

آشنایی با SPSS

مدارس

دکتر فرج اله رهنورد

کارشناسی ارشد مدیریت دولتی

تحليل آماری (Statistical Analysis)

هر آنالیز آماری شامل مراحل زیر است:

- توصیف داده‌ها: نمایش و خلاصه کردن داده‌ها
- آزمون فرضیات: نتیجه‌گیری در مورد گروه بزرگتر (جامعه) از روی اطلاعات به دست آمده در یک نمونه
- ارزیابی روابط: مطالعه روابط بین متغیرها

واژه های کلیدی در آمار

👉 آماره (Statistic): صفتی است از نمونه مانند میانگین (x) و واریانس نمونه (s)

👉 پارامتر (Parameter): برای توصیف ویژگی های جامعه به کار می رود مانند میانگین جامعه (μ)

👉 واریانس: پرکاربردترین شاخص پراکندگی است. هرچه ارزش واریانس بیشتر باشد، مقادیر بیشتر پراکنده هستند. اگر واریانس صفر باشد تمامی نمونه ها دارای مقدار واحدی هستند.

👉 انحراف معیار (std deviation) به پراکندگی مشاهدات در يك نمونه اشاره دارد، در حالیکه خطای معیار (std error) به پراکندگی يك آماره مربوط می باشد.

واژه های کلیدی در آمار

- نما (Mode): مقداری از داده‌هاست که بیشترین فراوانی را دارد.
- میانه (Median): مقداری از متغیر است که نیمی از مقادیر کوچکتر از آن و نیم دیگر بزرگتر از آن هستند.
- صداک‌ها (Percentile): مقادیری از متغیر هستند که در مورد درصد نمونه‌ها را نشان می‌دهند. برای مثال می‌توانید مقداری را پیدا کنید که 25% نمونه‌ها در زیر آن قرار گیرند.
- چارک‌ها (Quartiles): از آنجائی که صداک‌های 25، 50، و 75 نمونه را تقریباً به چهار گروه برابر تقسیم می‌کنند، مجموعاً به آنها چارک‌ها گویند.

Statistics

Age		
N	Valid	291
	Missing	9
Percentiles	25	1.0000
	50	2.0000
	75	3.0000

Analyze → Descriptive Statistics → Frequencies

انتخاب متغیر

OK ← Continue ← Quartiles ← Statistics..

طرز محاسبه صدك ها برای هريك از گروه‌ها

Analyze → Descriptive statistics → Explore..

Statistics..

وارد کردن متغیر مورد نظر در محل **Dependent List**
و متغیری که دارای زیرگروه‌های مورد نظر است در محل **Factor List**

Percentiles → Continue → OK

در گروه مذکور ۱۰٪ از نمونه‌ها مقاومت کم (۲) دارند

صدك‌های مقاومت افراد در برابر تغییر Percentiles

		Percentiles						
جنسیت		5	10	25	50	75	90	95
Weighted Average(Definitic	مذکر Resistanc	1.0000	2.0000	3.0000	4.0000	4.0000	5.0000	5.0000
	مونث	1.0000	1.0000	2.0000	3.0000	4.0000	5.0000	5.0000
Tukey's Hinges	مذکر Resistanc			3.0000	4.0000	4.0000		
	مونث			2.0000	3.0000	4.0000		

واژه های کلیدی در آمار

شاخص های پراکندگی: (Measures of Variability) سعی می کنند میزان گستردگی مشاهدات را اندازه گیری نمایند .

– دامنه تغییرات: (Range) ساده ترین شاخص پراکندگی است که اختلاف بین بزرگترین و کوچکترین مقدار داده ها را نشان می دهد.

– واریانس: (Variance) پرکاربردترین شاخص پراکندگی است. فرمول محاسبه واریانس در يك نمونه (S^2) به شکل زیر است:

$$\text{واریانس} = \frac{\text{مجموع توان دوم فاصله هریک از مقادیر از میانگین}}{\text{تعداد نمونه ها}} \quad (1-)$$

اگر واریانس صفر باشد تمامی نمونه ها دارای مقدار واحدی هستند. هرچه واریانس بیشتر باشد، مقادیر بیشتر پراکنده هستند.

– انحراف معیار: (Standard Deviation) جذر واریانس را انحراف معیار گویند. اندازه انحراف معیار به واحد اندازه گیری بستگی دارد. مثلاً انحراف معیار سن با واحد روز بزرگتر از انحراف معیار سن با واحد سال است.

واژه های کلیدی در آمار

نمره استاندارد (Standard score): نشان می‌دهد که یک مشاهده چند انحراف معیار بیشتر یا کمتر از میانگین می‌باشد.

$$\text{نمره استاندارد} = \frac{\text{مقدار-میانگین}}{\text{انحراف معیار}} = \frac{X - \mu}{\delta}$$

دستور محاسبه نمره استاندارد:

Analyze → Descriptive statistics → Descriptives

انتقال متغیر (های) مورد نظر به محل Variables

OK ← Save standardized values as variables

نماد E در خروجی SPSS

👉 SPSS هنگام نمایش اعداد خیلی کوچک یا خیلی بزرگ از علائم علمی استفاده می کند. عددی که بعد از حرف E قرار می گیرد نشان می دهد که محل اعشار چقدر باید جابجا شود.

👉 اگر عددی که بعد از E می آید منفی باشد، علامت اعشار را به سمت چپ حرکت دهید. اگر مثبت باشد علامت اعشار را به سمت راست حرکت دهید. به عنوان مثال:

$$-1.1E-02 = -0.011$$

طرز تبدیل يك متغير

در بعضی موارد برای تامین شروط استفاده از تحلیل‌های آماری، لازم است لگاریتم متغیر مورد نظر را حساب کرده و جایگزین متغیر قبلی کنیم.

مثلاً اگر قرار است متغیر Ram را به LogRam تبدیل کنیم، به صورت زیر عمل می‌کنیم:

Transform → **Compute**

نام متغیر جدید را در محل **Target Variable** تایپ می‌کنیم

از قسمت ماشین حساب فرمول $\text{Ln}[\text{numexpr}]$ را به محل **Numeric Expression** منتقل می‌کنیم

بجای علامت سؤال ظاهر شده متغیر مورد را منتقل می‌کنیم → **OK**

طرز کد گذاری مجدد يك متغير

Transform → Recode → Into different variables

انتقال متغير مورد نظر به محل Output variable

تايپ نام جديد متغير و انتخاب change

Old & New Values → Label تايپ عنوان جديد متغير در محل

تايپ ارزش (کد) قديمی در محل Old value و ارزش جديد در محل New value و انتخاب Add

Continue → OK

طرز گرد کردن اعداد

برای گرد کردن اعداد در روش کدگذاری کافی است در بخش Old value گزینه Range را انتخاب کنید

دامنه مورد نظر را تایپ کنید مثلاً

1 though 1.49

در بخش New value تایپ کنید 1

به همین ترتیب دیگر دامنه های مورد نظر را مشخص کنید و بقیه فرایند کدگذاری را ادامه دهید.

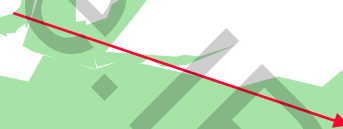
طرز تبدیل متغیر string به Numeric

از آنجائی که بعضی عملیات آماری در مورد متغیرهای نوع String قابل اجرا نیست، لازم است چنین متغیرهایی را به متغیرهای Numeric تبدیل کنیم.

Transform



Automatic recode



OK



تایپ اسم متغیر مورد نظر

نمودارهای مستطیلی (Boxplot)

یکی از ساده ترین روش های مقایسه زیر گروه های يك متغیر استفاده از نمودار مستطیلی است. این نمودار به طور همزمان میانه، دامنه بین چارکی و کوچکترین و بزرگترین مقادیر را برای گروه ها را نشان می دهد.

مرز پایینی مستطیل نماینده صدك ۲۵ می باشد و مرز بالایی نماینده صدك ۷۵ می باشد. طول عمودی مستطیل نشان دهنده دامنه بین چارکی است یعنی ۵۰ درصد نمونه ها مقداری درون مستطیل دارند. خط سیاه داخل مستطیل نشان دهنده میانه است.

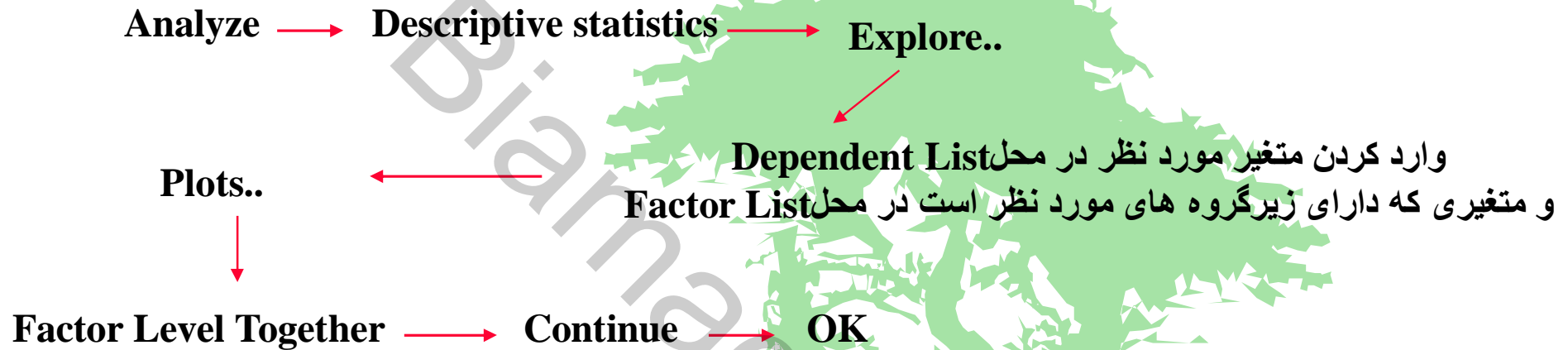
از روی نمودار مستطیلی می توان استنباط های زیر را داشت:

– از روی میانه می توان ایده ای در مورد محل تجمع داده ها به دست آورد.

– از روی طول مستطیل می توان گفت که چقدر داده ها از هم اختلاف دارند (پراکندگی)

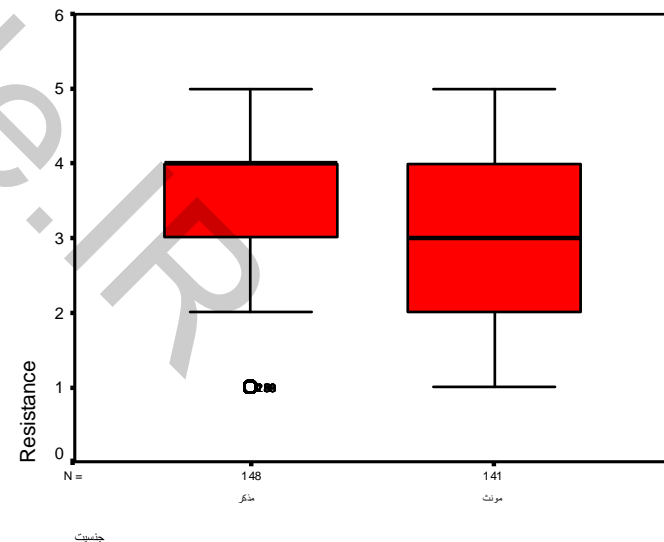
– اگر خط میانه به سمت پایین مستطیل متمایل باشد، نشان دهنده چولگی مثبت و در صورتی که به لبه بالایی مستطیل نزدیک تر باشد نشانگر چولگی منفی (Negative Skewness) می باشد.

طرز ترسیم نمودار مستطیلی



www.spowpowerplant.blogfa.com

وبلاگ یک مهندس



طرز ترسیم نمودار ستونی برای مقایسه دو

متغیر

Graphs → Bar.. → Simple → Summaries of separate variables

انتقال دو متغیری که نمودارهای آنها باید

مقایسه شوند به محل Bars Represent

Define

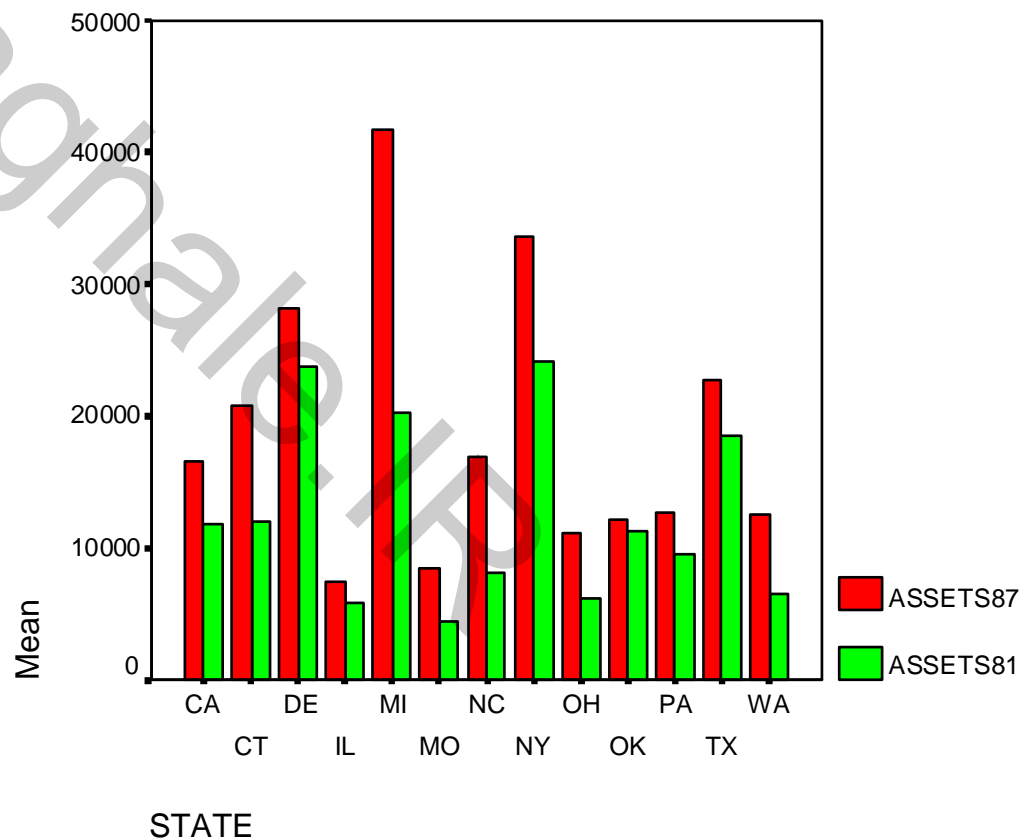
انتقال متغیری که دارای زیر گروه ها

می باشد به محل Category Axis

OK

داده ها در فایل ind8187

موجود هستند



مقیاس های اندازه گیری

1. مقیاس اسمی (Nominal Scale): ضعیف ترین سطح اندازه گیری مقیاس اسمی است. در این سطح اعداد یا نشانه ها برای طبقه بندی اشیاء یا مشاهدات به کار می روند. اطلاعات با مقیاس اسمی را نمی توان به صورت معنی داری از کم به زیاد ردیف نمود. مانند جنسیت

2. مقیاس رتبه ای (Ordinal Scale): (برای نشان دادن طبقه هایی با رتبه مساوی به کار می رود مثلاً رضایت شغلی

خیلی کم کم تا حدودی زیاد خیلی زیاد

3. مقیاس فاصله ای: زمانی به کار می رود که مقدار یک متغیر را بتوان قید کرد اما صفر مطلق ندارد (مقادیر منفی نیز پوشش داده می شود). مانند درجه حرارت. نمی توان نسبت بین دو مقدار را در متغیرهای فاصله ای محاسبه کرد.

4. مقیاس نسبت: مشابه مقیاس فاصله ای است با این تفاوت که در نسبت، مبنای صفر مورد نظر است. مثلاً وزن یا طول کمتر از صفر وجود ندارد، اما درجه حرارت هم مقادیر منفی و هم مثبت را شامل می شود. در این حالت می توان گفت که یک مقدار در مقایسه با دیگری چقدر کوچکتر یا بزرگتر است. مانند سن یا سنوات تحصیل.

انواع متغیرها

☞ متغیر وابسته (Dependent): متغیری است که تغییر پذیری آن وابسته به دیگر متغیرهاست.

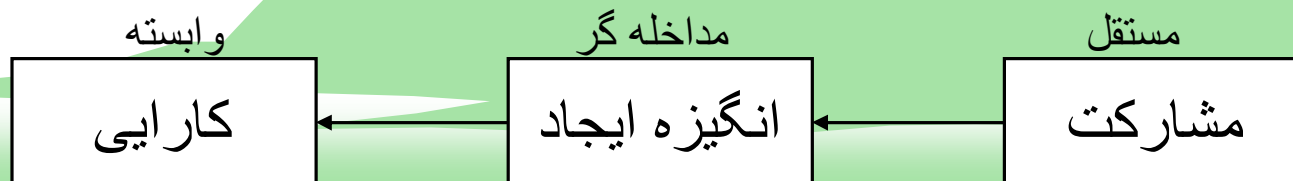
☞ متغیر مستقل (Independent): متغیری است که دلیل تغییر در متغیر وابسته را باید در آن جستجو کرد.

☞ متغیر تعدیل کننده (Moderating): این متغیر بر رابطه بین



☞ متغیر کنترل: متغیری است که تاثیر گذاری آن را در جریان پژوهش کنترل می کنیم.

☞ متغیر مداخله گر (Intervening): متغیری است که نحوه تاثیر گذاری متغیر مستقل را بر متغیر وابسته نشان می دهد.



طرز انتخاب تعداد خاصی از نمونه ها با ویژگی

مشخص

فرض کنیم بخواهیم گروهی از نمونه را در محاسبات وارد کنیم که زن هستند، در این صورت به شرح زیر عمل می‌کنیم.

Data → Select Cases → If condition is satisfied

If ← Gender = 2 که در آن شماره 2 معرف زن بودن نمونه است

Continue

OK

طرز خواندن يك فایل excel در spss

- از روی فایل دستور Open و سپس data را انتخاب کنید.
- در پنجره ای که باز می شود در قسمت File type فرمت excel را انتخاب کنید
- فایل مورد نظر را انتخاب و open را بزنید.
- در پنجره ای که باز می شود در قسمت Range آدرس متغیرها را تایپ کنید مثلاً a4: av382 و Ok را بزنید.

طرز تهیه الگوی یکسان (Template) برای

متغیرها

فرض کنید تعدادی از متغیرها در ویژگی هایی نظیر مقیاس ۵ درجه ای لیکرت مثل هم باشند. در این صورت ضرورتی برای تایپ این مقیاس برای تک تک متغیرها وجود ندارد. برای اینکار کافی است، الگوی (Template) مورد نظر را تهیه و ذخیره کنیم و ویژگی های آن را به متغیرهای مورد نظر تعمیم دهیم.

Data → Templates → Define

Value labels

تایپ نام مورد نظر در محل Name

تعریف مقیاس های مورد نظر → Continue → Add → OK

روایی محتوا (Validity)

➡ روایی محتوا اطمینان می دهد که ابزار مورد نظر به تعداد کافی پرسش های مناسب برای اندازه گیری مفهوم مورد سنجش در بردارد.

➡ هر قدر عناصر مقیاس گسترده تر و قلمرو مفهوم مورد سنجش را بیشتر در برگیرند، روایی محتوا بیشتر خواهد بود. به بیان دیگر، روایی محتوا نشان می دهد که ابعاد و عناصر يك مفهوم) که باید سنجیده شود مانند کارایی (تا چه حد تحت پوشش دقیق قرار گرفته است.

➡ برای سنجش روایی زمانی که چند مفهوم در يك پرسش نامه مورد سنجش قرار می گیرند از تحلیل عاملی استفاده می شود.

طرز سنجش روایی پرسش نامه

برای سنجش روایی محتوایی پرسش نامه باید ثابت کنیم که شاخص های يك متغیر روی يك فاکتور سوار می شوند تا ثابت کنیم آنها تك بعدی هستند.

برای این کار باید شاخص های (متغیرهای مختلف را یکجا از طریق تحلیل عاملی مورد ارزیابی قرار دهیم.

اگر پرسش نامه از روایی لازم برخوردار باشد، باید شاخص های هر متغیر روی يك عامل مستقل سوار شوند.

در صورتی که بعضی از شاخص های (نشانگرهای يك متغیر روی يك عامل سوار نشوند، می توان آنها را حذف کرد.

اعتبار (Reliability)

اعتبار ابزار میزان پایایی و سازگاری آن را در اندازه گیری يك مفهوم نشان می دهد. یعنی توانایی ابزار در حفظ پایایی خود در طول زمان (علیرغم شرایط غیر قابل کنترل آزمون و وضعیت خود پاسخگویان) حاکی از پایداری آن و تغییر پذیری اندک آن می باشد.

برای سنجش اعتبار پرسش نامه از روش های زیر می توان بهره گرفت:

– آفای کرونباخ

– تست -تست مجدد (Test-Retest)

روش Test-retest

در این روش تعدادی پرسش نامه در مقیاس کوچک توزیع می شود و بعد از گذشت مثلاً يك هفته همان پرسش نامه میان همان گروه توزیع می گردد.

داده های به دست آمده در دو مرحله با استفاده از آزمون t نمونه های جفتی مورد مقایسه قرار می گیرند. و یا همبستگی متغیر ها در دو مرحله محاسبه می شود.

آلفای کرونباخ

- ❏ یکی از متداول ترین روش ها برای تعیین اعتبار است.
- آلفا مبنی بر سازگاری داخلی يك تست است. یعنی مبتنی بر همبستگی متوسط ارقام درون يك تست است.
- ❏ فرض بر آن است که ارقام يك مقیاس باهم همبستگی دارند زیرا آنها يك موجودیت مشترك را اندازه گیری می کنند.
- ❏ چون آلفا (α) می تواند به عنوان ضریب همبستگی تعبیر شود، دامنه آن می تواند از صفر تا 1 در نوسان باشد.

طرز تست کردن اعتبار پرسش نامه

برای آنکه بتوانیم اعتبار (Reliability) شاخص های مربوط به يك متغير را بسنجيم و اطمینان حاصل کنیم که شاخص های مورد نظر متغير واحدی را می سنجدند، لازم است به شرح زیر عمل کنیم.

آلفای به دست آمده نباید از 0.7 کمتر باشد تا بتوانیم ادعا کنیم که شاخص های انتخاب شده متغير مورد نظر را می سنجدند و می توان به آنها اعتماد کرد .

Analyze → Scale → Reliability Analysis

انتقال شاخص های مورد نظر به محل Items ← Statistics

Scale if item deleted → Continue → OK

نمونه برداری

👉 **نمونه برداری تصادفی طبقه ای:** جامعه آماری ابتدا به طبقه های معنی دار تقسیم می شوند. سپس آزمودنی ها به یکی از دو صورت زیر انتخاب می شوند:

– نسبت به تعداد کل آنها در جامعه آماری

– بر اساس معیار دیگری به غیر از تعداد کلی جامعه آماری

👉 **نمونه برداری خوشه ای:** نخست گروه هایی که اعضای نامتجانس

دارند شناسایی می شوند. سپس بعضی از آنها به طور تصادفی

انتخاب می شوند و سرانجام همه اعضای هر گروه که به طور

تصادفی انتخاب شده اند مورد مطالعه قرار می گیرند. مثلاً از بین

استان های کشور به طور تصادفی چند استان انتخاب می شود و

سپس از تمام ساکنان این استان ها نظر سنجی می شود.

فرضیه صفر و فرضیه مخالف

بین فرضیه علمی و فرضیه آماری تفاوت وجود دارد. آزمون فرضیه آماری، آزمون پارامترهای جامعه هستند. فرضیه آماری معمولاً به شکل فرضیه صفر و فرضیه جایگزین (مخالف) بیان می شود.

فرضیه صفر (Null Hypothesis) ادعایی است که در مورد جامعه که مبین عدم وجود تفاوت است.

فرضیه مخالف (Alternative Hypothesis) وضعیتی را مشخص می کند که فرضیه صفر صحیح نباشد.

فرضیه صفر را باید به صورتی تعریف کرد که يك وضعيت واحد را به صورت کامل توصیف کند. مثلاً فرضیه صفر نمی تواند این باشد که افراد با تحصیلات دانشگاهی در هفته 40 ساعت کار نمی کنند. (در مقابل این عبارت که ساعات کار افراد دارای تحصیلات عالی و افراد بدون تحصیلات دانشگاهی یکسان است يك فرضیه صفر مناسب

فرضیه صفر و فرضیه مخالف (ادامه)

همیشه دقت کنیم که یا فرض صفر را رد کنیم یا آن را رد نکنیم .
هیچگاه از این جمله استفاده نکنیم که " ثابت شد که فرضیه صفر صحیح نیست".

آزمون معنی داری شاخص های آماری، در واقع همان آزمون فرضیه صفر است.

هنگامی که سطح معنی داری مشاهده شده کمتر از $05/$ باشد فرضیه صفر را رد می شود . آنگاه می توان فرضیه مقابل پذیرفته می شود.

توان (Power) اصطلاح آماری است که برای توصیف توانایی رد فرضیه صفر وقتی که غلط است به کار می رود. احتمالی است که بین صفر تا 1 تغییر می کند. توان به حجم نمونه، واریانس، و سطح معنی داری بستگی دارد.

طرز ایجاد و شمارش متغیرهای چندجوابی

اگر سئوالی در پرسش نامه مطرح شده که افراد مجاز باشند چند گزینه را انتخاب کنند در مرحله تعریف چنین متغیری و وارد کردن داده ها به شرح زیر عمل می کنیم:

1. فرض کنیم از بین يك سوال هفت گزینه ای زیر هر فرد مجاز باشد حداکثر چهار مورد را انتخاب کند .

1. Health 2. Finance 3. Lack of basic services
4. Family 5. Personal 6. legal 7. Miscellaneous

در این صورت چهار متغیر را به تعداد جواب های ممکن به شرح زیر تعریف می کنیم:

prob1 prob2 prob3 prob4

2. در مرحله وارد کردن داده ها بترتیب از متغیر اول شروع می کنیم. یعنی اگر فرد i ام دو مورد ۳ و ۵ را انتخاب کرده باشد، به ترتیب در ردیف مربوطه در ستون متغیر prob1 کد ۳ و در ستون متغیر prob2 کد ۵ را وارد می کنیم. در این صورت در محل متغیرهای prob3 و prob4 چیزی وارد نمی کنیم.

طرز ایجاد و شمارش متغیرهای چندجوابی

3. دستور زیر را پیاده می کنیم:

Analyze → Multiple response → Define Sets..

انتخاب و انتقال متغیرهای مورد نظر (مثلاً Prob1, prob2, prob3, prob4) به محل Variables in Set

Categories ← وارد کردن دامنه کدگذاری در محل Range مثلاً ۱ الی ۷

تایپ نام عمومی متغیر (مثلاً prob) در قسمت Name و عنوان کلی متغیر در محل Label

Add

Close

طرز ایجاد و شمارش متغیرهای چندجوابی

4. فراوانی جواب های چندگانه را با دستور زیر را به دست آورید:

Analyze → Multiple Response → Frequencies → Ok

تبصره:

در صورتی پرسش های چندگانه توسط پاسخ دهنده بدون پاسخ گذاشته شده باشد، هیچ عددی وارد نکنید. در این حالت در محل Missing values گزینه زیر را انتخاب کنید:

Exclude cases listwise within categories

آزمون T يك نمونه ای

One Sample T test

👉 آزمون يك نمونه ای برای آزمون این فرضیه صفر به کار گرفته می شود که آیا يك نمونه مورد نظر به جامعه با میانگین مشخص تعلق دارد یا خیر .

👉 اگر سطح معنی داری مشاهده شده کمتر از 0.05 / باشد فرض صفر رد می شود. ($H_0: \mu = \mu_0$)

👉 به عنوان مثال، می توانید از این رویه برای آزمون این فرضیه استفاده کنید که آیا بچه هایی که به طور زودرس به دنیا آمده اند بهره هوشی آنها به طور متوسط 100 می باشد یا خیر. در این حالت فرضیه صفر به این صورت بیان می شود که بین بهره هوشی افراد زودرس با بهره هوشی جامعه (مثلاً 100) تفاوت وجود ندارد. یعنی در این قبیل موارد میانگین نمونه با میانگین جامعه) که به نام Test value برچسب خورده است (مقایسه می شود.

دستور آزمون T يك نمونه ای

Analyze → Compare Means → One-Sample T test

انتقال متغیر مورد نظر به محل Test Variable

انتخاب مقداری که باید به جامعه تعمیم داده شود در محل Test Value

OK

خروجی آزمون يك نمونه ای

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	474	\$34,419.57	\$17,075.66	\$784.31

One-Sample Test

	Test Value = 32000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Current Sal	3.085	473	.002	\$2,419.57	\$878.40	\$3,960.73

سطح معنی داری دو دامنه ای و يك دامنه ای

- هرگاه جهت تفاوت در آزمون فرضیه، مشخص نباشد آن را آزمون دو دامنه گویند.
- هرگاه جهت تفاوت مورد توجه باشد، باید سطح معنی داری يك طرفه حساب شود. مانند آزمون فرضیه صفر
- برای محاسبه سطح معنی داری يك طرفه از روی سطح معنی داری دو طرفه (مثلاً آن را α^* بخوانیم) در خروجی SPSS، برای حالت $(H_0: \mu_1 - \mu_2 \leq 0)$ و $H_a: \mu_1 - \mu_2 > 0$ (به روش زیر عمل می کنیم):
 - فرض صفر (H_0) را رد می کنیم اگر $2/\alpha^*$ کوچکتر از $0.05/0$ باشد جایی که مقدار t مثبت است.
 - فرض صفر (H_0) را رد می کنیم اگر $2/(\alpha^* - 1)$ کوچکتر از $0.05/0$ باشد جایی که مقدار t منفی است.
 - در حالت $H_0: \mu_1 - \mu_2 \geq 0$ و $H_a: \mu_1 - \mu_2 < 0$ جایی که t منفی و مقدار $2/\alpha^*$ کوچکتر از $0.05/0$ باشد، فرض صفر رد می شود. و جایی که t مثبت و مقدار $2/(\alpha^* - 1)$ کوچکتر از $0.05/0$ باشد، فرض صفر را رد می کنیم.

توان (Power) آزمون t

توان اصطلاح آماری است که برای توصیف توانایی رد فرضیه صفر وقتی که غلط است، به کار می رود. توان به مقدار اختلاف واقعی، حجم نمونه، واریانس اختلاف و سطح معنی داری که در آن می خواهید فرضیه صفر را رد کنید، بستگی دارد.

برای پیدا کردن توان يك آزمون، می توان به جدول توان که توسط کوهن (۱۹۸۸) تهیه شده مراجعه کرد.

برای افزایش توان يك آزمون موارد زیر توصیه شده است:

- هرچه سطح معنی داری بزرگتر باشد، توان آزمون بیشتر است
- هرچه اندازه نمونه بزرگتر باشد، توان آزمون بیشتر است.
- آزمون يك دامنه ای از آزمون دو دامنه ای پرتوان تر است.
- در آزمون يك دامنه، هرچه اختلاف بین میانگین واقعی (μ) و میانگین نمونه (μ_0) بیشتر باشد، توان آزمون بیشتر است.

آزمون T با نمونه های جفت

Paired Sample T test

- ➡ از آزمون T زمانی استفاده می کنیم که هر فرد یا حیوانی دو بار در دو وضعیت متفاوت مورد مشاهده قرار می گیرند.
- ➡ یکی از این مطالعات طرح قبل و بعد می باشد. به عنوان مثال، اگر بخواهیم به روش Test-Retest اعتبار پرسشنامه را بسنجید یا فشار خون فرد را قبل و بعد از اعمال یک تیمار مقایسه کنید.
- ➡ فرضیه صفر در طرح جفت این است که اختلافی بین مقادیر متوسط در دو عضو یک جفت جامعه وجود ندارد. به عبارت دیگر، اختلاف متوسط جامعه صفر است.

دستور آزمون T با نمونه های جفتی

Analyze



Paired Sample T Test



مورد نظر انتخاب دو متغیر که معرف مشاهدات دوتایی هستند و انتقال آنها به محل

Paired Variables



OK

در صورتی که سطح معنی داری مشاهده شده کمتر از 05/ باشد، می توان فرض صفر را رد کرد. یعنی تفاوت معنی داری بین دو حالت مورد نظر وجود دارد.

خروجی آزمون T با نمونه های جفتی

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Current Salary	\$34,419.57	474	\$17,075.66	\$784.31
	Beginning Salary	\$17,016.09	474	\$7,870.64	\$361.51

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Current Salary & Beginning Salary	474	.880	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	5% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Current Salary - Beginning Salary	17,403.48	\$10,814.62	\$496.73	16,427.41	18,379.56	35.036	473	.000

چند مشکل احتمالی در آزمون T با نمونه های

جفتی

- 👉 اگر می خواهید اثر دو درمان را بر روی يك فرد مقایسه کنید باید بین دو درمان زمان کافی وجود داشته باشد.
- 👉 اگر نمونه های شما يك فعالیت یا آزمایش را دوبار تکرار می کنند، ممکن است دفعه دوم به دلیل اثر آموزش بهتر از دفعه اول عمل کنند (حتی اگر مداخله مورد نظر اثری نداشته باشد، اثر آموزش می تواند موجب تغییراتی شود).
- 👉 در تحقیقاتی از این دست باید این موضوع در محدودیت های پژوهش نوشته شود.

آزمون T با نمونه های مستقل

Independent-sample T test

رویه T با نمونه های مستقل این فرضیه صفر را آزمون می کند که میانگین يك متغیر در جامعه برای دو گروه مختلف افراد باهم برابر است .

هنگامی که سطح معنی داری مشاهده شده کوچک باشد فرض صفر) میانگین های يك متغیر در گروه های مختلف باهم تفاوتی ندارند) رد می شود و نتیجه می گیرید که به نظر می رسد میانگین های دو گروه با هم برابر نیستند.

دستور آزمون T با دو نمونه مستقل

Analyze → Compare Means — Independent-sample T Test

انتقال متغیر مورد نظر به محل Test Variable و متغیری که شامل گروه های مورد
نظر است به محل Grouping variable

↓
Define Groups

↓
نوشتن کد هر دو گروه در محل های مورد نظر

↓
Continue

↓
OK

تفسیر نتایج در آزمون T با دو نمونه مستقل

👉 در خروجی دو ویرایش متفاوت از آزمون t دیده می شود. اولی با این فرض است که واریانس دو جامعه با هم برابرند و در دومی چنین فرضی وجود ندارد.

👉 با استفاده از آزمون Levene می توان فرضیه صفر متعلق بودن دو نمونه به جامعه هایی با واریانس های برابر را آزمون کرد. اگر سطح معنی داری مشاهده شده در آزمون کمتر از 05/ باشد می توان فرضیه صفر برابری واریانس دو جامعه را رد کرد. در این مثال برابری واریانس ها رد می شود. بنابراین، باید نتایجی را که دارای برچسب Equal variance not assumed هستند، استفاده کرد.

آنالیز واریانس يك طرفه (One-Way ANOVA)

توزیع t برای آزمون تفاوت بین دو گروه مناسب است. اگر سه یا چند گروه برای مقایسه داشته باشیم، نمی توان توزیع t را برای سنجش معنی دار بودن تفاوت گروه ها به کار گرفت.

اگر بخواهیم میانگین های جامعه را در بین چند گروه مقایسه کنیم از آنالیز واریانس یکطرفه استفاده می کنیم .

آنالیز واریانس یکطرفه نمونه ها را بر اساس مقادیر يك متغیر در گروه های مختلف مقایسه می کند. متغیری که جهت تشکیل گروه ها استفاده می شود فاکتور نامیده می شود.

مثلا فرض صفر می گوید که میانگین حقوق در طبقات شغلی مختلف یکسان است (بین طبقات شغلی مختلف از نظر حقوق دریافتی تفاوت وجود ندارد).

جدول آنالیز واریانس

ANOVA

Current Salary

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.9E+10	2	4.5E+10	434.481	.000
Within Groups	4.8E+10	471	1.0E+08		
Total	1.4E+11	473			

با نگاه کردن به سطح معنی داری مشاهده شده می توان فرض صفر یعنی برابری حقوق را در طبقات شغلی مختلف رد کرد. به عبارت دیگر، فرض صفر مبنی بر اینکه بین طبقات مختلف شغلی از نظر حقوق دریافتی تفاوت وجود ندارد، رد می شود.

آزمون مقایسه چندگانه در حقوق های دریافتی

Multiple Comparisons

Dependent Variable: Current Salary

Bonferroni

(I) Employment Category	(J) Employment Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Clerical	Custodial	-\$3,100.35	\$2,023.76	.379	-\$7,962.56	\$1,761.86
	Manager	\$36,139.26*	\$1,228.35	.000	\$39,090.45	-\$33,188.07
Custodial	Clerical	\$3,100.35	\$2,023.76	.379	-\$1,761.86	\$7,962.56
	Manager	\$33,038.91*	\$2,244.41	.000	\$38,431.24	-\$27,646.58
Manager	Clerical	\$36,139.26*	\$1,228.35	.000	\$33,188.07	\$39,090.45
	Custodial	\$33,038.91*	\$2,244.41	.000	\$27,646.58	\$38,431.24

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

تفسیر مقایسه چندگانه

👉 در آزمون مقایسه Bonferroni هر ردیف شامل مقایسه يك جفت می باشد. در مثال فوق ردیف اول مربوط به مقایسه کارمند دفتری با نگهبان و مدیر است. ردیف آخر مربوط به مقایسه مدیر با کارمند دفتری و نگهبان می باشد.

👉 جفت هایی از میانگین که به طور معنی داری باهم اختلاف دارند توسط ستاره مشخص شده است. سایر طبقات شغلی باهم اختلاف معنی داری ندارند.

دستور اجرای آنالیز واریانس یکطرفه

Analyze → Compare Means → One-Way ANOVA

انتقال متغیر Curent Salary به محل Dependent List

انتقال متغیر Jobcat به محل Factor

Post Hoc..

انتخاب Bonferroni

Continue

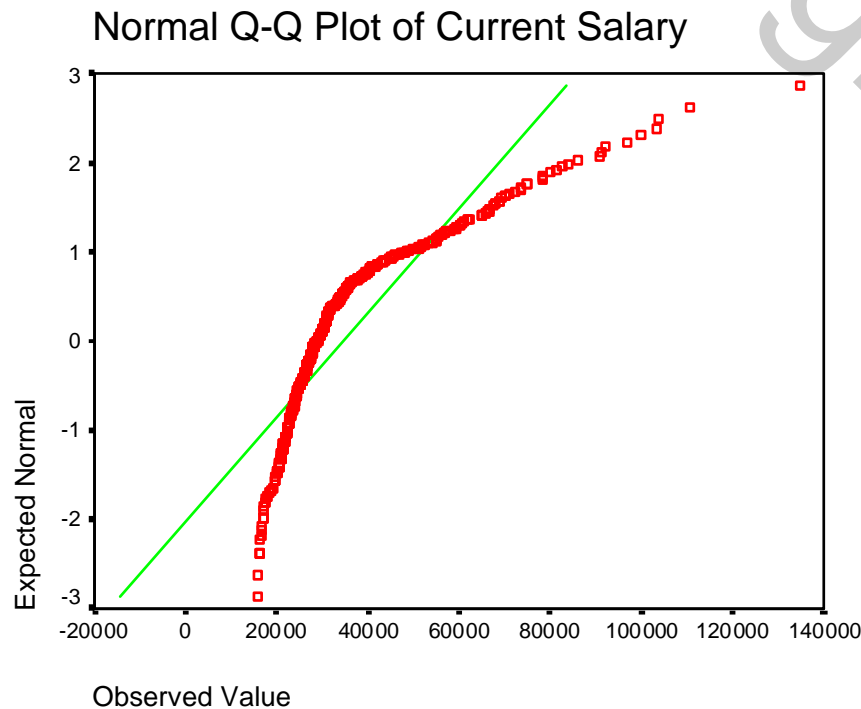
OK

شرط استفاده از آنالیز واریانس

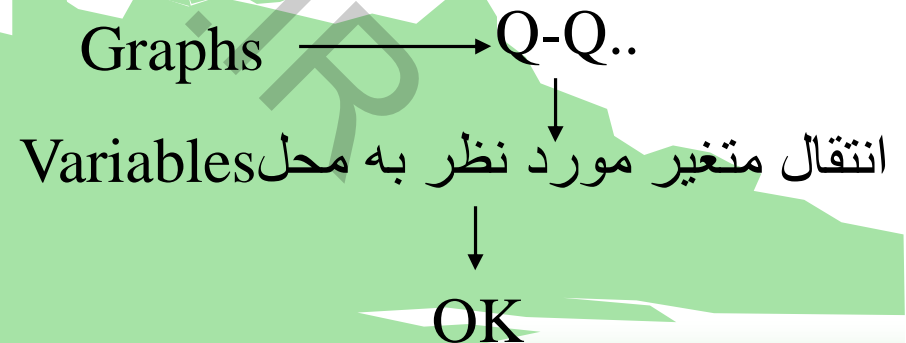
- ☞ از هر جامعه نمونه های تصادفی مستقل گرفته شده باشند (هریک از مشاهدات باید مستقل باشند).
- ☞ متغیرها بایستی به طور نرمال توزیع شده باشند
- ☞ گروه ها باید دارای واریانس یکسان باشند

بررسی نرمال بودن

1. نمودار Q-Q: این نمودار مقدار مشاهده شده و مقدار مورد انتظار (مربوط به حالتی که داده های نمونه متعلق به توزیع نرمال است) را نشان می دهد. اگر داده ها متعلق به یک توزیع نرمال باشند، نقاط باید اطراف یک خط صاف جمع شوند.



دستور رسم نمودار:



بررسی نرمال بودن

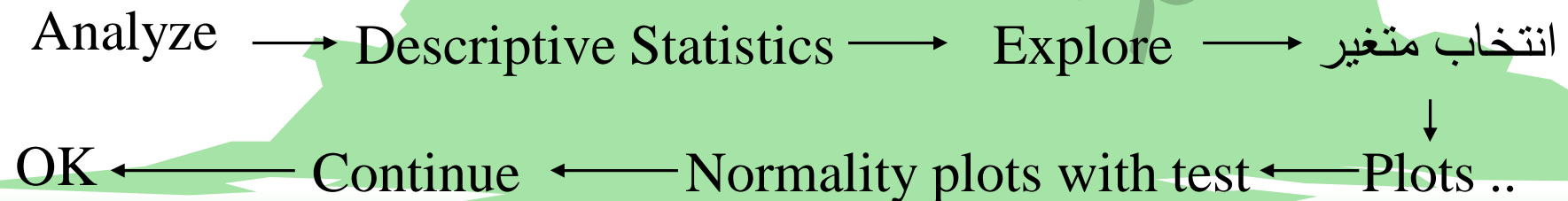
2. آزمون آماری: می توان این فرض صفر را آزمون کرد که آیا داده های مورد نظر، نمونه ای از جامعه نرمال می باشد. اگر سطح معنی دار مشاهده شده کوچک باشد، باید به فرض نرمال بودن شك کرد. در غیر اینصورت) اگر سطح معنی داری به اندازه کافی بزرگ باشد، فرض نرمال بودن غیر منطقی نیست.)

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
Current Salary	.208	474	.000

a. Lilliefors Significance Correction

دستور آزمون نرمال بودن:



تعبیر چولگی (Skewness) و کشیدگی (Kurtosis)

این دو شاخص هایی برای قضاوت در خصوص نرمال بودن توزیع هستند. یک توزیع نرمال باید دارای چولگی صفر و کشیدگی صفر باشد.

اگر چولگی مثبت باشد، توزیع چوله به راست است و اگر منفی باشد چوله به چپ.

اگر کشیدگی منفی باشد، توزیع از حالت زنگی شکل پهن تر است و اگر مثبت باشد توزیع باریک تر از منحنی نرمال خواهد بود.

آزمون برابری واریانس گروه‌های مورد

مقایسه

برای استفاده از بعضی تکنیک‌های آماری نظیر آزمون t با دو نمونه مستقل لازم است به آزمون برابری واریانس‌ها پرداخت. برای این منظور از آزمون **Levene's Test** استفاده می‌شود. یعنی فرضه صفر متعلق بودن دو نمونه به جامعه‌هایی با واریانس برابر را آزمون می‌کنیم. این آزمون نباید معنی‌دار باشد. در این صورت می‌توان گفت که واریانس گروه‌ها یکسان است. یعنی تفاوت معنی‌داری از نظر تغییرپذیری بین گروه‌ها وجود ندارد.

اگر آزمون **Levene** منجر به رد فرضه صفر برابری واریانس دو جامعه شود، باید از ستونی استفاده کنید که دارای برچسب **Equal variances not assumed** است.

دستور آزمون برابری واریانس ها

Analyze → Compare Means → Independent samples T test

انتقال متغیر مورد نظر به محل Test Variable

انتقال متغیری که مقادیر آن دو گروه مورد نظر را تعریف می کند به محل Grouping Variables

OK

تحلیل واریانس عاملی (دوطرفه)

Factorial ANOVA

روش سنتی پژوهش تجربی مبتنی بر مطالعه اثر یک متغیر مستقل در یک متغیر وابسته بوده است.

روش تحلیل واریانس عاملی این امکان را به وجود آورده است که در طرح پژوهشی خود نفوذ چند متغیرهای مستقل) که معمولاً عامل Factor نامیده می شوند) بر متغیر وابسته را بسنجیم.

تحلیل واریانس عاملی روشی است که اثرهای ساده و تعاملی دو یا چند متغیر مستقل را بر حسب یک متغیر وابسته مورد تحلیل قرار می دهد. به بیان دیگر، در تحلیل عاملی دو یا چند متغیر مستقل، به گونه مستقل یا در تعامل با یکدیگر تغییر می کنند تا تغییر پذیری متغیر وابسته را به وجود آورند.

يك مثال پژوهشی از تحلیل واریانس عاملی

- ➡ آیا متوسط کارایی در چند گروه تحصیلی یکسان است؟
- ➡ آیا متوسط کارایی مردها و زن ها یکسان است؟
- ➡ آیا رابطه بین متوسط کارایی و درجه تحصیلی در مردها و زن ها یکسان است؟
در سؤال 1 و 2 يك فاکتور (گروه تحصیلی یا جنسیت) (درگیر است در حالی که در سؤال سوم هر دو فاکتور به طور مزمان مورد نظر است.
- ➡ پژوهشگر در سؤال سوم يك فرضه تعاملی را مطرح می سازد و معتقد است که متوسط ساعات کار مردها و زن ها که دارای تحصیلات مختلف هستند به گونه متفاوت ظاهر می شود.

فوق لیسانس

لیسانس

دیپلم

مردها

زن ها

کارایی

این پژوهشگر 60 کارمند را از میان جامعه کارکنان به تصادف انتخاب می کند به گونه ای که به 6 گروه 10 نفری مطابق جدول زیر تقسیم شده باشند.

جنسیت	تحصیلات		
	A1	A2	A3
B1	A1B1	A2B1	A3B1
B2	A1B2	A2B2	A3B2

اثر اصلی (Main Effect)

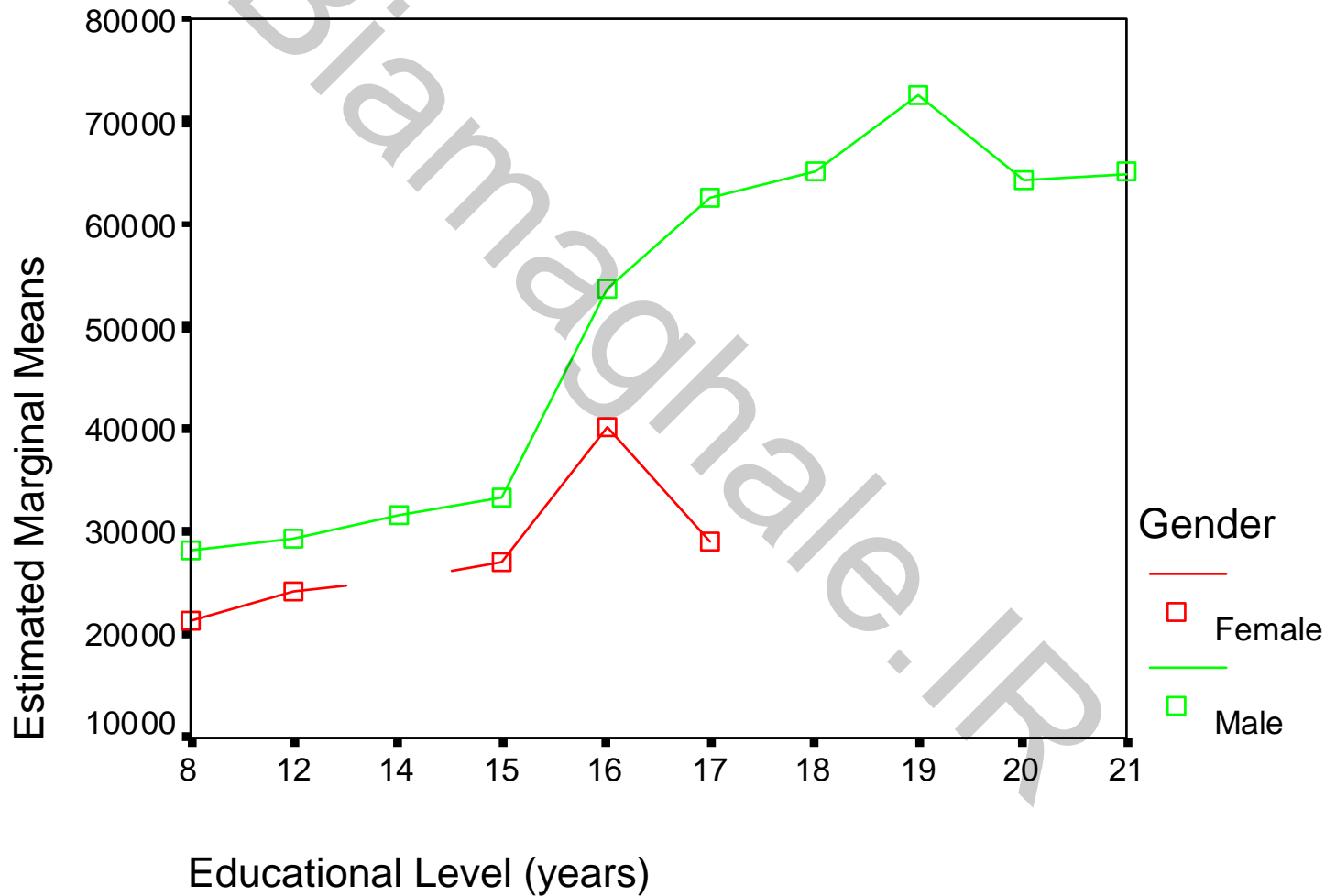
در آنالیز واریانس اثر اصلی عبارت است از اثر هر یک از فاکتورها صرف نظر از اثر سایر فاکتورها مثلاً اثر درجه تحصیلی (به تنهایی) بر روی متوسط کارایی اثر اصلی نامیده می شود.

اثر تعاملی: (Interaction Effect) اگر دو فاکتور را

همزمان در نظر بگیریم، آزمون اثر تعاملی نامیده می شود.

نمودار خطی میانگین های مشاهده شده

Estimated Marginal Means of Current Salary



Non-estimable means are not plotted

جدول آنالیز واریانس

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.5E+10 ^a	14	6.8E+09	72.704	.000
Intercept	8.7E+10	1	8.7E+10	932.754	.000
GENDER	3.3E+09	1	3.3E+09	35.054	.000
EDUC	4.8E+10	9	5.4E+09	57.684	.000
GENDER * EDUC	1.4E+09	4	3.5E+08	3.695	.006
Error	4.3E+10	459	9.3E+07		
Total	7.0E+11	474			
Corrected Total	1.4E+11	473			

a. R Squared = .689 (Adjusted R Squared = .680)

برای استخراج این جدول به فایل Employee data یا Bank مراجعه کنید.

تفسیر آنالیز واریانس دوطرفه

با نگاه کردن به سطح معنی دار مشاهده شده برای هر یک از نسبت های F مشخص می شود که آیا می توان فرضیه صفر آن را رد کرد:

– **فرضیه صفر اثر اصلی:** میانگین های جامعه (میزان حقوق) در تمام سطوح تحصیلی یکسان است. یعنی در سطوح تحصیلی مختلف تفاوتی در میزان حقوق افراد وجود ندارد.

– **فرضه صفر اثر اصلی:** میانگین های جامعه (میزان حقوق) در هر دو جنس زن و مرد یکسان است.

– **فرضیه صفر اثر تعاملی:** اثر تحصیلات بر روی حقوق در هر دو جنس زن و مرد در جامعه یکسان است (بین دو متغیر اثر تعاملی وجود ندارد).

همان طور که ملاحظه می شود سطح معنی داری مشاهده شده در مورد اثر تعاملی کوچکتر از $05/$ می باشد. بنابراین، فرض صفر رد می شود. یعنی اثر سطح تحصیلات بر روی حقوق در مردها و زن ها یکسان نیست. به عبارت دیگر، رابطه بین سطح تحصیلات و حقوق در مردها و زن ها متفاوت است.

تفسیر آنالیز واریانس دوطرفه (ادامه)

همان طور که ملاحظه می شود سطح معنی داری مشاهده شده در مورد اثر اصلی (متغیر جنسیت) (کوچتر از 0.05 / می باشد بنابراین، فرض صفر یعنی برابری میانگین حقوق در هر دو جنس مرد و زن رد می شود.

همان طور که ملاحظه می شود سطح معنی داری مشاهده شده در مورد اثر اصلی (متغیر تحصیلات) (کوچتر از 0.05 / می باشد بنابراین، فرض صفر یعنی برابری میانگین حقوق در هر میان افراد با سطوح تحصیلات مختلف رد می شود.

آزمون اثرات تعاملی بسیار مهم است. اگر اثر تعاملی پیدا شود، صحبت کردن در مورد جنسیت و سطح تحصیلی به نهایی معنی ندارد و هر دو باید باهم در نظر گرفته شوند.

مجموع مربعات ارائه شده برای ردیف Corrected Model می گوید چقدر از پراکندگی متغیر وابسته می تواند توسط اثرات متغیرهای مشاهده شده تبیین شود این درصد در زیر جدول به نام R^2 ارائه شده

دستور اجرای آنالیز واریانس دوطرفه

Analyze → General Linear Model

↓
Univariate

↓
انتقال متغیر وابسته به محل
Dependent Variable

↓
انتقال دو یا چند عامل مورد نظر به محل
Fixed Factors

↓
از روی Plots عامل اول را به محل Horizontal Axis و عامل دوم را به محل Separate Lines
و در صورت لزوم عامل سوم را به محل Separate Plots منتقل Add و Continue را انتخاب کنید.

↓
OK

تحلیل واریانس چند متغیره (MANOVA)

- ➡ رویه MANOVA تحلیل رگرسیونی و تحلیل واریانس را برای متغیرهای وابسته چندگانه فراهم می سازد.
- ➡ (متغیر) های عامل (Factor Variables) جامعه را به چند گروه تقسیم می کنند.
- ➡ در این رویه اثر متغیرهای مستقل (عامل) بر میانگین های گروه های مختلفی از متغیرهای وابسته آزمون می شود.
- ➡ در این روش تعامل بین عوامل (متغیرهای مستقل) و اثرات انفرادی عوامل سنجیده می شود.

اثر تعاملی (Interaction Effect)

در آنالیز واریانس دو طرفه اثر اصلی (Main Effect) عبارت است از اثر هر یک از فاکتورها صرف نظر از سایر فاکتورها. مثلاً اثر درجه تحصیلی بر روی متوسط ساعات کار یک اثر اصلی نامیده می شود، اما اگر دو فاکتور را همزمان وارد کنید آزمون اثر تعاملی نامیده می شود.

اگر در خروجی SPSS نمودار خطی میانگین های مشاهده شده همدیگر را قطع کنند اثر تعاملی وجود دارد.

برای آزمون اثر تعاملی می توان فرضه های آماری تشکیل داد مثلاً (اثر فاکتور 1 (فاکتور تحصیلات) بر روی متغیر وابسته در گروه های فاکتور) 2 فاکتور جنسیت (یکسان است):

$$\left\{ \begin{array}{l} H_0: \text{اثر تحصیلات بر روی متوسط ساعات کار در هر دو جنس زن و مرد یکسان است.} \\ H_a: \text{اثر تحصیلات بر روی متوسط ساعات کار در هر دو جنس زن و مرد یکسان نیست.} \end{array} \right.$$

اثر تعاملی (ادامه)

اگر سطح معنی داری مشاهده شده برای اثر تعاملی بزرگ تر از 05/ باشد (Sig. >0.05) نمی توان فرض صفر را رد کرد (بین دو متغیر اثر تعاملی وجود ندارد). (یعنی مثلاً اثر تحصیلات بر روی ساعات کار در مردها و زن ها یکسان به نظر می رسد).

اگر اثر تعاملی پیدا شود، دیگر نباید اثر هر يك از آنها را به تنهایی بر روی متغیر وابسته بسنجیم. اگر اثر تعاملی مشاهده نگردید می توان اثر اصلی هر يك از آنها را به روش آنالیز واریانس يك طرفه بر متغیر وابسته سنجید.

تحليل واريانس چند متغيره (MANOVA)

به علاوه، اثرات همپراش ها (covariates) و همچنين تعامل هاي همپراش ها با عوامل (متغيرهاي مستقل) مورد بررسي قرار مي گيرد.

براي تحليل رگرسيوني، متغيرهاي مستقل (Predictor) به عنوان متغيرهاي همپراش در نظر گرفته مي شوند.

اگر بيش از يك متغير وابسته مشخص شود، تحليل چند

متغيره واريانس از روش هاي $phillai$, $wilks$,

$Hotelling$ و Roy با آماره F استفاده مي کند. همچنين

تحليل تک متغيره (univariate) واريانس براي هر متغير وابسته صورت مي گيرد.

مثال

فرض کنید اثر دو متغیر مستقل I_1 و I_2 را بر روی سه متغیر وابسته $D1$ ، $D2$ ، و $D3$ مورد آزمایش قرار می دهیم. در این صورت خروجی SPSS در تحلیل واریانس چند متغیره به صورت زیر خواهد بود.

خروجی تحلیل واریانس چند متغیره

Multivariate Tests^c

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.928	1089.906 ^a	3.000	252.000	.000
	Wilks' Lambda	.072	1089.906 ^a	3.000	252.000	.000
	Hotelling's Trace	12.975	1089.906 ^a	3.000	252.000	.000
	Roy's Largest Root	12.975	1089.906 ^a	3.000	252.000	.000
I1	Pillai's Trace	.174	3.907	12.000	762.000	.000
	Wilks' Lambda	.832	4.004	12.000	667.021	.000
	Hotelling's Trace	.195	4.076	12.000	752.000	.000
	Roy's Largest Root	.152	9.638 ^b	4.000	254.000	.000
I2	Pillai's Trace	.088	1.924	12.000	762.000	.029
	Wilks' Lambda	.913	1.937	12.000	667.021	.028
	Hotelling's Trace	.093	1.946	12.000	752.000	.026
	Roy's Largest Root	.069	4.374 ^b	4.000	254.000	.002
I1 * I2	Pillai's Trace	.310	1.833	48.000	762.000	.001
	Wilks' Lambda	.717	1.854	48.000	750.305	.001
	Hotelling's Trace	.359	1.874	48.000	752.000	.000
	Roy's Largest Root	.218	3.468 ^b	16.000	254.000	.000

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept+I1+I2+I1 * I2

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	D1	70.101 ^a	24	2.921	2.624	.000
	D2	55.598 ^b	24	2.317	1.966	.006
	D3	93.502 ^c	24	3.896	3.842	.000
Intercept	D1	1566.989	1	1566.989	1407.720	.000
	D2	1357.000	1	1357.000	1151.896	.000
	D3	1386.230	1	1386.230	1366.885	.000
I1	D1	9.114	4	2.278	2.047	.088
	D2	5.888	4	1.472	1.249	.291
	D3	35.533	4	8.883	8.759	.000
I2	D1	19.216	4	4.804	4.316	.002
	D2	5.476	4	1.369	1.162	.328
	D3	4.241	4	1.060	1.045	.384
I1 * I2	D1	43.806	16	2.738	2.460	.002
	D2	24.397	16	1.525	1.294	.201
	D3	20.763	16	1.298	1.280	.210
Error	D1	282.738	254	1.113		
	D2	299.227	254	1.178		
	D3	257.595	254	1.014		
Total	D1	3334.000	279			
	D2	3099.000	279			
	D3	3235.000	279			
Corrected Total	D1	352.839	278			
	D2	354.824	278			
	D3	351.097	278			

a. R Squared = .199 (Adjusted R Squared = .123)

b. R Squared = .157 (Adjusted R Squared = .077)

c. R Squared = .266 (Adjusted R Squared = .197)

دستور اجرای MNOVA

Analyze → General linear Model → Multivariate

انتقال متغیرهای وابسته به محل
Dependent Variables

انتقال متغیرهای مستقل به محل
Fixed factors

Plots..

انتقال يك متغیر به محل Horizontal Axis و متغیر دیگر به محل Separate Lines

Add → Continue → OK

آزمون مربع کای دو

Chi-square test

H_0 : هستند متغیر مستقل دو

H_a : دارد متغیر وجود دو تفاوت معنی داری بین

- اگر بخواهیم استقلال بین دو متغیر را که محاسبه میانگین آنها ممکن نیست) تنها می توان در مورد تعداد مقادیر مختلف آن صحبت کرد مانند متغیرهای اسمی (آزمون کنیم از آماره کای دو استفاده می کنیم.
- آماره کای دو بر مقادیر مشاهده شده و مورد انتظار که از طریق جدول توافقی به دست می آیند استوار است.
- در جدول توافقی مقدار مشاهده شده عبارت است از تعدادی از نمونه ها که در يك خانه قرار داند.
- مقدار مورد انتظار عبارت است از تعدادی که در صورت مستقل بودن دو متغیر پیش بینی می شود.

جدول توافقی

👉 جزء اصلی جدول توافقی تعداد نمونه هایی است که در هر يك از خانه های جدول قرار می گیرند. روش های آماری که در این فرضیه های صفر به کار می روند بر اساس مقایسه موارد مشاهده شده در هر خانه با تعداد مورد انتظار آن عمل می کند. تعداد مورد انتظار به طور ساده تعدادی از نمونه هاست که در صورت صحیح بودن فرضیه صفر انتظار می رود در هر يك از خانه ها پیدا شود.

👉 فرضیه صفر در جدول توافقی به صورت مستقل بودن دو متغیر بیان می شود (دو متغیر مورد نظر مستقل از یکدیگر هستند.)

شرط استفاده از آماره کای دو

1. مشاهدات باید مستقل باشند. یعنی نمونه به صورت تصادفی از جامعه انتخاب شده باشد.

2. بیش از 20% سلول های جدول توافقی دارای فراوانی مورد انتظار کمتر از 5 نباشد. یعنی

شاخص Minimum expected frequency کمتر از يك نباشد.

دستور اجرای آماره کای دو

Analyze → Descriptive Statistics → Crosstabs..

انتقال متغیرهای مورد نظر یکی به محل Rows و دیگری به محل Columns

Chi-square ← Statistics

انتخاب گزینه phi and cramer's v اگر یکی از متغیرهای مورد نظر اسمی باشد و Kendall's tau برای متغیرهای رتبه ای

Cells ← continue

continue ← Unstandardised و Total, Expected, Observed های انتخاب گزینه

OK

خروجی آزمون کای دو (جدول توافقی)

Is Life Exciting or Dull * Married? Crosstabulation

			Married?		Total
			yes	no	
Is Life Exciting or Dull	Dull	Count	21	44	65
		Expected Count	33.5	31.5	65.0
		% of Total	2.1%	4.4%	6.5%
		Residual	-12.5	12.5	
	Routine	Count	241	218	459
		Expected Count	236.4	222.6	459.0
		% of Total	24.2%	21.9%	46.1%
		Residual	4.6	-4.6	
	Exciting	Count	251	221	472
		Expected Count	243.1	228.9	472.0
		% of Total	25.2%	22.2%	47.4%
		Residual	7.9	-7.9	
Total	Count	513	483	996	
	Expected Count	513.0	483.0	996.0	
	% of Total	51.5%	48.5%	100.0%	

برای به دست آوردن این
جدول از فایل GSS
متغیرهای Married
و Life را انتخاب کنید.

خروجی آزمون کای دو - (χ^2) ادامه

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.303 ^a	2	.006
Likelihood Ratio	10.475	2	.005
Linear-by-Linear Association	4.477	1	.034
N of Valid Cases	996		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 31.52.

تست شرایط استفاده از χ^2

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.102	.006
	Cramer's V	.102	.006
N of Valid Cases		996	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

تفسیر نتایج در آزمون کای دو

اگر سطح معنی داری مشاهده شده کمتر از $05/$ باشد می توان فرض صفر یعنی استقلال دو متغیر را رد کرد.

مقدار کای دو (ارزش عددی به دست آمده در روش پیرسون) مقیاسی از قدرت رابطه دو متغیر نیست. برای پیدا کردن قدرت رابطه دو متغیر از مقیاس های Φ و V Cramer's استفاده می کنیم. در مقیاس کرامر مقدار به دست آمده بین صفر و ۱ در نوسان است. صفر نشانگر عدم وجود رابطه و ۱ معرف رابطه کامل است.

در زیر جدول آزمون کای میزان تحقق شرایط استفاده از آزمون قید شده است.

آزمون کای دو (خی دو (یک نمونه ای

👉 از آزمون کای دو می توان جهت آزمون توزیع یک متغیر نیز استفاده کرد. یعنی می توانید ببینید که آیا توزیع مقادیر مشاهده شده در یک جدول فراوانی با تعداد مقادیر مورد انتظار یکی می باشد.

👉 مثلاً می توان آزمون کرد که آیا احتمال افرادی که زندگی را هیجان انگیز می دانند یا معمولی و خسته کننده می دانند با هم برابر است؟

خروجی آزمون خی دو یک نمونه ای

Is Life Exciting or Dull

	Observed N	Expected N	Residual
Dull	65	332.3	-267.3
Routine	459	332.3	126.7
Exciting	473	332.3	140.7
Total	997		

Test Statistics

	Is Life Exciting or Dull
Chi-Square ^a	322.865
df	2
Asymp. Sig.	.000

باتوجه به پایین بودن سطح

معنی داری مشاهده شده

فرض صفر یعنی برابری برداشت افراد

از زندگی رد می شود.

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 332.3.

دستور آزمون کای دو یک نمونه ای

Analyze



Nonparametric Test



Chi-Square..



انتقال متغیر مورد نظر به محل Test Variable List



انتخاب گزینه All Categories Equal



OK

توان آزمون کای دو

می دانیم که توانایی ما برای رد فرض صفر وقتی که غلط است (توان آزمون) به اندازه نمونه بستگی دارد. این قاعده برای آزمون کای دو نیز درست است. مقدار آماره کای دو به تعداد مشاهدات در نمونه بستگی دارد.

اگر نمونه کوچکی داشته باشیم، ممکن است قادر به رد فرض صفر حتی اگر آن غلط باشد نباشیم. به طریق مشابه برای نمونه های بزرگ ممکن است به رد فرض صفر برسیم حتی اگر انحراف از استقلال خیلی کوچک باشد.

وقتی یکی از متغیرها یا هر دوی آنها در جدول توافقی بر مبنای مقیاس رتبه ای سنجیده می شود، آزمون کای دو مانند دیگر آماره ها برای کشف انحراف از استقلال قوی

پارامتری غیر آزمون های Non-Parametric Tests

👉 آزمون های غیر پارامتری زمانی به کار گرفته می شوند که نمونه ها کوچک باشند و فرض های ضروری آزمون های پارامتری به شدت صادق نباشند.

👉 آزمون های پارامتری به آزمون هایی گفته می شوند که به فرض نرمال بودن جامعه نیاز دارند. در حالیکه آزمون های غیر پارامتری دارای توزیع نرمال نیستند.

👉 مزیت آزمون های غیر پارامتری آن است که به فرض های کمتری در مورد داده ها نیاز دارند. عیب آنها آن است که توان کمتری نسبت به سایر آزمون ها دارند.

آزمون های غیر پارامتری

Non-Parametric Tests

👉 آزمون Sign جایگزین غیر پارامتری آزمون t جفتی است که توان کمتری دارد. در حالی که آزمون مشابه آن یعنی Wilcoxon از توان بیشتری برخوردار است.

👉 آزمون Mann-Whitney جایگزین غیر پارامتری آزمون t با نمونه های مستقل می شود.

👉 آزمون Kruskal-Wallis جایگزین غیر پارامتری برای آنالیز واریانس یکطرفه است.

👉 آزمون Runs بررسی می کند آیا مشاهدات کنار هم در یک سری از هم مستقل هستند.

آزمون رتبه ای فریدمن (Friedman)

آزمون رتبه ای فریدمن (1937) یک آزمون رتبه ای برای K نمونه همبسته است.

Analyze → Nonparametric Tests → K Related Samples

انتقال متغیرها به محل Test variables ← Friedman ← OK

Ranks

	Mean Rank
مرتبط بودن اطلاعات	3.75
دقت اطلاعات	3.13
صحت اطلاعات	2.98
ایمنی اطلاعات	3.25
سرعت اطلاعات	4.15
اقتصادی بودن اطلاعات	3.75

Friedman Test

آزمون دو جمله ای

آزمون دو جمله ای برای آزمون این فرضیه به کار می رود که آیا یک متغیر از یک جامعه دو جمله ای با احتمال مشخصی تصادفی است یا خیر. متغیر مورد نظر تنها دو مقدار می تواند داشته باشد.

Analyze → Nonparametric Tests → Binomial

انتقال متغیر مورد نظر به محل Test Variable list ← OK

Binomial Test

	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
V7	Group 1	47	.22	.50	.000 ^a
	Group 2	171	.78		
	Total	218	1.00		

a. Based on Z Approximation.

اندازه گیری رابطه

- ➡ روش های آماری مختلفی برای مطالعه رابطه بین متغیرها استفاده می شود که به نوع متغیر و مانند اینها بستگی دارد. مانند کای دو، رگرسیون، و همبستگی
- ➡ برای مشخص کردن قدرت و طبیعت رابطه بین دو متغیر باید از طریق محاسبه شاخص رابطه (Measures of association) اقدام کرد.
- ➡ يك شاخص رابطه عددی است که بزرگی آن قدرت رابطه بین دو متغیر را مشخص می کند.

شاخص های رابطه برای متغیرهای اسمی (Nominal)

برای متغیرهای اسمی دو نوع شاخص رابطه وجود دارد. اول شاخص های رابطه بر اساس آماره کای دو و دوم شاخص های کاهش نسبی خطا که به آنها PRE می گویند.

شاخص های رابطه بر مبنای کای دو عبارت انداز:

– ضریب فی (phi Coefficient)

– ضریب توافق (Coefficient of Contingency)

– کرامر (Cramer's V)

– هر سه شاخص قدرت رابطه یکسانی را نشان می دهند، اما تفسیر شاخص هایی که بر مبنای کای دو محاسبه می شوند، مشکل است.

شاخص های مزبور جهت رابطه را بیان نمی کنند زیرا متغیرهای اسمی ترتیبی ندارند و صحبت از رابطه منفی یا مثبت در مورد آنها معنی ندارد.

شاخص های رابطه برای متغیرهای اسمی (ادامه)

👉 شاخص های رابطه بر مبنای کاهش نسبی در خطا

– لامدا (λ) پرکاربردترین شاخص PRE است.

– مقدار صفر برای لامدا به این معنی است که متغیر مستقل کمکی در پیش بینی متغیر وابسته نمی کند.

– هنگامی که دو متغیر از نظر آماری مستقل هستند

مقدار لامدا برابر صفر است، اما صفر بودن لامدا لزوماً به معنی مستقل بودن نیست. بنابراین، هنگامی که لامدا برابر صفر است سایر شاخص های رابطه ممکن است مقدار متفاوتی را پیدا کنند.

شاخص های رابطه برای مقیاس های رتبه ای (Ordinal)

لاندای به عنوان یک شاخص رابطه در تمام متغیرها با هر مقیاس قابل استفاده است.

گامای گودمن و کروسکال (Goodman and Kruskal's Gamma) گامای مثبت معرف رابطه مثبت و برعکس است.

تاو b-کندال (Kendal's tau-b) که می تواند مقادیری بین -1 تا +1 داشته باشد.

تاو C-کندال (Kendal's tau-c)

دسامرز (Somers's d)

شاخص های رابطه بر مبنای همبستگی

👉 هنگامی که متغیرهای شما در مقیاسی اندازه گیری شده اند که رتبه در آن معنا دارد، می توان با محاسبه ضرایب همبستگی قدرت رابطه دو متغیر را اندازه گیری کرد.

👉 دو ضریب همبستگی پر کاربرد عبارت انداز:

- ضریب همبستگی پیرسون که دارای مقادیری بین -1 تا $+1$ و قدرت رابطه خطی بین متغیرهای فاصله ای را بیان می کند.

- ضریب همبستگی اسپیرمن که مترادف غیر پارامتری ضریب همبستگی پیرسون است. در ضمن، این ضریب قدرت رابطه متغیرهای رتبه ای را اندازه گیری می کند.

ضرایب همبستگی پیرسون و اسپیرمن

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval Pearson's R	.219	.028	7.070	.000 ^c
Ordinal by Ordinal Spearman Correlation	.208	.030	6.695	.000 ^c
N of Valid Cases	995			

- Not assuming the null hypothesis.
- Using the asymptotic standard error assuming the null hypothesis.
- Based on normal approximation.

دستور:

Analyze → Descriptive Statistics → Crosstab..

Statistics ← انتخاب متغیرهای مورد نظر در ردیف و ستون

Correlations → Continue → OK

دستور محاسبه همبستگی بین متغیرها

Analyze → Correlate → Bivariate..

انتقال دو متغیر مورد نظر به محل Variables

انتخاب روش محاسبه همبستگی با توجه به نوع متغیر:

1. متغیرهای فاصله ای: پیرسون

2. متغیرهای رتبه ای: اسپیرمن

3. متغیرهای رتبه ای: کندال

OK

مقایسه ضرایب همبستگی پیرسون و اسپیرمن

Correlations

		Is Life Exciting or Dull	R's Highest Degree
Is Life Exciting or Dull	Pearson Correlation	1.000	.219**
	Sig. (2-tailed)	.	.000
	N	997	995
R's Highest Degree	Pearson Correlation	.219**	1.000
	Sig. (2-tailed)	.000	.
	N	995	1496

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

Correlations

			Is Life Exciting or Dull	R's Highest Degree
Spearman's rho	Is Life Exciting or Dull	Correlation Coefficient	1.000	.208**
		Sig. (2-tailed)	.	.000
		N	997	995
	R's Highest Degree	Correlation Coefficient	.208**	1.000
		Sig. (2-tailed)	.000	.
		N	995	1496

** . Correlation is significant at the .01 level (2-tailed).

همبستگی جزئی (Partial Correlation)

در صورتی که بخواهیم در تعیین همبستگی دو متغیر اثر دیگر متغیرها را کنترل کنیم، می توان از رویه همبستگی جزئی استفاده کرد.

اگر می خواهید همبستگی جزئی دو به دو متغیرها را اندازه گیری کنید، گزینه Zero-order correlations را انتخاب کنید.

Analyze → Correlate.. → Partial..

انتقال دو یا چند متغیر مورد نظر به محل Variables و دیگر متغیرهای کنترل به محل Controlling for

انتخاب گزینه Zero-order correlations از روی Options

OK

رگرسیون و همبستگی

☞ آنالیز رگرسیون زمانی استفاده می شود که متغیر وابسته دارای مقادیر متعدد است و در مقیاس فاصله ای یا نسبی اندازه گیری شده باشد.

☞ در بررسی رابطه دو متغیر قدم اصلی رسم نموداری از مقادیر است. با مشاهده یک نمودار می توانید بگویید آیا بین مقادیر دو متغیر الگویی مشاهده می شود.

☞ اگر نقاط حول یک خط راست جمع شده باشند، بین دو متغیر رابطه خطی وجود دارد. اگر ضریب زاویه خط مثبت باشد، رابطه "مثبت" خواهد بود. اگر ضریب زاویه خط منفی باشد، رابطه مشاهده شده "منفی" است.

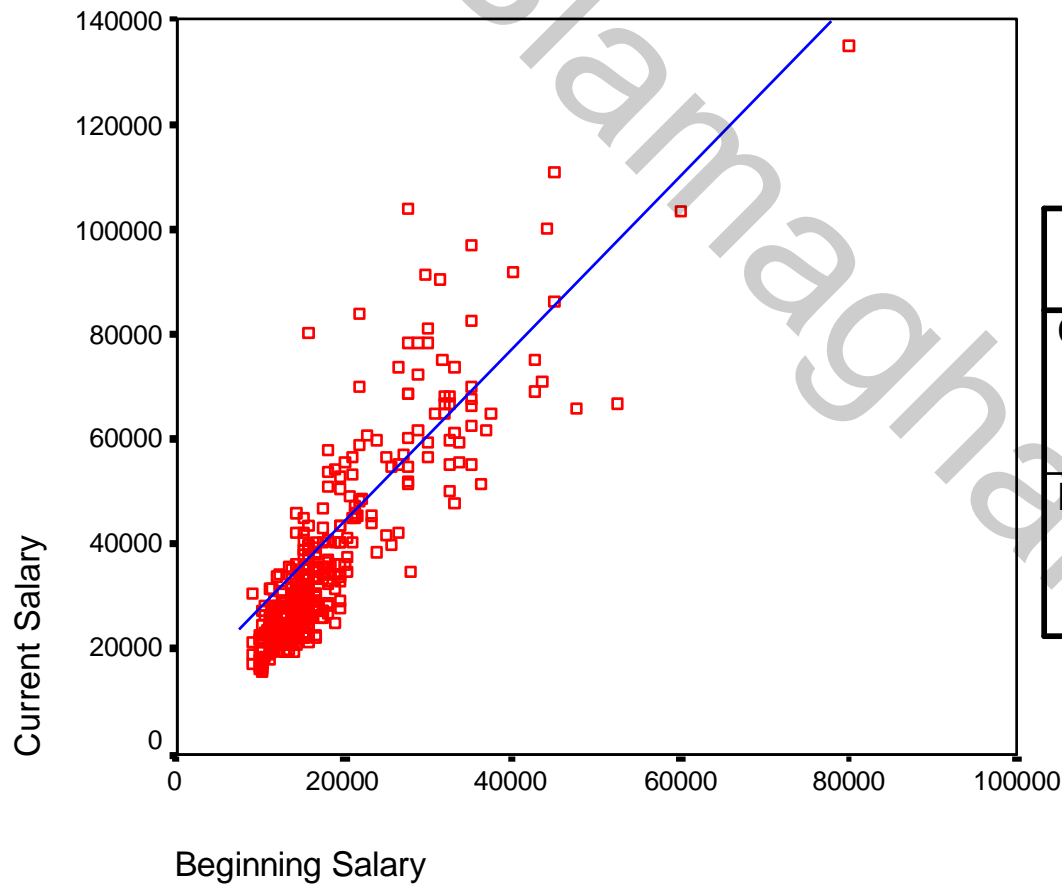
رگرسیون و همبستگی (ادامه)

👉 قدر مطلق ضریب همبستگی پیرسون بین دو متغیر بیانگر آن است که چقدر نقاط حول يك خط صاف تجمع پیدا کرده اند.

👉 هر دو مقدار بزرگ مثبت (نزدیک $+1$) و بزرگ منفی (نزدیک -1) بیانگر آن است که رابطه خطی قوی بین دو متغیر وجود دارد یا در حقیقت نقاط به خط رگرسیون نزدیک هستند.

👉 اگر رابطه خطی بین دو متغیر وجود نداشته باشد، مقدار ضریب همبستگی پیرسون نزدیک صفر خواهد بود، اما ضریب همبستگی صفر به این معنی نیست که هیچ نوع رابطه بین دو متغیر وجود ندارد.

ضریب همبستگی و نمودار رابطه دو متغیر



Correlations

	Current Salary	Beginning Salary
Current Salary Pearson Correlation	1.000	.880*
Sig. (2-tailed)	.	.000
N	474	474
Beginning Salary Pearson Correlation	.880*	1.000
Sig. (2-tailed)	.000	.
N	474	474

** Correlation is significant at the 0.01 level (

دستور:

Simple

Scatter..

Graphs

Define

انتقال دو متغیر به محل محورهای X و Y

OK

روش به دست آوردن رگرسیون خطی

Analyze → Regression → Linear..

انتقال متغیر فاصله ای وابسته به محل Dependent

انتقال متغیرهای مستقل به محل Independent

انتخاب روش مناسب مانند Enter یا Stepwise

OK

خروجی تحلیل رگرسیونی

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.912 ^a	.832	.831	\$7,024.15

a. Predictors: (Constant), Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category

b. Dependent Variable: Current Salary

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.1E+11	4	2.9E+10	581.575	.000 ^a
	Residual	2.3E+10	469	4.9E+07		
	Total	1.4E+11	473			

a. Predictors: (Constant), Previous Experience (months), Beginning Salary Educational Level (years), Employment Category

b. Dependent Variable: Current Salary

خروجی تحلیل رگرسیونی (ادامه)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3068.271	1782.508		-1.721	.086
	Educational Level (years)	601.303	155.934	.102	3.856	.000
	Employment Category	5930.283	640.029	.269	9.266	.000
	Beginning Salary	1.342	.070	.618	19.035	.000
	Previous Experience (months)	-19.031	3.327	-.117	-5.720	.000

a. Dependent Variable: Current Salary

برای انجام این تحلیل رگرسیونی از فایل Employee data استفاده کنید.

شرایط استفاده از رگرسیون

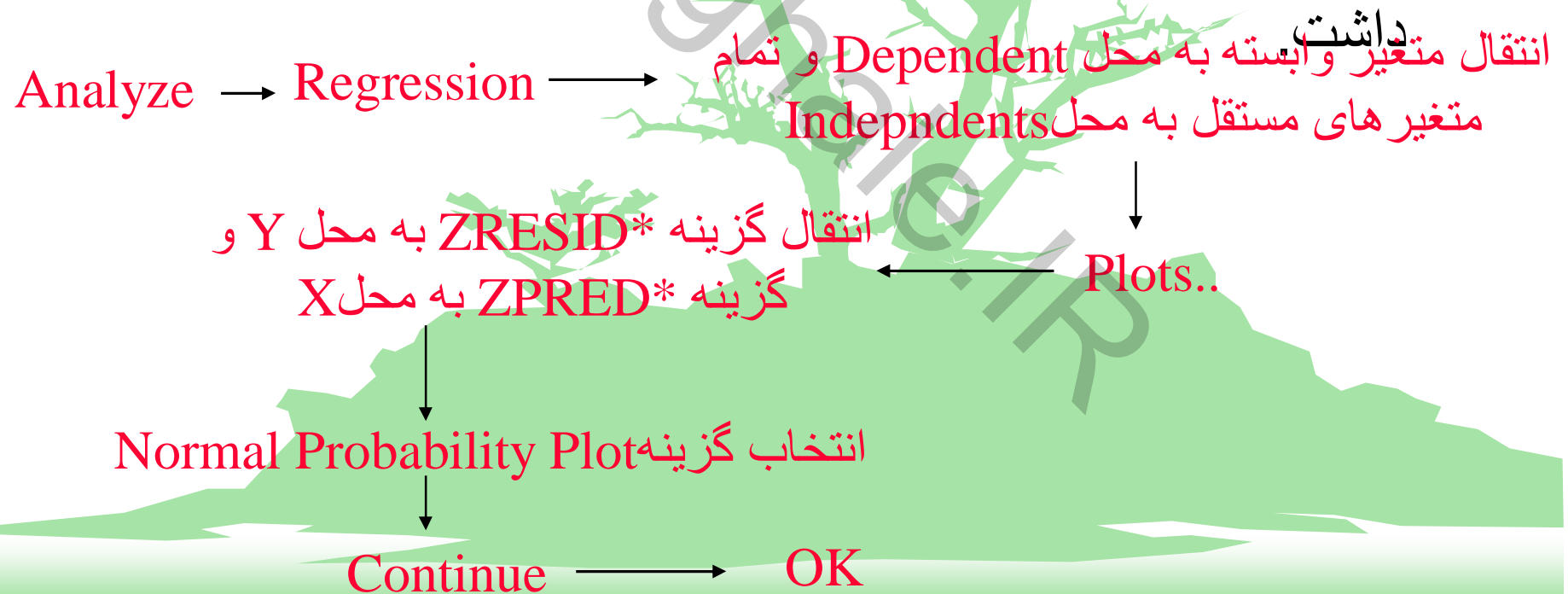
1. تمام مشاهدات باید مستقل باشند. (Independence)
2. برای تمام مقادیر متغیرهای مستقل، توزیع مقادیر متغیر وابسته باید نرمال باشد. (Normality)
3. واریانس توزیع متغیر وابسته باید برای تمام مقادیر متغیرهای مستقل یکسان باشد. (Constant Variance)
4. رابطه بین متغیر وابسته و متغیرهای مستقل باید در جامعه خطی باشد. (Linearity)
5. رابطه خطی مشترک چند گانه بین متغیرهای وابسته برقرار نباشد. (Multicollinearity)

در ضمن، متغیرها باید حداقل در مقیاس رتبه ای اندازه گیری شده باشند.

بررسی نرمال بودن

اگر در جامعه آماری متغیر وابسته برای هر يك از مقادير متغیر های مستقل به صورت نرمال توزیع شده است، پس توزیع باقیمانده ها (residuals) نیز باید به تقریب نرمال باشد.

برای سنجش نرمال بودن باقیمانده ها باید در بخش متغیرهای مستقل تمامی آنها را وارد کنیم و در نهایت فقط يك نمودار خواهیم



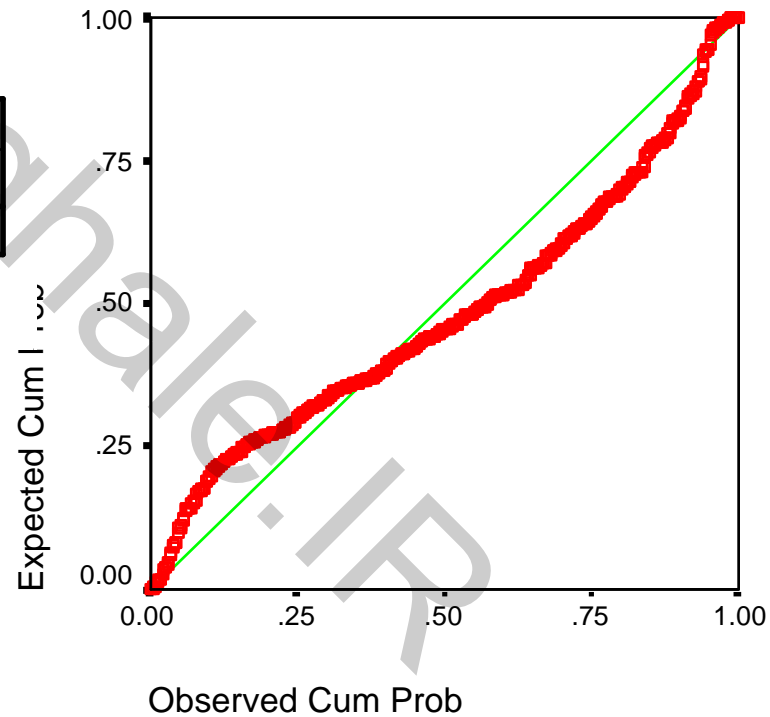
بررسی نرمال بودن) ادامه

Normal P-P Plot of Regression Standardized Residuals
Dependent Variable: Current Salary

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
Current Salary	.208	474	.000

a. Lilliefors Significance Correction



Normality plot with tests

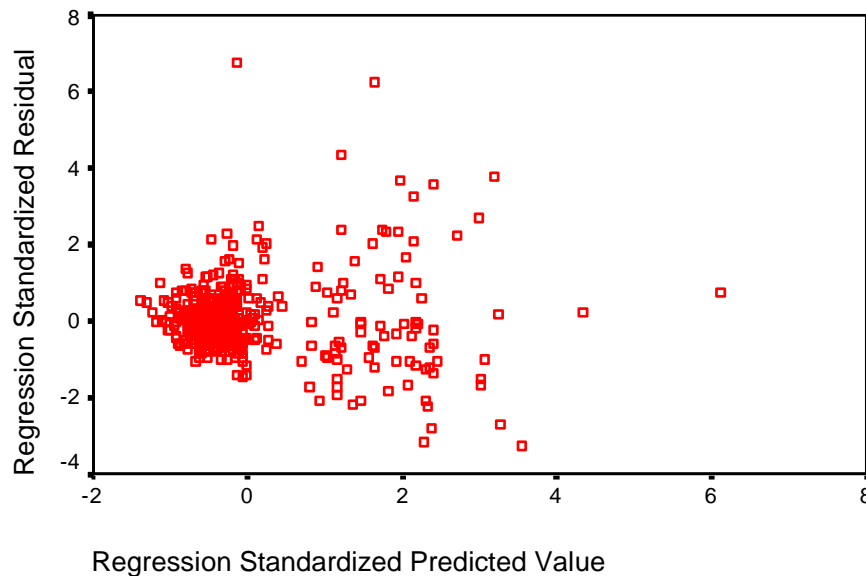
Explore → Plots

بررسی ثابت بودن واریانس

برای سنجش اینکه آیا واریانس متغیر وابسته برای تمام مقادیر متغیرهای مستقل ثابت است، می توان از نمودار باقیمانده ها (Residuals) را در مقابل مقادیر پیش بینی شده (Predicted values) استفاده کرد.

Scatterplot

Dependent Variable: Current Salary



اگر واریانس ثابت باشد الگویی در نقاط مربوط به داده ها مشاهده نخواهید کرد.

در روش بررسی نرمال بودن نیز این نمودار ظاهر می شود

اگر واریانس متغیر وابسته ثابت به نظر نرسد چه باید کرد؟

اگر واریانس متغیر وابسته با افزایش مقدار متغیر مستقل به طور خطی افزایش پیدا کند، با جذر گرفتن از متغیر وابسته آن را تبدیل کنید.

اگر انحراف معیار متغیر وابسته با افزایش مقادیر متغیر وابسته به صورت خطی افزایش می یابد، با لگاریتم گرفتن آن را تبدیل کنید و سپس رگرسیون را با متغیر تبدیل شده اجرا کنید.

انتخاب نام مناسب برای متغیری که می خواهید
تبدیل کنید در محل Target variable

انتقال تابع تبدیل به محل Numeric expression

OK

انتقال متغیر مورد نظر به محل علامت؟

بررسی خطی بودن

باید ثابت کنیم متغیر وابسته با متغیرهای مستقل به طور خطی رابطه دارد.

اگر در نمودار پراکنندگی الگوی خاصی تشکیل نشود، روابط بین دو متغیر

خطی خواهد بود. در غیر اینصورت باید از طریق انتقال يك متغیر با

استفاده از توابعی مانند Log ، جذر گرفتن، Tan ، و مانند اینها بین آنها

رابطه خطی برقرار کرد. به عنوان مثال ممکن است رابطه خطی به شکل

زیر درآید $Y = a + b \text{Log } x$:



Scatter.. ← Graphs

Matrix → Define

انتقال کلیه متغیرها اعم از وابسته و مستقل به

محل Matrix variables

OK

بررسی هم خطی بودن

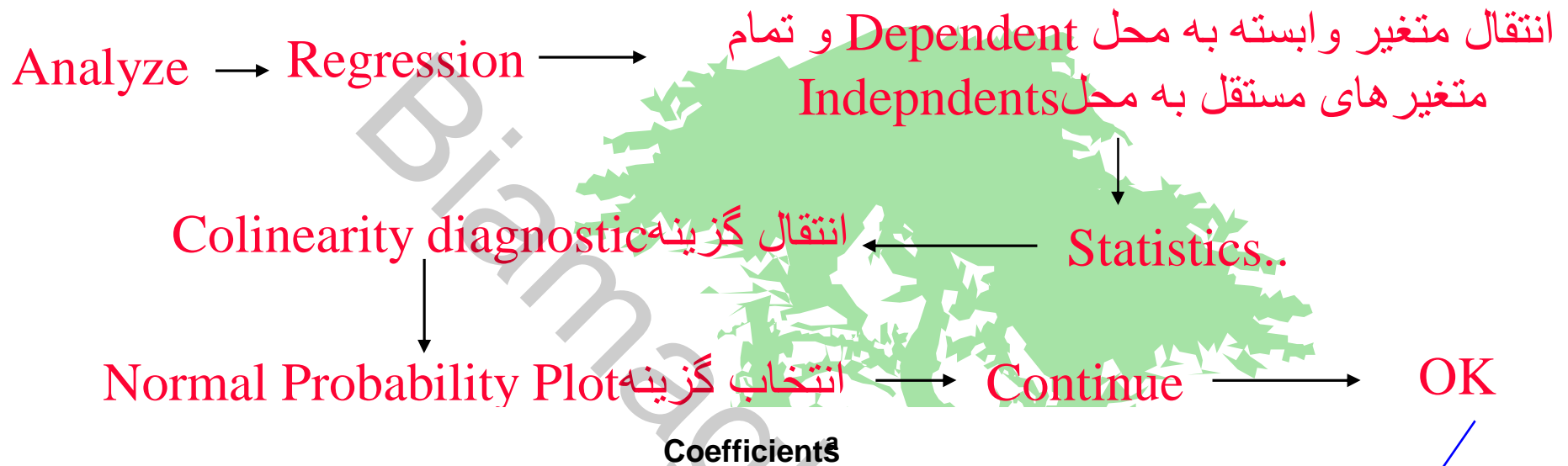
☞ قدرت رابطه خطی بین متغیرهای مستقل توسط شاخصی اندازه گیری می شود که تولرانس نامیده می شود. مقدار آن بین صفر و ۱ تغییر می کند.

☞ برای هر متغیر مستقل تولرانس نسبتی از پراکندگی آن متغیر است که توسط روابط خطی آن متغیر با دیگر متغیرهای مستقل در مدل توجیه نمی شود.

☞ مقدار نزدیک به صفر (کمتر از ۱) معرف آن است که یک متغیر مستقل تقریباً یک ترکیب خطی از سایر متغیرهای مستقل است. یعنی داده ها دارای روابط خطی مشترک چند گانه (multicollinear) هستند.

☞ مقدار نزدیک به ۱ معرف آن است که متغیر مورد نظر تابع خطی از دیگر متغیرها نیست.

بررسی هم خطی بودن (ادامه)



Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3068.271	1782.508		-1.721	.086		
	Educational Level (years)	601.303	155.934	.102	3.856	.000	.515	1.940
	Employment Category	5930.283	640.029	.269	9.266	.000	.426	2.348
	Beginning Salary	1.342	.070	.618	19.035	.000	.339	2.950
	Previous Experience (months)	-19.031	3.327	-.117	-5.720	.000	.861	1.161

a. Dependent Variable: Current Salary

$$\text{Variance Information Function} \equiv \text{VIF} = 1/\text{Tolerance}$$

برآزش مدل رگرسیونی

(Goodness of fit of the model)

👉 در خروجی رگرسیون ضریب تعیین (R^2 R square) معرف آن است که چه نسبتی از تغییر پذیری یا واریانس کلی در متغیر وابسته توسط متغیرهای مستقل تبیین می‌شود.

👉 اگر R^2 مساوی ۱ باشد نشان دهنده آن است که متغیر وابسته کاملاً از طریق متغیرهای مستقل قابل پیش بینی است. مقدار صفر معرف آن است که متغیر وابسته به طور خطی به متغیرهای وابسته مرتبط نیست.

👉 **Multiple R** قدر مطلق ضریب همبستگی بین مقادیر مشاهده شده متغیر وابسته و مقادیر پیش بینی شده (Predicted) بر اساس مدل رگرسیونی متغیر وابسته را نشان می‌دهد.

آزمون فرضیه های رگرسیون

آزمون صفر بودن ضریب زاویه

برای اثبات اینکه رابطه خطی بین متغیر وابسته و متغیرهای مستقل وجود دارد، می توان این فرض صفر را آزمون کرد که ضریب زاویه جامعه برابر با صفر است.

جدول آنالیز واریانس برای آزمون فرض صفر به کار می رود) هیچ رابطه خطی بین متغیر وابسته و متغیرهای مستقل در جامعه آماری وجود ندارد).

اگر سطح معنی داری کمتر از $05/$ باشد می توان فرض صفر را که هیچ رابطه خطی بین متغیر وابسته و متغیرهای مستقل وجود ندارد را رد کرد.

آزمون صفر بودن ضریب زاویه

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.1E+11	4	2.9E+10	581.575	.000 ^a
	Residual	2.3E+10	469	4.9E+07		
	Total	1.4E+11	473			

a. Predictors: (Constant), Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category

b. Dependent Variable: Current Salary

چون سطح معنی داری مشاهده شده کوچک می باشد، لذا فرضیه صفر مبنی بر اینکه رابطه خطی بین متغیر وابسته و متغیرهای مستقل وجود ندارد، رد می شود.

آزمون فرضیه های رگرسیون

آزمون صفر بودن ضرایب رگرسیونی

مقادیر ضرایب رگرسیونی (Regression coefficient) در ستونی در ستونی با عنوان B ظاهر می شوند. این ضرایب بهترین تخمین ما از ضرایب رگرسیونی در جامعه آماری است.

برای تعمیم ضرایب رگرسیون به جامعه باید این فرضیه صفر را تست کنیم که آیا این ضرایب در جامعه صفر هستند

$$(H_0: \beta_1 = \beta_2 = \beta_3 = 0)$$

اگر سطح معنی داری مشاهده شده کوچک باشد، فرضیه صفر یعنی صفر بودن ضرایب رگرسیونی در جامعه رد می شود.

آزمون صفر بودن ضرایب رگرسیونی

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3068.271	1782.508		-1.721	.086
	Educational Level (years)	601.303	155.934	.102	3.856	.000
	Employment Category	5930.283	640.029	.269	9.266	.000
	Beginning Salary	1.342	.070	.618	19.035	.000
	Previous Experience (months)	-19.031	3.327	-.117	-5.720	.000

a. Dependent Variable: Current Salary

با توجه به پایین بودن سطح معنی داری مشاهده شده، فرضیه صفر (صفر بودن ضرایب رگرسیونی رد می شود).

معادله رگرسیونی

☞ در صورت صادق بودن شرایط رگرسیونی، تحلیل نهایی از خروجی رگرسیون باید در شکل يك مدل ریاضی (معادله رگرسیونی) ارائه شود. مثلاً با توجه به مثال فوق خواهیم داشت:

$$\text{Current Salary} = 1.34 \text{ Beginning salary} + 5930.28 \text{ Employment Category} - 19.03 \text{ Previous Experience} + 601.30 \text{ Educational Level} - 3068.27$$

☞ نظر به اینکه در معادله فوق به دلیل همسنگ نبودن ضرایب رگرسیونی، میزان تاثیرگذاری نسبی متغیرهای مستقل مشخص نیست، ضروری است از ضرایبی که در ستون Beta ذکر شده اند، استفاده شود.

☞ مثلاً ۶۲ درصد از تغییر پذیری در متغیر وابسته توسط حقوق اولیه تبیین می شود. ۲۷ درصد از تغییر پذیری در متغیر وابسته توسط طبقه شغلی تعیین می شود. و به همین قیاس در

میزان تاثیر هر یک از متغیرها بر تغییرات در

تعاملی برای اثرات رگرسیونی تحلیل

Regression Analysis for Interactive Effects

فرض کنید می خواهیم اثر متغیر سن (Age) را به عنوان يك متغیر
تعدیل کننده (Moderator) در رابطه خطی $SR = \alpha + \beta DP + e$
بسنجیم. به کمک تحلیل رگرسیونی رابطه خطی دو متغیر را می
سنجیم و R Square آن را یادداشت می کنیم.

متغیر Age را اگر به صورت پیوسته نباشد) به صورت متغیر
مصنوعی (Dummy) تعریف می کنیم. مثلاً اگر Age در سه طبقه
جوان (1) ، میانسال (2) ، و پیر (3) طبقه بندی شده باشد به کمک
دستور Recode متغیرهای مصنوعی جدید را به صورت زیر
تعریف می کنیم:

1	→	0	
2	→	1	→ Age1
3	→	0	

1	→	0	
2	→	0	→ Age2
3	→	1	

تحلیل رگرسیونی برای اثرات تعاملی (ادامه)

👉 سپس به کمک دستور Compute دو متغیر جدید زیر را تعریف می کنیم:

DP*Age1 DP*Age2

👉 حال باید رابطه خطی زیر را از طریق رگرسیون مورد آزمایش قرار دهیم:

$$SR = \alpha + \beta_1 DP + \beta_2 Age1 + \beta_3 Age2 + \beta_4 Age1 * DP + \beta_5 Age2 * DP + e$$

👉 اگر ضراب رگرسیونی از نظر آماری معنی دار باشند، در این صورت

فرض آماری صفر که هیچ اثر تعاملی وجود ندارد رد می شود. مقدار

ضرایب رگرسیونی اطلاعاتی را در خصوص ماهیت تعامل بیان می کند .

R Square جدید نیز باید بیشتر از R Square قبلی (محاسبه شده در

مرحله اول باشد) تا بتوان گفت متغیر Age به عنوان يك متغیر مداخله گر

عمل می کند.

تحلیل رگرسیونی برای اثرات تعاملی (ادامه)

در صورتی که متغیر تعدیل کننده يك متغیر پیوسته باشد، دیگر نیازی به محاسبه متغیرهای مصنوعی (Dummy) نیست. مثلاً نخست رابطه خطی زیر را به کمک رگرسیون برآورد می کنیم:

$$SR = \alpha + \beta_1 DP + \beta_2 RO + \beta_3 LS + \beta_4 EH + \beta_5 EF + e$$

سپس به کمک دستور Compute متغیرهای زیر را تعریف می کنیم:

$$DP * Age, RO * Age, LS * Age, EH * Age, EF * Age$$

آنگاه از طریق رگرسیون با در نظر گرفتن متغیرهای جدید به عنوان متغیر مستقل رابطه خطی جدید را برآورد می کنیم. اگر ضریب تعیین بیشتر باشد ($R \text{ Square Change} > 0$) و ضرایب رگرسیونی معنی دار باشند، متغیر Age را می توان به عنوان متغیر تعدیل کننده مطرح ساخت.

تحلیل رگرسیونی برای اثرات تعاملی (ادامه)

در تحلیل رگرسیونی برای اثرات تعاملی بهتر است یافته های پژوهش در قالب جدولی مانند زیر خلاصه شود.

	R Square	R Square Change	Beta
Age*DP			
Age*RO			
Age*LS			
Age*EH			
Age*EF			

تحليل عاملی (Factor analysis)

👉 تحليل عاملی يك تكنيك آماری است که برای تشخیص حداقل عواملی که می توانند يك پدیده پیچیده را تشریح کنند.

👉 هدف تحليل عاملی تشخیص عوامل ناپیدا بر مبنای عوامل مشاهده شده است. این عوامل باید ساده و در عین حال معنی دار باشند.

👉 هدف تحليل عاملی پیدا کردن عوامل مشترکی است که متغیرها را بتوان به صورت تابعی از آنها نوشت.

👉 تحليل عاملی ابزاری است برای سنجش روائی پرسش نامه. مقصود آن است که پرسش نامه چیزی را اندازه بگیرد که برای اندازه گیری آن ساخته شده است.

تحلیل عاملی (Factor analysis)

عامل متغیر جدیدی است که از طریق ترکیب خطی نمره های اصلی متغیرهای مشاهده شده برپایه

فرمول زیر برآورد می شود: $F_i = \sum W_{ij} X_j$

امید است که تعداد کمی از عامل ها (یعنی ترکیب خطی نمره های اصلی متغیرهای مشاهده شده) بتواند تقریباً همه اطلاعاتی را که توسط مجموعه بزرگتری از متغیرها به دست می آید دربرگرفته در نتیجه توصیف ویژگی های یک پدیده را ساده سازد.

شرایط مناسب ماتریس داده ها برای تحلیل عاملی

👉 حجم گروه نمونه :

- کامری (۱۹۷۳) گروه نمونه ۱۰۰ نفری را ضعیف، ۲۰۰ نفری را نسبتاً مناسب، ۳۰۰ نفری را خوب، ۵۰۰ نفری را خیلی خوب، و ۱۰۰۰ نفری را عالی می داند.
- حجم گروه نمونه باید دست کم ۱۰ یا ۲۰ برابر عده متغیرهایی باشد که در ماتریس همبستگی وارد می شود.
- برای اطمینان از کفایت نمونه از شاخص KMO استفاده می شود.
- مقادیر کوچک KMO بیانگر آن است که همبستگی بین زوج ها نمی تواند توسط متغیرهای دیگر تبیین شود. کیزر (Kaiser, 1977) معتقد است که KMO بزرگتر از 9/ عالی، در دامنه 8/ شایسته، در دامنه 7/ بهتر از متوسط، در دامنه 6/ متوسط، در دامنه 5/ بدبختی و پایین تر از آن غیر قابل قبول است.

شرایط مناسبیت ماتریس داده ها برای تحلیل عاملی

👉 معنا دار بودن ماتریس: آیا ماتریس داده ها شامل اطلاعات مفید و با معنایی است یا نه؟

- اگر همبستگی بین متغیرها کوچک باشد، احتمال آنکه در عامل های مشترک سهم باشند، اندک است.
- برای ارزشیابی ماتریس همبستگی از آزمون کرویت بارتلت (Bartlett's test of sphericity) استفاده می شود.
- آزمون کرویت برای رد فرضیه صفر مبنی بر درست بودن ماتریس همانی در جامعه است (یعنی این فرضیه که ماتریس همبستگی مشاهده شده متعلق به جامعه ای با متغیرهای ناهمبسته است).
- ماتریس همانی (Identity matrix) ماتریسی است که همه عناصر قطری آن ۱ و همه نصار غیر قطری آن صفر باشد.

شرایط مناسبیت ماتریس داده ها برای تحلیل عاملی

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.834
Bartlett's Test of Sphericity	Approx. Chi-Square	2367.490
	df	435
	Sig.	.000

مراحل تحلیل عاملی

تشکیل ماتریس همبستگی: (Correlation matrix) ماتریس همبستگی به ما کمک می کند تا عوامل مشترك را شناسایی کنیم. اگر همبستگی بین متغیرها کوچک باشد، غیرمحمتمل است که آنها در عوامل مشترك باهم سهیم باشند. آزمون کرویت بارتلت می تواند برای آزمایش این فرضیه صفر به کار گرفته شود که ماتریس همبستگی يك ماتریس واحد (Identity) است. ماتریس واحد به ماتریسی گفته می شود که تمام اعداد قطر آن برابر 1 و جوانب آن صفر باشد. اگر فرض صفر مبنی بر اینکه ماتریس همبستگی جامعه آماری يك ماتریس واحد است، رد نشود. در این صورت نمی توان از تحلیل عاملی استفاده کرد.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

مراحل تحلیل عاملی (ادامه)

احصاء عوامل: (Factor Extraction) برای تخمین اولیه از تحلیل مولفه های اصلی استفاده می شود. در این روش ترکیب خطی متغیرهای مشاهده شده تشکیل می شود. روش مولفه های اصلی تنها کاری که انجام می دهد عبارت است از انتقال متغیرهای همبسته به مجموعه ای از متغیرهای غیر همبسته. در این مرحله تعداد عوامل بر حسب ارزش ویژه (Eigenvalue) تعیین می شود. عواملی که دارای ارزش ویژه بزرگتر از یک هستند در مدل عاملی وارد می شوند. ارزش ویژه نشان می دهد که یک عامل تاچه حد در تبیین واریانس مشترك زیربنایی متغیرها سهم است. اگر عاملی تمام واریانس مشترك موجود در هر متغیر را تبیین کند هرکدام دارای بار عاملی ۱ و مقدار ارزش ویژه برای عامل ۱ برابر) n تعداد متغیرها (خواهد بود.

درصدی از کل واریانس را که به وسیله یک عامل تعیین می شود را با pct of var نشان می دهند. برپایه ملاک کیسر (۱۹۶۰) تعداد عامل های مشترك باید مساوی با تعداد مولفه های اصلی دارای ارزش ویژه بزرگتر از ۱ باشد.

جدول ماتریس عوامل

جدول ماتریس عوامل (Factor matrix) ضرایبی را نشان می دهد

که می توان هر متغیر را به کمک آنها بر حسب عوامل مشترك نوشت.

بخشی از واریانس يك متغیر که توسط عوامل مشترك قابل تبیین است

را اشتراك (Communality) آن متغیر گویند.

مقدار اشتراك می تواند از صفر تا 1 در نوسان باشد. صفر نشان دهنده

آن است که عوامل مشترك هیچ واریانسی را بیان نمی کنند.

عدد ۱ نشان می دهد که تمام واریانس

	F_1	F_2	F_3
V_1	a_{11}	a_{12}	a_{13}
V_2	a_{21}	a_{22}	a_{23}
.
.
.
V_n	a_{n1}	a_{n2}	a_{n3}

متغیر توسط عوامل مشترك قابل تبیین

است. آن قسمت از واریانس که توسط

عوامل مشترك قابل تبیین نیست به

عوامل انحصاری متغیر نسبت داده

می شود.

مراحل تحلیل عاملی (ادامه)

چرخش (Rotation): در این مرحله متغیرها به صورت تابعی از عوامل احصاء شده دسته بندی می شوند. چرخش بارهای عاملی این امکان را فراهم می سازد تا عامل های با معنا به دست آید. نسبتی از واریانس که توسط يك عامل تبیین می شود را به وسیله شاخص اشتراك (Communality) مشخص می کنند. مقدار آن بین صفر و ۱ در نوسان است.

صفر معرف آن است که عوامل مشترك هیچ واریانسی از متغیر را تعیین نمی کند و ۱ معرف آن است که تمام واریانس متغیر به وسیله عوامل مشترك تبیین می شود.

هرچند روش های مختلفی برای چرخش بارهای عاملی وجود دارد، اما روش Varimax عامل هایی را تولید می کند که با مجموعه کوچکتری از متغیرها دارای همبستگی قوی و با مجموعه دیگری از متغیرها دارای همبستگی ناچیز باشد. به همین دلیل عمومیت بسیار دارد و کاربرد آن زیادتر است.

روش های چرخش عامل ها

➡ عامل های چرخش یافته باید متعامد (Orthogonal) یا متمایل (Oblique) باشد.

➡ اگر پژوهشگر بخواهد نتایج حاصل از تحلیل او بهترین برآزش را با داده ها داشته باشد باید از چرخش متمایل استفاده کند. اما اگر علاقه بیشتری به **تعمیم پذیری نتایج** داشته باشد، باید چرخش متعامد را به کار برد.

➡ چرخش متمایل نتایجی به دست می دهد که با داده های گروه نمونه بهترین برآزش را دارد، چرخش متعامد راه حلی به دست می دهد که با داده های گذشته و آینده بهترین برآزش را داراست.

➡ در روش متعامد عامل ها ناهمبسته هستند در حالی که در روش متمایل عامل ها همبستگی دارند.

➡ تفاوت دو روش ناچیز و فاقد معناست. بنابراین، چرخش عامل ها به روش متعامد هم ساده تر و هم قابل درک تر و تفسیر پذیر تر است. **به نظر می**

رسد روش واریماکس نیز بهترین روش چرخش متعامد باشد.

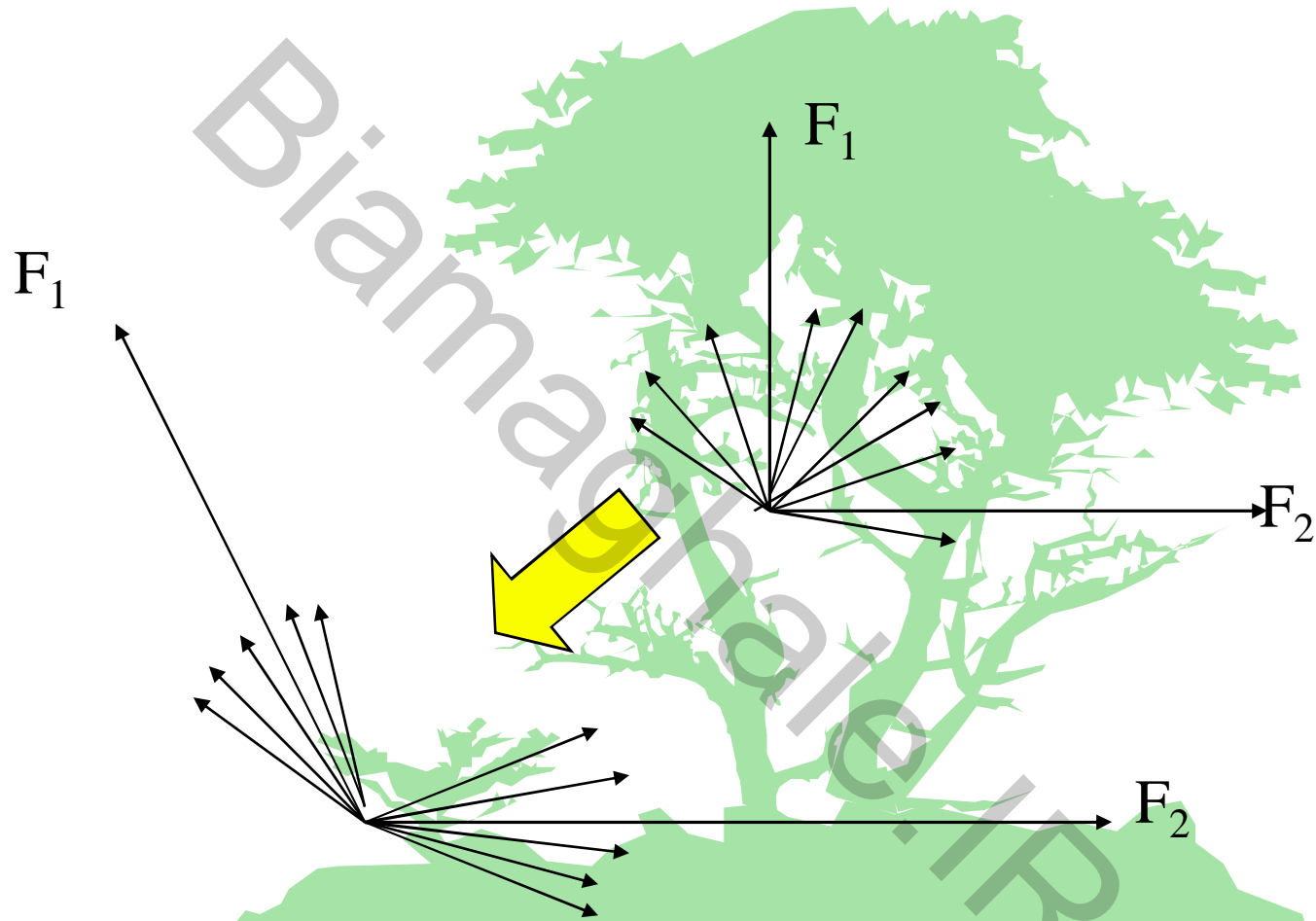
تعداد عامل ها

➡ يك تصميم عمده در اجراى تحليل عاملى مربوط به تعداد عامل ها است.

➡ براى چرخش عامل ها مى توان تعداد معينى از عامل را انتخاب كرد. بنا بر اين، مى توان ساختارهاى چرخش يافته را بر اساس تعداد مختلفى از عامل را با يكديگر مقايسه كرده و مناسب ترين آنها را انتخاب كرد.

➡ هرسازه اى كه داراى كمينه شرايط لازم براى بار عاملى دست كم روى سه متغير نباشد، از تحليل حذف مى شود.
(Santos & Clegg, 1999).

نمایش هندسی چرخش عوامل



www.spowpowerplant.blogfa.com

وبلاگ یک مهندس

جدول کل واریانس تعیین شده

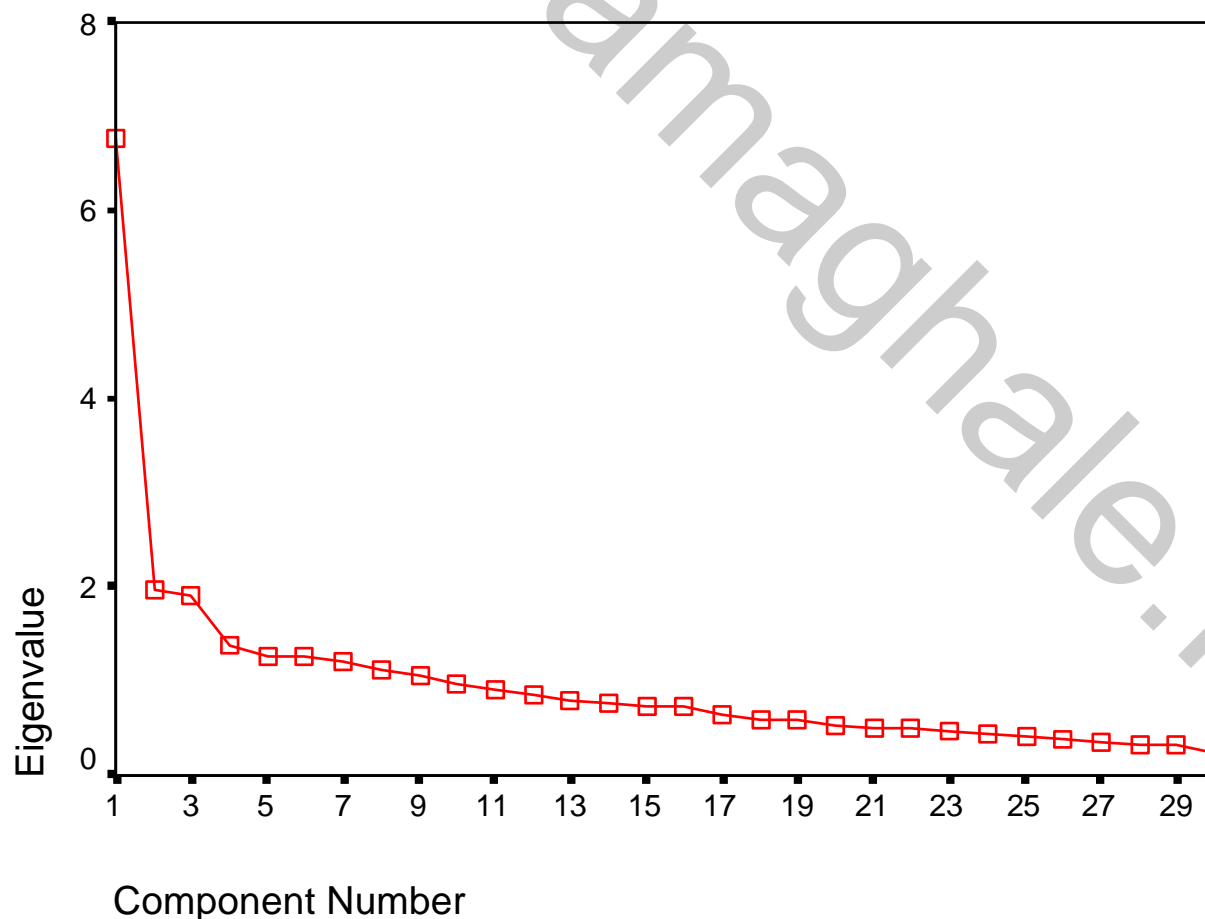
Total Variance Explained

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	6.752	22.506	22.506	6.752	22.506	22.506	3.603	12.009	12.009
2	1.965	6.550	29.055	1.965	6.550	29.055	3.089	10.296	22.305
3	1.902	6.339	35.394	1.902	6.339	35.394	2.751	9.171	31.476
4	1.391	4.637	40.032	1.391	4.637	40.032	2.243	7.477	38.953
5	1.277	4.258	44.290	1.277	4.258	44.290	1.601	5.337	44.290
6	1.252	4.173	48.463						
7	1.205	4.017	52.480						
8	1.116	3.719	56.199						
9	1.065	3.549	59.748						
10	.985	3.282	63.030						
11	.904	3.012	66.042						
12	.850	2.833	68.875						
13	.804	2.680	71.555						
14	.776	2.586	74.141						
15	.737	2.457	76.598						
16	.723	2.410	79.009						
17	.641	2.137	81.146						
18	.601	2.002	83.148						
19	.578	1.928	85.075						
20	.536	1.787	86.862						
21	.511	1.702	88.564						
22	.496	1.654	90.218						
23	.478	1.592	91.810						
24	.428	1.426	93.236						
25	.411	1.369	94.605						
26	.389	1.295	95.900						
27	.348	1.160	97.060						
28	.326	1.088	98.148						
29	.310	1.033	99.182						
30	.245	.818	100.000						

Extraction Method: Principal Component Analysis.

نمودار تشخیص تعداد عامل های معتبر

Scree Plot



در روش اسکری نمودار مقدار ویژه برای هر عامل ترسیم می شود. در نقطه ای که شکل منحنی برای مقادیر ویژه به صورت افقی در آید، آن نقطه اسکری (سنگ ریزه) نامیده می شود و عامل هایی که سمت چپ آن قرار دارند عامل های واقعی و بقیه عامل های خطا فلامداد می شوند.

انواع تحلیل عاملی

👉 تحلیل عاملی اکتشافی: در این روش پژوهشگر هیچگونه فرضیه قبلی در باره نتایج ندارد و در پی اکتشاف عوامل تاثیر گذار است. بنابراین، تحلیل اکتشافی بیشتر به عنوان يك روش تدوین و تولید نظریه و نه آزمون نظریه در نظر گرفته می شود.

👉 تحلیل عاملی تاییدی: در این روش پژوهشگر سعی می کند تاییدی بر يك ساختار عاملی مفروض به دست آورد. یعنی تعیین می کند که داده ها با يك ساختار عاملی معین که در فرضیه آمده هماهنگ است یا نه. تحلیل عاملی تاییدی برای سنجش روایی شاخص های (نشانگرهای) (يك سازه در پرسش نامه نیز به کار گرفته می شود تا معلوم شود هماهنگی و همسویی لازم بین شاخص ها) سؤال ها) وجود دارد.

تفسیر نتایج تحلیل عاملی

Rotated Component Matrix^a

	Component				
	1	2	3	4	5
Q29	.672				
Q28	.602	.307			
Q27	.593	.411			
Q19	.572				
Q23	.572				
Q20	.540				
Q21	.493				
Q26	.480	.300			
Q30	.380			.314	
Q14		.700			
Q18		.685			
Q16		.649			
Q13		.585			
Q17	.348	.555			
Q10			.597		
Q24	.390		.593		
Q11			.567		
Q6			.543		
Q15			.519		
Q4			.487	.385	
Q3			.474		
Q12		.320	.418		
Q5				.734	
Q8				.571	
Q7				.526	
Q22		.310		.410	
Q9				.372	
Q1					.673
Q2					.611
Q25	.309				.567

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 9 iterations.

همان طور که در ماتریس دیده می شود متغیرهای مورد نظر روی ۵ عامل سوار می شوند. با توجه به ویژگی های هر خوشه (گروه) از متغیرهایی که روی یک عامل سوار می شوند، باید نام مناسبی انتخاب شود.

تفسیر نتایج تحلیل عاملی

اگر مجذور بارهای عاملی را مثلاً برای عامل ۱ جمع کنیم، حاصل برابر با $6/75$ یعنی برابر با همان ارزش ویژه این عامل است.

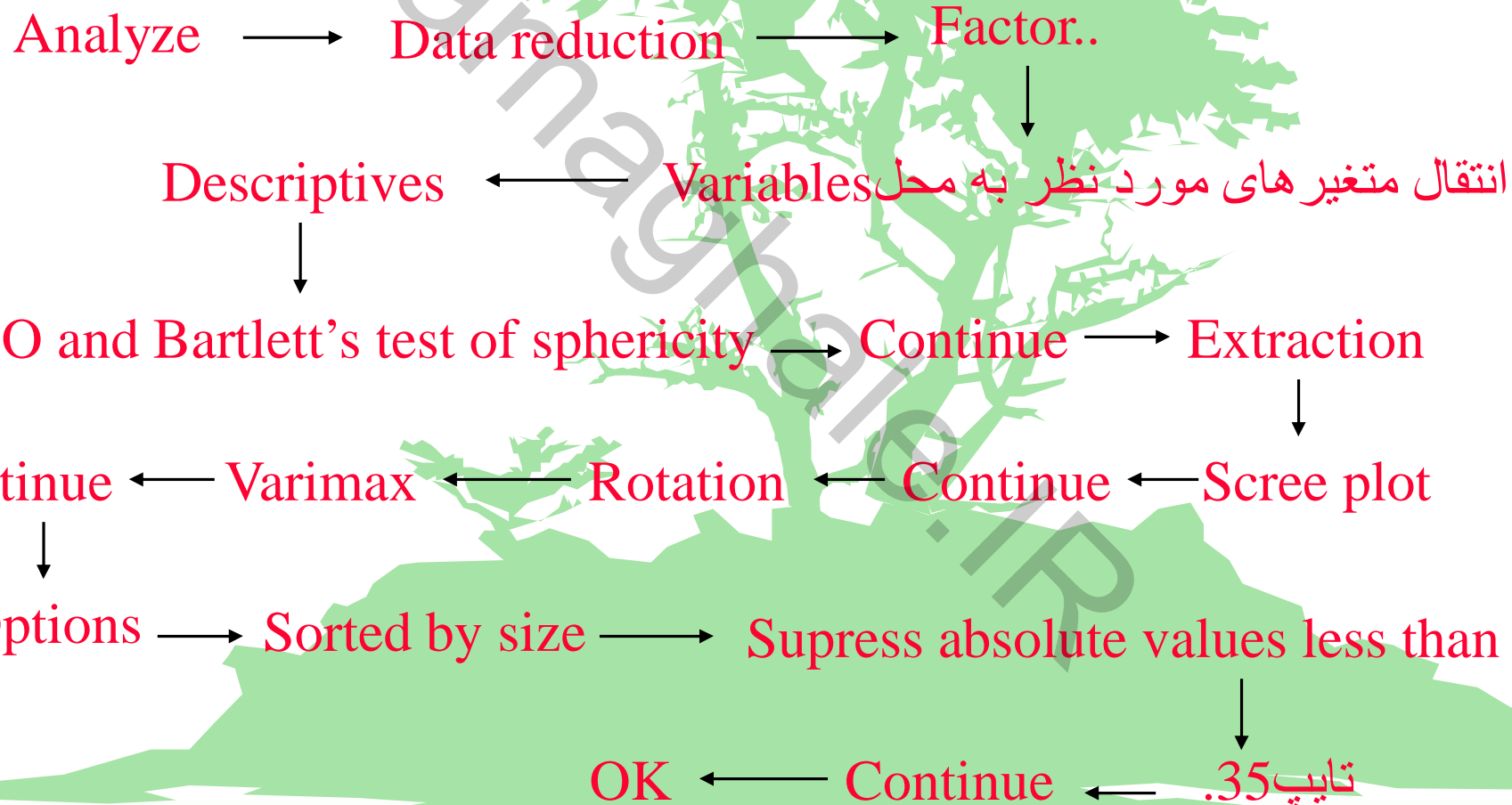
اگر عامل ۱ همه واریانس مشترک موجود در متغیرها را تبیین کند، هر کدام دارای بار عاملی ۱ و مقدار ارزش ویژه برای عامل ۱ برابر با n یعنی تعداد متغیرها خواهد بود.

در خروجی ماتریس چرخش یافته هر عامل معرف يك خوشه روشن و متمایز است و احتمالاً معرف يك سازه با معناست.

تفسیر نتایج تحلیل عاملی

- ➡ بار عاملی نشان دهنده وزن (ضریب بتای) هر عامل است .
- ➡ بارهای عاملی به واقع ضرایب استاندارد شده در یک معادله رگرسیون است که در آن متغیر اصلی به عنوان متغیر وابسته و عامل ها به عنوان متغیرهای مستقل عمل می کند.
- ➡ وجود بارهای منفی نشان می دهد که برخی متغیرها بیانگر عکس چیزی است که توسط آن عامل مشخص می شود .
- بارهای منفی از طریق بیان این مطلب که آن عامل **چه چیز نیست** به روشن ساختن تفسیر کمک می کند.

دستور اجرای تحلیل عاملی



تحلیل خوشه ای (Cluster analysis)

- 👉 تحلیل خوشه ای روشی است برای گروه بندی داده ها. به عبارت دیگر، تحلیل خوشه ای یک تکنیک طبقه بندی برای تشکیل گروه های همگون در مجموعه پیچیده ای از داده هاست.
- 👉 سه هدف عمده یعنی اکتشاف، تایید و ساده سازی را می توان در رابطه با تحلیل خوشه ای مطرح کرد.
- 👉 تحلیل خوشه ای در حقیقت یک آزمون آماری نیست بلکه شامل تعدادی از الگوریتم های طبقه بندی است.
- 👉 در تحلیل خوشه ای گروه بندی چیزها بر اساس شباهت ها یا فاصله ها (عدم شباهت ها) انجام می شود.

روش های عمده خوشه بندی

روش سلسله مراتبی: يك روش متداول در تحلیل خوشه ای است که کاربرد فراوان دارد. در این روش خوشه ها به دو طریق تشکیل می شوند:

- از طریق گروه بندی مشاهده ها به صورت بزرگتر و بزرگتر تا همه مشاهده ها اعضای يك خوشه واحد باشند) خوشه بندی تراکمی.
- مشاهده ها در يك خوشه واحد قرار گرفته و سپس خوشه ها تقسیم می شوند (خوشه بندی بخش پذیر).

روش تراکمی چون خوشه بندی با مشاهده های انفرادی شروع می شود در ابتدا تعداد خوشه ها برابر با تعداد مشاهده هاست.

روش خوشه بندی غیر سلسله مراتبی: در این روش هم متغیرها و هم مشاهده ها به طور همزمان گروه بندی می شوند. روش k-means یکی از این روش هاست.

کاربرد خوشه بندی

خوشه بندی اشخاص (cases)

خوشه بندی متغیرها: این روش مشابه تحلیل عاملی است، اما از لحاظ روشی که متغیرها به گونه گسسته به گروه ها اختصاص می یابند، با آن متفاوت است. تفاوت اساسی در به کار بردن واریانس يك متغیر است. تحلیل عاملی معمولاً واریانس بین چند منبع یا عامل را بخش بندی می کند، در حالی که تحلیل خوشه ای کل واریانس را به يك منبع بنیادی اختصاص می دهد.

پرسش پژوهش در تحلیل خوشه ای

☞ مفهوم سازی اولیه مساله خوشه بندی خیلی مهم است، زیرا استفاده و مقصود نهایی طبقه بندی را تعیین می کند.

☞ اکتشاف: تحلیل خوشه ای می تواند به عنوان يك تكنيك اکتشافی به کار گرفته شود. پرسش کلی آن است که داده های مشاهده شده چگونه به صورت ساختاری مفید و با معنا سازمان داده شود؟

☞ تایید: اگر دانش یا نظریه های پیشین، طبقه بندی خاصی را پیشنهاد کند، خوشه بندی را می توان برای آزمون آن به کاربرد.

☞ ساده سازی: برای ساده کردن يك مجموعه پیچیده از داده ها به کار گرفته می شود.

دستور اجرای تحلیل خوشه ای

Analyze → Classify → Hierarchical cluster Analysis

انتخاب و انتقال متغیرهای استاندارد شده به محل Variables

انتخاب و انتقال متغیری که بر مبنای آن باید خوشه بندی صورت گیرد به محل

Label Cases by

Plots..

Continue

Dendrogram

OK

مراحل تحلیل خوشه ای

👉 انتخاب نمونه برای خوشه سازی نظیر خریداران، بیماران، محصولات، کارمندان

👉 تعریف متغیرهایی که بتوان بر مبنای آنها به اندازه گیری اشیاء، حوادث، یا افراد پرداخت (مانند وضعیت مالی، وابستگی سیاسی، ویژگی های بهره وری).

👉 محاسبه شباهت ها بین موجودیت ها از طریق همبستگی و دیگر تکنیک ها

👉 به حداکثر رساندن شباهت درون خوشه و تفاوت های بین خوشه ای

👉 معتبر سازی و مقایسه خوشه.

تفسیر نتایج

- 👉 در خروجی نمودار سلسله مراتبی مشاهدات را برحسب ویژگی های مشترك نشان می دهد.
- 👉 لازم است برای هر شاخه با توجه به ویژگی های مشترك نام مناسب انتخاب شود. و این کار آنقدر ادامه پیدا کند تا کلیه شاخه ها به طور مناسب ناگذاری شود.
- 👉 شکل گیری دنباله ای خوشه ها و فواصل نسبی آنها در نموداری که Dendogram نامیده می شود نشان داده شده است.
- 👉 همان طور که دیده می شود n خوشه این مجموعه داده ها را توضیح می دهد. اولین خوشه شامل و همگی آنها .. هستند. خوشه شماره 2 شامل است

تحلیل مسیر (Path analysis)

- تحلیل مسیر کاربرد رگرسیون چند متغیری در ارتباط با تدوین بارز مدل های علی (causal models) است.
- هدف تحلیل مسیر به دست آوردن برآوردهای کمی روابط علی بین مجموعه ای از متغیرهاست.
- تحلیل مسیر بیان می کند که کدام مسیر مهمتر و یا معنادارتر است.
- ضرایب مسیر بر اساس ضریب استاندارد شده رگرسیون محاسبه می شود.
- یک متغیر به صورت تابعی از دیگر متغیرها فرض می شود و مدل رگرسیونی آن ترسیم می شود.
- در مرحله دوم متغیر مستقلی که دارای بیشترین ضریب رگرسیونی است) با توجه به ستون (Beta) به عنوان متغیر وابسته در نظر گرفته می شود و معادله رگرسیونی آن برآورد می شود و الی آخر.

انواع روابط بین متغیرها در نمودار تحلیل مسیر

👉 اثر مستقیم: بیانگر آن است که یا X علت Y و یا Y علت X است.

👉 اثر غیر مستقیم: رابطه بین X و Y وقتی غیر مستقیم است که X علت Z است و Z نیز به نوبه خود در Y اثر دارد.

👉 اثر کاذب: رابطه بین X و Y وقتی کاذب (Spurious) است که Z علت هر دو متغیر X و Y باشد.

👉 اثرات تحلیل نشده: رابطه بین دو متغیر وقتی تحلیل نشده است که هر دوی آنها برونزا (exogenous) بوده و بنابراین تبیین تغییر پذیری بین آنها توسط مدل امکان پذیر نباشد.

محدویت های تحلیل مسیر

همه متغیرهای مداخله گر باید در تحلیل رگرسیون چند متغیری به عنوان متغیرهای وابسته عمل کنند. بنابراین، همه آنها باید دارای مقیاس فاصله ای باشند. اندازه های رتبه ای تحلیل مسیر را ناممکن می سازند.

تحلیل مسیر می تواند فرضیه های علی را ارزشیابی کند اما هرگز نمی تواند جهت علیّت را محقق سازد.

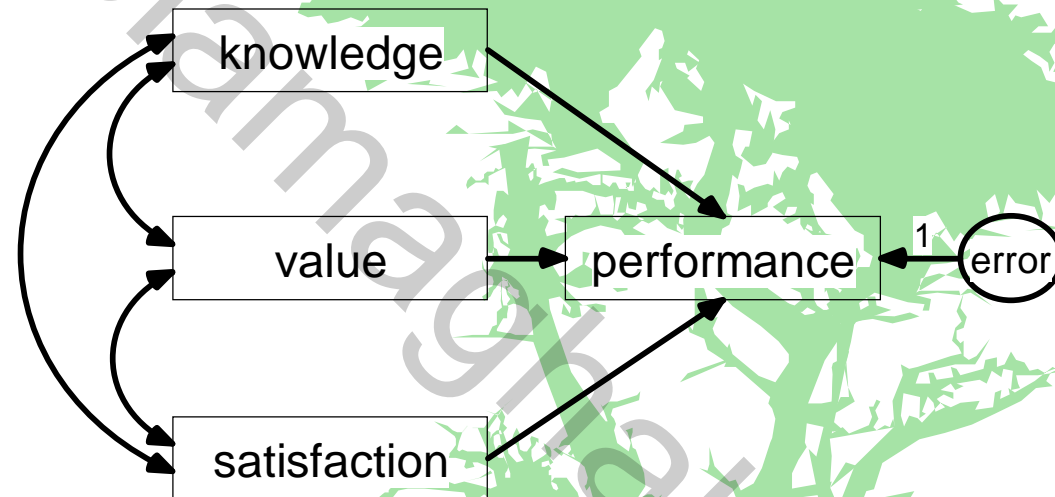
تحلیل مسیر در مرحله اکتشافی قابلیت اندکی دارد.

استنباط رابطه علیت

برای پی بردن به وجود رابطه علیت بین دو متغیر باید شرایط زیر وجود داشته باشد:

- وجود همبستگی بین دو متغیر
- حذف روابط کاذب: اگر تاثیر تمام متغیرها کنترل شود و هنوز رابطه بین دو متغیر (مستقل و وابسته) وجود داشته باشد، می‌گوییم رابطه کاذب نیست. یعنی در مرحله تست همبستگی دو متغیر دیگر متغیرها به عنوان متغیر کنترل انتخاب شوند.
- توالی زمانی (Time order): یعنی علت اول اتفاق بیافتد و بعد معلول حادث شود.

مدل ساختاری در تحلیل مسیر



Example 4
Conventional linear regression
Job performance of farm managers
(Model Specification)