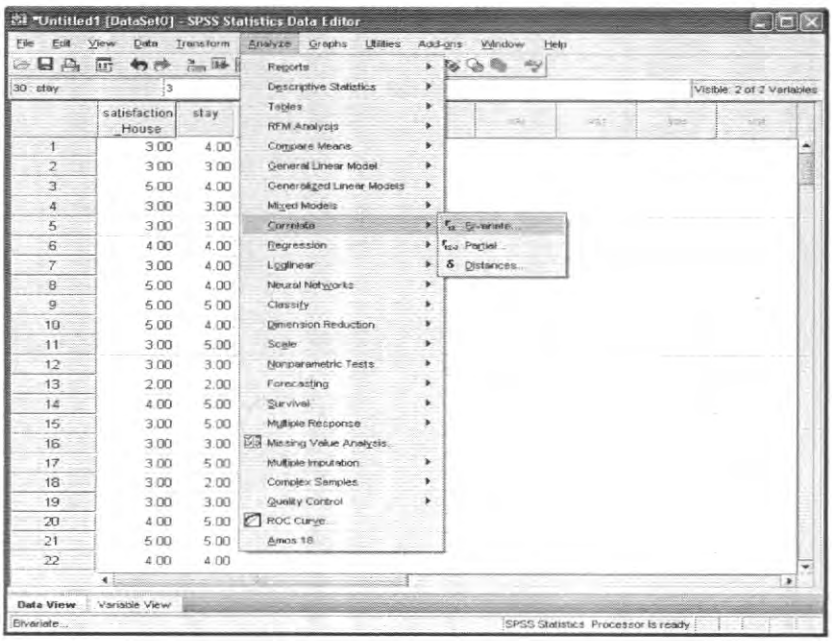
مثال: سطح رضایت از سکونت گاههای روستایی را با ماندگاری در روستا از طریق پرسشنامه بدست آورده ایم که رتبه های زیر بدست آمده اند. با توجه به رتبه های بدست آمده آیا می توان گفت که بین این دو متغیر رابطه وجود دارد؟ (رتبه ۱ به معنای کمترین و ۵ به معنای بیشترین سطح رضایت از سکونت گاههای روستایی و ماندگاری در روستا می باشد)

فرض های آماری در این صورت عبارتند از:

$$\begin{cases} H_0: \text{بین ماندگاری در روستا و رضایت از سکونتگاه رابطه وجود ندارد} \\ H_1: \text{بین ماندگاری در روستا و رضایت از سکونتگاه رابطه وجود دارد} \end{cases}$$

پس از وارد کردن داده ها مسیر شکل (۷-۳) را برای اجرای آزمون استفاده می کنیم.

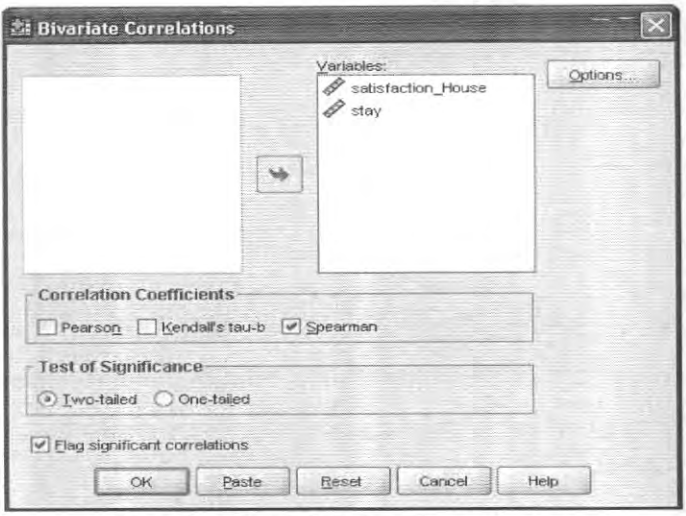
Analyze ⟶ Correlate ⟶ Bivariate



شکل (۷-۳) پنجره SPSS Statistics Data Editor 2

پس از انتخاب این منو پنجره Bivariate Correlate باز می گردد. روی متغیرهای satisfaction_House و stay کلیک کرده و آنها را همانند شکل (۷-۴) به کادر

Variables منتقل کنید. سپس نوع ضریب همبستگی را در کادر Correlate Coefficients تغییر داده و گزینه Spearman را فعال کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۷-۴) پنجره Bivariate Correlate 2

نتیجه شامل یک خروجی می باشد. این خروجی (جدول ۷-۲) ضریب همبستگی اسپرمن، sig و تعداد مشاهدات را نشان می دهد. با توجه به اینکه ضریب همبستگی اسپرمن(۰/۳۵۵) و sig آزمون در سطح اطمینان ۹۹ درصد معنی دار است، H_0 رد می شود و می توان گفت که بین دو متغیر ماندگاری در روستا و رضایت از سکونتگاه رابطه مثبت و معنی داری وجود دارد.

(جدول ۷-۲) نتایج ضریب همبستگی اسپرمن

Correlations

			satisfaction_House	stay
Spearman's rho	satisfaction_House	Correlation Coefficient	1.000	.355**
		Sig. (2-tailed)	.	.000
		N	200	200
	stay	Correlation Coefficient	.355**	1.000
		Sig. (2-tailed)	.000	.
		N	200	200

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

۷-۴. ضریب همبستگی تاو _ بی کندال

یکی از متداولترین روشهای همبستگی رتبه ای استفاده از ضریب همبستگی کندال است. هنگامی که مقیاس حداقل یکی از متغیرها ترتیبی باشد و حجم نمونه کوچک(کمتر از ۲۰) باشد، می توان از ضریب همبستگی کندال استفاده کرد. ضریب همبستگی کندال را با حرف یونانی τ (تاو) نمایش می دهند و از طریق رابطه (۷-۴) قابل محاسبه می باشد.

$$\tau = \frac{2s}{N(N-1)} \qquad (۷-۴)$$

که در آن S مجموع ضرایب همبستگی جفت و جور شده اعضای دو جامعه مورد نظر و N تعداد جامعه می باشد.

مثال: محققی قصد دارد تا به بررسی ارتباط بین جمعیت روستاها با میزان کیفیت زندگی ساکنانشان بپردازد. بنابراین ضمن آگاهی از جمعیت روستاها از طریق دهیاران هر روستا کیفیت زندگی را با طیف لیکرت پنج سطحی مورد سئوال قرار می دهد. حال با توجه به رتبه‌های بدست آمده می توان گفت که از نگاه دهیاران بین این دو متغیر رابطه وجود دارد؟

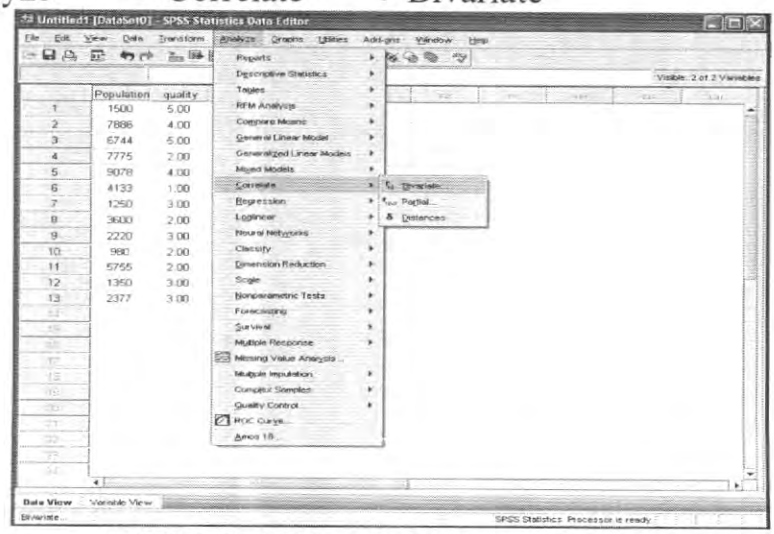
فرض های آماری در این صورت عبارتند از:

H_0: بین جمعیت روستاها با میزان کیفیت زندگی ساکنانش از نگاه دهیاران رابطه وجود ندارد

H_1: بین جمعیت روستاها با میزان کیفیت زندگی ساکنانش از نگاه دهیاران رابطه وجود دارد

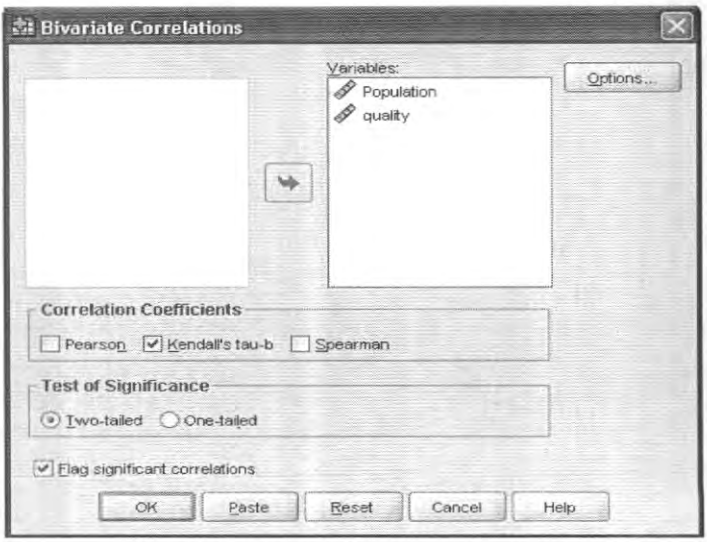
پس از وارد کردن داده ها مسیر شکل (۷-۵) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Correlate ⟶ Bivariate



شکل (۷-۵) پنجره SPSS Statistics Data Editor 3

پس از انتخاب این منو پنجره Bivariate Correlate باز می گردد. روی متغیرهای Populationو quality کلیک کرده و آنها را همانند شکل (۷-۶) به کادر Variables منتقل کنید. سپس نوع ضریب همبستگی را در کادر Correlate Coefficients تغییر داده و گزینه Kendalls tau-b را فعال کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۷-۶) پنجره Bivariate Correlate 3

نتیجه شامل یک خروجی می باشد. این خروجی (جدول ۷-۳) ضریب همبستگی تاو _ بی کندال، sig و تعداد مشاهدات را نشان می دهد. با توجه به اینکه sig آزمون بزرگ تر از ۵ درصد است، H_0 رد نمی شود و می توان گفت بین جمعیت روستاها با میزان کیفیت زندگی ساکنانش از نگاه دهیاران رابطه وجود ندارد.

جدول (۷-۳) نتایج ضریب همبستگی تاو بی کندال

Correlations

			Population	quality
Kendall's tau_b	Population	Correlation Coefficient	1.000	.085
		Sig. (2-tailed)	.	.704
		N	13	13
	quality	Correlation Coefficient	.085	1.000
		Sig. (2-tailed)	.704	.
		N	13	13

فصل هشتم

رگرسیون

در عمل بسیاری از پیشامدها و پدیده ها تابعی از پیشامد های دیگر و تأثیر پذیر از کیفیت و کمیت آنها می باشد. مثلاً دادن کود باعث افزایش محصولات کشاورزی می‌شود؛ بارندگی تحت تأثیر دما قرار دارد؛ میزان درآمد در ارتباط مستقیم با ساعت کار می‌باشد. بنابراین یکی از هدفهای مهم در تحقیق های علمی کشف وجود یا عدم وجود رابطه بین پدیده‌ها و متغیر های مختلف می باشد.

تحلیل رگرسیون یک روش آماری برای بررسی و مدل سازی رابطه بین متغیرها است که تقریباً در تمامی علوم مورد استفاده قرار می گیرد. در حقیقت تحلیل رگرسیون یک جنبه از آمار کاربردی است که به محقق امکان می دهد 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. اساس تفکر رگرسیون این است که ببینیم آیا رابطه خطی معنی داری بین دو یا چند متغیر وجود دارد یا نه؟

انواع رگرسیون:

۸-۱. رگرسیون خطی ساده

این رگرسیون زمانی مورد استفاده قرار می گیرد که متغیر وابسته تنها با یک متغیر مستقل در ارتباط باشد. یا به عبارتی متغیرها نباید از دو تا بیشتر باشد. رگرسیون خطی ساده از طریق رابطه (۸-۱) محاسبه می گردد.

$$Y= a + bx \qquad (۱-۸)$$

یعنی متغیر y به صورت خطی وابسته به متغیر x می باشد که اگر صحت چنین رابطه‌ای تأیید شود می توانیم به ازای مقادیر مختلف x برآوردی از مقادیر y مورد انتظار ارائه دهیم.

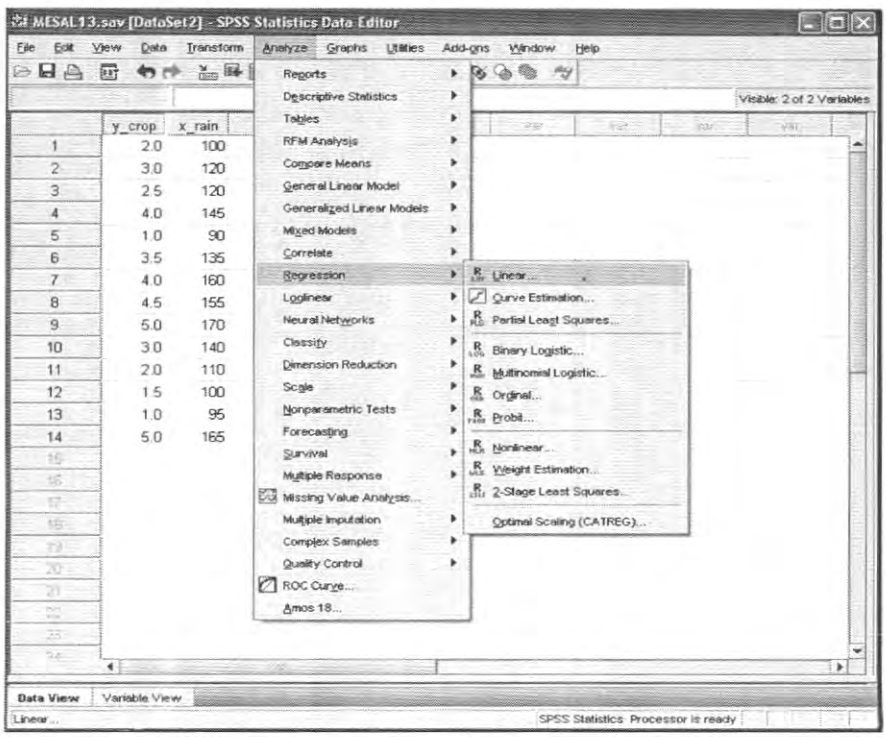
فرض ها در رگرسیون به شکل زیر می باشند:

$$\begin{cases} H_0: \beta = 0 & \text{هیچ رابطه خطی بین دو متغیر وجود ندارد} \\ H_1: \beta \neq 0 & \text{رابطه خطی بین دو متغیر وجود دارد} \end{cases}$$

مثال: می خواهیم بدانیم که بین میزان محصول(بر حسب تن) و بارندگی رابطه خطی معنی داری وجود دارد یا نه. برای این منظور به گردآوری داده ها پرداخته ایم. حال با توجه به این داده ها آیا می توان گفت که بین این دو متغیر رابطه خطی معنی داری وجود دارد یا خیر؟

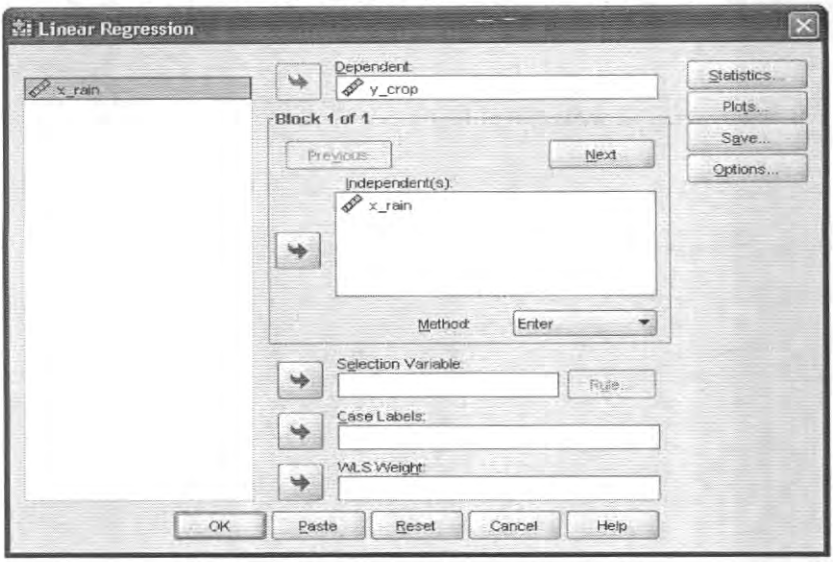
پس از وارد کردن داده ها مسیر شکل (۸-۱) را برای اجرای آزمون استفاده می کنیم.

Analyze ⟶ Regression ⟶ Linear



شکل (۸-۱) پنجره SPSS Statistics Data Editor 1

پس از انتخاب این منو پنجره Linear Regression باز می گردد. همانند شکل (۸-۲) متغیر y_crop را به کادر Dependent و متغیر x_rain را به کادر Independent(s) منتقل کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۸-۲) پنجره Linear Regression 1

نتیجه شامل چهار خروجی می باشد. خروجی اول (جدول ۸-۱) متغیرهای مستقل وارد شده (که در رگرسیون خطی ساده تنها یک متغیر است)، متغیرهای حذف شده و روش مورد استفاده در تعیین رگرسیون را نشان می دهد.

جدول (۸-۱) پنجره Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	x_rain[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: y_crop

خروجی دوم (جدول ۸-۲) به ترتیب ضریب همبستگی چندگانه، ضریب تعیین(میزان تغییرپذیری در متغیر وابسته که می توان به وسیله رگرسیون آن را توضیح داد)، ضریب تعیین تعدیل شده و خطای معیار تخمین را نشان می دهد.

جدول (۸-۲) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.974[a]	.950	.945	.3242

a. Predictors: (Constant), x_rain

تفاوت بین R Square و Adjusted R Square : تفاوت بین ضریب تعیین و ضریب تعیین تعدیل شده می تواند ناشی از حجم نمونه و تعداد متغیرها باشد. در صورتی که نمونه کوچک باشد، ضریب تعیین تعدیل شده برای تفسیر مناسب تر است. با بزرگ شدن حجم نمونه این دو ضریب به هم نزدیک می شوند. با افزودن تعداد متغیرها به مدل رگرسیون، مقدار R Square ممکن است افزایش یابد. برای جلوگیری از این وضعیت و کنترل تورم(تورش)R Square ، آماره دیگری به نام Adjusted R Square مطرح می شود که مشکلات R Square را برطرف می سازد.

مقدار Adjusted R Square به کمک رابطه(۸-۲) به دست می آید. در این رابطه n تعداد مشاهدات و k تعداد متغیرهای مستقل است.

$$r^2 = 1 - \left[\frac{n-1}{k-1}(1 - r^2)\right] \qquad (۸-۲)$$

خطای معیار تخمین(S_e): این شاخص میزان پراکندگی نقاط را حول خط رگرسیون اندازه گیری می کند. هرچه مقدار این شاخص بیشتر باشد پراکندگی نقاط حول خط رگرسیون بیشتر خواهد بود. شاخص خطای معیار تخمین از طریق رابطه () محاسبه می-گردد.

$$S_e = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n-k-1}} \qquad (۸-۳)$$

خروجی سوم (جدول ۸-۳) که برای آزمون رگرسیون از طریق روش خرد کردن مجموع انحرافات y محاسبه شده است و مبتنی بر آماره آزمون F می باشد، مقادیر مجموع مربعات، درجه آزادی، میانگین مربعات، مقدار آماره فیشر و سطح معنی داری مدل رگرسیون گزارش شده است. بر مبنای مقدار Sig به دست آمده که کمتر از ۵ درصد می‌باشد، فرض H_0 رد می شود. یعنی بین میزان محصول و بارندگی یک رابطه خطی وجود دارد.

جدول (۸-۳) ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.739	1	23.739	225.848	.000[a]
	Residual	1.261	12	.105		
	Total	25.000	13			

a. Predictors: (Constant), x_rain

b. Dependent Variable: y_crop

خروجی چهارم (جدول ۸-۴) در ستون B به ترتیب مقدار ثابت و ضریب متغیر مستقل در معادله رگرسیون ارایه شده است. ضرایب استاندارد نشده در ستون B و ضرایب استاندارد شده در ستون Beta مشخص می گردند. در ضرایب استاندارد نشده مقیاس متغیرها با همدیگر یکسان نیستند در صورتی که در ضرایب استاندارد شده مقیاس متغیرها یکسان شده و امکان مقایسه متغیرها وجود دارد. بنابراین جهت مقایسه اثرات چندین متغیر مستقل روی متغیر وابسته از ضرایب استاندارد شده استفاده می شود. در مثال فوق به ازای یک واحد تغییر در میزان بارندگی، ۰/۹۷۴ تغییر در میزان محصول ایجاد می شود. با توجه به اینکه Sig آزمون تساوی ضریب رگرسیون و مقدار ثابت با صفر کوچک تر از ۵ درصد است. بنابراین فرض تساوی این دو ضریب با صفر رد می شود و نباید آن ها را از معادله رگرسیون خارج کرد. نهایتاً از طریق این جدول می توان معادله خطی را به این صورت نوشت:

$$Y= -3.311+ 0.049(x)$$

جدول (۸-۴) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.311	.429		-7.721	.000
	x_rain	.049	.003	.974	15.028	.000

a. Dependent Variable: y_crop

۸-۲. رگرسیون خطی چند گانه یا چند متغیری

این روش نوعی تحلیل رگرسیون است که یک متغیر وابسته را در ارتباط با چندین متغیر مستقل بررسی می کند. مثلاً مهاجرت می تواند تابعی از متغیرهای کمبود درآمد، کمبود زمین، سن، فاصله از شهر و ... باشد. در محاسبات رگرسیون خطی چندگانه این قابلیت وجود دارد تا سهم هر متغیر مستقل را بر روی متغیر وابسته به صورت جداگانه بسنجیم. بطور کلی از رگرسیون چند متغیری برای بررسی رابطه متغیرها در مقیاس فاصله ای استفاده می شود. همچنین این تحلیل به تعداد زیادی مشاهده نیاز دارد تا فرض نرمال بودن برقرار شود. رابطه خطی رگرسیون چند گانه به صورت زیر می باشد:

$$\acute{Y} = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$ رابطه (۸-۴)

انواع روش های ورود متغیرها در رگرسیون

۱. **روش Enter**: رگرسیون به شیوه Enter، روشی است که در آن کلیه متغیرهای مستقل هم زمان وارد تحلیل شده و اثرات کلیه متغیرهای مستقل بر روی متغیر وابسته بررسی می گردد.

۲. **روش Stepwise**: در این روش متغیرهای مستقل یکی یکی به معادله رگرسیون اضافه می شوند و اگر نقش معنی داری در رگرسیون نداشته باشند (با توجه به sig آزمون خطی بودن رابطه متغیرها) از آن حذف می شوند.

۳. **روش Remove**: این روش رویکردی در انتخاب متغیرها است که در آن کلیه متغیرهای وارد شده در یک مرحله حذف می شوند.

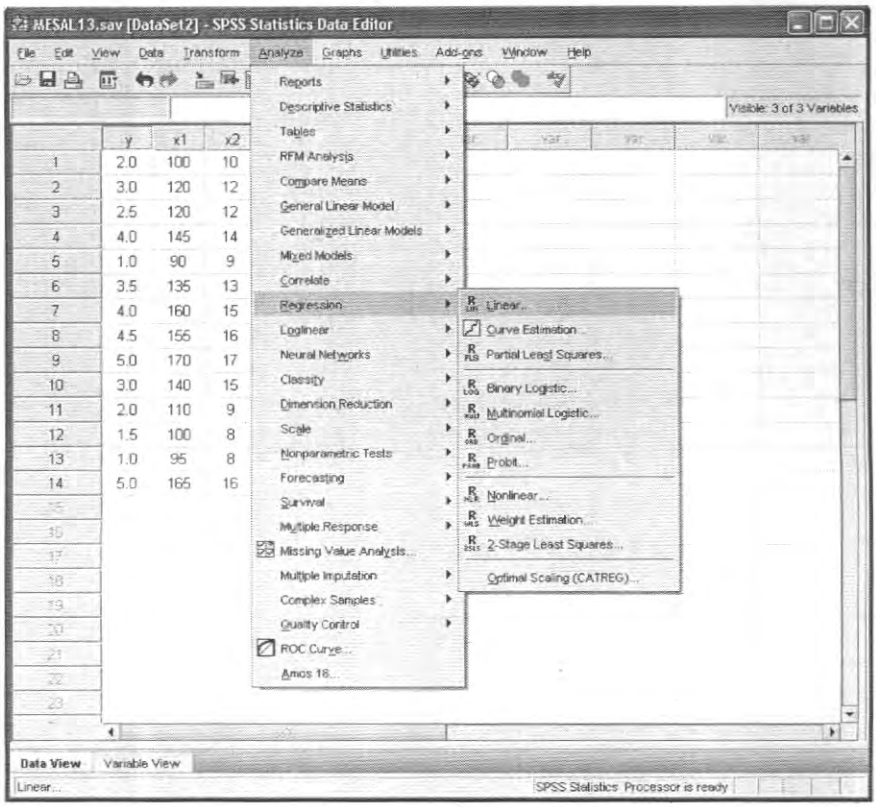
۴. **روش Backward**: در این روش همانند روش Enter، ابتدا کلیه های متغیرهای مستقل وارد معادله شده واثر کلیه متغیرها بر روی متغیر وابسته سنجیده می شود. سپس به مرور متغیرهای ضعیف تر یکی پس از دیگری از معادله خارج شده و حذف می گردند.

۵. **روش Forward**: این روش شبیه روش Stepwise است. در این روش نیز متغیرهای قوی تر انتخاب و وارد معادله می گردند. این عمل تا زمانی ادامه می یابد که خطای آزمون به ۵ درصد برسد. نتایج به دست آمده در این روش همانند روش Stepwise می باشد.

مثال: در ادامه مثال قبلی متغیر دیگر به نام درجه حرارت را نیز لحاظ می کنیم حال می‌خواهیم ببینیم که آیا بین این دو متغیر (بارندگی و درجه حرارت) با میزان محصول رابطه خطی وجود دارد ؟

پس از وارد کردن داده ها مسیر شکل (۸-۳) را برای اجرای آزمون استفاده می کنیم.

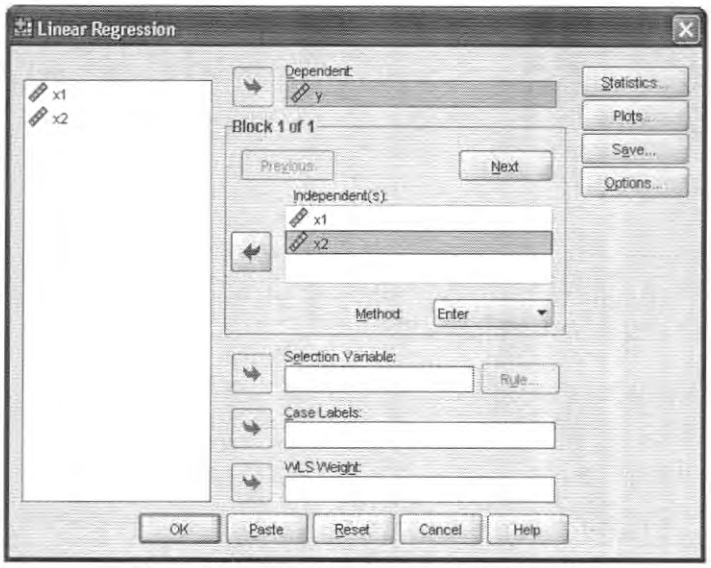
Analyze ⟶ Regression ⟶ Linear



	y	x1	x2
1	2.0	100	10
2	3.0	120	12
3	2.5	120	12
4	4.0	145	14
5	1.0	90	9
6	3.5	135	13
7	4.0	160	15
8	4.5	155	16
9	5.0	170	17
10	3.0	140	15
11	2.0	110	9
12	1.5	100	8
13	1.0	95	8
14	5.0	165	16

شکل (۸-۳) پنجره SPSS Statistics Data Editor 2

پس از انتخاب این منو پنجره Linear Regression باز می گردد. همانند شکل (۸-۴) متغیر y را به کادر Dependent و متغیرهای x1 و x2 را به کادر Independent(s) منتقل کنید. جهت اجرای آزمون گزینه Ok را کلیک کنید.



شکل (۸-۴) پنجره Linear Regression 2

نتیجه شامل چهار خروجی می باشد. خروجی اول (جدول ۸-۵) متغیرهای مستقل وارد شده (که در رگرسیون خطی ساده تنها یک متغیر است)، متغیرهای حذف شده و روش مورد استفاده در تعیین رگرسیون را نشان می دهد.

جدول (۸-۵) Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	x2, x1[a]	.	Enter

a. All requested variables entered.

خروجی دوم (جدول ۸-۶) به ترتیب ضریب همبستگی چندگانه، ضریب تعیین، ضریب تعیین تعدیل شده و خطای معیار تخمین را نشان می دهد.

جدول (۸-۶) Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.942	.3334

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.942	.3334

a. Predictors: (Constant), x2, x1

خروجی سوم (جدول ۸-۷) مقادیر مجموع مربعات، درجه آزادی، میانگین مربعات، مقدار آماره فیشر و سطح معنی داری مدل رگرسیون را ارایه می دهد. با توجه به اینکه Sig به دست آمده کمتر از ۵ درصد می باشد، H_0 رد می شود. یعنی بین متغیر وابسته(میزان محصول) و متغیرهای مستقل(بارندگی و درجه حرارت) یک رابطه خطی وجود دارد.

جدول (۸-۷) ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.777	2	11.889	106.960	.000[a]
	Residual	1.223	11	.111		
	Total	25.000	13			

a. Predictors: (Constant), x2, x1

b. Dependent Variable: y

با توجه به خروجی چهارم (جدول ۸-۸) می توان فهمید که تأثیر کدام یک از این دو عامل بیشتر است. این کار از دو طریق میزان Sig یا از طریق مقدار Beta امکان پذیر می باشد، در صورتی که مقدار sig هر متغیر کوچک تر از ۵ درصد باشد یا مقدار Beta آن بالاتر باشد، تأثیر آن متغیر بیشتر است. در مثال حاضر sig برای متغیر میزان بارندگی کوچک تر از ۵ درصد و مقدار Beta آن نیز بالاتر از متغیر درجه حرارت می باشد. در نتیجه تأثیر بارندگی در افزایش محصول بیشتر است. بنابراین از طریق این جدول می توان معادله خطی را به این صورت نوشت:

$$\acute{Y} = -3.186 + 0.042(x_1) + 0.066(x_2)$$

جدول (۸-۸) Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.186	.489		-6.515	.000
	x1	.042	.013	.829	3.255	.008
	x2	.066	.111	.150	.590	.567

a. Dependent Variable: y

۸-. رگرسیون لجستیک ۳

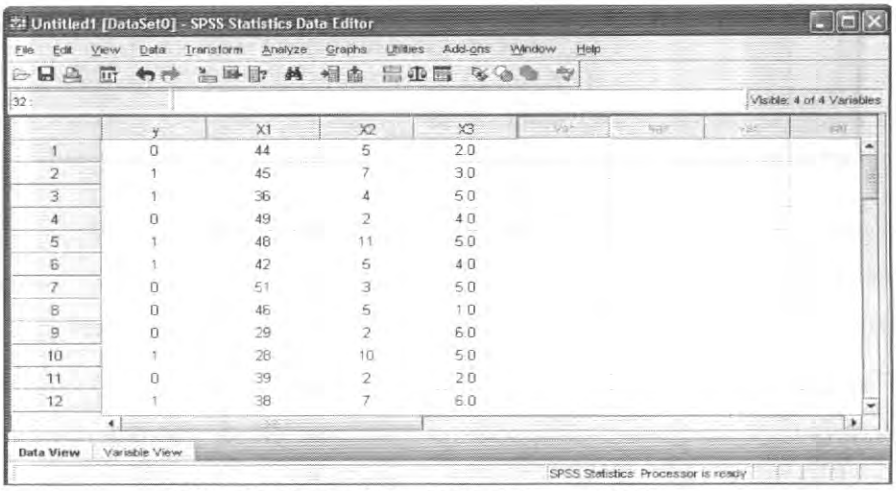
از رگرسیون لجستیک زمانی استفاده می گردد که متغیر وابسته دارای دو نتیجه باشد (مانند: بلی، خیر، پذیرش، عدم پذیرش و غیره) و تنها یکی از دو ارزش صفر یا یک را بپذیرد. در این صورت، صفر به معنای عدم وقوع حادثه و یک به معنای وقوع حادثه مورد نظر است. مثلاً محققی می خواهد اثر قبول شدن در کنکور را با توجه به مدت مطالعه، سن و معدل دیپلم بررسی کند. در این حالت عدم قبولی در کنکور یا قبولی در آن، دو حالت ممکن برای متغیر وابسته این تحقیق است.

به طور کلی اگر بخواهیم اثر چند متغیر را به عنوان متغیرهای مستقل، بر روی یک متغیر به عنوان متغیر وابسته تحقیق که از نوع اسمی دو وجهی می باشد بسنجیم، باید به جای رگرسیون چندگانه از رگرسیون لجستیک استفاده کنیم. در این روش براساس متغیرهای مستقل مورد استفاده می توان احتمال هر یک از سطوح متغیر دو وجهی وابسته را محاسبه کرد.

رگرسیون لجستیک شبیه به رگرسیون معمولی است با این تفاوت که روش تخمین ضرائب یکسان نمی باشد. در رگرسیون لجستیک، به جای حداقل کردن مجذور خطاها که در رگرسیون معمولی انجام می گیرد، احتمالی را که یک واقعه رخ می دهد حداکثر می کند. همچنین در تحلیل رگرسیون معمولی، برای بررسی معنی دار بودن رابطه از آمارهای استاندارد F و t استفاده می شود در حالی که در رگرسیون لجستیک از آمارهای "کای دو" (X^2) و "والد" (Wald) استفاده می شود. آماره "کای دو" به منظور تعیین میزان اثر گذاری متغیر (متغیرهای) مستقل بر متغیر وابسته و به طور کلی برازش کل مدل می باشد و قابل مقایسه با آماره F در تحلیل رگرسیون معمولی است. آزمون والد

نیز معنی دار بودن متغیرهای وارد شده در معادله رگرسیون را بررسی می کند و قابل مقایسه با آمار t در رگرسیون معمولی است.

مثال: محققی قصد دارد تا عوامل مؤثر بر پذیرش سیستم های آبیاری تحت فشار را با توجه به ویژگی هایی همچون سن، سطح تحصیلات و میزان کل اراضی پیش بینی کند. برای این منظور باید ابتدا پذیرش سیستم های آبیاری تحت فشار و عدم پذیرش آنها (متغیر وابسته) را در میان افراد به صورت کدهای ۰ و ۱ و سایر متغیرهای مستقل نیز همانند شکل (۸-۵) در ستون های دیگر پنجره Data view وارد گردند.



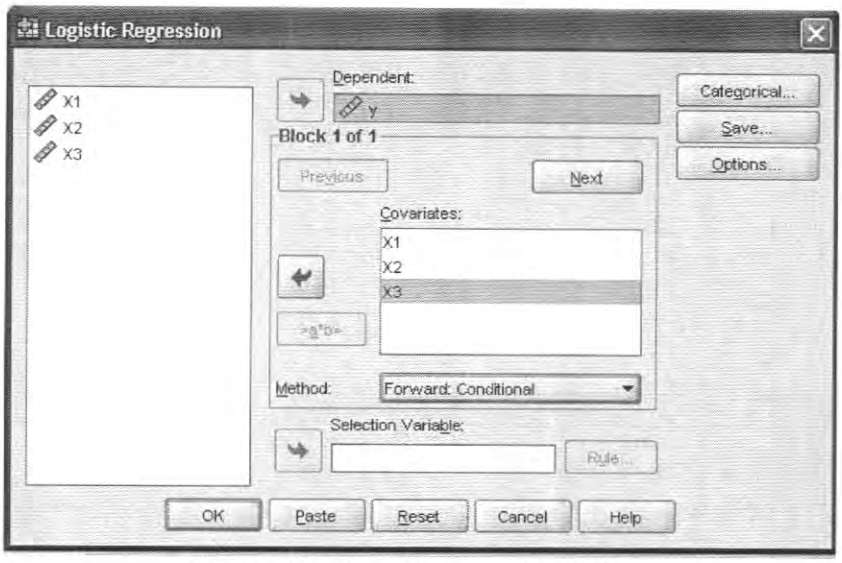
	y	X1	X2	X3
1	0	44	5	2.0
2	1	45	7	3.0
3	1	36	4	5.0
4	0	49	2	4.0
5	1	48	11	5.0
6	1	42	5	4.0
7	0	51	3	5.0
8	0	46	5	1.0
9	0	29	2	6.0
10	1	28	10	5.0
11	0	39	2	2.0
12	1	38	7	6.0

شکل (۸-۵) پنجره SPSS Statistics Data Editor 3

پس از وارد کردن داده ها مسیر زیر را اجرا کنید.

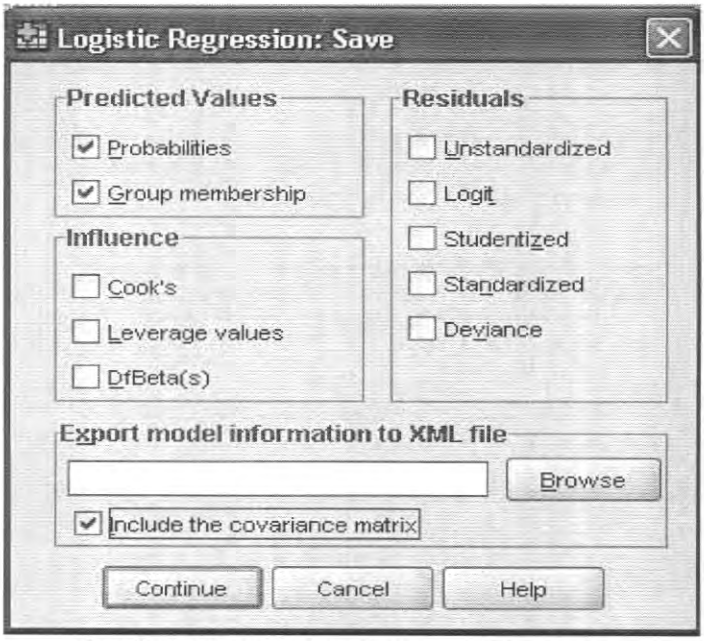
Analyze ⟶ Regression ⟶ Binary Logistic

پس از انتخاب این منو پنجره Linear Regression باز می گردد. همانند شکل (۸-۶)، متغیر وابسته را به قسمت Dependent و متغیرهای مستقل را به قسمت Covariates انتقال دهید. در قسمت Method گزینه Forward: Conditional را به عنوان روش کار انتخاب کنید.



شکل (۸-۶) پنجره Linear Regression

در همین پنجره دکمه Save را کلیک کنید تا پنجره مربوطه باز گردد، در این پنجره در قسمت Predicted Values گزینه های Probabilities و Group membership را همانند (شکل ۸-۷) فعال کنید. گزینه Probabilities تعلق فرد به گروههای دوگانه را براساس متغیرهای مستقل پیش بینی می کند. چنانچه احتمال پذیرش بیش از ۰/۵ باشد فرد سیستم های آبیاری تحت فشار را به عنوان روش آبیاری خود پذیرفته است و چنانچه این احتمال کمتر از ۰/۵ باشد فرد این سیستم ها را نپذیرفته است. همچنین گزینه Group membership این احتمال را براساس تعلق فرد به گروه اول (که در این مثال پذیرش سیستم های آبیاری تحت فشار می باشد) نشان می دهد. سپس دکمه Continue را کلیک کنید تا به پنجره اصلی باز گردید. حال برای نمایش نتایج گزینه Ok را کلیک کنید.



شکل (۸-۷) پنجره Logistic Regression Save

نتیجه شامل ۱۱ خروجی می باشد که در این بخش خروجی های با اهمیت تر به تفکیک ارائه می شود. خروجی اول جدولی تحت عنوان Case Processing Summary است که تعداد افراد مورد بررسی، داده های گم شده و درصد افراد تحلیل شده را نشان می‌دهد. بر طبق نتایج جدول (۸-۹) تعداد افراد مورد بررسی ۲۵ نفر بوده و هیچگونه داده گم شده ای در آن وجود نداشته است و ۱۰۰ درصد افراد در تحلیل وارد شده اند.

جدول (۸-۹) Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	25	100.0
	Missing Cases	0	.0
	Total	25	100.0
Unselected Cases		0	.0
Total		25	100.0

خروجی دوم (جدول ۸-۱۰) نشان می دهد که متغیر وابسته با کدهای ۱ برای پذیرش و ° برای عدم پذیرش اختصاص یافته است.

جدول (۸-۱۰) Dependent Variable Encoding

Original Value	Internal Value
عدم پذیرش	0
پذیرش	1

خروجی ششم (جدول ۸-۱۱) کای دو (χ^2) مدل را نشان می دهد با توجه به اینکه کای دو مدل در گام دوم (Step2) برابر با 20/433 بوده که sig آن کمتر از ۵ درصد می باشد، بنابراین متغیرهای مستقل بر متغیر وابسته تأثیر داشته و می توان گفت که به طور کلی برازش کل مدل مناسب است.

جدول (۸-۱۱) Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	13.909	1	.000
	Block	13.909	1	.000
	Model	13.909	1	.000
Step 2	Step	6.524	1	.011
	Block	20.433	2	.000
	Model	20.433	2	.000

خروجی هفتم جدولی با عنوان Model Summary می باشد در این جدول ستونی تحت عنوان Nagelkerke R Square درج شده است که معادل R^2 در رگرسیون خطی می باشد. نتایج (جدول ۸-۱۲) نشان می دهد که ۷۴/۸ درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل رگرسیون لجستیک تبیین می شود.

جدول (۸-۱۲) **Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	20.388[a]	.427	.572
2	13.864[b]	.558	.748

خروجی بعدی جدولی تحت عنوان Classification Table می باشد که ابعاد دو وجهی متغیر وابسته را نشان می دهد این جدول نمایشی از مقادیر مشاهده شده و مقادیر پیش بینی شده است. براساس این خروجی (جدول ۸-۱۳) مشخص می شود که چقدر پیش بینی مدل درست می باشد. به عبارت دیگر درصدهای این جدول حساسیت مدل در تعیین افراد مربوط به یک حالت را نشان می دهد. درصد کل در این جدول نشان می-دهد که این مدل تا چند درصد کل افراد را به درستی از همدیگر تفکیک کرده است. بر این اساس، در گام اول این جدول، برای ۲۵ نفر آزمودنی مورد نظر، ۸۰ درصد کل افراد را به درستی تفکیک کرده است و در گام دوم حساسیت مدل برای تعیین افراد پذیرنده سیستم های آبیاری تحت فشار به ۸۴ درصد افزایش یافته است.

جدول (۸-۱۳) **Classification Table**

Observed			Predicted		
			y		
			پذیرش	عدم پذیرش	Percentage Correct
Step 1	y	پذیرش	7	4	63.6
		عدم پذیرش	1	13	92.9
	Overall Percentage				80.0
Step 2	y	پذیرش	9	2	81.8
		عدم پذیرش	2	12	85.7
	Overall Percentage				84.0

a. The cut value is .500

خروجی بعدی جدول Variables in the Equation می باشد. در این جدول مقدار ثابت، ضریب B، نسبت برتری (EXP(B))، آماره والد و غیره را نشان می دهد. در این قسمت ضریب مثبت متغیر سطح سواد نشان دهنده این است که تحصیلات رابطه مثبت و معنی داری با پذیرش سیستم ها داشته است. ضریب منفی متغیر سن نیز بیانگر این است که افراد با سن بیشتر، کمتر این سیستم ها را پذیرفته اند. به عبارت دیگر با افزایش سن، پذیرش سیستم های آبیاری تحت فشار کمتر شده است. با توجه به نتایج جدول (۸-۱۴) می توان معادله بهینه رگرسیون لجستیک را به شرح ذیل نوشت:

برتری لگاریتمی = ۳/۱۳۴ − ۰/۲۰۵(سن) + ۱/۲۵۰(سطح سواد)

جدول (۸-۱۴) Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	سطح سواد	.802	.332	5.815	1	.016	2.229
	Constant	-4.002	1.725	5.385	1	.020	.018
Step 2[b]	سن	-.205	.105	3.785	1	.052	.815
	سطح سواد	1.250	.601	4.332	1	.037	3.490
	Constant	3.134	3.266	.920	1	.337	22.956

a. Variable(s) entered on step 1: سطح سواد.

b. Variable(s) entered on step 2: سن.

فصل نهم

تحلیل عاملی

هدف بیشتر مطالعات تحلیل عاملی خلاصه کردن ماتریس همبستگی‌ها به شیوه‌ای است که بتوان آنها را بر حسب چند عامل زیر بنایی تبیین کرد. به عبارت دیگر تحلیل عاملی سعی در شناسایی ابعاد زیر بنایی تحقیق به منظور تبیین الگوی همبستگی بین متغیرهای مشاهده شده دارد. خلاصه نمودن اطلاعات در این روش به ترتیبی صورت می‌پذیرد که نتیجه خلاصه شده از نظر مفهوم معنی دار است. بنابراین از این روش زمانی که هدف محقق یافتن عامل های مکنون جهت خلاصه کردن متغیرها از بین مجموعه ای از آنها باشد، استفاده می شود.

عامل متغیر جدیدی است که از طریق ترکیب خطی مقادیر اصلی متغیرهای مشاهده شده به صورت رابطه (۹-۱) برآورد می شود. در این رابطه X_i بیانگر متغیر i ام، W_{ji} ضریب نمره عاملی متغیر i ام و از نظر عامل j ام، p تعداد متغیرها و F_j عامل j ام است.

$$F_j = \sum W_{ji} X_i = W_{j1} X_1 + W_{j2} X_2 + \cdots + W_{jp} X_p \qquad (۹-۱)$$

تحلیل عاملی بر دو نوع است که شامل تحلیل عاملی اکتشافی (Exploratory factor analysis) و تحلیل عاملی تأییدی (Confirmatory factor analysis) است.

در تحلیل عاملی تأییدی پیش فرض اساسی محقق آن است که هر عامل با زیر مجموعه خاصی از متغیرها ارتباط دارد، به عبارت دیگر محقق در این روش پیش فرض معینی دارد و بررسی می کند که آیا همبستگی های میان متغیرهای مشاهده شده با ساختار عوامل فرضی هماهنگ هستند یا خیر. اما در تحلیل عاملی اکتشافی همان طوری که از نام آن پیداست محقق درصدد کشف ساختار زیربنایی مجموعه ایی از متغیرهای مشاهده شده است که به ایجاد همبستگی هایی در میان این متغیرها می انجامد. بنابراین پیش فرض اولیه محقق آن است که هر متغیری ممکن است با هر عاملی ارتباط داشته باشد.

۹-۱. شاخص KMO و آزمون بارتلت

قبل از انجام تحلیل عاملی ابتدا باید دانست که آیا داده های موجود برای تحلیل عاملی مناسب اند یا خیر؟ از این رو با استفاده از شاخص KMO و آزمون بارتلت می توان از وجود این مسئله اطمینان حاصل کرد.

شاخص KMO: این شاخص نشان می دهد که آیا انسجام درونی داده ها مناسب است یا خیر. این شاخص در دامنه صفر تا یک قرار دارد. اگر مقدار شاخص از ۰/۶ کمتر باشد نشان می دهد که داده های گردآوری شده برای تحلیل عاملی چندان مناسب نیست. مقدار بالاتر از ۰/۷ نشان دهنده این مطلب است که انسجام درونی داده ها جهت تحلیل عاملی مناسب می باشد. این شاخص از طریق رابطه (۹-۲) محاسبه می گردد.

$$KMO = \frac{\sum\sum r_{ij}^2}{\sum\sum r_{ij}^2 + \sum\sum a_{ij}^2} \qquad (۹-۲)$$

در این رابطه r_{ij} ضریب همبستگی بین متغیرهای i و j، و a_{ij} ضریب همبستگی جزیی بین آنها است.

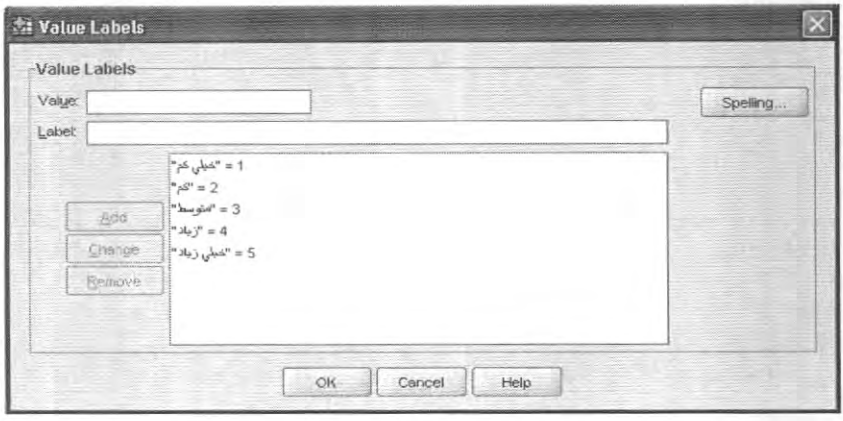
آزمون بارتلت: این آزمون بررسی می کند که آیا داده ها موجود برای شناسایی ساختار و کشف روابط بین متغیرها مناسب هستند یا خیر. در صورتی که sig آزمون بارتلت کوچک تر از ۵ درصد باشد تحلیل عاملی برای شناسایی ساختار مدل عاملی مناسب است.

نکته: تحلیل عاملی علاوه بر اینکه می تواند برای کاهش داده ها و شناسایی ساختار مورد استفاده قرار گیرد، برای سنجش اعتبار یک مقیاس یا شاخص نیز می تواند بکار گرفته شود.

۹-۲. اجرای تحلیل عاملی (اکتشافی)

مثال: محققی با هدف شناسایی نقش عوامل مؤثر در عدم استفاده از سیستم های آبیاری تحت فشار در میان اغلب کشاورزان مناطق روستایی با مراجعه به محل و ارزیابی مشارکتی روستایی، فهرستی از متغیرهای تأثیر گذار را در این باره شناسایی می کند. سپس از طریق توزیع پرسشنامه بین کشاورزان، داده های مربوط به هر متغیر براساس طیف لیکرت گردآوری می شود. وی برای اینکه مجموع این متغیرها را در قالب عواملی معنادار خلاصه کند تا از این طریق بتواند توازنی میان سرمایه و کار میان بخش ها برقرار کند و نهایتاً بتواند عواملی که بیشترین تأثیر را در روند کند بکارگیری این فناوری ها توسط کشاورزان داشته شناسایی کند از تحلیل عاملی استفاده می کند. که در ادامه مراحل آن تشریح می شود.

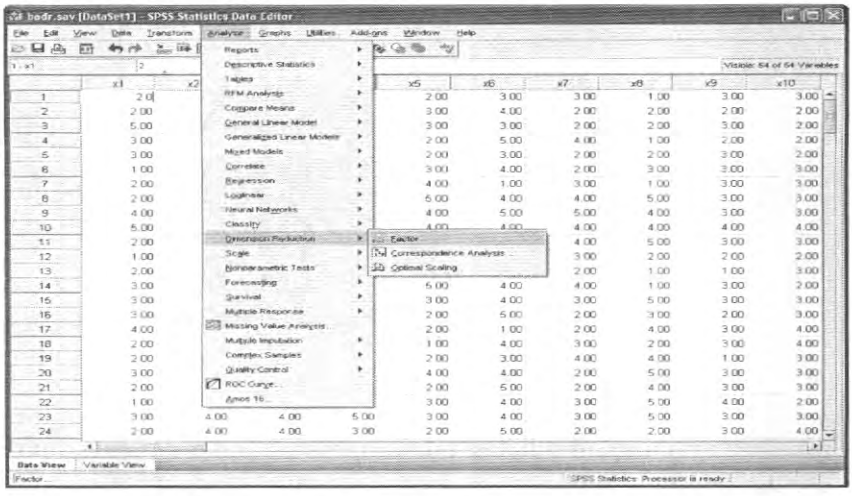
ابتدا متغیرهای موجود را طبق آنچه که در فصول قبل توضیح داده شد در پنجره Variable View تعریف کنید. در پنجره Variable View روی خانه Value کلیک کرده و طیف لیکرت را همانند شکل (۹-۱) ارزش گذاری کنید. این کار را روی یک خانه در ستون Value انجام دهید و سپس برای ارزش گذاری سایر متغیرها، روی ستون ارزش گذاری شده کلیک راست کرده و گزینه کپی را انتخاب کنید و روی سایر خانه های ستون Value آن را Paste کنید.



شکل (۹-۱) پنجره Value Labels

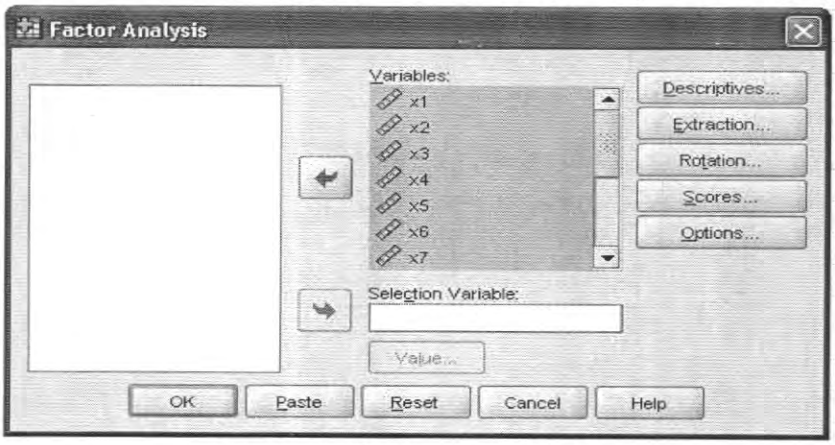
مسیر شکل (۹-۲) را اجرا کنید تا کادر Factor Analysis باز شود.

Analyze ⟶ Data Reduction ⟶ Factor



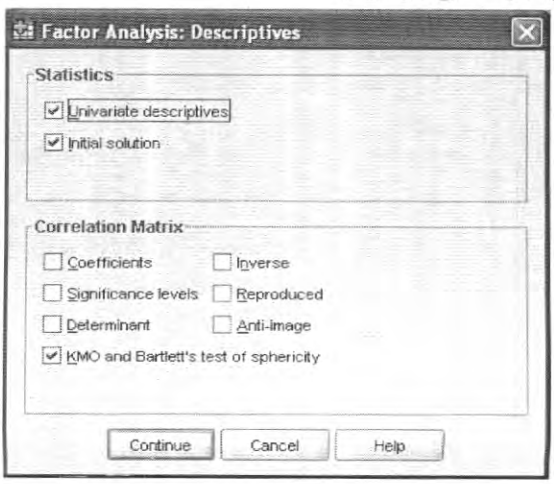
شکل (۹-۲) پنجره SPSS Statistics Data Editor

پس از انتخاب این منو پنجره Factor Analysis باز می گردد. همانند شکل (۹-۳) متغیرهای X_1 تا X_{14} را به کادر Variable انتقال دهید.



شکل (۹-۳) پنجره Factor Analysis

روی دکمه Descriptives کلیک کنید تا پنجره مربوطه باز شود. نرم افزار Spss به صورت پیش فرض گزینه Initial Solution را جهت ارائه اشتراک اولیه، مقادیر ویژه و درصد تبیین شده فعال کرده است آن را تغییر ندهید و در ادامه گزینه‌های Univariate Descriptives را جهت ارائه آمار توصیفی و KMO and Bartlett's Test of Sphericity را به منظور محاسبه شاخص KMO و آزمون بارتلت در قسمت Correlation Matrix همانند شکل (۹-۴) انتخاب کنید. سپس گزینه Continue را کلیک کنید تا به پنجره اصلی باز گردید.

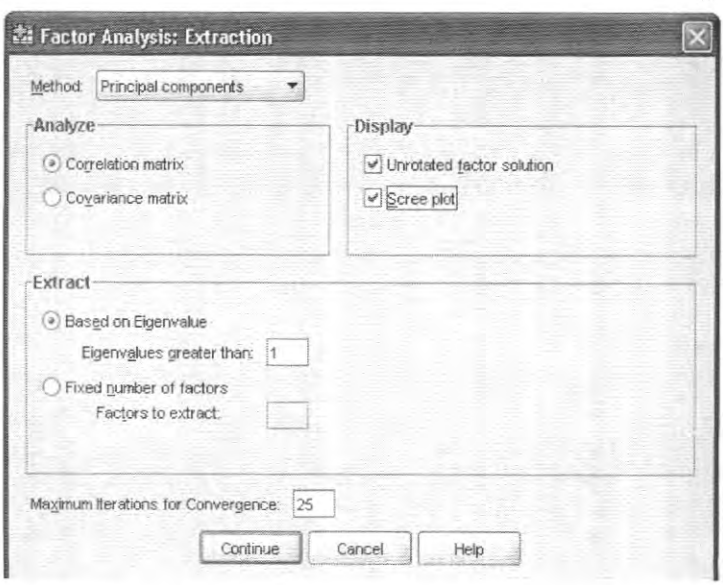


شکل (۹-۴) پنجره Factor Analysis Descriptive

سپس روی دکمه Extraction کلیک کنید تا این پنجره باز گردد. در قسمت Method شمار زیادی از روش های استخراجی وجود دارد که می توانید با کلیک روی فلش رو به پایین مقابل Method آنها را ملاحظه کنید و بر حسب نوع تحلیل روش مناسبی را انتخاب کنید. در صورتی که هدف اجرای تحلیل عاملی اکتشافی باشد (مانند مسئله ارائه شده) گزینه principal components (مؤلفه های اصلی) که تکنیک کاهش داده ها است، روشی مناسب است. البته نرم افزار به صورت پیش فرض آن را انتخاب می کند، بنابراین آن را تغییر ندهید.

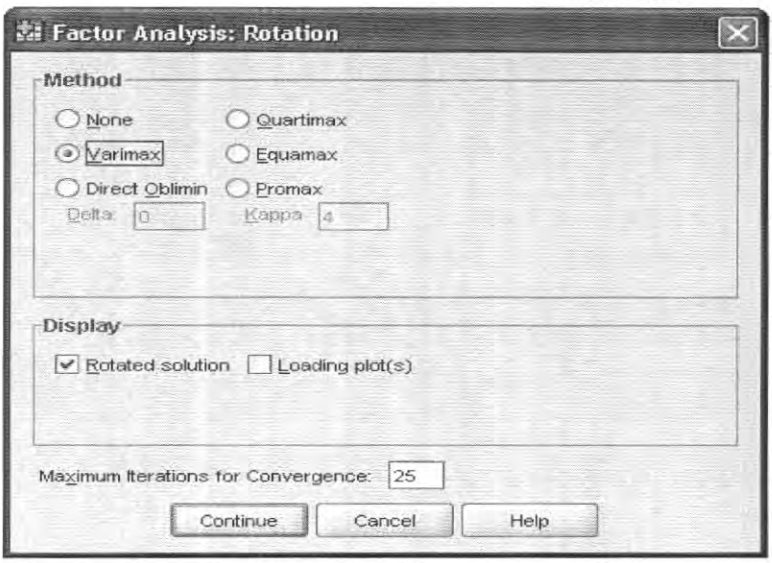
نکته: برای انجام تحلیل عاملی تأییدی باید پیش فرض Method را به Maximum likelihood تغییر داد و سپس تعداد عوامل (Factors of number) را نیز مشخص کرد.

برای نمایش نمودار مقادیر ویژه می توانید گزینه Scree plot را نیز انتخاب کنید سپس برای بازگشت به پنجره اصلی، Continue را کلیک کنید. بنابراین در پنجره Extraction علاوه بر پیش فرض های نرم افزار فقط گزینه Scree plot را برای نمایش نموداری همه عواملی که برای استخراج مورد توجه قرار می گیرند، فعال کنید شکل (۹-۵).



شکل (۹-۵) پنجره Factor Analysis Extraction

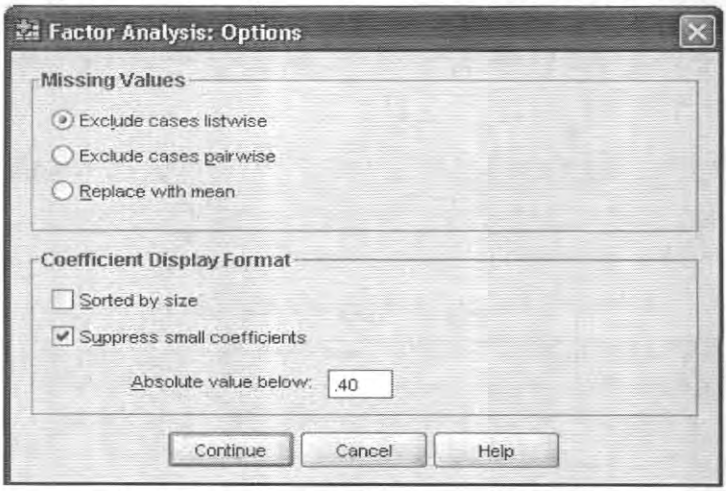
پس از بازگشت به پنجره اصلی برای انجام یک چرخش روی دکمه Rotation کلیک کنید تا این پنجره برای شما باز شود. هدف از اجرای چرخش این است که تصویر روشنی از ساختار عامل ها ارائه دهد. از اینرو لازم است که از یک روش چرخش استفاده کنیم. برنامه SPSS پنج گزینه را برای چرخش ارائه می دهد. رایج ترین روش در این پنجره Varimax است. از ویژگی های این روش برقراری استقلال در میان عامل ها می باشد به گونه ایی که تغییرات را در میان عامل ها به شکل یکنواخت توزیع می کند و هر عامل نسبت تقریباً یکسانی از تغییرات را توضیح می دهد. برای همین منظور گزینه Varimax را همانند شکل (۹-۶) فعال کنید. سپس گزینه Continue را کلیک کنید تا به پنجره اصلی باز گردید.



شکل (۹-۶) پنجره Factor Analysis Rotation

در ادامه دکمه Options را کلیک کنید. در این پنجره برای تعیین بارهای عوامل گزینه Suppress small coefficients را انتخاب کنید. پنهان کردن مقادیر زیر ۰/۴۰ منطقی به نظر می رسد، زیرا این نوعی معیار استاندارد است که توسط پژوهشگران برای تعیین بارهای عوامل مهم به کار می رود. بنابراین پس از انتخاب گزینه Suppress small coefficients، پیش فرض نرم افزار را همانند شکل (۹-۷) از ۰/۱۰ به ۰/۴۰

تغییر دهید سپس برای بازگشت به پنجره اصلی، Continue را کلیک کنید. پس از انجام تمامی این مراحل بر روی Ok کلیک کنید تا نتایج برای شما اجرا گردد.



شکل (۹-۷) پنجره Factor Analysis Option

نتیجه شامل هشت خروجی می باشد. خروجی اول ارائه دهنده آمار توصیفی شامل میانگین، انحراف معیار و تعداد آزمودنی هایی است که مورد تحلیل قرار گرفته اند.

خروجی دوم (جدول ۹-۱) مقدار شاخص KMO و مقدار آماره آزمون بارتلت، درجه آزادی و sig آزمون را نشان می دهد. با توجه به این که مقدار شاخص KMO برابر با ۰/۸۸۱ است، انسجام درونی داده ها مناسب می باشد. همچنین مقدار sig آزمون بارتلت کوچک تر از ۵ درصد است که نشان می دهد تحلیل عاملی برای شناسایی ساختار مدل عاملی مناسب است.

جدول (۹-۱) KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.881
Bartlett's Test of Sphericity	Approx. Chi-Square	1524.400
	df	91
	Sig.	.000

خروجی سوم (جدول ۹-۲) به ترتیب اشتراک اولیه (Initial) و اشتراک استخراجی (Extraction) را نشان می دهد. به دلیل این که ستون اشتراک اولیه، اشتراک ها را

قبل از استخراج عامل ها بیان می کند، تمامی اشتراک های اولیه برابر یک می باشد. اگر متغیر خاصی ضریب اشتراک پایینی داشته باشد، به این معناست که این عوامل استخراج شده نمی توانند واریانس بیشتری را در آن متغیر تبیین کنند. این متغیرها ممکن است از تحلیل کنار گذاشته شوند. جدول (۹-۲) نشان می دهد که مقادیر اشتراک استخراجی متغیرها بالا بوده و به نظر نمی رسد هیچکدام از متغیرهای مشاهده شده از تحلیل کنار گذاشته شوند البته برای اینکه نسبت به این مسئله اطمینان حاصل کنیم باید سایر خروجی ها را در ادامه مشاهده کنیم.

جدول (۹-۲) Communalities

	Initial	Extraction
توپوگرافی(پستی، بلندی) زمین	1.000	.861
شرایط جوی(بادخیزی شدید منطقه و...)	1.000	.857
جنس و بافت خاک(میزان نفوذپذیری آب و ...)	1.000	.841
کمبود زمین	1.000	.587
عدم توان برای سرمایه گذاری	1.000	.535
عدم دسترسی به اطلاعات و نتایج تحقیقات کشاورزی	1.000	.627
مالکیت آب	1.000	.659
تقطیع و پراکندگی اراضی و به صرفه نبودن	1.000	.683
برگزار نشدن دوره های آموزشی در زمینه احداث سیستم ها	1.000	.522
کمبود اعتبار و وام جهت راه اندازی سیستم ها	1.000	.657
کمبود مروجان آگاه از مسائل سیستم ها	1.000	.644
عدم آگاهی کافی از سیستم های آبیاری تحت فشار	1.000	.688
بیمه نبودن سیستم های آبیاری تحت فشار	1.000	.636
هزینه زیاد سیستم های آبیاری تحت فشار	1.000	.479

Extraction Method: Principal Component Analysis.

خروجی چهارم (جدول ۹-۳) حاوی سه قسمت است، قسمت اول Initial (Eigenvalues) مربوط به مقادیر ویژه و تعیین کننده عامل هاست. بدین ترتیب تنها عواملی که مقادیر ویژه بزرگتر از یک دارند در تحلیل باقی می مانند. عواملی که از تحلیل خارج می شوند عواملی هستند که حضور آنها باعث تبیین بیشتر واریانس نمی گردد.

قسمت دوم Extraction Sums of Squared Loadings است که مربوط به مقدار ویژه عوامل استخراجی بدون چرخش است و قسمت سوم Rotation Sums of Squared Loadings که نشان از مقدار ویژه عوامل استخراجی با چرخش است.

همچنان که مشاهده می‌کنید، سه عامل دارای مقادیر ویژه بزرگتر از یک هستند که در تحلیل باقی می مانند. این سه عامل ۶۶/۲۶ درصد از واریانس متغیرها را تفسیر و توضیح می‌دهند.

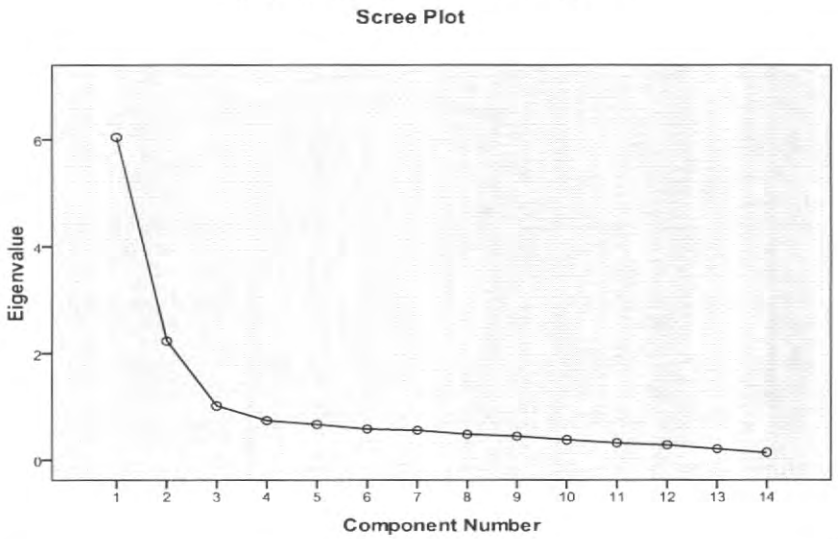
جدول (۹-۳) Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.040	43.140	43.140	6.040	43.140	43.140	3.161	22.578	22.578
2	2.228	15.916	59.056	2.228	15.916	59.056	3.120	22.283	44.862
3	1.009	7.208	66.265	1.009	7.208	66.265	2.996	21.403	66.265
4	.732	5.225	71.490						
5	.661	4.720	76.210						
6	.574	4.097	80.307						
7	.547	3.909	84.216						
8	.470	3.357	87.573						
9	.434	3.103	90.676						
10	.371	2.647	93.323						
11	.315	2.248	95.571						
12	.281	2.010	97.581						
13	.204	1.460	99.041						
14	.134	.959	100.000						

Extraction Method: Principal Component Analysis.

خروجی پنجم (شکل ۹-۸) نمایش نموداری همه عواملی است که برای استخراج مورد توجه قرار گرفته اند. همانطوریکه این نمودار نشان می دهد تنها سه عامل اول دارای مقادیر ویژه بزرگتر از یک هستند و سایر عوامل مقدار ویژه کمتر از یک دارند.

شکل (۹-۸) نمایش نموداری عوامل مهم



خروجی ششم (جدول ۹-۴) ماتریس اجزا را نشان می دهد که شامل بارهای عاملی هر یک از متغیرها در سه عامل باقی مانده است.

جدول (۹-۴) Component Matrix

	Component		
	1	2	3
توپوگرافی(پستی، بلندی) زمین	.552	.739	
شرایط جوی(بادخیزی شدید منطقه و...)	.528	.757	
جنس و بافت خاک(میزان نفوذپذیری آب و ...)	.564	.720	
کمبود زمین	.648		
عدم توان برای سرمایه گذاری	.697		
عدم دسترسی به اطلاعات و نتایج تحقیقات کشاورزی	.708		
مالکیت آب	.761		
تقطیع و پراکندگی اراضی و به صرفه نبودن	.622		.481
برگزار نشدن دوره های آموزشی در زمینه احداث سیستم ها	.700		
کمبود اعتبار و وام جهت راه اندازی سیستم ها	.615		-.464

کمبود مروجان آگاه از مسائل سیستم ها	.696		
عدم آگاهی کافی از سیستم های آبیاری تحت فشار	.713		-.402
بیمه نبودن سیستم های آبیاری تحت فشار	.689		
هزینه زیاد سیستم های آبیاری تحت فشار	.654		

تفسیر این بارگذاری بدون چرخش دشوار است. بنابراین عامل‌ها را می چرخانیم. خروجی هفتم (جدول ۹-۵) جدولی به نام Rotated Component Matrix را نشان می دهد که دارای بارهای چرخش یافته است.

جدول (۹-۵) Rotated Component Matrix

	Component		
	1	2	3
توپوگرافی(پستی، بلندی) زمین			.914
شرایط جوی(بادخیزی شدید منطقه و...)			.917
جنس و بافت خاک(میزان نفوذپذیری آب و ...)			.900
کمبود زمین		.490	.576
عدم توان برای سرمایه گذاری	.504	.517	
عدم دسترسی به اطلاعات و نتایج تحقیقات کشاورزی	.720		
مالکیت آب		.697	
تقطیع و پراکندگی اراضی و به صرفه نبودن	.814		
برگزار نشدن دوره های آموزشی در زمینه احداث سیستم ها	.422	.559	
کمبود اعتبار و وام جهت راه اندازی سیستم ها		.794	
کمبود مروجان آگاه از مسائل سیستم ها	.724		
عدم آگاهی کافی از سیستم های آبیاری تحت فشار		.780	
بیمه نبودن سیستم های آبیاری تحت فشار	.742		
هزینه زیاد سیستم های آبیاری تحت فشار	.448	.519	

خروجی پایانی جدولی با عنوان Component Transformation Matrix است که از آن برای تفسیر نتایج استفاده نمی شود و تنها برخی اطلاعات کلی به ما می دهد.

در نهایت با توجه به تحلیل عاملی صورت گرفته شده روی ۱۴ متغیر مؤثر در عدم استفاده از سیستم های آبیاری تحت فشار در مناطق روستایی، سه عامل بعنوان عامل

های اصلی شناسایی شدند. این سه عامل را می توان با توجه به ادبیات موضوع نامگذاری کرد. (جدول ۹-۶) نام عامل ها، متغیرهای مربوط به هر عامل و بارهای عاملی را نشان می دهد.

جدول (۹-۶) نام عامل ها، متغیرهای مربوط به هر عامل و بارهای عاملی

نام عامل	متغیر	بار عاملی
عامل ساختاری	تقطیع و پراکندگی اراضی و به صرفه نبودن سیستم های آبیاری تحت فشار	۰/۸۱۴
	بیمه نبودن سیستم های آبیاری تحت فشار	۰/۷۴۱
	کمبود مروجان آگاه از مسائل سیستم های آبیاری تحت فشار	۰/۷۲۴
	عدم دسترسی به اطلاعات و نتایج تحقیقات کشاورزی	۰/۷۱۹
عامل اقتصادی- اجتماعی	کمبود اعتبار و وام (نسبت به هزینه سیستم ها)	۰/۷۹۳
	عدم آگاهی کافی از سیستم های آبیاری تحت فشار	۰/۷۸۰
	مالکیت آب	۰/۶۹۷
	برگزار نشدن دوره های آموزشی در زمینه احداث سیستم ها	۰/۵۵۹
	هزینه زیاد سیستم های آبیاری تحت فشار	۰/۵۱۸
	عدم توان برای سرمایه گذاری(نداشتن وثیقه جهت دریافت تسهیلات)	۰/۵۱۶
عامل طبیعی	شرایط جوی(بادخیزی شدید منطقه و...)	۰/۹۱۷
	توپوگرافی(پستی، بلندی) زمین	۰/۹۱۴
	جنس و بافت خاک(میزان نفوذپذیری آب و...)	۰/۹۰۰
	کمبود زمین	۰/۵۷۶

فصل دهم

تحلیل خوشه‌ای

تحلیل خوشه‌ای یکی از روش های پرکاربرد در بسیاری از شاخه های علمی است. این روش به محقق امکان می دهد تا بر مبنای همگنی موجود در بین موارد یا موضوعات مورد مطالعه، آنها را به شیوه ایی مناسب طبقه بندی کرده و سپس مورد تفسیر و تحلیل قرار دهد. به عبارت دیگر تحلیل خوشه ای تکنیکی برای سطح بندی مناطق، شهرها، روستاها و... براساس ویژگی های مختلف موجود در هر مکان است. بطوری که در این سطح بندی مکان های واقع در یک سطح شباهت زیادی با همدیگر داشته اما با مکان های سطوح دیگر تفاوت قابل توجهی دارند. سطح بندی مکان های همگن در این روش به شیوه های مختلفی صورت می گیرد. مهم ترین روشی که در تحلیل خوشه ای برای تعیین شباهت های بین مکان های مختلف مطرح می شود، تعیین ضریب همبستگی بین هر جفت از مکان ها یا متغیرها است. اندازه گیری فاصله به ویژه فاصله اقلیدیسی از دیگر روش‌هایی است که در سطح وسیع برای تعیین شباهت های بین مکان ها مورد استفاده قرار می گیرد.

۱۰-۱. الگوریتم های عمومی مورد استفاده در روش تحلیل خوشه ای

۱- روش خوشه ای سلسله مراتبی

۲- روش خوشه ای غیر سلسله مراتبی

در این کتاب سعی شده روش خوشه ای سلسله مراتبی که کاربرد بیشتری در مطالعات اجتماعی و اقتصادی دارد مورد بحث قرار گیرد.

روش خوشه ای سلسله مراتبی دارای یک ساختار درختی است. این روش خود به دو شیوه مختلف انجام می گیرد که عبارتند از: روش تراکمی و روش تفکیک پذیری.

در روش تراکمی هر مکان با خوشه ای خاص آغاز می شود سپس دو مکان با هم ترکیب شده و خوشه ای جدید می سازند. در بعضی موارد نیز ممکن است دو خوشه دو مکانه با هم ترکیب شوند و خوشه ای جدید به وجود آورند. نهایتاً تمام مکان ها با هم ترکیب شده و به صورت ساختاری درختی، یک خوشه بزرگ را پدید می آورند. در روش خوشه‌ای سلسله مراتبی تراکمی پنج روش عمده برای تشکیل خوشه های تراکمی وجود دارد که عبارتند از:

۱- پیوند تکی
۲- پیوند کامل
۳- پیوند متوسط
۴- روش وارد
۵- روش مرکز ثقل

این روش ها از نظر نحوه محاسبه فاصله بین خوشه ها با هم متفاوتند و انتخاب هر کدام از این روش ها بستگی به نوع داده های موجود و هدف برنامه ریزی دارد.

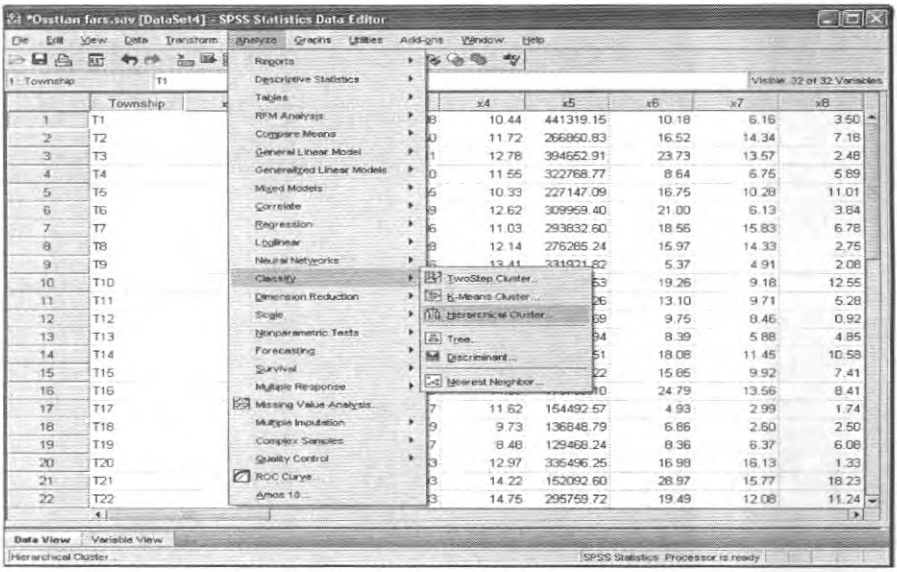
در صورتی که فرایند خوشه ای کردن در جهت عکس روش تراکمی انجام گیرد، آنگاه از روش تفکیک پذیری استفاده می شود. بعبارت دقیق تر در روش خوشه ای سلسله مراتبی تفکیکی، روند خوشه ای کردن از یک خوشه بزرگ که تمام مکان ها را پوشش می دهد آغاز می گردد و در هر مرحله مکان هایی که از همگنی کمتری برخوردارند از خوشه اولیه جدا شده و خوشه های جدید ایجاد می شود. این فرایند تا جایی ادامه می یابد تا مکان هایی که شباهت های بیشتری با یکدیگر دارند در یک خوشه خاص قرار گیرند.

۱۰-۲. اجرای تحلیل خوشه ای

مثال: محققی با هدف برنامه ریزی کشاورزی در شهرستان های استان فارس فهرستی از متغیرها را برای شناسایی و طبقه بندی شهرستان های این استان انتخاب می کند، وی پس از ترکیب متغیرها و شاخص سازی، ابتدا برای تعیین مناطق همگن از نظر توسعه کشاورزی از تحلیل خوشه ای استفاده می کند.

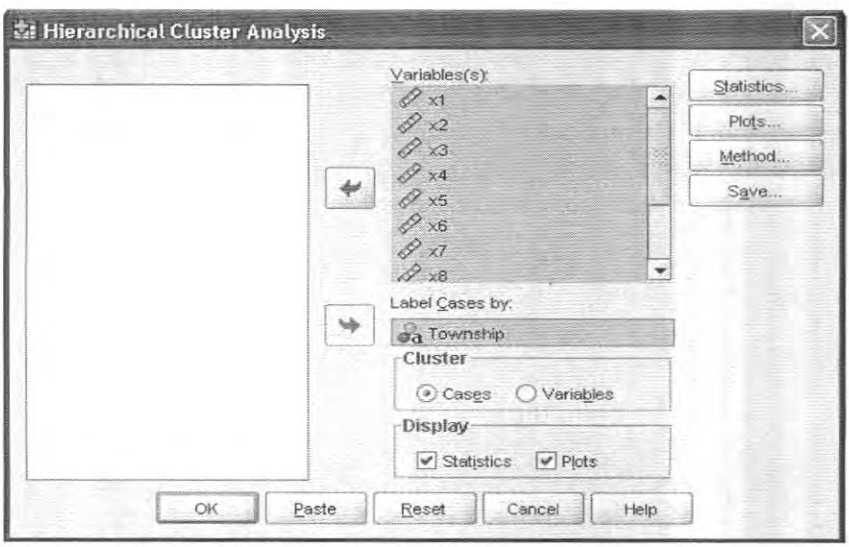
مسیر زیر را همانند شکل (۱۰-۱) اجرا کنید تا کادر Hierarchical Cluster باز شود. (نام شهرستان ها به صورت کد از T1 تا T22 نوشته شده است)

Analyze ⟶ Classify ⟶ Hierarchical Cluster



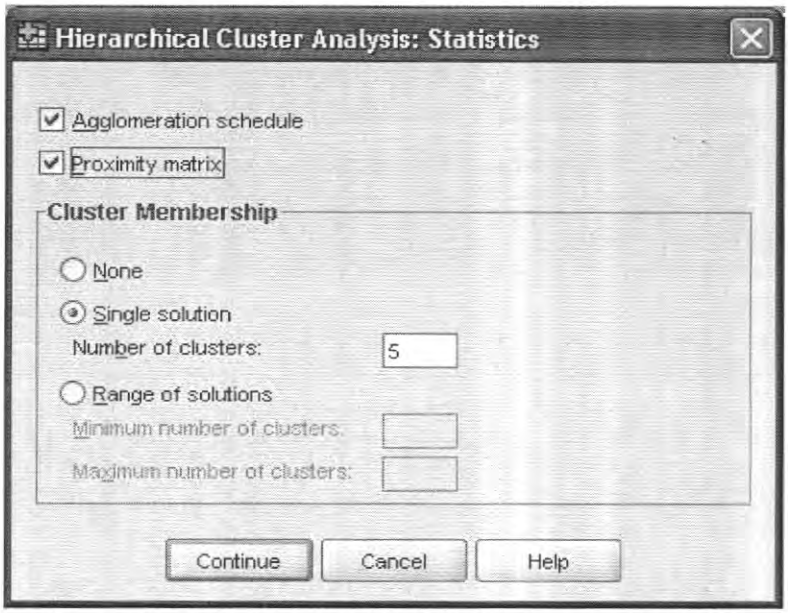
شکل (۱۰-۱) پنجره SPSS Statistics Data Editor

با اجرای این دستور، پنجره ای تحت عنوان Hierarchical Cluster Analysis ظاهر می شود. در این قسمت شاخص های توسعه کشاورزی را به کادر Variables منتقل کنید. سپس مطابق شکل (۱۰-۲) نام شهرستان ها را به کادر Label Cases by انتقال دهید.



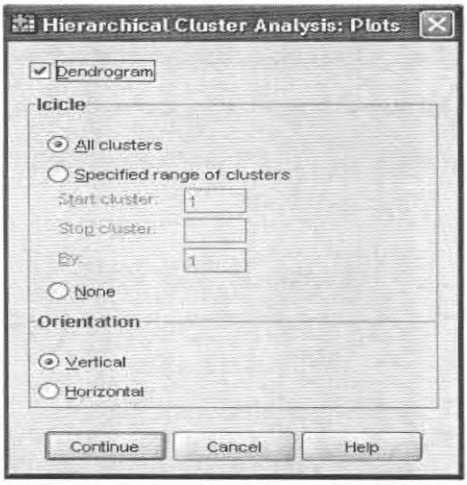
شکل (۱۰-۲) پنجره ای Hierarchical Cluster Analysis

پس از این مرحله ابتدا روی دکمه Statistics کلیک کنید تا پنجره مربوطه باز شود. در این قسمت علاوه بر پیش فرض نرم افزار که گزینه Agglomeration schedule را فعال کرده، می توان گزینه Proximaty matrix را برای نمایش ماتریس فواصل و همگنی بین مکان ها فعال کرد. همچنین برای تعیین تعداد خوشه ها و نمایش آن در بخش Output می توان در بخش Cluster Membership بر روی گزینه Single solution کلیک کرده و تعداد خوشه ها را همانند شکل (۱۰-۳) مشخص کرد. پس از انجام این مرحله گزینه Continue را کلیک کنید تا به پنجره اصلی باز گردید.



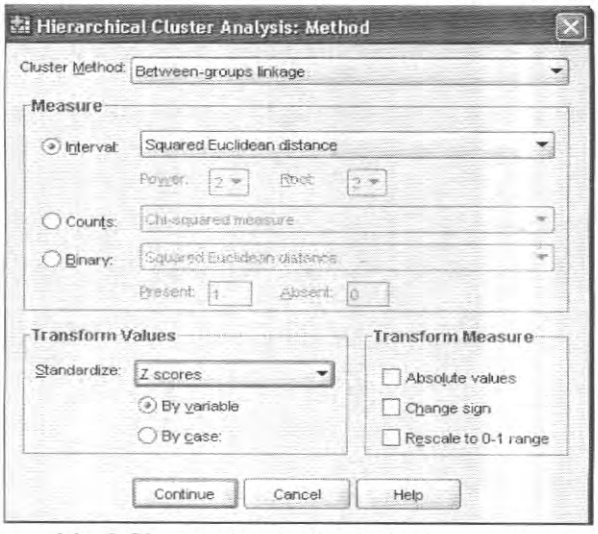
شکل (۱۰-۳) پنجره Hierarchical Cluster Analysis Statistics

سپس گزینه Plots را کلیک کنید تا پنجره مربوطه باز شود. همانند شکل (۱۰-۴) گزینه Dendrogram را برای نمایش نمودار درختی مکان های همگن فعال کنید. در قسمت Orientation نیز می توانید نوع نمایش تحلیل را به صورت های عمودی و افقی انتخاب کنید. با کلیک روی گزینه Continue به پنجره اصلی باز گردید.



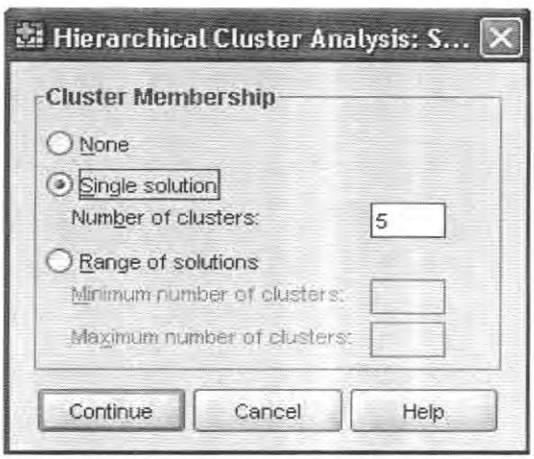
شکل (۱۰-۴) پنجره Hierarchical Cluster Analysis Plots

در ادامه گزینه Method را کلیک کنید. با توجه به این مطلب که در تحلیل خوشه ای باید از داده های استاندارد شده استفاده گردد لذا در صورتی که شاخص ها دارای مقیاس های متفاوت باشند باید ابتدا آنها را استاندارد سازی کرد. بنابراین در قسمت Transform Values بر روی کادر مستطیلی روبه روی Standardize کلیک کرده و گزینه Z scores را همانند شکل (۱۰-۵) انتخاب کنید. سپس گزینه Continue را کلیک کنید تا به پنجره اصلی باز گردید.



شکل (۱۰-۵) پنجره Hierarchical Cluster Analysis Method

نهایتاً با کلیک بر روی گزینه Save پنجره مربوط به آن باز می گردد در صورتی که بخواهیم نتایج حاصل از تحلیل خوشه ای در پنجره Data View نیز ظاهر گردد، با کلیک بر روی گزینه Single solution تعداد خوشه هایی که انتخاب کرده اید همانند شکل (۱۰-۶) دوباره انتخاب کنید. در این صورت نرم افزار به طور خودکار متغیری تحت عنوان CLU5-1 در پنجرهVariable view تعریف می کند. برای بازگشت به پنجره اصلی گزینه Continue را کلیک کنید. پس از انجام تمامی این مراحل بر روی Ok کلیک کنید تا نتایج برای شما اجرا گردد.



شکل (۱۰-۶) پنجره Hierarchical Cluster Analysis Single

نتایج خروجی شامل پنج جدول و یک نمودار درختی است. خروجی اول جدولی تحت عنوانCase Processing Summary است که اطلاعاتی دربارهٔ تعداد مکان های مورد تحلیل و داده های گمشده نشان می دهد. خروجی دوم جدولی با عنوان Proximaty matrix است. این جدول مربع فاصله اقلیدیسی هر شهرستان را نسبت به شهرستان بعدی بیان می کند. خروجی سوم (جدول ۱۰-۱) ترکیب تراکمی خوشه ها را تحت عنوانAgglomeration schedule نشان می دهد که از طریق پیوند متوسط انجام گرفته است. لازم به ذکر است که نرم افزار SPSS به صورت پیش فرض از روش پیوند متوسط استفاده می کند.

جدول (۱۰-۱) **Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	8	10.095	0	0	11
2	6	22	14.691	0	0	7
3	13	18	15.368	0	0	6
4	7	11	15.881	0	0	5
5	7	20	17.062	4	0	10
6	13	19	19.437	3	0	13
7	6	15	20.389	2	0	11
8	14	16	20.525	0	0	15
9	4	9	20.669	0	0	12
10	2	7	21.482	0	5	17
11	3	6	24.636	1	7	15
12	4	17	25.219	9	0	13
13	4	13	27.943	12	6	19
14	5	10	28.790	0	0	16
15	3	14	32.734	11	8	16
16	3	5	35.811	15	14	18
17	2	12	37.966	10	0	18
18	2	3	40.594	17	16	19
19	2	4	54.558	18	13	20
20	1	2	62.371	0	19	21
21	1	21	80.783	20	0	0

خروجی چهارم (جدول ۱۰-۲) اعضای هر خوشه را در جدولی مستقل تحت عنوان Cluster Membership نشان می دهد. در این جدول به خوبی طبقه بندی شهرستان‌ها در گروههای همگن نشان داده می شود.

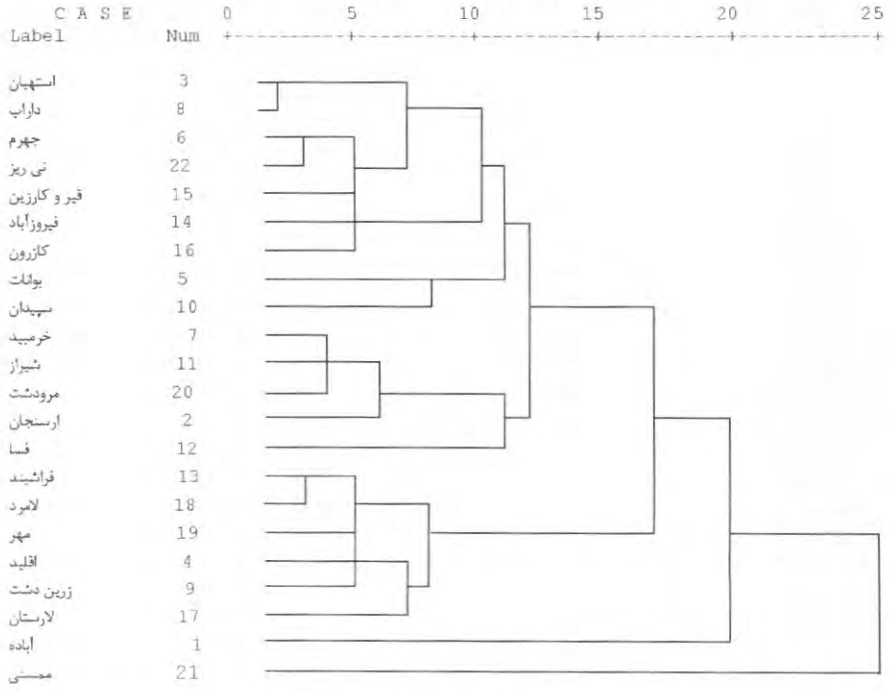
جدول (۱۰-۲) Cluster Membership

Case	5 Clusters
1: آباده	1
2: ارسنجان	2
3: استهبان	3
4: اقلید	4
5: بوانات	3
6: جهرم	3
7: خرمبید	2
8: داراب	3
9: زرین دشت	4
10: سپیدان	3
11: شیراز	2
12: فسا	2
13: فراشبند	4
14: فیروزآباد	3
15: قیر و کارزین	3
16: کازرون	3
17: لارستان	4
18: لامرد	4
19: مهر	4
20: مرودشت	2
21: ممسنی	5
22: نی ریز	3

خروجی پنجم قندیل عمودی است. این قندیل ها اطلاعاتی در مورد این که چگونه موارد در هر مرحله از تحلیل در خوشه ها ترکیب شده اند را نشان می دهد. خروجی پایانی (شکل ۱۰-۷)، نمودار درختی نحوه ترکیب مکان ها یا متغیرها به طور سلسله مراتبی می

باشد که شهرستان های استان فارس را از نظر توسعه کشاورزی در پنج سطح همسان دسته بندی کرده است.

شکل (۱۰-۷) خوشه بندی شهرستان های استان فارس براساس توسعه کشاورزی



نهایتاً با استفاده از تحلیل خوشه ای صورت گرفته، می توان سطوح توسعه کشاورزی را در شهرستان های استان فارس در پنج دسته همگن طبقه بندی کرد. البته در این چنین تحقیقاتی پیشنهاد می گردد که محقق ابتدا از تحلیل عاملی استفاده کند تا عامل های مهم استخراج گردند و سپس نتایج حاصل از عامل های ترکیبی را وارد تحلیل خوشه‌ای سازد تا بدین طریق شهرستان ها براساس عامل های ترکیبی مهم طبقه بندی گردند.

فصل یازدهم

روایی و پایایی

به طور کلی ابزار سنجش به دو طبقه تقسیم می شود: ۱) استاندارد و میزان شده، ۲) محقق ساخته. ابزار سنجش و آزمون های استاندارد و میزان شده معمولاً از روایی و پایایی مناسبی برخوردارند، از این رو محققان می توانند آنها را با اطمینان به کار گیرند؛ ولی ابزار محقق ساخته فاقد چنین اطمینانی هستند و محقق باید از روایی و پایایی آنها اطمینان حاصل کند.

۱۱-۱. روایی

روایی بیانگر آن است که آیا ابزار ما قادر به سنجش و اندازه گیری متغیر یا سازه ای که برای آن ساخته شده است می باشد یا خیر؟ در واقع روایی با این مسئله سر و کار دارد که یک ابزار اندازه گیری تا چه حد چیزی را اندازه گیری می کند که ما فکر می کنیم. بنابراین روایی، میزان انطباق بین تعریف مفهومی و نظری متغیر با تعریف عملیاتی آن است. بعبارت دیگر منظور از روایی این است که مقیاس و محتوای ابزار یا سئوالات مندرج در ابزار دقیقاً متغیرها و موضوع مورد مطالعه را بسنجد.

۱۱-۲. روش های تعیین روایی

الف) منطقی

این روش خود شامل دو روش زیر می باشد:

- ظاهری: یعنی پرسشنامه باید حداقل از نظر ظاهری دارای روایی باشد.
- محتوایی: در این حالت تعداد سئوالات و کیفیت آنها از نظر کارشناسان مورد ارزیابی قرار می گیرد.

ب) روایی وابسته به معیار

این روش نیز دارای دو زیر مجموعه می باشد:

• روایی همزمان: در این نوع روایی پژوهشگر همزمان با پاسخ دهی به سئوالات توسط پاسخ دهنده و مقایسه آن با رفتار فرد، روایی را بررسی می کند.

• روایی پیش بین: در این نوع روایی پژوهشگر بعد از گذشت مدت زمانی پی به روایی پرسشنامه می برد. به عبارت دیگر روایی پرسشنامه را با توجه به نتایجی که بعد از مدت زمان مشخص به دست می آید مقایسه می کنند.

ج) روایی سازه

مهمترین نوع روایی سازه تحلیل عاملی است. برای این منظور از دو آزمون KMO (شاخص کفایت نمونه گیری) و آزمون کرویت بارتلت استفاده می گردد. (مبحث فوق در فصل ۹ شرح داده شده است.)

۱۱-۳. پایایی

پایایی ابزار که از آن به اعتبار، دقت و اعتماد پذیری نیز تعبیر می شود، عبارت است از اینکه اگر یک وسیله اندازه گیری که برای سنجش متغیر و صفتی ساخته شده در شرایط مشابه در زمان یا مکان دیگر مورد استفاده قرار گیرد، نتایج مشابهی از آن حاصل شود. به عبارت دیگر ابزار پایا یا معتبر ابزاری است که از خاصیت تکرار پذیری و سنجش نتایج یکسان برخوردار باشد.

۱۱-۴. روش های تعیین پایایی

در برنامه SPSS روش های مختلفی برای ارزیابی پایایی ابزار سنجش وجود دارد که شامل گزینه های: alpha ، Split-half ، Guttmann ، Parallel ، Strict Parallel می باشد. هر کدام از این روش ها براساس همسانی درونی داده ها به ارزیابی پایایی ابزار می پردازند که در این مبحث تنها به توضیح دو مورد اول (alpha و Split-half) که نسبت به سایرین کاربردی تر و از اهمیت بیشتری برخوردارند خواهیم پرداخت.

alpha: یکی از مهمترین روش های تعیین پایایی روش آلفا نام دارد که به آن ضریب آلفای کرونباخ نیز گفته می‌شود. در این روش تعیین پایایی که واضع آن کرونباخ می‌باشد، پایایی ابزار سنجش براساس همسانی درون داده ها صورت می پذیرد. به عبارت

دیگر در این روش اجزا یا قسمت های آزمون برای سنجش ضریب پایایی آزمون به کار می روند. از ضریب آلفای کرونباخ می توان برای سئوالات دو ارزشی صفر و یک و همچنین گزینه های چند ارزشی (مانند طیف پنج گزینه ای لیکرت) استفاده کرد. کلی-ترین صورت تحلیل واریانس سئوال از طریق ضریب آلفای کرونباخ از طریق رابطه (۱۱-۱) به دست می آید.

$$r_a = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum S_i^2}{S_t^2}\right) \qquad (۱-۱۱)$$

که در آن، n برابر است با تعداد سئوالات آزمون،

S_i^2 واریانس سئوال I (واریانس داده های موجود در یک ستون ماتریس داده ها)

S_t^2 واریانس کل آزمون یا به عبارت دیگر واریانس ستون مجموع در ماتریس داده‌هاست.

در صورتی که همبستگی بین سئوالات کم باشد میزان پایایی نیز کم است.

Split-half: در این روش با تقسیم سئوالات به دو نیمه، ضریب همبستگی بین هر دو نیمه ی سئوالات محاسبه می گردد، ضریب همبستگی حاصل از دو نیمه آزمون، ضریب پایایی هر یک از دو نیمه خواهد بود. برای برآورد پایایی تمام آزمون تصحیح لازم با استفاده از فرمول «اسپیرمن – براون» اعمال می شود که از طریق رابطه (۱۱-۲) محاسبه می گردد.

$$r_{tt} = \frac{2r_{\frac{11}{12}}}{1+r_{\frac{11}{12}}} \qquad (۲-۱۱)$$

که در این رابطه ($r_{\frac{11}{12}}$) ضریب اعتبار بین هر دو نیمه آزمون و (r_{tt}) ضریب اعتبار تمام آزمون است. این روش نیز همانند ضریب آلفای کرونباخ متکی بر میزان همبستگی درونی سئوالات می باشد.

اگر روش دو نیمه کردن آزمون را با روش ضریب آلفا مقایسه کنیم، خواهیم دید که روش دو نیمه کردن، نوعی روش آلفاست. به عبارت دیگر ضریب آلفا معرف میانگین ضرایب پایایی حاصل از تمامی اجزای ممکن یک آزمون است، یعنی در این روش هر سئوال آزمون با تک تک سئوالات دیگر مقایسه می شود. در حالی که در روش دو نیمه کردن، یک نیمه آزمون با نیمه دیگرش مقایسه می گردد.

مثال: محققی برای بررسی میزان رضایت از زندگی شهری پرسشنامه ای را در چهار سئوال بین بیست نفر مورد سنجش قرار داده است. طیف پاسخ به سئوالات، طیف پنج سطحی می باشد که براساس خیلی زیاد ۵، زیاد ۴، متوسط ۳، کم ۲ و خیلی کم ۱ طراحی شده است. وی می خواهد بداند که آیا پرسشنامه ای که طراحی کرده است دارای روایی و پایایی مناسبی است یا خیر؟ چهار سئوال این پرسشنامه عبارتند از:

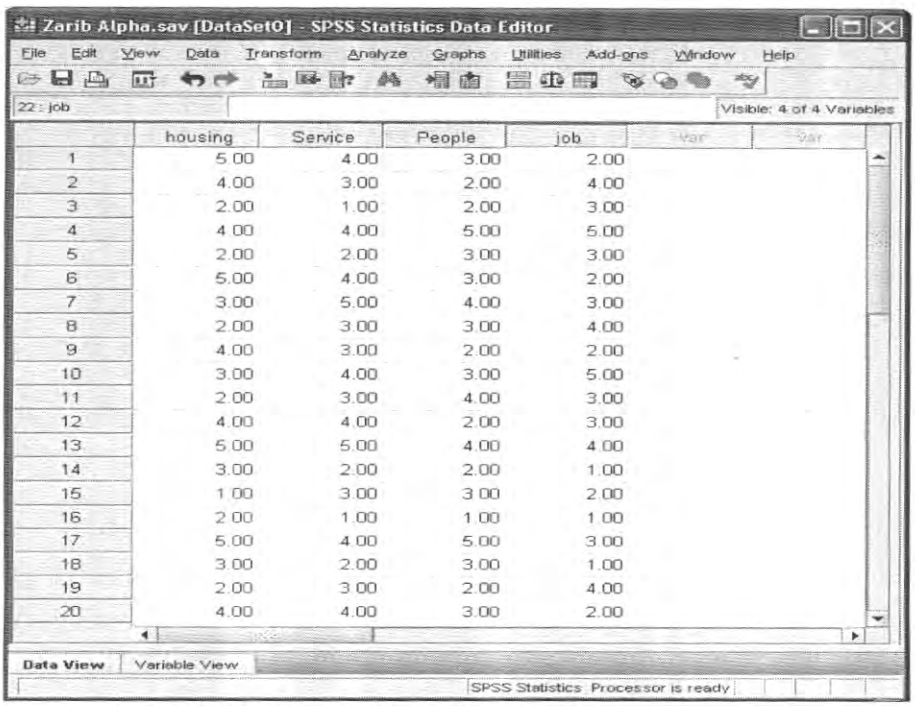
میزان رضایت از مسکن؟

میزان رضایت از خدمات؟

میزان رضایت از مردم؟

میزان رضایت از شغل؟

بنابراین ابتدا داده های گردآوری شده را همانند شکل (۱۱-۱) وارد نرم افزار SPSS کنید.


Zarib Alpha.sav [DataSet0] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

22 : job — Visible: 4 of 4 Variables

	housing	Service	People	job	var	var
1	5.00	4.00	3.00	2.00		
2	4.00	3.00	2.00	4.00		
3	2.00	1.00	2.00	3.00		
4	4.00	4.00	5.00	5.00		
5	2.00	2.00	3.00	3.00		
6	5.00	4.00	3.00	2.00		
7	3.00	5.00	4.00	3.00		
8	2.00	3.00	3.00	4.00		
9	4.00	3.00	2.00	2.00		
10	3.00	4.00	3.00	5.00		
11	2.00	3.00	4.00	3.00		
12	4.00	4.00	2.00	3.00		
13	5.00	5.00	4.00	4.00		
14	3.00	2.00	2.00	1.00		
15	1.00	3.00	3.00	2.00		
16	2.00	1.00	1.00	1.00		
17	5.00	4.00	5.00	3.00		
18	3.00	2.00	3.00	1.00		
19	2.00	3.00	2.00	4.00		
20	4.00	4.00	3.00	2.00		

Data View | Variable View

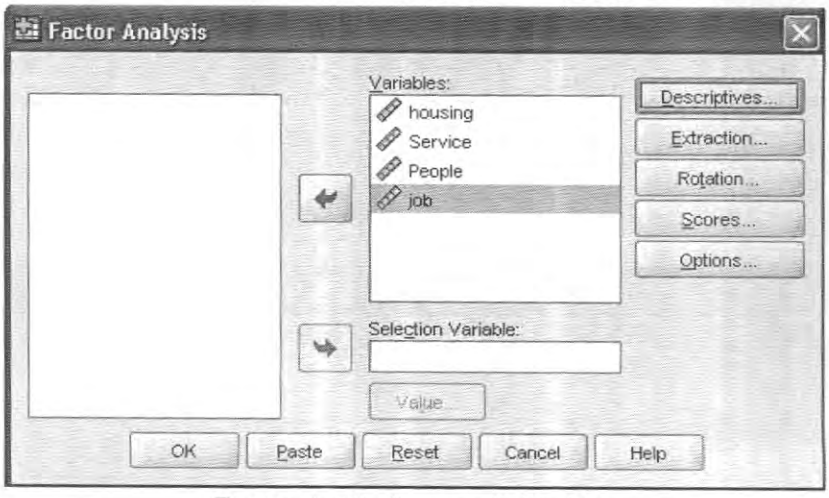
SPSS Statistics Processor is ready

شکل (۱۱-۱) پنجره SPSS Statistics Data Editor

سپس مسیر زیر را اجرا کنید تا پنجره Factor Analysis باز گردد.

Analyze ⟶ Data Reduction ⟶ Factor

همانند شکل (۱۱-۲) متغیرها را به کادر Variables منتقل کنید. در ادامه روی دکمه Descriptives کلیک کنید تا پنجره مربوطه باز شود. گزینه KMO and Bartlett's Test of Sphericity را به منظور محاسبه شاخص KMO و آزمون بارتلت در قسمت Correlation Matrix همانند آنچه در فصل ۹ گفته شد، انتخاب کنید. سپس برای بازگشت به پنجره اصلی، Continue را کلیک کنید. پس از انجام این مرحله بر روی Ok کلیک کنید تا نتایج برای شما اجرا گردد.



شکل (۱۱-۲) پنجره Factor Analysis

اولین خروجی، جدول KMO and Bartlett's Test می باشد که نشان می دهد ابزار ما قادر به سنجش و اندازه گیری متغیر و سازه ای که برای آن ساخته شده است می باشد و دارای روایی نسبتاً مناسبی است (جدول ۱۱-۱).

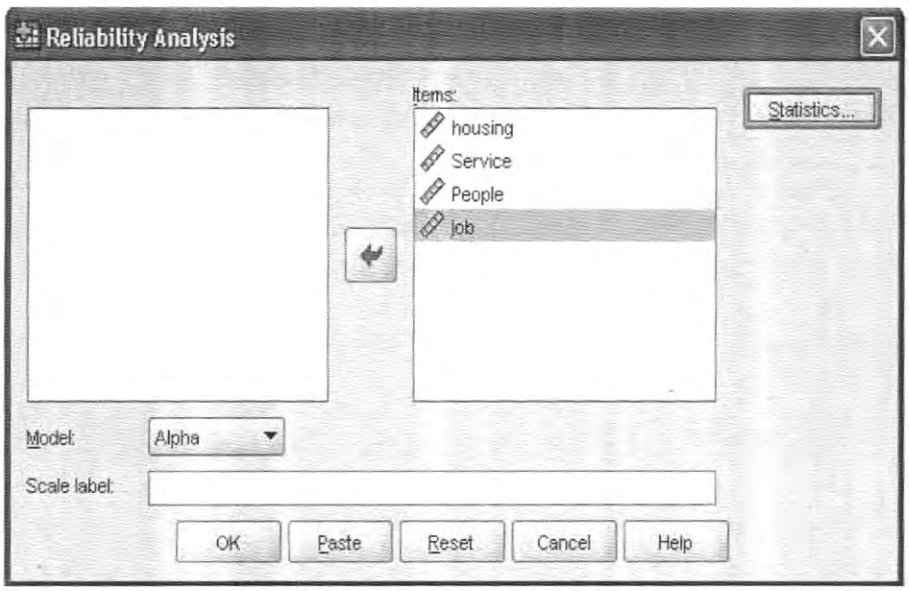
جدول (۱۱-۱) KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.602
Bartlett's Test of Sphericity	Approx. Chi-Square	21.957
	df	6
	Sig.	.001

در ادامه فرایند تحقیق برای اینکه به ارزیابی پایایی ابزار بپردازیم از کاربردی ترین روش سنجش پایایی یعنی ضریب آلفای کرونباخ استفاده می کنیم. بنابراین مسیر زیر را اجرا کنید تا پنجره مربوطه باز گردد.

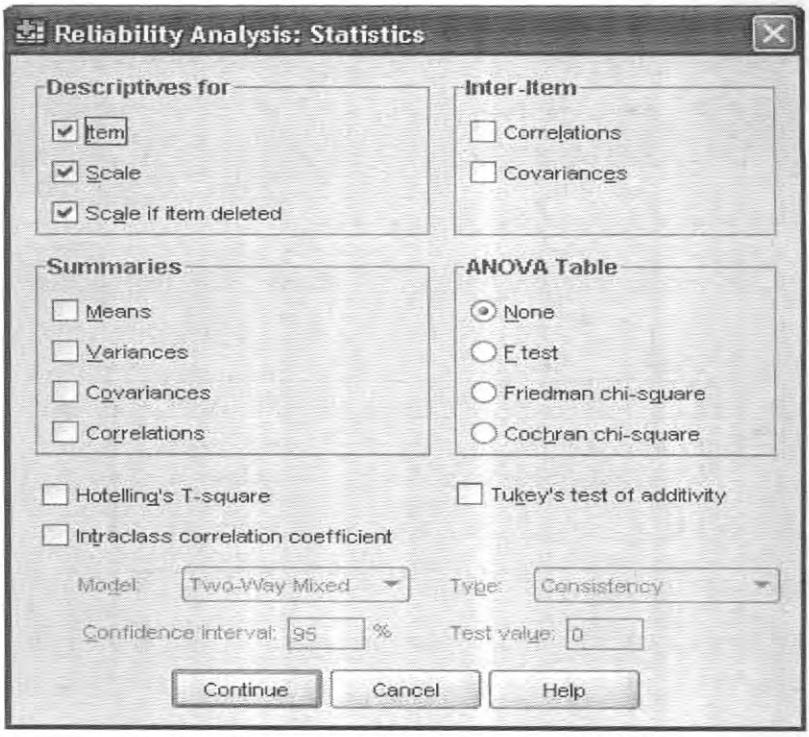
Analyze ⟶ Scale ⟶ Reliability Analysis

با اجرای این دستور پنجره ای تحت عنوان Reliability Analysis باز می گردد. سپس همانند شکل (۱۱-۳) متغیرها را به کادر Items انتقال دهید. در پایین کادر گزینه ای به نام Model وجود دارد که شما می توانید جهت سنجش پایایی از روش های مختلف موجود در کادر روبه روی آن استفاده کنید. اما چون هدف ما سنجش پایایی ابزار با استفاده از آلفای کرونباخ است بنابراین گزینه alpha را که نرم افزار به صورت پیش فرض انتخاب کرده است، تغییر ندهید.



شکل (۱۱-۳) پنجره Reliability Analysis

سپس بر روی دکمه Statistics کلیک کنید تا پنجره مربوطه باز گردد. همانند شکل (۱۱-۴) گزینه های Item ، Scale ، Scale if item deleted را از کادر Descriptives for فعال کنید. پس از انجام این مراحل، برای بازگشت به پنجره اصلی، Continue را کلیک کنید. سپس بر روی Ok کلیک کنید تا نتایج برای شما اجرا گردد.



شکل (۱۱-۴) پنجره Reliability Analysis Statistics

نتیجه شامل پنج خروجی می باشد، خروجی اول جدولی تحت عنوان Case Processing Summary می باشد که تعداد داده ها و درصد داده های مورد تحلیل قرار گرفته را نشان می دهد. جدول (۱۱-۲) نشان می دهد که تمامی داده ها در تحلیل وارد شده اند.

جدول (۱۱-۲) Case Processing Summary

		N	%
Cases	Valid	20	100.0
	Excluded[a]	0	.0
	Total	20	100.0

خروجی دوم، ضریب آلفای کرونباخ و تعداد متغیرها (سئوالات) را نشان می دهد. در صورتی که ضریب آلفا بیش از ۰/۷۰ باشد، می توانیم استدلال کنیم که آزمون از پایایی قابل قبولی برخوردار می باشد. نتایج حاصل از جدول (۱۱-۳) نشان می دهد که پایایی آزمون برابر با ۰/۷۲۷ به طور تقریبی ۰/۷۳ است. بنابراین آزمون از پایایی قابل قبولی برخوردار می باشد.

جدول (۱۱-۳) Reliability Statistics

Cronbach's Alpha	N of Items
.727	4

خروجی سوم (جدول ۱۱-۴) برخی از آمار توصیفی هر کدام از متغیرها را به تفکیک ارائه می دهد که شامل مواردی همچون: میانگین، انحراف معیار و تعداد نمونه های تحلیل شده می باشد.

جدول (۱۱-۴) Item Statistics

	Mean	Std. Deviation	N
housing	3.2500	1.25132	20
People	2.9500	1.05006	20
Service	3.2000	1.15166	20
job	2.8500	1.22582	20

خروجی چهارم جدولی با نام Item-Total Statistics می باشد. در ستون اول این جدول متغیرها (سئوالات)، در ستون دوم میانگین مجموع متغیرها در صورت حذف متغیر مورد نظر، در ستون سوم واریانس متغیرها در صورت حذف متغیر مورد نظر، در ستون چهارم میزان همبستگی متغیر با کل متغیرهای موجود و در ستون آخر ضریب آلفای کرونباخ در صورت حذف متغیر مورد نظر را نشان می دهد. به عبارت دقیق تر این

جدول (جدول ۱۱-۵) بررسی می کند که در صورت حذف هر کدام از این متغیرها چقدر آزمون دچار تغییر و دگرگونی می گردد. برای مثال در این آزمون با حذف متغیر job ، ضریب آلفا تا حدود ۰/۷۶ افزایش خواهد یافت. بنابراین در تحقیقاتی که آزمون از پایایی مناسبی برخوردار نیست، می توان با استفاده از اطلاعات این جدول، متغیرهایی (سئوالات) که با حذف آنها آزمون پایایی بیشتری به دست می آورد را در صورت کم اهمیت بودن از فرایند پرسشنامه اصلی خارج کرد.

جدول (۱۱-۵) Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
housing	9.0000	7.684	.410	.733
People	9.3000	7.589	.588	.631
Service	9.0500	6.261	.782	.501
job	9.4000	8.147	.349	.765

خروجی پایانی نیز پاره ای از اطلاعات کلی راجع به آمار توصیفی ارائه می دهد و از اهمیت چندانی برخوردار نمی باشد.

منابع و مآخذ

۱- آذر، عادل و مؤمنی، منصور(۱۳۷۲)، آمار و کاربرد آن در مدیریت(قسمت اول)، انتشارات سمت، تهران.

۲- افشانی، سید علیرضا، نوریان، مرتضی، نوریان، ابوذر، مهدوی زفرقندی، مهدی (۱۳۸۹)، مرجع کاربردی SPSS 17، انتشارات بیشه، تهران.

۳- بهبودیان، جواد (۱۳۷۸)، آمار ناپارامتری، انتشارات دانشگاه شیراز، شیراز.

۴- تریولا، ماریو (۱۳۷۷)، آمار کاربردی، ترجمه محمد صادق تهرانیان و ابوالقاسم بزرگ نیا، انتشارات جهاد دانشگاهی، مشهد.

۵- تقدیسی، احمد و بسحاق، محمدرضا(۱۳۸۹)، چالش های کشاورزی و تأثیر آنها بر جمعیت روستایی با تأکید بر تغییرات سطح زیر کشت (مطالعه موردی: دهستان سیلاخور شرقی- شهرستان ازنا)، فصل نامه پژوهش های روستایی، شماره ۲، صص ۱۶۱- ۱۳۷.

۶- تقوایی، مسعود، بسحاق، محمدرضا، سالاروند، اسماعیل(۱۳۸۹)، تحلیلی بر عوامل مؤثر در عدم استفاده از سیستم های آبیاری تحت فشار در روستاهای ایران(مطالعه موردی: مناطق روستایی شهرستان ازنا)، مجله مطالعات جغرافیایی مناطق خشک، سال اول، شماره ۲، صص ۲۳-۱۱ .

۷- جمشیدی، خدیجه (۱۳۸۶)، آمار و کاربرد آن در مدیریت(قسمت اول)، چاپ نوزدهم، انتشارات دانشگاه پیام نور، تهران.

۸- حافظ نیا، محمدرضا(۱۳۸۷)، مقدمه ای بر روش تحقیق در علوم انسانی، چاپ چهاردهم، انتشارات سمت، تهران.

۹- حکمت نیا، حسن و میرنجف موسوی (۱۳۸۲)، کاربرد مدل در برنامه ریزی شهری و ناحیه‌ای، انتشارات علم نوین، یزد.

۱۰- خراسانی زاده، علی (۱۳۷۵)، مقدمه ای بر کاربرد نرم افزار Spss در پژوهش آماری، انتشارات قائم، تهران.

۱۱- دلاور، علی (۱۳۹۰)، روش تحقیق در روانشناسی و علوم تربیتی، چاپ سی و چهارم، نشر ویرایش، تهران.

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

۱۳- رضایی، عبدالمجید و سلطانی، افشین (۱۳۷۷)، مقدمه ای بر تحلیل رگرسیون کاربردی، مرکز نشر دانشگاه صنعتی اصفهان، اصفهان.

۱۴- زرگر، محمود (۱۳۸۰)، راهنمای جامع Spss 10 ، انتشارات بهینه، تهران.

۱۵- ساعی، علی (۱۳۷۷)، آمار در علوم اجتماعی با کاربرد نرم افزار SPSS/PC+ در پژوهش‌های اجتماعی، مؤسسه نشر جهاد، تهران.

۱۶- صالحی صدقیانی، جمشید، ابراهیمی، ایرج (۱۳۸۶)، تحلیل آماری پیشرفته، چاپ دوم، نشر هستان، تهران.

۱۷- طالعی، علی رضا (۱۳۸۱)، مفاهیم و کاربردهای آمار، انتشارات دانشگاه تهران، تهران.

۱۸- طاهری، ابوالقاسم(۱۳۷۶)، روش تحقیق و مأخذشناسی، انتشارات دانشگاه پیام نور، تهران.

۱۹- عبدی، عباس (۱۳۷۰)، آزمون فرضیه در علوم رفتاری با ضمیمه برنامه آماری Spss، انتشارات حنانه، تهران.

۲۰- عسگری، محمد (۱۳۸۰)، مقدمه ای بر آمار توصیفی در علوم رفتاری، انتشارات نور علم، همدان.

۲۱- فرگوسن، جرج، اوتاکانه، یوشیو (۱۳۸۰)، تحلیل آماری در روانشناسی و علوم تربیتی، ترجمه علی دلاور و سیامک نقشبند، انتشارات ارسباران، تهران.

۲۲- کلانتری، خلیل (۱۳۸۹)، پردازش و تحلیل داده ها در تحقیقات اجتماعی- اقتصادی، چاپ چهارم، انتشارات فرهنگ صبا، تهران.

۲۳- کنوور، دبلیو، جی (۱۳۷۲)، آمار ناپارامتری کاربردی، ترجمه سید مقتدی هاشمی پرست، مرکز نشر دانشگاهی، تهران.

۲۴- گائور، آجای اس و گائور، سانجایا اس (۱۳۸۸)، روش های آماری برای تمرین و پژوهش با مدیریت SPSS، انتشارات فرهنگ سبز، تهران.

۲۵- گرین، جودیت و الیویرا، مانوئلادی (۱۳۸۰)، کاربرد آزمون های آماری در پژوهش های علوم رفتاری، ترجمه علی دلاور و مهرداد پرّمان، انتشارات ارسباران، تهران.

۲۶- مهدوی، مسعود (۱۳۷۳)، آمار و روشهای تجزیه و تحلیل داده ها در جغرافیا، نشر قومس، تهران.

۲۷- مهدوی، مسعود، طاهرخانی، مهدی (۱۳۸۳)، کاربرد آمار در جغرافیا، نشر قومس، تهران .

۲۸- مؤمنی، منصور و فعال قیومی، علی(۱۳۸۶)، تحلیل های آماری با استفاده از SPSS، انتشارات کتاب نو، تهران.

۲۹- نوروسیس، ماریا (۱۳۷۸)، کتاب آموزشی آنالیز آماری داده ها با SPSS 8، ترجمه اکبر فتوحی و فریبا اصغری، کانون نشر علوم، تهران.

۳۰- هومن، حیدرعلی (۱۳۷۸)، استنباط آماری در پژوهش رفتاری، انتشارات پارسا، تهران.

31- Bailey, K. D., (1988), Methods of Social Research, New York.

32- Berg, B. L., (1989) Qualitative Research Methods for the Social Science, Allyn and Bacon, Boston.

33- Bogdan,R. C., Biklen, S. K., (1992) Qualitative Research for Education: An Introduction to Theory and Methods, 2nd edn, Allyn and Bacon, Boston.

34- Bryman, A., Cramer, D., (1999) Quantitative data analysis with SPSS release and for windows, Routledge and Newyork.

35- Cattell, R. B., (1978) The Scientific Use of Factor Analysis in Behavioral Sciences, Plenum, New York.

36- Darrn, G., Paul, M., (2001) SPSS for windows step by step, Third edition, USA.

37- Dennis, C., (1973) The Essrntials of Factor Analysis, Holt Rinehart and Winston, New York.

38- Draper, N., Harry, S., (1986) Applied Regression Analysis, Inc, New York.

39- Ebdon, D., (1985) Statistics in Geography, Billing and Sons Ltd, Oxford, Uk.

40- Fruchter, B., (1964) Intoduction to Factor Analysis, Princeton, Inc, New York.

41- Ghosh, B. N., (1982) Scientific Methods and Social Research, Sterling Publishers, Pvt. Ltd, New York.

42- Gnanadesikan, M., Scheaffer, R. L., Witmer, J. R., (1997) An activity based statistics course, Journal of Statistics Education, Vol. 5, No. 2, pp. 118-129 .

43- Gopal, M. H., (1984) An Introduction to Research Procedure in Social Sciences, Asia Publishing House, Bombay.

44- Kinnear, Paul. R., (2000) SPSS for Windows made simple release 10, Psychology press ltd.

45- Kothari, C. R., (1992) Quantitative Techniques, Pashupati Printers, Delhi.

46- Mahmood, A., (1993) Statistical Methods in Geographical Studies, Rajesh Publications, New Delhi.

47- Marascuilo, L. A., Sweeney, M., (1977) Nonparametric and Distribution- free Methods for the Social Sciences, Monterey, Canada.

48- Maxwell, A. E., (1987) Analyzing Qualitative Data, John Wiley and Sons, New York.

49- Mendenhall, W., (1989) Statistics for Management and Economics, PWS-KENT Publishing Company, USA.

50- Morrison, D., (1987) Measurment Problems in Cluster Analysis, Management Science, Vol. 4, pp. 39-48.

51- Ott, R. L., (1993) An Introduction to Statistical Methods and Data Analysis, Duxbury, California.

52- Shaw, G., Wheeler, D., (1994) Statistical Techniques in Geographical Analysis, Halsted Press, New York.

53- Siegel, S., (1986) Nonparametric Statistics for the Behavioral Sciences, New York.

54- Snedecor, G. W., Cochran, W.G., (1980), Statistical Methods, 7 th ed, The Iova state university, Press Ames.

55- Spiegel, M., (1972) Theory and Problems of Statistics, McGraw-Hill Book Company, U.K.

56- Sweet, A. S., (1999) Data analysis with SPSS, Allyn & Bacon, USA.

57- Taylor, B., (2006) Research Methodology, Prentice-Hall, New Delhi.

58- Tryon, R. C., Bailey, D. E., (1980) Cluster Analysis, McGraw-Hill, New York.

پژوهشگاه متخصص
بحران، حادثه، طبیعی و پدافند غیر عامل
شاخص پیشرو

مدیریت بحران

The Test of Research Assumptions by SPSS

Dr. Masoud Taghvaei
Mohammad Reza Bosshaq
Smail Salarvand

ISBN:978-600-93834-6-7