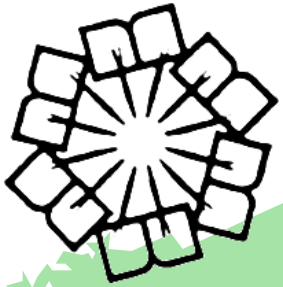
بنام خدا



موسسه عالی آموزش و پژوهش مدیریت و برنامه ریزی

آشنایی باSPSS

مدرس

دکتر فرج اله رهنورد

کارشناسی ارشد مدیریت دولتی

تحلیل آماری(Statistical Analysis)

- هر آنالیز آماری شامل مراحل زیر است:
  - توصیف داده‌ها :نمایش و خلاصه کردن داده ها
  - آزمون فرضیات :نتیجه گیری در مورد گروه بزرگتر) جامعه (از روی اطلاعات به دست آمده در یك 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 نمونه‌ها دارای مقدار واحدی هستند.
- انحراف معیار (std deviation) به پراکندگی مشاهدات در یك نمونه اشاره دارد، در حالیکه خطای معیار (std error) به پراکندگی یك 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.
- صدك ها :(Percentile) مقادیری از متغیر هستند که در مورد درصد نمونه‌ها را نشان می دهند .براي مثال می توانید مقداری را پیدا کنید که 25% نمونه ها در زیر آن قرار گیرند.
- چارك‌ها :(Quartiles) از آنجائی که صدك های25 ، 50، و 75 نمونه را تقریباً به چهار گروه برابر تقسیم می کنند، مجموعاً به آنها چارك‌ها گویند.

Analyze → Descriptive Statistics → Frequencies

↓

انتخاب متغیر

↓

Statistics.. → Quartiles → Continue → OK

طرز محاسبه صدك ها برای هريك از گروه‌ها

Analyze → Descriptive statistics → Explore..

وارد کردن متغیر مورد نظر در محل Dependent List

و متغیری که دارای زیرگروه های مورد نظر است در محل Factor List

Statistics..

Percentiles → Continue → OK

در گروه مذکر 10 %از نمونه ها مقاومت کم (2 دارند)

صدك هاي مقاومت افراد در برابر تغيير

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 می‌دهد.
  - واریانس :(Variance) پرکاربرد ترین شاخص پراکندگی است .فرمول محاسبه واریانس در یك نمونه (S^2) به شکل زیر است:

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

اگر واریانس صفر باشد تمامی نمونه ها دارای مقدار واحدی هستند .هرچه واریانس بیشتر باشد، مقادیر بیشترپراکنده 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 یك مشاهده چند انحراف معیار بیشتر یا کمتر از میانگین می باشد.

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

دستور محاسبه نمره استندارد:

Analyze → Descriptive statistics → Descriptives

انتقال متغیر) های) مورد نظر به محلVariables

Save standardized values as variables → OK

نماد E در خروجیSPSS

- SPSSهنگام نمایش اعداد خیلی کوچك 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 یك متغیر

- 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 می کنیم:

Transform → Compute

نام متغیر جدید را در محل Target Variable تایپ می کنیم

از قسمت ماشین حساب فرمول Ln [numexpr را به محل Numeric Expression منتقل می کنیم

بجای علامت سئوال ظاهر شده متغیر مورد را منتقل می کنیم

OK

طرز کد گذاری مجدد یك متغیر

Transform → Recode → Into different variables

انتقال متغیر مورد نظر به محلOutput variable

تایپ نام جدید متغیر و انتخابchange

Old & New Values → تایپ عنوان جدید متغیر در محلLabel

تایپ ارزش) کد) قدیمی در محل Old value و ارزش جدید در محل New value و انتخاب Add

Continue → OK

طرز گرد کردن اعداد

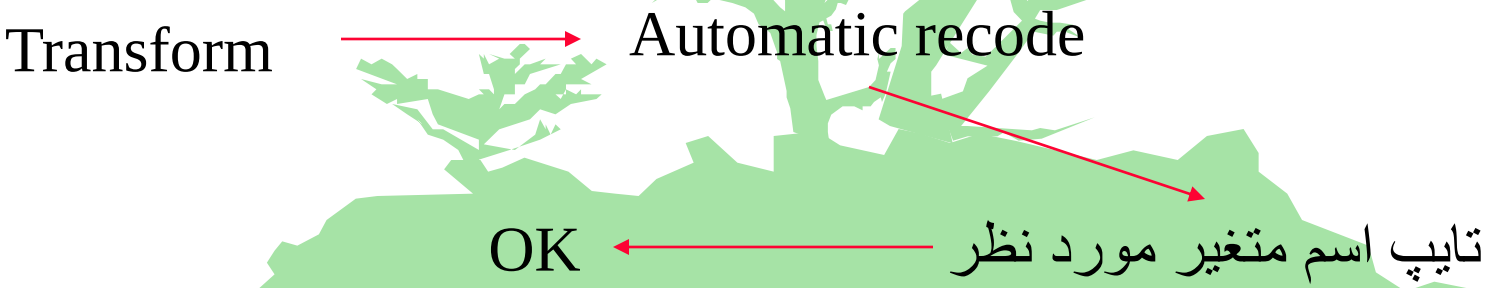
- برای گرد کردن اعداد در روش کدگذاری کافی است در بخش Old value گزینه Range را انتخاب کنید
- دامنه مورد نظر را تایپ کنید مثلاً

1 though 1.49

- در بخش New value تایپ کنید1
- به همین ترتیب دیگر دامنه های مورد نظر را مشخص کنید و بقیه فرایند کدگذاری را ادامه دهید.

طرز تبدیل متغیر string به متغیرNumeric

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



نمودار های مستطیلی(Boxplot)

- یکی از ساده ترین روش های مقایسه زیر گروه های یك متغیر استفاده از نمودار مستطیلی است. أین 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 25 می باشد و مرز بالایی نماینده صدك 75 می‌باشد. طول عمودی مستطیل نشان دهنده دامنه بین چارکی است یعنی 50درصد 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 چولگي مثبت و در صورتی که به لبه بالایی مستطیل نزدیك تر باشد نشانگر چولگی منفی (Negative Skewness)می باشد.

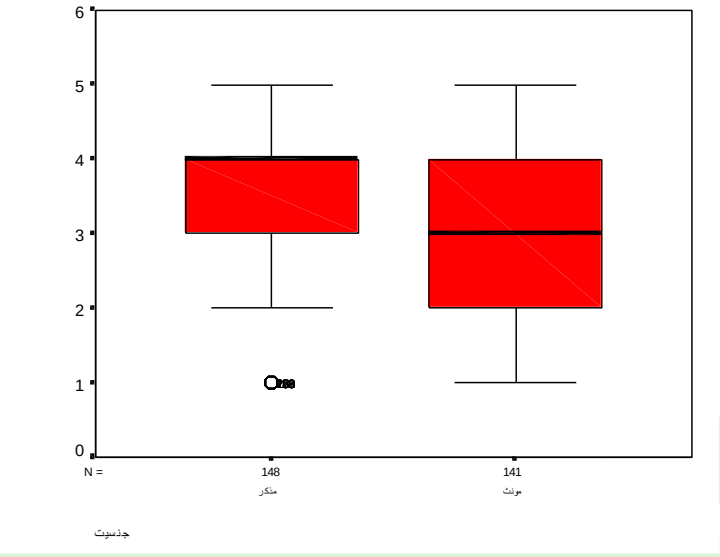
طرز ترسیم نمودار مستطیلی

Analyze → **Descriptive statistics** → **Explore..**

وارد کردن متغیر مورد نظر در محل Dependent List
و متغیری که دارای زیرگروه های مورد نظر است در محل Factor List

Plots..

Factor Level Together → **Continue** → **OK**



طرز ترسیم نمودار ستونی برای مقایسه دو متغیر

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

↓

Define

انتقال دو متغیری که نمودارهای آنها باید مقایسه شوند به محل Bars Represent

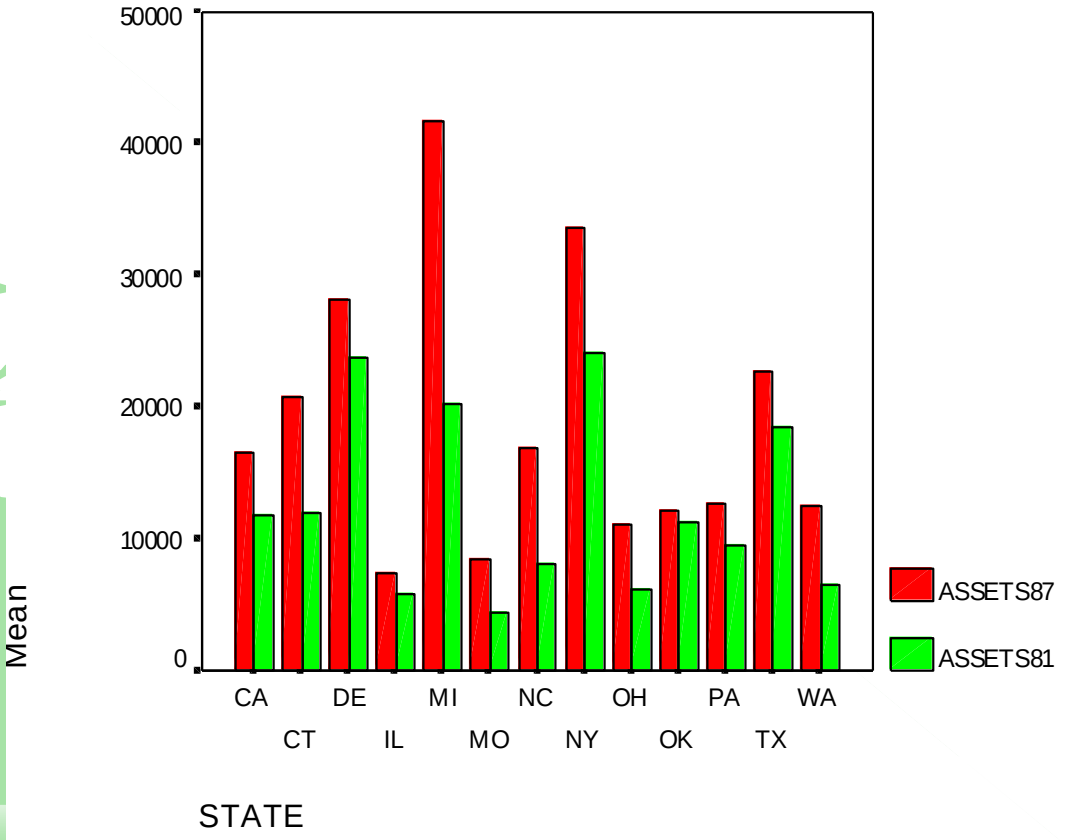
↓

انتقال متغیری که دارای زیر گروه ها می باشد به محل Category Axis

↓

OK

داده ها در فایل ind8187 موجود هستند



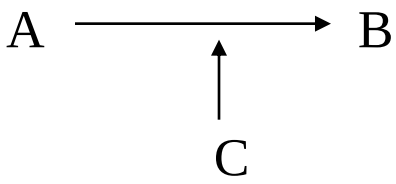
مقیاس های اندازه گیری

1. مقایس 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 می‌روند .اطلاعات با مقیاس اسمی را نمی توان به صورت معنی داری از کم به زیاد ردیف نمود .مانند جنسیت
2. مقیاس رتبه ای (ترتیبی :(Ordinal Scale) (برای نشان دادن طبقه هایی با رتبه مساوی به کار می رود مثلاً رضایت شغلی

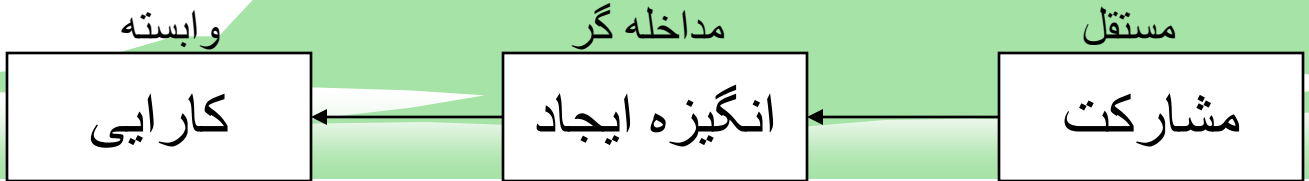
   خیلی کم کم تاحدودی زیاد خیلی زیاد
3. مقیاس فاصله ای: زمانی به کار می رود که مقدار یك متغیر را بتوان قید کرد اما صفر مطلق ندارد (مقادیر منفی نیز پوشش داده می شود). مانند درجه حرارت. نمی توان نسبت بین دو مقدار را در متغیرهای فاصله ای محاسبه کرد .
4. مقیاس نسبت: مشابه مقیاس فاصله ای است با این تفاوت که 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 متوان گفت که یك مقدار در مقایسه با دیگری چقدر کوچکتر یا بزرگتر است .مانند سن یا سنوات تحصیلی

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 متغیرهای مستقل و وابسته تاثیر اقتضایی دارد.

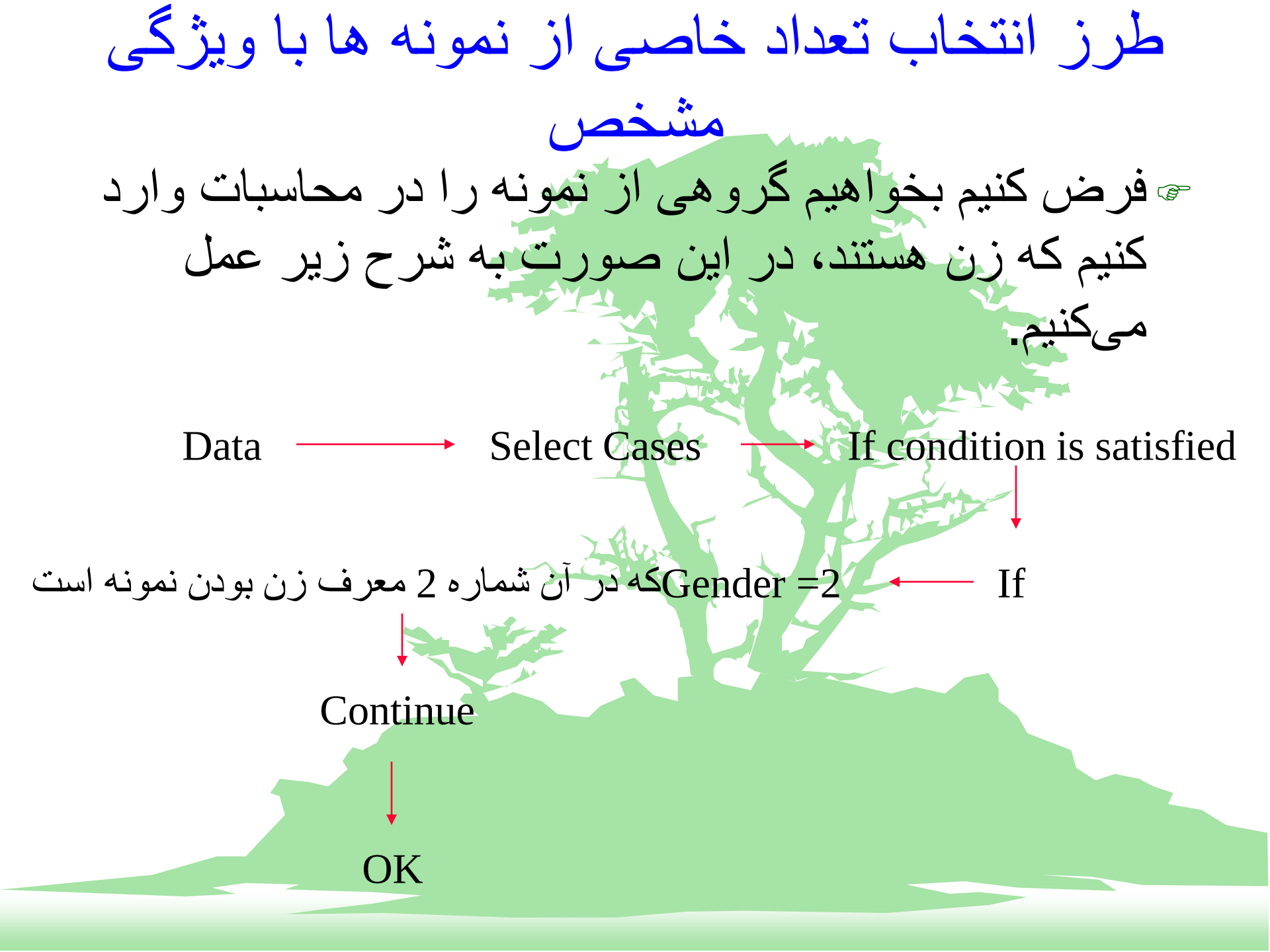


- متغیر کنترل :متغیری است که تاثیر گذاری آن را در جریان پژوهش کنترل می کنیم.
- متغیر مداخله گر :(Intervening) متغیری است که نحوه تاثیر گذاری متغیر مستقل را بر متغیر وابسته نشان می دهد.



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.

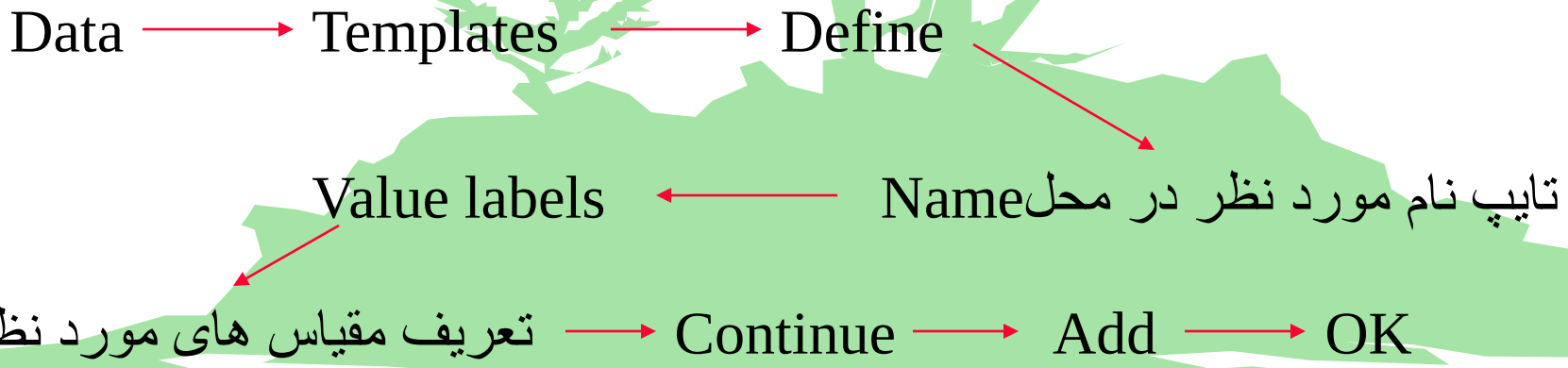


طرز خواندن يك فايل excel درspss

- از روی فایل دستور Open و سپس data را انتخاب کنید.
- در پنجره ای که باز می شود در قسمت File type فرمت excelرا انتخاب کنید
- فایل مورد نظر را انتخاب و open را بزنید.
- در پنجره ای که باز می شود در قسمت Range آدرس متغیر ها 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 5 درجه ای لیکرت مثل هم باشند. در این صورت ضرورتی برای تایپ این مقیاس برای تك تك 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 تعیمم دهیم .



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.
- هرقدر 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 یك مفهوم) که باید سنجیده شود مانند کارایی (تا چه حد تحت پوشش دقیق قرار گرفته است .
- برای سنجش روایی زمانی که چند مفهوم در یك 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 یك متغیر روی یك فاکتور سوار می شوند تا ثابت کنیم آنها تك بعدی هستند.
- برای این کار باید شاخص های) سئوالات (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 یك عامل مستقل سوار شوند.
- در صورتی که بعضی از شاخص های) نشانگرهای (یك متغیر روی یك عامل سوار نشوند، می توان آنها را حذف کرد.

اعتبار(Reliability)

- اعتبار ابزار میزان پایایی و سازگاری آن را در اندازه گیری یك مفهوم نشان می دهد. یعنی توانایی ابزار در حفظ پایایی خود در طول زمان) علیرغم (شرایط غیرقابل کنترل آزمون و وضعیت خود پاسخگویان) حاکی از پایداری آن و تغییر پذیری اندك آن می باشد.
- برای سنجش اعتبار پرسش نامه از روش های زیر می توان بهره گرفت:
  - آلفای کرونباخ
  - تست -تست مجدد(Test-Retest)

روشTest-retest

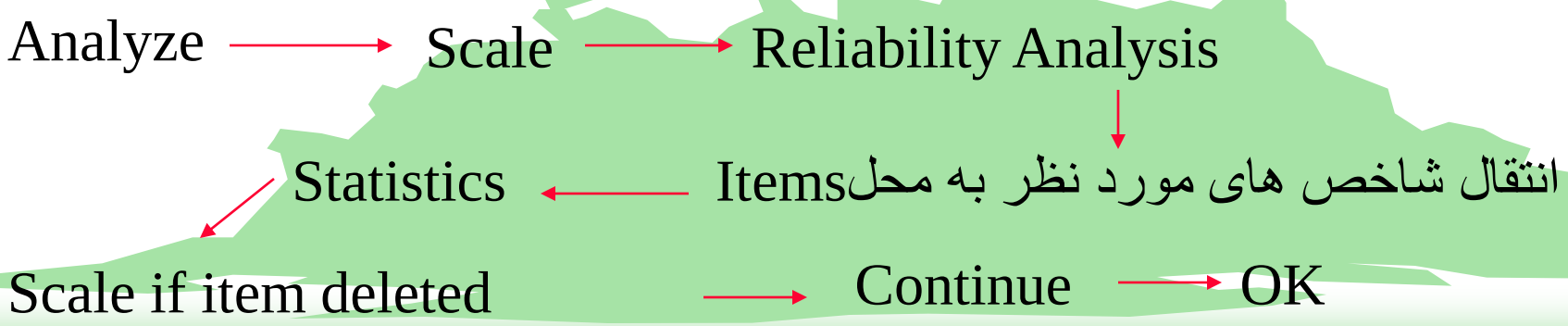
- در این روش تعدادی پرسش نامه در مقیاس کوچك توزیع می شود و بعداز گذشت مثلاً یك هفته همان پرسش نامه میان همان گروه توزیع می گردد.
- داده های به دست آمده در دو مرحله با استفاده از آزمون t نمونه های جفتی مورد مقایسه قرار می گیرند. و یا همبستگی متغیر ها در دو مرحله محاسبه می شود.

آلفای کرونباخ

- یکی از متداول ترین روش ها برای تعیین اعتبار است. آلفا مبنی بر سازگاری داخلی یك تست است. یعنی مبتنی بر همبستگی متوسط اقلام درون یك تست است .
- فرض برآن است که اقلام یك مقیاس باهم همبستگی دارند زیرا آنها یك 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 یك متغیر را بسنجیم و اطمینان حاصل کنیم که شاخص های مورد نظر متغیر واحدی را می سنجند، 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 متغیر مورد نظر را می سنجند و می توان به آنها اعتماد کرد .



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
  - براساس 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 معمولا به شکل فرضیه صفر و فرضیه جایگزین) مخالف) بیان می شود.
- فرضیه صفر (Null Hypothsis) ادعایی است که در مورد جامعه که مبیّن 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 یك وضعیت 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 یك 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 معنی‌داری 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 یك سئوال 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 3 و 5 را انتخاب کرده باشد، به ترتیب در ردیف مربوطه در ستون متغیر prob1 کد 3 و در ستون متغیر prob2 کد 5 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 1 الی 7

↓

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 هاي 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 اين فرضيه صفر به كار گرفته می شود كه آيا يك نمونه مورد نظر به جامعه با ميانگين مشخص تعلق دارد يا خير .
- اگر سطح معنی داری مشاهده شده كمتر از 05/ باشد فرض صفر رد می شود .(H_0: $\mu=\mu_0$)
- به عنوان مثال، می توانيد از اين رويه برای آزمون اين فرضيه استفاده كنيد كه آيا بچه هايي كه به طور زودرس به دنيا آمده اند بهره هوشی آنها به طور متوسط 100 می باشد يا خير. در اين حالت فرضيه صفر به اين صورت بيان می شود كه بين بهره هوشی افراد زودرس با بهره هوشی جامعه (مثلاً 100) تفاوت وجود ندارد. يعنی در اين قبيل موارد ميانگين نمونه با ميانگين جامعه) كه به نام Test value برچسب خورده است (مقايسه 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 محلTest Value

OK

خروجی آزمون یک نمونه ای

سطح معنی داری دو دامنه ای و یك 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 یك طرفه حساب شود .مانند آزمون فرضیه صفر
- برای محاسبه سطح معنی داری یك طرفه از روی سطح معنی داری دو طرفه (مثلاً آن را α* بخوانیم) در خروجیSPSS ، 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 α*/ 2 کوچکتر از0/ 05 باشد جایی که مقدار tمثبت است.
  - فرض صفر (H_0) را رد می کنیم اگر (1- α*)/2 کوچکتر از0/ 05 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 α*/ 2 کوچکتر از0/ 05 باشد، فرض صفر رد می شود. و جایی که t مثبت و مقدار (1- α*)/2 کوچکتر از0/ 05 باشد، فرض صفر را رد می کنیم.

توان (Power) آزمونt

- توان اصطلاح آماری است که برای توصیف توانایی رد فرضه 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 یك آزمون، می توان به جدول توان که توسط کوهن (1988) (تهیه شده مراجعه کرد.
- برای افزایش توان یك 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 یك دامنه ای از آزمون دو دامنه ای پرتوان تر است.
  - در آزمون یك 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 حیواني 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 بعداز اعمال يك تيمار مقايسه كنيد.
- فرضیه صفر در طرح جفت این است که اختلافی بین مقادیر متوسط در دو عضو يك جفت جامعه وجود ندارد. به عبارت دیگر، اختلاف متوسط جامعه صفر است.

دستور آزمون T با نمونه های جفتی

Analyze

↓

Paired Sample T Test

↓

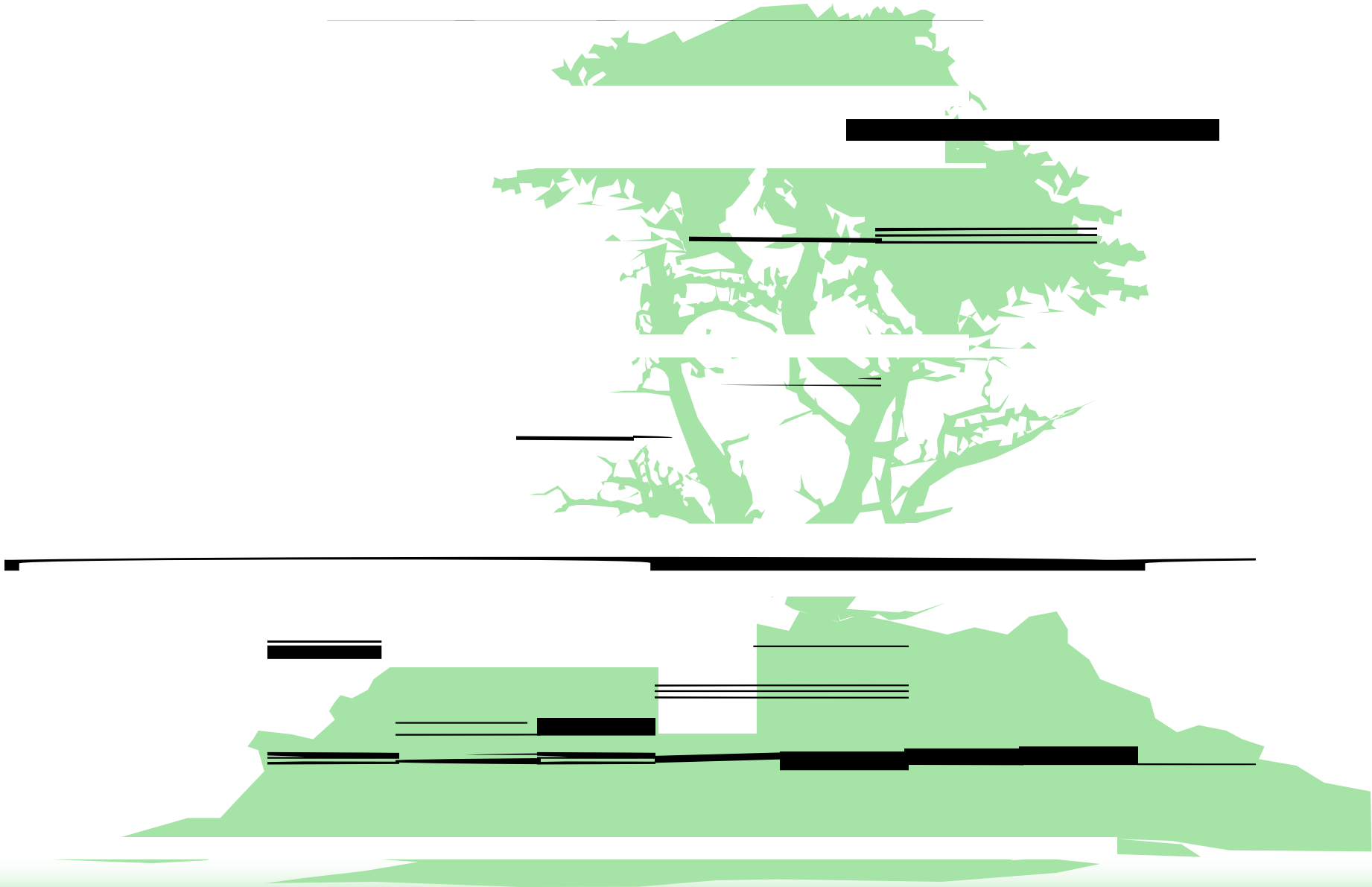
مورد نظر انتخاب دو متغیر که معرف مشاهدات دوتایی هستند و انتقال آنها به محل Paired Variables

↓

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



چند مشکل احتمالی در آزمون T با نمونه های جفتی

- اگر می خواهید اثر دو درمان را بر روی یك فرد مقایسه کنید باید بین دو درمان زمان کافی وجود داشته باشد.
- اگر نمونه های شما یك فعالیت یا آزمایش را دوبار تکرار می‌کنند، 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 یك متغیر در جامعه برای دو گروه مختلف افراد باهم برابر است .
- هنگامی که سطح معنی داری مشاهده شده کوچك باشد فرض صفر) میانگین های یك متغیر در گروه های مختلف باهم تفاوتی ندارند) رد می شود و نتیجه می گیرید که به نظر می‌رسد میانگین های دو گروه با هم برابر نیستند.

دستور آزمون T با دو نمونه مستقل

Analyze → Compare Means → Indepndent-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

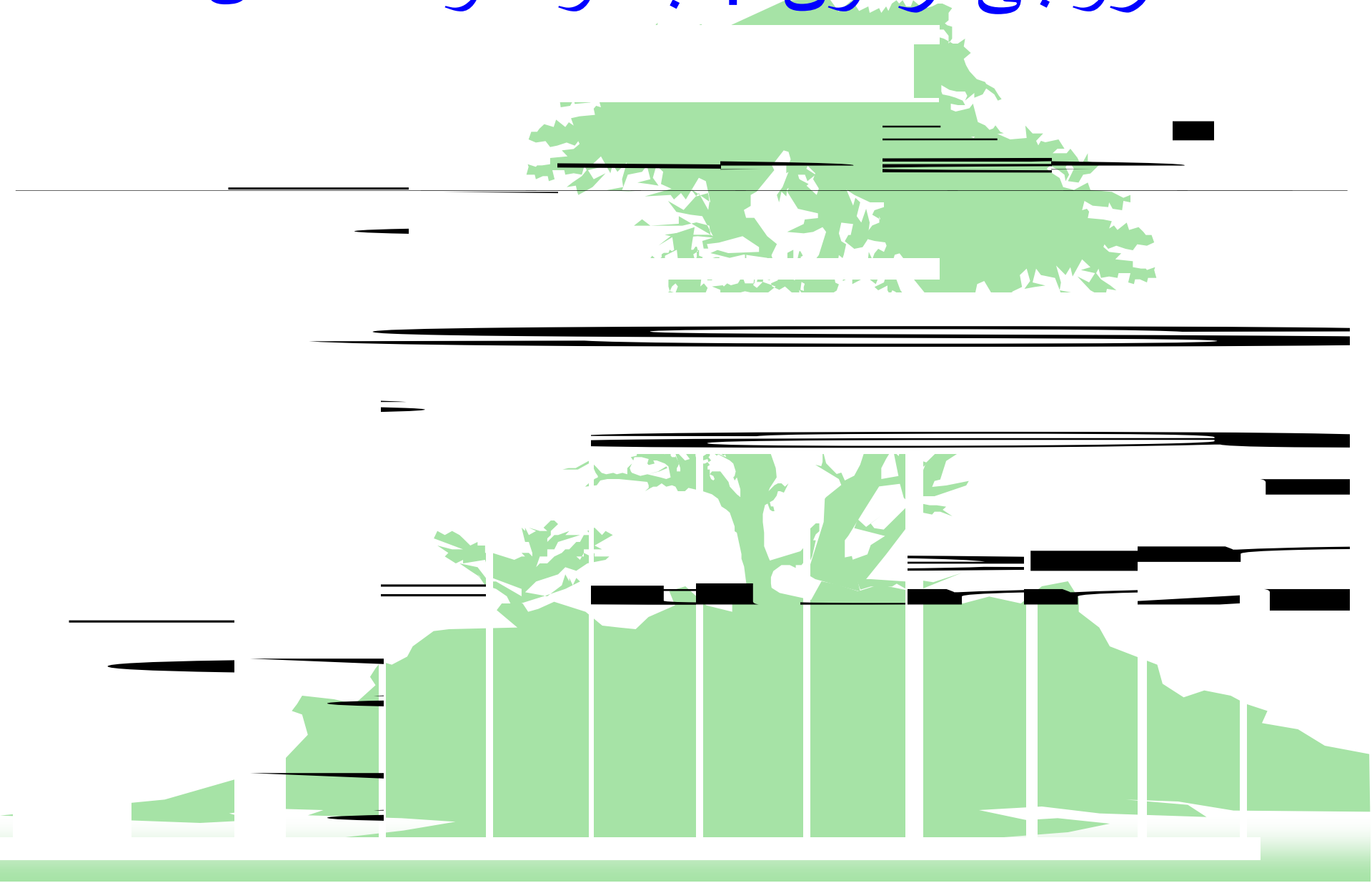
↓

Continue

↓

OK

خروجی آزمون T با دو نمونه مستقل



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 assumedهستند، استفاده کرد .

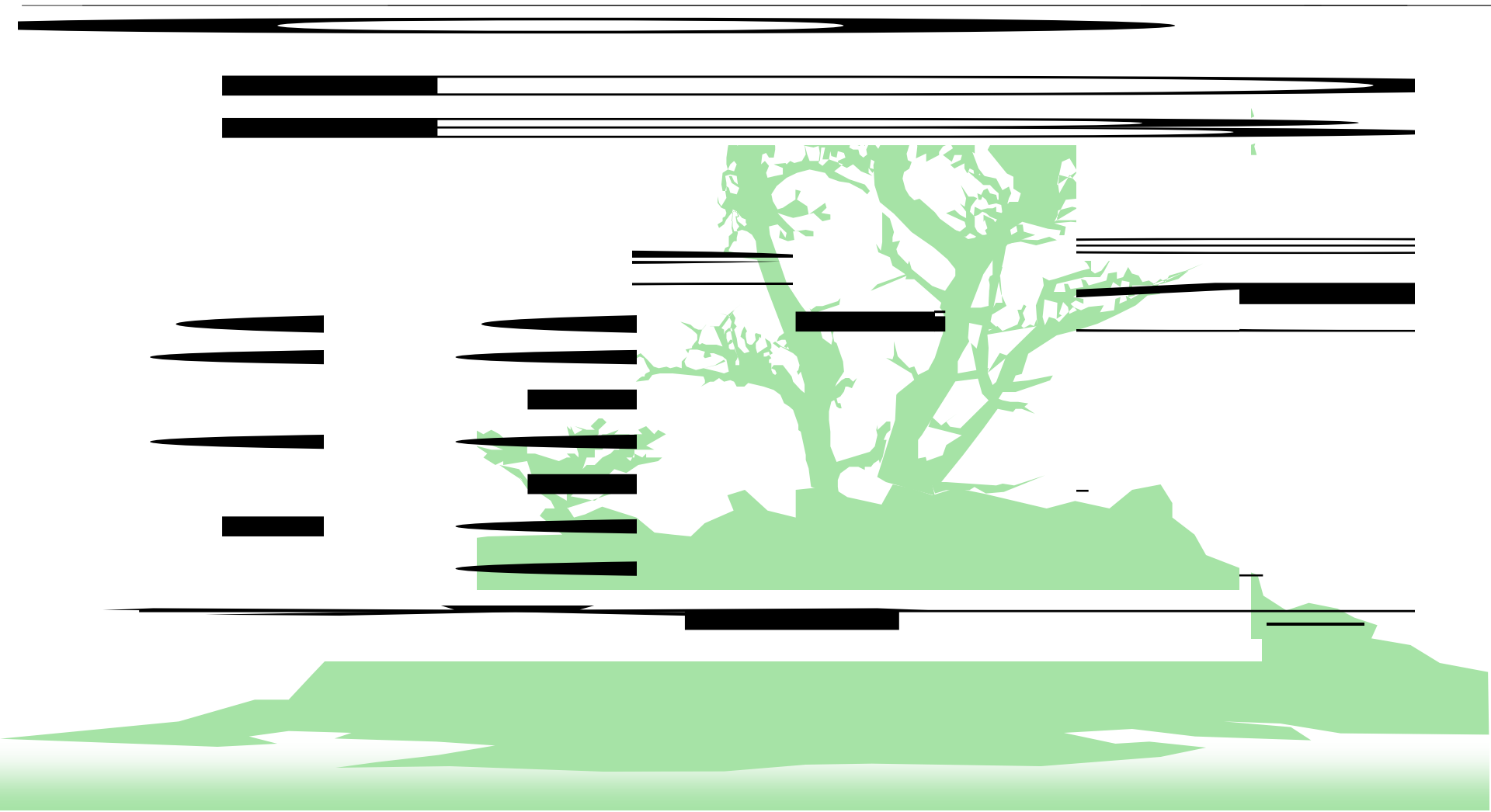
آنالیز واریانس یك طرفه(One-Way ANOVA)

- 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 براساس مقادیر یك 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

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



تفسیر مقایسه چندگانه

- در آزمون مقایسه Bonferroni هر ردیف شامل مقایسه یك 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.

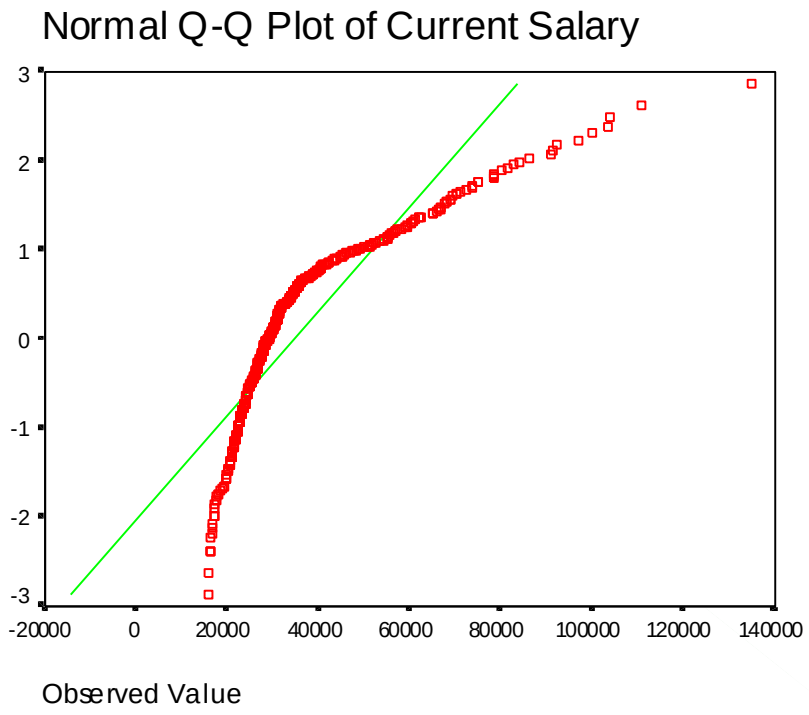
دستور اجرای آنالیز واریانس یکطرفه

شرط استفاده از آنالیز واریانس

- از هر جامعه نمونه های تصادفی مستقل گرفته شده باشند (هریك از مشاهدات باید مستقل باشند.(
- متغیر ها 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 یك توزیع نرمال باشند، نقاط باید اطراف یك خط صاف جمع شوند.



دستور رسم نمودار:

Graphs ← Q-Q..

انتقال متغیر مورد نظر به محل Variables

OK

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 کوچك 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.(

دستور آزمون نرمال بودن:

Analyze ← Descriptive Statistics ← Explore ← انتخاب متغیر

↓

OK ← Continue ← Normality plots with test ← Plots ..

تعبیر چولگی (Skewness) و کشیدگی (Kurtosis)

- این دو شاخص هایی برای قضاوت در خصوص نرمال بودن توزیع هستند .یك 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 باریك تر از منحنی نرمال خواهد بود .

آزمون برابری واریانس گروه های مورد مقایسه

- برای استفاده از بعضی تکنیك های آماری نظیر ازمون t با دو نمونه مستقل لازم است به آزمون برابری واریانس ها پرداخت . برای أین منظور از آزمون Levene’s Test استفاده می شود. یعنی فرضه صفر متعلق بودن دو نمونه به جامعه هایی با واریانس برابر را آزمون می کنیم. این آزمون نباید معنی دار باشد. در این صورت می توان گفت که واریانس گروه ها یکسان است. یعنی تفاوت معنی داری از نظر تغییر پذیری بین گروه ها 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 assumedاست.

دستور آزمون برابری واریانس ها

Analyze → Compare Means → Independent samples T test

انتقال متغیر مورد نظر به محل Test Variable

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

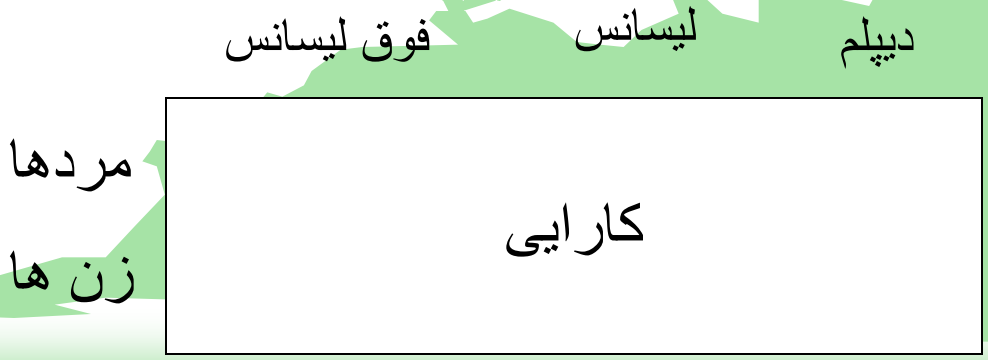
OK

تحلیل واریانس عاملی) دوطرفه(
Factorial ANOVA

- روش سنتی پژوهش تجربی مبتنی بر مطالعه اثر یك متغیر مستقل در یك 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 Factorنامیده می شوند) بر متغیر وابسته را بسنجیم.
- تحلیل واریانس عاملی روشی است که اثرهای ساده و تعاملی دو یا چند متغیر مستقل را برحسب یك 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.

یك 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 سئوال 1 و 2 یك فاکتور (گروه تحصیلی یا جنسیت (درگیر است در حالی که در سئوال سوم هردو فاکتور به طور مزمان مورد نظر است.
- پژوهشگر در سئوال سوم یك 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.

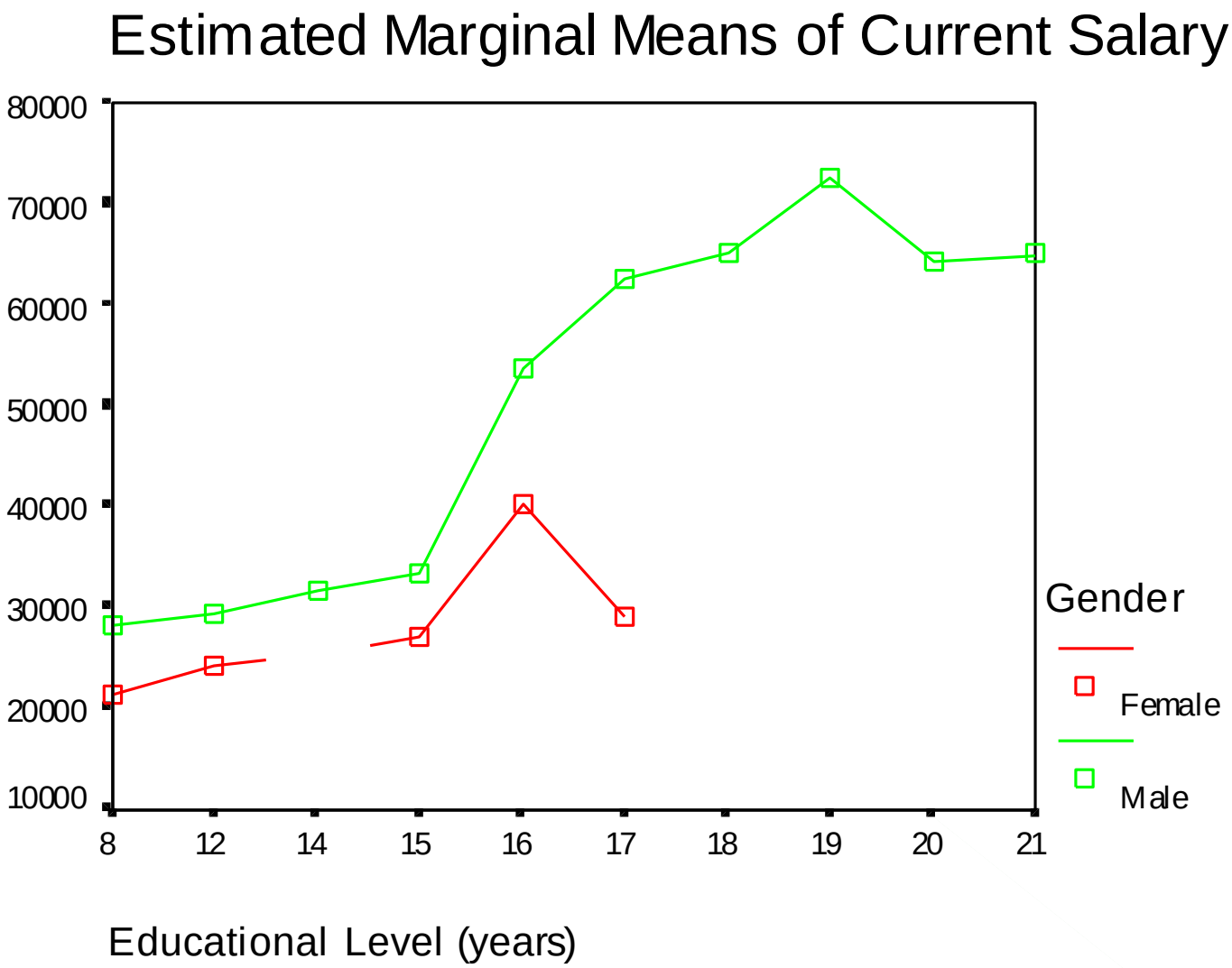


- این پژوهشگر 60 کارمند را از میان جامعه کارکنان به تصادف انتخاب می کند به گونه ای که به 6 گروه 10 نفری مطابق جدول زیر تقسیم شده باشند.

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 هریك 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



جدول آنالیز واریانس

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 هریك 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 هردو 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 (کوچکتر از 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 (کوچکتر از 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 متغیر های مشاهده شده تبیین شود. این درصد در زیر جدول به نام R squared 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 هریك از فاکتورها صرف نظر از سایر فاکتورها. مثلاً اثر درجه تحصیلی بر روی متوسط ساعات کار یك اثر اصلی نامیده می شود، اما اگر دو فاکتور را همزمان وارد کنید آزمون اثر تعاملی نامید 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 گروه‌های فاکتور) 2 فاکتور جنسیت (یکسان است:(

H_0: اثر تحصیلات بر روی متوسط ساعات کار در هر دو جنس زن و مرد یکسان است.

H_a: اثر تحصیلات بر روی متوسط ساعات کار در هر دو جنس زن و مرد یکسان نیست.

اثر تعاملی (ادامه(

- اگر سطح معنی داری مشاده 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 یك از آنها را به تنهایی بر روی متغیر وابسته بسنجیم .اگر اثر تعاملی مشاهده نگردید می توان اثر اصلی هریك از آنها را به روش آنالیز واریانس یك طرفه بر متغیر وابسته سنجید.

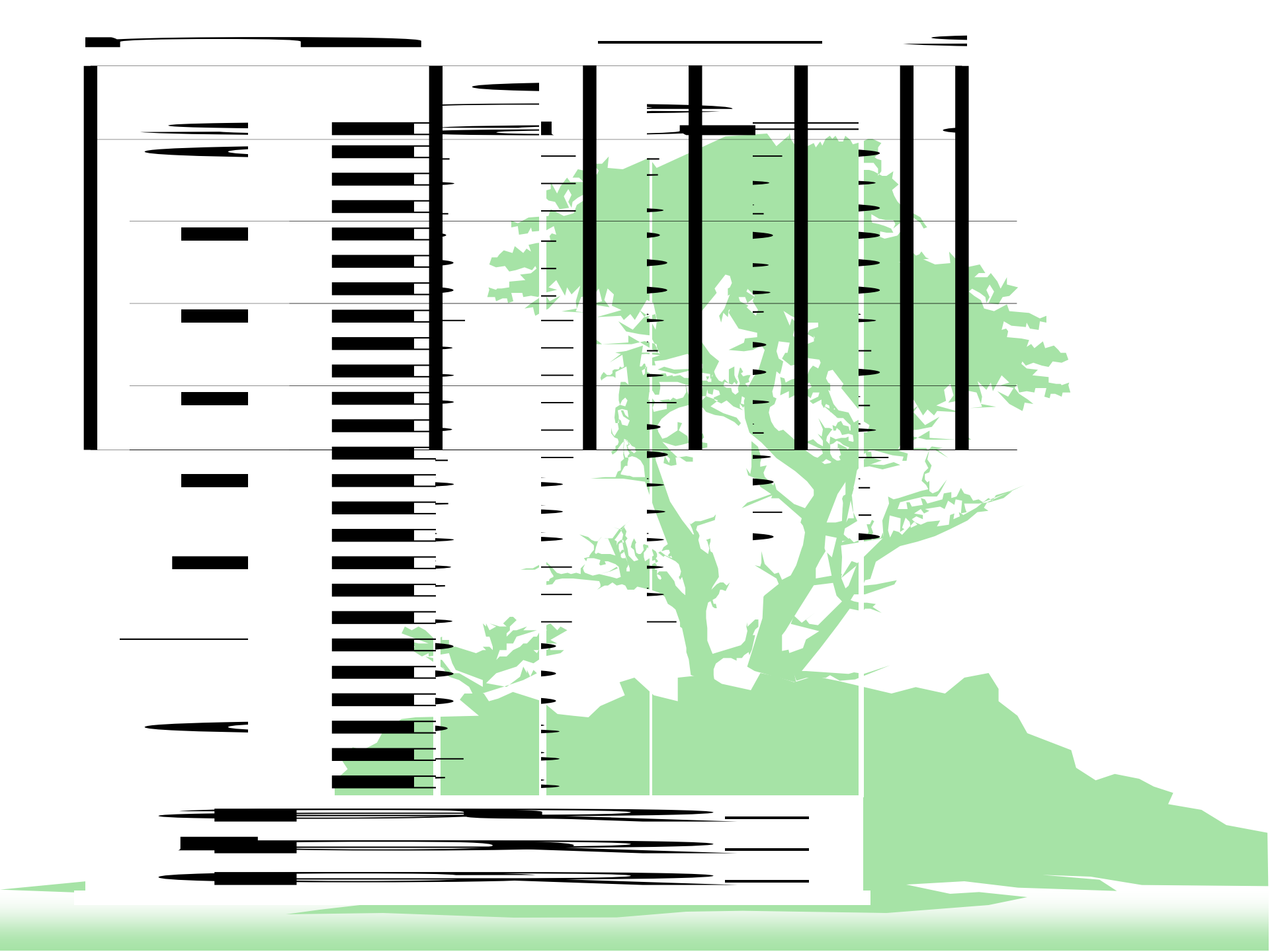
تحلیل واریانس چند متغیره(MANOVA)

- به علاوه، اثرات همپراش ها (covariates) و همچنین تعامل های همپراش ها با عوامل) متغیرهای مستقل) مورد بررسی قرار می گیرد.
- برای تحلیل رگرسیونی، متغیرهای مستقل (Predictor) به عنوان متغیرهای همپراش درنظر گرفته می شوند.
- اگر بیش از یك متغیر وابسته مشخص شود، تحلیل چند متغیره واریانس از روش های phillai, wilks, Hotellingو Roy با آماره F استفاده می کند .همچنین تحلیل تك متغیره (univariate) واریانس برای هر متغیر وابسته صورت می‌گیرد.

مثال

- فرض کنید اثر دو متغیر مستقل I_1 و I_2 را بر روی سه متغیر وابستهD1 ، 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



دستور اجرای MNOVA

Analyze → General linear Model → Multivariate

انتقال متغیر های وابسته به محل Dependent Variables

↓

انتقال متغیر های مستقل به محل Fixed factors

↓

Plots..

↓

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

↓

Add → Continue → OK

H_0: متغیر مستقل دو هستند

H_a: متغیر وجود دو تفاوت معنی داری بین دارد

آزمون مربع کای دو
C hi-square test

- 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 متغیر های 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 یك خانه قرار داند.
- مقدار مورد انتظار عبارت است از تعدادی که در صورت مستقل بودن دو متغیر پیش ینی می‌شود.

جدول توافقی

- جزء اصلی جدول توافقی تعداد نمونه هایی است که در هر یك از خانه های جدول قرار می گیرند. روش های آماری که در این فرضیه های صفر به کار می روند براساس مقایسه موارد مشاهده شده در هرخانه 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 هریك 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 یك نباشد.

دستور اجرای آماره کای دو

Analyze → Descriptive Statistics → Crosstabs..

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

Statistics → Chi-square

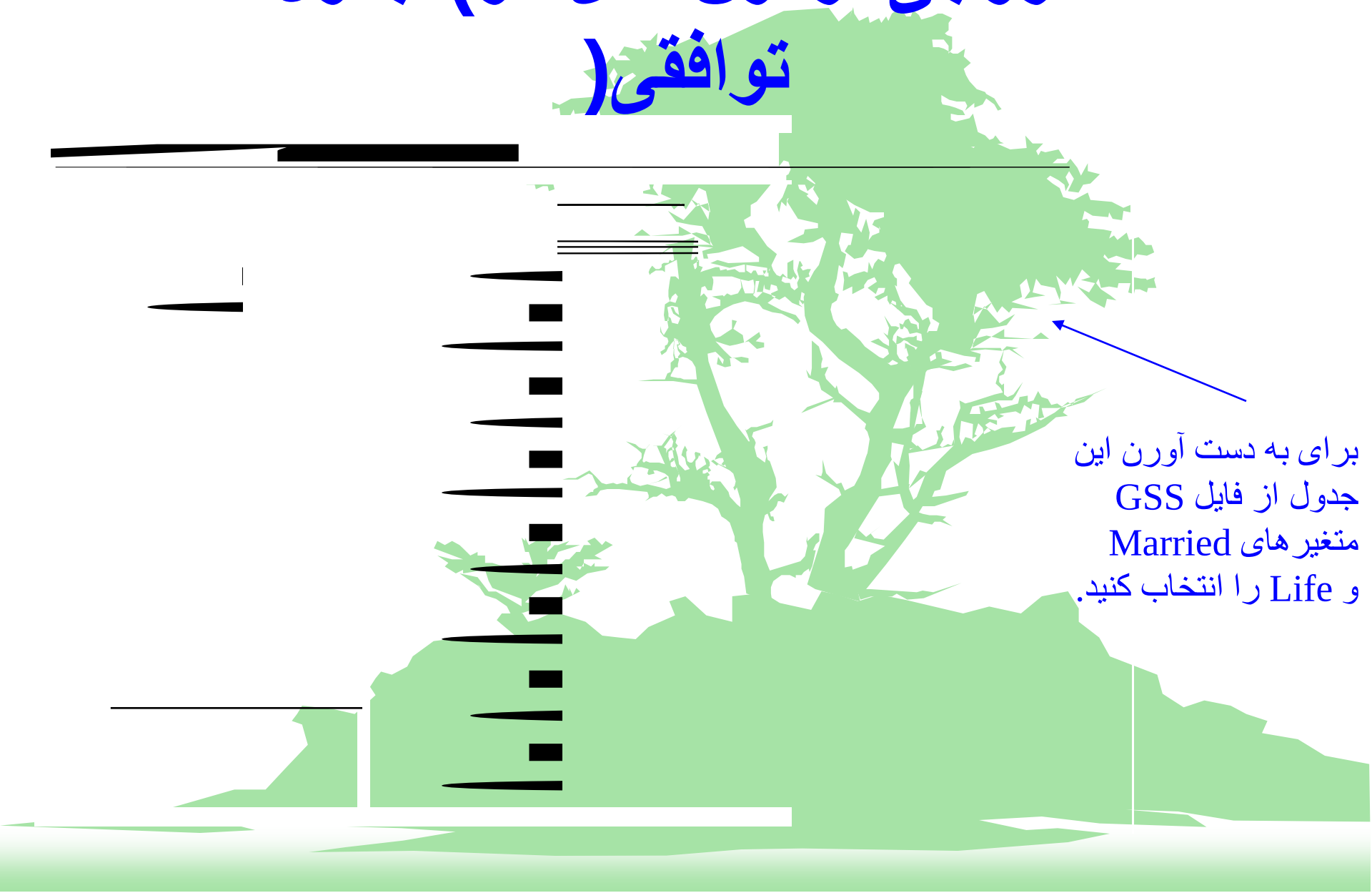
انتخاب گزینه phi and cramer’s v اگر یکی از متغیرهای مورد نظر اسمی باشد و Kendall’s tauبرای متغیرهای رتبه ای

continue → Cells

انتخاب گزینه هایObserved ، Expected ،Total وUnstandardised → continue

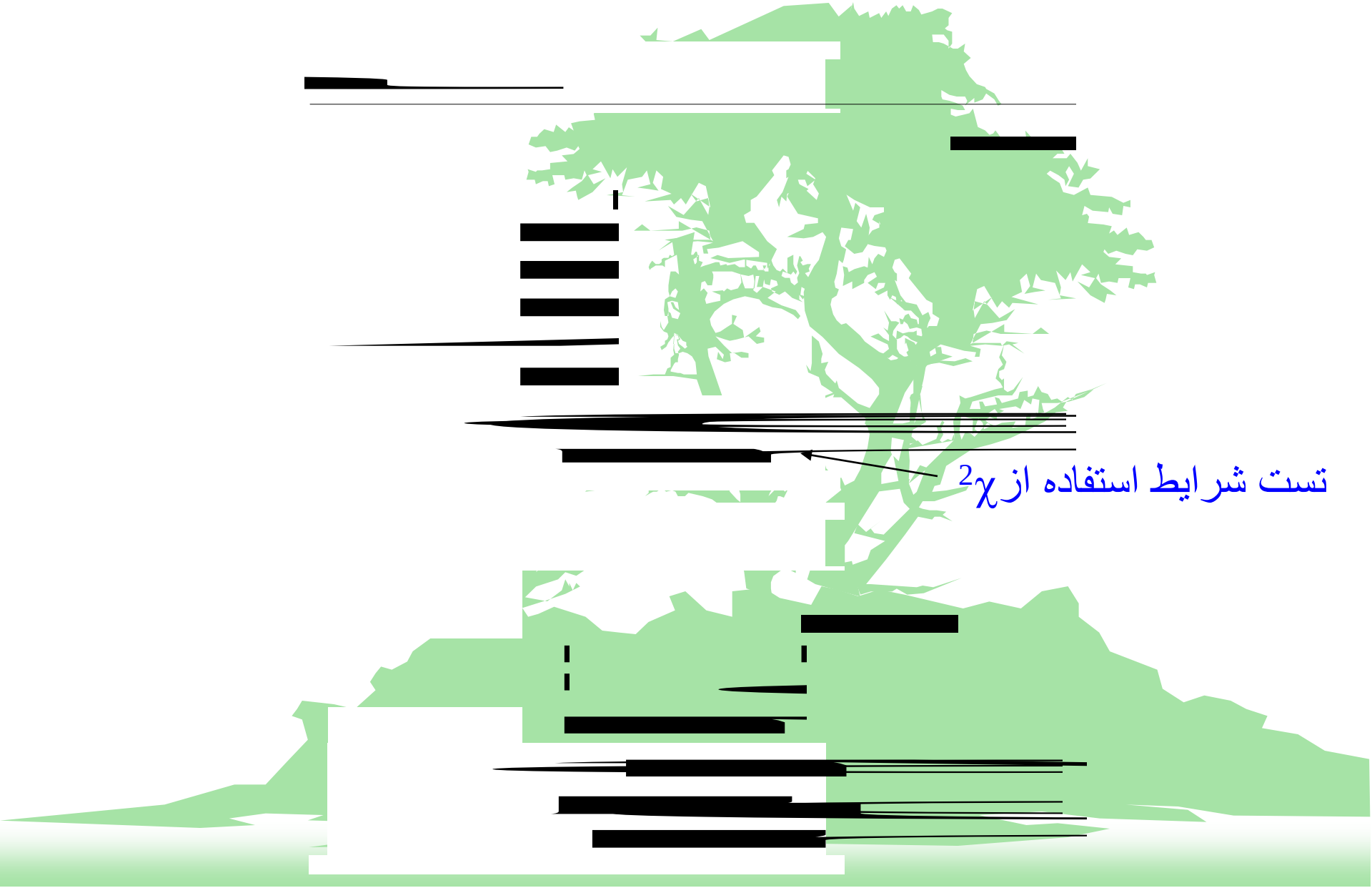
OK

خروجی آزمون کای دو) جدول توافقی(



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

خروجی آزمون کای دو- (χ^2)ادامه



تست شرایط استفاده از χ^2

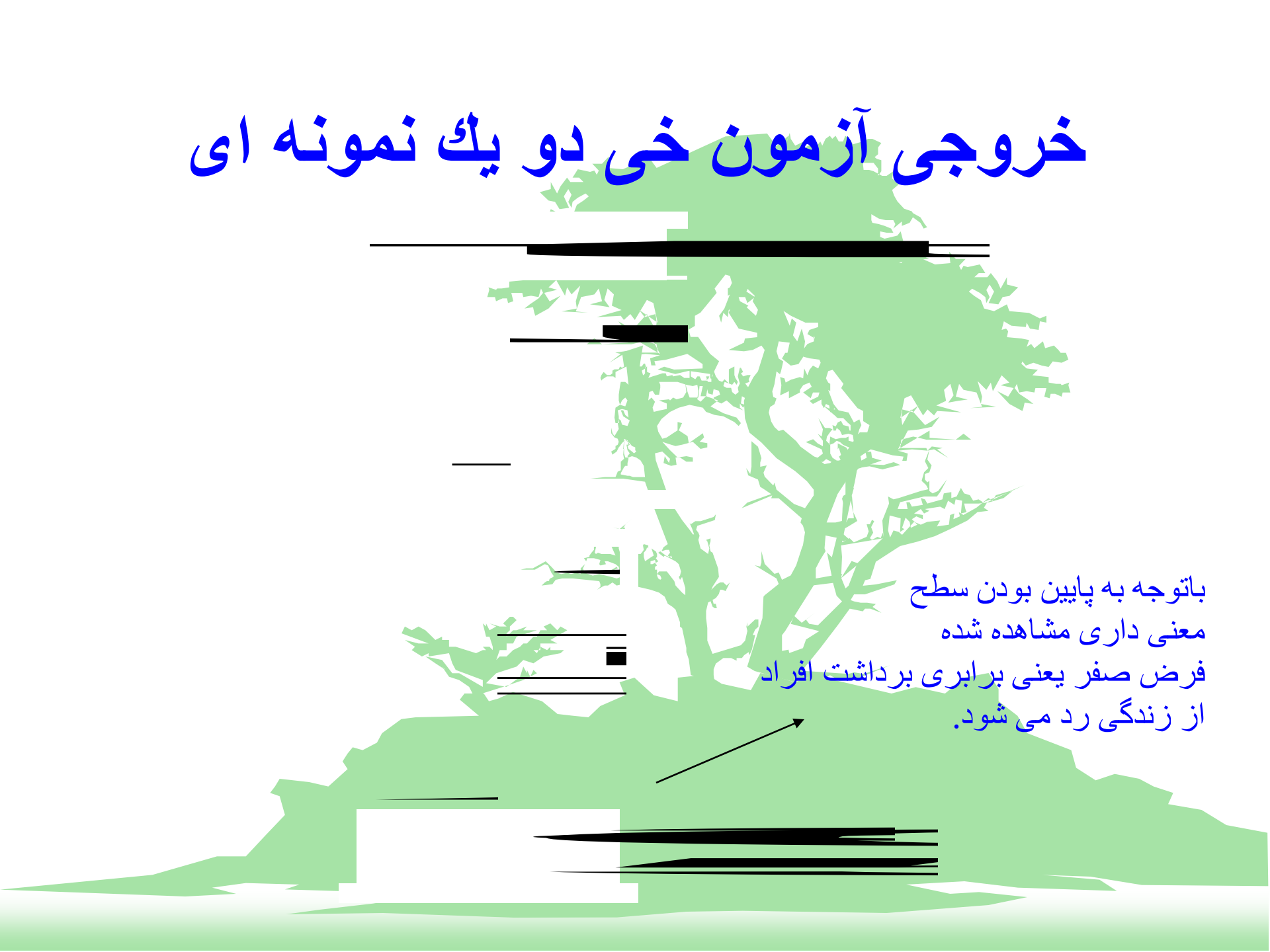
تفسیر نتایج در آزمون کای دو

- 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 Phi و Cramer’s V استفاده می کنیم. در مقیاس کرامر مقدار به دست آمده بین صفر و 1 در نوسان است. صفر نشانگر عدم وجود رابطه و 1 معرف رابطه کامل است.
- در زیر جدول آزمون کای میزان تحقق شرایط استفاده از آزمون قید شده است.

آزمون کای دو) خی دو (یك نمونه ای

- از آزمون کای دو می توان جهت آزمون توزیع یك متغیر نیز استفاده کرد. یعنی می توانید ببینید که آیا توزیع مقادیر مشاهده شده در یك 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 باهم برابر است؟

خروجی آزمون خی دو یك نمونه ای



باتوجه به پایین بودن سطح
معنی داری مشاهده شده
فرض صفر یعنی برابری برداشت افراد
از زندگی رد می شود.

دستور آزمون کای دو یك نمونه ای

توان آزمون کای دو

- مي دانيم كه توانايی ما برای رد فرض صفر وقتی كه غلط است (توان آزمون) به اندازه نمونه بستگی دارد. أين قاعده برای آزمون كاي دو نيز درست است. مقدار آماره كاي دو به تعداد مشاهدات در نمونه بستگی دارد.
- اگر نمونه كوچكی داشته باشيم، ممكن است قادر به رد فرض صفر حتی اگر آن غلط باشد نباشيم. به طريق مشابه براي نمونه های بزرگ ممكن است به رد فرض صفر برسيم حتی اگر انحراف از استقلال خيلی كوچك باشد.
- وقتی يكی از متغيرها يا هردوی آنها در جدول توافقی برمبنای مقياس رتبه ای سنجيده می شود، آزمون كای دو مانند ديگر آماره ها برای كشف انحراف از استقلال قوی

پارامتری غیر آزمون های
Non-Parametric Tests

- آزمون های غیر پارامتری زمانی به کار گرفته می شوند که نمونه ها کوچك 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 پاامتری آزمون t جفتی است که توان کمتری دارد. در حالی که آزمون مشابه آن یعنی Wilcoxonاز 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 یك سری از هم مستقل هستند.

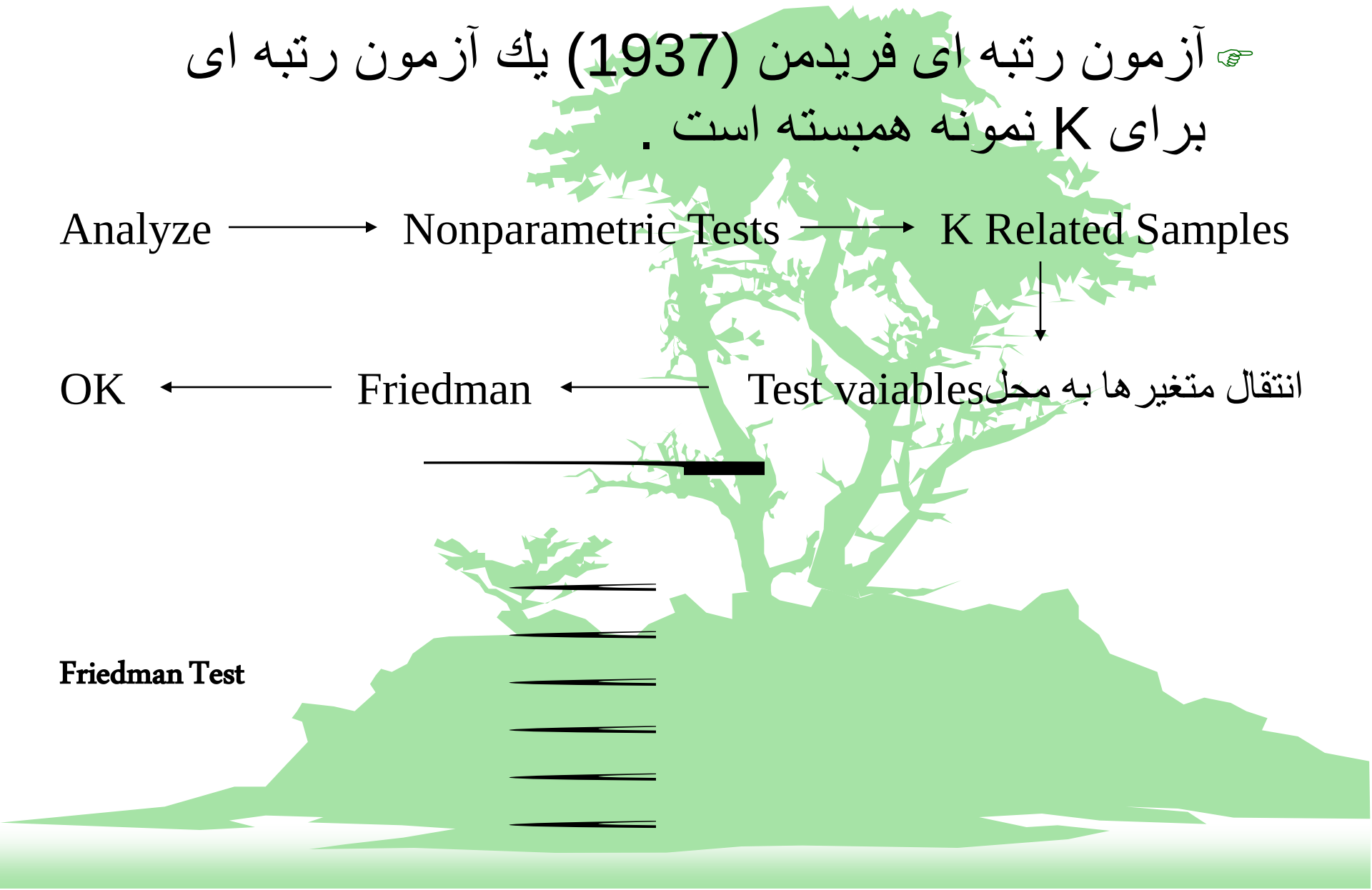
آزمون رتبه ای فریدمن(Friedman)

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

Analyze → Nonparametric Tests → K Related Samples

↓

انتقال متغیر ها به محل Test vaiables → Friedman → OK



Friedman Test

آزمون دو جمله ای

- آزمون دو جمله ای برای آزمون این فرضه به کار می رود که آیا یك متغیر از یك 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

اندازه گیری رابطه

- روش های آماری مختلفی برای مطالعه رابطه بین متغیر ها 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 .
- یك 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 براساس آماره کای دو و دوم شاخص های کاهش نسبی خطا که به آنها PRE می گویند.
- شاخص های رابطه برمبنای کای دو عبارت انداز:
  - ضریب فی(phi Coefficient)
  - ضریب توافق(Coeficient of Contingency)
  - کرامر(Cramer's V)
  - هر سه شاخص قدرت رابطه یکسانی را نشان می دهند، اما تفسیر شاخص هایی که برمبنای 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 (Lambda) 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 مقیاس‌های رتبه ای(Ordinal)

- لاندا به عنوان يك 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 tou-b) که می تواند مقادیری بین -1 تا 1+ داشته باشد.
- تاو C-کندال(Kendal’s tou-c)
- dسامرز(Somer’s d)

شاخص های رابطه برمبنای 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 هبستگی پرکاربرد 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

دستور:

Analyze → Descriptive Statistics → Crosstab..

↓

انتخاب متغیرهای مورد نظر در ردیف و ستون

↓

Statistics

↓

Correlations → Continue → OK

دستور محاسبه همبستگی بین متغیرها

Analyze → Correlate → Bivariate..

انتقال دو متغیر مورد نظر به محلVariables

انتخاب روش محاسبه همبستگی با توجه به نوع متغیر:

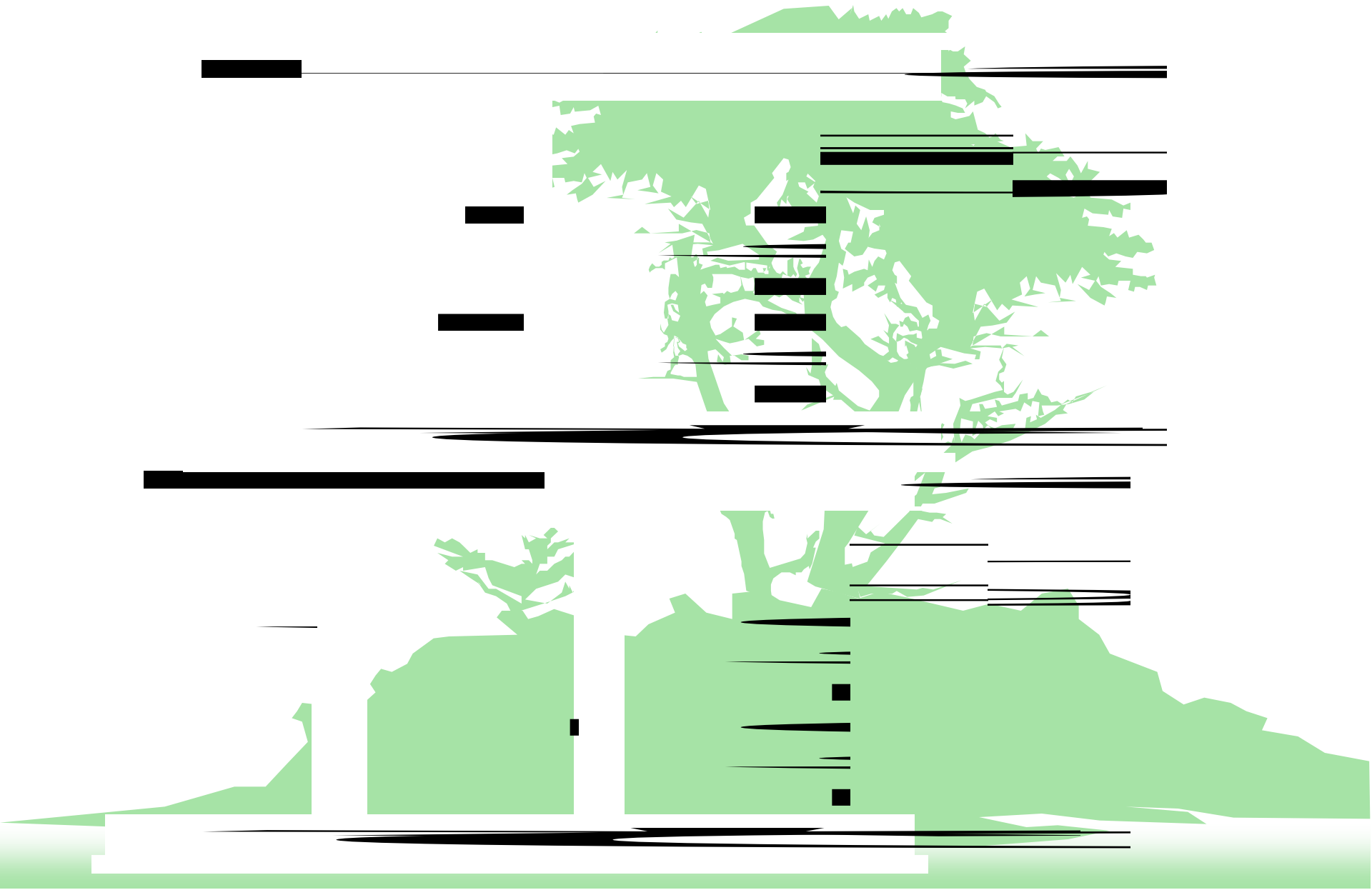
1.متغیرهای فاصله ای: پیرسون

2.متغیرهای رتبه ای :اسپیرمن

3.متغیرهای رتبه ای :کندال

OK

مقایسه ضرایب همبستگی پیرسون و اسپیرمن



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 محلControlling for

↓

انتخاب گزینه Zero-order correlations از رویOptions

↓

OK

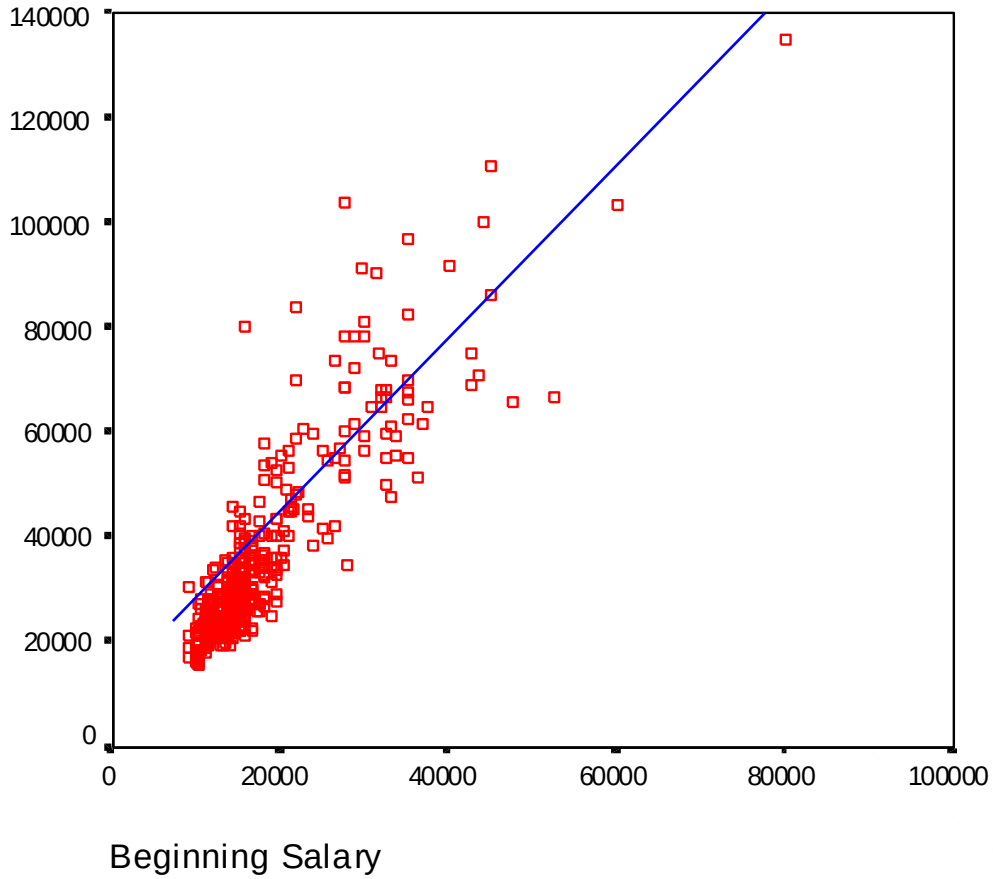
رگرسیون و همبستگی

- آنالیز رگرسیون زمانی استفاده می شود که متغیر وابسته دارای مقادیر متعدد است و در مقیاس فاصله ای یا نسبتی اندازه گیری شده باشد.
- در بررسی رابطه دو متغیر قدم اصلی رسم نموداری از مقادیر است .با مشاهده یك نمودار می توانید بگویید آیا بین مقادیر دو متغیر الگویی مشاهده می شود.
- اگر نقاط حول یك 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 یك خط صاف تجمع پیدا کرده اند.
- هردو مقدار بزرگ مثبت) نزدیك 1+) و بزرگ منفی) نزدیك (-1بیانگر آن است که رابطه خطی قوی بین دو متغیر وجود دارد یا در حقیقت نقاط به خط رگرسیون نزدیك هستند .
- اگر رابطه خطی بین دو متغیر وجود نداشته باشد، مقدار ضریب همبستگی پیرسون نزدیك 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

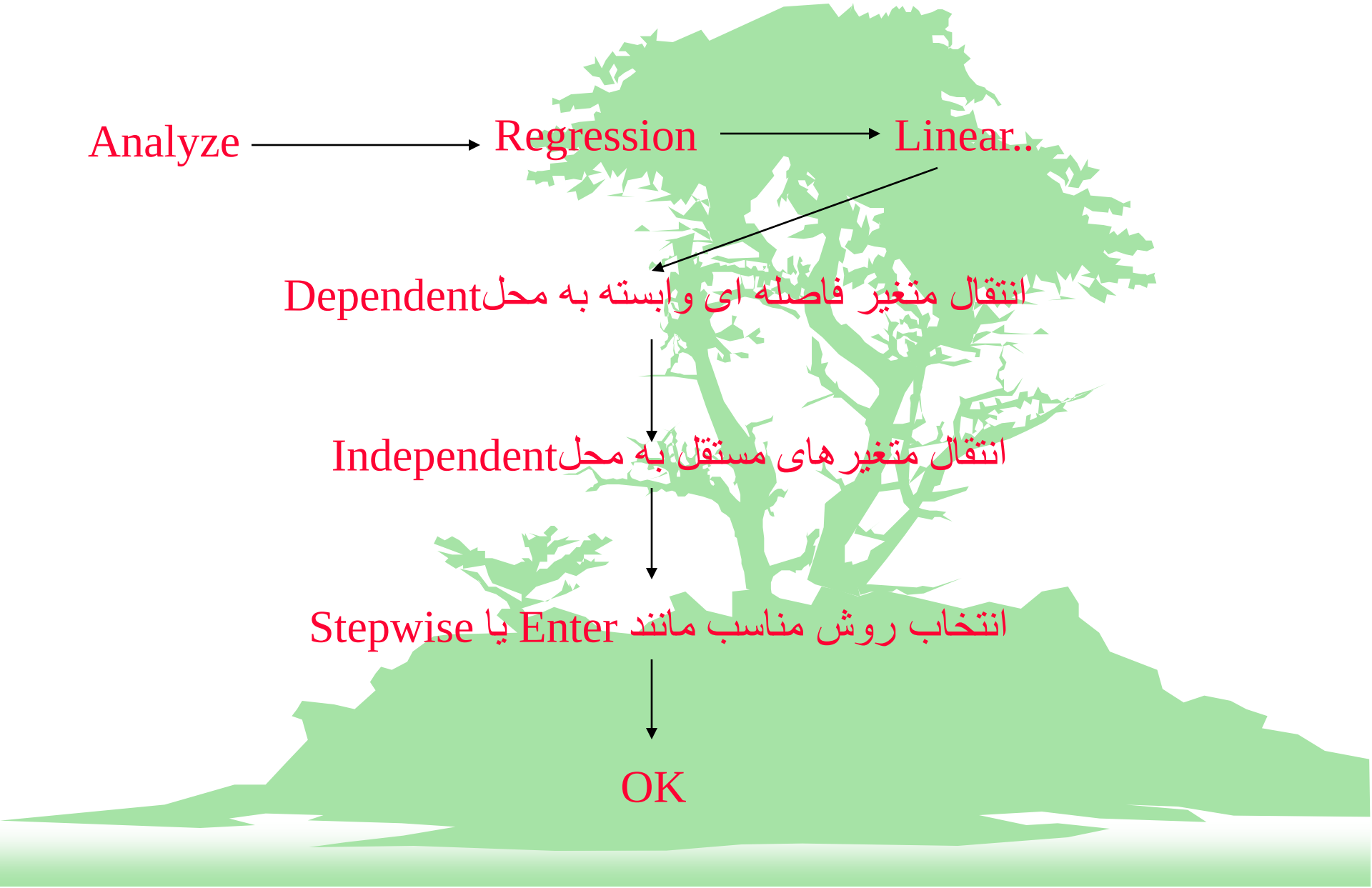


دستور:

Graphs ← Scatter.. ← Simple

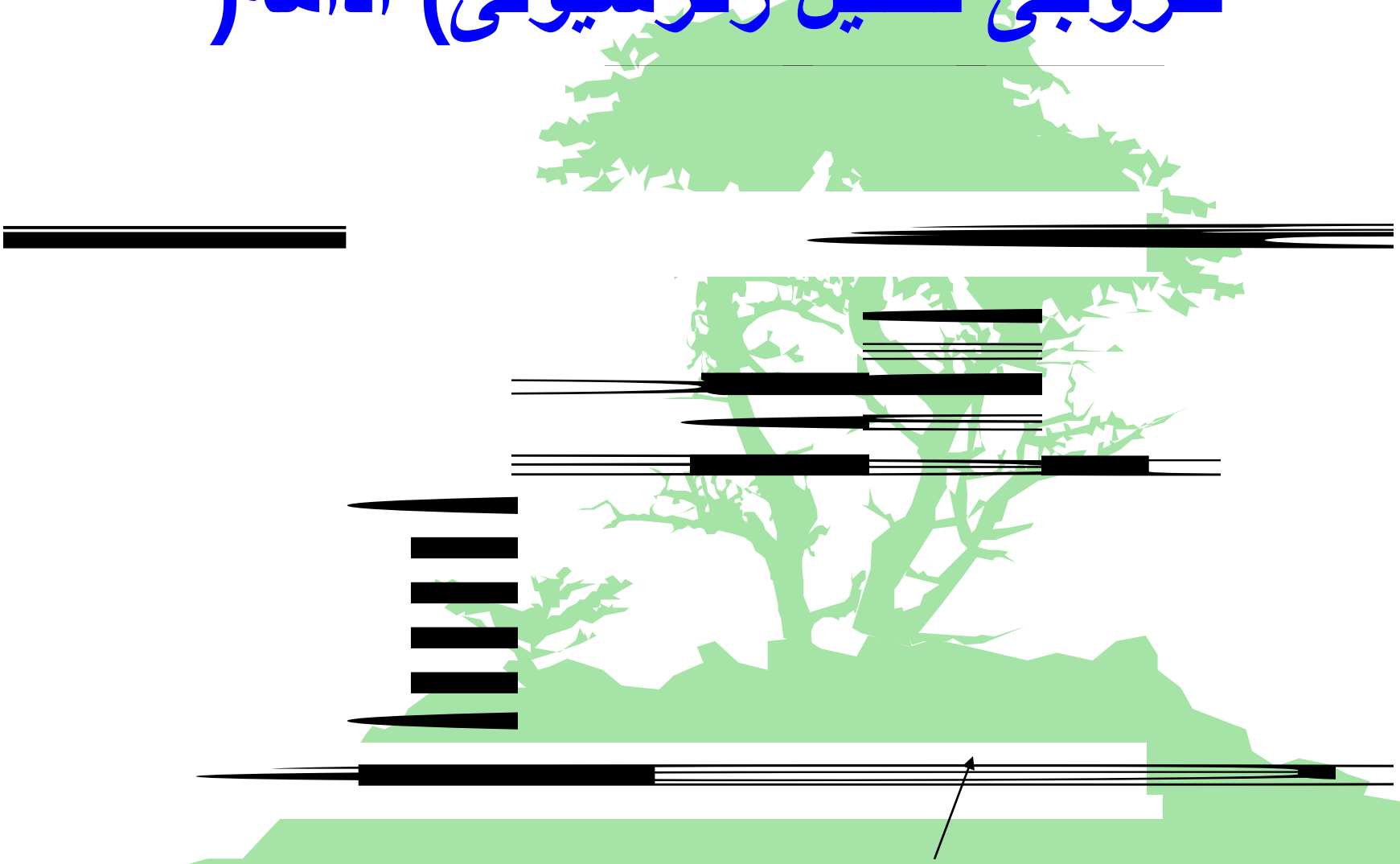
Simple → Define → انتقال دو متغیر به محل محورهای X و Y → OK

روش به دست آوردن رگرسیون خطی



خروجی تحلیل رگرسیونی

خروجی تحلیل رگرسیونی) ادامه(



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

شرایط استفاده از رگرسیون

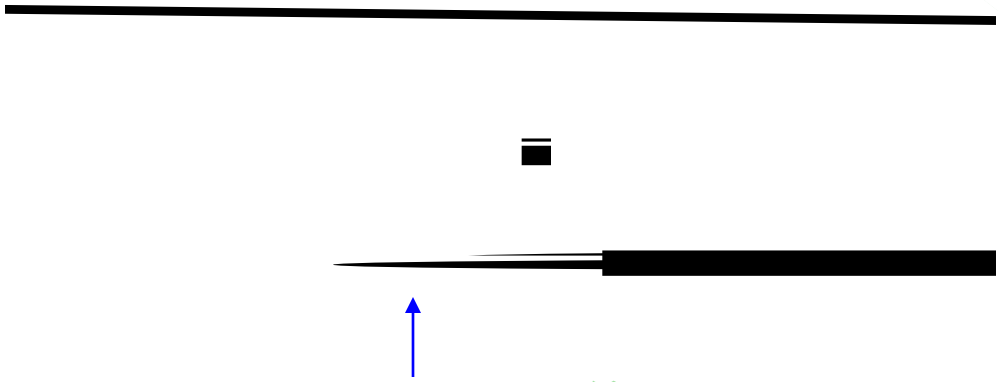
1. تمام مشاهدات باید مستقل باشند.(Independence)
2. برای تمام مقادیر متغیرهای مستقل، توزیع مقادیر متغیر وابسته باید نرمال باشد.(Normality)
3. واریانس توزیع متغیر وابسته باید برای تمام مقادیر متغیرهای مستقل یکسان باشد.(Constant Variance)
4. رابطه بین متغیر وابسته و متغیرهای مستقل باید در جامعه خطی باشد.(Linearity)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

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

بررسی نرمال بودن

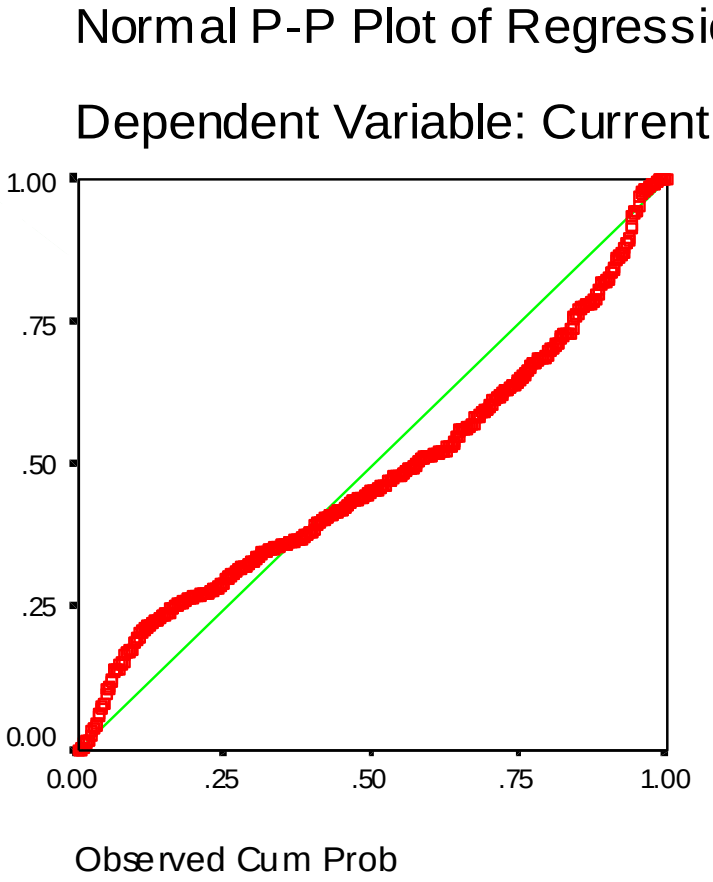
- اگر در جامعه آماری متغیر وابسته برای هریك از مقادیر متغیر های مستقل به صورت نرمال توزیع شده است، پس توزیع باقیمانده ها (residuals) نیز باید به تقریب نرمال باشد.
- برای سنجش نرمال بودن باقیمانده ها باید در بخش متغیرهای مستقل تمامی آنها را وارد کنیم و در نهایت فقط یك نمودار خواهیم داشت.

بررسی نرمال بودن) ادامه(



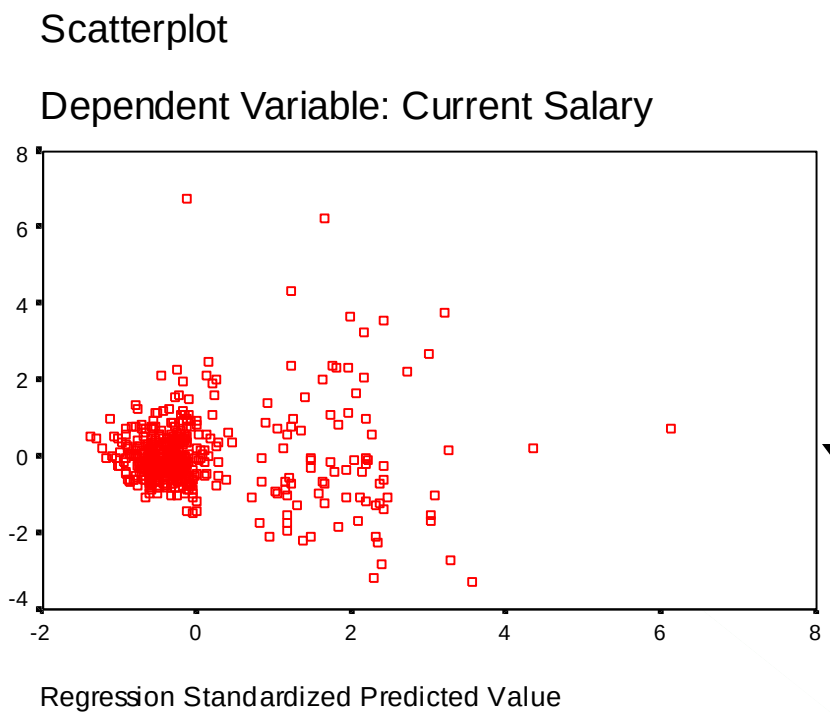
Normality plot with tests

Explore → Plots



بررسی ثابت بودن واریانس

- برای سنجش اینکه آیا واریانس متغیر وابسته برای تمام مقادیر متغیرهای مستقل ثابت است، می توان از نمودارباقیمانده ها (Residuals)را در مقابل مقادیر پیش بینی شده (Predicted values)استفاده کرد.
- 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 تبیدل کنید و سپس رگرسیون را با متغیر تبدیل شده اجرا کنید.

Transform ← Compute.. ← انتخاب نام مناسب براي متغیری که می خواهید تبیدل کنید در محلTarget variable

↓

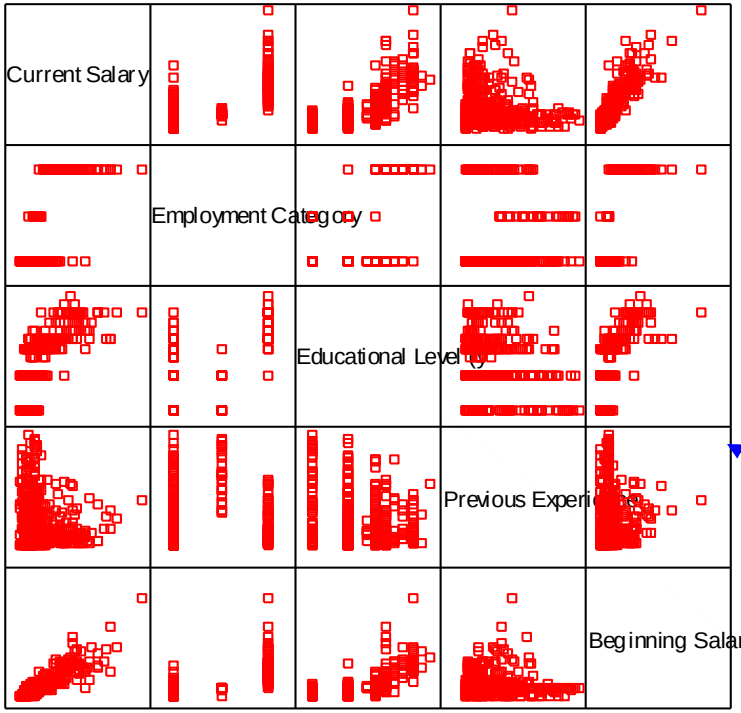
انتقال تابع تبدیل به محلNumeric expresion

↓

انتقال متغیر مورد نظر به محل علامت؟ ← OK

بررسی خطی بودن

- باید ثابت کنیم متغیر وابسته با متغیرهای مستقل به طور خطی رابطه دارد.
- اگر در نمودار پراکندگی الگوی خاصی تشکیل نشود، روابط ین دو متغیر خطی خواهد بود. در غیر اینصورت باید از طریق انتقال یك متغیر با استفاده از توابعی مانندLog ، جذر گرفتن، Tan، و مانند اینها بین آنها رابطه خطی برقرار کرد. به عنوان مثال ممکن است رابطه خطی به شکل زیر درآید x $Y=a+bLog$:



Scatter.. ← Graphs

↓

Matrix → Define

↓

انتقال کلیه متغیرها اعم از وابسته و مستقل به محل Matrix variables

↓

OK

بررسی هم خطی بودن

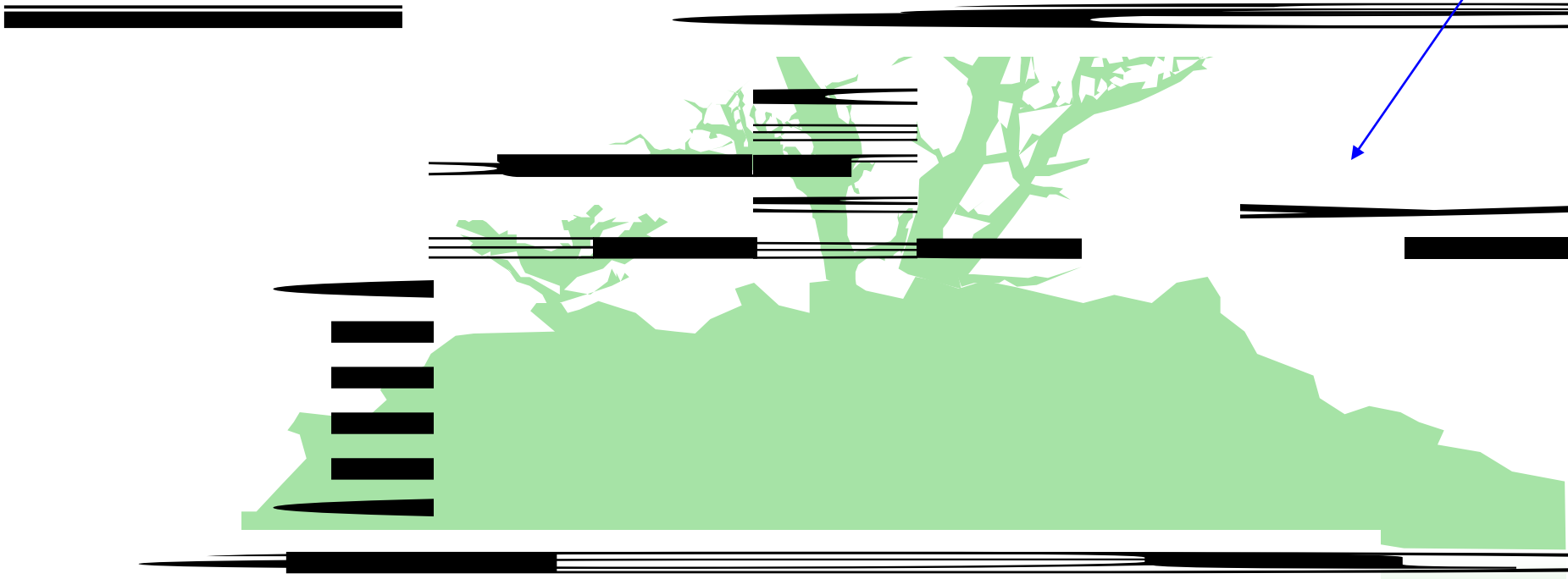
- قدرت رابطه خطی بین متغیرهای مستقل توسط شاخصی اندازه گیری می شود که تولرانس نامیده می شود. مقدار آن بین صفر و 1 تغییر می‌کند.
- برای هر متغیر مستقل تولرانس نسبتی از پراکندگی آن متغیر است که توسط روابط خطی آن متغیر با دیگر متغیرهای مستقل در مدل توجیه نمی شود.
- مقدار نزدیك به صفر (کمتر از 1/) معرف آن است که یك متغیر مستقل تقریباً یك ترکیب خطی از سایر متغیرهای مستقل است. یعنی داده ها دارای روابط خطی مشترك چند گانه (multicolinear) هستند.
- مقدار نزدیك به 1 معرف آن است که متغیر مورد نظر تابع خطی از دیگر متغیرها نیست.

بررسی هم خطی بودن) ادامه(

Analyze → Regression → انتقال متغیر وابسته به محل Dependent و تمام متغیر های مستقل به محلIndepndents

Statistics.. → انتقال گزینهColinearity diagnostic

Normal Probability Plot انتخاب گزینه → Continue → OK



Variance Information Function≡VIF = 1/Tolerance

برازش 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 متغیر های مستقل تبیین مشود .
- اگر R^2 مساوی 1 باشد نشان دهنده آن است که متغیر وابسته کاملاً از طریق متغیر های مستقل قابل پیش بینی است. مقدار صفر معرف آن است که متغیر وابسته به طور خطی به متغیر های وابسته مرتبط نیست.
- Multiple Rقدر مطلق ضریب همبستگی بین مقادیر مشاهده شده متغیر وابسته و مقادیر پیش بینی شده (Predicted) براساس مدل رگرسیونی متغیر وابسته را نشان می دهد .

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

چون سطح معنی داری مشاهده شده کوچك می باشد، لذا فرضیه صفر مبنی بر اینکه رابطه خطی بین متغیر وابسته و متغیر های 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 کوچك باشد، فرضیه صفر یعنی صفر بودن ضرایب رگرسیونی در جامعه رد می شود.

آزمون صفر بودن ضرایب رگرسیونی

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

معادله رگرسیونی

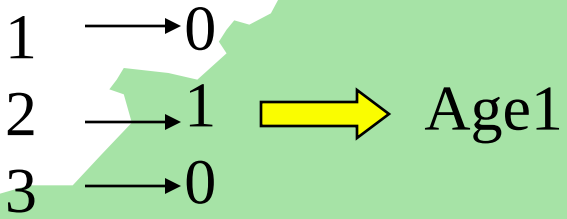
- در صورت صادق بودن شرایط رگرسیونی، تحلیل نهایی از خروجی رگرسیون باید در شکل یك مدل ریاضی) معادله رگرسیونی (ارائه شود. مثلاً با توجه به مثال فوق خواهیم داشت:

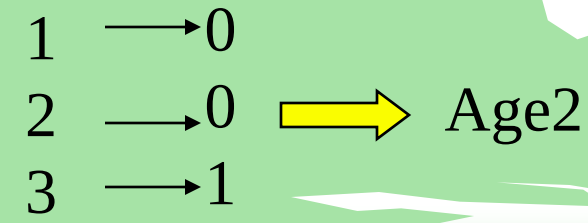
Current Salary = 1.34 Beginning salary + 5930.28 Employment Category -19.03 Previous Experience+ 601.30 Educational Level- 3068.27

- 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 62 درصد از تغییر پذیری در متغیر وابسته توسط حقوق اولیه تبیین می شود. 27درصد از تغییر پذیری در متغیر وابسته توسط طبقه شغلی تعیین می شود. و به همین قیاس در

تعاملی برای اثرات رگرسیونی تحلیل
Regression Analysis for Interactive Effects

- فرض کنید می خواهیم اثر متغیر سن (Age) را به عنوان یك متغیر تعدیل کننده (Moderator) در رابطه خطی $SR = \alpha+\beta DP+ e$ بسنجیم. به کمك 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 کمك دستور Recode متغیرهای مصنوعی جدید را به صورت زیر تعریف می کنیم:





تحلیل رگرسیونی برای اثرات تعاملی (ادامه(

- سپس به کمك 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 یك متغیر مداخله گر عمل می کند.

تحلیل رگرسیونی برای اثرات تعاملی (ادامه(

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

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

- سپس به کمك دستور Compute متغیرهای زیر را تعریف می کنیم:

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

- آنگاه از طریق رگرسیون با درنظر گرفتن متغیرهای جدید به عنوان متغیر مستقل رابطه خطی جدید را برآورد می کنیم .اگر ضریب تعیین بیشتر باشد (R 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 پژوش در قالب جدولی مانند زیر خلاصه شود.

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

تحلیل عاملی(Factor analysis)

- تحلیل عاملی یك تكنیك آماری است كه برای تشخیص حداقل عواملی كه می توانند یك پدیده پیچیده را تشریح كنند.
- هدف تحلیل عاملی تشخیص عوامل ناپیدا برمبنای عوامل مشاهده شده است. این عوامل باید ساده و در عین حال معنی دار باشند.
- هدف تحلیل عاملی پیدا كردن عوامل مشتركی است كه متغیرها را بتوان به صورت تابعی از آنها نوشت .
- تحلیل عاملی ابزاری است برای سنجش روائی پرسش نامه. مقصود آن است كه پرسش نامه چیزی را اندازه بگیرد كه 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: $F_i = \Sigma W_{ij} X_j$
- امید است که تعداد کمی از عامل ها) یعنی ترکیب خطی نمره های اصلی متغیرعای مشاده شده (بتواند تقریباً همه اطلاعاتی را که توسط مجموعه بزرگتری از متغیرها به دست می آید دربرگرفته در نتیجه توصیف ویژگی های یك پدیده را ساده سازد.

شرایط مناسبت ماتریس داده ها برای تحلیل عاملی

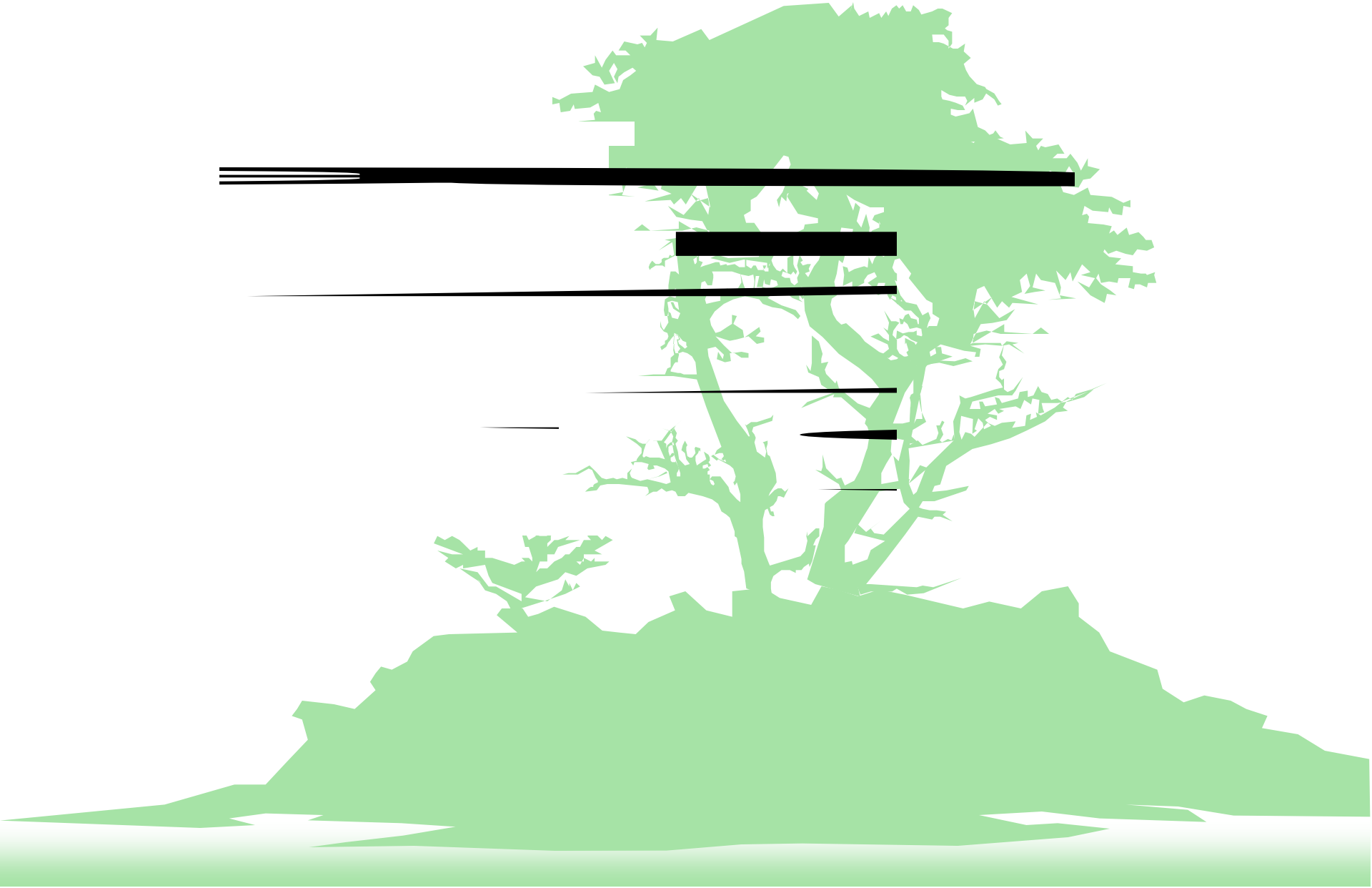
☞ حجم گروه نمونه :

- کامری (1973) گروه نمونه 100 نفری را ضعیف، 200 نفری را نسبتاً مناسب، 300 نفری را خوب، 500 نفری را خیلی خوب، و 1000 نفری را عالی می داند.
- حجم گروه نمونه باید دست کم 10 یا 20 برابر عده متغیرهایی باشد که در ماتریس همبستگیٍ وارد می شود.
- برای اطمینان از کفایت نمونه از شاخص KMO استفاده می شود.
- مقادیر کوچك 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 غیرقابل قبول است .

شرایط مناسبت 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 کوچك باشد، احتمال آنکه در عامل های مشترك سهیم باشند، اندك است.
  - برای ارزشیابی ماتریس همبستگی از آزمون کرویت بارتلت (Bartlett's test of sphericity)استفاده می شود.
  - آزمون کرویت براي رد فرضیه صفر مبنی بر درست بودن ماتریس همانی در جامعه است (یعنی این فرضه که ماتریس همبستگِی مشاهده شده متعلق به جامعه ای با متغیرهای ناهمبسته است.)
  - ماتریس همانی (Identity matrix) ماتریسی است که همه عناصر قطری آن 1 و همه نصار غیر قطری آن صفر باشد.

شرایط مناسبت ماتریس داده ها برای تحلیل عاملی



مراحل تحلیل عاملی

- تشکیل ماتریس همبستگی :(Correlation matrix) ماتریس همبستگی به ما کمك می کند تا عوامل مشترك را شناسایی کنیم .اگر همبستگی بین متغیرها کوچك باشد، غیرمحتمل 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 یك 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 یك 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 مي شود. در أين روش تركيب خطی متغيرهای مشاهده شده تشكيل می شود. روش مولفه های اصلی تنها کاری که انجام می دهد عبارت است از انتقال متغيرهای همبسته به مجموعه ای از متغيرهای غير همبسته. در اين مرحله تعداد عوامل برحسب ارزش ويژه (Eigenvalue) تعيين می شود . عواملی که دارای ارزش ويژه بزرگتر از يك هستند در مدل عاملی وارد می شوند .ارزش ويژه نشان می دهد كه يك عامل تاچه حد در تبيين واريانس مشترك زيربنايی متغيرها سهيم است .اگر عاملی تمام واريانس مشترك موجود در هر متغير را تبيين كند هركدام دارای بار عاملی 1 و مقدار ارزش ويژه برای عامل 1 برابر) n تعداد متغيرها (خواهد بود .
- درصدی از كل واريانس را كه به وسيله يك عامل تعيين می شود را با pct of varنشان می دهند .برپايه ملاك كيسر (1960) تعداد عامل های مشترك بايد مساوی با تعداد مولفه های اصلی دارای ارزش ويژه بزرگتر از 1 باشد.

جدول ماتریس عوامل

- جدول ماتریس عوامل (Factor matrix) ضرایبی را نشان می دهد که می توان هر متغیر را به کمك آنها برحسب عوامل مشترك نوشت.
- بخشی از واریانس یك 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 1 نشان می دهد که تمام واریانس 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.

	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}	....

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 یك عامل تبیین می شود را به وسیله شاخص اشتراك (Communality) مشخص می کنند. مقدار آن بین صفر و 1 در نوسان است.
- صفر معرف آن است که عوامل مشترك هیچ واریانسی از متغیر را تعیین نمی کند و 1 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.

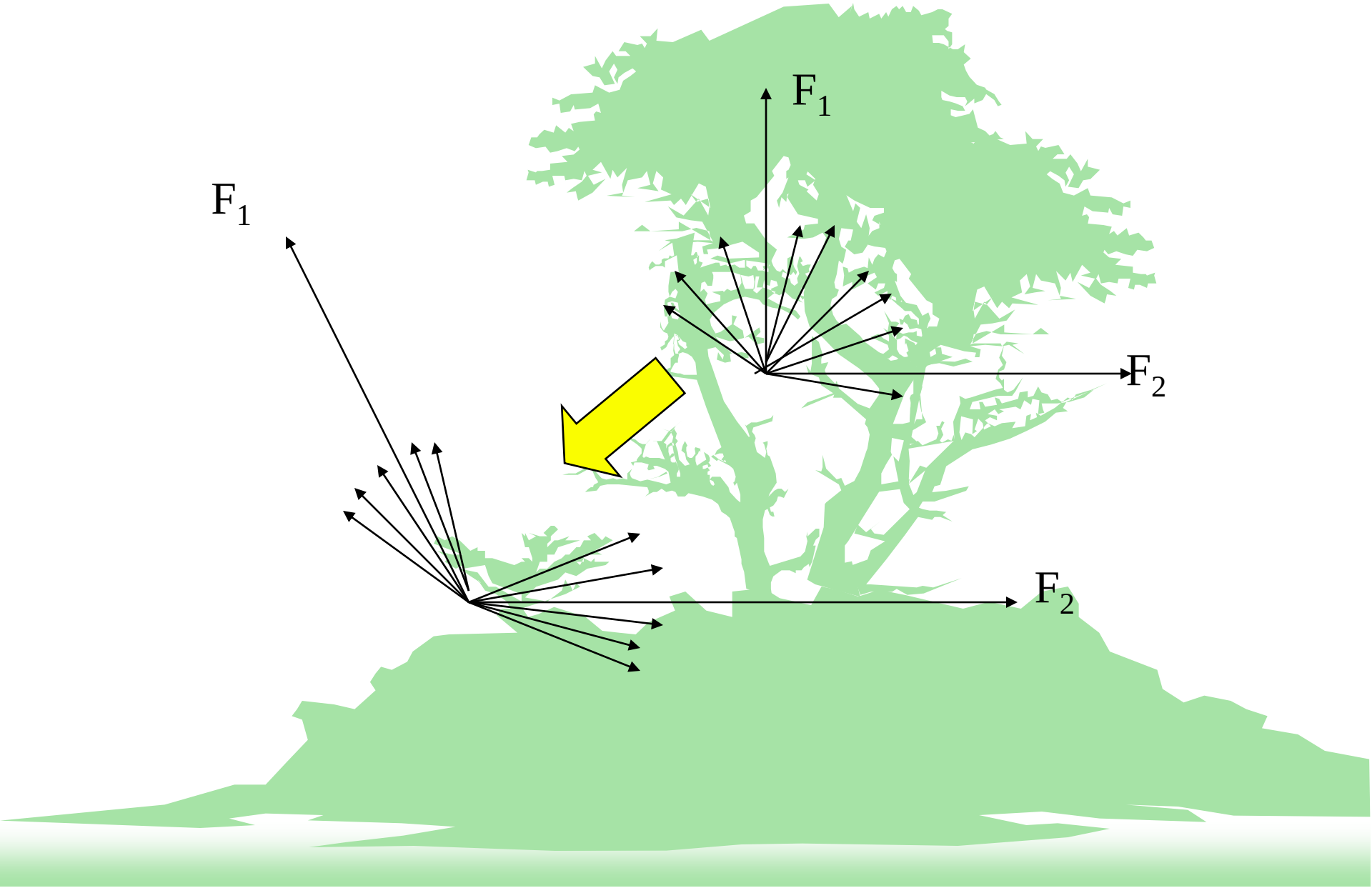
روش های چرخش عامل ها

- عامل های چرخش یافته باید متعامد (Orthogonal) یا متمایل (Oblique)باشد.
- اگر پژوهشگر بخواهد نتایج حاصل از تحلیل او بهترین برازش 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 برازش را دارد، چرخش متعامد راه حلی به دست می دهد که با داده های گذشته و آینده بهترین برازش 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 درك تر و تفسیر پذیر تر است. به نظر می رسد روش واریماکس نیز بهترین روش چرخش متعامد باشد.

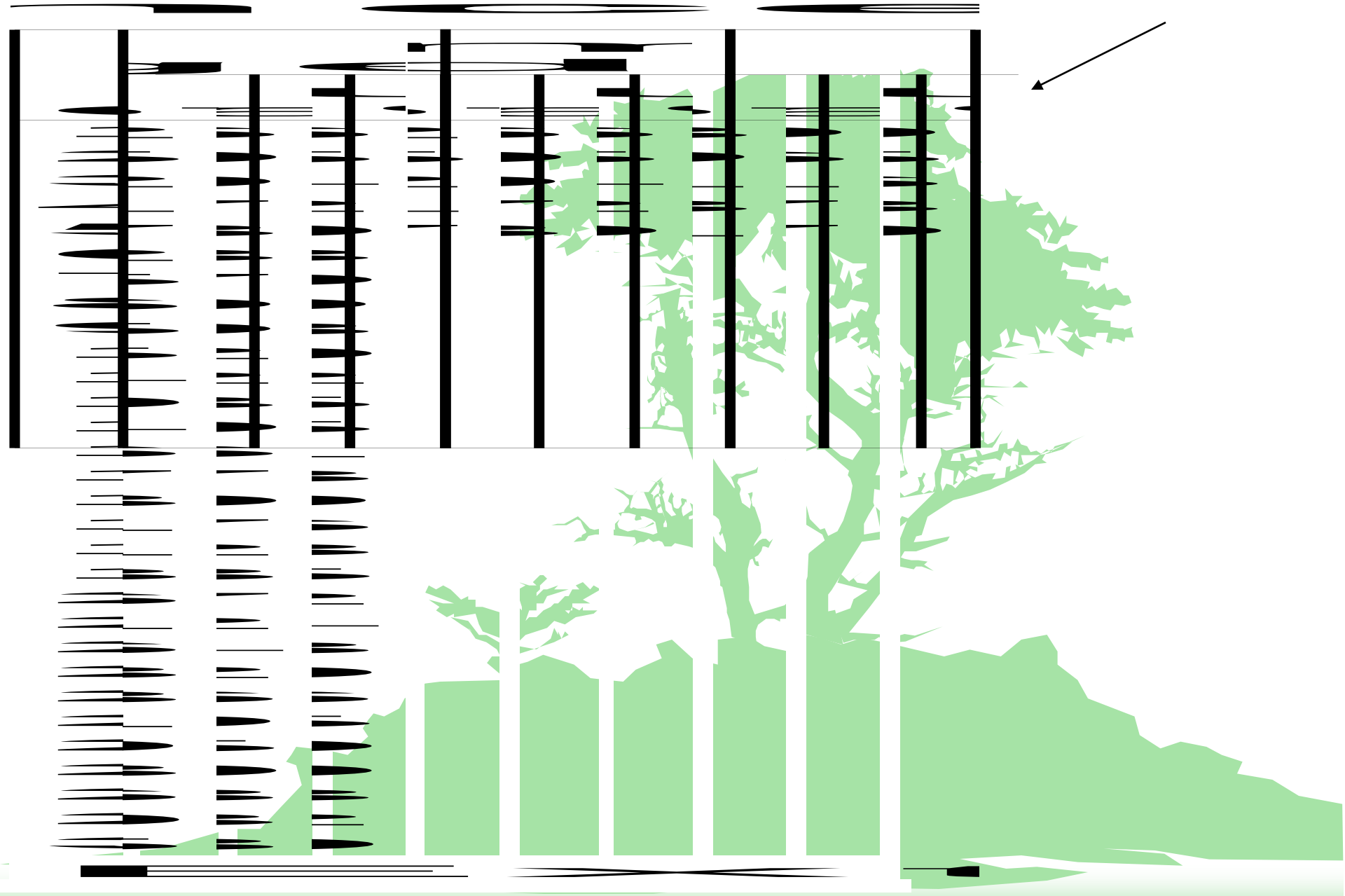
تعداد عامل ها

- يك تصميم عمده در اجرای تحليل عاملی مربوط به تعداد عامل هاست.
- برای چرخش عامل ها می توان تعداد معينی از عامل را انتخاب کرد .بنابراين، می توان ساختارهای چرخش يافته را براساس تعداد مختلفی از عامل را با يکديگر مقايسه کرده و مناسب ترين آنها را انتخاب کرد.
- هرسازه ای که دارای کمينه شرايط لازم برای بارعاملی دست کم روی سه متغير نباشد، از تحليل حذف می شود (Santos & Clegg, 1999).

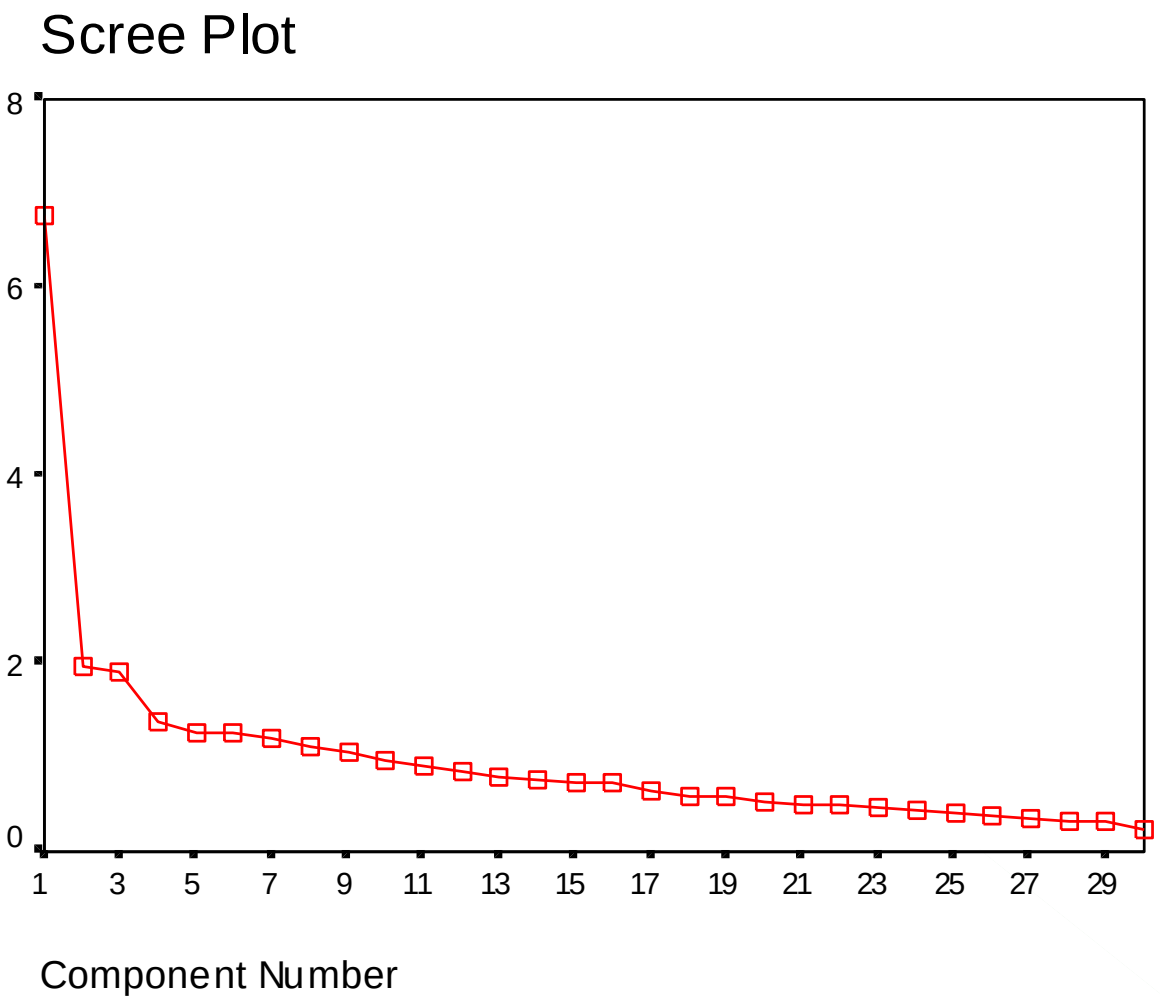
نمایش هندسی چرخش عوامل



جدول کل واریانس تعیین شده



نمودار تشخیص تعداد عامل های معتبر

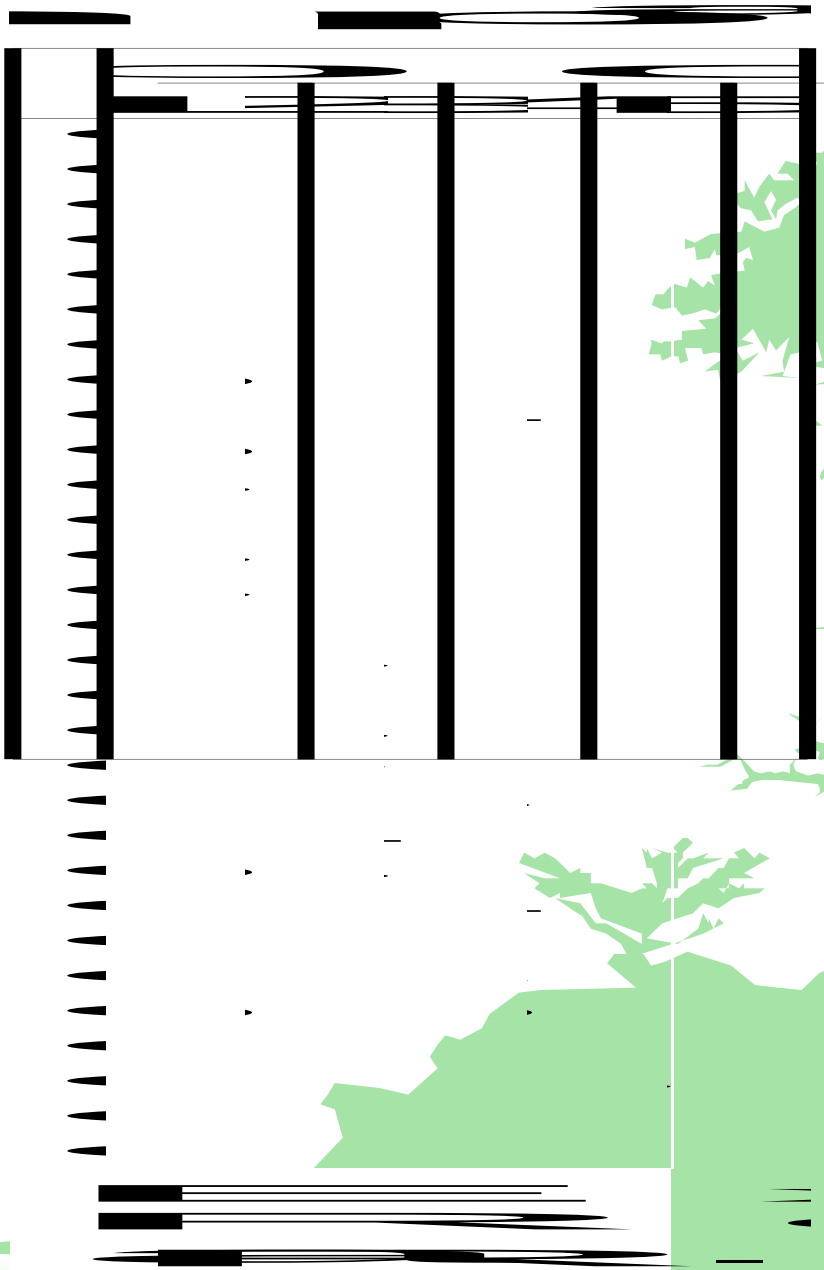


در روش اسکری نمودار مقدار ویژه برای هر عامل ترسیم مي شود. در نقطه ای که شکل منحنی براي مقادیر ویژه به صورت افقی در آید، آن نقطه اسکری (سنگ ریزه) نامیده می شود و عامل هایی که سمت چپ آن قرار دارند عامل های واقعی و بقیه عامل های خطا فلمداد می شوند.

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

تفسیر نتایج تحلیل عاملی



همان طور که در ماتریس دیده می شود متغیرهای مورد نظر روی 5 عامل سوار می شوند.

با توجه به ویژگی های هر خوشه) گروه (از متغیرهایی که روی یك عامل سوار می شوند، باید نام مناسبی انتخاب شود.

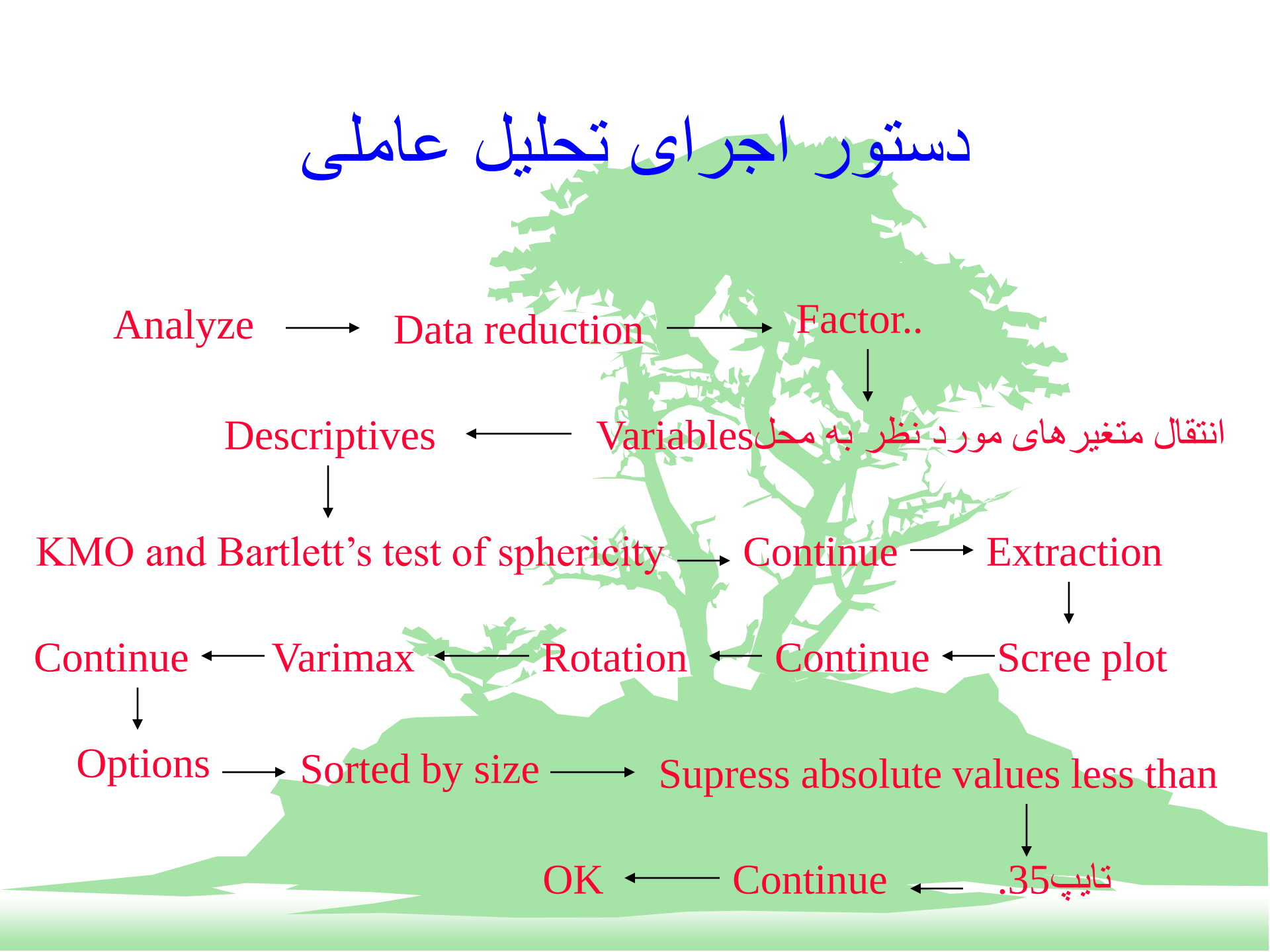
تفسیر نتایج تحلیل عاملی

- اگر مجذور بارهای عاملی را مثلاً برای عامل 1 جمع کنیم، حاصل برابر با 6/75 یعنی برابر با همان ارزش ویژه این عامل است.
- اگر عامل 1 همه واریانس مشترك موجود در متغیرها را تبیین کند، هرکدام دارای بار عاملی 1 و مقدار ارزش ویژه برای عامل 1 برابر با n یعنی تعداد متغیرها خواهد بود.
- در خروجی ماتریس چرخش یافته هر عامل معرف یك خوشه روشن و متمایز است و احتمالاً معرف یك 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 یك 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 ازطریق بیان أین مطلب که آن عامل چه چیز نیست به روشن ساختن تفسیر کمك 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 یك تكنیك طبقه بندی برای تشكیل گروه های همگون در مجموعه پیچیده ای از داده هاست.
- سه هدف عمده یعنی اكتشاف، تایید و ساده سازی را می توان در رابطه با تحلیل خوشه ای مطرح كرد.
- تحلیل خوشه ای در حقیقت یك آزمون آماری نیست بلكه شامل تعدادی از الگوریتم های طبقه بندی است.
- در تحلیل خوشه ای گروه بندی چیزها براساس شباهت ها یا فاصله ها (عدم شباهت ها (انجام می شود.

روش های عمده خوشه بندی

- روش سلسله مراتبی :يك روش متداول در تحليل خوشه ای است كه كاربرد فراوان دارد. در اين روش خوشه ها به دو طريق تشكيل می شوند:
  - از طريق گروه بندی مشاهده ها به صورت بزرگتر و بزرگتر تا همه مشاهده ها اعضای يك خوشه واحد باشند) خوشه بندی تراكمی.(
  - مشاهده ها در يك خوشه واحد قرار گرفته و سپس خوشه ها تقسيم می‌شوند (خوشه بندی بخش پذير.(
- در روش تراكمی چون خوشه بندی با مشاهده های انفرادی شروع می‌شود در ابتدا تعداد خوشه ها برابر با تعداد مشاهده هاست.
- روش خوشه بندی غير سلسله مراتبی: در اين روش هم متغيرها و هم مشاهده‌ها به طور همزمان گروه بندی می شوند. روش k-means يكی از اين روش هاست.

کاربرد خوشه بندی

- خوشه بندی اشخاص(cases)
- خوشه بندی متغیر ها: این روش مشابه تحلیل عاملی است، اما از لحاظ روشی که متغیر ها به گونه گسسته به گروه ها اختصاص می یابند، با آن متفاوت است. تفاوت اساسی در به کاربردن واریانس یك 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 یك منبع بنیادی اختصاص می‌دهد.

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 یك تكنیك 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 یك 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 محلVariables

انتخاب و انتقال متغیری که برمبنای آن باید خوشه بندی صورت گیرد به محل Label Cases by

Plots..

Dendrogram

Continue

OK

مراحل تحلیل خوشه ای

- انتخاب نمونه برای خوشه سازی نظیر خریداران، بیماران، محصولات، کارمندان
- تعریف متغیر هایی که بتوان برمبنای آنها به اندازه گیری اشیا، حوادث، یا افراد پرداخت) مانند وضعیت مالی، وابستگی سیاسی، ویژگی های بهره وری.(
- محاسبه شباهت ها بین موجودیت ها از طریق همبستگی و دیگر تکنیك 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 ویژگیِ 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 می‌دهد. 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 علّی (causal models) است.
- هدف تحلیل مسیر به دست آوردن برآوردهای کمی روابط علّی بین مجموعه ای از متغیرهاست.
- تحلیل مسیر بیان می کند که کدام مسیر مهمتر و یا معنادارتر است.
- ضرایب مسیر براساس ضریب استاندارد شده رگرسیون محـــــاسبه می شود.
- يك 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 Yعلت 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 (exogenious) بوده و بنابراین تبیین تغییر پذیری بین آنها توسط مدل امکان پذیر نباشد.

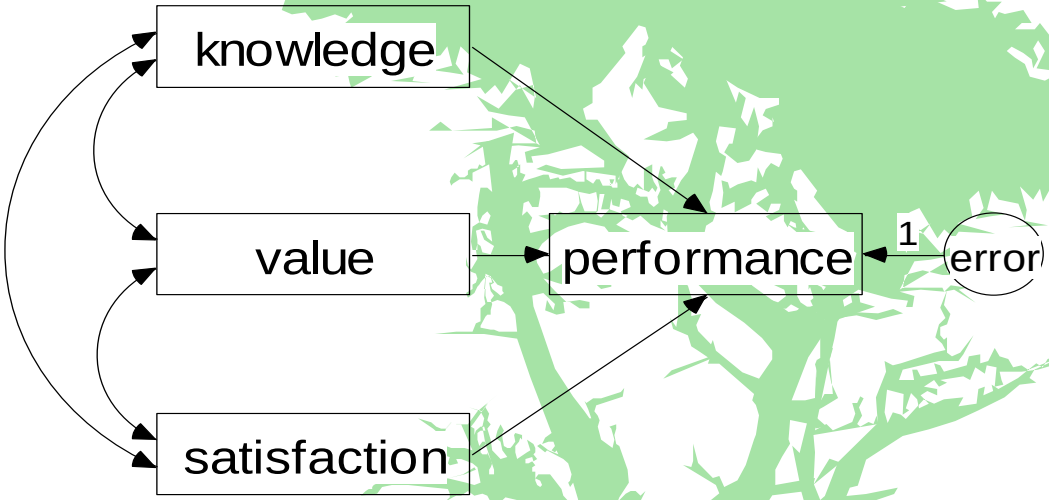
محدویت های تحلیل مسیر

- همه متغیر های مداخله گر باید در تحلیل رگرسیون چند متغیری به عنوان متغیر های 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 علّی 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، می گوییم رابطه کاذب نیست. یعنی در مرحله تست همبستگی دو متغیر دیگر متغیرها به عنوان متغیر کنترل انتخاب شوند.
  - توالی زمانی :(Time order) یعنی علّت 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)