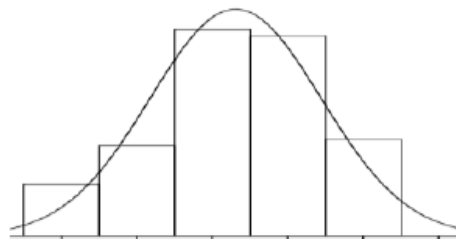


## بررسی نرمال بودن داده ها در SPSS

در اغلب آزمون های پارامتری، مفروضات مقدماتی بسیاری وجود دارد که تا این مفروضات تامین نشوند؛ نتایج به دست آمده از آزمون، نامعتبر خواهد بود. در میان این مفروضات مهمترین و شایعترین فرض، فرض نرمال بودن داده هاست. منظور از نرمال بودن توزیع داده ها این است که هیستوگرام فراوانی داده ها تقریبا به صورت منحنی نرمال باشد.



برای بررسی اینکه آیا متغیر یا متغیرهایی که در اختیار دارید، دارای توزیع نرمال هستند یا نه؟ باید چند گام اساسی بردارید.

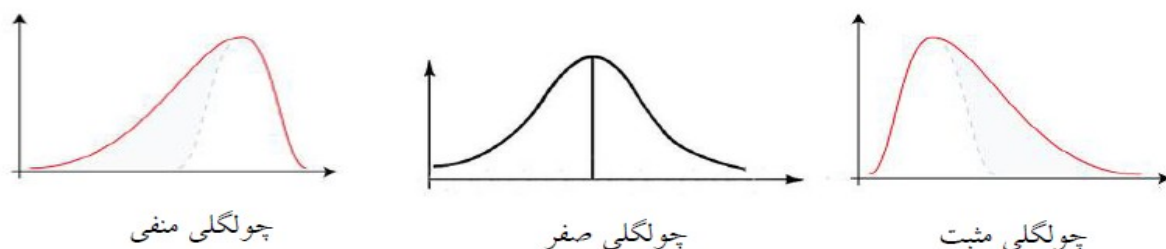
### ❖ تحلیل توصیفی (بررسی شاخص های توزیع)

ضریب چولگی و ضریب کشیدگی، دو شاخص اساسی توزیع داده ها هستند. که با داشتن این شاخص ها تا حدودی می توان به نرمال بودن یا نبودن توزیع داده ها پی برد.

۱- چولگی یا **skewness** معیاری از تقارن یا عدم تقارن تابع توزیع می باشد. در آمار فرمول های متعددی برای محاسبه ضریب چولگی داده ها وجود دارد که یکی از آنها به صورت زیر است.

$$\text{Skew} = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x_j - \bar{x}}{s} \right)^3$$

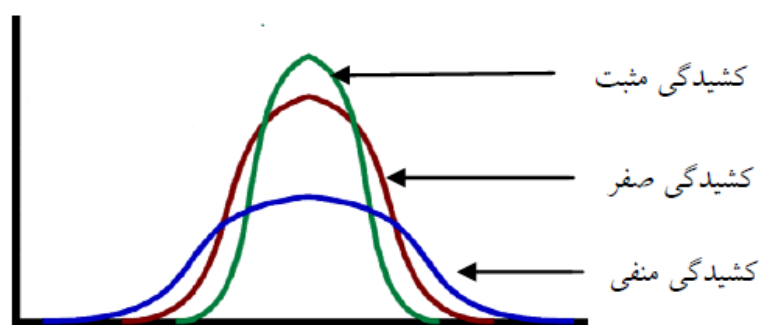
در یک توزیع کاملاً متقارن، چولگی صفر است. برای یک توزیع نامتقارن، ضریب چولگی مثبت یا منفی است (هرچه مقدار چولگی از صفر بیشتر فاصله داشته باشد، عدم تقارن، شدیدتر است)



۲- کشیدگی یا **kurtosis** نشان دهنده ارتفاع یک توزیع است. به عبارت دیگر کشیدگی معیاری از بلندی منحنی در نقطه ماکزیمم است. همیشه کشیدگی را با کشیدگی توزیع نرمال مقایسه می کنند. مقدار کشیدگی برای توزیع نرمال برابر ۳ می باشد. فرمول های متعددی برای محاسبه ضریب کشیدگی وجود دارد که یکی از آنها به صورت زیر است.

$$\text{Kurtosis} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_j - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

بنابراین، کشیدگی مثبت، یعنی قله توزیع مورد نظر از توزیع نرمال بالاتر است و کشیدگی منفی، نشانه پایین تر بودن قله توزیع نسبت به توزیع نرمال است.



در حالت کلی چنانچه مقدار چولگی و کشیدگی داده ها خارج از فاصله (۲، -۲) باشند (البته ممکن است بعضی از آماردانان این بازه را کوچکتر یا بزرگتر در نظر بگیرند)، داده ها از توزیع نرمال برخوردار نیستند و می بایست قبل از

هر گونه آزمونی که مشروط به فرض نرمال بودن داده ها ست، آن ها را به توزیع نرمال نزدیک کرد. روش هایی وجود دارد که با تغییراتی روی داده ها، توزیع آن ها نرمال می شود. این روش ها را در ادامه همین بحث بررسی می کنیم.

## ❖ محاسبه چولگی و کشیدگی در SPSS

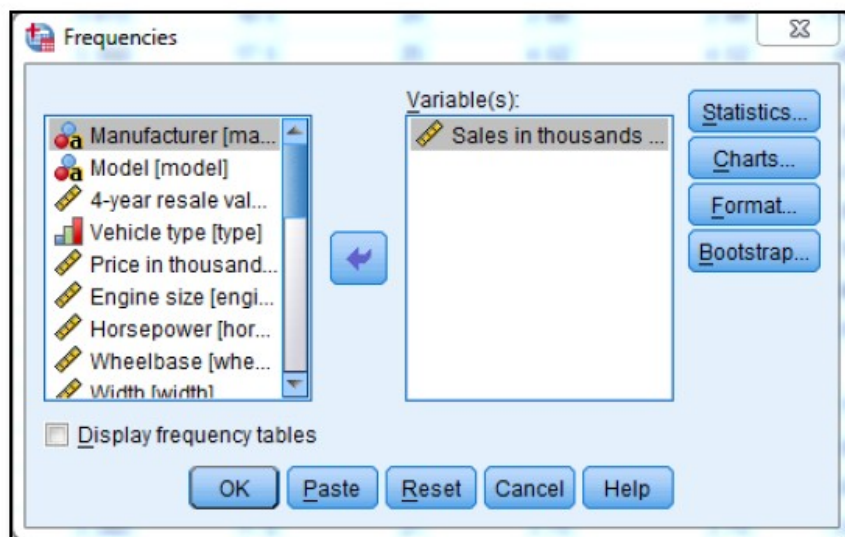
**مثال:** در فایل داده های مربوط به فروش خودرو به نام **Car\_Cales.sav** از مجموعه داده های **SPSS**، متغیری به نام **Sales** (قیمت فروش خودرو) وجود دارد که می خواهیم نرمال بودن آن را بررسی کنیم. برای این کار فایل داده های **Car\_Cales.sav** را از پوشه داده های **SPSS** به آدرس زیر باز کنید:

**C:\Program file\IBM\SPSS\Statistics\۲۰\Samples\English\Car\_Cales.sav**

فرمان زیر را در **SPSS** اجرا کنید:

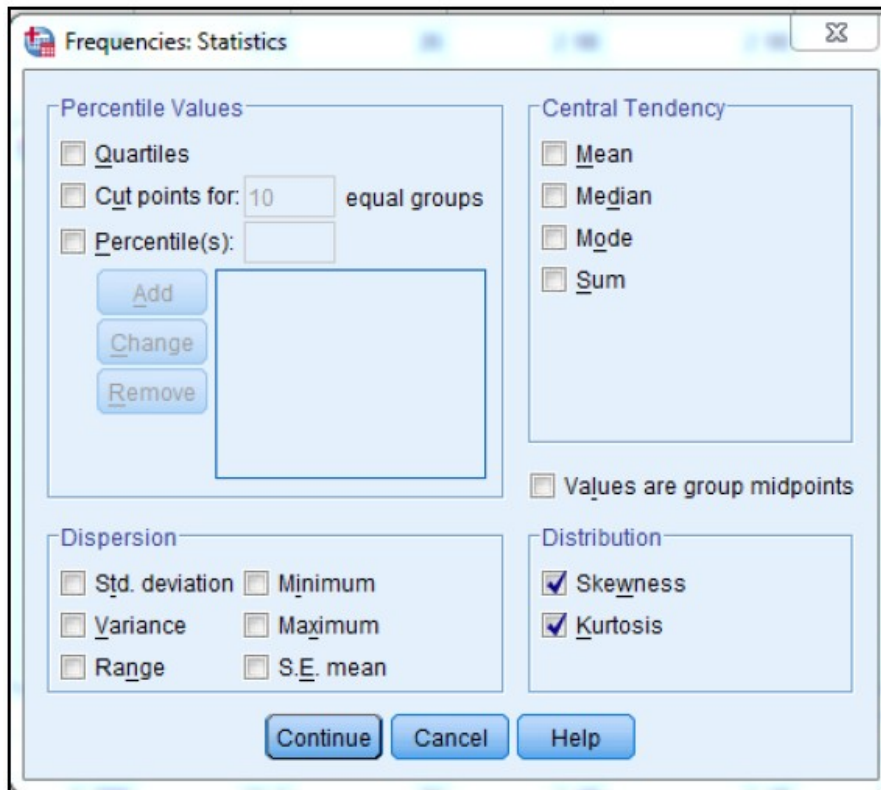
**Analyze/Descriptive Statistics/Frequencies**

کادر محاوره زیر را مشاهده خواهید کرد.



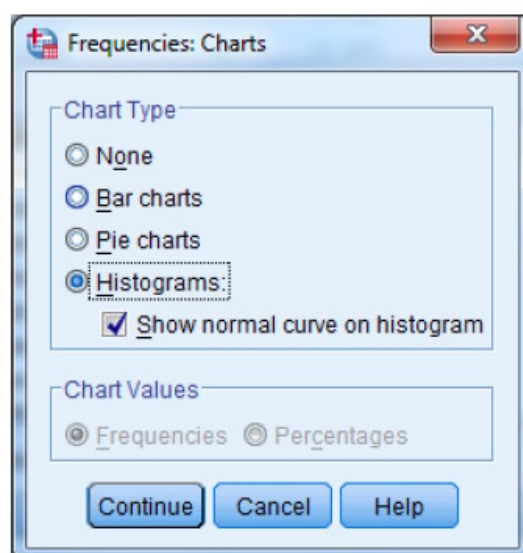
ابتدا متغیر **Sales** را از فهرست متغیرها در کادر سمت چپ، به کادر سمت راست انتقال دهید. سپس گزینه **Statistics** را کلیک کرده تا به کادر محاوره آن منتقل شوید.

در کادر جدید گزینه های **Skewness** و **kurtosis** را فعال کنید.



کلید **Continue** را کلیک کنید تا به کادر محاوره قبلی باز گردید.

روی کلید **Charts** کلیک کنید تا کادر محاوره آن مانند شکل زیر باز شود.



گزینه **Histograms** که منجر به رسم نمودار مستطیلی (هیستوگرام) می شود، را انتخاب کنید. متعاقب آن، گزینه

**Show normal curve on Histogram** را نیز فعال کرده و کلید **Continue** را کلیک کنید.

قبل از انتخاب **Ok** در کادر محاوره **frequency**، گزینه **Display frequency tables** را غیرفعال کنید. دلیل این

مورد آن است که در این مثال به جدول توزیع فراوانی نیازی نیست.

**OK** را کلیک کنید و نتایج را مانند زیر مشاهده نمایید :

Statistics		
Sales in thousands		
N	Valid	157
	Missing	0
Skewness		3.409
Std. Error of Skewness		.194
Kurtosis		17.557
Std. Error of Kurtosis		.385

مقدار چولگی مشاهده شده برابر  $3/409$  است و در بازه  $(2, -2)$  قرار ندارد. یعنی به لحاظ چولگی، متغیر **Sales**

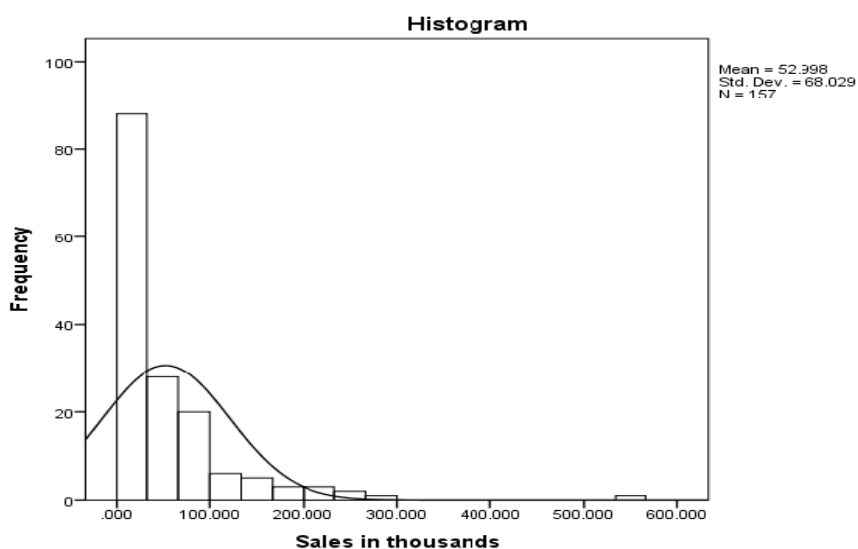
نرمال نبوده و توزیع آن نامتقارن و به سمت راست چولگی دارد. مقدار کشیدگی آن  $17/557$  شده است و در بازه

$(2, -2)$  قرار ندارد. این نشان می دهد که توزیع متغیر نسبت به توزیع نرمال، کشیدگی زیادی دارد و از حالت نرمال

خارج شده است.

با توجه به این دو شاخص که در حالت نرمال قرار ندارند، حدس می زنیم متغیر **Sales** توزیعی غیر نرمال داشته باشد.

این موضوع در نمودار به دست آمده در زیر نیز کاملاً مشهود است.



## ❖ تحلیل استنباطی

پس از بررسی مقدماتی متغیر **Sales**، حدس ما این بود که توزیع متغیر **Sales**، نرمال نباشد. اما برای اثبات این ادعا، باید از آزمون های آماری استفاده کنیم. هنگام بررسی نرمال بودن داده ها، ما فرض صفر را مبتنی بر اینکه توزیع داده ها نرمال است، در نظر می گیریم و مثل همه آزمون ها، آن را در سطح خطای ۵ درصد تست می کنیم.

در نرم افزار **SPSS** آزمونهایی وجود دارند که به کمک آنها می توان نرمال بودن یا نبودن توزیع داده ها را آزمون کرد. یکی از این آزمون ها، آزمون شاپیرو -ویلک است و دیگری آزمون کلموگروف -اسمیرنوف که هر دوی آنها را برای بررسی نرمال بودن متغیر **Sales** در مثال بالا، مورد استفاده قرار می دهیم. همچنین آزمون دیگری با نام اندرسون -دارلینگ نیز برای بررسی نرمال بودن داده ها استفاده می شود که ما در این مقاله کمتر به آن توجه داریم.

دو آزمون کلموگروف -اسمیرنوف و شاپیرو -ویلک مشابه هم هستند. ولی در عمل تفاوت اندکی با یکدیگر دارند.

الف - اگر حجم نمونه کم است، هر دو آزمون نتایج یکسانی دارند.

ب - اگر حجم نمونه بیشتر از ۲۰۰۰ است، آزمون کلموگروف -اسمیرنوف مناسب تر از شاپیرو است.

## ❖ آزمون شاپیرو ویلک :

شاخص آزمون شاپیرو ویلک (W) از فرمول زیر محاسبه می شود :

$$w = \frac{\left( \sum_{i=1}^n a_i X_{(i)} \right)^r}{\sum_{i=1}^n (X_i - \bar{X})^r}$$

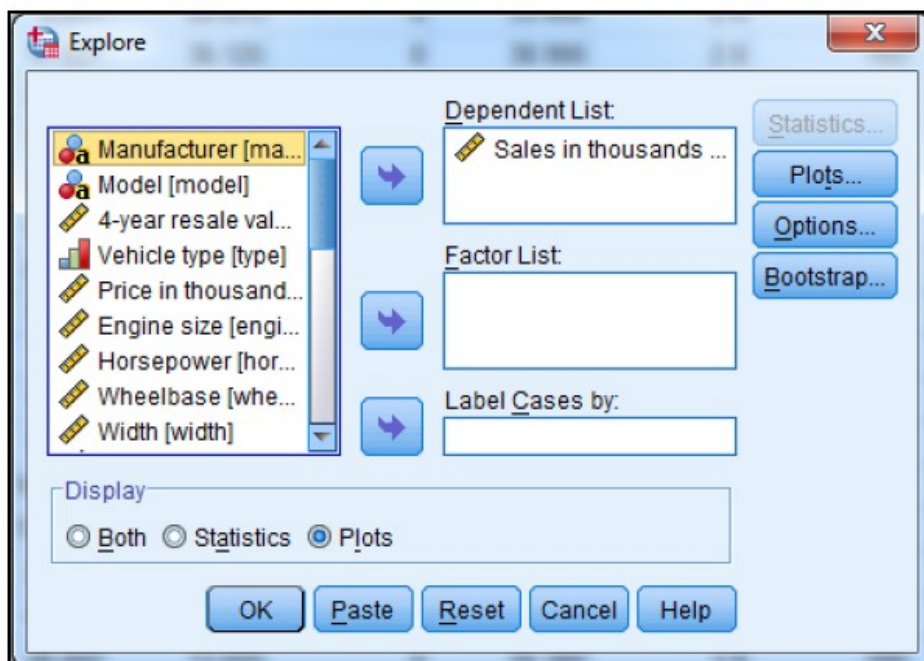
که در آن  $X_i$  ها آماره های مرتب هستند.

## ❖ مثال :

مسیر زیر را دنبال کنید **SPSS** : در مثال قبل، برای انجام این آزمون در

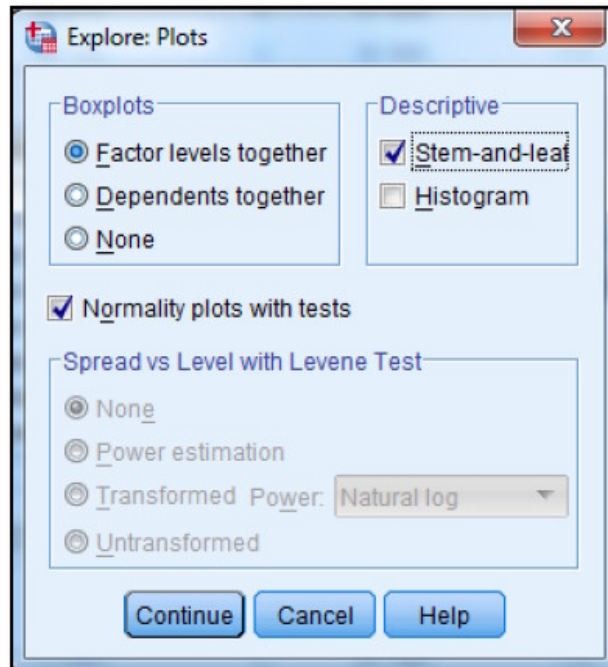
### Analyze/Descriptive Statistics/Explore

به کادر محاوره ای مانند شکل زیر منتقل خواهید شد.



متغیر Sales را مطابق شکل، وارد فهرست کنید : **Dependent List** .

کلید **Plots** را کلیک کنید تا به کادر محاوره ای مانند شکل زیر منتقل شوید.



در این کادر محاوره، گزینه هایی برای رسم نمودار جعبه ای و نمودار ساق و برگ انتخاب شده است، آنها را تغییر ندهید. این نمودارها به صورت بصری، نرمال بودن یا نبودن متغیر مورد آزمون را تایید می کنند.

اگر گزینه ی **Normality plots with tests** را علامت دار کنید، هر دو آزمون شاپیرو ویلک و کلموگروف - اسمیرنوف، انجام خواهد شد. کلید **Continue** را کلیک نمایید تا به کادر محاوره اول برگردید.

در قسمت **Display**، گزینه **Plots** را انتخاب کنید و **OK** را کلیک نمایید و نتایج به دست آمده را در فایل خروجی **SPSS**، مانند زیر مشاهده کنید.

نتایج شامل جدولی است تحت عنوان **Tests of Normality** است که سطح معناداری را برای هر دو آزمون گفته شده، مشخص کرده است.



### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Sales in thousands	.218	157	.000	.667	157	.000

a. Lilliefors Significance Correction

معمولا چنانچه سطح معنی داری در این دو آزمون که در این جدول با **sig** نمایش داده شده، بیشتر از ۵ درصد باشد، می توان داده ها را با اطمینان بالایی نرمال فرض کرد. در غیر این صورت نمی توان گفت توزیع داده ها، نرمال است. با توجه به جدول فوق و مقادیر سطح معنی داری **sig** که در هر دو آزمون کلموگروف -اسمیرنوف و شاپیرو -ویلک، کمتر از ۵ درصد شده است، فرض نرمال بودن توزیع متغیر **Sales** رد می شود. به عبارتی همان گونه که در تحلیل توصیفی پیش بینی کرده بودیم، نمی توان گفت توزیع داده های این متغیر، نرمال است.

نمودار ساق و برگ و نمودار جعبه ای در زیر این موضوع را تایید می کنند.

### Sales in thousands Stem-and-Leaf Plot

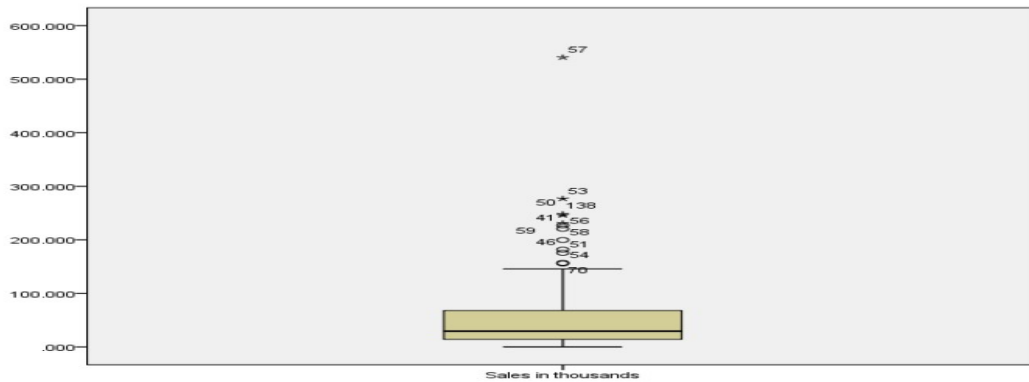
#### Frequency Stem & Leaf

```

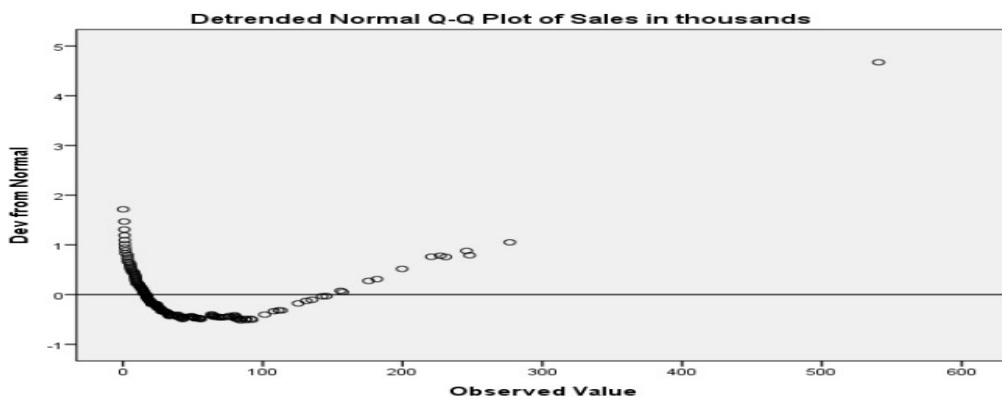
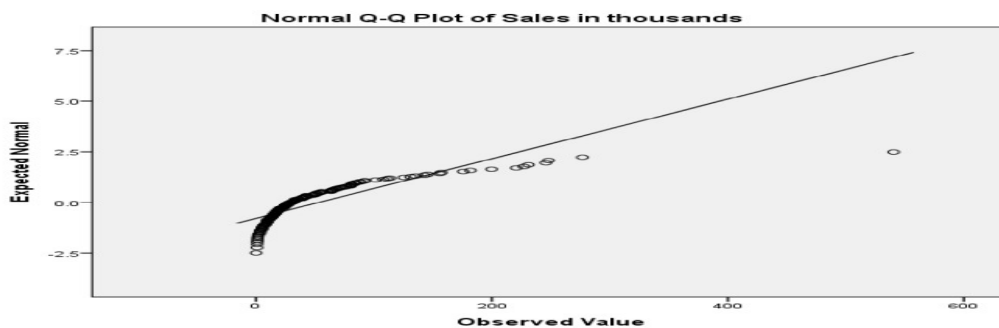
۳۱.۰۰۰ ۰ . ۰۰۰۱۱۱۱۱۱۱۳۳۳۳۴۵۵۵۵۶۶۷۷۸۸۸۹۹۹۹۹۹
۲۷.۰۰۰ ۱ . ۱۱۱۲۲۲۳۳۴۴۴۴۵۵۵۶۶۶۶۷۷۷۸۸۸۹۹
۲۱.۰۰۰ ۲ . ۰۰۰۱۲۳۴۴۴۴۵۵۶۶۶۷۷۷۷۸۹
۱۶.۰۰۰ ۳ . ۰۱۱۲۲۲۳۳۵۵۸۹۹۹۹
۹.۰۰۰ ۴ . ۱۲۲۲۲۷۸۹۹
۷.۰۰۰ ۵ . ۱۱۱۳۴۵۵
۸.۰۰۰ ۶ . ۳۳۳۵۵۶۷۸
۶.۰۰۰ ۷ . ۰۱۳۶۶۹
۹.۰۰۰ ۸ . ۰۰۰۱۳۳۴۸۸
۲.۰۰۰ ۹ . ۱۲
۲.۰۰۰ ۱۰ . ۱۷
۲.۰۰۰ ۱۱ . ۱۳
۱.۰۰۰ ۱۲ . ۵
۲.۰۰۰ ۱۳ . ۱۵
۲.۰۰۰ ۱۴ . ۲۵
۱۲.۰۰۰ Extremes (>=۱۵۶)

```

Stem width: ۱۰.۰۰۰  
Each leaf: ۱ case (s)



دو نمودار زیر که به نمودارهای **Q-Q** معروف هستند، برای بررسی نرمال بودن یا نبودن توزیع داده‌ها استفاده می‌شوند. در این نمودارها یک خط نرمال به عنوان معیار توزیع نرمال ترسیم شده است. هر مقدار نقطه‌ای که نماینده داده‌ها هستند، به این خط نزدیک تر باشند، نشانه نزدیکی توزیع داده‌ها به توزیع نرمال است و بر عکس. در هر دو نمودار زیر می‌بینید که نقاط مربوط به داده‌ها با خط معیار نرمال بسیار مقایرت دارند. این بدان معناست که داده‌ها نمی‌توانند توزیع نرمال داشته باشند.



## ❖ آزمون کولموگروف-اسمیرنوف

در نرم افزار SPSS آزمون کلموگروف-اسمیرنوف برای بررسی همتوزیعی داده ها با چهار توزیع نرمال، یکنواخت، پواسن و نمایی طراحی شده است. ما در اینجا قصد داریم تنها از این آزمون برای بررسی نرمال بودن داده ها استفاده کنیم.

در این آزمون، فرض صفر بر نرمال بودن توزیع داده ها بنا شده است. بنا بر این اگر فرض صفر رد شود به معنی غیرنرمال بودن داده ها است.

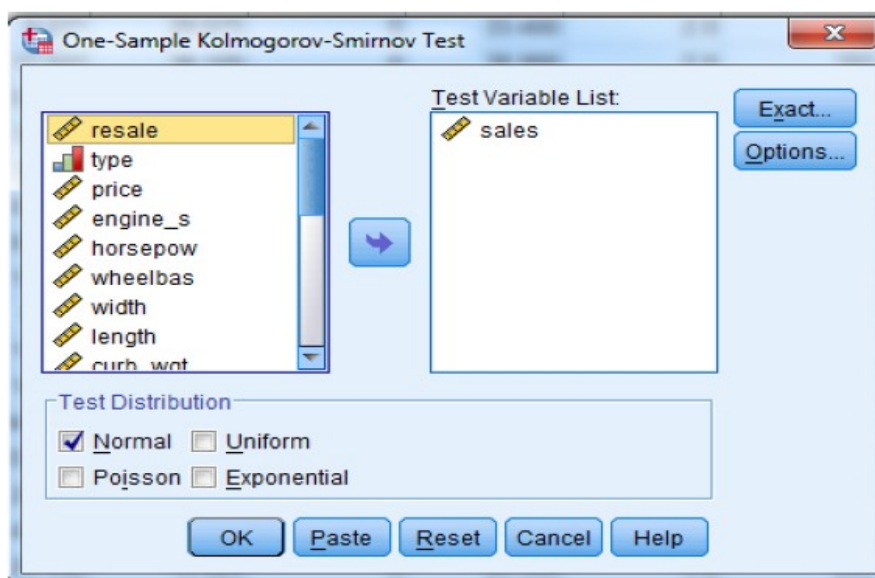
## ❖ مثال

فرض کنید در مثال قبل می خواهیم یک بار دیگر و تنها به کمک آزمون کلموگروف – اسمیرنوف نرمال بودن یا نبودن توزیع متغیر sales را بررسی کنیم.

مراحل زیر را برای انجام این آزمون دنبال کنید:

از فرمان **Analyze/Nonparametric Test/1 Samples K-S...**

به کادر محاوره آزمون کلموگروف – اسمیرنوف مانند شکل زیر وارد شوید.



متغیر sales را به فهرست متغیرهای آزمون ( **Test Variable List:** ) وارد کنید.

توجه داشته باشید گزینه **Normal** در قسمت **Test Distribution** که به طور پیش فرض انتخاب شده، تغییر نکرده باشد.

اگر گزینه **Options** را انتخاب کنید، می توانید بعضی از شاخص های توصیفی را محاسبه کنید. در اینجا ما از این گزینه استفاده نکرده ایم.

کلید **OK** را به کلیک کنید تا آزمون انجام شود. نتایج را در فایل خروجی و مانند جدول زیر مشاهده خواهید کرد.

**One-Sample Kolmogorov-Smirnov Test**

		Sales in thousands
N		157
Normal Parameters <sup>a,b</sup>	Mean	52.99808
	Std. Deviation	68.029422
Most Extreme Differences	Absolute	.218
	Positive	.197
	Negative	-.218
Kolmogorov-Smirnov Z		2.737
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

به جدول نتایج و مقدار **Asymp. Sig. (2-tailed)** دقت کنید. می بینید که مقدار  $0/000$  نشان از رد فرض صفر دارد.

یعنی این فرضیه که "نمونه مورد نظر از توزیع نرمال به دست آمده است"، رد می شود. به عبارتی توزیع این متغیر،

نرمال نیست.

## ❖ اقدامات بعدی

در ابتدای این مبحث دیدیم که شاخص های توزیع، نشان از غیر نرمال بودن متغیر Sales داشتند و هر دو آزمون کلموگروف - اسمیرنوف و شاپیرو - ویلک، فرض نرمال بودن داده ها را رد کردند. نمودارهای متعددی نیز این موضوع را به درستی تایید کردند. اکنون برای ما مسجل است که این متغیر نمی تواند نرمال باشد. وقتی با چنین پدیده ای روبرو می شوید ممکن است دو تصمیم بگیرید:

۱- این متغیر را با آزمون های پارامتری تحلیل کنید.

۲- بدلیل اهمیت موضوع، ناگزیر باشید این متغیر را به هر صورت ممکن به متغیری نرمال تبدیل نمایید.

## ❖ روش های نرمال کردن متغیرها

روش های گوناگونی برای نرمال کردن داده ها وجود دارد. که از مهمترین آنها می توان به تبدیل لگاریتمی، تبدیل باکس - کاکس، استفاده از نمودار احتمال لگاریتمی، استفاده از نمودار فینی و غیره اشاره کرد.

### ✓ تبدیل لگاریتمی

در این روش اگر فرض کنید  $x_i$  یک داده باشد، با تبدیل

$$y = Ln(x_i)$$

به جای  $x_i$  از  $y_i$  که مقدار لگاریتم نپین  $x_i$  است استفاده می شود. انتظار زیادی می رود که  $y_i$  دارای توزیع نرمال باشد. به این روش، تبدیل لگاریتمی ساده می گویند. در مواردی ممکن است با تبدیل لگاریتمی ساده

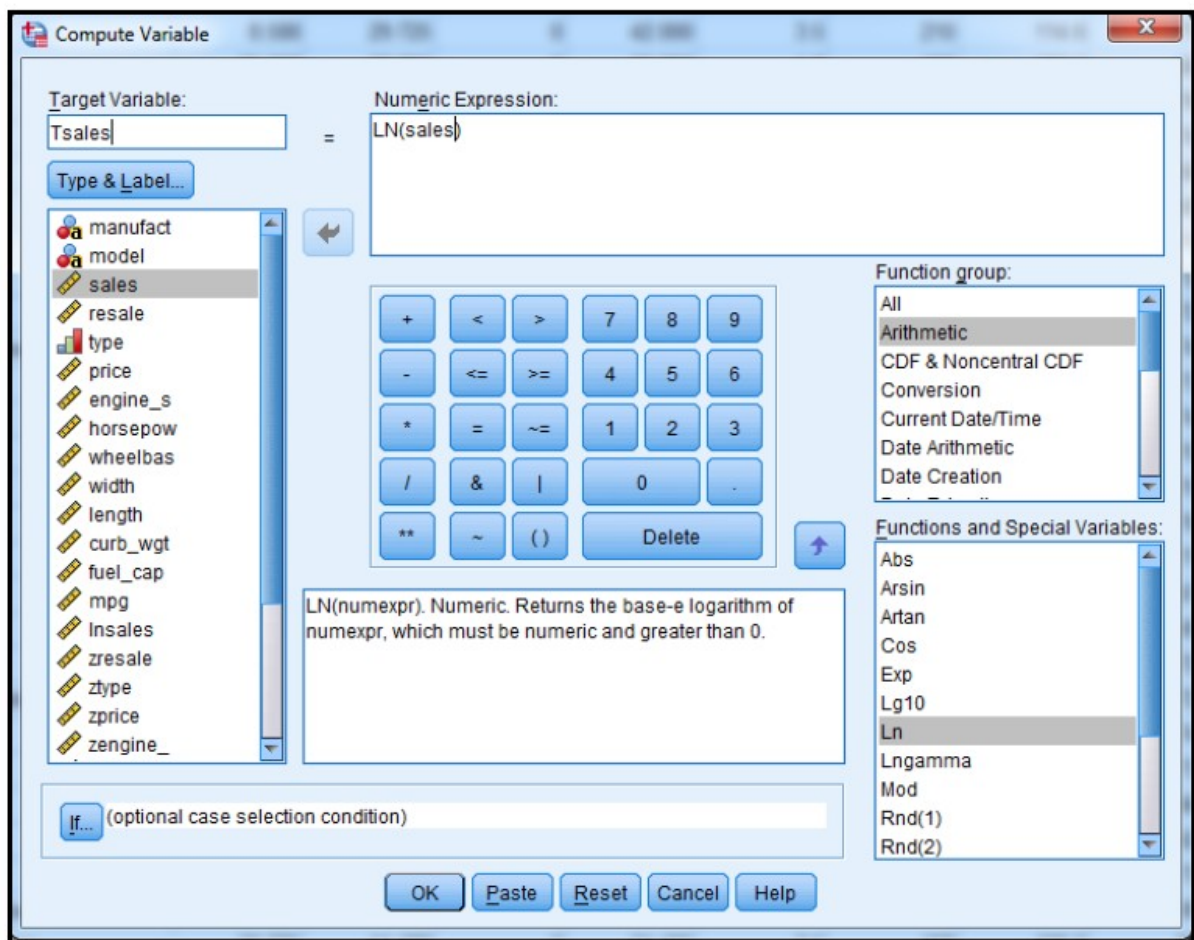
آن را به نتوان توزیع داده ها را به توزیع نرمال نزدیک کرد. کافی است با افزودن و یا کاستن عددی به  $x_i$ ، توزیع نرمال نزدیک نمود. این روش، که به روش تبدیل لگاریتمی سه پارامتری معروف است، به صورت زیر تعریف

$$y = Ln(ax_i \pm b)$$

می شود:

که معمولاً مقدار **a** را یک در نظر می‌گیرند و تنها برای **b**، یک مقدار مثبت یا منفی تعیین می‌کنند.

در مثال بالا می‌خواهیم به کمک تبدیل لگاریتمی، داده‌های متغیر **sales** را تغییر داده و بررسی کنیم آیا با تبدیل فوق، توزیع این متغیر نرمال خواهد شد. برای این کار از منوی اصلی گزینه **Transform** و متعاقب آن گزینه **Compute Variables...** را انتخاب کنید تا کادر محاوره‌ای مانند شکل زیر را مشاهده نمایید.



در قسمت **Target Variable:** یک نام جدید دلخواه برای متغیری که قرار است محتوی لگاریتم داده‌ها باشد،

انتخاب کنید. ما برای آن نام **Tsales** را برگزیده ایم.

در بخش **Function Group:**، گروه توابع محاسباتی **Arithmetic** را انتخاب کنید. در کادر زیر این بخش مجموعه

ای از توابع محاسباتی را مشاهده خواهید کرد. شما باید تابع لگاریتم نپیر را که با حروف اختصاری **Ln** نشان داده شده

است، انتخاب نمایید. برای این کار دو بار روی آن کلیک کنید تا به کادر **Numeric Expression** منتقل شود. در این حال آن را به صورت **Ln(?)** مشاهده می نمایید.

اینک کافی است متغیر **Sales** را از فهرست متغیرها به جای علامت سوال منتقل کنید. این کار را با تایپ کردن نام متغیر به جای علامت سوال، نیز می توانید انجام دهید.

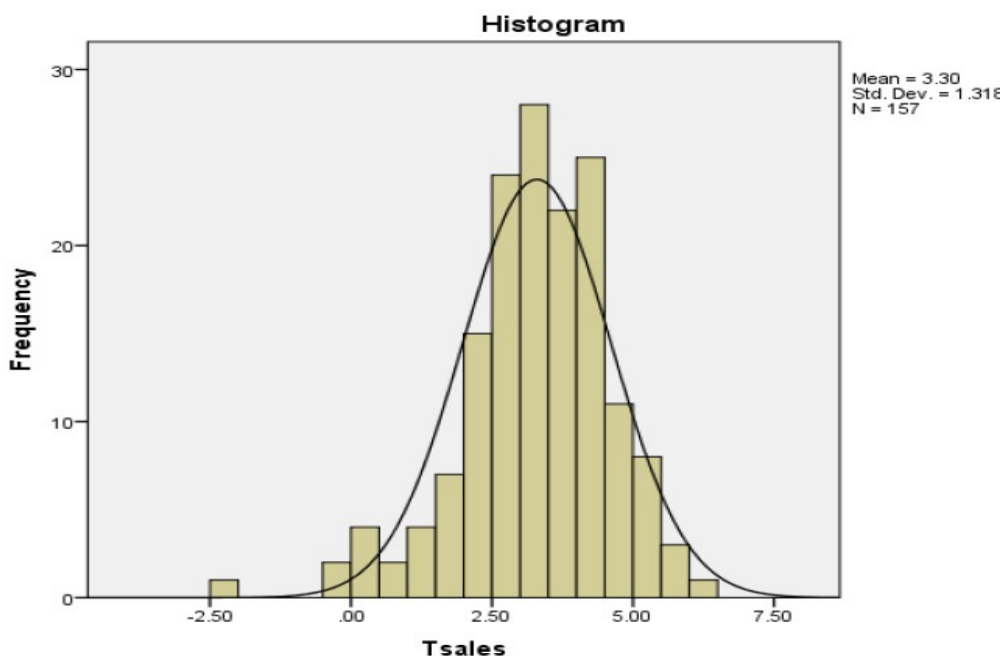
با انجام این مراحل، شما از نرم افزار خواسته اید، لگاریتم نپیرین داده های متغیر **Sales** را محاسبه کرده و در متغیر **Tsales** ذخیره نماید.

با زدن کلید **OK** می توانید متغیر جدید را که به فهرست متغیرها اضافه شده، مشاهده نمایید. این متغیر ممکن است دارای توزیع نرمال باشد.

برای بررسی نرمال بودن متغیر جدید **Tsales** ما ابتدا هیستوگرام آن را از مسیر

### Analyze/Frequencies

ترسیم کرده و با انتخاب گزینه های مناسب، توزیع آن را با منحنی نرمال مقایسه کرده ایم. شکل زیر گویای آن است که با احتمال زیادی می توان گفت؛ با این تبدیل، توزیع داده ها نرمال شده است.



برای اثبات اینکه داده ها دارای توزیع نرمال هستند، کافی است متغیر **Tsales** را به کمک یکی از آزمون های گفته شده در بالا، مورد بررسی قرار دهیم. ما نرمال بودن آن را به کمک آزمون کلموگروف-اسمیرنوف انجام دادیم که نتایج آن در جدول زیر آمده است.

### One-Sample Kolmogorov-Smirnov Test

		Tsales
N		157
Normal Parameters <sup>a,b</sup>	Mean	3.2959
	Std. Deviation	1.31821
	Absolute	.066
Most Extreme Differences	Positive	.042
	Negative	-.066
Kolmogorov-Smirnov Z		.826
Asymp. Sig. (2-tailed)		.503

a. Test distribution is Normal.

b. Calculated from data.

همانگونه که شاهد هستید، معیار تصمیم (**Asymp. Sig. (2-tailed)**) نشان می دهد که با مقدار  $0/503$  فرض نرمال بودن داده ها، پذیرفته است. به عبارتی واقعا داده ها دارای توزیع نرمال هستند.

### ✓ تبدیل باکس-کاکس :

در این روش ابتدا تبدیلاتی برای پیدا کردن یک مقدار لامبدا ( $\lambda$ ) انجام می شود. سپس به صورت کلی به ازای هر مشاهده  $X_t$  از تبدیل زیر استفاده می شود.

$$T(X_t) = \frac{X_t^\lambda - 1}{\lambda}$$

در استفاده از این تبدیل لازم است بدانید؛

(۱) اگر مقدار  $\lambda$  برابر صفر باشد نیاز به تبدیل نیست،

(۲) اگر مقدار  $\lambda$  برابر ۱- بدست آید، از فرمول بالا استفاده نمی کنیم بلکه از معکوس داده ها استفاده می کنیم.



## ❖ تبدیل باکس - کاکس در SPSS

### ❖ روش اول

اگر از قبل مقداری برای لامبدا به دست آورده اید، برای استفاده از روش باکس - کاکس در SPSS کافیست از منوی **transform** گزینه **compute variable** را انتخاب و در قسمت **target** یک متغیر جدید تعریف کرده و در بخش **Numeric expression:** فرمول های کاکس باکس را اعمال کنید.

### ❖ روش دوم

فرض کنید به عنوان مثال می خواهیم متغیر **Sales** را به روش باکس - کاکس، تبدیل کرده و بررسی کنیم آیا متغیر تبدیل یافته می تواند توزیع نرمال داشته باشد؟ برای این کار مراحل زیر را دنبال می کنیم.

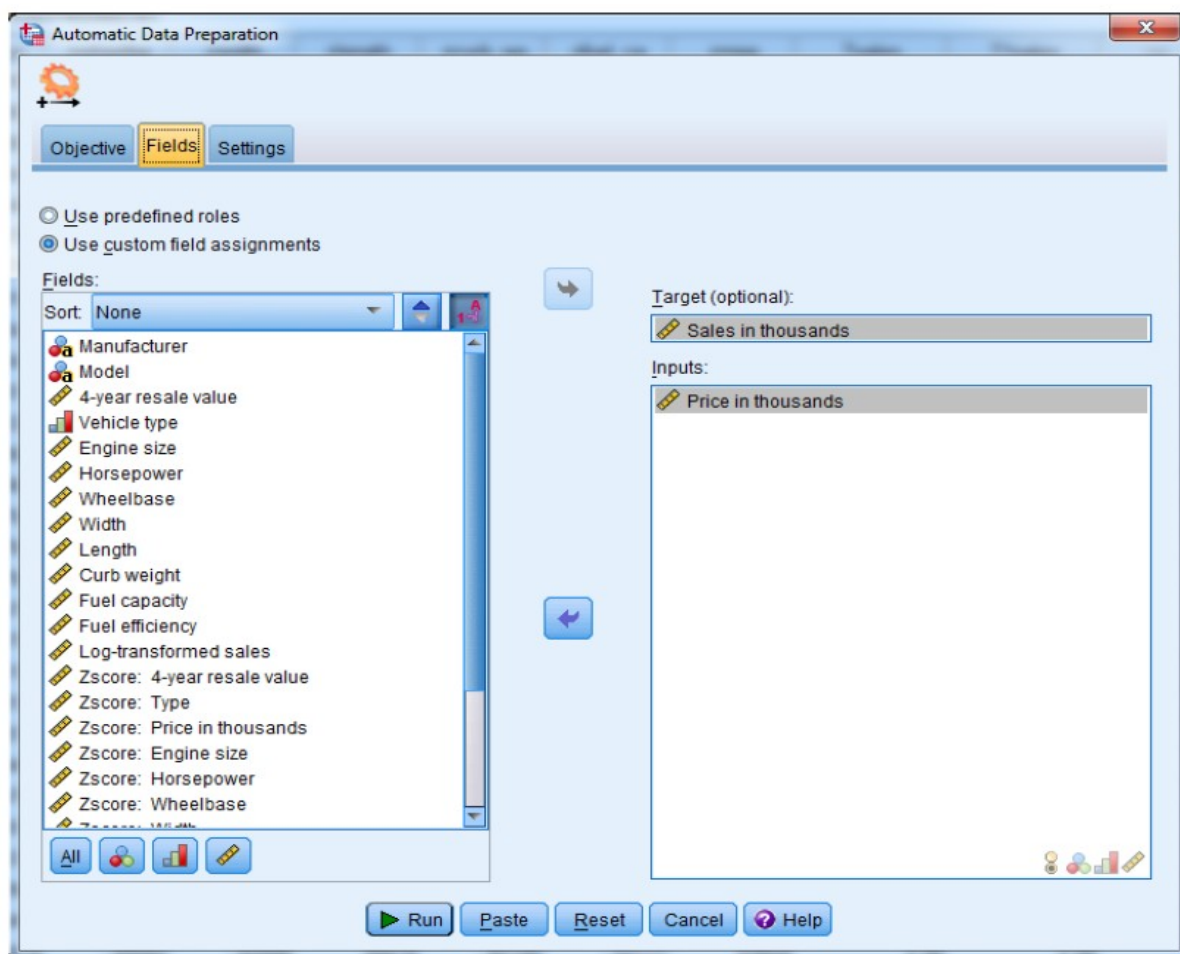
از منوی اصلی **SPSS**، گزینه **Transform** و متعاقب آن گزینه **Prepare Data for Modeling** را انتخاب کنید و سپس گزینه **Automatic...** را کلیک کنید تا کادر محاوره ای مانند شکل زیر مشاهده نمایید.



این کادر محاوره، شامل سه برگه است که شما در وحله اول، باید روی برگه **Objective** واقع باشید. در این برگه،

گزینه **Customize analysis** را انتخاب نمایید.

اینک روی برگه **Fields** کلیک نمایید تا به شکل زیر، آن را ببینید.



اگر متغیرها را در کادر **Target (optional):** مشاهده می کنید، آنها را به کادر **Fields:** منتقل نمایید.

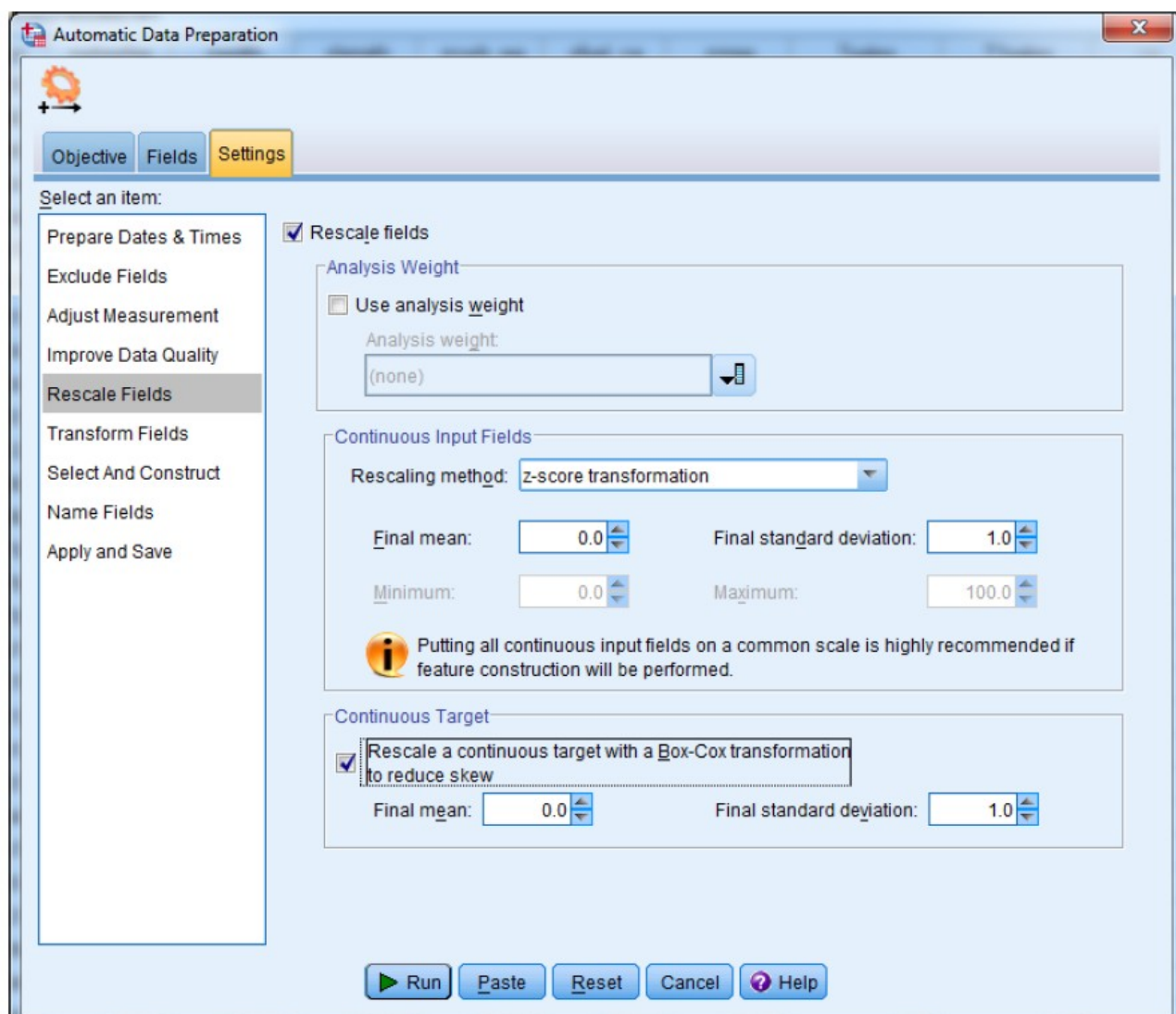
اینک متغیر یا متغیرهایی را که می خواهید به کمک تبدیل باکس - کاکس، تغییر دهید، در قسمت **Target**

**(optional):** وارد نمایید. ما در این مسئله، متغیر **Sales** را به این قسمت منتقل می کنیم.

یک متغیر پیوسته و دلخواه، از فهرست متغیرها، به بخش **inputs:** وارد نمایید. ما متغیر **Price** را برای این

کادر انتخاب کرده ایم.

حال به برگه سوم یعنی برگه **Settings** بروید که کادر محاوره آن مانند شکل زیر است.



ابتدا در فهرست **Select an item:**، آیتم **Rescale Fields** را انتخاب کنید. سپس گزینه:

### **Rescale a continuous target with a Box-Cox transformation to reduce skew**

را علامت دار کنید. سایر گزینه های این کادر محاوره را تغییر ندهید.

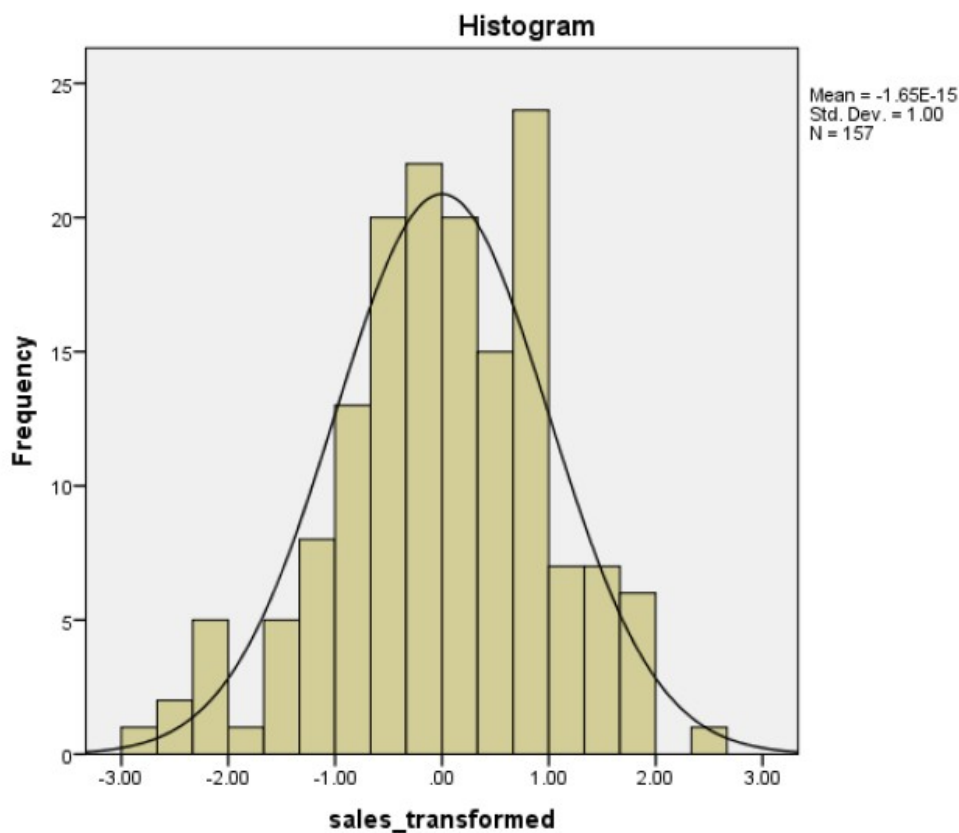
اکنون کلید **Run** را بزنید تا تبدیل خواسته شده، انجام شود.

می توانید تبدیل یافته ی متغیرها را که به فهرست متغیرهای نرم افزار، اضافه شده اند، مشاهده نمایید. متغیر مورد نظر ما را که به کمک تبدیل باکس – کاکس به دست آمده است، در فهرست متغیرها با نام **sales\_transformed** مشاهده می کنید.

اکنون ما برای اطمینان از اینکه تبدیل فوق توانسته است توزیع متغیر **Sales** را نرمال کند، ابتدا هیستوگرام متغیر **sales\_transformed** را از مسیر

### Analyze/Frequencies

ترسیم کرده و با انتخاب گزینه های مناسب، هیستوگرام توزیع آن را با منحنی نرمال مقایسه کرده ایم. شکل زیر به لحاظ بصری، نرمال بودن توزیع داده های جدید را تایید می کند.



برای اثبات اینکه داده ها دارای توزیع نرمال هستند، کافی است متغیر **sales\_transformed** را به کمک یکی از آزمون های گفته شده در بالا، مورد بررسی قرار دهیم. ما نرمال بودن آن را به کمک آزمون کلموگروف – اسمیرنوف انجام داده ایم که نتایج آن در جدول زیر آمده است.

		sales_transformed
N		157
Normal Parameters <sup>a,b</sup>	Mean	.0000
	Std. Deviation	1.00000
	Absolute	.041
Most Extreme Differences	Positive	.034
	Negative	-.041
Kolmogorov-Smirnov Z		.516
Asymp. Sig. (2-tailed)		.953

a. Test distribution is Normal.

همانگونه که شاهد هستید، معیار تصمیم **(Asymp. Sig. (2-tailed))**، با مقدار  $0/935$  قویا نشان می دهد که فرض نرمال بودن داده ها، پذیرفته است. به عبارتی واقعا داده ها دارای توزیع نرمال هستند.

اگر چه روش باکس - کاکس، کاربرد بسیاری دارد، اما باید به معایب آن نیز توجه داشت. این روش در برخی موارد نمی تواند تبدیل مناسبی برای نرمال کردن داده ها پیدا کند. ضمن اینکه از این روش تنها برای داده هایی با مقادیر مثبت، استفاده می شود.

برای رفع چنین مشکلاتی یک تبدیل دیگر به نام تبدیل جانسون (**Johnson-Transformation**) ارائه شده که از پیچیدگی و کیفیت بیشتری نسبت به تبدیل **BOX-COX** برخوردار است. البته بهتر است اول روش باکس – کاکس را انجام دهید و اگر تبدیل مناسبی به دست نیامد، سپس از تبدیل جانسون استفاده نمایید. برای انجام این تبدیل، اگر با نرم افزار **MINITAB** آشنایی جزئی داشته باشید، می توانید این تبدیل را به راحتی روی داده ها اعمال کنید تا داده ها توزیع نرمال پیدا کنند.

اطلاعات بیشتر در مورد این تبدیل را در مقاله ی زیر مطالعه کنید.

**"Transforming non normal Data to Normality in Statistical Process Control" Y. Chou, A.M. Polansky, and R.L. Mason (۱۹۹۸)**

برای اطلاع بیشتر از نرمال سازی داده ها، به کتاب "تحلیل داده های اکتشافی" تألیف دکتر علی اصغر حسینی پاک و مهندس

محمد شرف الدین، انتشارات دانشگاه تهران مراجعه شود.

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.