## آزمونهای ناپارامتری:

### - معادل ناپارامتری آزمون فرض در مورد میانگین های دو جامعه :

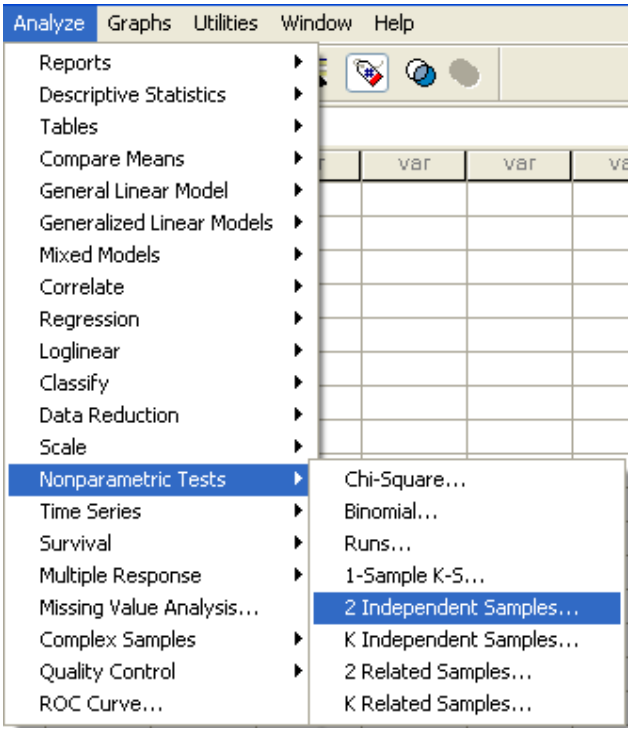
### ۱- آزمون من- ویتنی

زمانی که بخواهیم دو جامعه مستقل را به شیوه ناپارامتری آزمون و مقایسه کنیم, از آزمون من- ویتنی استفاده می کنیم .

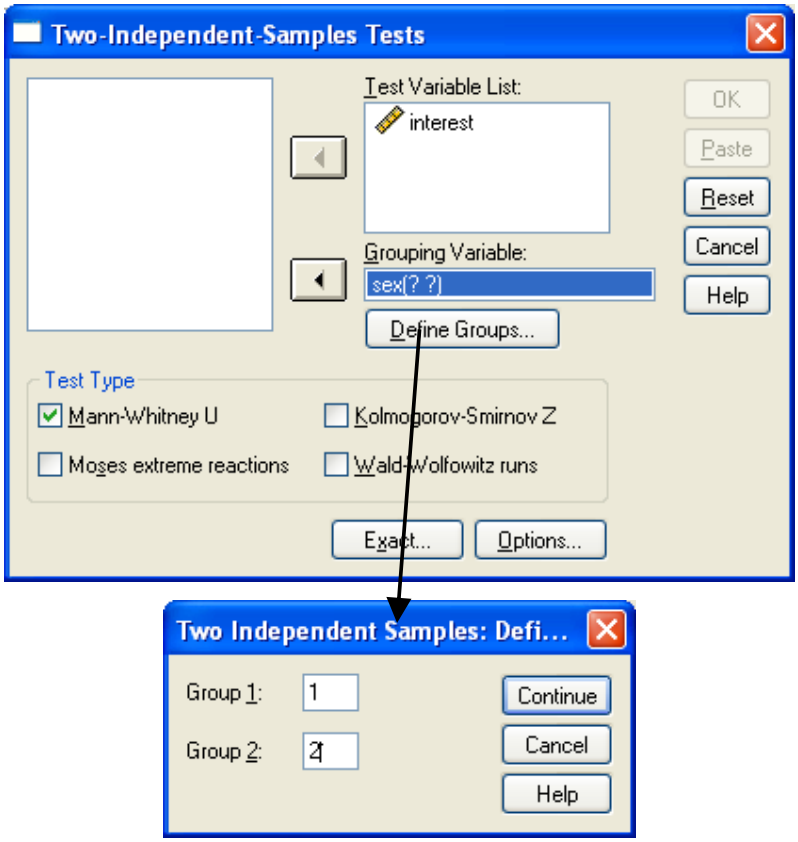
**مثال :** در یک بررسی می‌خواهیم ببینیم که آیا میزان علاقهٔ آقایان و خانمها به آشپزی یکسان است یا خیر؟ بدین منظور داده‌های زیر را از طریق پرسشنامه و دادن رتبه به میزان علاقهٔ افراد به آشپزی بدست آورده‌ایم:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **خانمها** | ۵ | ۴ | ۴ | ۳ | ۵ | ۴ | ۳ | ۵ | ۴ | ۴ |
| **آقایان** | ۲ | ۳ | ۱ | ۲ | ۳ | ۵ | ۳ | ۴ | ۳ | ۲ |

برای انجام این مثال مسیر زیر را طی کنید:



متغیر مورد آزمون را به قسمت Test variable list ببرید و متغیری که گروهبندی را مشخص می کند به قسمت Grouping variable ببرید و مانند قبل گروهها را برای آن تعریف کنید.




بر دکمه **continue** کلیک کنید و در قسمت **Test type** گزینه **Mann-Whitney U** را تیک بزنید و پس از آن بر دکمه Ok کلیک کرده تا خروجی به صورت زیر ظاهر گردد.

## Mann-Whitney Test

**Ranks**

| | sex | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| interest | ۱,۰۰ | ۱۰ | ۱۳,۸۰ | ۱۳۸,۰۰ |
| | ۲,۰۰ | ۱۰ | ۷,۲۰ | ۷۲,۰۰ |
| | Total | ۲۰ | | |

**Test Statistics[b]**

| | interest |
|---|---|
| Mann-Whitney U | ۱۷,۰۰۰ |
| Wilcoxon W | ۷۲,۰۰۰ |
| Z | -۲,۵۷۷ |
| Asymp. Sig. (۲-tailed) | .۰۱۰ |
| Exact Sig. [۲*(۱-tailed Sig.)] | .۰۱۱[a] |

a. Not corrected for ties.

b. Grouping Variable: sex

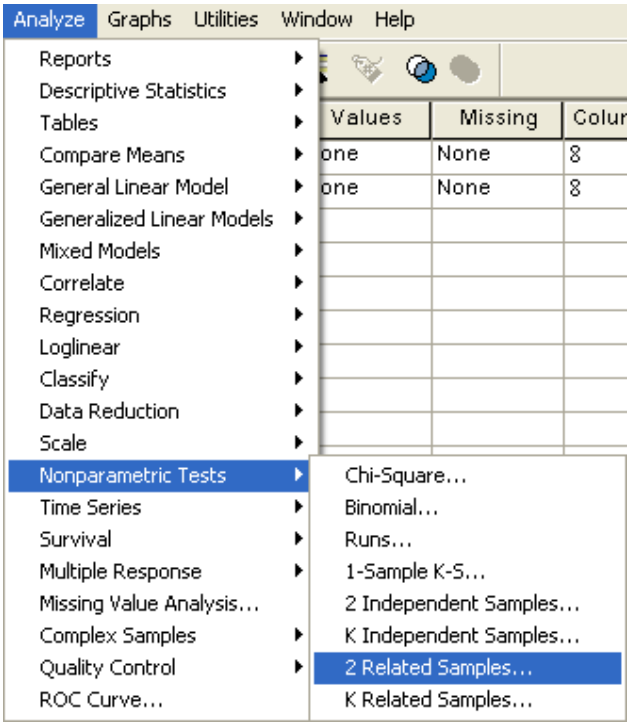
**تذکر :** در استفاده از آزمون من- ویتنی توجه داشته باشید که مقیاس اندازه گیری نمونه ها باید رتبه ای (با تعداد مقوله های کم ) و یا فاصله ای باشد.

## ۲- آزمون های علامت, ویلکاکسون( آزمون رتبهٔ علامت دار) و مک نمار:

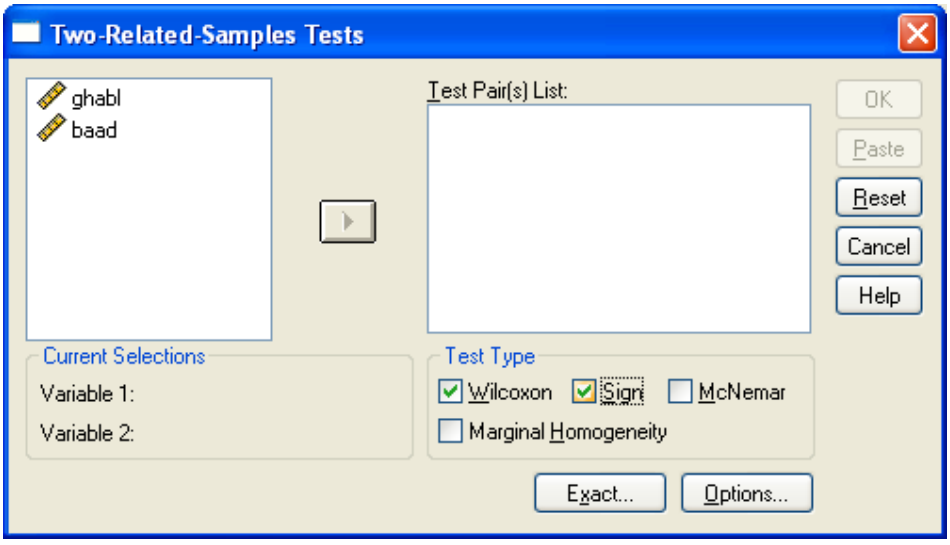
زمانی که بخواهیم میانگینهای نمونه های جفتی(زوجی) را از طریق آزمونهای ناپارامتری مقایسه و آزمون کنیم, از آزمون های **علامت, ویلکاکسون, و مک نمار** استفاده می کنیم. اما هر یک از این سه آزمون فرقهایی با یکدیگر دارند.

**مثال (آزمون علامت و ویلکاکسون):** می خواهیم مثال مربوط به وزن دانش آموزان ( حاصل انجام یک ماده ورزش مرتب ) را با استفاده از آزمون های ناپارامتری آزمون کنیم.

برای این کار مسیر زیر را طی کنید:




تا کادر مکالمه ای زیر باز شود .مانند حالت پارامتری دو متغیر ghabl و baad را انتخاب و آنها را به قسمت Test Pair(s) list منتقل کنید. در قسمت Test type گزینه های Wilcoxon و Sign را تیک دار کنید .




Ok را کلیک کنید تا نتایج بصورت زیر برای شما ظاهر شود :

# Wilcoxon Signed Ranks Test

**Ranks**

| | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| baad – ghabl | Negative Ranks | 7[a] | 4.79 | 33.50 |
| | Positive Ranks | 1[b] | 2.50 | 2.50 |
| | Ties | 0[c] | | |
| | Total | 8 | | |

a. baad < ghabl

b. baad > ghabl

c. baad = ghabl

**Test Statistics[b]**

| | baad – ghabl |
|---|---|
| Z | -2.200[a] |
| Asymp. Sig. (2-tailed) | .028 |

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

# Sign Test

**Frequencies**

| | | N |
|---|---|---|
| baad – ghabl | Negative Differences[a] | 7 |
| | Positive Differences[b] | 1 |
| | Ties[c] | 0 |
| | Total | 8 |

a. baad < ghabl

b. baad > ghabl

c. baad = ghabl

**Test Statistics[b]**

| | baad – ghabl |
|---|---|
| Exact Sig. (2-tailed) | .070[a] |

a. Binomial distribution used.

b. Sign Test

**تذکر:** در مقایسه های زوجی در صورتی که دادهٔ پرت داشتیم ( از نمودار پراکنش دو متغیر استفاده کنید ) یا نمی توانستیم از آزمون t زوجی استفاده کنیم, می توانیم از روشهای ناپارامتری جهت آنالیز استفاده کنیم. توجه داشته باشید که آزمون علامت نسبت به داده های پرت کاملا ایمن است, در حالی که آزمون ویلکاکسون اینگونه نیست. (فرق آزمونهای علامت و ویلکاکسون )

**تذکر:** توجه داشته باشید که در آزمون های ویلکاکسون و علامت الزامی در مورد شکل توزیع نداریم ولی متغیر مورد بررسی می باید که پیوسته باشد.

## مثال ( آزمون مک نمار) :

یک محقق نظر افراد را در مورد موافق یا مخالف بودن افراد نسبت به یک کاندیدا را در قبل و بعد از سخنرانی مورد بررسی قرار می دهد. داده های حاصل به صورت زیر است:

| بعد از انجام سخنرانی / قبل از انجام سخنرانی | موافق | مخالف |
|---|---|---|
| موافق | ۲۰ | ۶۰ |
| مخالف | ۱۰ | ۱۰ |

می خواهیم بدانیم که آیا نظر افراد در دو وضعیت متفاوت است یا خیر؟

با استفاده از آزمون مک نمار و بدست آوردن خرجی زیر در SPSS متوانیم نتیجه بگیریم که نگرش افراد در قبل و بعد از سخنرانی متفاوت بوده است.( چرا؟ )

**McNemar Test**

**Crosstabs**

**before & after**

| before | after 1 | after 2 |
|---|---|---|
| 1 | 20 | 60 |
| 2 | 10 | 10 |

**Test Statistics[b]**

| | before & after |
|---|---|
| N | 100 |
| Chi-Square[a] | 34.300 |
| Asymp. Sig. | .000 |

a. Continuity Corrected

b. McNemar Test

**تمرین :**

در یک مطالعه مربوط به تعیین میزان توانایی تداعی پستانداران به جز انسان, ۱۹ میمون همسن به طور تصادفی به ۲ گروه ۹ و ۱۰ تایی تقسیم می شوند. دو گروه مزبور را برای به خاطر آوردن یک محرک مربوط به صدا, با دو روش آموزش مختلف تعلیم می دهند. نمرات میمون ها در آزمون بعد از آموزش عبارتند از :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| روش۱ | ۱۶۷ | ۱۴۹ | ۱۳۷ | ۱۷۸ | ۱۷۹ | ۱۵۵ | ۱۶۴ | ۱۰۴ | ۱۵۱ | ۱۵۰ |
| روش۲ | ۹۸ | ۱۲۷ | ۱۴۰ | ۱۰۳ | ۱۱۶ | ۱۰۵ | ۱۰۰ | ۹۵ | ۱۳۱ | |

آیا داده ها قویا تفاوت را در قدرت حافظه میمونهای تعلیم داده شده با دو روش نشان میدهد ؟

**تطابق توزیع ها :**

همانطور که در بخشهای پیش گفته شد یکی از اهداف بررسی یک جامعه این است که محقق بخواهد توزیع نمونه را با یک توزیع فرضی مانند توزیع نرمال بررسی کند. به عبارتی سئوال در مورد تطابق ( Goodness-of-fit ) به این گونه آزمونها در اصطلاح **آزمونهای نیکویی برازش** می گویند.

**الف – داده های کمی (آزمون کلموگروف – اسمیرنف ) :**

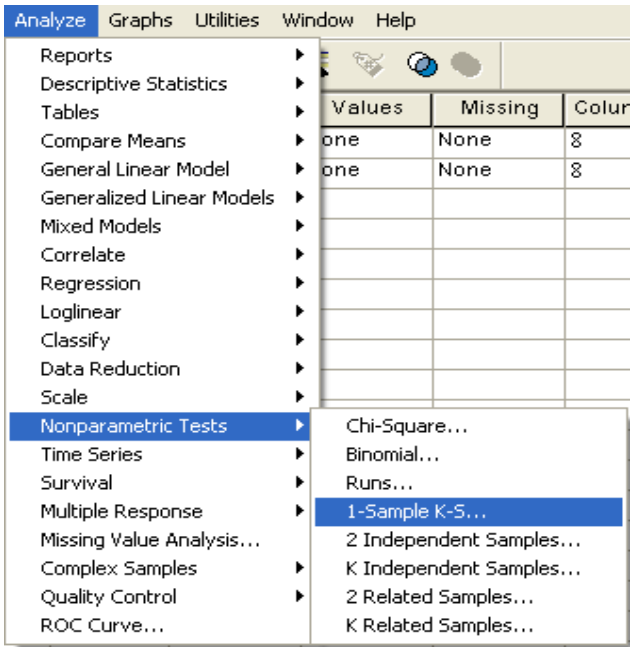
زمانی که برای داده های کمی بخواهیم آزمون نیکویی برازش را بکار ببریم از آزمون کلموگروف – اسمیرنف استفاده می کنیم . به عبارتی می خواهیم فرضهای زیر را آزمون کنیم :

$$\begin{cases} H_0 : F(x) = F_0(x) \\ H_1 : F(x) \neq F_0(x) \end{cases}$$

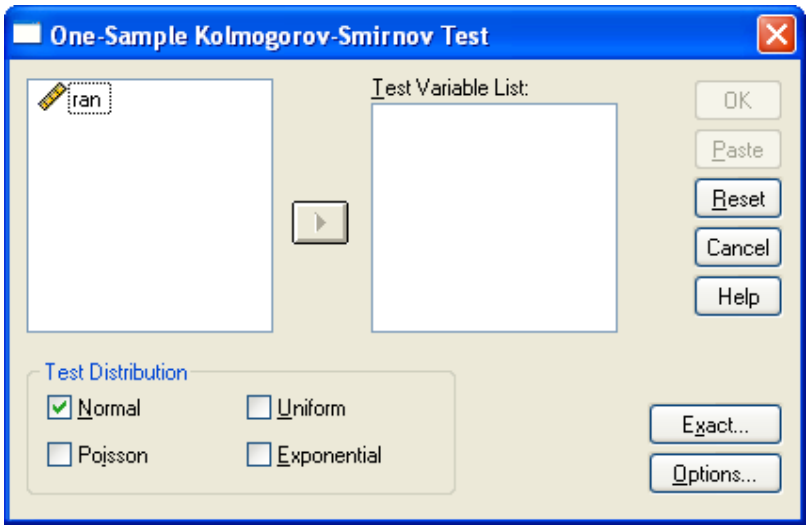
**مثال :** می خواهیم ببینیم که داده های زیر از توزیع نرمال پیروی می کنند یا خیر؟

| | ran |
|---|---|
| 1 | 3.50 |
| 2 | 5.33 |
| 3 | 3.87 |
| 4 | 8.19 |
| 5 | 8.72 |
| 6 | 7.05 |
| 7 | 6.37 |
| 8 | 8.76 |
| 9 | 2.81 |
| 10 | -.46 |
| 11 | 5.70 |
| 12 | 9.25 |
| 13 | 5.11 |
| 14 | 5.14 |
| 15 | 10.38 |
| 16 | 1.86 |
| 17 | 7.17 |
| 18 | 3.69 |
| 19 | 10.69 |
| 20 | 10.44 |

برای انجام این کار مسیر زیر را طی می کنیم :




تا کادر محاوره ای زیر ایجاد شود :




متغیر ran را به قسمت Test variable list وارد کنید . گزینه Normal را تیک بزنید و سپس دکمه Ok را کلیک کنید تا خروجی را به صورت زیر ببینید :

**NPar Tests**

**One-Sample Kolmogorov-Smirnov Test**

| | | RAN |
|---|---|---|
| N | | 20 |
| Normal Parameters[a,b] | Mean | 6.1788 |
| | Std. Deviation | 3.08160 |
| Most Extreme Differences | Absolute | .095 |
| | Positive | .073 |
| | Negative | -.095 |
| Kolmogorov-Smirnov Z | | .424 |
| Asymp. Sig. (2-tailed) | | .994 |

a. Test distribution is Normal.

b. Calculated from data.

**تذکر:** آزمون کلموگروف – اسمیرنف از مقایسه تابع تجمعی احتمال مشاهدات و تابع احتمال توزیع مفروض, فرض پیروی مشاهدات از توزیع احتمالی خاص را پیروی می کند .

## ب – داده های دو حالتی اسمی :

فرض کنید سکه ای را ۳۰ مرتبه پرتاب کرده اید و نتایج به صورت زیر بدست آمده است ( ۱= شیر , ۰= خط ) :

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1

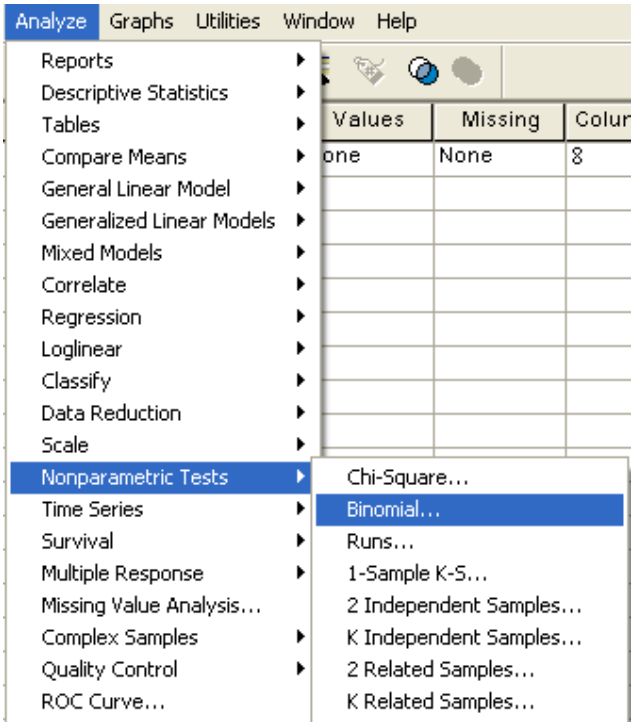
می خواهیم آزمون کنیم که آیا سکه سالم است یا خیر ؟ می خواهیم فرض های زیر را آزمون کنیم:

$H_0 = p = 0.5$

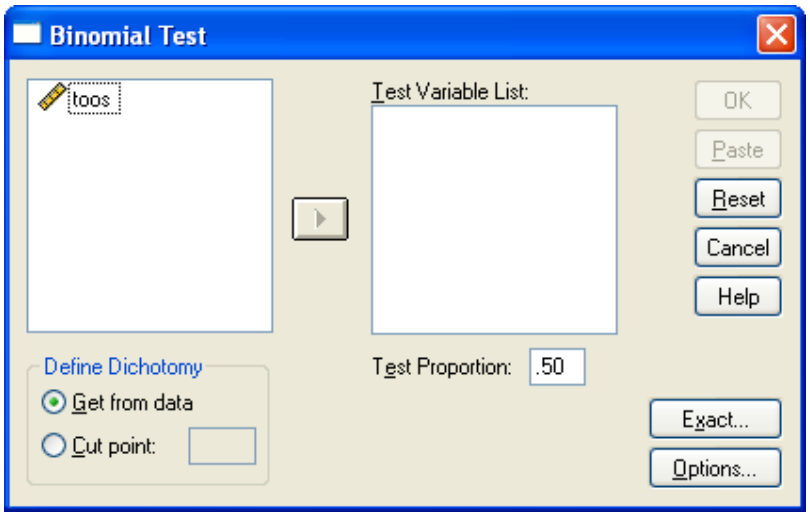
$H_1 = p \neq 0.5$

برای آزمون این فرض از آزمون دوجمله ای بهره می گیریم :

برای انجام این کار داده های بالا را به عنوان متغیر toss وارد کرده و مسیر زیر را طی کنید :




کادر مکالمه ای زیر را مشاهده می کنید :




متغیر toss را به قسمت Test variable list وارد کرده و مقدار Test proportion را ۰/۵ وارد کنید و در قسمت Define Dichotomy گزینه اول را انتخاب کنید و سپس دکمه Ok را کلیک کنید تا خروجی زیر ظاهر شود :

**NPar Tests**

**Binomial Test**

| | | Category | N | Observed Prop. | Test Prop. | Asymp. Sig. (2-tailed) |
|---|---|---|---|---|---|---|
| TOOS | Group 1 | 1.00 | 18 | .60 | .50 | .362[a] |
| | Group 2 | .00 | 12 | .40 | | |
| | Total | | 30 | 1.00 | | |

a. Based on Z Approximation.

با توجه به این خروجی به سئوالی که در ابتدای بحث مطرح شد پاسخ دهید.

## تمرین :

دو نوع اسباب بازی مکانیکی و غیر مکانیکی داریم.می خواهیم ببینیم که آیادرکودکان ۵ ساله تمایل به سمت هرکدام از این اسباب بازیها یکسان است؟ بدین منظور از ۲۰۰ کودک ۵ ساله خواسته ایم که به دلخواه یکی از این اسباب بازی ها را انتخاب کنند و نتایج به صورت زیر بدست آمده است:

| | |
|---|---|
| مکانیکی | ۱۲۰ |
| غیرمکانیکی | ۸۰ |

از این داده ها چه استنباطی دارید ؟

**تمرین :**

وزن ۲۰ بازیکن فوتبال که به تصادف از تیمهای فوتبال یک شهر انتخاب شده اند به صورت زیر است:

| | | | |
|---|---|---|---|
| ۸۱ | ۷۱/۵ | ۷۵ | ۶۴ |
| ۸۳ | ۷۰ | ۶۹/۵ | ۶۷ |
| ۹۴ | ۵۸/۵ | ۹۵ | ۹۰ |
| ۹۷ | ۸۸ | ۶۲ | ۵۹ |
| ۸۰ | ۶۱/۵ | ۶۰ | ۵۹/۵ |

**آیا میتوان گفت که :**

**الف** –۶۰ درصد از بازیکنان فوتبال این شهر کمتر از $65kg$ وزن دارند؟

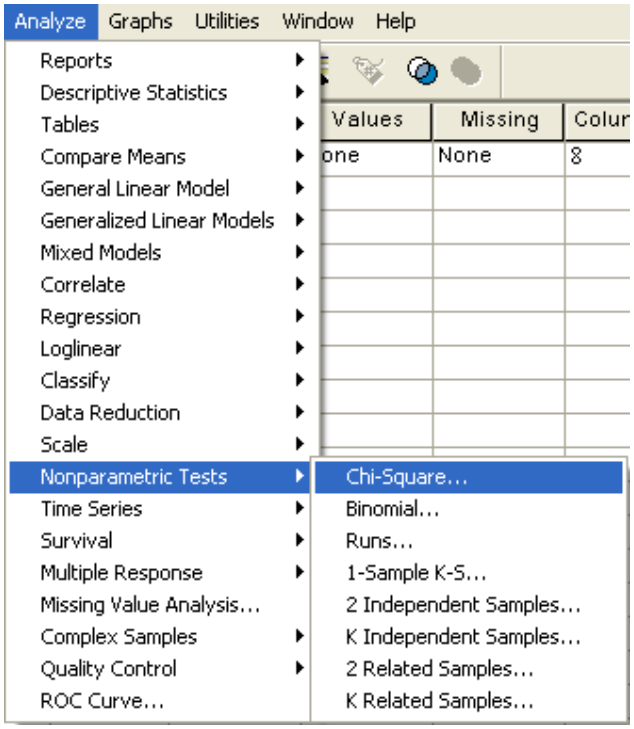
**ب**– ۴۰ درصد از بازیکنان فوتبال این شهر بیشتراز $60kg$ وزن دارند ؟

## پ ) آزمون تطابق توزیع با سه طبقه یا بیشتر ( آزمون Chi – Square ) :

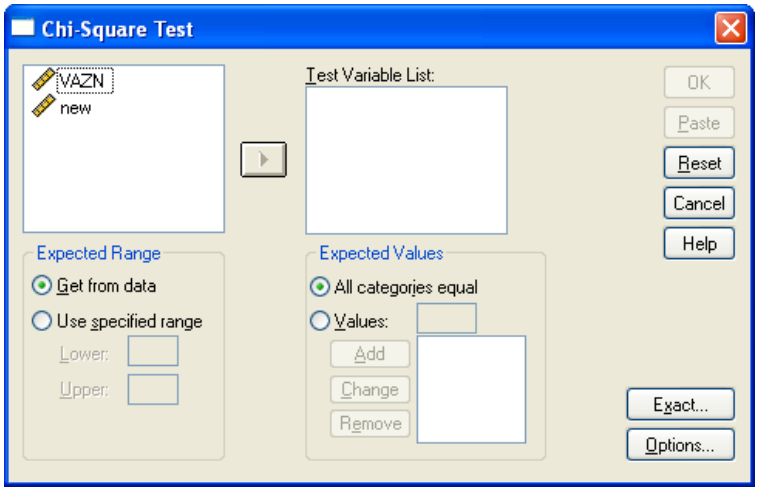
فرض کنید می خواهیم آزمون کنیم که آیا نسبت با سوادی در گروه هایی با درآمدهای مختلف یکسان است یا خیر؟ یا توزیع نمرات دانشجویان یک درس در بازه های مختلف یکسان است یا خیر؟ و به طور کلی می خواهیم توزیع نسبت مقادیر یک متغیر در رده های مختلف را با توزیع از پیش تعیین شده ای مقایسه کنیم . برای این کار آزمون Chi – Square را به کار می بریم.

**مثال :** داده های صفحهٔ ۲۰ جزوه را در نظر بگیرید. می خواهیم آزمون کنیم که آیا اوزان کره ها در ۷ رده وزنی ساخته شده بطور یکسانی توزیع شده اند یا خیر ؟

برای این کار مسیر زیر را طی کنید :




تا کادر مکالمه ای زیر باز شود :




متغیرnew را به قسمت Test variable list وارد کرده و Ok را کلیک کنید تا خروجی را به صورت زیر مشاهده کنید :

**Chi-Square Test**

**Test Statistics**

| | NEW |
|---|---|
| Chi-Square[a] | 7.950 |
| df | 6 |
| Asymp. Sig. | .242 |

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.7.

در قسمت Expected Range می توانید برد داده ها, و در قسمت Expected Values نیز می توانید نسبت رده را خودتان تعیین کنید .

## آزمون تصادفی بودن ( گردش ) :

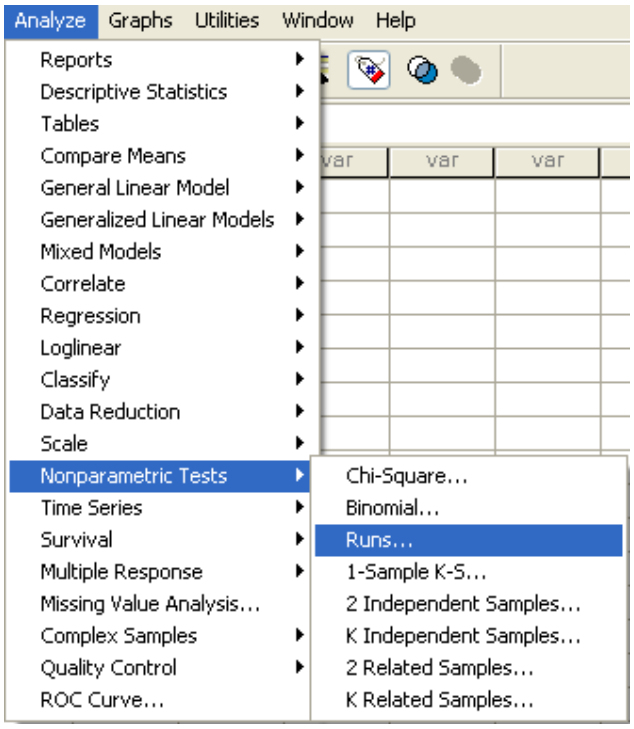
در بسیاری از آنالیزهای آماری یکی از فرضهای اساسی "تصادفی بودن مشاهدات"است. برای مثال در بررسی مانده های یک مدل آماری , برقراری فرض تصادفی بودن (استقلال مشاهدات ) حیاتی است.انحراف از فرض تصادفی بودن می تواند به دلیل وجود روندهای افزایشی یا کاهشی, رفتارهای دوره ای یا افزایش تغییرپذیری و برخی علل دیگر باشد.

**مثال :** فرض کنید نتایج ۲۰ بار پرتاب یک سکه بصورت زیر است (۱= شیر , ۰= خط ) :

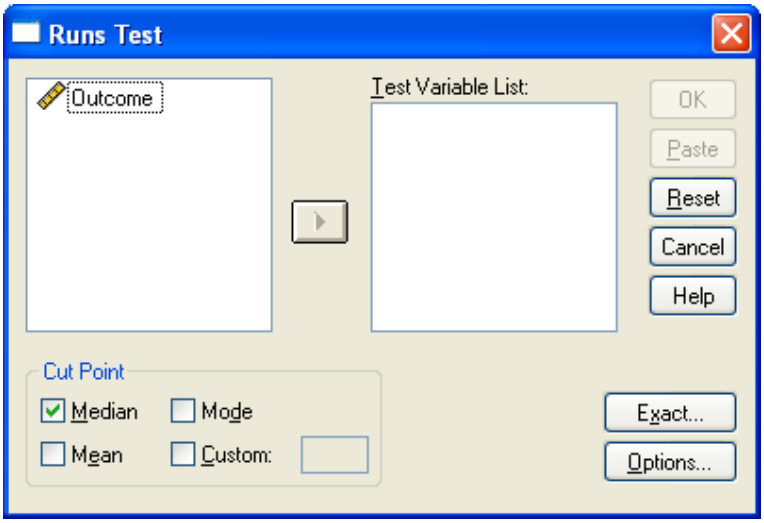
0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1

می خواهیم بدانیم که آیا این نتایج تصادفی است یا خیر ؟

برای انجام این کار ابتدا داده ها را به عنوان متغیر Outcome وارد SPSS کنید و سپس مسیر زیر را طی کنید :




تا کادر محاوره ای زیر باز شود :




در کادر محاوره ای بالا متغیر Outcome را به قسمت Test variable list برده و در قسمت Cut point خط مبنای تشخیص نوسانات تصادفی را انتخاب کنید ( برای مثال میانه را انتخاب کنید) و دکمه Ok را کلیک کنید و خروجی را به صورت زیر مشاهده کنید :

## NPar Tests

**Runs Test**

| | OUTCOME |
|---|---|
| Test Value[a] | .50 |
| Cases < Test Value | 10 |
| Cases >= Test Value | 10 |
| Total Cases | 20 |
| Number of Runs | 20 |
| Z | 3.905 |
| Asymp. Sig. (2-tailed) | .000 |

a. Median

با توجه به این خروجی به سئوالی که در ابتدای بحث مطرح شد پاسخ دهید .

با انتخاب هر خط مبنایی, کوچکتر از آن, کد **صفر** و به مقادیر بزرگتر یا مساوی با آن, کد **یک** خواهند داشت و با ساختن دنباله مشاهدات با مقادیر صفر و یک آزمون گردش اجرا می شود .

**تمرین :**

تعداد دانشجویانی را که ساعت ۸ صبحِ ۲۰ روز تحصیلی, یک اتوبوس از خوابگاه به دانشگاه می برد, اعداد زیر می باشد :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ۲۱ | ۲۴ | ۲۳ | ۱۵ | ۳۲ | ۲۸ | ۲۶ | ۱۷ | ۲۰ | ۲۸ |
| ۳۰ | ۲۴ | ۱۳ | ۳۵ | ۲۶ | ۲۱ | ۱۹ | ۲۹ | ۲۷ | ۱۸ |

آیا این داده ها یک نمونه تصادفی اند؟

# ارتباط و رابطه بین متغیرها

در تجزیه و تحلیل آماری بجز اطلاعاتی که در مورد هر متغیر بدست می آوریم، علاقه‌مندیم تا نحوهٔ ارتباط متغیرها را شناسایی کنیم:

– آیا رابطه ای بین قد شخص با وزن او وجود دارد؟

– آیا معدل دانش آموزان رابطه ای با شغل والدین دارد؟

– آیا رنگ پوست یک مجرم تاثیری در رای دادگاه علیه وی دارد؟

–آیا بعد از یک عمل جراحی شدت ضربان قلب با مقدار داروی بیهوشی مصرف شده ارتباط دارد؟

–آیا مقدار محصول یک مزرعه با نوع کود مصرفی در آن ارتباط دارد؟

در حقیقت در اکثر بررسی های آماری با سئوالاتی ازاین دسته به وفور برخورد خواهیم کرد. ولی با تأملی در متغیرهای به کار رفته در هر مورد می بینیم که نمی توان تمامی سئوالات فوق را در یک دسته قرار داد. در واقع بسته به نوع متغیرهایی که مایل به کشف رابطه بین آنها هستیم، روش های آماری مختلفی را به کار خواهیم گرفت. برای شروع کار فرض می کنیم تنها با ۲ متغیر روبرو هستیم، بسته به نوع دو متغیر نحوه انجام آنالیزهای آماری را شرح می دهیم:

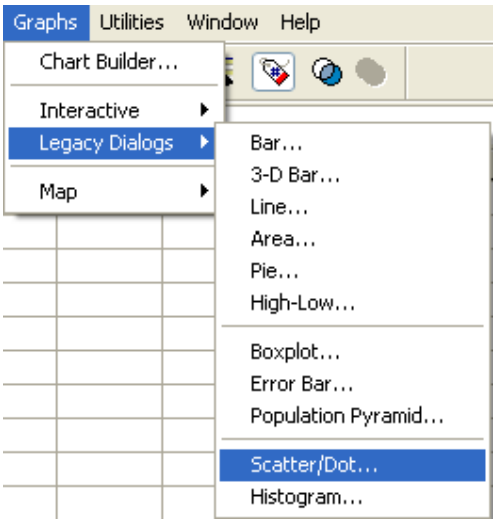
## ۱–ارتباط بین متغیرهای عددی:

نمودار پراکنش وسیله ی مناسبی برای کشف رابطه بین متغیرهای عددی است. قبل از بدست آوردن ضریب همبستگی پیرسون ( برای دو متغیر عددی ) بایستی که نمودار پراکنش داده ها را رسم کنیم.

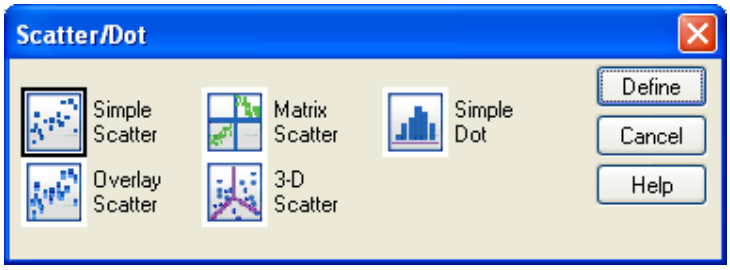
**مثال (***نمودار پراکنش (Scatter plot) و بررسی معنی داری رابطه* **):** جهت بررسی رابطهٔ بین ضریب هوشی فرد و نمره ریاضی از ۱۰ نفر، این صفات در جدول زیر ثبت شده است :

| نمره ریاضی | ۳۱ | ۲۵ | ۲۰ | ۲۴ | ۱۸ | ۲۸ | ۱۸ | ۲۰ | ۱۶ | ۲۷ |
|---|---|---|---|---|---|---|---|---|---|---|
| ضریب هوشی | ۱۲۰ | ۱۱۲ | ۱۱۰ | ۱۲۳ | ۱۰۵ | ۱۲۵ | ۱۱۴ | ۱۱۳ | ۱۰۶ | ۱۲۷ |

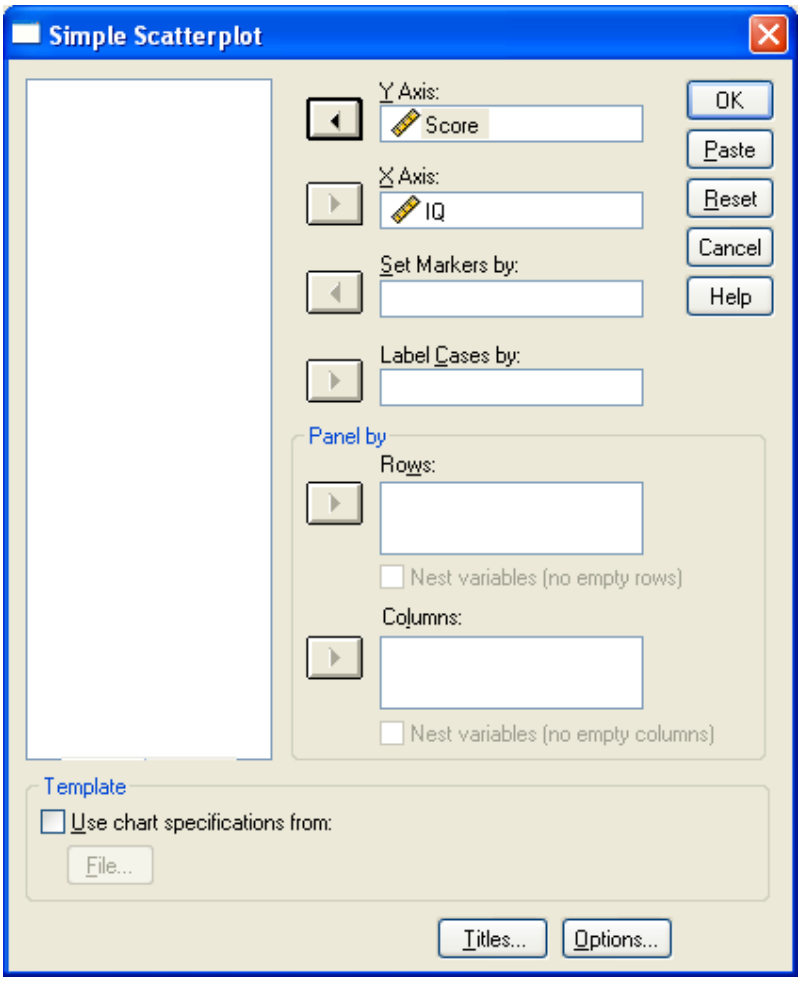
ابتدا داده ها را در دو ستون، در محیط SPSS وارد کنید.حال مسیر زیر را برای رسم نمودار پراکنش طی کنید تا کادر مربوطه باز شود. متغیر ها را به صورت زیر نامگذاری و وارد کنید :



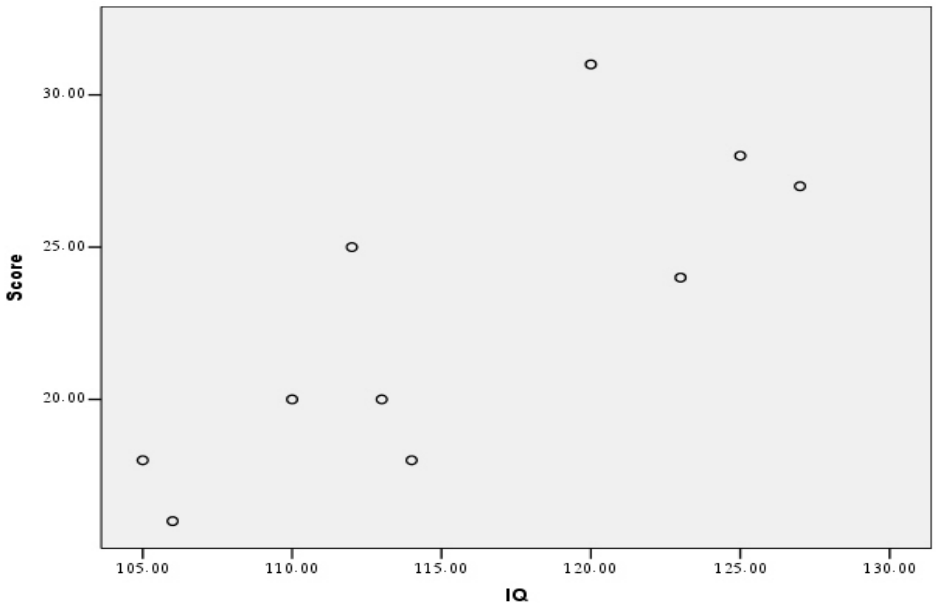

در کادر زیرگزینه Simple Scatter را انتخاب کنید و دکمه Define را کلیک کنید:




تا کادر مکالمه ای زیر باز شود:




دکمه Ok را کلیک کنید. نمودار رسم شده به صورت زیر خواهد بود:



ملاحظه می کنید که با افزایش مقادیر IQ، مقادیر متناظر Score هم افزایش می یابند و بـه نظـر می رسد مقادیر IQ وScore با هم رابطه دارند. با اینکه نمودار پراکنش وسیله خوبی برای بیـان رابطـه بین دو متغیر عددی است، ولی علاقه مندیم بوسیله یک شاخص آماری شدت این رابطه را بیان کنیم.

شاخص آماری که برای نشان دادن سطح ارتباط دو متغیر عددی به کار می رود ضریب همبستگی است که عددی است بین ۱- و ۱. هر چه ضریب همبستگی به صفر نزدیکتر باشد. نشان دهنده عدم وجود رابطه بین دو متغیر است ضریب همبستگی مثبت، رابطه هم جهت را نشان می‌دهد و ضریب همبستگی منفی، مبین رابطه عکس است. در SPSS از ۳ نوع ضریب همبستگی استفاده می شود:

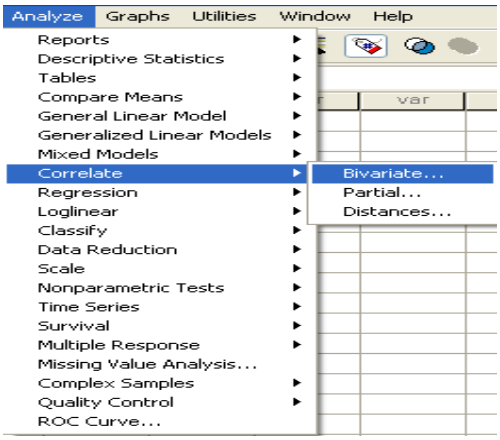
۱) *ضریب همبستگی ساده (Pearson)* : در حالتی که هر دو متغیر دارای مقیاس فاصله ای باشند به کار می رود.

۲) *ضریب همبستگی Spearman* : به جای مقادیر متغیرها، ضریب همبستگی Spearman با رتبه های دو متغیر محاسبه می شود. ضریب همبستگی Spearman در مواردی که حداقل یکی از متغیرها ordinal باشد، به کار می رود.

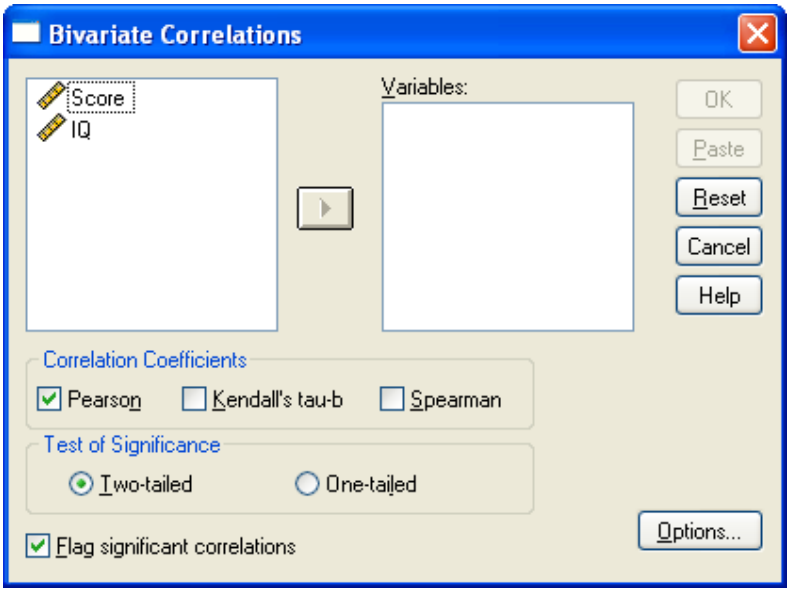
۳) *ضریب همبستگی kendall's tau-b* : همانند ضریب همبستگی Spearman یک ضریب همبستگی ناپارامتری است و هنگامی که مقیاس حداقل یکی از متغیرها ordinal باشد به کار خواهد رفت.

برای داده های IQ و Score مثال قبل ضریب همبستگی ساده ( Pearson ) و Spearman را محاسبه کنید.

برای محاسبه ضریب همبستگی ساده مسیر زیر را طی کنید:



در کادر ظاهر شده، هر دو متغیر IQ وScore را انتخاب کنید (توجه کنید که در قسمت Correlation coefficient، ضریب همبستگی Pearson انتخاب شده باشد):




سپس دکمه ok را فشار دهید. خروجی به صورت زیر خواهد بود:

## Correlations

**Correlations**

| | | IQ | Score |
|---|---|---|---|
| IQ | Pearson Correlation | 1 | .795** |
| | Sig. (2-tailed) | | .006 |
| | N | 10 | 10 |
| Score | Pearson Correlation | .795** | 1 |
| | Sig. (2-tailed) | .006 | |
| | N | 10 | 10 |

**. Correlation is significant at the 0.01 level

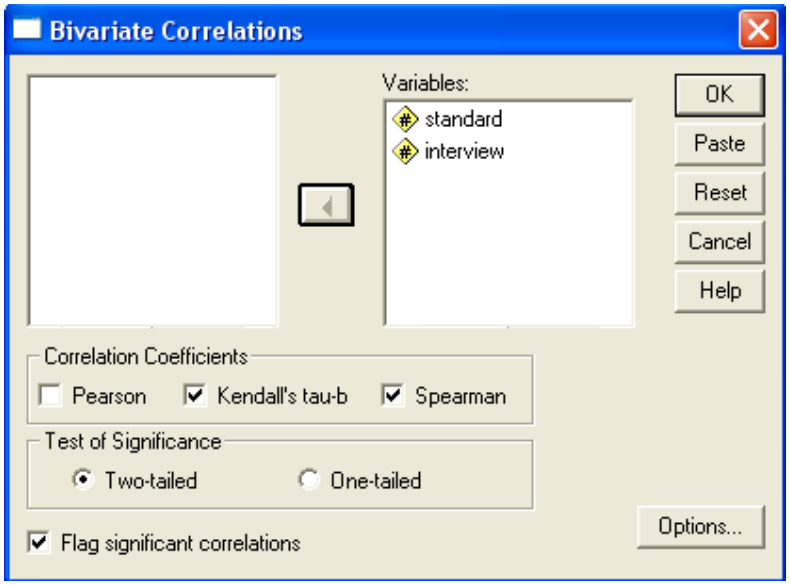
مشاهده می کنید که ضریب همبستگی IQ وIQ، همچنین Score و Score برابر یک است و ضریب همبستگی IQ و Score هم ۰.۷۹۵ است در ضمن آزمون معناداری ضریب همبستگی هم انجام شده است که صفر آزمون به صورت $H_0 : \rho = 0$ می باشد. اگر فرض $H_0$ رد می شود به این معنی است که ضریب همبستگی صفر نیست (اصطلاحاً معنادار است.)

**مثال:** مصاحبه کننده ای که مسئول استخدام تعداد زیادی ماشین نویس است می خواهد قدرت رابطهٔ بین رتبه های داده شده بر مبنای یک مصاحبه و نمرات یک آزمون استعداد را تعیین کند. داده های ۶ متقاضی عبارتند از :

| **رتبهٔ مصاحبه** | ۵ | ۲ | ۳ | ۱ | ۶ | ۴ |
|---|---|---|---|---|---|---|
| **نمرهٔ استاندارد** | ۴۷ | ۳۲ | ۲۹ | ۲۸ | ۵۶ | ۳۸ |

ضریب همبستگی اسپیرمن و کندال-تاو را بدست می آوریم و معنی داری رابطهٔ را نیز بررسی میکنیم:

ابتدا متغیرها را با نام standard وinterview در SPSS وارد کنید. در کادر مکالمه ای قبل متغیرها را وارد کرده و گزینه های kendall's tua-b و spearman را تیک دار کنید.




حال گزینهٔ Ok را کلیک کنید تا خروجی به صورت زیر نمایش داده شود:

**Correlations**

| | | | standard | interview |
|---|---|---|---|---|
| Kendall's tau_b | standard | Correlation Coefficient | 1.000 | .867* |
| | | Sig. (2-tailed) | . | .015 |
| | | N | 6 | 6 |
| | interview | Correlation Coefficient | .867* | 1.000 |
| | | Sig. (2-tailed) | .015 | . |
| | | N | 6 | 6 |
| Spearman's rho | standard | Correlation Coefficient | 1.000 | .943** |
| | | Sig. (2-tailed) | . | .005 |
| | | N | 6 | 6 |
| | interview | Correlation Coefficient | .943** | 1.000 |
| | | Sig. (2-tailed) | .005 | . |
| | | N | 6 | 6 |

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

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

خروجی را تفسیر کنید(؟)

**تمرین:** محققی مایل است بداند آیا بین سن افراد و قدرت بدنی افراد رابطه ای خطی وجود دارد یا نه؟ داده های جمع آوری شده توسط وی به صورت زیر است:

| سن | قدرت بدنی |
|---|---|
| ۱۰ | ۱۲ |
| ۱۵ | ۴۵ |
| ۲۰ | ۷۵ |
| ۲۵ | ۷۰ |
| ۳۰ | ۶۳ |
| ۳۵ | ۶۰ |
| ۴۰ | ۵۵ |
| ۴۵ | ۵۰ |
| ۵۰ | ۴۰ |
| ۵۵ | ۳۰ |
| ۶۰ | ۲۸ |
| ۶۵ | ۲۵ |
| ۷۰ | ۲۰ |
| ۷۵ | ۱۵ |
| ۸۰ | ۵ |

آیا بین سن افراد و قدرت بدنی افراد رابطه ای خطی وجود دارد؟

## ۲-ارتباط بین متغیرهای رسته ای:

در این بخش با سوالاتی بدین صورت روبرو هستیم:

– آیا رابطه ای بین جنسیت و سیگاری بودن افراد وجود دارد؟ یعنی به عنوان مثال مردان بیش از زنان گرایش به سیگار داشته باشند؟

– آیا گرایش سیاسی یک شخص (دمکرات – جمهوری خواه) با گرایش مذهبی او (مسیحی – یهودی – مسلمان – غیره)رابطه ای دارد؟

### - جداول توافقی:

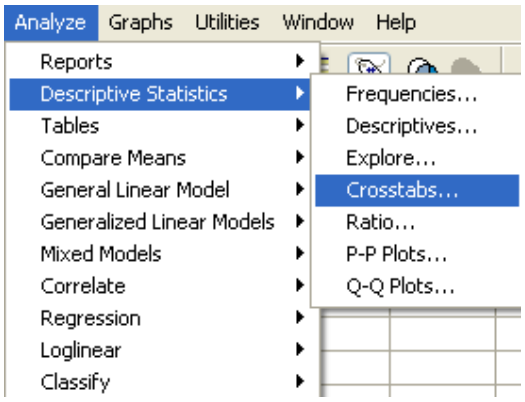
اولین قدم در آنالیز مسائلی به این صورت، رسم جدول توافقی است. جدول توافقی دو متغیر رسته ای، به تعداد رسته های متغیر اول سطر و به تعداد رسته های متغیر دوم، ستون دارد و در هر خانه این جدول تعداد فراوانیهای مربوط به آن حالت قرار می گیرد (به جای فراوانی، ممکن است درصد فراوانی نوشته شود)

داده های زیر نتایج یک بررسی بر روی ۳۰ نفر است که جنسیت و سیگاری بودن (یا نبودن) پرسیده شده است:

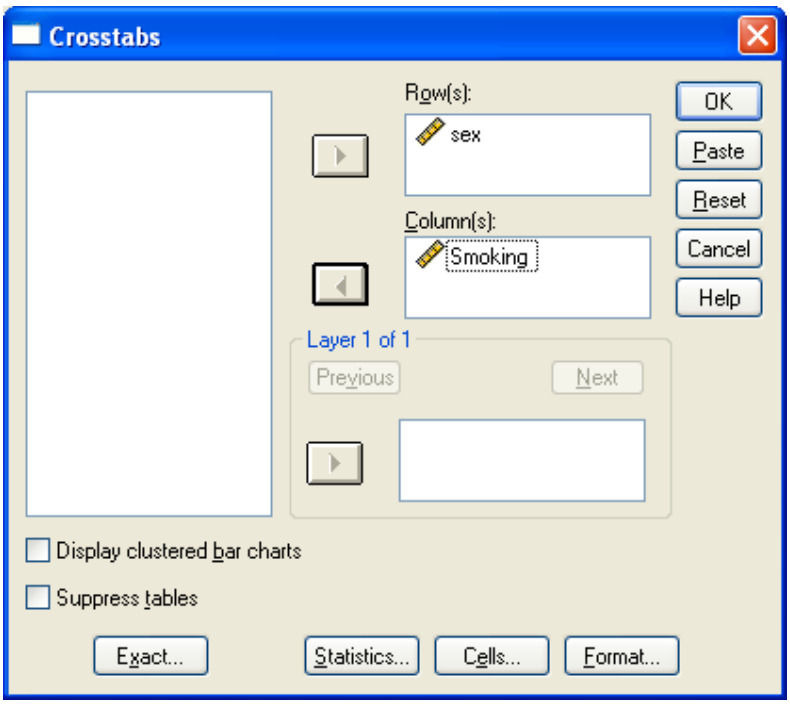
| | sex | Smoking |
|---|---|---|
| 1 | Female | No |
| 2 | Female | No |
| 3 | Female | No |
| 4 | Female | Yes |
| 5 | Female | No |
| 6 | Female | Yes |
| 7 | Female | No |
| 8 | Female | No |
| 9 | Female | No |
| 10 | Female | No |
| 11 | Female | No |
| 12 | Female | Yes |
| 13 | Female | No |
| 14 | Female | No |
| 15 | Male | Yes |

| | sex | Smoking |
|---|---|---|
| 16 | Male | No |
| 17 | Male | No |
| 18 | Male | Yes |
| 19 | Male | Yes |
| 20 | Male | Yes |
| 21 | Male | No |
| 22 | Male | No |
| 23 | Male | Yes |
| 24 | Male | Yes |
| 25 | Male | No |
| 26 | Male | Yes |
| 27 | Male | Yes |
| 28 | Male | No |
| 29 | Male | Yes |
| 30 | Male | Yes |

برای رسم جدول توافقی داده های فوق مسیر زیر را طی کنید:




تا کادر زیر باز شود :




مطابق تصویر، متغیر sex را به قسمت Row و متغیر Sigari را به قسمت Column(s) ببرید و دکمه Ok را کلیک کنید. خروجی به صورت زیر خواهد بود:

**Crosstabs**

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| SEX * SIGARI | 30 | 100.0% | 0 | .0% | 30 | 100.0% |

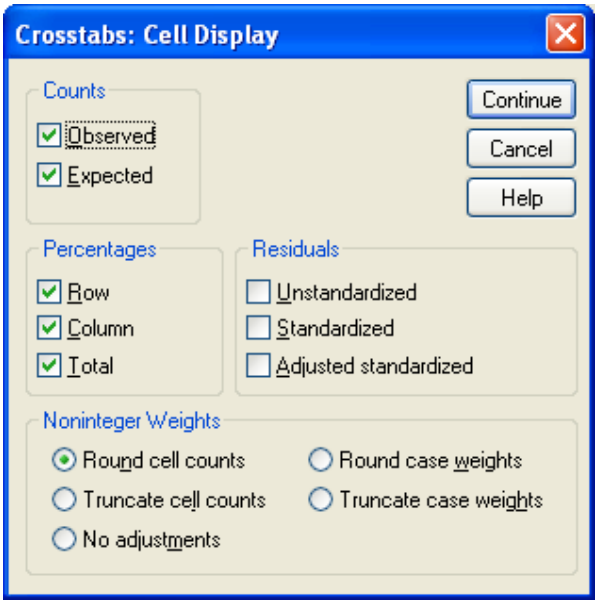
**SEX * SIGARI Crosstabulation**

Count

| | | SIGARI | | Total |
|---|---|---|---|---|
| | | بلي | خير | |
| SEX | زن | 3 | 11 | 14 |
| | مرد | 10 | 6 | 16 |
| Total | | 13 | 17 | 30 |

جدول اول تعداد داده های موجود، گمشده و کل داده ها را نشان می دهد و جدول دوم، جدول توافقی دو متغیر فوق است که تعداد مردان سیگاری، زنان سیگاری، مردان غیرسیگاری و زنان غیرسیگاری را نشان می دهد.

برای مشاهده درصدها و مقادیر موردانتظار در کادر اصلی Crosstabs روی دکمه Cells... کلیک کنید تا کادر زیر باز شود:




مطابق تصویر بالا، مقادیر موردانتظار، درصدهای سطری، ستونی و کل را انتخاب کنید.

Continue و سپس Ok را کلیک کنید.

جدول توافقی حاصل به صورت زیر است:

**sex * Smoking Crosstabulation**

| | | | Smoking | | |
|---|---|---|---|---|---|
| | | | Yes | No | Total |
| sex | Female | Count | 3 | 11 | 14 |
| | | Expected Count | 6.1 | 7.9 | 14.0 |
| | | % within sex | 21.4% | 78.6% | 100.0% |
| | | % within Smoking | 23.1% | 64.7% | 46.7% |
| | | % of Total | 10.0% | 36.7% | 46.7% |
| | Male | Count | 10 | 6 | 16 |
| | | Expected Count | 6.9 | 9.1 | 16.0 |
| | | % within sex | 62.5% | 37.5% | 100.0% |
| | | % within Smoking | 76.9% | 35.3% | 53.3% |
| | | % of Total | 33.3% | 20.0% | 53.3% |
| Total | | Count | 13 | 17 | 30 |
| | | Expected Count | 13.0 | 17.0 | 30.0 |
| | | % within sex | 43.3% | 56.7% | 100.0% |
| | | % within Smoking | 100.0% | 100.0% | 100.0% |
| | | % of Total | 43.3% | 56.7% | 100.0% |

## رسم جدول توافقی بدون داشتن داده های خام:

فرض کنید در مثال بالا به جای اطلاعات هر ۳۰ نفر تنها تعداد مردان سیگاری، زنان سیگاری، مردان غیر سیگاری و زنان غیر سیگاری را داشته باشیم. داده ها را به ترتیب زیر در SPSS وارد می کنیم:

| sex | smoking | frequency |
|---|---|---|
| Male | Yes | 10.00 |
| Male | No | 6.00 |
| Female | Yes | 3.00 |
| Female | No | 11.00 |

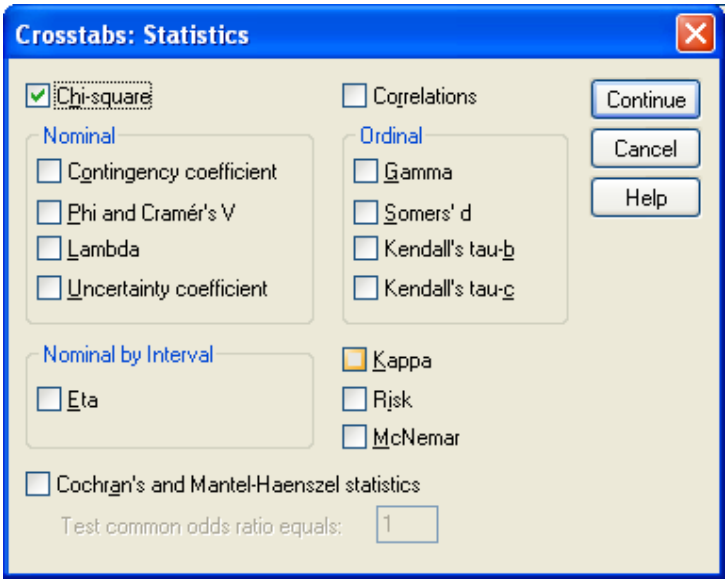
سپس به داده ها بر حسب متغیر freq وزن می دهیم و مراحل رسم جدول توافقی را طی می کنیم .

## آزمون استقلال متغیرها:

پس از تحلیل توصیفی داده ها و خلاصه کردن در قالب جدول توافقی، علاقه مندیم بدانیم آیا بین متغیرها رابطه ای وجود دارد یا نه؟ دو متغیر در صورتی که مستقل باشند، هیچ رابطه ای با هم ندارند. مثلاً اگر دو متغیر مثال قبلی مستقل باشند. سیگاری بودن شخص رابطه ای با جنسیت او ندارد یا به عبارتی جنسیت تاثیری در گرایش مردم به سیگار ندارد. ولی در صورتیکه دو متغیر مستقل نباشند، با هم رابطه دارند. برای آزمون استقلال دو متغیر از *آزمون خی دو* استفاده خواهیم کرد.

منتها استفاده از آزمون خی دو شرایطی دارد مه داده ها بایستی انرا دارا باشند. ایـن شـرایط را بایـد در جدول توافقی داده عها جستجو کرد. با توجه به جدول توافقی حاصل، آیا می توانیم از آزمـون خـی دو استفاده کنیم؟؟؟ . . .

برای انجام آزمون خی دو در کادر Crosstabs روی دکمه ...Statistics کلیک کنیـد مطـابق شـکل گزینه chi-square را انتخاب کنید Continue و سپس ok را کلیک کنید:




نتیجه آزمون استقلال برای داده های مثال قبل به صورت زیر خواهد بود:

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 5.129[b] | 1 | .024 | | |
| Continuity Correction[a] | 3.593 | 1 | .058 | | |
| Likelihood Ratio | 5.336 | 1 | .021 | | |
| Fisher's Exact Test | | | | .033 | .028 |
| Linear-by-Linear Association | 4.958 | 1 | .026 | | |
| N of Valid Cases | 30 | | | | |

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.07.

نتایج آزمون خی دو پیرسون در سـطر اول آمـده اسـت. فـرض صـفر در ایـن آزمـون اسـتقلال متغیرهاست و در صورتیکه فرض صفر رد شود، نشان دهنده وجود رابطه بین دو متغیر خواهد بود.

**تمرین ۱ :**

محققی مایل است بداند آیا بین جنسیت افراد و چپ دست بودن-راست دست بودن افراد رابطه ای وجود دارد یا نه؟ بدین منظور داده های زیر را جمع آوری کرده است:

| | دختر | پسر |
|---|---|---|
| چپ دست | ۲ | ۱۰ |
| راست دست | ۱۱ | ۱ |

به نظر شما این رابطه معنی دار است یا خیر؟

**تمرین ۲ :**

میخواهیم بدانیم آیا رابطهٔ بین گروه خونی افراد و فشار خون آنها رابطه ای وجود دارد یا نه؟ بدین منظور داده های زیر را جمع آوری کرده ایم:

| | A | B | O | AB |
|---|---|---|---|---|
| + | ۲۵ | ۱۲ | ۱۷ | ۸ |
| − | ۸ | ۱۷ | ۱۲ | ۲۵ |

به نظر شما این رابطه معنی دار است یا خیر؟

## ۳- مقایسه چند جمعیت (آنالیز واریانس)

با کاربرد آزمون t برای مقایسه میانگین دوجامعه آشنا شدیم، آنالیز واریانس به ما کمک می کند تا میانگین های چند جامعه را با هم مقایسه کنیم. متغیری که قرار است میانگین آن در چند زیر جامعه مقایسه شود متغیر وابسته و متغیر یا متغیرهایی که باعث تقسیم جامعه به چند زیر جامعه می شوند، متغیر(های) مستقل نامیده می شوند. متغیر مستقل معمولاً از نوع اسمی است. جواب سئوالات زیر با آنالیز واریانس داده می شود:

- آیا درآمد نهایی فیلم های امسال به نوع فیلم بستگی دارد یا نه؟
- آیا شغل والدین در معدل دانش آموزان یک مدرسه تاثیر دارد یا نه؟
- آیا چهار روش مختلف آموزش ریاضی در دانش آموزان چهار کلاس یک مدرسه، تاثیر یکسانی دارد یا خیر؟

در سئوالات فوق می بینید که متغیرهای مستقل باعث تقسیم جامعه به چند زیر گروه یا زیر جامعه می شوند و هدف آزمون این مساله است که آیا میانگین متغیر وابسته در زیر گروهها یکسان است یا خیر. به عبارتی فرض صفر در آنالیز واریانس، یکسان بودن میانگین تمام زیر گروههاست.

در آنالیز واریانس فرض بر این است که داده ها از توزیع نرمال آمده اند و واریانس داده ها در زیر گروهها یکسان است. در واقع درستی فرض های فوق، برای معنا دار بودن آنالیز واریانس لازم است. لذا همواره قبل از انجام آنالیز واریانس فرض های فوق را آزمون خواهیم کرد مخصوصاً فرض تساوی واریانس ها.

اگر شکی در نرمال بودن داده ها داشته باشیم یا تعداد داده ها کم باشد می توان به جای آنالیز واریانس از معادل ناپارامتری آن استفاده کرد.

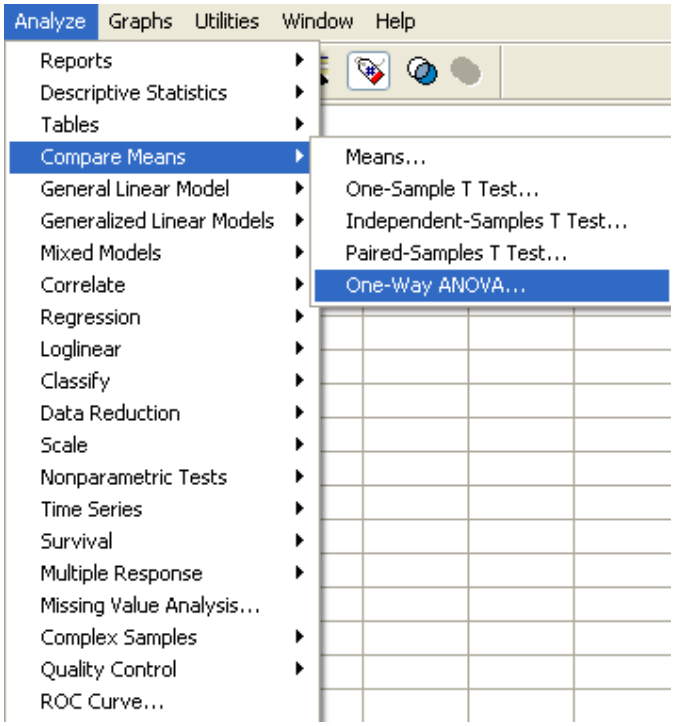
مثال: داده های زیر مقدار نیکوتین ۲۰ سیگار است که به طور تصادفی از میان ۳ نوع سیگار مختلف انتخاب شده اند، آیا هر سه نوع سیگار، مقدار نیکوتین یکسان دارند؟

| | nikotin | sigar |
|---|---|---|
| 1 | 74.0 | 1 |
| 2 | 73.5 | 1 |
| 3 | 74.5 | 1 |
| 4 | 74.0 | 1 |
| 5 | 73.5 | 1 |
| 6 | 74.0 | 1 |
| 7 | 64.5 | 2 |
| 8 | 65.0 | 2 |
| 9 | 64.0 | 2 |
| 10 | 65.0 | 2 |
| 11 | 64.5 | 2 |
| 12 | 64.5 | 2 |
| 13 | 64.0 | 2 |
| 14 | 65.0 | 2 |
| 15 | 65.5 | 3 |
| 16 | 65.0 | 3 |
| 17 | 64.5 | 3 |
| 18 | 66.0 | 3 |
| 19 | 64.5 | 3 |
| 20 | 65.5 | 3 |

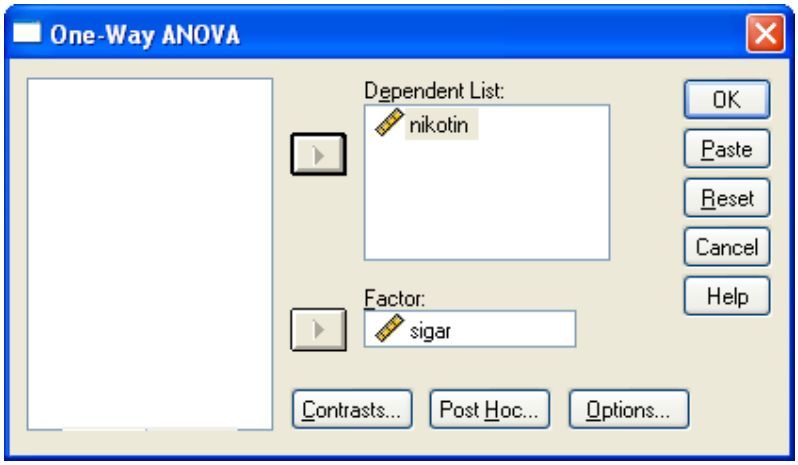
✓ **به نحوه وارد کردن داده ها توجه کنید.**

در این مسأله مقدار نیکوتین ( Nikotin ) متغیر وابسته و نوع سیگار ( cigar ) متغیر مستقل است.

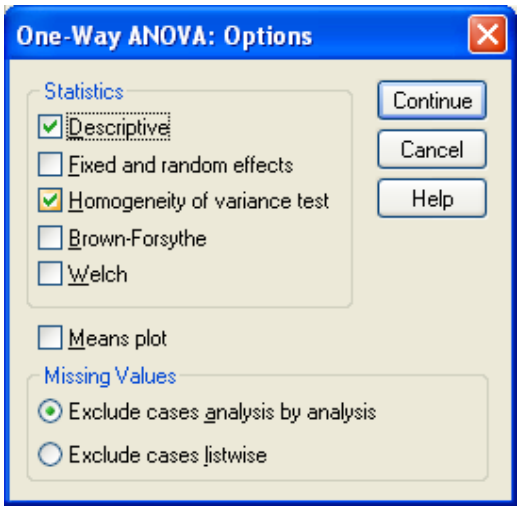
برای انجام آنالیز واریانس مسیر زیر را طی کنید:




تا کادر زیر باز شود:




متغیرها را مطابق شکل در قسمتهای مربوطه وارد کنید. سپس روی دکمه ...Options کلیک کنید تـا کادر زیر باز شود:




گزینه های Homogeneity of variance test و Descriptive را انتخاب کنید سپس دکمه Continue و سپس OK را کلیک کنید. خروجی به صورت زیر خواهد بود:

**Descriptives**

NIKOTIN

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| 1 | 6 | 73.917 | .3764 | .1537 | 73.522 | 74.312 | 73.5 | 74.5 |
| 2 | 8 | 64.563 | .4173 | .1475 | 64.214 | 64.911 | 64.0 | 65.0 |
| 3 | 6 | 65.167 | .6055 | .2472 | 64.531 | 65.802 | 64.5 | 66.0 |
| Total | 20 | 67.550 | 4.3070 | .9631 | 65.534 | 69.566 | 64.0 | 74.5 |

**Test of Homogeneity of Variances**

NIKOTIN

| Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|
| 1.506 | 2 | 17 | .250 |

**ANOVA**

NIKOTIN

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 348.690 | 2 | 174.345 | 788.174 | .000 |
| Within Groups | 3.760 | 17 | .221 | | |
| Total | 352.450 | 19 | | | |

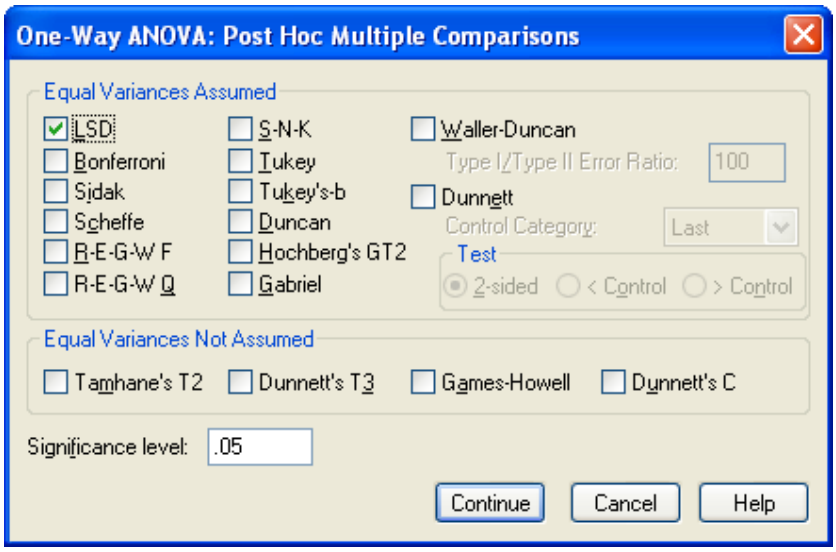
در کادر Descriptive آماره های توصیفی برای زیر گروهها به صورت جداگانه محاسبه شده است.

در کادر دوم آزمون تساوی واریانس زیر گروهها انجام شده است که همگنی واریانسها را نشان می دهد.

نتیجه نهایی آنالیز واریانس در کادر ANOVA آمده است. با توجه به سطح معناداری آزمون چه نتیجه ای می گیریم؟

نتیجه آنالیز واریانس مشخص می کند که آیا میانگین زیر گروهها یکسان است یا خیر؟ اگر نتیجه آنالیز واریانس متفاوت بودن میانگین ها را نتیجه دهد، علاقه مندیم بدانیم این تفاوت کلی ناشی از کدام تفاوتهای جزئی است. به عبارتی دیگر کدام تفاوت های دوگانه باعث ایجاد تفاوت کلی شده است. به عنوان مثال در مساله بالا نتیجه آنالیز واریانس تفاوت در میانگین نیکوتین ۳ نوع سیگار مورد آزمایش را نشان می‌دهد، حال علاقه مندیم با تحلیل بیشتر داده ها، بدانیم این تفاوت کلی از کجا نشأت گرفته است، لذا میانگین ها را دو بدو مقایسه می‌کنیم.

در کادر آنالیز واریانس روی دکمه Post Hoc کلیک کنید تا کادر زیر باز شود:




گزینه LSD را انتخاب کنید Continue و سپس Ok را کلیک کنید.

ملاحظه می کنید در خروجی جدولی به صورت زیر اضافه شده است:

**Post Hoc Tests**

**Multiple Comparisons**

Dependent Variable: NIKOTIN
LSD

| (I) SIGAR | (J) SIGAR | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 9.354* | .2540 | .000 | 8.818 | 9.890 |
| | 3 | 8.750* | .2715 | .000 | 8.177 | 9.323 |
| 2 | 1 | -9.354* | .2540 | .000 | -9.890 | -8.818 |
| | 3 | -.604* | .2540 | .029 | -1.140 | -.068 |
| 3 | 1 | -8.750* | .2715 | .000 | -9.323 | -8.177 |
| | 2 | .604* | .2540 | .029 | .068 | 1.140 |

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

در این جدول آزمون تساوی میانگین زیر گروهها دوبه دو انجام شده است.

# آنالیز واریانس دوطرفه :

در SPSS برای طرح های پیچیده تر از آنالیز واریانس یک طرفه از یک روند آماری به نام مدل عمومی خطی( GLM ) استفاده می کنیم.

در این بخش به آنالیز واریانس دو طرفه (یک نوع مدل خطی ) می پردازیم :

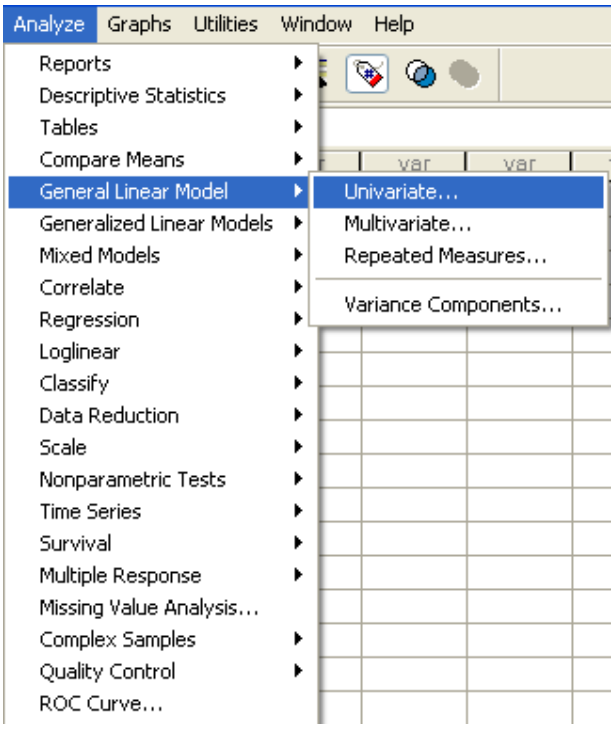
فرض کنید بجای بررسی تاثیر یک فاکتور بر متغیر پاسخ بخواهیم اثر دو فاکتور را بر متغیر پاسخ بررسی کنیم, در این صورت با یک طرح دو عاملی روبرو خواهیم بود .

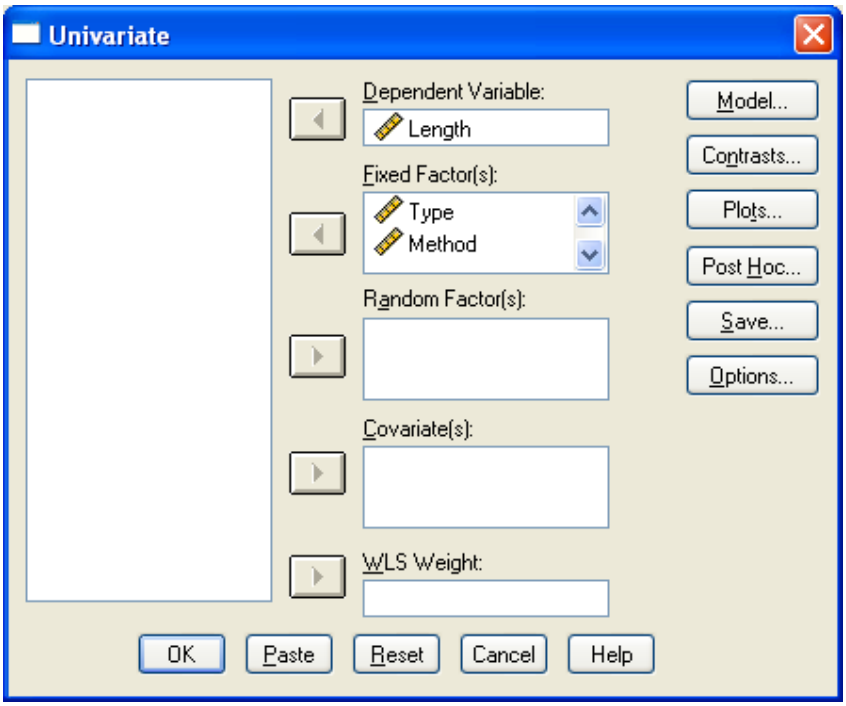
**مثال :**

می خواهیم بدانیم که آیا روش درمان یک بیماری خاص و نوع داروی مصرفی جهت درمان آن بیماری در طول مدت درمان آن بیماری تاثیری دارد یا خیر؟ برای ارزیابی این فرضیه ۶۰نفر از افراد مبتلا به این بیماری خاص را مورد آزمایش قرار داده ایم. داده ها به صورت زیرند :

ابتدا باید داده هارا در SPSS وارد کنیم. به نحوه داده وارد کردن توجه کنید.(با نامهای method ،type و length و سپس مسیر زیر را طی کنید ،تا کادر مکالمه ای مربوط به آن باز شود و متغیرها را به صورت نشان داده شده وارد کنید:

| | | نوع دارو | | | |
|---|---|---|---|---|---|
| | | ۱ | ۲ | ۳ | ۴ |
| روش درمان | ۱ | ۶ | ۷ | ۸ | ۸ |
| | | ۵ | ۸ | ۷ | ۹ |
| | | ۴ | ۲ | ۱۰ | ۱۰ |
| | | ۷ | ۳ | ۵ | ۱ |
| | | ۲ | ۵ | ۶ | ۶ |
| | ۲ | ۳ | ۴ | ۲ | ۳ |
| | | ۲ | ۶ | ۵ | ۴ |
| | | ۱ | ۵ | ۶ | ۵ |
| | | ۴ | ۳ | ۳ | ۷ |
| | | ۵ | ۲ | ۴ | ۱ |
| | ۳ | ۴ | ۶ | ۵ | ۸ |
| | | ۵ | ۸ | ۹ | ۶ |
| | | ۲ | ۵ | ۲ | ۳ |
| | | ۷ | ۴ | ۳ | ۹ |
| | | ۱ | ۹ | ۱ | ۱ |






توجه داشته باشید که هر دو فاکتور method و type تثبیت شده اند . اگر چنانچه هر کـدام از آنها تصادفی بودند آنگاه آن را به قسمت Random Factor(s) وارد می کردیم.

پس از زدن کلید Ok خروجی را به صورت زیر مشاهده می کنید :

**Between-Subjects Factors**

| | | N |
|---|---|---|
| Method | 1.00 | 20 |
| | 2.00 | 20 |
| | 3.00 | 20 |
| Type | 1.00 | 15 |
| | 2.00 | 15 |
| | 3.00 | 15 |
| | 4.00 | 15 |

**Tests of Between-Subjects Effects**

Dependent Variable: lenght

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 97.333[a] | 11 | 8.848 | 1.508 | .160 |
| Intercept | 1421.067 | 1 | 1421.067 | 242.227 | .000 |
| Method | 48.433 | 2 | 24.217 | 4.128 | .022 |
| Type | 20.933 | 3 | 6.978 | 1.189 | .324 |
| Method * Type | 27.967 | 6 | 4.661 | .795 | .579 |
| Error | 281.600 | 48 | 5.867 | | |
| Total | 1800.000 | 60 | | | |
| Corrected Total | 378.933 | 59 | | | |

a. R Squared = .257 (Adjusted R Squared = .087)

آیا هیچ نشانی از اینکه یکی از این عوامل در طول مدت درمان موثر باشند وجود دارد؟
آیا دو عامل اثر متقابل دارند؟

## تمرین ۱)

مهندسین برق برای اندازه گیری روشنایی لامپ تصویر تلویزیون, جریان لازم برای بدست آوردن سطح روشنایی مشخص را ملاک قرار میدهند. فرض کنید میخوا هیم اثر نوع شیشه ونوع فسفر, در روشنایی لامپ تصویر را بررسی کنیم. در این مسئله متغیر پاسخ, مقدار جریان لازم است. داده های بدست آمده بصورت زیراست :

| نوع شیشه | | نوع فسفر | | |
|---|---|---|---|---|
| | | ۱ | ۲ | ۳ |
| | ۱ | ۲۸۰ | ۳۰۰ | ۲۹۰ |
| | | ۲۹۰ | ۳۱۰ | ۲۸۵ |
| | | ۲۸۵ | ۲۹۵ | ۲۹۰ |
| | ۲ | ۲۳۰ | ۲۶۰ | ۲۲۰ |
| | | ۲۳۵ | ۲۴۰ | ۲۲۵ |
| | | ۲۴۰ | ۲۳۵ | ۲۳۰ |

آیا هیچ نشانی از اینکه یکی از این عوامل در روشنایی موثر باشد, وجود دارد؟
آیا دو عامل اثر متقابل دارند؟

## تمرین۲)

از چهار شیمیدان خواسته شده است که درصد الکل متیلیک یک ترکیب خاص شیمیایی را تعیین کنند. هر یک از این شیمیدانها سه اندازه گیری انجام داده و نتایج زیر را گزارش کرده اند :

| شیمیدان | | درصد الکل متیلیک | | |
|---|---|---|---|---|
| | ۱ | ۸۴/۹۹ | ۸۴/۰۴ | ۸۴/۳۸ |
| | ۲ | ۸۵/۱۵ | ۸۵/۱۳ | ۸۴/۸۸ |
| | ۳ | ۸۴/۷۲ | ۸۴/۴۸ | ۸۵/۱۶ |
| | ۴ | ۸۴/۲۰ | ۸۴/۱۰ | ۸۴/۵۵ |

آیا گزارش شیمیدانها تفاوت معنی دار دارد؟ ( در سطح $\alpha = 0.05$ آزمون را انجام دهید. )

## تمرین ۳)

می خواهیم بدانیم که آیا روش درمان یک بیماری خاص در طول مدت درمان آن بیماری تـاثیری دارد یا خیر؟ برای ارزیابی این فرضیه ۲۵ نفر از افراد مبتلا به بیماری خاص به ۴ گروه تقسیم و هر کـدام از گروها با یک روش مورد درمان قرار گرفتند . داده های حاصل به صورت زیرند:

| | روشهای درمان | | | |
|---|---|---|---|---|
| | ۱ | ۲ | ۳ | ۴ |
| طول مدت درمان | ۸ | ۷ | ۹ | ۵ |
| | ۷ | ۵ | ۷ | ۷ |
| | ۶ | ۸ | ۵ | ۸ |
| | ۸ | ۴ | ۸ | ۷ |
| | ۷ | ۵ | ۸ | ۵ |
| | ۹ | ۶ | ۷ | ۸ |
| | | | | ۷ |

آیا داده ها دلیلی بر تفاوت طول مدت درمان در روشهای درمان متفاوت است یا خیر؟

## تمرین ۴)

آزمایشی برای تعیین اثر دمای کوره یا اثرهای موقعیت کوره بر چگالی یک آنودکربنی اجرا شده است .داده ها را در زیر نشان داه شده است.

| | | دما | | |
|---|---|---|---|---|
| | | ۸۰۰ | ۸۲۵ | ۸۵۰ |
| موقعیت | ۱ | ۵۷۰ | ۱۰۶۳ | ۵۶۵ |
| | | ۵۶۵ | ۱۰۸۰ | ۵۱۰ |
| | | ۵۸۳ | ۱۰۴۳ | ۵۹۰ |
| | ۲ | ۵۲۸ | ۹۸۸ | ۵۲۶ |
| | | ۵۴۷ | ۱۰۲۶ | ۵۳۸ |
| | | ۵۲۱ | ۱۰۰۴ | ۵۳۲ |

داده ها را تحلیل کنید ؟

# رگرسیون:

در بحث ضریب همبستگی چگونگی بررسی رابطه میان متغیرهای عددی و محاسبه ضریب همبستگی را بیان کردیم. ولی اغلب در مسائلی که با متغیرهای عددی روبرو هستیم، علاقه مندیم علاوه بر بررسی ساده رابطه دو متغیر مقادیر یک متغیر (متغیر وابسته) را به صورت تابعی از یک یا چند متغیر مستقل دیگر بنویسیم. روشهای تحلیل رگرسیون به ما کمک می کنند تا چگونگی این رابطه را تعیین کنیم.

**-رگرسیون خطی ساده:**

در ساده ترین حالت فرض می کنیم متغیر وابسته تنها با یک متغیر مستقل در ارتباط باشد و رابطه این دو به صورت زیر باشد:

$$y = \alpha + \beta x + \varepsilon$$

یعنی y ها به صورت خطی وابسته به مقادیر x اند. اگر صحت چنین رابطه ای تأیید شود، توانسته ایم داده ها را به صورت بسیار خوبی خلاصه کنیم. علاوه بر این برای مقادیر مختلف x می توانیم برآوردی از مقادیر y مورد انتظار ارائه دهیم.

داده های زیر مخارج تبلیغات فروش هفتگی ۱۲ فروشگاه مختلف در یک شهر است.

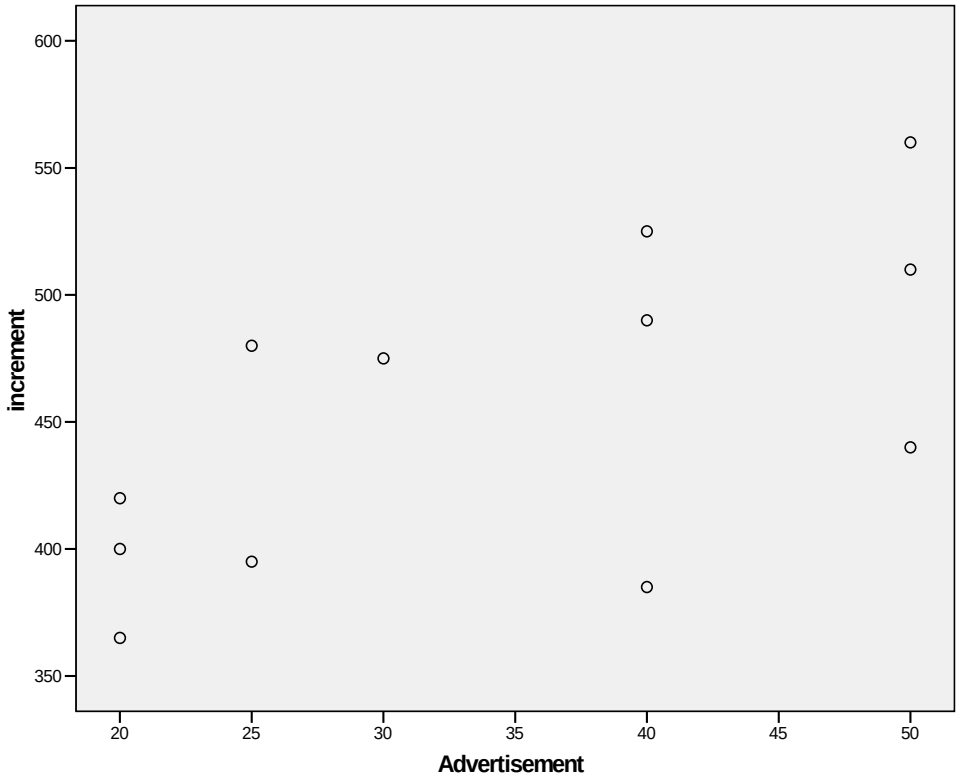
| مخارج تبلیغ | ۴۰ | ۲۰ | ۲۵ | ۲۰ | ۳۰ | ۵۰ | ۴۰ | ۲۰ | ۵۰ | ۴۰ | ۲۵ | ۵۰ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| فروش | ۳۸۵ | ۴۰۰ | ۳۹۵ | ۳۶۵ | ۴۷۵ | ۴۴۰ | ۴۹۰ | ۴۲۰ | ۵۶۰ | ۵۲۵ | ۴۸۰ | ۵۱۰ |

رابطه خط رگرسیونی بین دو متغیر را در صورت وجود بیابید. (رابطه خطی)

نخست داده ها را در دو ستون در محیط SPSS وارد می کنیم:

| Advertisement | increment |
|---|---|
| 40 | 385 |
| 20 | 400 |
| 25 | 395 |
| 20 | 365 |
| 30 | 475 |
| 50 | 440 |
| 40 | 490 |
| 20 | 420 |
| 50 | 560 |
| 40 | 525 |
| 25 | 480 |
| 50 | 510 |

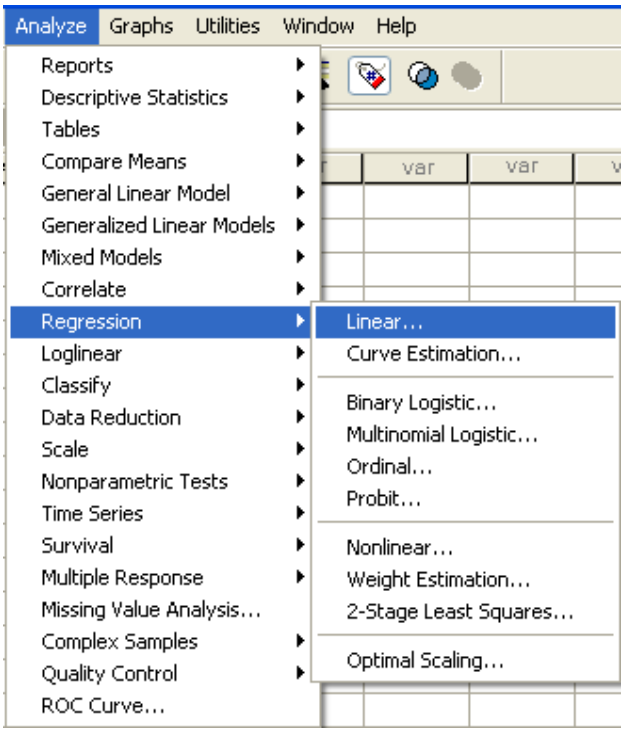
قدم اول در تحلیل رگرسیونی ساده همانند بخش ضریب همبستگی، رسم نمودار پراکنش است. نمودار پراکنش می تواند وجود یا عدم وجود رابطه خطی بین متغیرها را به خوبی نشان دهد. نمودار پراکنش را برای دو متغیر رسم کنید. نمودار پراکنش به صورت زیر خواهد شد:



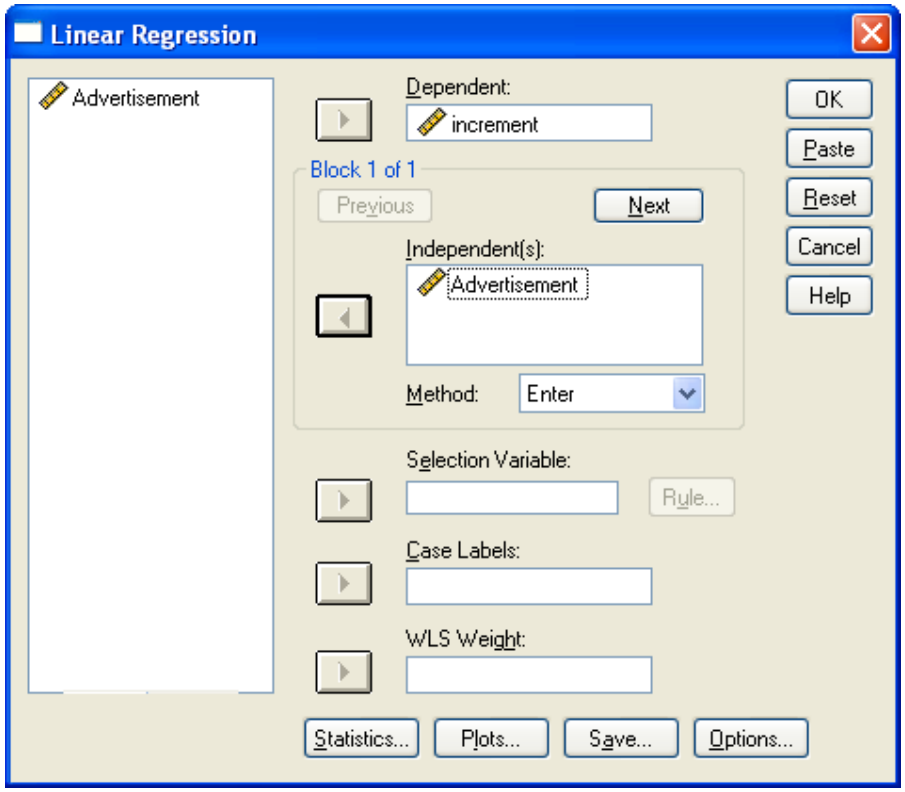

نمودار پراکنش نشاندهنده رابطه نسبی دو متغیر است. یعنی به نظر می رسد y با x رابطه خطی دارد.

مرحلهٔ بعد پیدا کردن برآورد ضریبهای معادله خط رگرسیون $(\beta,\alpha)$ است.

برای انجام تحلیل رگرسیونی به صورت کامل مسیر زیر را طی کنید:




تا کادر زیر باز شود :



متغیـــر increment را در قســـمت Dependent و متغیـــر Advertisement را در قســـمت Independents وارد کنید و دکمه Ok را فشار دهید، خروجی به صورت زیر خواهد بود:

**Variables Entered/Removed[b]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | TABLIGH[a] | . | Enter |

a. All requested variables entered.

b. Dependent Variable: FOROOSH

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .635[a] | .403 | .343 | 50.226 |

a. Predictors: (Constant), TABLIGH

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 17030.04 | 1 | 17030.044 | 6.751 | .027[a] |
| | Residual | 25226.21 | 10 | 2522.621 | | |
| | Total | 42256.25 | 11 | | | |

a. Predictors: (Constant), TABLIGH

b. Dependent Variable: FOROOSH

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 343.706 | 44.766 | | 7.678 | .000 |
| | TABLIGH | 3.221 | 1.240 | .635 | 2.598 | .027 |

a. Dependent Variable: FOROOSH

جدول دوم، حاوی اطلاعاتی در مورد کارایی مدل است. هر چه مقدار (Adjust R Square) یا R Square به یک نزدیکتر باشد مدل کاراتر است. در این مثال مقدار R Square چندان زیاد نیست.

جدول سوم، جدول اصلی تحلیل رگرسیونی است. آزمون صورت گرفته به آزمون معناداری خط رگرسیون مشهور است. اگر نتیجه آزمون رد فرض $H_0$ باشد (.sig کمتر از 0.05 باشد) خط رگرسیون معنادار است.( که در این مورد رگرسیون با معنی است. )

جدول چهارم، جدول ضرایب خط رگرسیونی است و آزمون های صورت گرفته هم، آزمون برابری این ضرایب با صفر می باشد که در صورت رد شدن آزمون، ضرایب مخالف صفر خواهند بود (یا به عبارتی این ضرایب در خط رگرسیون وجود دارند و بی معنی نیستند.)

- **رگرسیون چند متغیره خطی:**

اگرتعداد متغیرهای مستقل بیش از یکی باشد. مثلاً دو متغیر باشد رابطه خط رگرسیون به صورت زیر خواهد بود:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

و هدف پیدا کردن برآورد ضرایب فوق است.

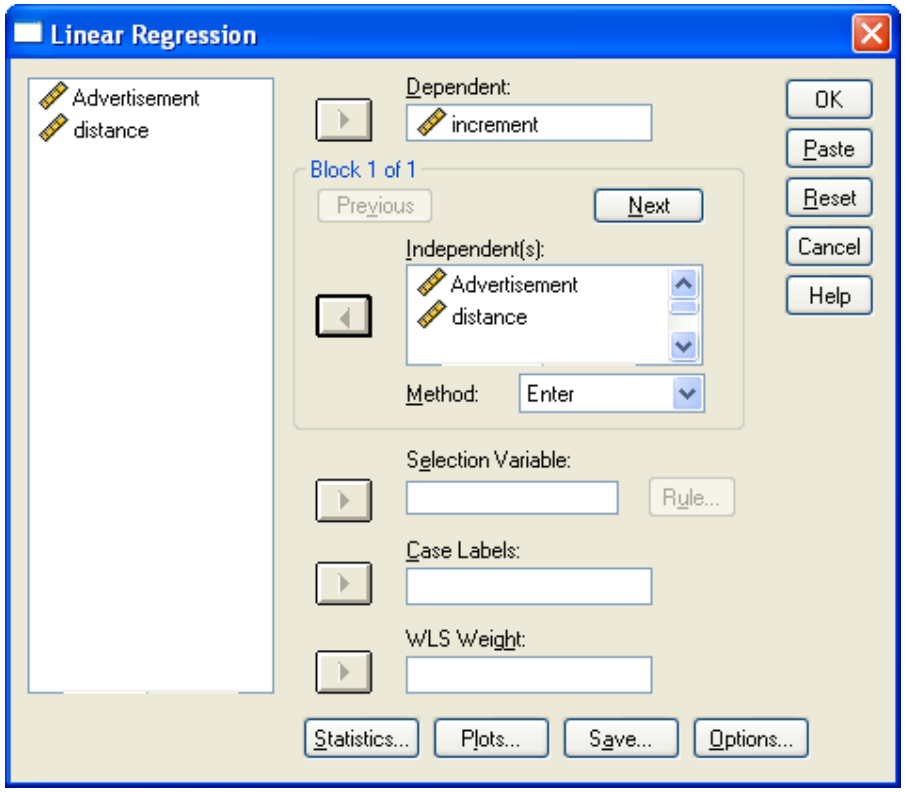
فرض کنید در مساله قبل طبق تجارب گذشته حدس می زنیم که میزان فروش هفتگی علاوه بر هزینه تبلیغ به فاصله فروشگاه تا میدان مرکزی شهر هم بستگی دارد. داده های زیر فاصله ۱۲ فروشگاه مربوطه از مرکز شهر بر حسب کیلومتر است:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ۶/۳ | ۰/۲ | ۹/۴ | ۵/۳ | ۳/۴ | ۲/۶ | ۷/۱ | ۸/۳ | ۲/۴ | ۰/۷ | ۳/۱ | ۴/۲ |

داده ها را در ستون جدیدی وارد کنید:

| Advertisement | increment | distance |
|---|---|---|
| 40 | 385 | 4.20 |
| 20 | 400 | 3.10 |
| 25 | 395 | .70 |
| 20 | 365 | 2.40 |
| 30 | 475 | 8.30 |
| 50 | 440 | 7.10 |
| 40 | 490 | 2.60 |
| 20 | 420 | 3.40 |
| 50 | 560 | 5.30 |
| 40 | 525 | 9.40 |
| 25 | 480 | .20 |
| 50 | 510 | 6.30 |

مسیری که باید طی کنیم همان مسیر قبلی است اما این بار در کادر تحلیل رگرسیونی متغیر مستقل دوم را هم اضافه می کنیم:



خروجی در این حالت به صورت زیر خواهد بود:

**Variables Entered/Removed[b]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | FASELE, TABLIGH[a] | . | Enter |

a. All requested variables entered.

b. Dependent Variable: FOROOSH

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .650[a] | .422 | .294 | 52.079 |

a. Predictors: (Constant), FASELE, TABLIGH

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 17846.45 | 2 | 8923.223 | 3.290 | .085[a] |
| | Residual | 24409.80 | 9 | 2712.201 | | |
| | Total | 42256.25 | 11 | | | |

a. Predictors: (Constant), FASELE, TABLIGH

b. Dependent Variable: FOROOSH

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 344.412 | 46.436 | | 7.417 | .000 |
| | TABLIGH | 2.736 | 1.560 | .539 | 1.754 | .113 |
| | FASELE | 3.592 | 6.546 | .169 | .549 | .597 |

a. Dependent Variable: FOROOSH