آمـــوزش جـــامع نـــــرم‌افـــــزار SPSS

مدرس: رضا بهرامی

کارشناس ارشد آمار کاربردی

(دانشگاه علامه طباطبایی تهران)

شماره تماس: ۰۹۳۹۷۴۹۴۰۵۲

http://statcamp.blogfa.com
http://spss-lisrel.org

مقدمه

با پیشرفت علوم و گسترش تکنولوژی، اهمیت استفاده از روش‌های آماری در علوم مختلف بیش از پیش مورد توجه قرار گرفته است و آموختن آمار کاربردی در هر رشته جزء ملزومات گردیده است. فرآیند آنالیز آماری کمک می‌کند تا پژوهشگر بتواند از داده‌های اولیه، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد. اگر حجم داده‌ها بزرگ باشد، استفاده از روش‌های مختلف آنالیز آماری بسیار خسته کننده و مشکل خواهد بود. امروزه انواع نرم افزارهای مختلف آماری موجود، قادرند انواع آنالیزهای آماری را انجام دهند.

نرم افزار **SPSS** یکی از قدیمی ترین، برنامه‌های کاربردی در زمینه تجزیه و تحلیل‌های آماری است؛ نرم افزاری آماری با قابلیت‌های انجام توصیفی زیبا و گویا از اطلاعات، شامل رسم نمودارها و چارت‌های گوناگون و محاسبات مربوط به میانگین، انحراف معیار واریانس، میانه و ....

کلمه **SPSS** مخفف Statistical Package for Social Science (نرم افزار آماری برای علوم اجتماعی) می‌باشد. این نرم‌افزار که یکی از نرم‌افزارهای تخصصی آمار است، بیشتر به بحث‌های آماری در حیطهٔ علوم اجتماعی، مدیریت روانشناسی و علوم رفتاری و ... می‌پردازد.

برخی از قابلیت‌های نرم افزار **SPSS** به شرح زیر است:

- تهیه خلاصه‌های آماری مانند گراف‌ها، جداول، آماره‌ها و ...
- انواع توابع ریاضی مانند قدر مطلق، تابع علامت، لگاریتم، توابع مثلثاتی و...
- تهیه انواع جداول سفارشی مانند جداول فراوانی، فراوانی تجمعی، درصد فراوانی و...
- انواع توزیع‌های آماری شامل توزیع‌های گسسته و پیوسته
- تهیه انواع طرح‌های آماری
- انجام آنالیز واریانس یکطرفه، دوطرفه، چندطرفه و آنالیز کوواریانس
- ایجاد داده‌های تصادفی و پیوسته
- محاسبه انواع آماره‌های توصیفی
- انواع آزمون‌های مرتبط با مقایسه میانگین بین دو یا چند جامعه مستقل و وابسته
- قابلیت مبادله اطلاعات با نرم‌افزارهای دیگر
- برازش انواع مختلف رگرسیون

فصل اول:

آشنایی با محیط SPSS

چگونگی نصب و راه اندازی نرم افزار:

نصب نرم افزار

برای شروع نصب نرم افزار لازم است که ابتدا وارد سیستم عامل windows شوید و سپس CD نرم افزار را داخل CD Rom قرارادهید و پس از انتخاب شاخه SPSS با تقه زدن مضاعف بر روی آیکون Setup نصب نرم افزار آغاز می شود و با انتخاب گزینه های لازم می توان نصب را به پایان رساند (مراحل کامل نصب در آزمایشگاه کامپیوتر بیان واجرا خواهد شد).

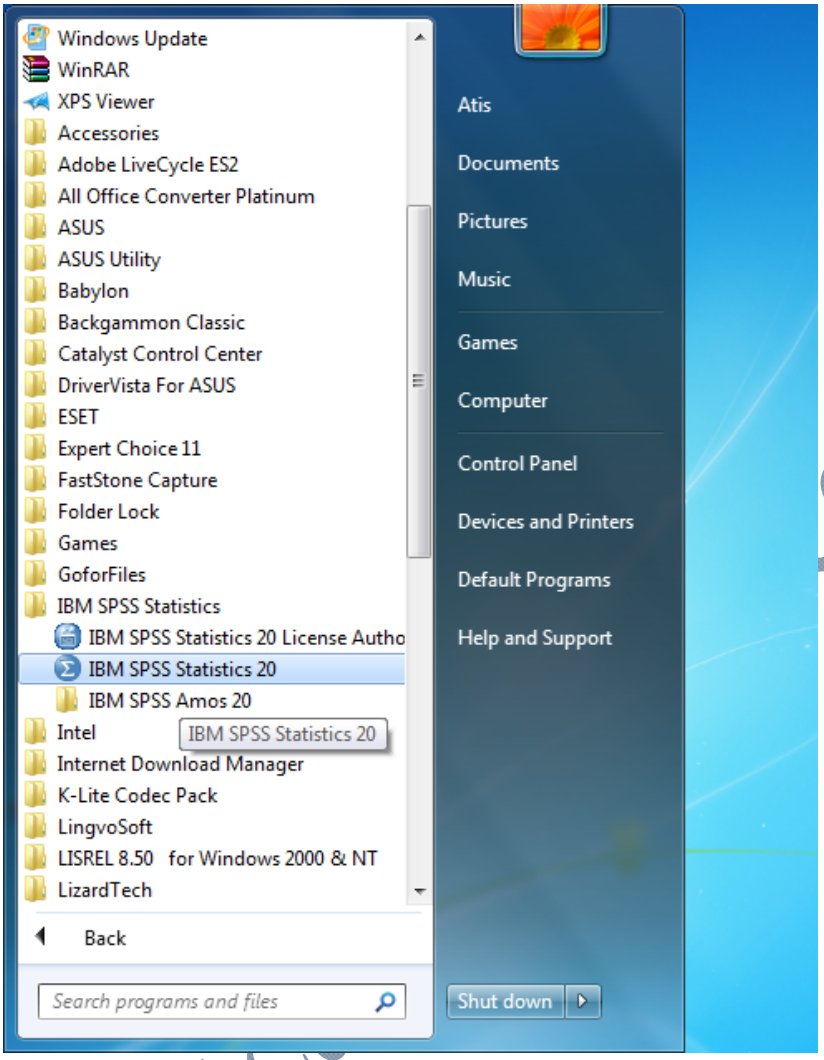
راه اندازی نرم افزار

اگر برنامه SPSS در دستگاه شما نصب یا در شبکه موجود باشد، چندین راه برای دسترسی به آن وجود دارد. اگر آیکون میانبر برای SPSS بر روی Desktop وجود داشته باشد، بر روی آن کلیک مضاعف کنید تا نرم افزار شروع به کار کند.

راه دیگر این است که بر روی یک فایل SPSS کلیک مضاعف کنید.

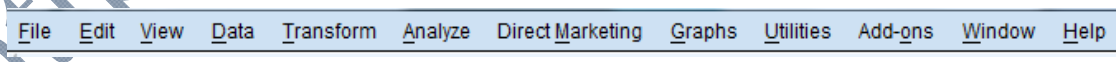
راه دیگر دسترسی به نرم افزار از منوی Start-taskbar-program را انتخاب نموده و در آن شاخه]IBM SPSS Statistics را انتخاب نموده و با کلیک روی IBM SPSS Statistics20 نرم افزار را اجرا می‌نماییم.

در زیر نحوه اجرا این حالت ارائه گردیده است.



پس از اجرای بسته ی نرم افزاری SPSS پنجره ی ویرایشگر داده‌ها بر روی صفحه‌ی نمایش ظاهر می شود.

علاوه بر آن کلیه ی منوهای اصلی نرم افزار، که مجموعه ی قابلیت های کلی آن را نمایش می دهد، بر روی یک نوار نشان داده می شود.



نوار منو شامل ۱۰ منوی اصلی است که هر یک بخشی از فعالیت‌های نرم افزار را بر عهده دارند. این منوها عبارت اند از:

۱- منوی فایل(File): ساختن فایل جدید داده‌ها، فراخوانی و ذخیره سازی انواع فایل‌ها در این نرم افزار و سایر نرم افزارها(مثلاً SAS یا EXCEL) چاپ گزارش‌ها و نمودارها.

۲- منوی ویرایش(Edit): انجام انواع ویرایش ها مانند کپی، حذف یا الصاق و جستجو در فایل داده ها.

۳- منوی نمایش(View): تغییرات در صفحه ی نمایش داده ها شامل تغییر فونت، سفارشی کردن نوار ابزار و غیره.
٤- منوی داده‌ها(Data): تغییر مشخصه های متغیرها، اضافه کردن متغیرها یا نمونه های جدید، مرتب کردن، جدا کردن، ادغام کردن، خرد کردن، وزن دادن و انتخاب بخشی از داده ها در یک فایل.
٥- منوی تبدیل(Transform): تبدیل، رتبه بندی و کدگذاری داده ها و جانهی مقادیر گمشده.
٦- منوی تحلیل(Analyze): انواع تحلیل های آماری مانند آمار توصیفی، مقایسه ی میانگین ها، رگرسیون، تحلیل واریانس و غیره.
۷- منوی نمودارها(Graphs): نمودارهایی مانند نمودار میله ای، دایره ای و خط ی، هیستوگرام، نمودار نرمال، نمودار پراکنش و غیره.
۸- منوی امکانات(Utilities): اطلاعاتی درباره ی فایل داده، سیستم مدیریت فایل خروجی، تغییر شکل.
۹- منوی پنجره(Window): مرتب کردن، گزینش و کنترل خصوصیات پنجره های ظاهری پنجره ها و ویرایش نوار ابزار نرم افزار
۱۰- منوی کمک رسانی(Help): راهنمایی‌های کلی در مورد نحوه‌ی استفاده از نرم افزار و بخش‌های مختلف آن، امکان جستجوی یک عنوان خاص، نمایش ترکیب فرمان های SPSS، و خلاصه‌ای از تعاریف اصطلاحات و فرمان ها.

معرفی نرم افزار و قابلیت‌های آن

ساختار کلی SPSS بر مبنای پنجره های ثابت و متحرک شکل گرفته است. این نرم افزار شامل ٤ پنجره اصلی است:

۱- data editor window
۲- output window
۳- syntax window
٤- help window

پنجره data editor

مهمترین و اصلی‌ترین پنجره نرم افزار پنجره data editor است. زیرا که برای فعال کردن سایر بخش‌های عملیاتی لازم و ضروری است که فایل داده‌ای در محیط SPSS فعال باشد و به همین دلیل است که پس از هر بار ورود به نرم افزار ابتدا پنجره data editor مشاهده می شود که خالی از داده می‌باشد و هر گونه ویرایشی بر روی داده‌های فایل جاری نیز در این پنجره قابل اجرا است.

لازم به ذکر است که این پنجره حذف شدنی نیست و همواره بر روی صفحه نمایش (به صورت گسترده یا کوچک شده) فعال است و در هر لحظه امکان دسترسی به آن وجود دارد.

نکته دیگر آنکه در پایین پنجره data editor دو انتخاب وجود دارد:

۱- data view ۲- variable view

انتخاب variable view جهت نامگذاری و تعیین مشخصات متغیرها به کار می‌رود و توسط انتخاب Data view کاربر قادر خواهد بود داده‌های خود را وارد نماید (این دو انتخاب در بخش بعدی به صورت کامل و جامع توضیح داده خواهند شد).

پنجره output

پنجره خروجی که از نوع متن است می‌تواند به تعداد متناهی در حین انجام کار با نرم افزار ایجاد شود، به همین دلیل اولین پنجره از این نوع با شماره ۱ مشخص می شود و در صورت ایجاد پنجره های خروجی دیگر به ترتیب شماره ۲، ۳ و ... برای آنها بکار می‌رود.

در این پنجره کاربر قادر خواهد بود کلیه خروجی‌های حاصل از پردازش بر روی داده‌ها را که می‌تواند متن، نمودار، جدول و .. باشد را مشاهده نماید. همچنین در این پنجره امکان ویرایش خروجی‌ها وجود دارد.

پنجره syntax

این پنجره که از نوع متن است برای آماده سازی اجرای فرمان‌ها در زیر برنامه نویسی SPSS طرح ریزی شده است. در این پنجره می‌توان امکانات ویرایشی را نیز به راحتی به کاربرد. همچنین نشانه‌هایی برای اجرا و الحاق فرمان‌ها تعبیه شده است. کاربر از این پنجره نیز می‌تواند به تعداد متناهی ایجاد نماید.

پنجره help

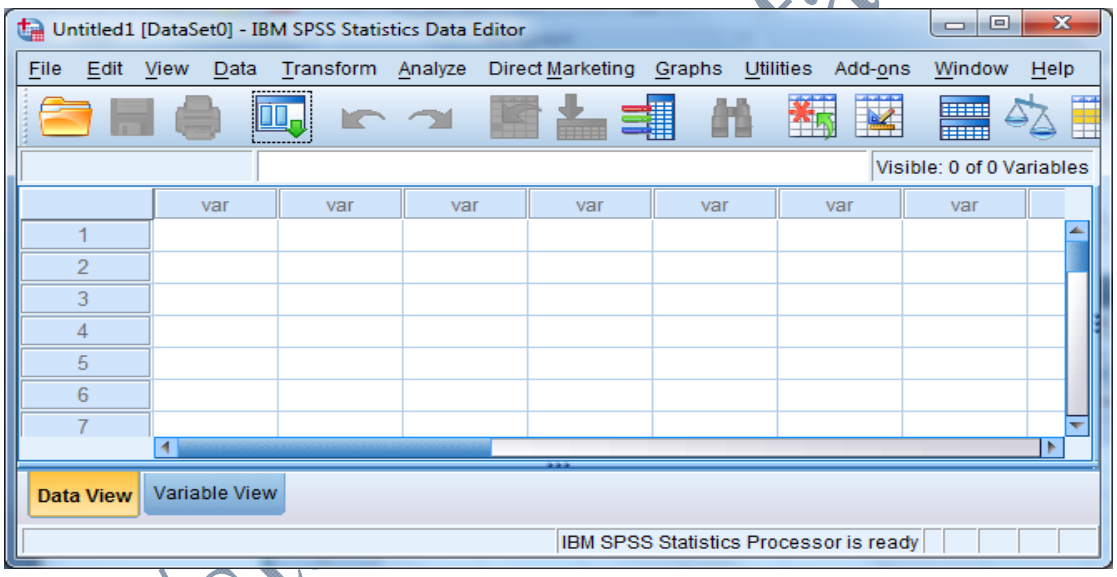
این پنجره حاوی اطلاعات مفید و اما مختصر در مورد نحوه کار با نرم افزار ماهیت فرمان‌ها و معرفی بخش‌های مختلف نرم افزار و برخی موارد آموزش مباحث مختلف آماری می باشد. برای استفاده از help نرم افزار می تواند از طرق مختلفی سود جست. یکی از راه‌های آن استفاده از منوی help می باشد. راه دیگر استفاده از دکمه help یی می باشد که در اکثر جعبه های گفتگو تعبیه شده است (که بجای فشردن آن می توان در هر کجای نرم افزار که باشد از کلید تابعی F1 استفاده نمایید) و اگر از موشواره استفاده می نمایید با فشردن کلید سمت راست آن (به شرطی که کلید اصلی موشواره کلید سمت چپ باشد) در برخی

نقاط نرم افزار با مشاهده عبارت what's this? و تقه زدن بر روی آن می توانید از help نرم افزار استفاده نمایید.

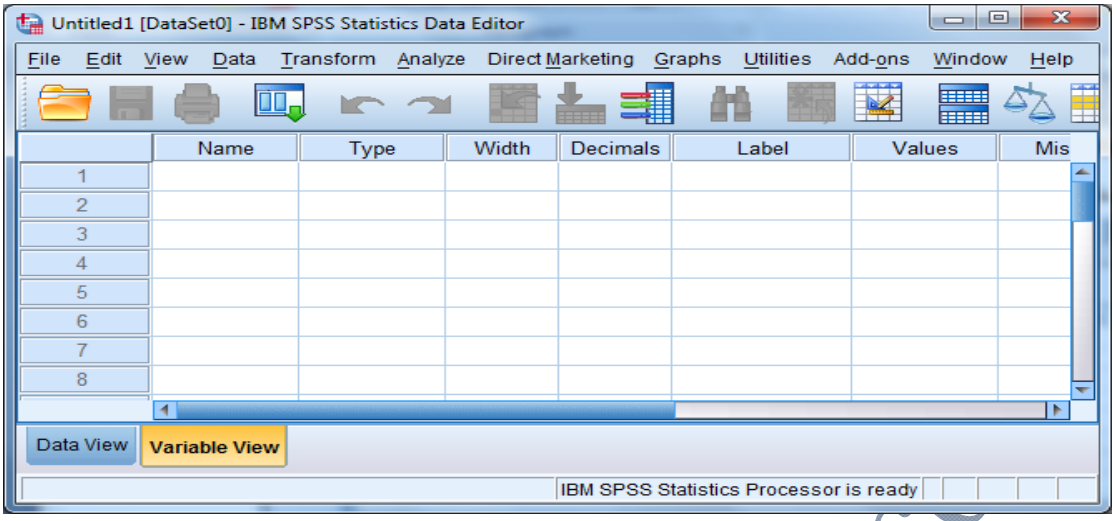
شروع کار با SPSS

وقتی برنامه SPSS را باز می‌کنیم، پنجره ای به نام Untitled- SPSS Editor (ویرایشگر SPSS – بدون نام) نشان داده می‌شود که شامل دو پنجره مختلف به نام‌های زیر می‌باشد:

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



۲- ***Variable view*** (نمای متغیر): این قسمت برای نام‌گذاری متغیرها، مشخصات و جزئیات مورد نظر آنها استفاده می شود.



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

مثال(۱): معدل دیپلم(کمی) و نوع دیپلم(کیفی) ۱۵ نفر از دانشجویان سال اول دانشگاه در چند رشته مختلف دانشگاهی به صورت زیر می باشد. به منظور استفاده از این اطلاعات در تجزیه و تحلیل‌ها، اطلاعات (داده‌ها) را به صورت زیر به نرم افزار وارد می کنیم.

کد ۱ = دیپلم ریاضی کد ۲ = دیپلم تجربی کد ۳ = دیپلم انسانی

برای این کار بعد از وارد شدن به نرم افزار SPSS به صورت زیر عمل می کنیم:

❖ در ستون ***Name*** ابتدا نامی برای متغیر انتخاب می کنیم(در اینجا داده‌های مربوط به معدل دانشجویان).

❖ قواعد نام‌گذاری متغیرها:

برای انتخاب نام متغیرها در SPSS موارد زیر بایستی رعایت شود:

۱- نباید بیشتر از هشت کاراکتر شود.

۲- باید با یک حرف یا @ شروع شود.

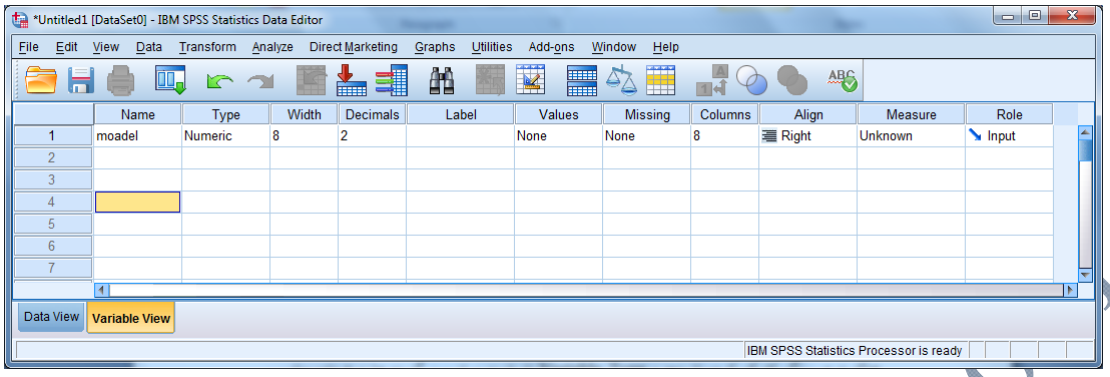
۳- می‌تواند شامل حروف، اعداد یا یکی از کاراکترهای @، #، - و یا $ باشد.

٤- نباید شامل فاصله و کاراکترهای خاص همانند ؟ باشد.

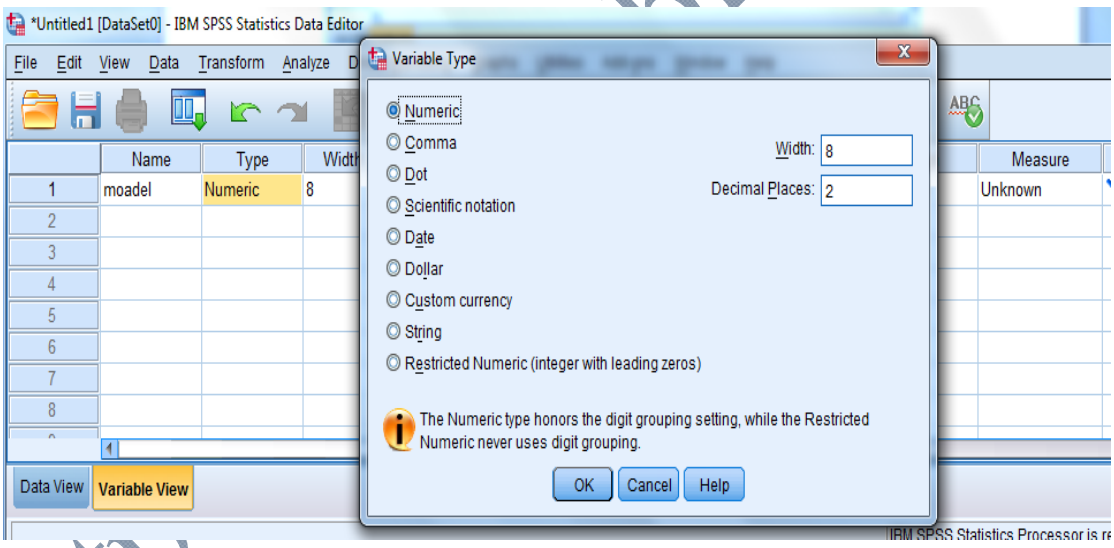
۵- نباید شامل کلمات کلیدی همانند AND، NOT، EQ، BY، ALL و کلماتی از این دست که SPSS از آنها به عنوان عبارات محاسباتی استفاده می‌کند، باشند.

نمونه	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
دیپلم	1	2	2	2	3	1	1	3	3	3	2	2	1	1	2
معدل	17	16.5	18.30	17.89	15.64	17.30	14	16.70	17.80	16	18.89	17.99	17.60	18	19

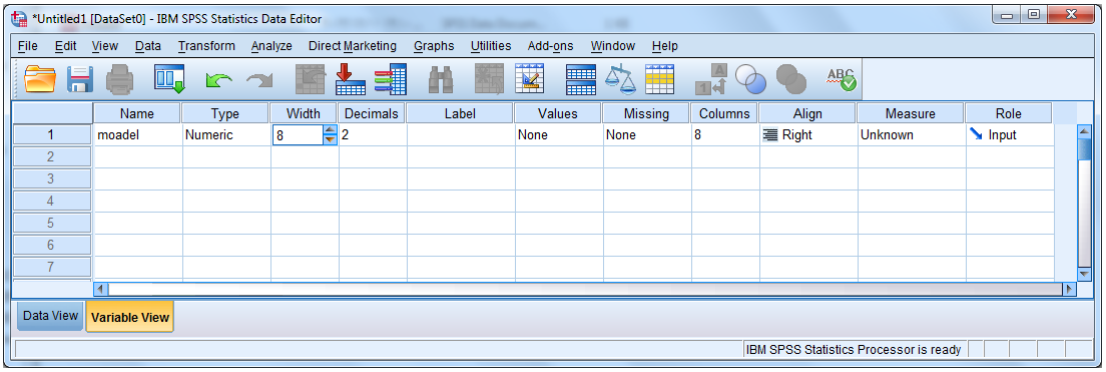
بعد از وارد کردن نام، سایر ستون‌ها با پیش فرض‌هایی که نرم افزار طراحی کرده است به صورت زیر نمایش داده می شوند که ما می‌توانیم به دلخواه و با توجه به نوع داده، تغییراتی در ستون‌ها ایجاد کنیم.



در مرحله بعد، درستون Type نوع متغیر را تعیین می کنیم. در این ستون روی مربع کوچک خاکستری رنگ کلیک کرده تا پنجره Variable Type باز شود. و از بین گزینه ها نوع داده مناسب را برای متغیر مورد نظر تعیین می کنیم. با توجه به اینکه معدل دانشجویان عددی می باشد همان گزینه Numeric (داده عددی) را بدون تغییر باقی می گذاریم.

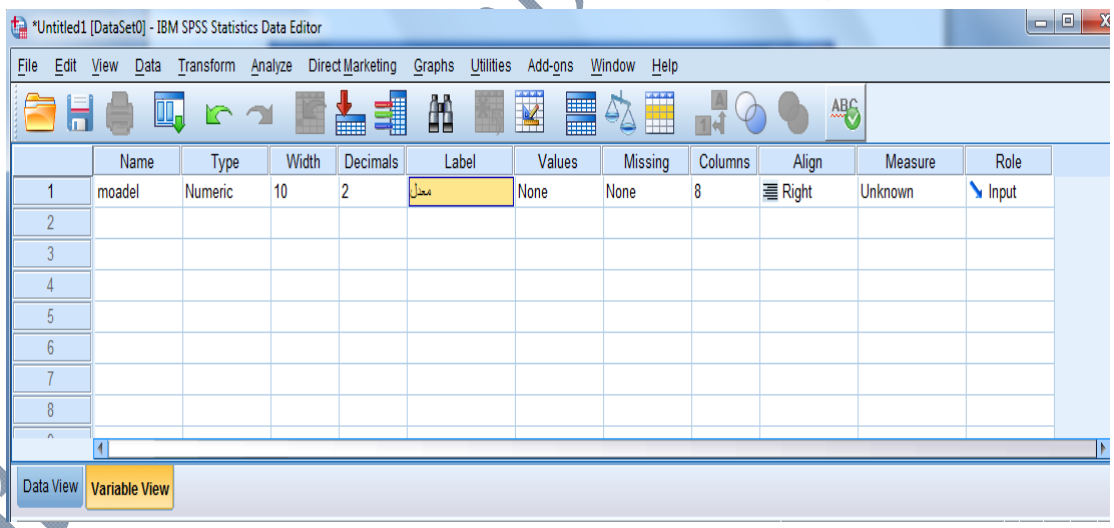


از ستون بعدی (Width) برای تغییر دادن پهنای متغیر استفاده می شود. با کلیک روی ستون، دو پیکان کوچک بالا و پایین نشان داده می شود که می توانیم با بالا و پایین کردن، پهنای مورد نظر را تغییر دهیم.

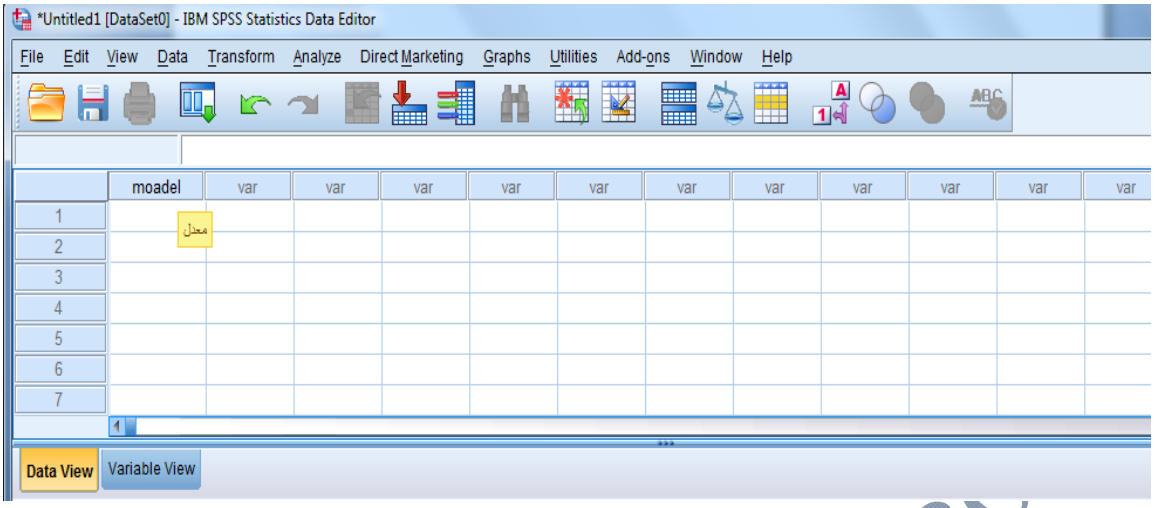


از ستون Decimals برای تعیین تعداد ارقام اعشار داده‌های مورد نظر استفاده می‌شود. مانند روش قبل روی ستون کلیک کرده و تعداد ارقام اعشار را تعیین می‌کنیم. برای این داده‌ها با توجه به اینکه معدل یک فرد می‌تواند به طور مثال ۱۸/۲۵ باشد به همین دلیل ستون مورد نظر را به همان صورت پیش فرض نگه می‌داریم. به این معنی که در ستون معدل در پنجره (Data View) اعداد مربوط به معدل دانشجویان تا دو رقم اعشار نشان داده می‌شوند.

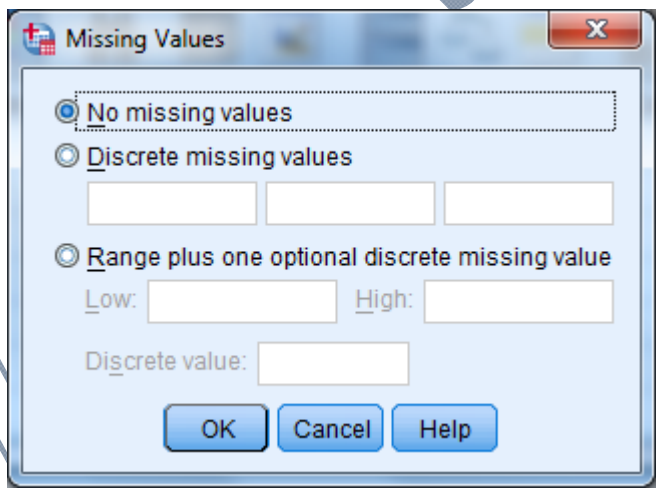
در ستون Lable می توان برای متغیر مورد نظر یک برچسب انتخاب کرد. برای این کار در ستون عنوان مورد نظر را تایپ می کنیم.



با این کار وقتی در پنجره Data view ، ماوس را روی عنوان moadel نگه داریم عنوان تایپ شده در Lable مشاهده می‌شود.

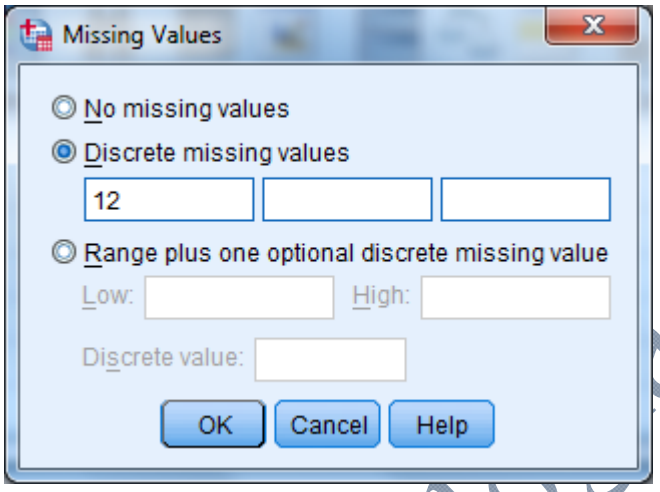


- ستون Values برای متغیرهای گروه‌بندی مورد استفاده قرار می‌گیرد. چون معدل دیپلم دانشجویان متغیر گروه‌بندی نمی‌باشد، اطلاعاتی در این قسمت اضافه نمی‌کنیم.
- ستون بعدی (Missing) مربوط به داده‌های گمشده می‌باشد. در این قسمت با کلیک بر روی ستون مورد نظر پنجره Missing Values باز می‌شود.

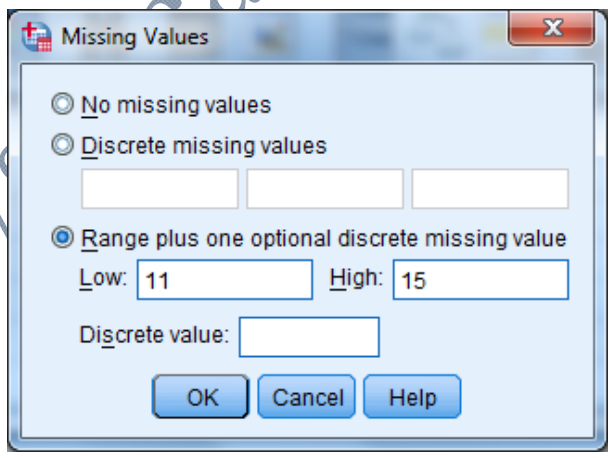


زمانی که در بین اطلاعات جمع آوری شده داده گمشده‌ای وجود نداشته باشد پیش فرض No missing values به همان حالت خود باقی می ماند. اما اگر داده گمشده وجود داشته باشد، برای مشخص کردن آن باید گزینه Discrete missing values را فعال کرد و شماره سطر مربوط به داده مورد نظر را در مستطیل‌های زیر وارد کرد. به طور مثال اگر معدل یکی از ۱۵ دانشجو در دسترس نباشد، به صورت زیر عمل می‌کنیم:

اگر عدد، مربوط به خانه شماره ۱۲ که بیان کننده معدل دانشجوی دوازدهم است، باشد و به بیان دیگر داده گمشده باشد باید شماره ۱۲ را به صورت زیر در مربع مورد نظر وارد کرد. به همین ترتیب اگر داده گمشده دیگری داشتیم، شماره های آنها را در مستطیلهای بعدی وارد می کنیم. (برای حداکثر۳ داده گمشده در مستطیل‌های بالایی)



اگر تعداد بیشتری داده گمشده وجود داشته باشد، داده ها را به ترتیب کوچک یا بزرگی تنظیم می کنیم، سپس با فعال کردن قسمت Range plus one optional discrete missing value در مستطیلهای پایینی شماره های داده گمشده را مشخص می کنیم. (از شماره .... تا شماره....) که به طور مثال در زیر نشان داده شده است.

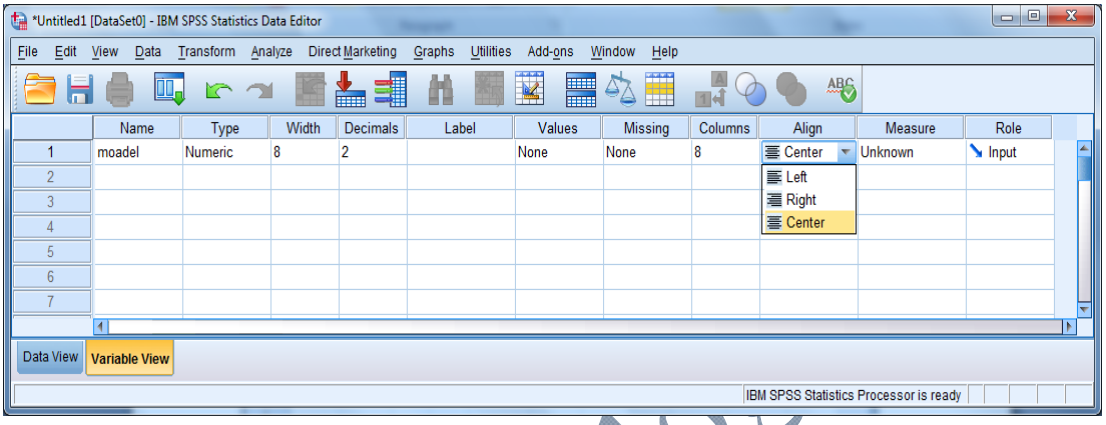


و اگر داده های گمشده به صورتی بودند که یک سری از آنها پشت سر هم و یکی از آنها جدا بود، شماره داده گمشده جدا را در قسمت Discrete value اضافه می کنیم.

اما در این مثال چون داده گمشده ای وجود ندارد ستون Missing بدون تغییر باقی می ماند.

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

تراز کردن داده‌ها در ستون Align قابل انجام شدن می‌باشد.



داده ها به طور پیش فرض در پنجره Data view راست چین هستند ولی زمانی پیش می آید که می خواهیم داده ها چپ چین و یا وسط چین باشند. برای این کار ابتدا مکان نمای ماوس را روی ستون Align قرار داده و گزینه مورد نظر را انتخاب می کنیم.

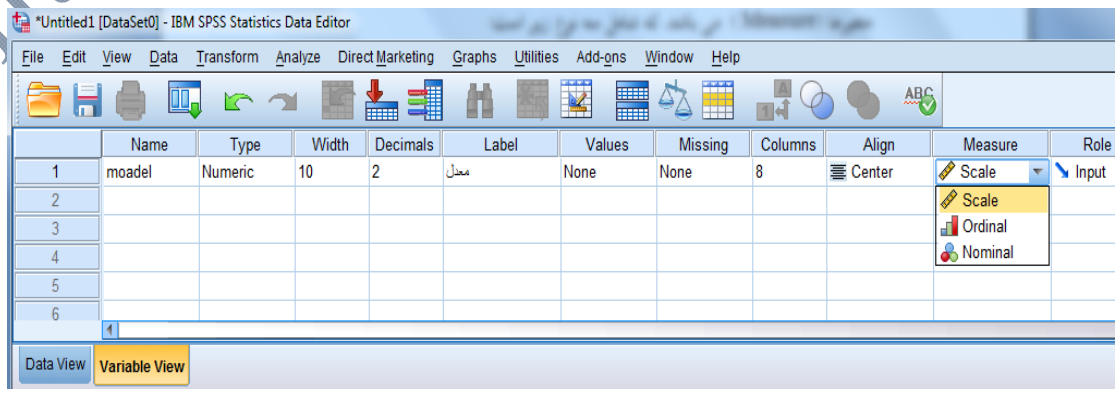
آخرین ستون مورد بررسی در پنجره Variable view مربوط به ستون مقیاس اندازه گیری متغیرها (Measure) می باشد. که شامل سه نوع زیر است:

Scale = داده‌های فاصله‌ای و نسبتی

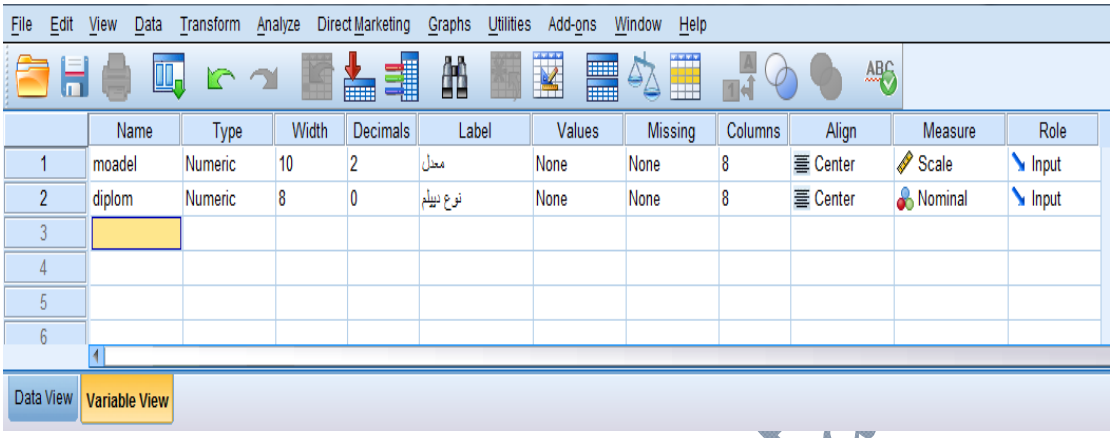
Ordinal = داده‌های رتبه‌ای

Nominal = داده‌های اسمی

داده‌های مربوط به معدل دانشجویان از نوع Scale می‌باشند.

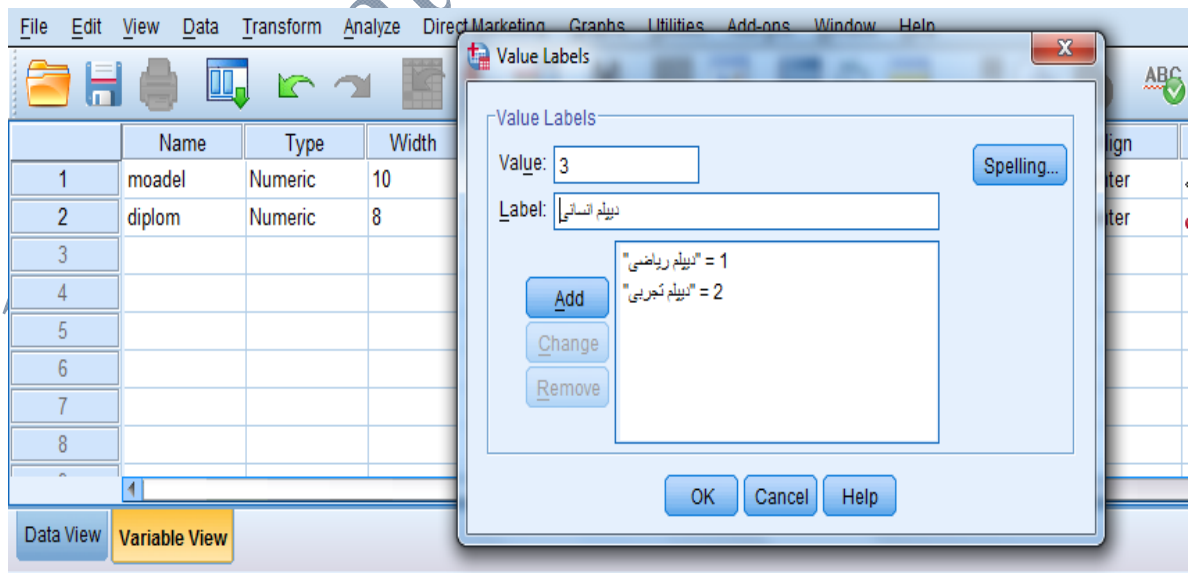


به همین ترتیب متغیر نوع دیپلم را در پنجره Variable view تعریف کرده و متناسب با نوع داده، ستون‌های مختلف را تنظیم می کنیم.

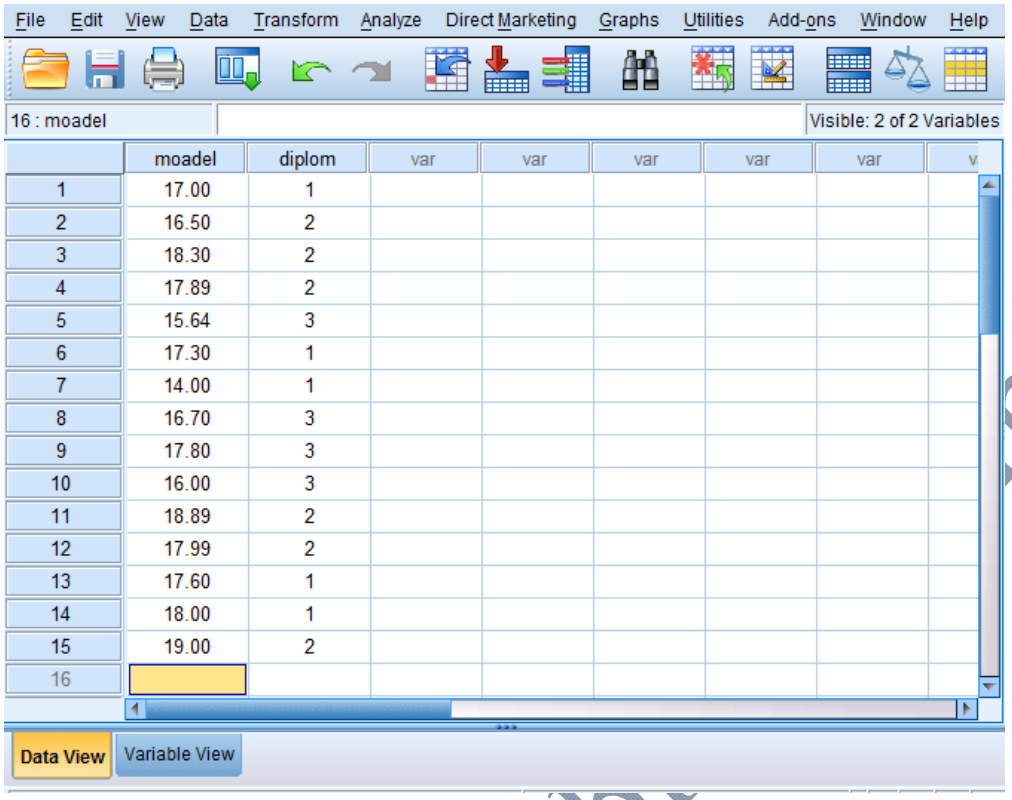


با توجه به اینکه نوع دیپلم، داده اسمی می باشد، در این قسمت می توانیم در ستون Values هر کدام از کدهای دیپلم را با یک برچسب نشان داد.

مانند روش زیر در سطر مربوط به Values ابتدا داده مورد استفاده در پنجره Data view را وارد کرده، سپس در سطر مربوط به Value label اسم مورد نظر را تایپ کرده و سپس روی گزینه Add کلیک می کنیم. بعد از وارد کردن اطلاعات روی گزینه Ok کلیک می کنیم. این کار باعث می شود که در خروجی‌های ما به طور مثال به جای نمایش کد ۱ معادل آن یعنی دیپلم ریاضی مشاهده می شود.



سپس در پنجره ***Data view*** اعداد را به صورت زیر وارد می نماییم.

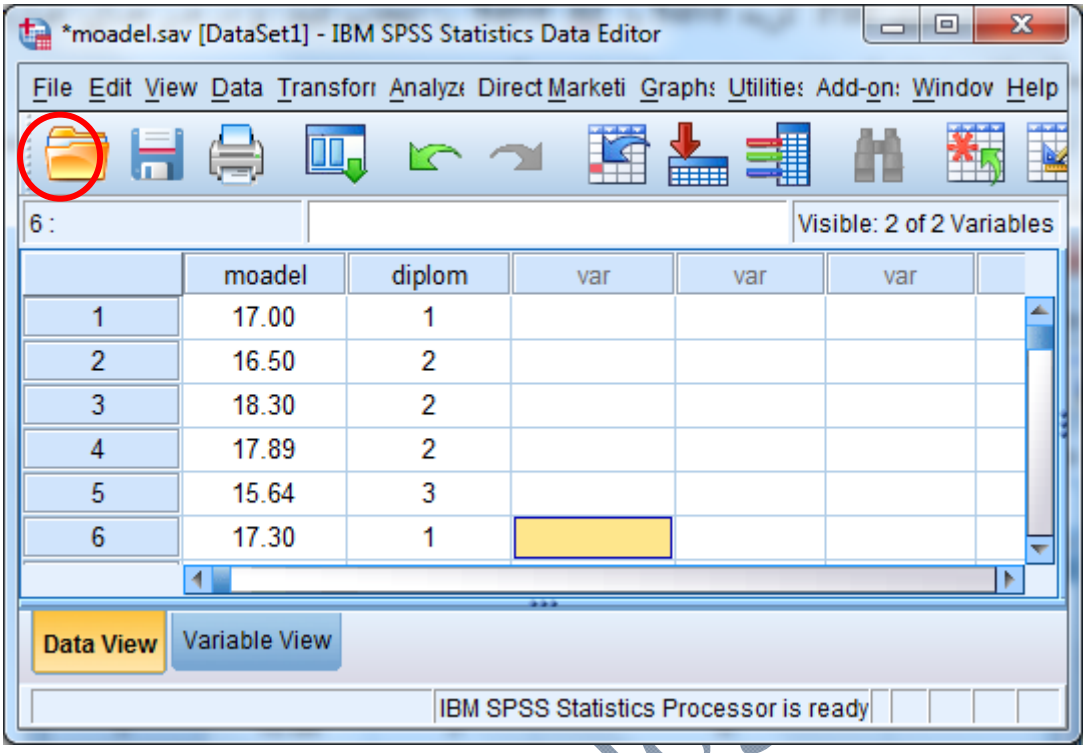


ذخیره کردن و باز کردن فایل در برنامه SPSS

توصیه می شود که بعد از اجرای برنامه، حتماً ابتدا فایل و پروژه‌ی خود را ذخیره کنید و سپس مراحل بعدی تکمیل کارتان را ادامه دهید و هر از گاهی می‌توانید با زدن همزمان دو کلید Ctrl+S ذخیره سازی را آپدیت کنید. چون زمان زیادی صرف وارد کردن داده‌ها می‌شود، باید پرونده داده‌ها را به محض بررسی دقیق، ذخیره کنیم. اگر حجم بزرگی از داده‌ها را وارد کنیم ذخیره سازی هر چند دقیقه یک بار توصیه می‌شود.

در رابطه با ذخیره برنامه وقتی پروژه‌ای را که انجام داده‌اید خواستید بر روی هارد کامپیوترتان ذخیره کنید از منوی File گزینه Save یا Save az را انتخاب کنید و در هر جای هارد دیسک کامپیوترتان که مایل بودید می‌توانید ذخیره کنید.

از آیکون میانبری که در نوار ابزار زیر منوی Edit به شکل یک فلاپی هست(در شکل زیر با دایره قرمز مشخص شده است) هم می‌توانید استفاده کنید.

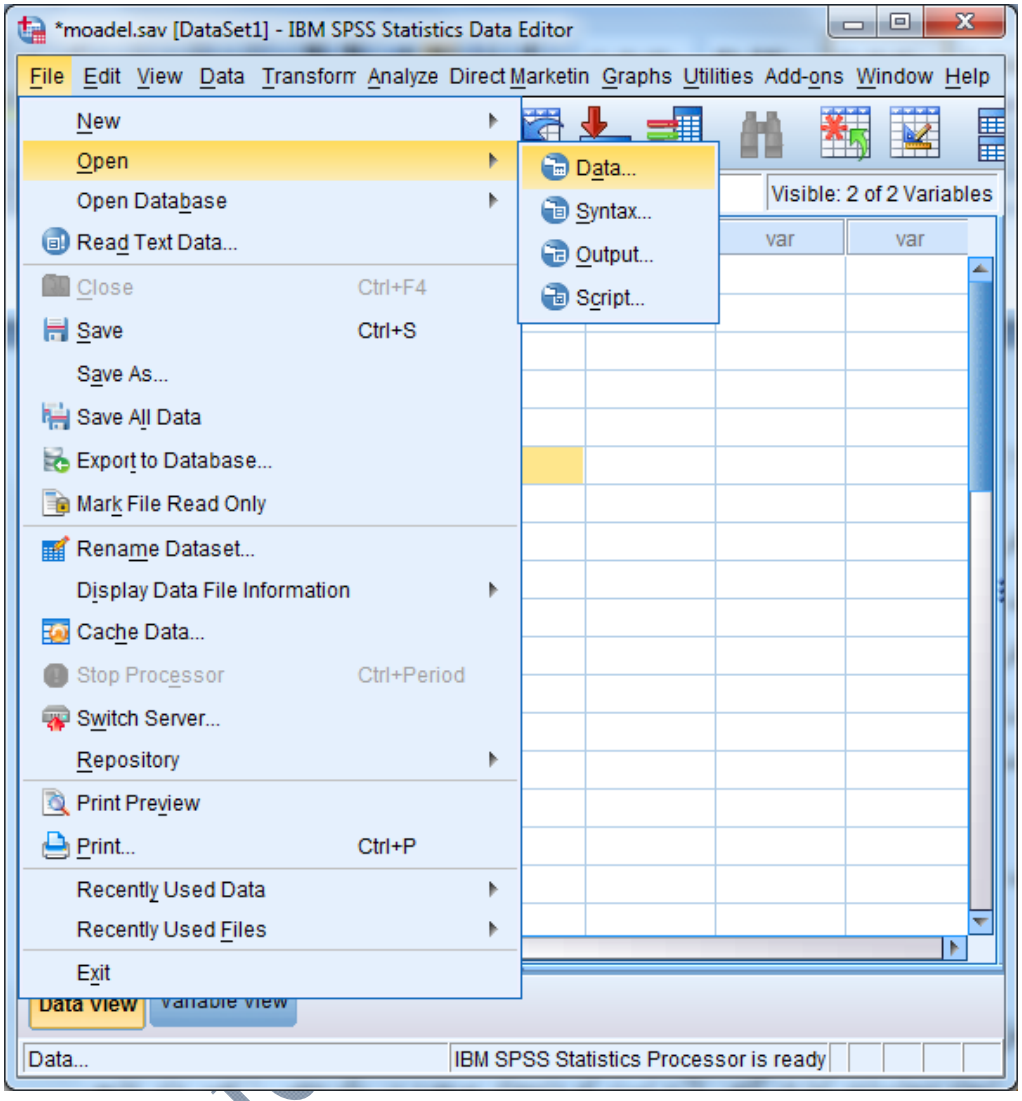


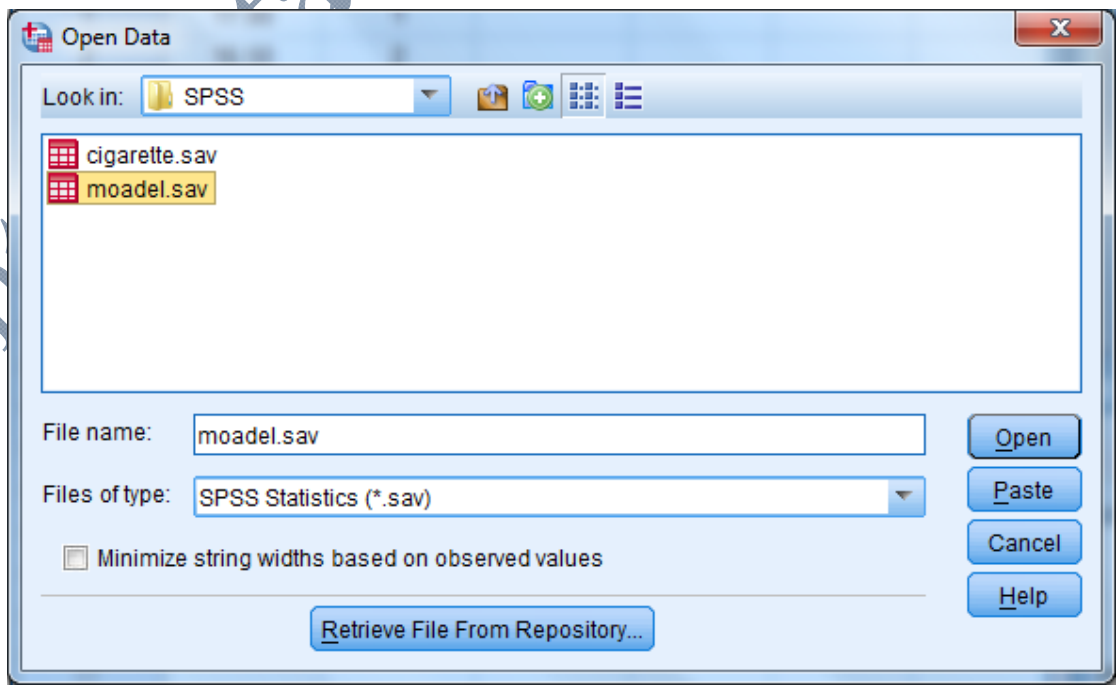
هنگام ذخیره خود سیستم پسوند « sav» را به صورت پیش فرض داده است که همان پسوند برنامه است، اگر خواستید که بعد بتوانید فایل خودتان را بوسیله این برنامه باز کنید با همین پسوند ذخیره کنید اما اگر خواستید که فایلتان را بعد از ذخیره با برنامه دیگری باز کنید در قسمت Save az type پسوندتان را عوض کنید.

تذکر۱: برای ذخیره فایلتان روی دیسکت، ابتدا فایل مورد نظر را روی هارد ذخیره کنید بعد آن را روی فلاپی کپی کنید.

تذکر۲: اگر پروژه ای را که شما انجام داده‌اید با SPSS20 بوده باشد با SPSS نسخه‌های پایین‌تر نمی‌توان باز کرد یعنی SPSS فایلی که با نسخه پایین‌تر از نسخه خودش ذخیره شده باز می‌کند ولی فایل‌هایی که با نسخه بالاتر ذخیره شده باشند را باز نمی‌کند.

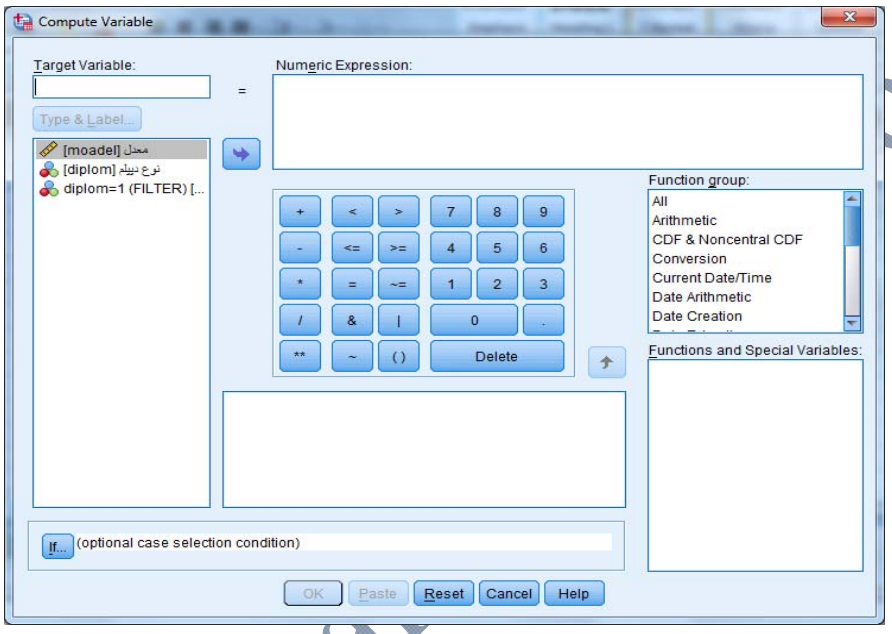
برای باز کردن یک فایل توسط برنامه SPSS بعد از باز شدن برنامه از مسیر File / Open / Data رفته و سپس از محل ذخیر، فایل مورد نظر را یافته و روی دکمه Open کلیک می‌کنیم.


*moadel.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketin Graphs Utilities Add-ons Window Help
New
Open
Open Database
Read Text Data...
Close Ctrl+F4
Save Ctrl+S
Save As...
Save All Data
Export to Database...
Mark File Read Only
Rename Dataset...
Display Data File Information
Cache Data...
Stop Processor Ctrl+Period
Switch Server...
Repository
Print Preview
Print... Ctrl+P
Recently Used Data
Recently Used Files
Exit
Data...
Syntax...
Output...
Script...
Visible: 2 of 2 Variables
var
var
Data View
Variable View
Data...
IBM SPSS Statistics Processor is ready


Open Data
Look in: SPSS
cigarette.sav
moadel.sav
File name: moadel.sav
Files of type: SPSS Statistics (*.sav)
Minimize string widths based on observed values
Retrieve File From Repository...
Open
Paste
Cancel
Help

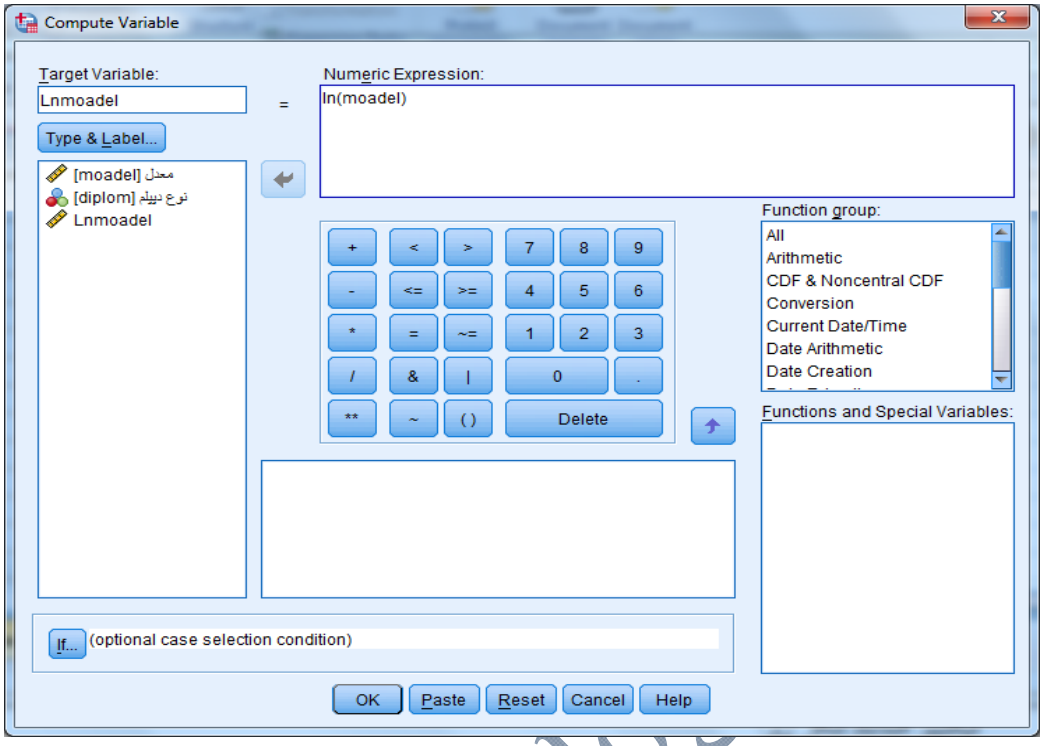
تبدیل و انتخاب داده ها به وسیله دستور Transform…. Compute:

برای انجام محاسبات آماری گاهی لازم است که داده‌های موجود در یک متغیر را تغییر داد. نتایج حاصل از این تغییر را می‌توان در متغیر قبلی جایگزین نمود و یا اینکه در یک متغیر جدید ذخیره نمود. این کار توسط دستور Transform انجام می‌شود. برای انجام این تبدیلات بر روی متغیرها پس از انتخاب مسیر Transform / Compute ، کادر زیر به نمایش درخواهد آمد.

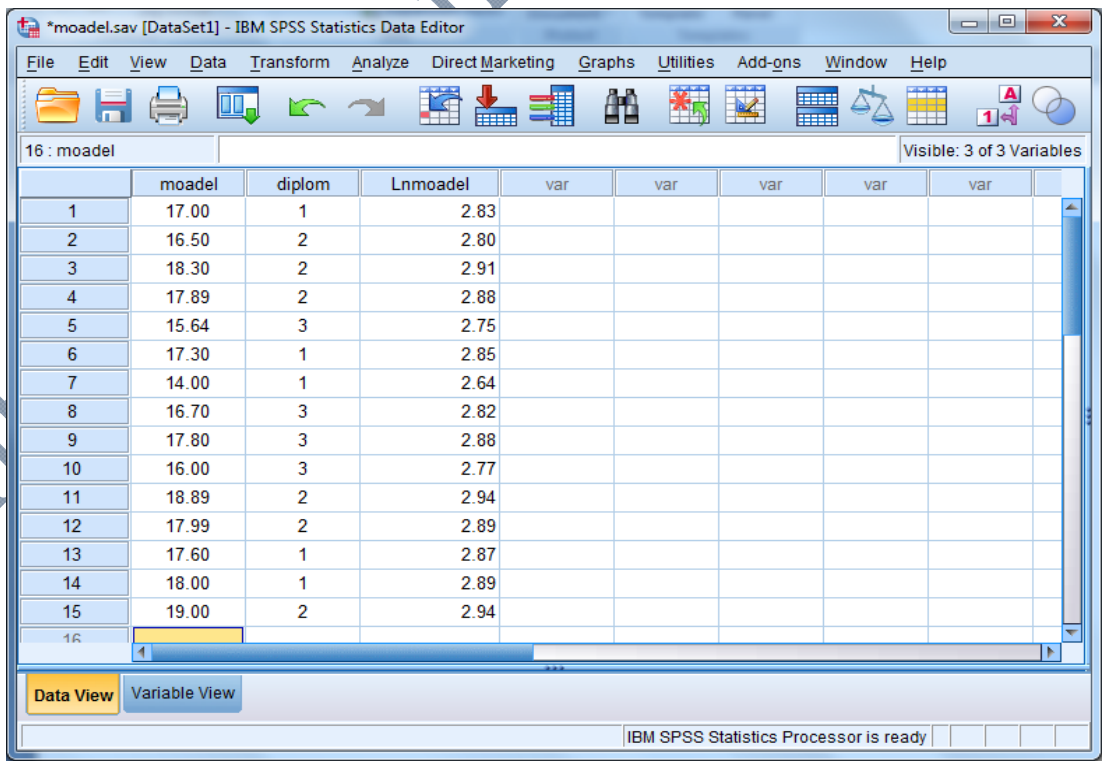


در قسمت Target Variable نام متغیر هدف را نوشته و در قسمت Numeric Expression عملیات ریاضی مورد نظر را وارد خواهیم کرد.

به عنوان مثال در رابطه با مثال معدل دانشجویان فرض کنید که بخواهیم از معدل دانشجویان لگاریتم بگیریم. برای اینکه متغیر جدیدی بسازیم نام متغیر هدف را به صورت Lnmoadel تعریف می‌کنیم. در قسمت Numeric Expression نیز عبارت ln(moadel) را یادداشت میکنیم. همانند شکل زیر:



سپس Ok را انتخاب می‌کنیم. در صفحه Data View ستون جدیدی با نام Lnmoadel به صورت زیر ایجاد خواهد شد.



	moadel	diplom	Lnmoadel	var	var	var	var	var
1	17.00	1	2.83					
2	16.50	2	2.80					
3	18.30	2	2.91					
4	17.89	2	2.88					
5	15.64	3	2.75					
6	17.30	1	2.85					
7	14.00	1	2.64					
8	16.70	3	2.82					
9	17.80	3	2.88					
10	16.00	3	2.77					
11	18.89	2	2.94					
12	17.99	2	2.89					
13	17.60	1	2.87					
14	18.00	1	2.89					
15	19.00	2	2.94					
16								

کدبندی مجدد و گروه‌بندی داده‌ها(*Recod*):

دستور Recod یکی از مفیدترین دستورهایی است که جهت تبدیل یک متغیر گسسته به کار می‌رود. برای مثال ما نمی‌توانیم در یک جدول توافقی، متغیر معدل که پیوسته است را در یک سط بیاوریم و متغیر نوع دیپلم را که گسسته است در ستون قرار دهیم. زیرا تعداد رده‌های موجود در متغیر معدل برابر با تعداد مشاهدات است. در نتیجه یک جدول توافقی بزرگ به دست می‌آید. از این رو متغیر سن را باید گروه بندی کنیم.

گزینه Recod از منوی Transform جهت کدگذاری یک متغیر پیوسته به رده های جدا از هم به کار می‌رود.

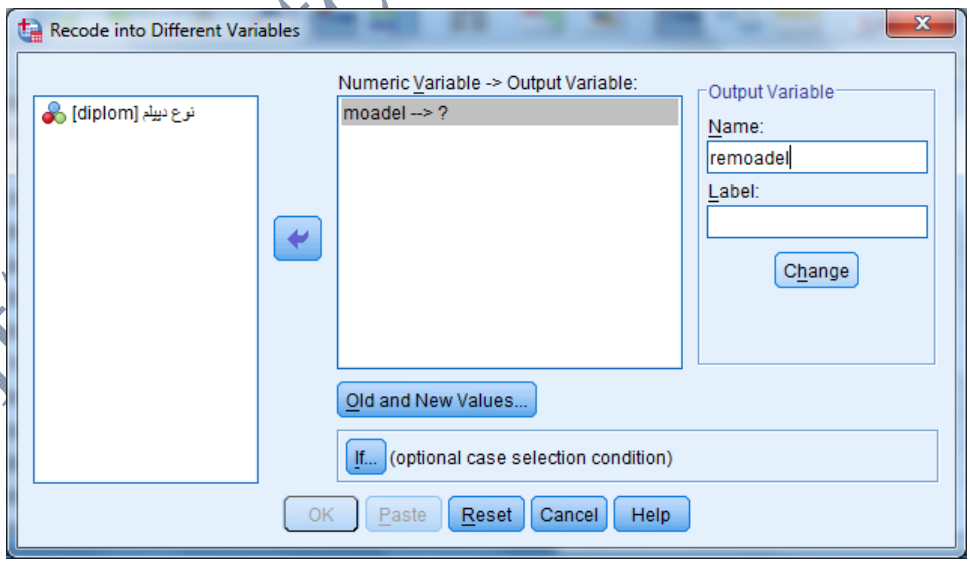
دو دسته Recod در SPSS وجود دارد.

۱- Recod into same variable: مقادیر تبدیل شده متغیر بر روی همان متغیر ذخیره می‌گردد.

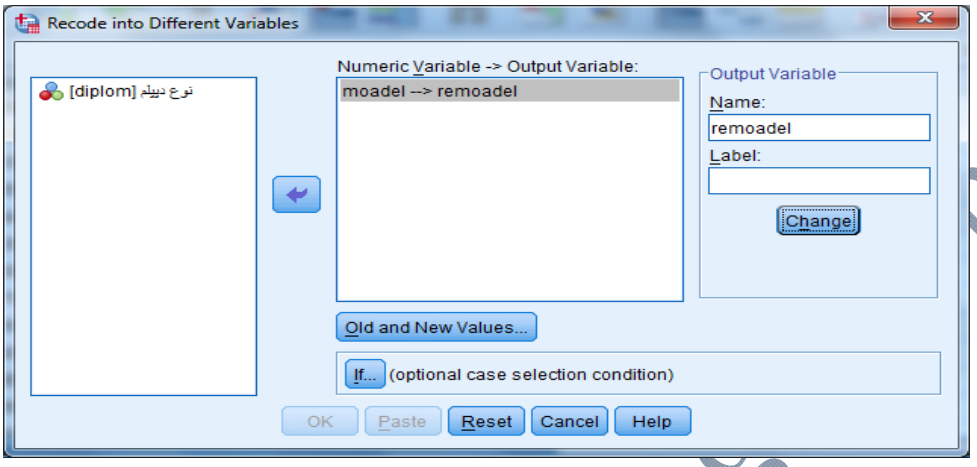
۲- Recod into Different variable: مقادیر تبدیل شده متغیر بر روی متغیر جدید ذخیره می‌گردد.

برای مثال، در مثال معدل دانشجویان فرض کنید که بخواهیم دانشجویان را بر اساس معدل به ٤ گروه(ضعیف(کمتر از ١٤)، متوسط(١٤ تا ١٦)، خوب(١٦ تا ١٨) و عالی(١٨ تا ٢٠)) تقسیم بندی کنیم.

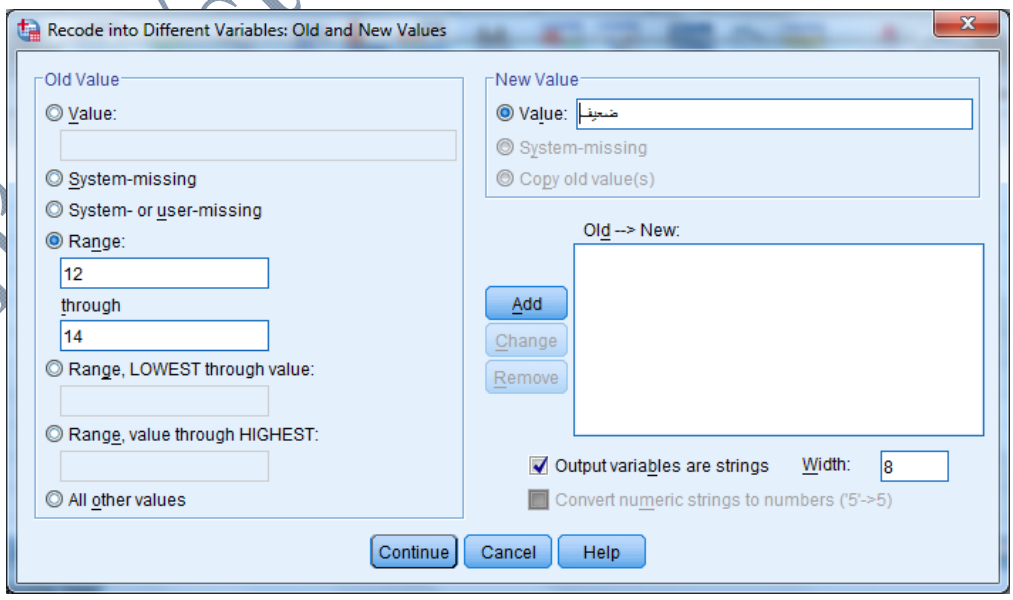
برای این کار با انتخاب مسیر Transform / Recode Inti Different Variables کادر زیر نمایش داده خواهد شد.



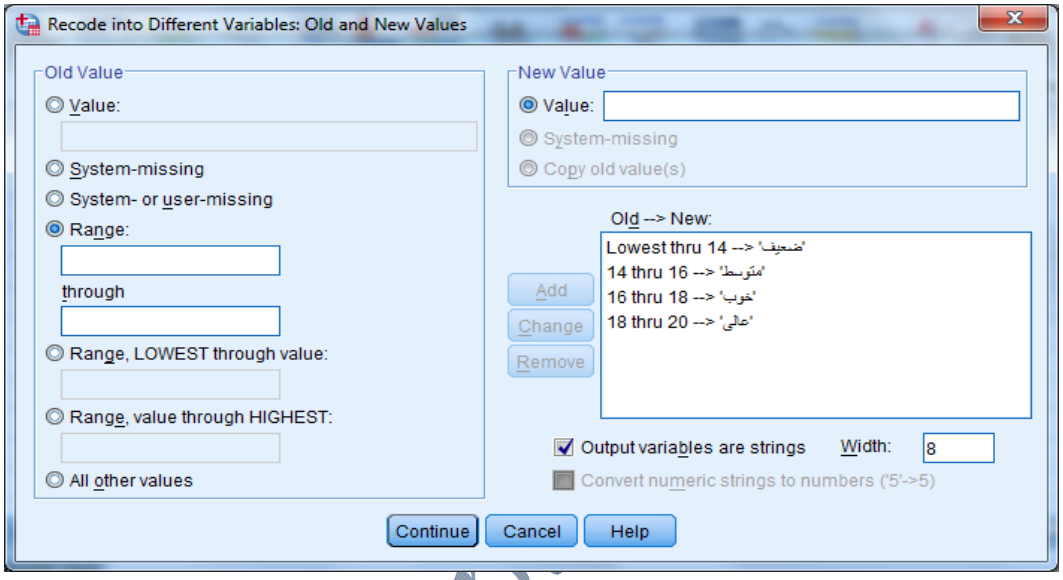
متغیر moadel را انتخاب و با استفاده از نشانگر  آن را به سمت راست منتقل می‌کنیم. سپس در کادر Name یک نام برای متغیر جدید انتخاب می‌کنیم(remoadel). با کلیک بر روی گزینه Change نام جدید به صورت شکل زیر ثبت خواهد شد.



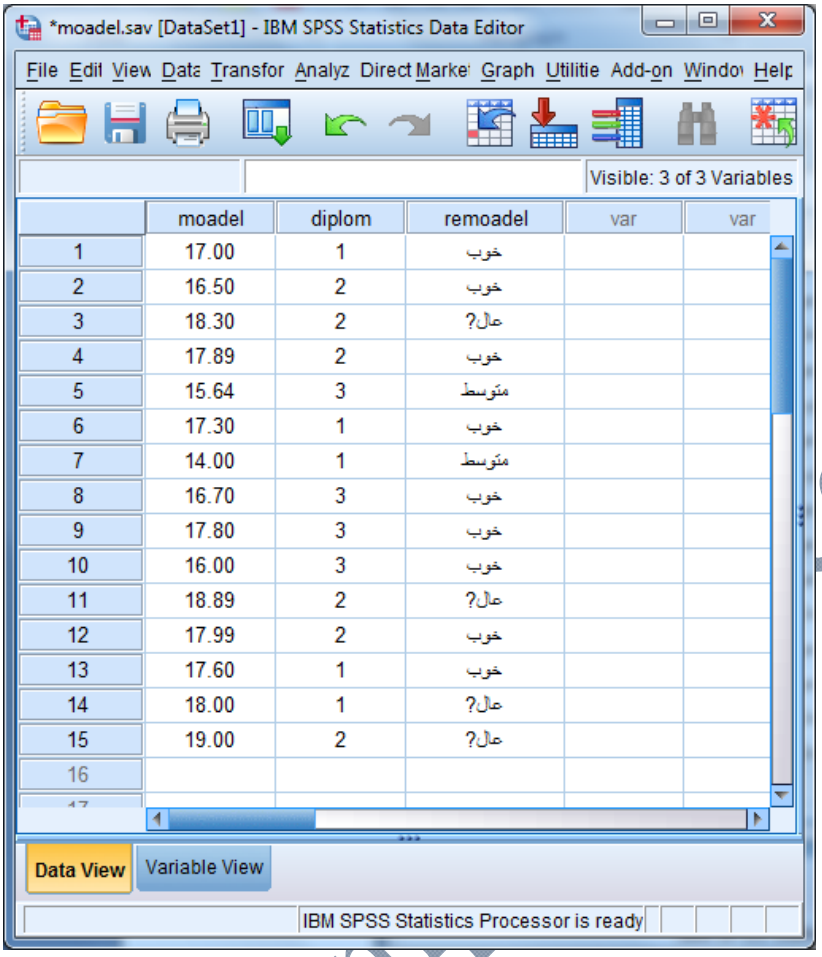
روی گزینه Old and New Values کلیک می‌کنیم تا کادر گفتگوی آن باز شود. چون می‌خواهیم در این مثال متغیر خروجی ما رشته‌ای باشد، در قسمت پایین سمت راست، داخل مربع Output variables are string را تیک می‌زنیم. با این کار نرم افزار متغیر خروجی جدید را به عنوان یک متغیر رشته‌ای می‌شناسد. در بخش Old value بر روی گزینه Range, LOWEST through value کلیک و عدد ۱۴ را تایپ می‌کنیم. سپس در قسمت New Variable کلمه ضعیف را تایپ می‌کنیم و دکمه Add را می‌زنیم.



با این عمل این تغییر را در کادر $Old \rightarrow New$ مشاهده می‌کنیم. سپس بر روی گزینه Range کلیک کرده و در کادر اول عدد ۱۴و در کادر دوم عدد ۱۶ را می‌نویسیم و در کادر New value عبارت متوسط را تایپ می‌کنیم و دکمه Add را می‌زنیم و این کار را برای دو قسمت دیگر نیز تکرار می‌کنیم. همانند شکل زیر:



سپس بر روی گزینه Continue و Ok کلیک میکنیم. متغیر جدیدی با نام remoadel به صورت زیر ایجاد خواهد شد.

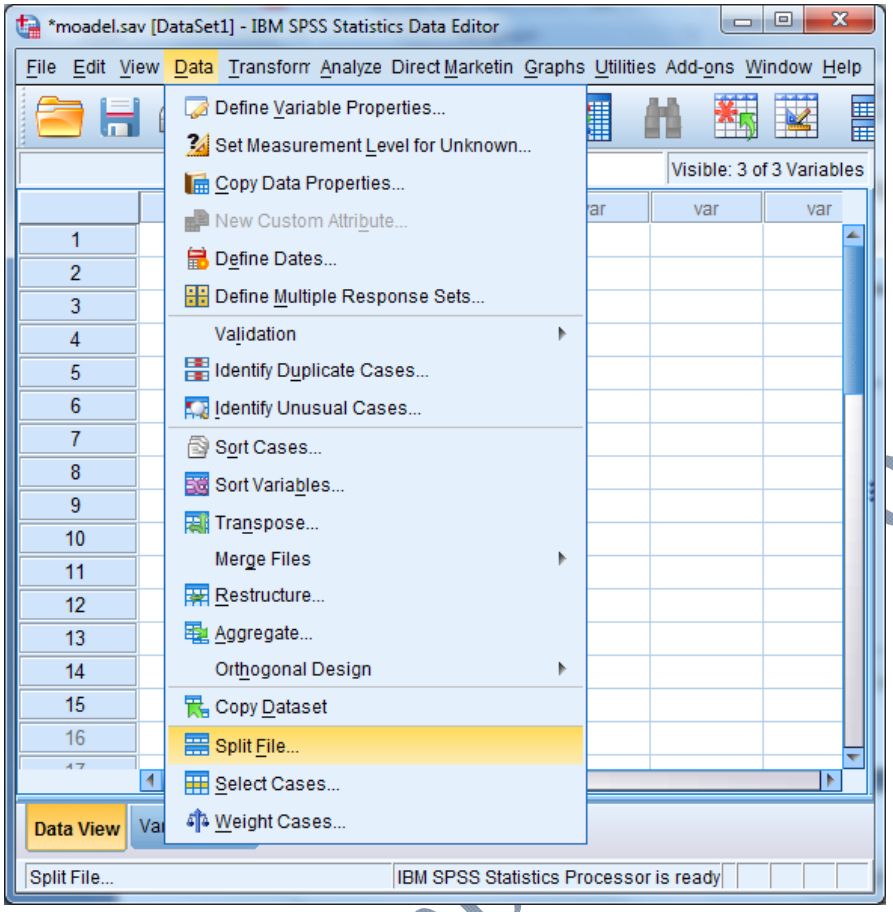


شکستن فایل‌ها با Split File

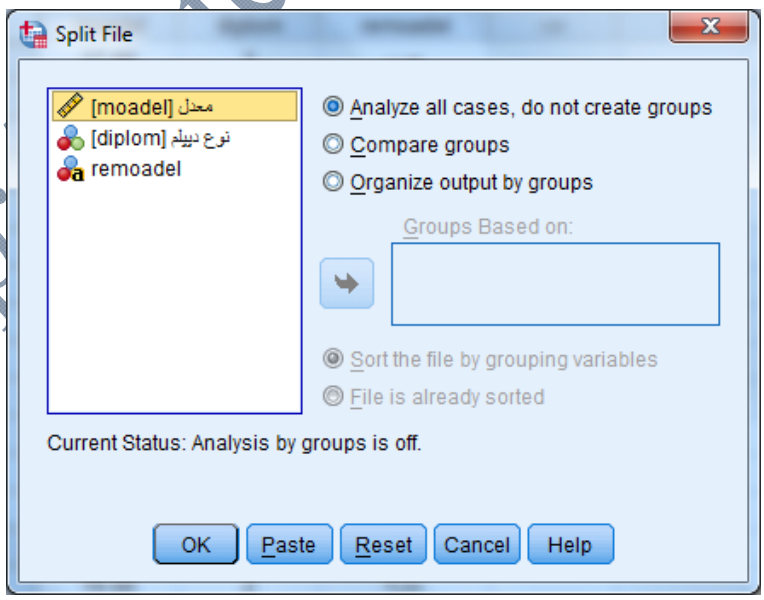
گاهی مناسب است که یک فایل بر اساس سطوح یک متغیر گروه‌بندی(همانند جنسیت یا تحصیلات) شکسته شود تا تجزیه و تحلیل آماری بعد از آن به تفکیک هر یک از سطوح متغیر گروه‌بندی انجام شود.

برای مثال داده‌های مربوط به معدل دانشجویان را می‌توان بر اساس متغیر دیپلم(diplom) شکست.

برای اینکار از مسیر Data / Split file به صورت شکل زیر خواهیم رفت.



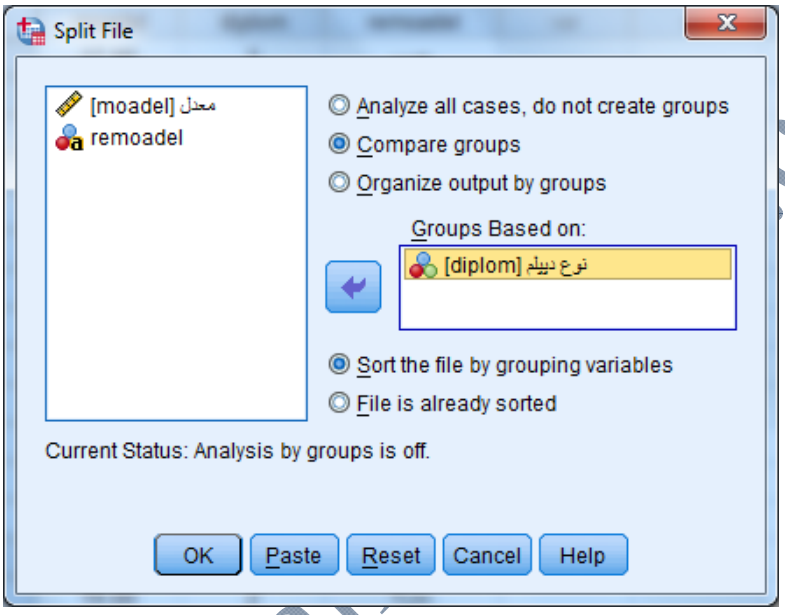
با انتخاب مسیر بالا کادر گفتگویی به صورت زیر مشاهده خواهد شد.



یکی از دو گزینه Compare group و یا Organize output by group را انتخاب می‌کنیم. تفاوت این دو گزینه در این است که با انتخاب Compare group هر یک از جداول خروجی دارای ردیفی برای هر یک

از سطوح متغیر نوع دیپلم است. در حالیکه با انتخاب گزینه Organize output by group خروجی ابتدا شامل تمامی جداول مربوط به سطح اول متغیر نوع دیپلم و سپس به همین ترتیب برای سطوح بعدی است.

در ادامه متغیر گروه‌بندی که قرار است شکستن بر اساس آن صورت گیرد که در اینجا متغیر نوع دیپلم(diplom) است را انتخاب و با استفاده از نشانگر به سمت راست و به کار Groups Based on: به صورت زیر، منتقل می‌کنیم.



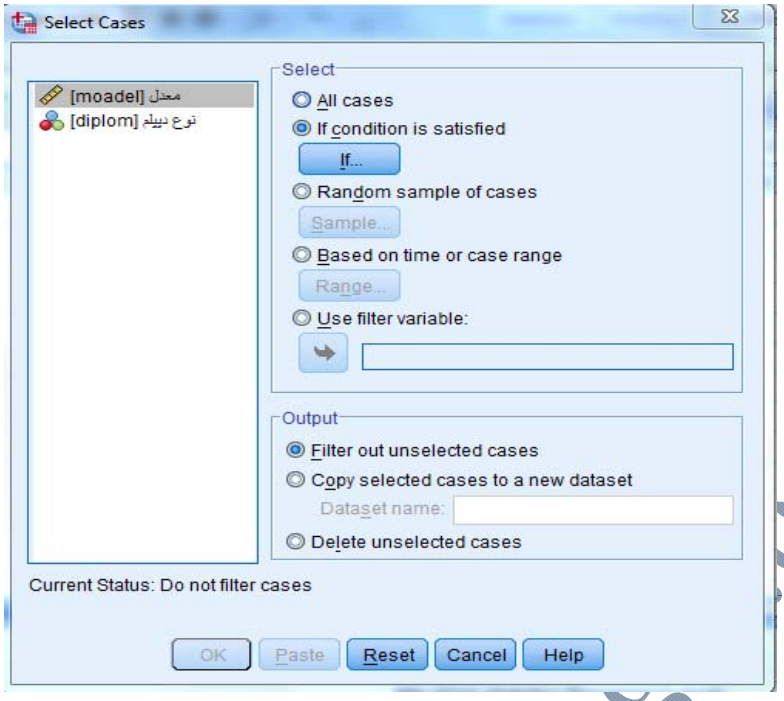
سپس بر روی گزینه Ok کلیک کنید.

اثرات این دستور این است که بعدا وقتی SPSS بر روی داده‌های این فایل تجزیه و تحلیل انجام می‌دهد، آنالیزهای جداگانه‌ای برای هر یک از متغیرهای گروه‌بندی انجام خواهد گرفت.

برای خروج از این حالت و اگر بخواهیم که مجدداً آنالیز بر روی تمامی داده‌ها به صورت یکجا انجام شود، همانند قبل مسیر Data / Split file را انتخاب و این‌بار گزینه Analyze all cases, do not create groups را انتخاب می‌کنیم.

انتخاب نمونه به وسیله دستور Select Cases

انتخاب نمونه این امکان را به ما می‌دهد که تحلیل را بر روی گروه مشخصی از مشاهدات محدود کنیم. با اجرای دستور Select Cases از منوی Data کادر گفتوی زیر باز می‌شود.



بر اساس شکل بالا چندین روش برای انتخاب نمونه وجود دارد که در ادامه به توضیح آنها خواهیم پرداخت:

All Cases: تمام مشاهدات را برای تجزیه و تحلیل انتخاب می‌کند. این گزینه پیش فرض است.

If Condition is satisfied: این امکان را می‌دهد که مشاهدات را بر اساس یک شرط منطقی با استفاده از زیر کادر if انتخاب کنیم.

Random Sample of Cases: این امکان را می‌دهد که یک نمونه تصادفی از مشاهدات را انتخاب کنیم.

Based on time or case range: این امکان را می‌دهد که دامنه‌ای از مشاهدات را بر اساس ترتیب‌شان در کاربرگ نمایش داده‌ها انتخاب کنیم.

Use filter variable: با مشخص کردن یک فیلتر این امکان را می‌دهد که مشاهداتی را انتخاب کنید که دارای مقدار صفر برای متغیر فیلتر نیستند.(متغیر فیلتر متغیری است که فقط دارای دو مقدار صفر و یک باشد.)

بیشترین کاربرد مربوط به موارد If Condition is satisfied و Random Sample of Cases است.

بخش Output دارای سه قسمت است.

Filter out unselected cases: با انتخاب این گزینه مشاهدات به طور موقت انتخاب می‌شوند و مشاهدات انتخاب نشده از بین نرفته و امکان دستیابی مجدد به آنها وجود دارد.

Copy selected cases to a new dataset: با انتخاب این گزینه مشاهدات انتخاب شده در یک پنجره جدید از ویرایشگر داده‌ها قرار می‌گیرند تا تحلیل‌گر تنها نمونه‌ی انتخاب شده را با یک نام جدید به طور کامل مشاهده کند.

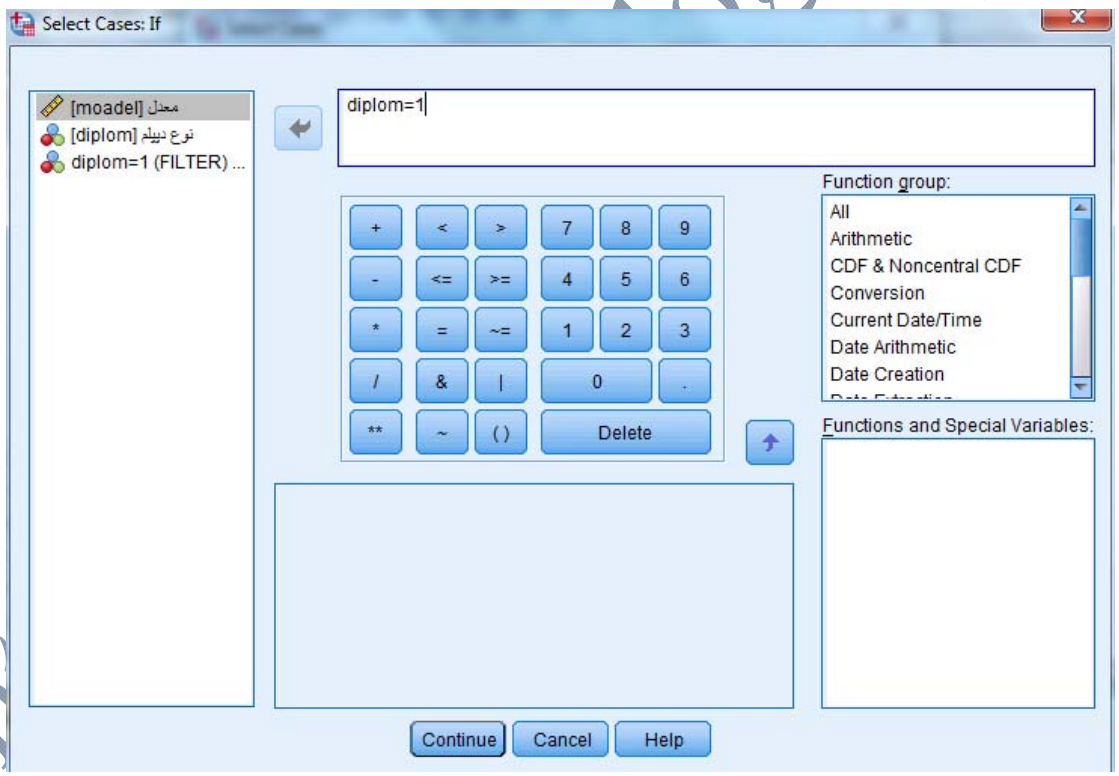
Delete unselected cases: با انتخاب این گزینه مشاهدات به طور دائم انتخاب می‌شوند. مشاهدات انتخاب نشده حذف و امکان دسترسی مجدد به آنها وجود ندارد.

به عنوان مثال در مورد مثال معدل دانشجویان می‌خواهیم تجزیه و تحلیل را بر روی دانش‌آموزانی انجام دهیم که دارای دیپلم ریاضی باشند(کد ۱). بدین منظور از مسیر زیر استفاده خواهیم کرد:

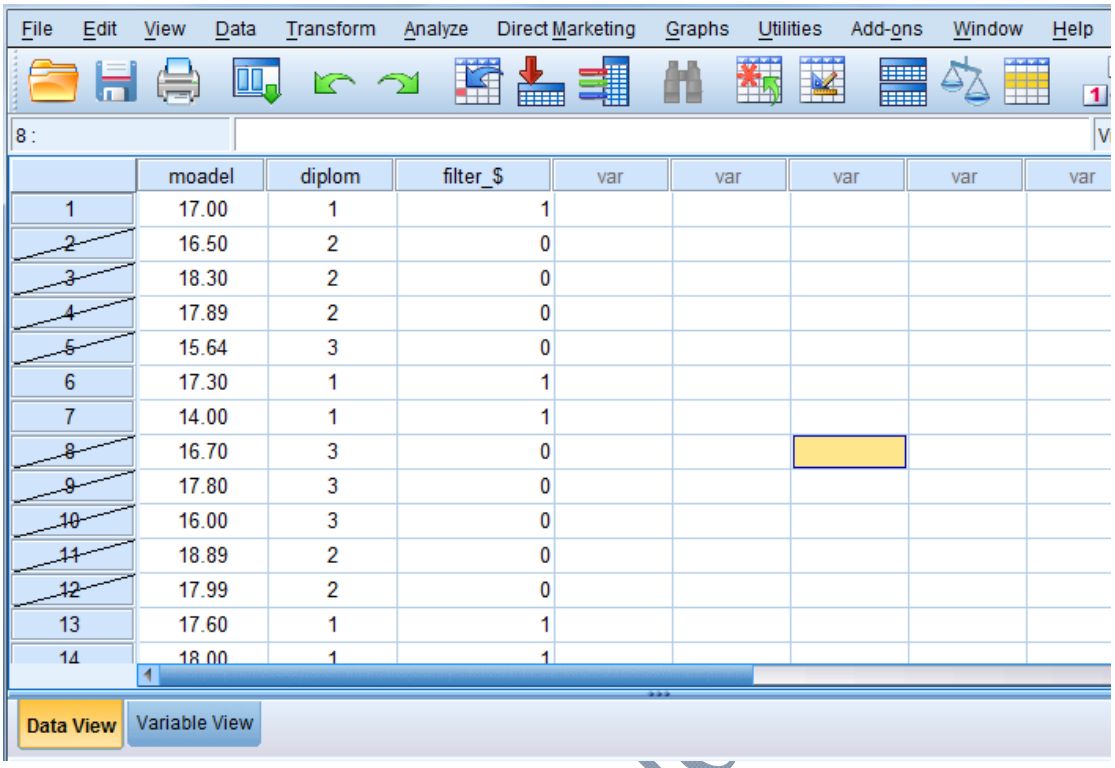
Data / Select cases / If condition is satisfied

عبارت diplom=1 را در کادر تایپ می‌کنیم.

در بخش Output گزینه Filter out unselected cases را انتخاب می‌کنیم. با این کار مشاهدات انتخاب نشده در فایل موجود به صورت فعال باقی می‌مانند و می‌توانیم تمامی مشاهدات اولیه را مجددا به دست آوریم.



سپس بر روی گزینه های Continue / Ok کلیک می‌کنیم. اکنون مشاهدات انتخاب نشده با علامت (/) بر روی ردیفشان به صورت زیر مشخص شده‌اند.



برای غیر فعال کردن انتخاب نمونه‌ها دوباره می‌توان مسیر بالا را رفته و این بار گزینه Select All Cases را انتخاب کنیم.

وزن دادن متغیرها بر حسب فراوانی‌های‌شان در spss:

فرض کنید که از پنجاه زن و پنجاه مرد در مورد مصرف یا عدم مصرف سیگار سوال می‌شود. پاسخ‌های آنهارا می‌توان به صورت یک جدول توافقی(Cross tabs) خلاصه کرد.

هدف از تشکیل یک جدول توافقی نشان دادن هرنوع رابطه‌ای است که ممکن است بین دو متغیر کیفی یا اسمی وجود داشته باشد.

مثال(۲): در مثال حاضر متغیرهای کیفی، جنس (با سطوح مرد و زن) و مصرف سیگار(با سطوح بله و خیر) است. از روی جدول مشخص است که در حقیقت رابطه‌ای بین این دو متغیر وجود دارد. به طوری که واضح است، نسبت بیشتری از پاسخ دهندگان مرد، سیگار مصرف می‌کنند. داده‌های مربوط به این مثال در جدول زیر داده شده است.

جنسیت	مصرف سیگار	
	بلی	خیر
مرد	۴۰	۱۰
زن	۲۰	۳۰

یک جدول توافقی نیز باید به گونه‌ای تغییر داده شود که در جهت ورود به SPSS مناسب شود. SPSS انتظار دارد که مجموعه داده‌ها به شکل ماتریسی مرتب شوند که ردیف‌های آن شرکت کنندگان باشند و ستون‌های آن تشکیل دهنده متغیرها باشد. روشن است که نحوه قرارگیری داده‌ها در جدول فوق به این شکل نیست و دو ستون حاوی مقادیر یک متغیره بوده و یک ردیف حاوی پاسخ‌های شرکت کنندگان متعددی است.

یک راه حل این است که متغیرها را به صورت ستون‌هایی از کدها در آوریم. با این کار این نیاز در ویرایشگر داده‌ها که هر ستون، مربوط به یک متغیر باشد، پاسخ داده می شود .اما در مثال حاضر باید برای SPSS روشن سازیم که داده‌های هریک از خانه ها نشان دهنده پاسخ یک فرد نیست بلکه مجموع پاسخ چندین نفر است. برای این کار، فراوانی های خانه‌ها را به ستون سومی در ویرایشگر داده‌ها منتقل کرده و دستور Cases Weight را اجرا می‌نماییم.

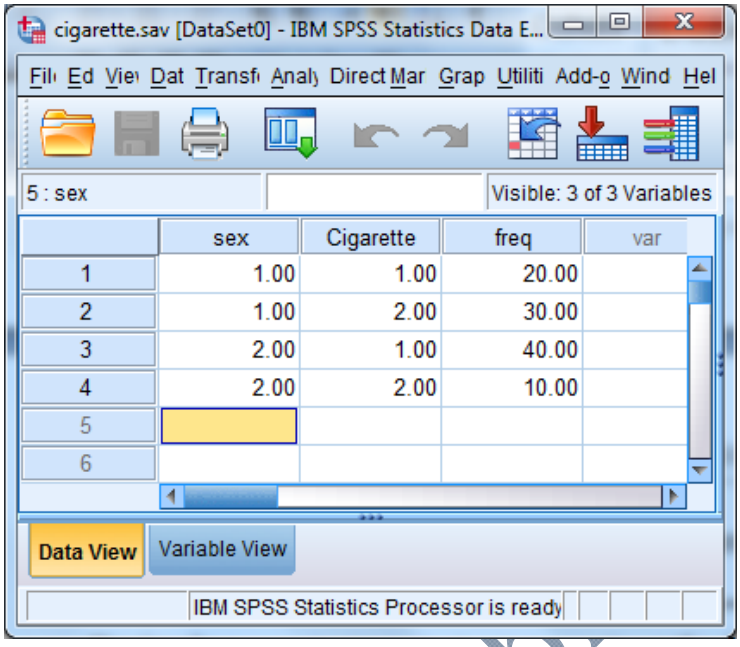
کدهای متغیر sex با مقادیر ۱(زن) و ۲(مرد) و برچسب‌های Female و Male در ستون Values از Variable View تعریف می‌شود و به همین ترتیب کدهای متغیر Cigarette با مقادیر ۱(مصرف سیگار)و ۲ (عدم مصرف سیگار) و برچسب‌های Yes و No مشخص می‌شود. ستون سوم را با عنوان freq تشکیل داده و فراوانی‌های مربوطه را در آن قرار می‌دهیم.

سپس به SPSS دستور خواهیم داد که به هر یک از ترکیب‌های کدهای فوق، براساس فراوانی‌های متغیر سوم(freq) وزن دهد که برای این کار از مسیر Data / Weight Cases استفاده می‌کنیم.

برای وارد کردن مجموعه داده‌های جدول فوق به ویرایشگر داده‌ها در Variable View: مراحل زیر را دنبال کنید:

- در ستون Values، متغیر اول را با نام sex نامگذاری کرده و مقادیر و برچسب‌های مقادیر آن را وارد نمایید. تعداد رقم‌های اعشار را در ستون Decimals به صفر تبدیل کنید.
- متغیر دوم را Cigarette نامگذاری کنید و تعداد رقم‌های اعشار را در ستون Decimals صفر کرده و برچسب‌های مقادیر آن را در ستون value وارد نمایید.
- متغیر سوم را freq نامگذاری کرده و تعداد رقم‌های اعشار را در ستون Decimals صفر کنید. این متغیر حاوی فراوانی‌های موردها در جدول توافقی است.

- بر روی دکمه Data View کلیک کرده و مقادیر مناسب و فراوانی‌ها را همانند شکل زیر وارد نمایید.

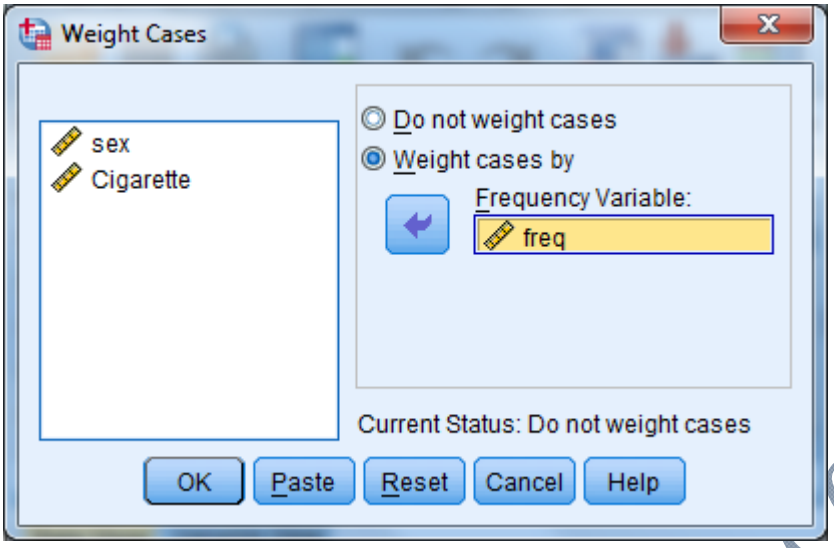


در این حالت صفحه Data View نشان می‌دهد که داده‌ها دارای دو ردیف مرد و دو ردیف زن بوده که هر کدام دارای دو حالت مصرف سیگار و عدم مصرف سیگار است و در نهایت فراوانی‌های ۲۰، ۳۰، ۴۰ و ۱۰ مشاهده می‌شود. در این جدول مشخص است که ۲۰ مورد زن سیگاری، ۳۰ زن غیر سیگاری، ۴۰ مرد سیگاری و ۱۰ مرد غیر سیگاری وجود دارد. حال SPSS را چگونه متوجه این مطلب کنیم؟

پاسخ این سؤال، استفاده از دستور Weight Cases است که یک متغیر را به عنوان ستونی از فراوانی‌ها (به جای ستون مقادیر) تعریف می‌کند.

برای انجام این کار مراحل زیر را دنبال کنید:

- از منوی Data گزینه Weight Cases را انتخاب کنید.
- تا کادر مکالمه Weight Cases باز شود.
- بر روی گزینه Weight Cases by کلیک کنید.
- متغیر freq را انتخاب کنید . با این کار دکمه پیکان پررنگ شده که با کلیک کردن بر روی آن، متغیر freq به چهار گوش متنی Frequency Variable منتقل می‌شود.
- بر روی OK کلیک کنید تا دستور اجرا شود.

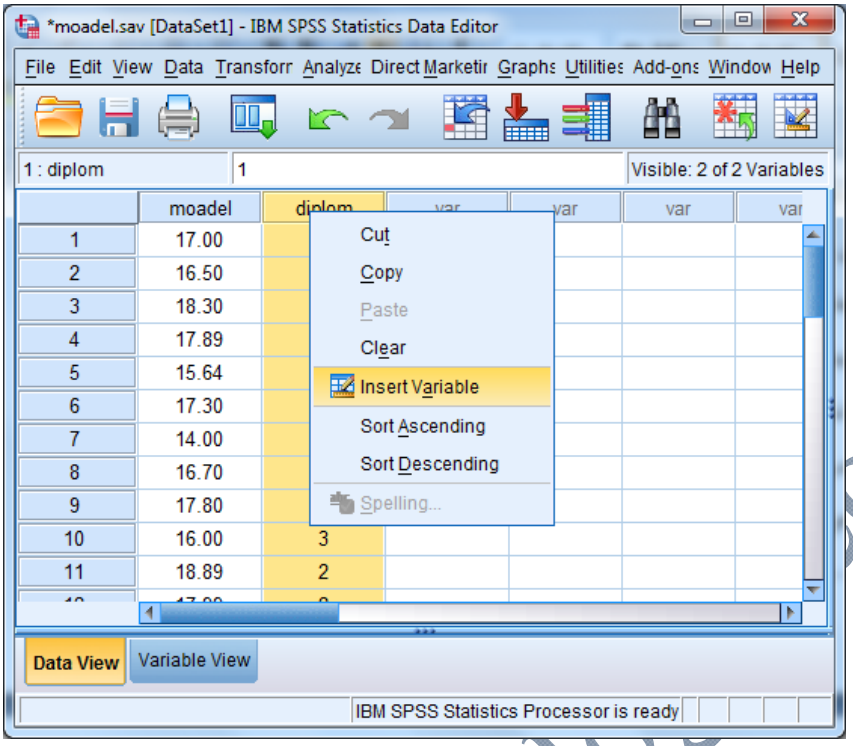


برای خارج شدن از این حالت دوباره مسیر بالا را رفته و این بار گزینه Do nit weight cases را انتخاب کنید.

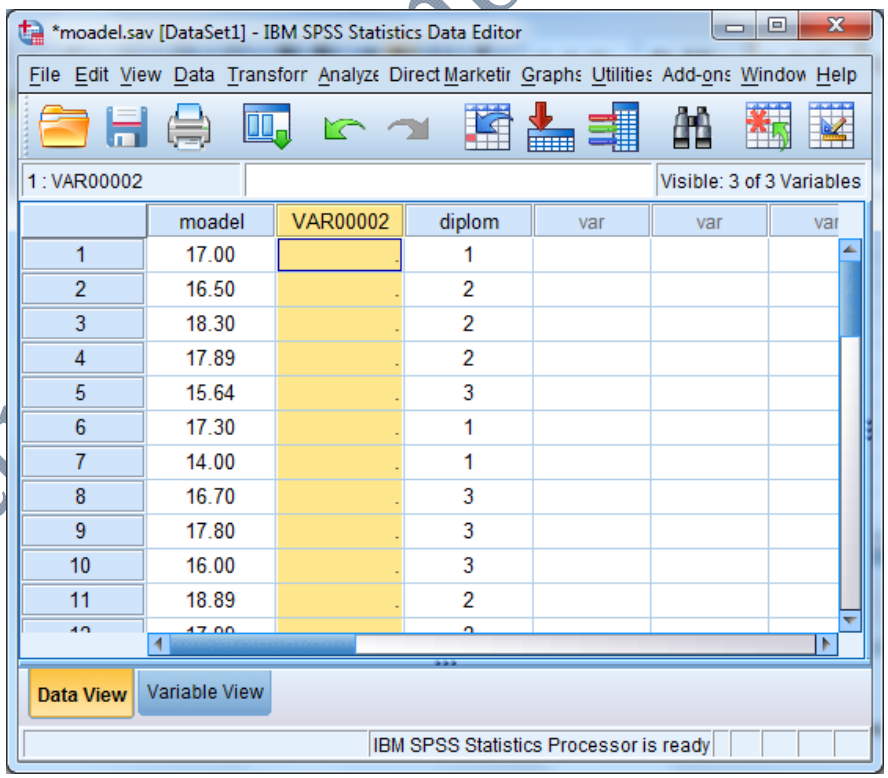
اضافه کردن متغیرها و موردها و تغییر ترتیب متغیرها:

اضافه کردن متغیر(ستون جدید) و مورد(سطر جدید) همیشه امکان‌پذیر است و می‌توان آنها را به ترتیب در سمت راست و پایین داده‌های موجود اضافه کرد. اما گاهی نیاز است که یک متغیر جدید را در کنار یک متغیر موجود و یک مورد جدید را در کنار موردهای قبلی ایجاد کرد. به همین ترتیب گاهی تغییر دادن محل قرارگیری متغیرها به منظور قرارگیری متغیرهای خاص کنار متغیرهای دیگر مفید است.

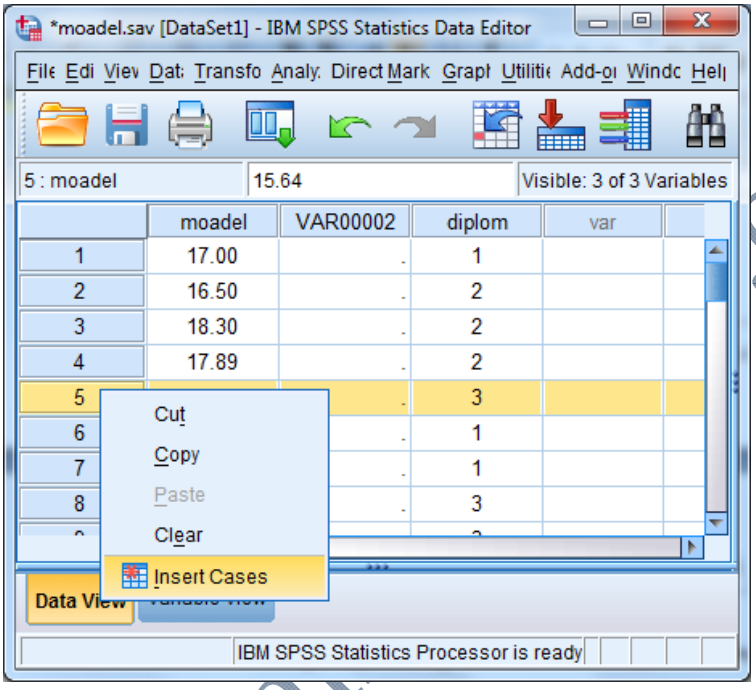
برای ایجاد یک متغیر جدید در یک محل مشخص کافی است که درسربرگ متغیرسمت راست آن کلیک کنید. با این کار کل ستون متغیر با رنگی متفاوت مشخص می‌شود. سپس بر روی نام متغیر انتخاب شده راست کلیک کرده و گزینه Insert variable را انتخاب می‌کنیم. با این کار یک ستون جدید در سمت راست متغیر انتخاب شده ایجاد می‌شود. برای نمونه فرض کنید در داده‌های معدل بخواهیم یک متغیر جدید بین دو متغیر معدل(moadel) و نوع دیپلم(diplom) ایجاد کنیم.



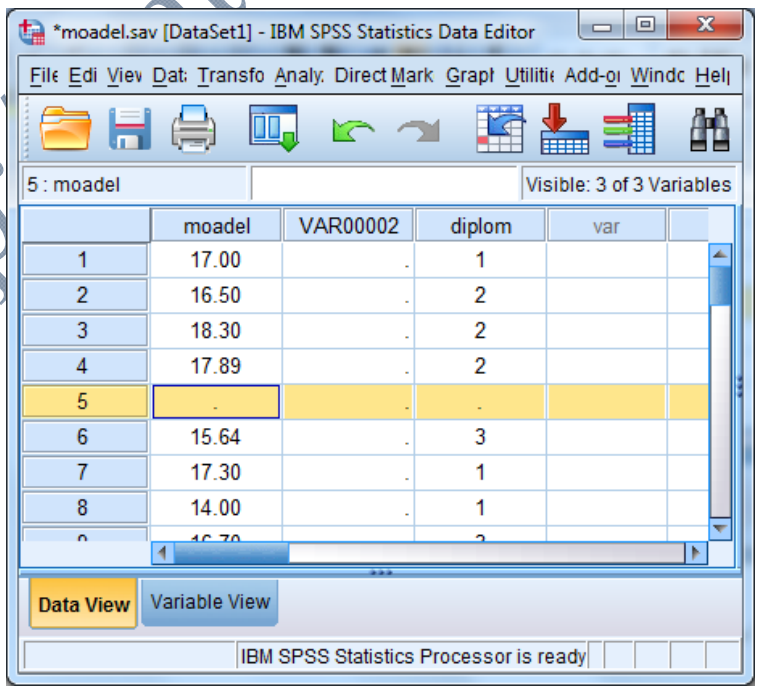
پس از ایجاد متغیر جدید صفحه Data View به صورت زیر در می‌آید.



و به همین ترتیب برای ایجاد یک مورد جدید در یک محل مشخص کافی است که در سمت چپ مورد کلیک کنید. با این کار کل سطر با رنگی متفاوت مشخص می‌شود. سپس بر روی شماره سطر انتخاب شده راست کلیک کرده و گزینه Insert Cases را انتخاب می‌کنیم. با این کار یک سطر جدید در بالای سطر انتخاب شده ایجاد می‌شود. برای نمونه فرض کنید در داده‌های معدل بخواهیم یک سطر جدید بین دو متغیر سطر ٤ و ٥ ایجاد کنیم.



پس از ایجاد مورد(سطر) جدید صفحه Data View به صورت زیر در می‌آید.



برای تغییر در ترتیب متغیرها، فرض کنید که بخواهید یک متغیر مشخص را از جای موقعیت فعلی حذف و در یک موقعیت جدید قرار دهید. برای این کار ابتدا ستون متغیر را همانند قبل انتخاب. سپس با راست کلیک بر روی آن متغیر را Cut کرده و با استفاده از روش بالا در محل مورد نظر یک متغیر جدید ایجاد و متغیر Cut شده را در محل مورد نظر ایجاد شده Paste می‌کنیم.

فصل دوم:

تحلیل توصیفی داده‌ها

در اصطلاح عامیانه آمار به معنای ثبت و نمایش اطلاعات عددی در مورد یک موضوع است. به عنوان مثال ثبت و نمایش تعداد بیکاران، تعداد تصادفات رانندگی، میزان محصولات کشاورزی، میزان صدور نفت، جمعیت شهر تهران و غیره. ولی علم آمار امروزه دارای مفهومی بسیار وسیعتر از این کاربرد عامیانه است. مفاهیم عامیانه آمار زیر مجموعه‌ای از آمار مصطلح بین آماردانان است. از نقطه نظر علمی، آمار به مجموعه روش‌هایی برای جمع آوری تنظیم و خلاصه کردن داده‌های عددی و غیر عددی و انجام استنباط و نتیجه-گیری بوسیله تجزیه و تحلیل آنها، اطلاق می‌شود.

با بیان دیگر می توان گفت که؛

آمار عبارت است از هنر و علم جمع‌اوری، تعبیر و تجزیه و تحلیل داده‌ها و استخراج تعمیم‌های منطقی در مورد پدیده‌های تحت بررسی.

با توجه به تعاریف بالا می توان گفت یک فرآیند تحلیل آماری شامل دو بخش عمده است. اولین قدم نمایش دادن و خلاصه کردن داده‌ها است تا توجه ما روی ویژگی‌های مهم داده ها متمرکز شود و جزئیات غیر ضروری کنار گذاشته شود. اما بخش دوم برای استخراج نکات کلی و استنباط هایی در مورد پدیده تحت مطالعه به کار می‌رود. بخش اول شامل روش‌های آمار توصیفی و بخش دوم در برگیرنده روش‌های موسوم به آمار استنباطی است.

آمار توصیفی

آمار توصیفی به آن دسته از روش‌های آماری گفته می‌شود که به پژوهشگر در طبقه‌بندی، خلاصه کردن، توصیف و تفسیر و برقراری ارتباط از طریق اطلاعات جمع‌آوری شده کمک می‌کند. مراحل اساسی توصیف داده‌ها عبارتست از:

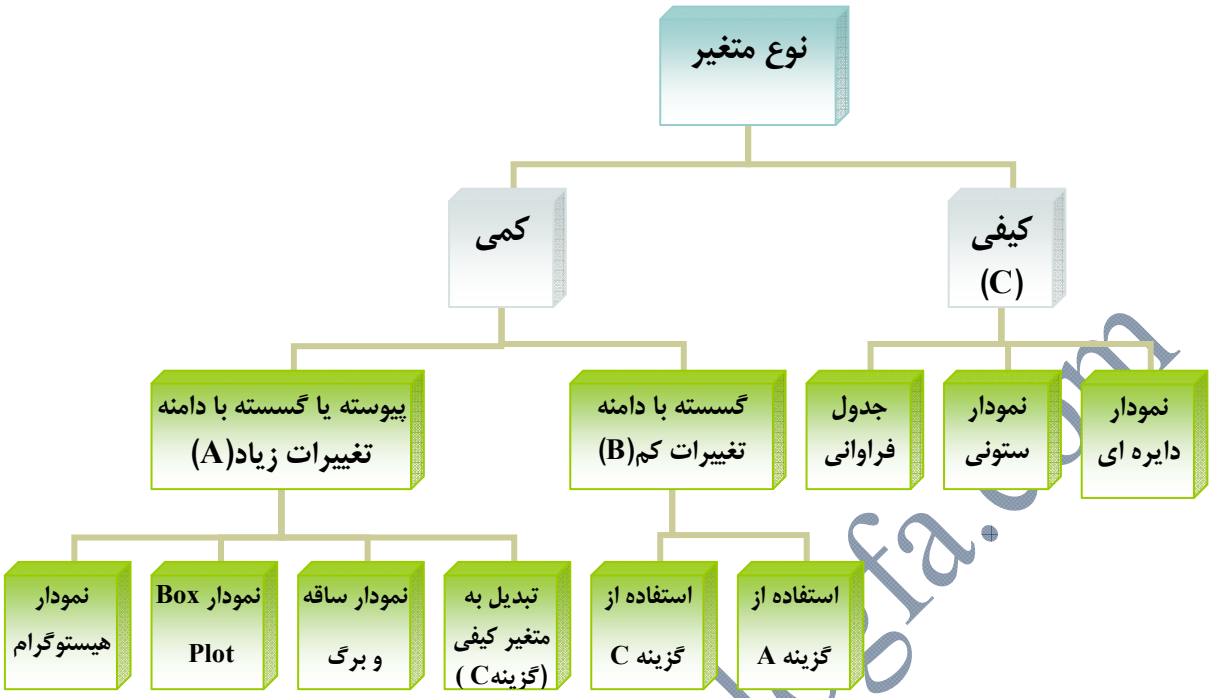
الف) خلاصه کردن و توصیف الگوی کلی

۱- فشرده کردن داده‌ها در قالب جدول‌های اماری

۲- نمایش آنها به وسیله نمودار

ب) محاسبه شاخص‌های آمار توصیفی

نقش آمار توصیفی در فرآیند تحلیل آماری بسیار مهم و حیاتی است. آمار توصیفی با خلاصه کردن داده‌ها، ویژگی‌های مهم آن را نمایان می‌سازد تا ایده‌های لازم را در ذهن پژوهشگر برای مرحله دوم تحلیل آماری(آمار استنباطی) ایجاد کند. برای استفاده از مراحل مختلف آمار توصیفی می‌توان از چارت زیر بهره بگیرید:



اینک مراحل مختلف آمار توصیفی را یک به یک و به طور مفصل بررسی می کنیم:

یک مجموعه داده آماری شامل مجموعه‌ای از مقادیر یک یا چند متغیر است. متغیرها می‌توانند عددی یا رسته‌ای (Categorical) باشند. متغیرهای عددی خود به دو دسته گسسته و پیوسته دسته بندی می‌شوند. این دسته بندی، روش‌های آماری را که برای داده ها مناسب اند، مشخص می کند. یکی از روش‌های خلاصه کردن و توصیف داده ها رسم یک نمودار آماری است. نوع نمودار مورد استفاده به نوع داده ها بستگی دارد و بسته به رسته ای بودن یا عددی بودن، نمودارهای مختلفی به کار برده می شود.

جداول فراوانی هم بسته به نوع متغیر، متفاوت خواهند بود، لذا مراحل فوق را برای انواع مختلف متغیر، جداگانه بررسی خواهیم کرد.

داده‌های کیفی(Categorical Data)

متغیرهای رسته‌ای به آن دسته از متغیرها اطلاق می شود که از نظر کیفی مقادیر آن به چندین رسته تقسیم می شود. برای مثال جنسیت، رنگ پوست، رشته تحصیلی، رتبه شغلی، شغل و ... نمونه هایی از متغیرهای رسته‌ای هستند. متغیرهای رسته ای به دو دسته کلی کیفی اسمی و کیفی رتبه ای تقسیم می‌شوند.

جدول فراوانی برای متغیرهای کیفی:

جداول فراوانی این نوع متغیرها، با فهرست کردن مقادیر مختلف متغیر، فراوانی مربوط به هر مقدار و درصد فراوانی هر مقدار، بدست خواهد آمد، با یک مثال نحوه ساختن این نوع جداول فراوانی را با SPSS می‌بینیم.

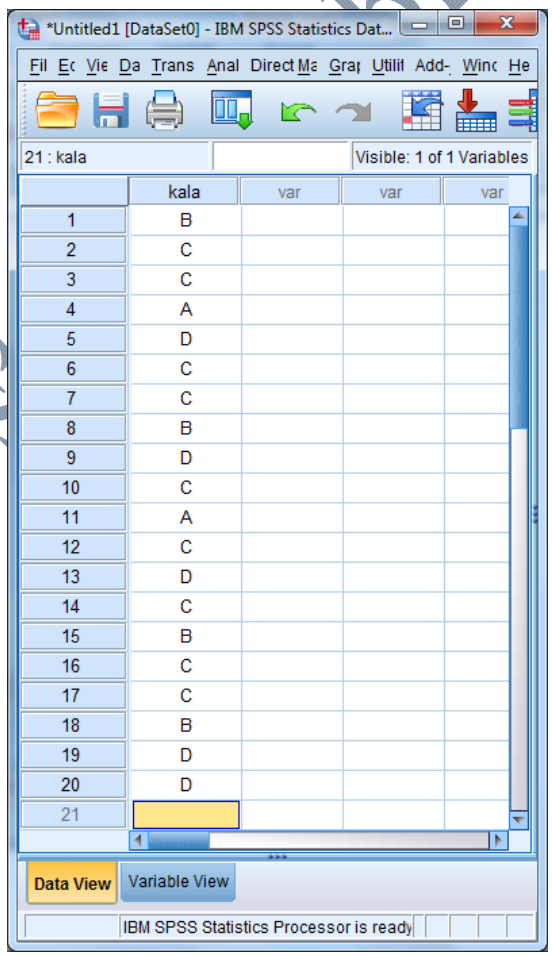
مثال(۳): صنعت‌گری چهار نوع قطعه D,C,B,A را تولید می‌کند اگر 20 قطعه تولید شده توسط وی به ترتیب زیر باشند.

B,C,C,A,D,C,C,B,D,C,A,C,D,C,B,C,C,B,D,D

یک جدول فراوانی برای داده های فوق می سازیم. سعی کنید با کمک آن به سئوالات زیر پاسخ دهید:

- چند عدد از قطعه C در این روز تولید شده است؟
- قطعات B و A چند درصد از کل تولید روزانه را در بر می‌گیرند؟

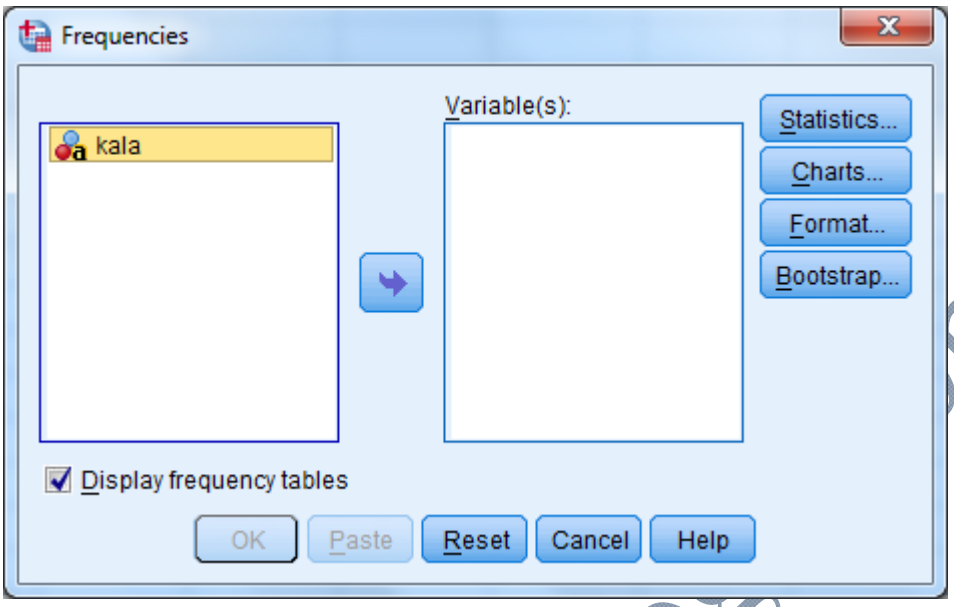
ابتدا داده‌ها را در محیط SPSS وارد می‌کنیم:



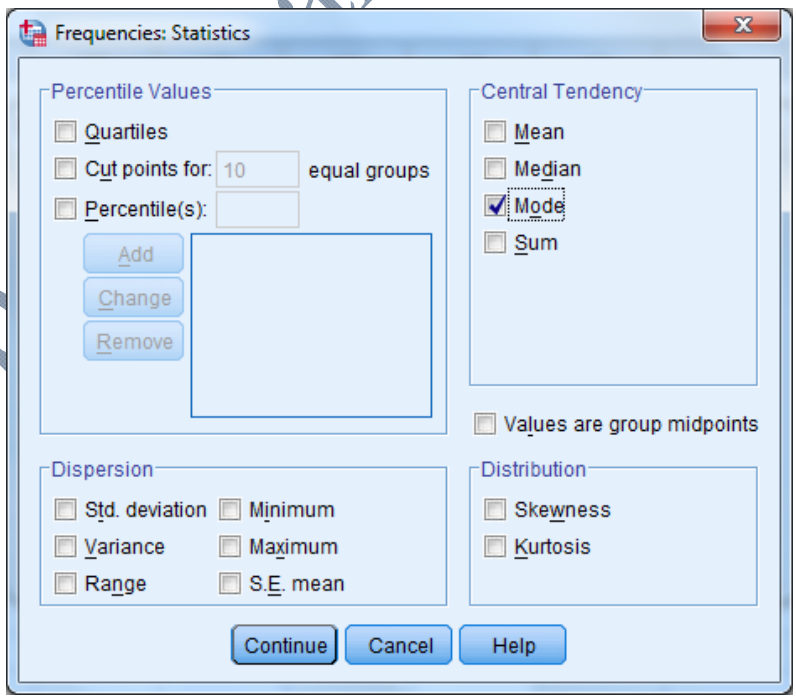
سپس برای رسم جدول توزیع فراوانی مسیر،

Analyze \ Descriptive Statistics \ Frequencies

را طی کنید تا کادر زیر باز شود.



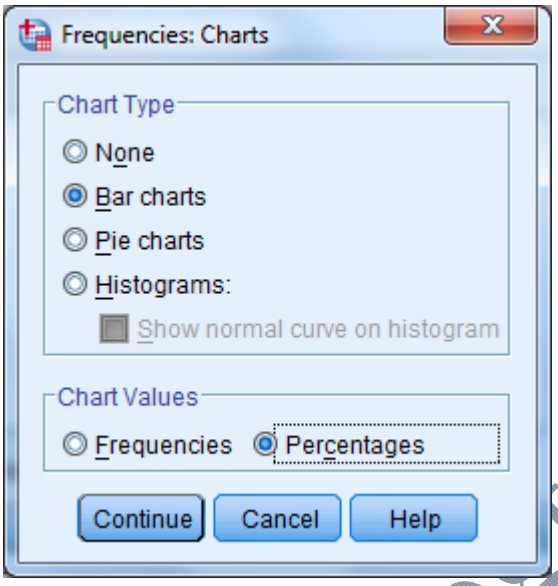
در کادر بالا متغیر kala را انتخاب کرده و با استفاده از دکمه آن را به کادر variable(s) انتقال می‌دهیم. برای رسم جدول فراوانی، در کادر کنار عبارت Display frequency tables تیک بگذارید. روی منوی Statistics کلیک کنید تا کادر زیر باز شود.



در این کادر بسته به نوع متغیرهایتان هر موردی را که بخواهید می‌توانید انتخاب نموده تا در خروجی نمایش داده شود. با توجه به کیفی بودن متغیر kala ما فقط Mode را انتخاب نموده‌ایم.

بر روی دکمه continue کلیک کنید.

بر روی گزینه charts کلیک کنید تا کادر زیر باز شود.



از قسمت Chart Type می‌توانید نوع نمودار خود را انتخاب کنید. در این مثال ما نمودار ستونی(Bar charts) را انتخاب کرده‌ایم. در قسمت Chart Value نیز میتوانید مشخص کنید که نمودار بر اساس فراوانی یا درصد فراوانی تشکیل شود. در این مثال ما درصد را انتخاب کرده‌ایم.

در نهایت بر روی دکمه های Continue و Ok کلیک می‌کنیم.

صفحه جداگانه ای تحت عنوان Output view باز خواهد شد که خروجی‌های SPSS همواره در آن ظاهر خواهد شد. خروجی این مثال به صورت زیر خواهد بود:

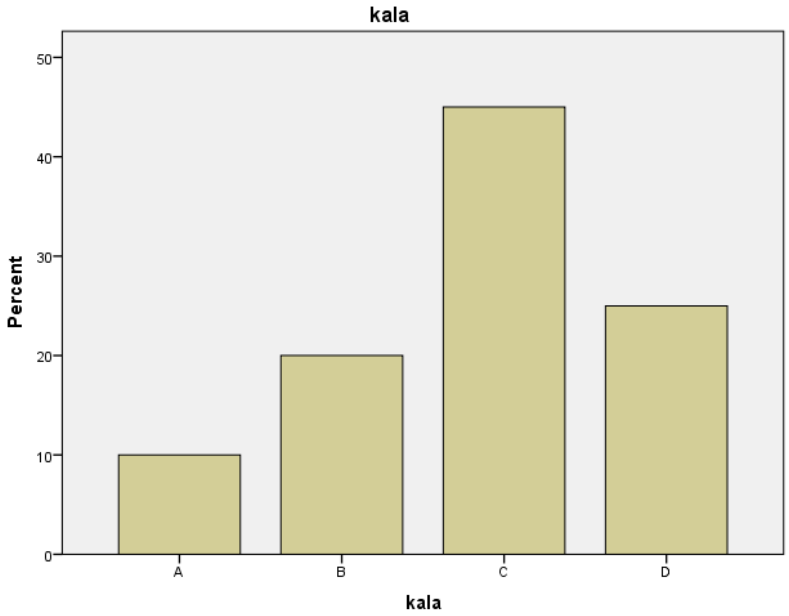
Statistics

kala

N	Valid	20
	Missing	0

kala

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	2	10.0	10.0	10.0
	B	4	20.0	20.0	30.0
	C	9	45.0	45.0	75.0
	D	5	25.0	25.0	100.0
	Total	20	100.0	100.0	



جدول اول تعداد داده‌های موجود(Valid) و داده‌های گمشده(Missing) را نشان میدهد. و جدول دوم جدول فراوانی متغیر است. ستون اول مقادیر متغیر، ستون دوم فروانی هر مقدار، ستون سوم درصد فراوانی آن مقدار و ستون چهارم هم درصد فراوانی تجمعی می‌باشد. از روی این جدول سعی کنید به سوال‌های مطرح شده در صورت مثال جواب دهید. برای متغیرهای کیفی رتبه ای هم مراحل رسم جدول فراوانی به همین صورت خواهد بود.

نمودارهای آماری برای متغیرهای کیفی:

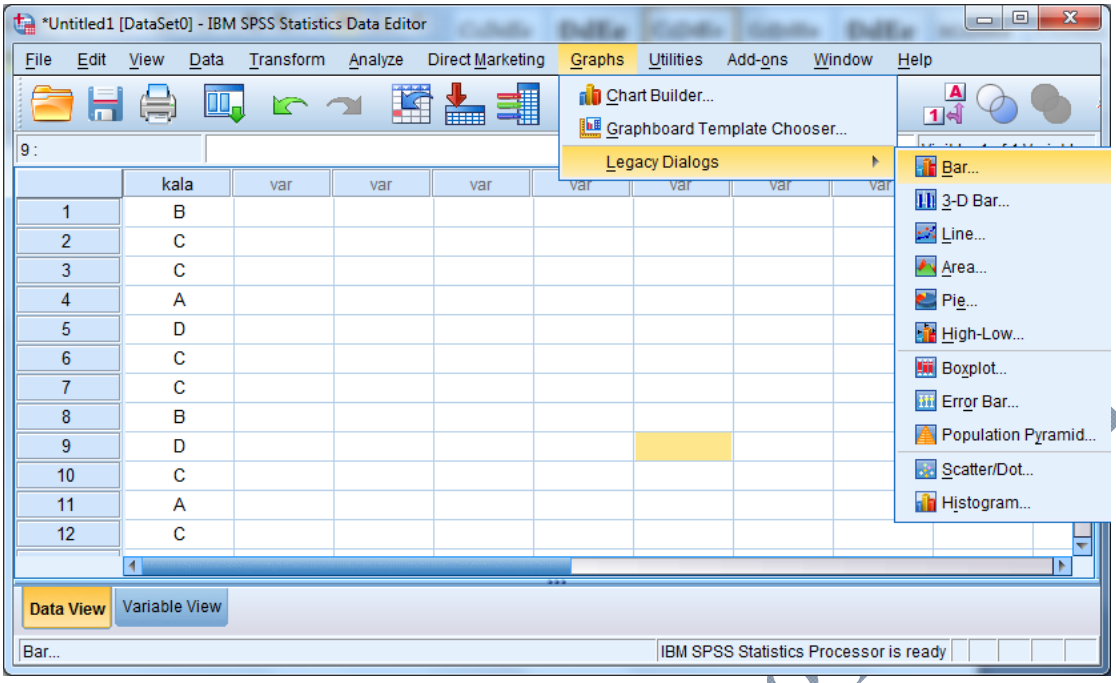
نمودارهای مناسب برای این نوع متغیرها عبارتند از :نمودار میله ای (Bar chart) و نمودار دایره ای(Pie chart) .

برای رسم این نمودارها می‌توان از دو راه زیر استفاده کرد:

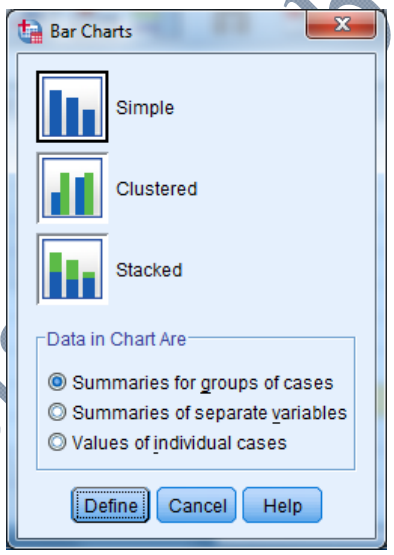
۱- از مسیری که برای رسم جدول فراوانی طی کردیم و در مثال قبل توضیح داده شد.

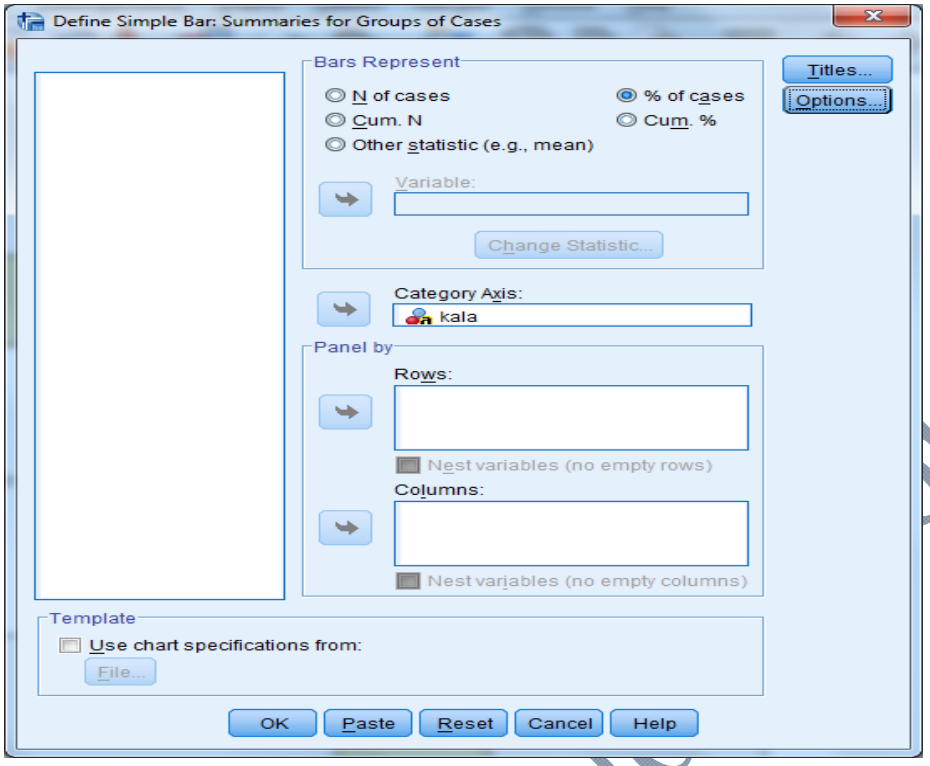
۲- راه دوم استفاده از منوی Graph است.

مسیر زیر را انتخاب میکنیم:



در کادر باز شده روی گزینه نشان داده شده در شکل زیر کلیک کرده و دکمه Define را کلیک کنید:





در کادر گفتگوی ظاهر شده متغیر مورد نظر را انتخاب کرده و دکمه Ok را کلیک کنید.

داده‌های عددی (Numerical Data)

داده‌های عددی دو نوع اند: گسسته و پیوسته. مقادیر متغیرهای گسسته اعداد حاصل از شمارش می‌باشد. برای مثال یک خانواده می‌تواند یک یا دو فرزند داشته باشد، اما تعداد فرزندان خانواده نمی‌تواند عددی ما بین این دو باشد. و در سمت مقابل، متغیرهای پیوسته فاقد واحدهای تفکیک پذیر هستند. برای مثال وزن یک متغیر پیوسته است.

متغیر عددی گسسته:

چون مقادیر یک متغیر گسسته، جدا از هم و معمولاً محدود است برای رسم جدول فراوانی یک متغیر عددی گسته همچون حالت متغیر رسته‌ای عمل می‌کنیم. اما اگر تعداد مقادیر متفاوتی که یک متغیر گسسته می‌گیرد زیاد باشد، برای رسم جدول فراوانی، با آن مثل یک متغیر پیوسته رفتار خواهیم کرد.

تمرین: فرض کنید تعداد قرصهای سرماخوردگی که یک خانواده در عرض زمستان مصرف کرده‌اند، در ۵۰ خانواده انتخاب شده به صورت زیر باشد:

۰،۰،۷،۵،۳،۳،۴،۵،۳،۲،۸،۳،۳،۲،۴،۴،۳،۶،۷،۴،۵،۴،۶،۴،۵

۲،۳،۴،۲،۷،۳،۵،۴،۶،۲،۳،۲،۴،۵،۴،۸،۴،۳،۲،۲،۶،۴،۵،۷،۸

- جدول فراوانی داده های فوق را با SPSS رسم کنید؟
- نمودارهای میله ای و دایره ای را برای داده های فوق رسم کنید؟
- مشخص کنید چند درصد از خانواده ها در طول زمستان ٦ قرص مصرف کرده اند؟ چند درصد حداکثر ٦ قرص مصرف کرده اند؟ چند درصد حداقل ٦ قرص مصرف کرده اند؟

چون متغیر ما عددی است می توانیم به جای نمودار میله ای از هیستوگرام (Histogram) که مخصوص داده‌های پیوسته است استفاده کنیم.

متغیر عددی پیوسته:

تبدیل داده‌های عددی پیوسته به گسسته:

نحوه رسم جداول فراوانی برای داده های پیوسته را با یک مثال بیان می کنیم.

مثال(٤): وزن‌های ٤٠ قالب کره که به نزدیکترین عدد صحیح گرد شده اند به قرار زیر است. جدول فراوانی داده های فوق را رسم کنید؟

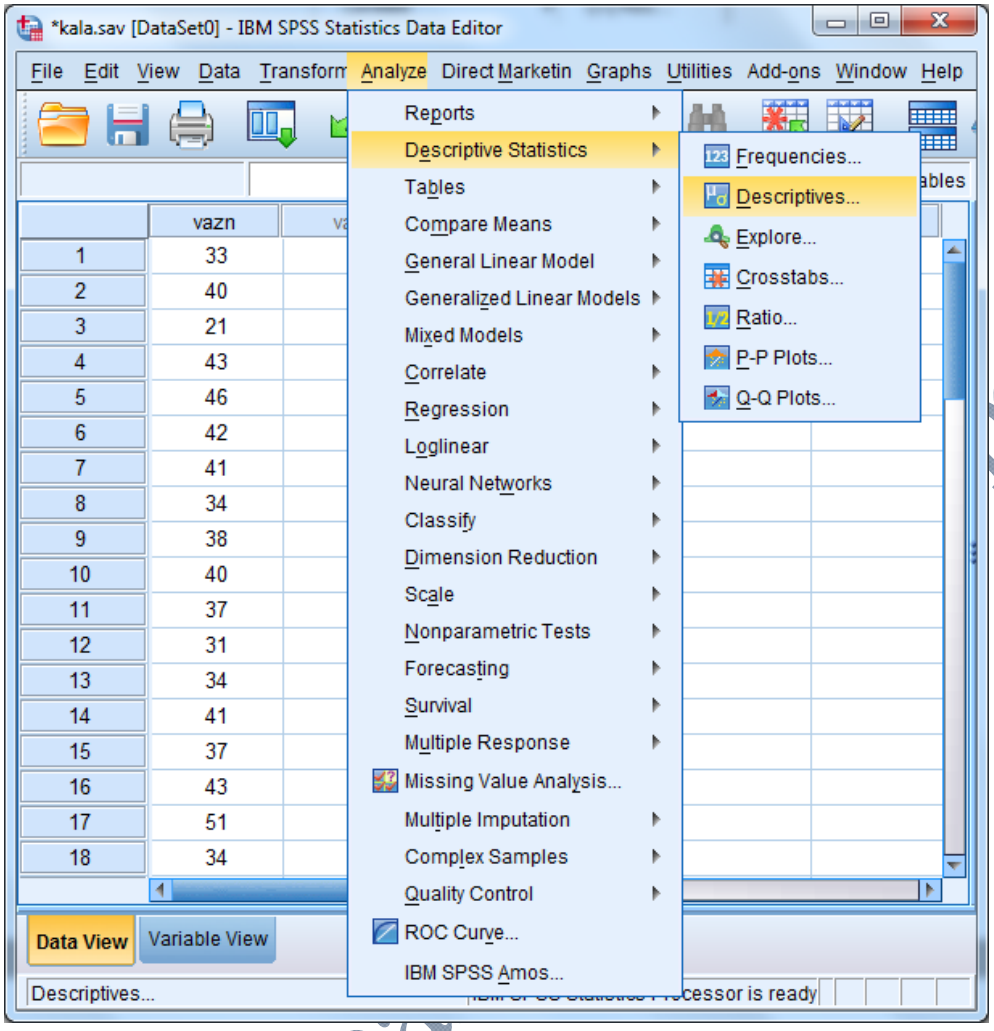
٥٢	٣٥	٢٤	٤٧	٣٦	٥١	٣٤	٣٨	٤٦	٣٣
٤٧	٣٦	٣٨	٥٠	٤٧	٣٤	٤١	٤٠	٤٢	٤٠
٢٦	٢٩	٣٠	٣٢	٣٠	٣٥	٣٧	٣٧	٤١	٢١
٣١	٣٠	٢٦	٣٥	٤٥	٢٣	٤٣	٣١	٣٤	٤٣

هرگاه داده‌های ما پیوسته باشد، داده‌ها را به تعدادی رده با طول مساوی تقسیم می‌کنیم و در هر رده فراوانی داده‌ها را می‌شماریم برای بدست آوردن تعداد رده‌ها در رسم جدول فراوانی به ترتیب زیر عمل کنید.

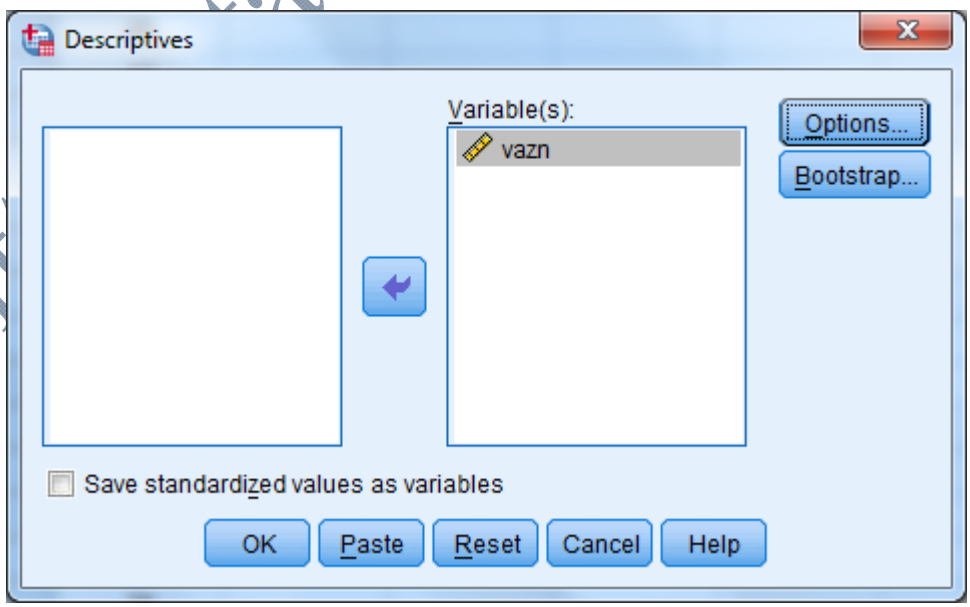
١- ابتدا حدود تغییرات داده ها را که از فرمول بدست می‌آید محاسبه می‌کنیم.

$$R=Max - Min$$

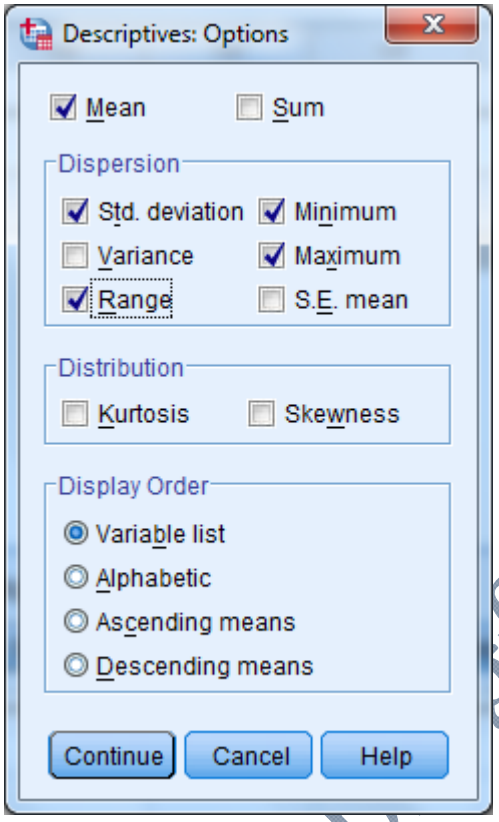
برای محاسبه دامنه تغییرات، کمترین داده ها و بیشترین داده‌ها از مسیر زیر استفاده می‌کنیم.



متغیر vazn را همانند شکل زیر به سمت راست و کادر variable(s) انتقال دهید.



سپس روی گزینه Options کلیک کنید تا کادر زیر ظاهر شود. گزینه های مورد نظر را تیک بگذارید.



Minimum، Maximum و Range را انتخاب کنید. دکمه Continue و سپس Ok را کلیک کنید. خروجی به صورت زیر خواهد بود.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
vazn	40	31	21	52	36.75	7.860
Valid N (listwise)	40					

۲- برای بدست آوردن تعداد رده ها یک قاعده عمومی وجود ندارد و معمولاً تعداد رده ها را بین ۵ تا ۵۲ رده اختیار می کنند. یک قاعده مفید استفاده از دستور استورگس Sturges به صورت زیر است:

$$m = ۱ + ۳/۲۲۲ \log(n)$$ (تعداد کل رده یا طبقات)

چون حاصل یک عدد اعشاری خواهد بود. آن را به بزرگترین عدد صحیح گرد می‌کنیم.

در مثال بالا داریم: $m = ۱ + ۳/۲۲۲ \log(۴۰) = ۶/۳۲۲$

پس تعداد طبقات را ۷ می گیریم.

۳- چون وزن‌ها به نزدیکترین عدد صحیح گرد شده‌اند بنابراین عدد ۵۳ در داده‌ها در واقع عددی بین ۵۲/۵ و ۵۳/۵ می باشد. عدد ۰/۵را تغییر پذیری مقادیر داده‌ها می‌نامیم که در ساختن حدود طبقات

مـورد اسـتفاده قـرار مـی‌گیـرد. طـول هـر طبقـه را هـم از فرمـول زیـر محاسـبه مـی‌کنـیم:

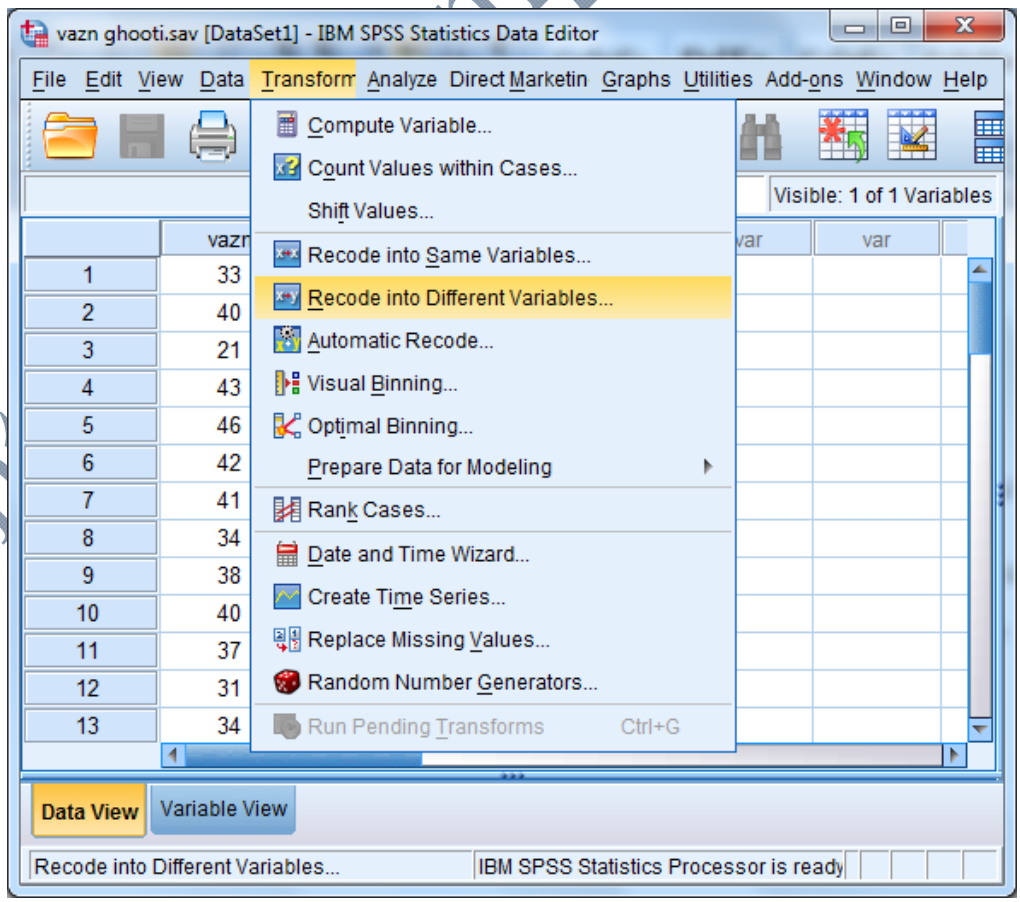
$L = Range / m = ۴/۵۷۴$ (طول رده)

پس طول هر طبقه را ۵ در نظر می‌گیریم. حال باید ۷ طبقه به طول ۵ بسازیم. طبقات مورد نظر بـه صـورت زیر خواهد بود:

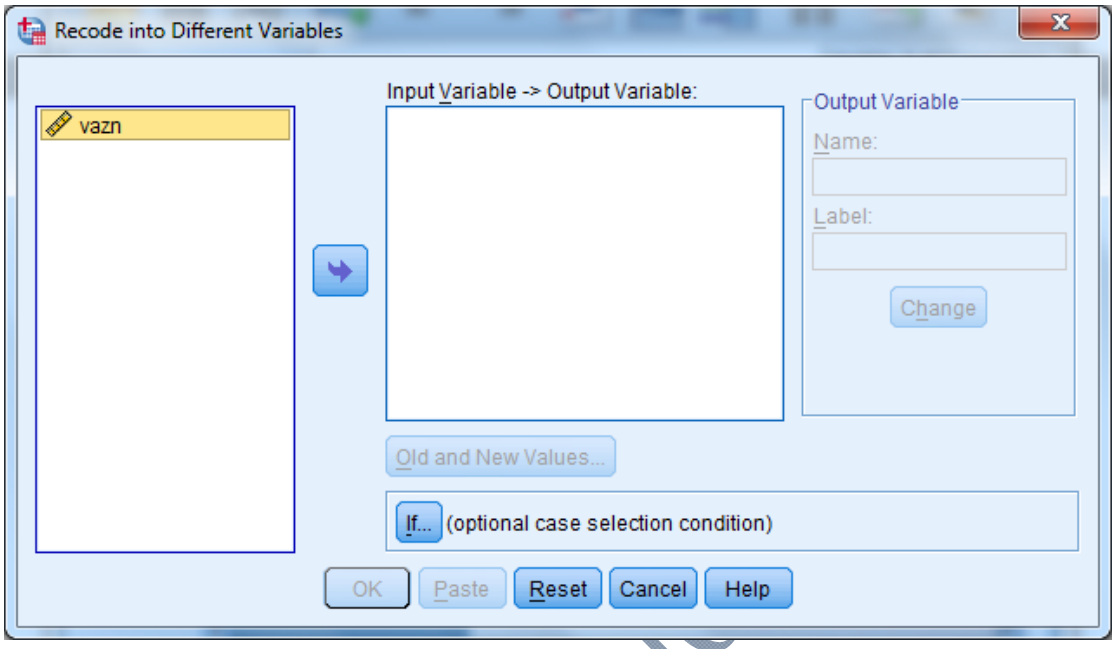
$۲۰/۵ - ۲۵/۵ \rightarrow ۱$
$۲۵/۵ - ۳۰/۵ \rightarrow ۲$
$۳۰/۵ - ۳۵/۵ \rightarrow ۳$
$۳۵/۵ - ۴۰/۵ \rightarrow ۴$
$۴۰/۵ - ۴۵/۵ \rightarrow ۵$
$۴۵/۵ - ۵۰/۵ \rightarrow ۶$
$۵۰/۵ - ۵۵/۵ \rightarrow ۷$

انتخاب حدود طبقات به صورت فوق باعث می‌شود که هر عدد دقیقاً در یک دسته قرار گیرد.

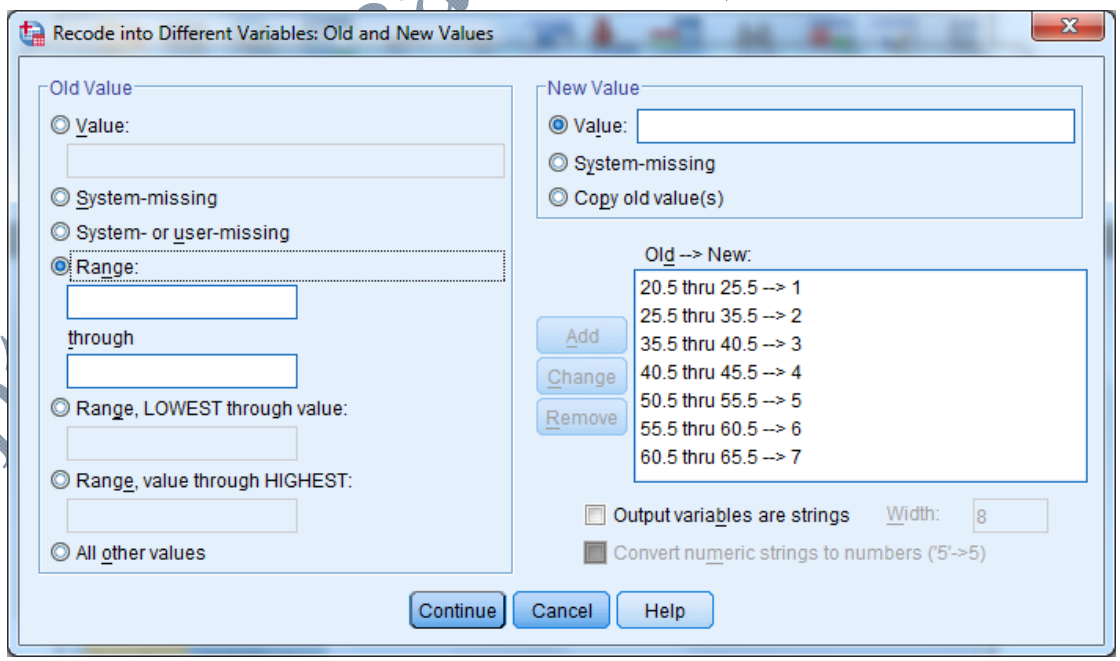
٤- حال متغیر جدیدی تعریف می کنیم که با توجه به واقع شدن داده در یکی از فواصل هفتگانه بالا یکـی از مقادیر ۱ تا ۷ را بپذیرد. برای تعریف این متغیر مسیر زیر را طی کنید:



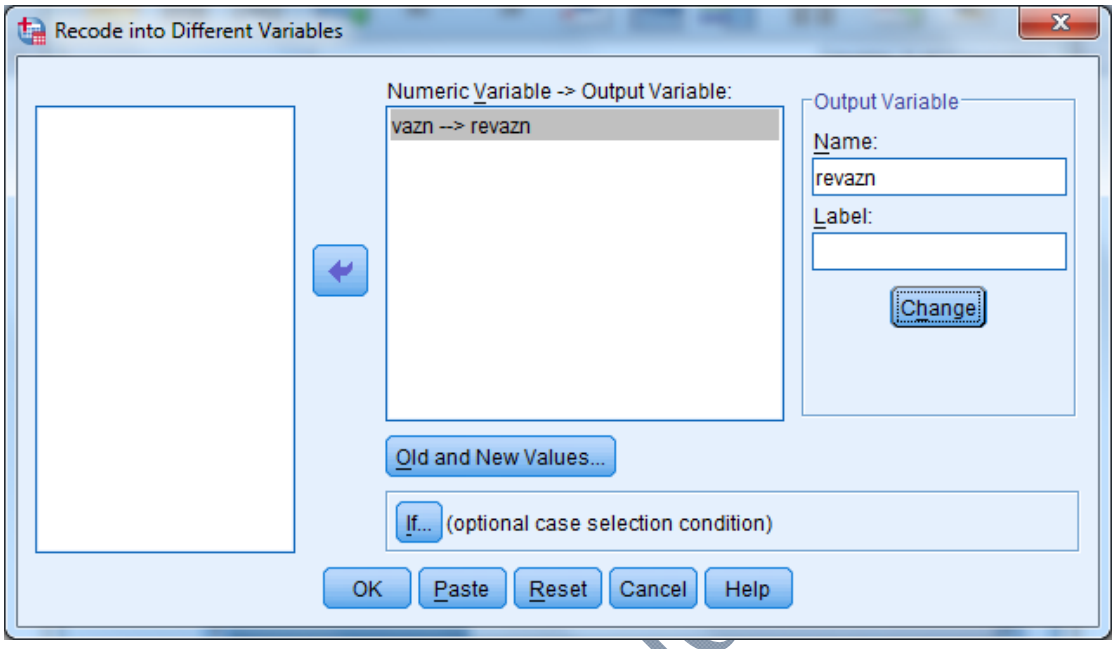
کادر گفتگوی زیر ظاهر می‌شود.



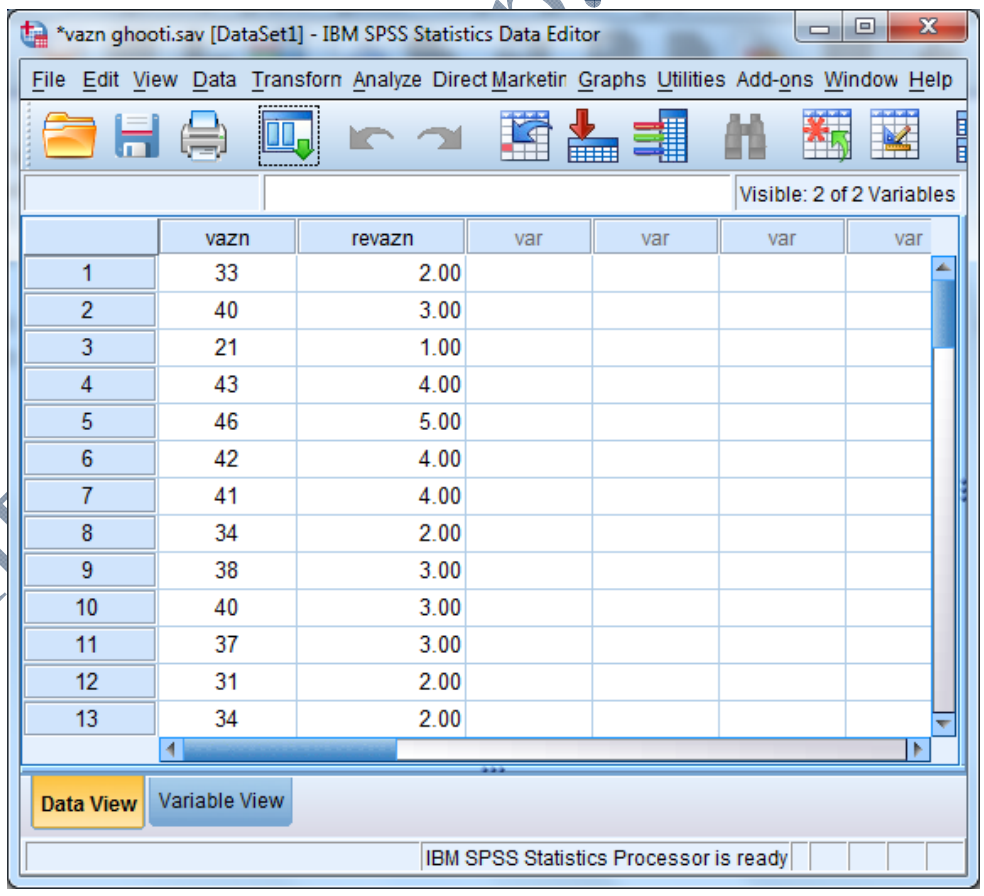
متغیر مورد نظر را انتخاب کرده و دکمه را فشار دهید تا به کادر سمت راست منتقل شود. سپس دکمه Old and New Values... کلیک کنید کادر گفتگوی زیر باز می شود که در آن مقادیر متغیر جدید را با توجه به مقادیر متغیر اولیه مشخص می کنیم:



پس از تعریف مقادیر دکمه Cotinue را کلیک کنید. مطابق تصویر زیـر در قسـمت Output variable نام متغیر جدید را تایپ کرده دکمه change را کلیک کنید.

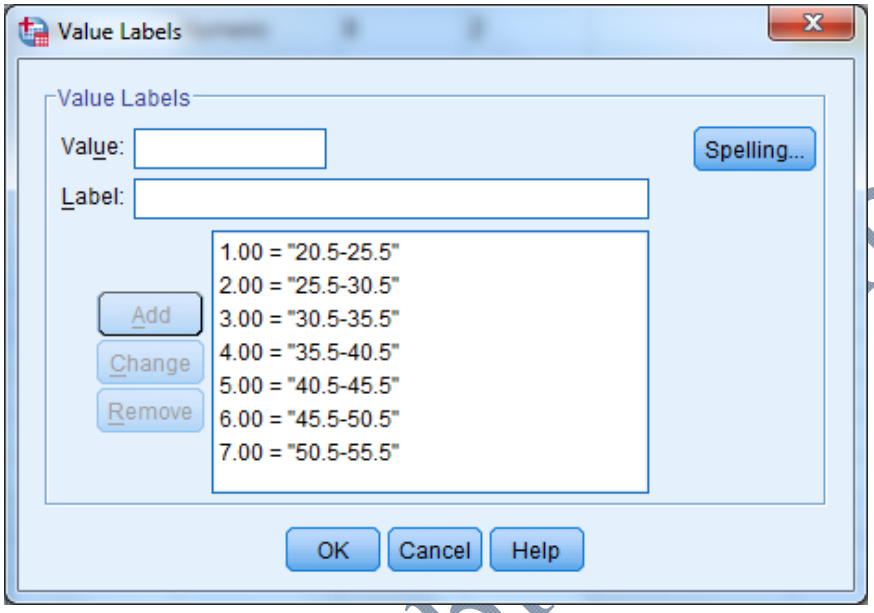


در پایان دکمه ok را کلیک کرده و نتیجه کار را به صورت زیر مشاهده کنید:



مشاهده می کنید که متغیر جدید با مقادیر ۱ تا ۷ ساخته شده است.

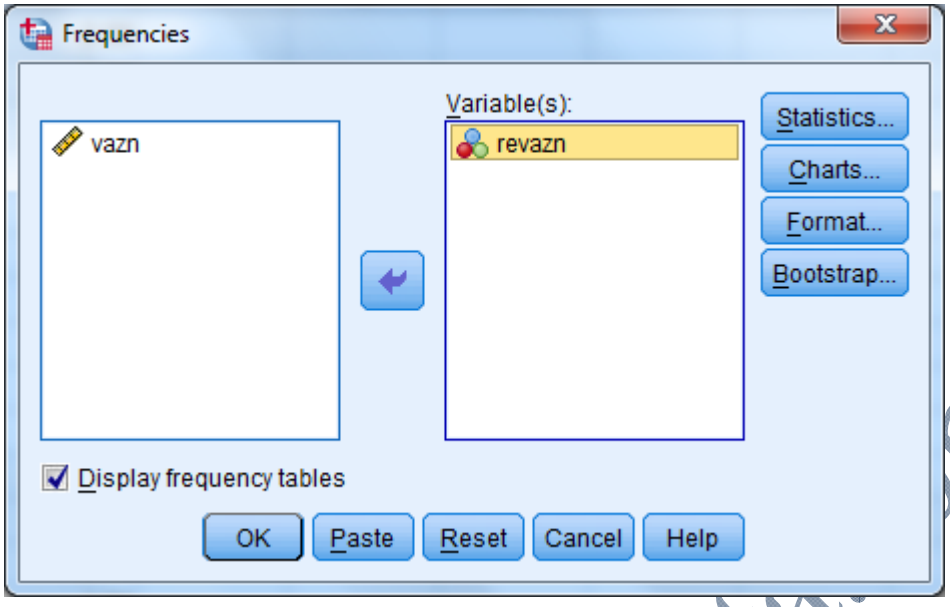
۵- به قسمت variable view بروید و در قسمت value برای مقادیر متغیر جدید، برچسب‌هایی به صورت زیر تعریف کنید:



به قسمت Data view برگردید.

اگر برای مقادیر یک متغیر، بر چسب تعریف شده باشد، نمایش مقادیر آن، به دو صورت خواهد بود: مقادیر متغیر و بر چسب های مقادیر. با کلیک روی دکمه زیر می‌توانید حالت نمایش را تغییر دهید:

حال متغیر پیوسته به یک متغیر گسسته با ۷ مقدار مختلف تبدیل شده است. برای رسم جدول فراوانی، همانند حالت گسسته عمل کنید (در کادر انتخاب متغیر، متغیر جدید کدبندی شده را انتخاب کنید).



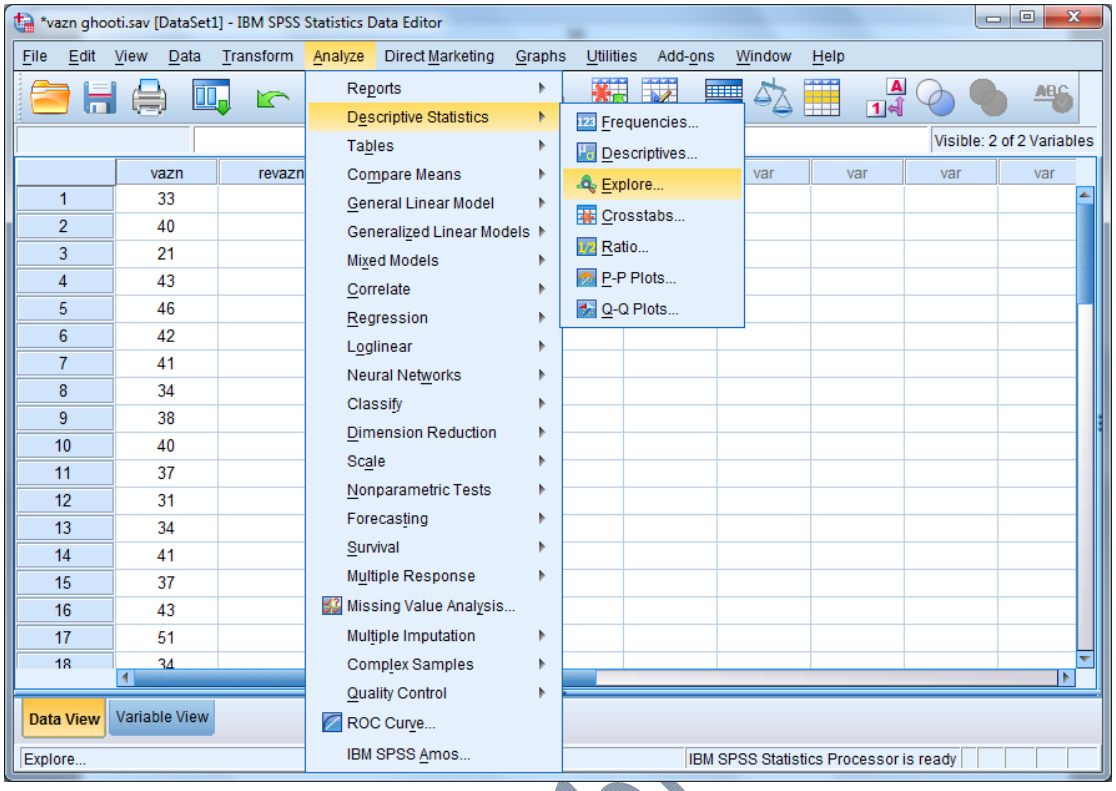
جدول فراوانی داده‌های مثال وزن قالب‌های کره به صورت زیر خواهد بود:

revazn

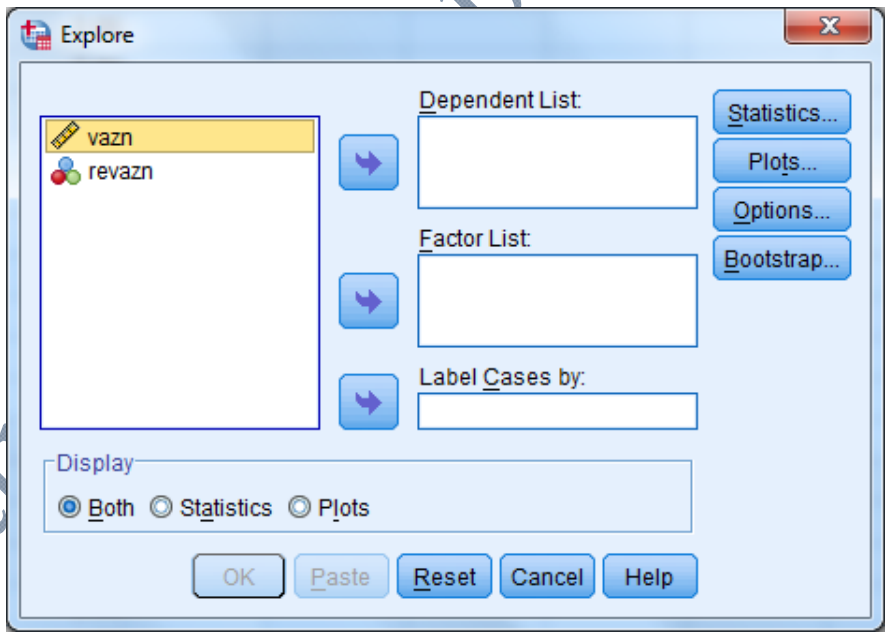
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20.5-25.5	3	7.5	7.5	7.5
	25.5-30.5	16	40.0	40.0	47.5
	30.5-35.5	8	20.0	20.0	67.5
	35.5-40.5	6	15.0	15.0	82.5
	40.5-45.5	5	12.5	12.5	95.0
	45.5-50.5	2	5.0	5.0	100.0
	Total	40	100.0	100.0	

رسم نمودار برای داده‌های کمی پیوسته:

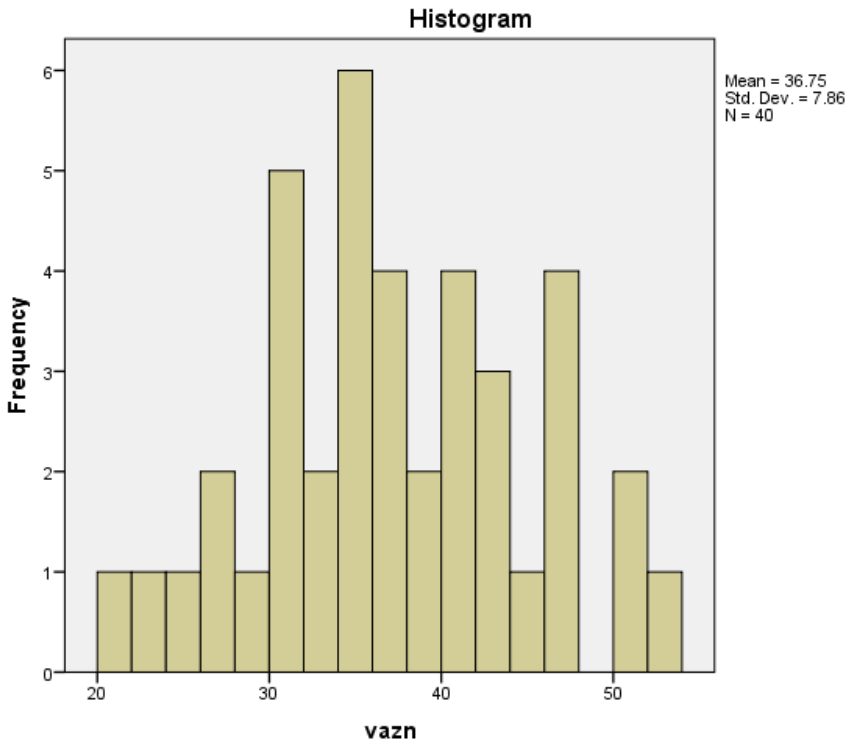
برای داده‌های پیوسته نمودارهای مختلفی به کار می‌روند که هر کدام کاربردهای خاص خود را دارد. ارجح ترین این نمودارها Histogram است. برای رسم هیستوگرام داده‌های پیوسته مسیر زیر را طی کنید:



کادر گفتگوی زیر ظاهر خواهد شد.



در قسمت Display (سمت چپ پایین کادر بالا) plots را انتخاب کنید. روی دکمه Plot...کلیک کنید در کادر گفتگو ظاهر شده Histogram را انتخاب کنید. Continue را کلیک کرده سپس ok را کلیک کنید. هیستوگرام داده‌های مثال قبل به صورت زیر خواهد بود:



یکی دیگر از نمودارهایی که برای متغیرهای پیوسته به کار می‌رود و در سال‌های اخیر کاربرد آن بسیار زیاد شده است نمودار جعبه‌ای یا Box plot است که به آن بعد از بحث شاخص‌های آماری خواهیم پرداخت.

محاسبه شاخص‌های آماری:

مرحله اول آمار توصیفی یعنی تشکیل جداول فراوانی و رسم نمودارهای آماری در SPSS بیان شد. مرحله دوم در آمار توصیفی، خلاصه کردن داده‌ها در قالب اعدادی است که موسوم به **شاخص‌های آماری** هستند. شاخص‌های آماری به دو دسته تقسیم می‌شوند: شاخص‌هائی که گرایش به مرکز یا مرکزیت داده‌ها را اندازه می‌گیرد(شاخصهای مرکزی). و شاخص‌هائی که برای اندازه‌گیری تغییر پذیری داده‌ها به کار می‌رود(شاخص‌های پراکندگی).

شاخص‌های مرکزی:

شاخص‌های مرکزی مهم عبارتند از

مد(Mode): مد داده‌ای است که بیشترین فراوانی را دارد. استفاده از این شاخص بیشتر در متغیرهای رسته ای است.

میانه(Median) و چندک‌ها: میانه به داده وسطی داده ها اطلاق می‌شود و در داده های کم تعداد یک شاخص پرکاربرد و کار آمد است. میانه داده ای است که تقریباً نصف داده‌ها از آن کمتر و نصف داده‌ها از آن بیشترند. تعریف چندک‌ها هم معادل میانه است، چندک مرتبه p، مقداری است که تقریباً $(۱۰۰\times P)$درصد داده ها از آن کمتر یا مساوی آن و $۱۰۰(۱-P)$ درصد داده ها از آن بیشرند. ساده ترین نوع چندکها، چارکها (Quartiles) و دهکها هستند.

چارک اول: مقداری است که یک چهارم داده ها از آن کمتر یا مساوی با آن هستند.

چارک دوم: معادل میانه است.

چارک سوم: مقداری است که سه چهارم داده‌ها از آن کمتر یا مساوی با آن هستند.

دهک اول: مقداری است که یک دهم داده‌ها از آن کمتر یا مساوی با آن هستند.

سایر دهک‌ها هم به همین صورت تعریف می‌شوند.

میانگین(Mean): پرکاربردترین و کاراترین شاخص برای اندازه گیری مرکزیت داده ها میانگین است. البته در صورتیکه تعداد داده ها کم باشد یا تعدادی داده پرت در میان داده ها مشاهده شود، دقت میانگین کاهش خواهد یافت لذا درصورتیکه یکی از حالات فوق اتفاق بیافتد باید در استفاده از میانگین هوشیار بود.

برای رفع مشکل داده‌های پرت، انواع دیگری از میانگین تعریف می‌شود که اثر اینگونه داده‌ها را کاهش می-دهد.

شاخص‌های پراکندگی:

غیر از شاخص‌هائی که گرایش داده‌ها را به یک مقدار مرکزی نشان می‌دهد، علاقه‌مند به شاخص‌هائی هستیم که به نوعی میزان پراکندگی داده‌ها را بیان کنند.

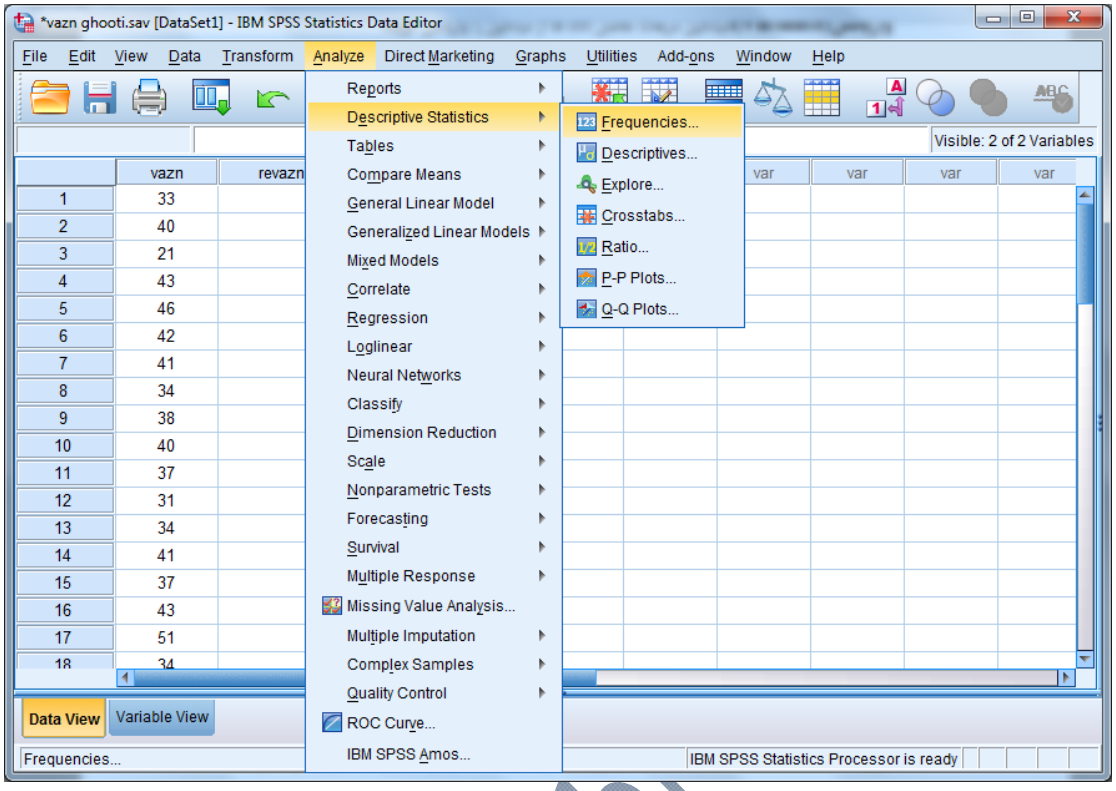
مهمترین شاخص‌های آماری پراکندگی عبارتند از:

دامنه تغییرات(Range): تفاضل بزرگترین و کوچکترین داده را دامنه تغییرات می نامند.

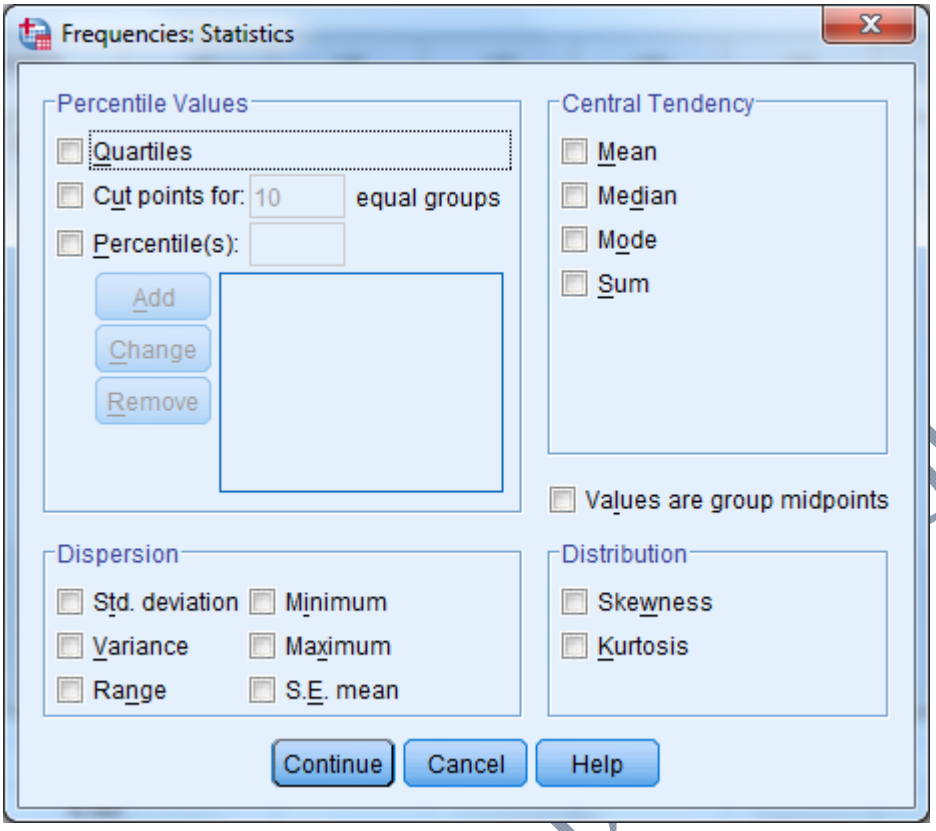
واریانس(Variance): میانگین مربعات تفاضل داده ها از میانگین را واریانس گویند.

انحراف معیار(Standard division): جذر واریانس را انحراف معیار گویند.

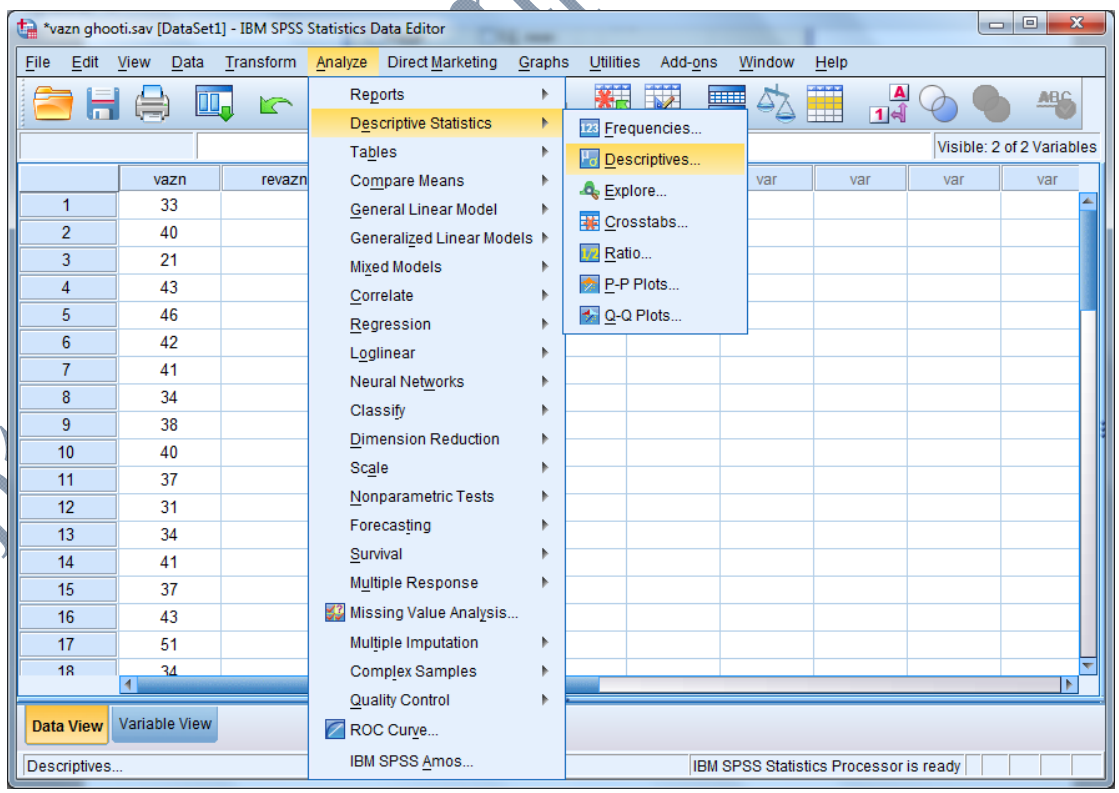
انحراف استاندارد میانگین(Standard Error of Mean): جذر حاصل تقسیم واریانس بر تعداد داده ها را انحراف استاندارد میانگین گویند. برای محاسبه شاخصهای بالا، ابتدا مسیر زیر را طی کنید:



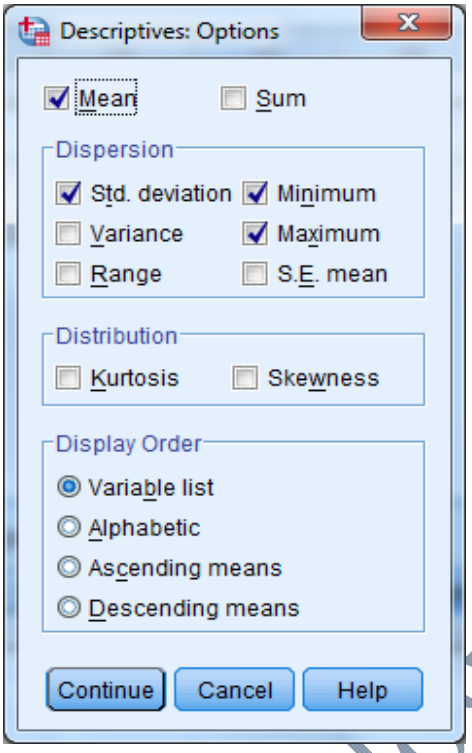
روی دکمه Statistics... کلیک کنید. کادر زیر باز خواهد شد. شاخص‌های مرکزی را می‌توانید از کادر Central Tendency و شاخص‌های پراکندگی را از کادر Dispersion و نوع شکل توزیع(تقارن و کشیدگی) را از کادر Distribution انتخاب کنید. پس از انتخاب شاخص‌های مورد نظر ف دکمه Continue و سپس OK را کلیک کنید.



و یا می‌توان از مسیر زیر استفاده کرد:

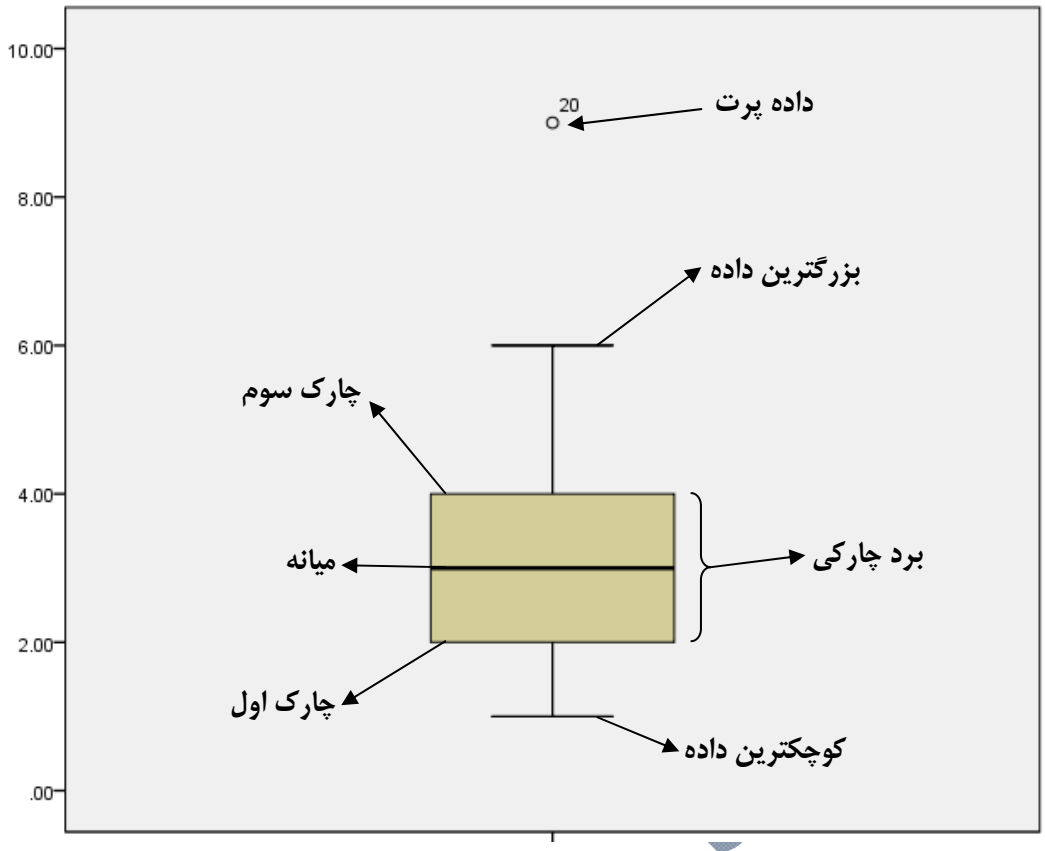


روی دکمه Option... کلیک کنید و شاخص‌های مورد نظرتان را انتخاب کنید:

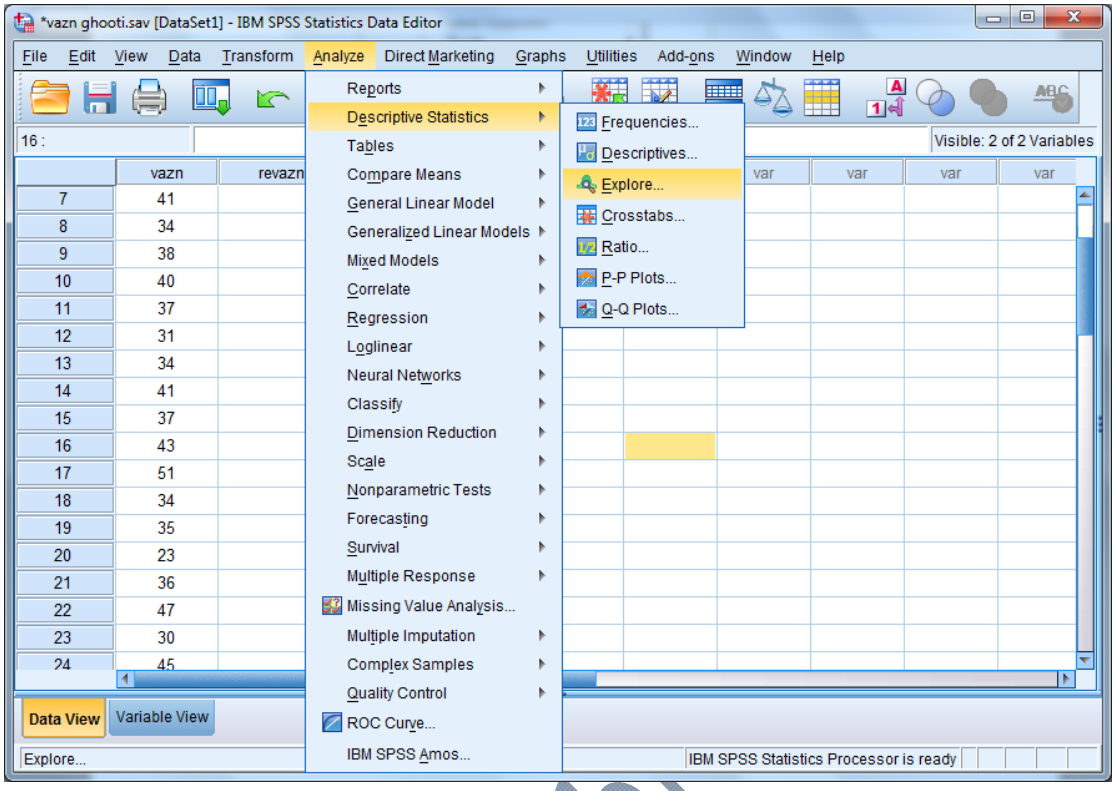


نمودار Box plot:

نمودار جعبه ای از نمودارهای پرتوان و پرکاربرد آماری است که در کنار گرایش به مرکز، پراکندگی داده‌ها، داده‌های پرت، تقارن و الگوی کلی داده ها را بیان می‌کند. شمایل کلی یک نمودار جعبه ای به صورت زیر است:

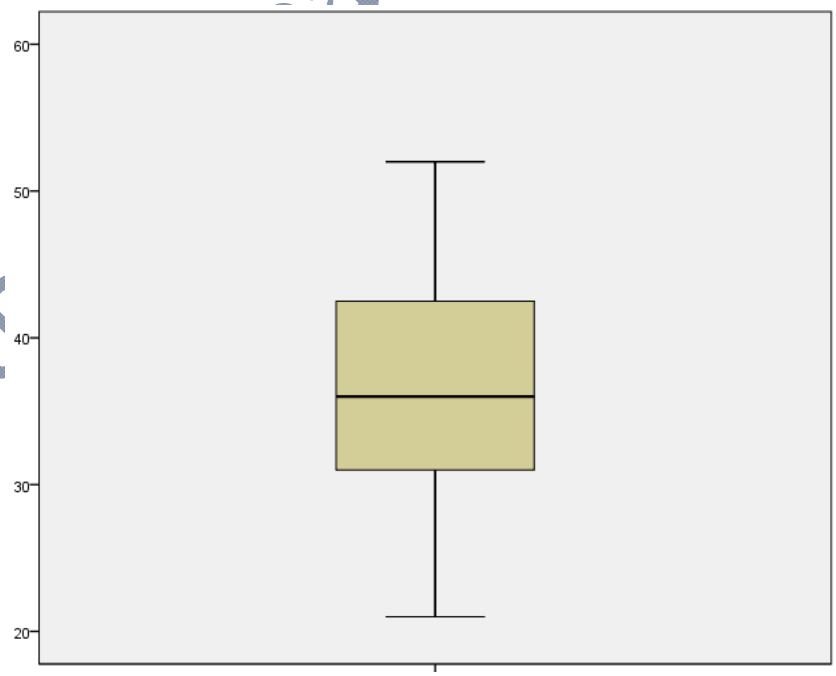


فاصله میان چارک اول و چارک سوم را دامنه میان چارکی می گویند. اگر داده ای بیش از۱/۵ برابر برد چارکی از چارک اول یا چارک سوم فاصله داشته باشد، داده پرت محسوب می‌شود و به صورت یک نقطه خارج از نمودار نشان داده می‌شود و نمودار بدون احتساب آن رسم می‌شود. برای رسم Box plot مسیر زیر را طی کنید:



روی Plot کلیک کنید و در کادر ظاهر شده نمودار Box plot را انتخاب کنید.

نمودار جعبه ای داده‌های مثال وزن قالب‌های کره‌ای در زیر رسم شده است. نحوه تقارن داده‌ها، میزان پراکندگی و وجود یا عدم وجود داده‌های پرت را به خوبی می‌توان در نمودار مشاهده کرد.



فصل سوم:
آمار استنباطی

آزمون فرض‌های آماری

همواره هدف از یک بررسی آماری، جمع‌آوری و تنظیم اطلاعات از قسمتی ازجامعه (نمونه) و تجزیه و تحلیل و نتیجه‌گیری از روی آنها در مورد کل جامعه است. این تجزیه و تحلیل و نتیجه گیری را استنباط آماری می‌گویند. در این بخش یکی از شاخه‌های مهم استنباط آماری یعنی آزمون فرض‌های آماری را مورد بررسی قرار می‌دهیم. هر فرض آماری در حقیقت حکمی درباره‌ی جامعه است که ممکن است درست یا نادرست باشد. به طور کلی، هدف آزمون فرض‌های آماری تعیین این موضوع است که با توجه به اطلاعات به دست آمده از داده‌های نمونه، حدسی که درباره خصوصیتی از جامعه می‌زنیم، قویاً قابل تایید است یا خیر.

برای انجام یک فرض آماری نیاز به یک سری تعاریف و مفاهیم داریم که در ذیل مختصراً به انها اشاره خواهیم کرد.

تعریف فرضیه: فرضیه اظهار نظری است که در ارتباط با پارامترهای یک یا چند جمعیت بیان می‌شود.

مثال: فرض کنید محققی ادعا می‌کند که میانگین فشار خون بیمارانی که از داروی جدید A استفاده می‌کنند در مقایسه با بیمارانی که تحت درمان استاندارد B قرار می‌گیرند کمتر است.

عبارت فوق یک فرضیه است که در ارتباط با میانگین فشار خون(پارامتر مجهول) دو گروه مطرح شده است.

آزمون فرض: با استفاده از آزمون فرض‌ها می‌توان نسبت به رد یا قبول فرضیه‌ای تصمیم گیری نمود. به عبارتی در آزمون یک فرضیه می‌توان تعیین نمود که آیا این اضهار نظرها یا فرضیه بیان شده با داده‌های موجود سازگار است. در این روش بر اساس اطلاعات جمع آوری شده (نمونه گیری) در ارتباط با فرضیه مورد نظر اظهار نظر خواهیم کرد.

مثال: فرض کنید محققی می‌خواهد بداند که آیا میانگین فشار خون بیمارانی که از داروی جدید A استفاده میکنند در مقایسه با بیمارانی که تحت درمان استاندارد B قرار می‌گیرند از نظر آماری متفاوت است؟

این مسئله در حقیقت بیان آزمون فرضیه است که با یک روش آماری خاص تحلیل می شود.

در هر آزمون آماری یک فرضیه اولیه وجود دارد که آنرا فرضیه صفر یا H_0 می‌گویند و بعنوان فرض عدم اختلاف شناخته میشود. در برابر این فرضیه، فرض H_1 یا فرض مقابل وجود دارد (ادعای مطرح شده) که بعنوان فرض وجود تفاوت یا اختلاف شناخته میشود.

در مثال بالا:

فرض صفر یا H_0 : میانگین فشار خون بیمارانی که از درمان A استفاده می‌کنند با بیمارانی که تحت درمان B قرار میگیرند برابر است.

فرض مقابل یا H_1: میانگین فشار خون بیمارانی که از درمان A استفاده میکنند با بیمارانی که تحت درمان B قرار میگیرند متفاوت (کمتر) است.

در هر آزمون فرضیه، از دیدگاه آماری، مرتکب دو نوع خطا می شویم:

خطای نوع اول که با α نمایش داده می‌شود: عبارت است از رد فرض صفر(به اشتباه) وقتی فرض صفر درست باشد. مقدار α در اختیار پژوهشگر است که معمولا ۰/۰۵ انتخاب می‌شود!

خطای نوع دوم که با β نمایش داده می‌شود: عبارت است از پذیرش فرض صفر(به اشتباه) وقتی فرض مقابل درست باشد.

توان آزمون: عبارت است از رد فرض صفر وقتی فرض مقابل درست باشد $(1-\beta)$.

در بین دو خطای مطرح شده خطای نوع اول یا α مهمتر از خطای نوع دوم یا β می‌باشد به همین دلیل ما α را قبل از شروع آزمایش ثابت فرض می‌کنیم. معمولاً مقدار α بسته به نوع آزمایش و نظر آزمایشگر برابر ۰/۰۵، ۰/۰۱ یا ۰/۱ در نظر گرفته می‌شود.

(توجه: پیش فرض نرم‌افزارهای آماری برای α مقدار ۰/۰۵ است.)

مفهوم P-value : معمولاً نتیجه هر آزمون آماری با P-value بیان می‌شود.

به عنوان مثال اگر متخصص آمار با استفاده از یک روش آماری فرضیه بالا را مورد آزمون قرار داده و P-value را ۰/۰۲ بیان کند چه نتیجه‌ای می‌گیریم به عبارتی مفهوم ۰/۰۲ چیست؟

به ما میگوید که اگر قرار باشد که بر اساس نمونه جمع آوری شده فرض صفر را به اشتباه رد کنیم (درمان A از B بهتر است) احتمال این اشتباه یا خطا تنها ۲ درصد یا ۲ صدم میباشد که این خطا به مراتب کمتر از خطایی است که ما اجازه داریم مرتکب شویم (یعنی $\alpha=0.05$).

و به صورت P-value<0.05 بیان می‌شود.

حال به استنباط‌هایی در مورد یک جامعه و دو جامعه می پردازیم:

استنباط در مورد یک جامعه:

زمانی که محققی در مورد یک جامعه به استنباط می‌پردازد و آزمون‌هایی را در این مورد انجام می‌دهد، عموماً دو هدف را دنبال می‌کند:

۱- ممکن است محقق بخواهد **از روی آماره‌های یک نمونه، در مورد پارامترهای یک جامعه استنباط انجام دهد.**

۲- ممکن است محقق بخواهد توزیع نمونه را با یک توزیع فرضی مانند توزیع نرمال مقایسه کند. در اصطلاح فنی، سئوال در مورد تطابق (Goodness-of-Fit) توزیع نمونه با یک توزیع فرضی خاص است. ابتدا در مورد هدف ۱ و سپس در مورد هدف ۲ بحث خواهیم کرد و در هر مورد تا استنباط در مورد دو جامعه پیش خواهیم رفت.

آزمون فرض در مورد میانگین یک جامعه:

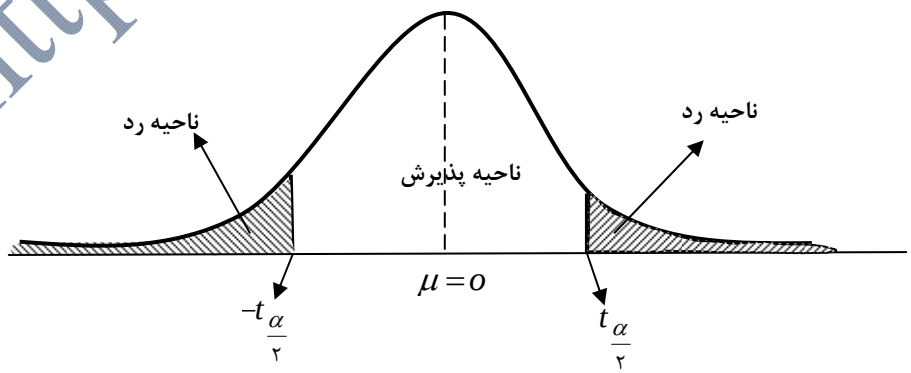
علاقه‌مندیم آزمون کنیم که آیا میانگین جامعه برابر عدد مشخصی هست یا خیر؟ فرض‌های آماری در اینصورت عبارتند از:

$$\begin{cases} H_o: & \mu = \mu_o \\ H_1: & \mu \neq \mu_o \end{cases}$$

آزمون مورد استفاده آزمون t یک نمونه‌ای می‌باشد. آماره آزمون به صورت زیر است:

$$t_{obs} = \frac{\bar{x} - \mu_o}{\frac{S}{\sqrt{n}}}$$

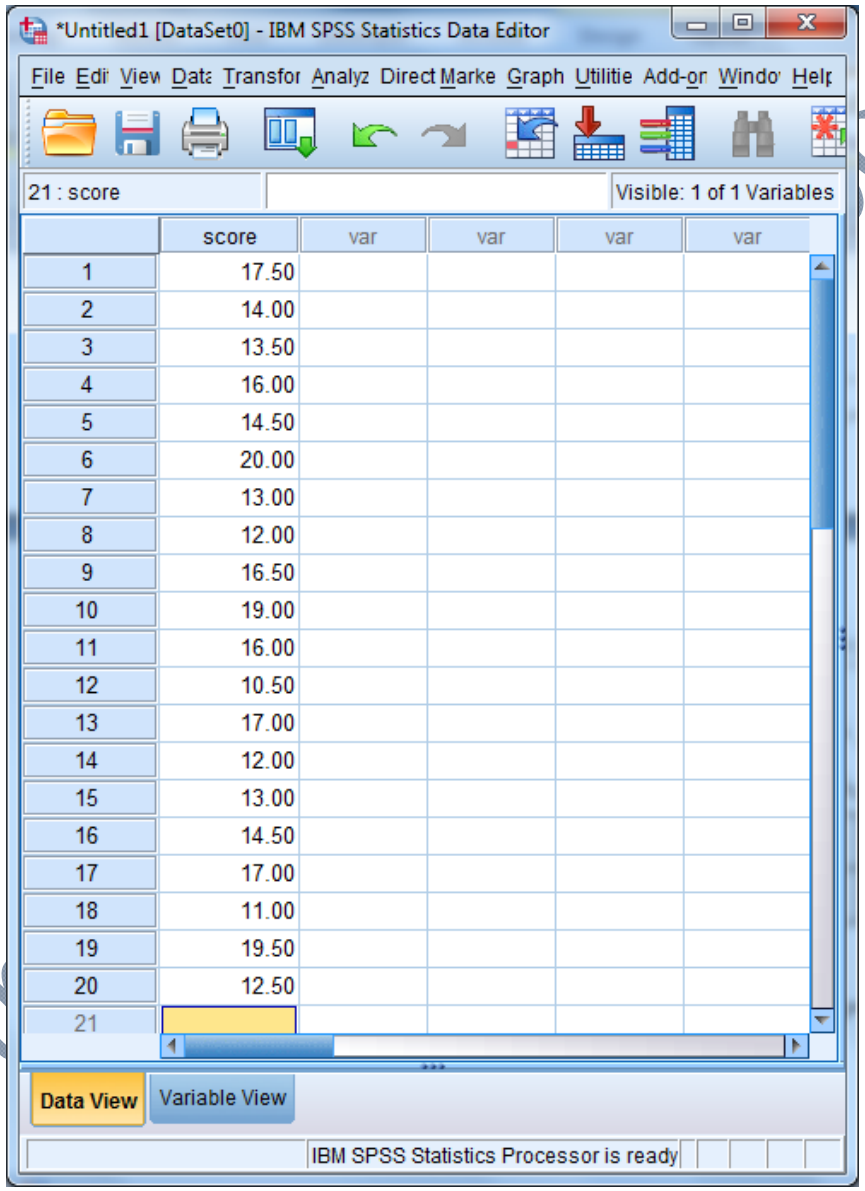
ناحیه رد و پذیرش برای آزمون بالا به صورت زیر خواهد بود.



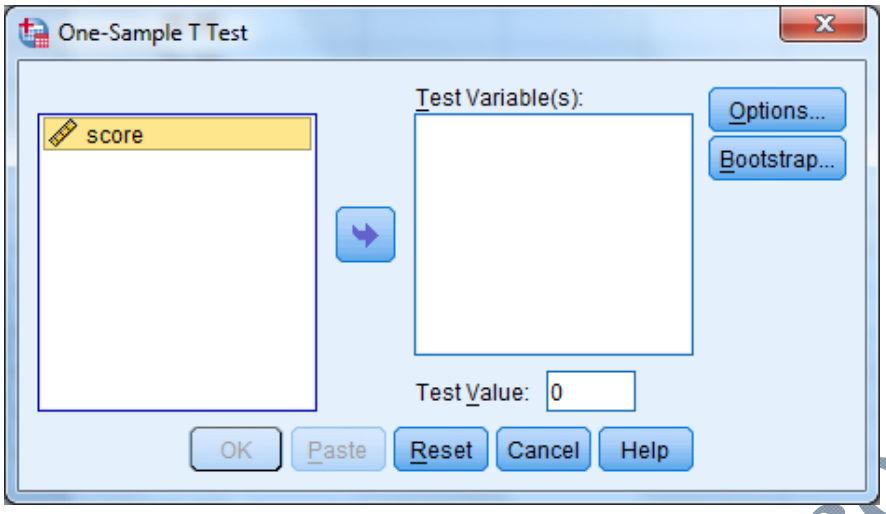
مثال(۵): داده‌های زیر نمره‌های ۲۰ دانش آموز کلاس در درس ریاضی است. آیا می‌توان گفت میانگین نمره‌های ریاضی دانش آموزان این کلاس ۱۲ است؟ ۱۵ چطور؟

۱۹	۱۶/۵	۱۲	۱۳	۲۰	۱۴/۵	۱۳	۱۶	۱۳/۵	۱۴	۱۷/۵
۱۲/۵	۱۹/۵	۱۱	۱۷	۱۴/۵	۱۳	۱۲	۱۷	۱۰/۵	۱۶

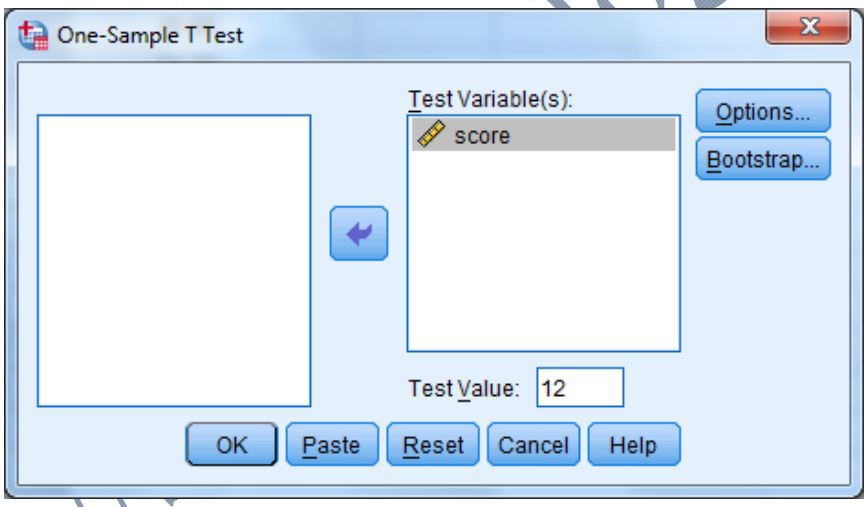
ابتدا داده‌ها را به صورت زیر وارد SPSS می‌کنیم.



سپس مسیر Analyze / Compare Means / One-Sample T Test… را انتخاب تا کادر زیر باز شود.



متغیر score را به سمت راست و داخل کادر Test Variable(s) انتقال دهید. و در قسمت Test Value مقداری را قصد آزمون دارید بنویسید. که در این مثال برای قسمت اول مقادیر ۱۲ و ۱۵ است. یک‌بار آزمون را برای مقدار ۱۲ و یک بار برای مقدار ۱۵ به صورت زیر انجام خواهیم داد.



نتایج مربوط به دو آزمون در زیر داده شده است.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
score	20	14.9500	2.82796	.63235

One-Sample Test

	Test Value = 12					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
score	4.665	19	.000	2.95000	1.6265	4.2735

همانطور که از داده‌های جدول بالا مشخص است، سطح معناداری در آزمون بالا برابر با مقدار ۰/۰۰۰ و کوچکتر از مقدار ۰/۰۵ است. در نتیجه در سطح اطمینان ۹۵ درصد فرض صفر پژوهش یعنی برابر بودن میانگین نمرات ریاضی با مقدار ۱۲ را نمی‌پذیریم.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
score	20	14.9500	2.82796	.63235

One-Sample Test

	Test Value = 15					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
score	-.079	19	.938	-.05000	-1.3735	1.2735

همانطور که از داده‌های جدول بالا مشخص است، سطح معناداری در آزمون بالا برابر با مقدار ۰/۹۳۸ و بزرگتر از مقدار ۰/۰۵ است. در نتیجه در سطح اطمینان ۹۵ درصد فرض صفر پژوهش یعنی برابر بودن میانگین نمرات ریاضی با مقدار ۱۵ را می‌پذیریم.

تمرین(۱): اداره بهداشت یک شهر، می‌خواهد تعیین کند که آیا میانگین تعداد باکتریها در واحد حجم آب شهر از سطح ایمنی یعنی ۲۰۰ بیشتر است یا نه؟ پژوهشگران ۱۰ نمونه از آب، هر یک به حجم واحد، را گرد آوری کرده و دیده‌اند که تعداد باکتریها عبارتند از:

۲۰۷، ۲۱۰، ۱۹۳، ۱۹٦، ۱۸۰، ۱۷۵، ۱۹۰، ۲۱۵، ۱۹۸، ۱۸٤

آیا داده ها دلیلی بر نگرانی به دست می دهند؟

تمرین(۲): از ۲۰ راننده تاکسی در یک شهر در مورد درآمد روزانهٔ آنها سئوال شده است. داده های حاصل بصورت زیرند(بر حسب هزار تومان):

۱٦۰،۱۷۰،۱٦۰،۱۵۰،۱۳۰،۲۲۰،۲۵۰،۲۱۰،۱۵۰،۱٦۰،۱۲۰،۲٦۰،۱۱۰،۸۰،۵۱۳،۱٤۰،۱٦۰،۱۵۵،۱٤۵،۱۵۰،۱٦۰

سازمان تاکسیرانی این شهر اعلام کرده است که متوسط درآمد رانندگان تاکسی در این شهر ۱۷۵ هزار تومان است. آیا این ادعا صحیح است؟

تمرین(۳): روانشناسی آزمونی را برای اندازه‌گیری مهارت‌های کلامی کودکان ٦ ساله طراحی کرد. متوسط نمره آنها ٦۰ است. او آزمون را به روی نمونه‌ای از کودکان(تک فرزند) اجرا کرد. اعداد به دست آمده از

آزمون در زیر داده شده است. در سطح ۰/۰۵ آزمون کنید که آیا مهارت‌های کلامی کودکان با مقدار متوسط(٦٠) تفاوتی دارد یا خیر؟

۶۰	۵۸	۶۲	۵۷	۶۴	۶۱	۵۹	۶۲	۵۸	۶۰
۶۰	۵۸	۶۲	۵۷	۶۳	۶۰	۶۳	۶۶	۵۶	۵۷

تمرین(٤): یک پرسشنامه مربوط به اندازه‌گیری خلق، دارای میانگین ٤٥ است. روانشناسی علاقه‌مند است تا بداند که محیط فرد تا چه اندازه بر خلق او تاثیر دارد. نمونه‌ای با اندازه ۲۰ نفر در اتاقی با رنگ تیره و بدون پنجره و هم چنین در حضور سایر عواملی مه بر خلق تاثیر منفی داشتند، به این پرسشنامه پاخ داده‌اند. نمرات مربوطه در جدول زیر داده شده است. در سطح اطمینان ۹۵ درصد آزمون کنید که آیا خلق افراد مورد بررسی با مقدار متوسط(٤٥) تفاوتی دارد یا خیر؟

آزمون فرض در مورد میانگین‌های دو جامعه مستقل:

در بررسی‌های آماری سئوالاتی به صورت زیر مطرح می‌شوند:

- آیا میانگین نمره ریاضی برای دو کلاس A و B یکسان است یا خیر؟
- آیا می توان گفت به طور متوسط قد زنان و مردان یکسان است یا خیر؟

در سئوالات فوق در حقیقت مایلیم بدانیم آیا میانگین دو جامعه مستقل، برابر است یا خیر؟

آزمون فرض مقایسه میانگین دو جامعه مستقل به صورت زیر است:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

آماره یا ملاکی که برای آزمون مقایسه میانگین دو گروه استفاده می‌شود(ملاک آزمون) به صورت زیر می‌باشد.

$$t_{obs} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

در فرمول فوق $\bar{x}_1$ و $\bar{x}_2$ به ترتیب میانگین نمونه گروه اول و دوم و n_1 و n_2 به ترتیب حجم نمونه گروه اول و دوم می‌باشند. S_p نیز انحراف معیار مشترک(ادغام شده) دو گروه است که به صورت زیر محاسبه میشود.

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

در فرمول فوق s_1^2 و s_2^2 به ترتیب واریانس گروه اول و دوم می‌باشند.

اکنون به منظور مقایسه آماری میانگین دو گروه باید شاخص t_{obs} را با مقداری از جدول توزیع t مقایسه نماییم.

اگر $|t_{obs}| > t_{\alpha/2}^{(n_1+n_2-2)}$ باشد، آنگاه فرض H_0 رد خواهد شد.

مثال(٦): فرض کنید محققی می‌خواهد بداند که آیا میانگین فشار خون بیمارانی که از داروی جدید A استفاده می‌کنند در مقایسه با بیمارانی که تحت درمان استاندارد B قرار میگیرند از نظر آماری متفاوت است یا خیر. برای این کار ۲۰ بیمار را انتخاب نموده و آنها را به صورت مساوی و به تصادف به یکی از دو گروه درمانی A یا B تخصیص میدهد. نتایج سنجش فشار خون این افراد در جدول زیر ثبت شده است.

A	۱۲۰	۱۲۲	۱۲۱	۱۳۴	۱۲۵	۱۳۲	۱۲۸	۱۴۰	۱۳۶	۱۲۷
B	۱۴۰	۱۴۲	۱۳۵	۱۳۸	۱۲۹	۱۴۶	۱۳۳	۱۴۲	۱۳۸	۱۳۴

ابتدا مسئله را به روش‌های معمول محاسباتی که در کلاس‌های درسی و در کتب درسی آورده شده است انجام خواهیم داد.

	A	B
میانگین	$\bar{X}_1 = 128.50$	$\bar{X}_2 = 137.70$
واریانس	$S_1^2 = 46.28$	$S_2^2 = 25.57$
حجم نمونه	$n_1 = 10$	$n_2 = 10$

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(9\times 46.28) + (9\times 25.57)}{10+10-2}} = 5.99$$

$$t_{obs} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{|128.5-137.7|}{5.99\sqrt{\frac{1}{10}+\frac{1}{10}}} = 3.43$$

$$t_{\alpha/2}^{(n_1-n_2-2)} = t_{0..025}^{10+10-2} = t_{0.025}^{18} = 2.10$$

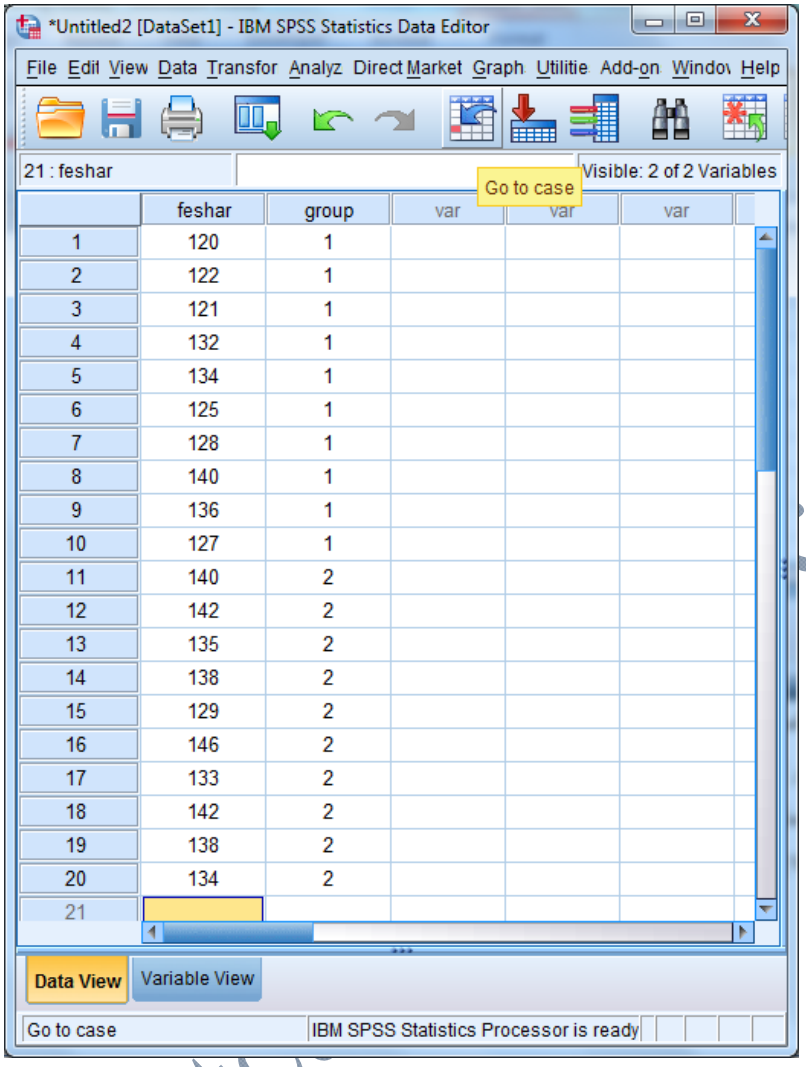
چون $t_{obs} = 3.43$ بزرگتر از مقدار t جدول است $(t^{18}_{0.025} = 2.10)$ بنابراین فرض H_0 (فرض برابری میانگین گروه‌ها) پذیرفته نخواهد شد.

اکنون مسئله بالا را با استفاده از نرم افزار انجام خواهیم داد.

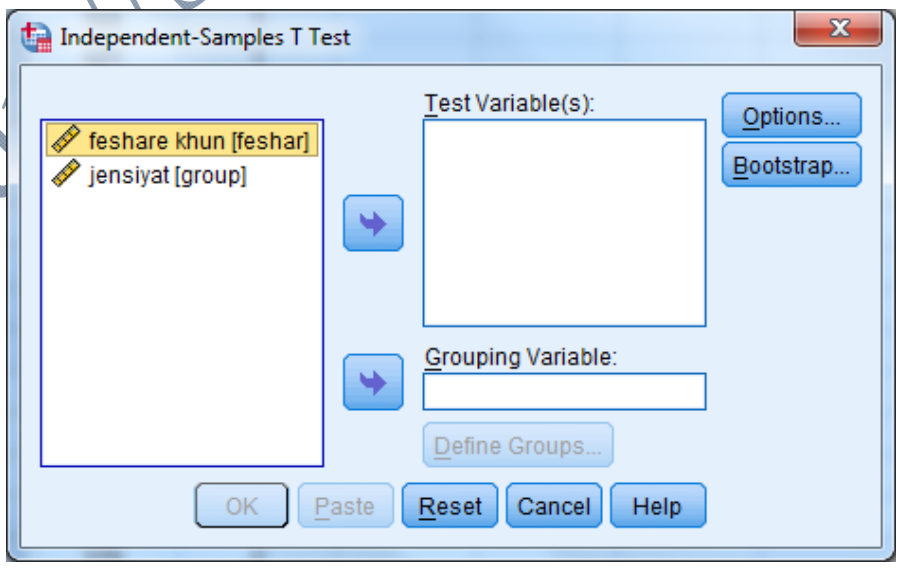
برای استفاده از آزمون مقایسه میانگین دو جامعه مستقل، فرض‌هایی در مورد جامعه حاکم است. داده‌ها باید از جامعه‌ای نرمال باشند و مهمتر اینکه باید واریانس داده‌ها در دو گروه یکسان باشد. برای مثال در داده‌های بالا باید ابتدا آزمون کنیم که آیا واریانس میانگین فشار خون دردو گروه درمانی یکسان است یا خیر؟ این فرض توسط آزمون Leven انجام می‌گیرد (SPSS خود این آزمون را انجام داده و نتیجه را در خروجی نشان می‌دهد).

روش ورود اطلاعات در نرم افزار SPSS

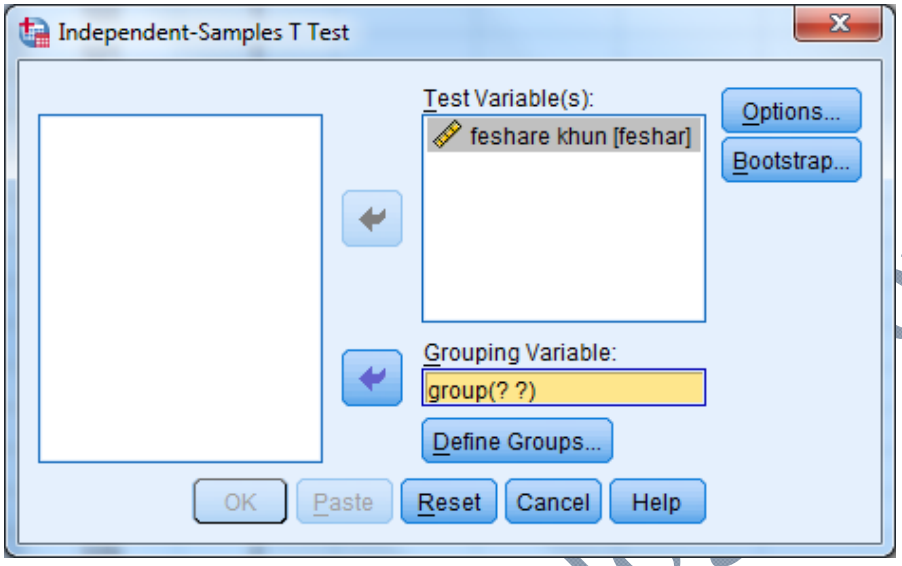
ابتدا دادهای گروه A را در ستون اول وارد نموده و سپس دادهای گروه دوم یا B را زیر آنها وارد می‌نماییم سپس متغیر دیگری (ستون دیگری) را تحت عنوان متغیر group در نظر گرفته و در آن ستون در مقابل دادهای ستون اول گروه درمانی A و B را با کدهای ۱و ۲ مشخص می‌نماییم. که در آن کد ۱ نمایانگر گروه درمانی A و کد ۲ نمایانگر گروه درمانی B است. (شکل زیر)



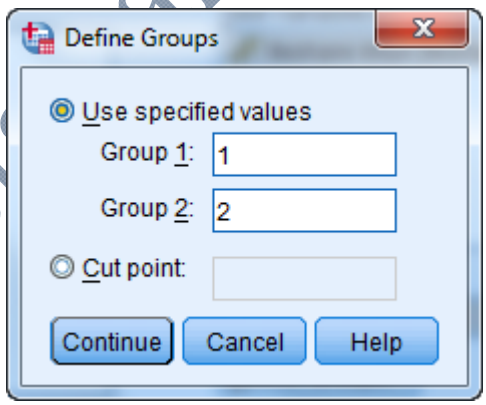
سپس مسیر Analyze / Compare means / Independent-Samples T Test.. را انتخاب کرده تا کادر زیر باز شود.



سپس متغیر پاسخ یا همان متغیر کمی(feshar) را در قسمت Test variable و متغیر کد بندی شده(jensiyat) را در قسمت Grouping variable همانند شکل زیر وارد نمایید.



در این حالت گزینه Define Groups... فعال می‌شود بر روی آن کلیک کرده تا کادر زیر باز شود. در کادر Group1: کد(۱) و در کادر Group2: کد(۲) را وارد کنید. سپس روی دکمه Continue و Ok کلیک کنید.



خروجی مربوط به این مثال در زیر داده شده است.

Group Statistics

	jensiyat	N	Mean	Std. Deviation	Std. Error Mean
feshare khun	A	10	128.50	6.803	2.151
	B	10	137.70	5.056	1.599

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
feshare khun	Equal variances assumed	1.374	.256	-3.432	18	.003	-9.200	2.680	-14.831	-3.569
	Equal variances not assumed			-3.432	16.619	.003	-9.200	2.680	-14.865	-3.535

می بینید که آزمون در دو حالت انجام شده است(سطر اول و دوم). سطر اول برای حالتی است که واریانس ها برابر فرض شده اند. دراین حالت آزمون Levene برای آزمون فرض برابری واریانس ها انجام شده است که در این حالت چون سطح معناداری آزمون لون برابر با ۰/۲۵۶ و بزرگتر از مقدار ۰/۰۵ شده است. بنابراین در سطح اطمینان ۹۵ درصد، فرض صفر برابری واریانس‌ها رد نمی‌شود، لذا نتایج سطر اول معتبر خواهد بود. سطر دوم برای حالتی است که برابری واریانس‌ها فرض نشده است. نتیجه نهایی آزمون برابری میانگین‌ها هم با توجه به سطح معناداری، تعیین می‌شود. در دو ستون آخر هم یک فاصله اطمینان ۹۵ درصدی برای تفاضل میانگین‌ها ساخته شده است.

اکنون با توجه به مقادیر به دست امده در بالا، نتیجه آزمون را تفسیر کنید و با روش قبل مقایسه کنید.

تمرین(۵): ۱۰ دانش آموز پسر و ۷ دانش آموز دختر به دلخواه انتخاب شده اند و وزن آنها اندازه گیری شده است. داده ها بصورت زیرند:

دختر	۵۴	۴۸	۶۵	۶۰	۴۵	۵۷	۴۵	۶۰	۶۳	۵۵
پسر	۷۵	۶۸	۸۰	۷۴	۷۲	۶۸	۷۰	۸۵	۹۰	۷۵

آزمون برابری وزن دختران و پسران را در سطح اطمینان ۹۵ درصد انجام دهید.

تمرین(٦): مایلیم بدانیم آیا نیروی چسبندگی دو نوع روغن تولیدی یک پالایشگاه یکسان است یا خیر؟ برای این کار از تولیدات کارخانه از هر نوع روغن ۸ قوطی انتخاب شده است که میزان چسبندگی آنها به صورت زیر است:

روغن A	۱۰/۲۹	۱۰/۲۸	۱۰/۲۷	۱۰/۳۲	۱۰/۳۰	۱۰/۲۷	۱۰/۲۸	۱۰/۲۷
روغن B	۱۰/۲٦	۱۰/۲۹	۱۰/۳۱	۱۰/۲۷	۱۰/۳۰	۱۰/۲٦	۱۰/۳۱	۱۰/۳۱

روغن B در سطح ۰/۰۱ آزمون کنید که آیا میانگین چسبندگی دو نوع روغن یکسان است؟

تمرین(۷): یک روانشناس تربیتی قصد مقایسه دو روش تدریس آمار را دارد. ۱۰ دانشجو به شیوه سنتی و با حداقل استفاده از کامپیوتر آموزش دیدند. گروه دیگر ۸ دانشجو بودند که در آزمون آنها از نرم افزارهای کامپیوتری استفاده شد. کل نمرات به دست امده در آزمون‌های مشابه به درصد تبدیل شدند. داده‌ها در جدول زیر نمایش داده شده است. در سطح ۰/۰۵ آزمون کنید که آیا بین دو روش تفاوتی وجود دارد یا خیر؟

آزمون فرض در مورد میانگین‌های دو جامعه وابسته(Paired t-test):

در برخی از آزمایش‌های مورد بررسی دو جامعه مورد بررسی وابسته هستند. این وابستگی می‌تواند بر اثر تکرار آزمایش(قبل و بعد از مداخله یک متغیر) و یا اینکه وابستگی خونی(دوقولوها) و یا نمونه‌های همتا(بیماران دیابتی یا سرطانی یا افراد با ضریب هوشی زیر ۲۰) شده باشد.

مثال (۷): برای مثال فرض کنید که هدف، بررسی اثر یک رژیم غذایی بر کاهش وزن یک گروه از افراد باشد. بر اساس آزمون t زوج شده ما ابتدا یک بار وزن افراد را قبل از رژیم غذایی اندازه گیری نموده و سپس یک بار دیگر وزن همان افراد را بعد از رژیم غذایی اندازه‌گیری می‌نماییم. در حقیقت در این روش ما هر فرد را کنترل خودش منظور می‌نماییم. این کار موجب می‌شود که اگر تفاوتی بین میانگین قبل از رژیم غذایی با میانگین بعد از رژیم غذایی مشاهد شود این تفاوت ناشی از اثر درمان باشد نه عوامل مخدوش کننده دیگر. دلیل این امر نیز آنست که چون هر فرد کنترل خودش محسوب می‌شود هنگام بررسی اثر درمان، عوامل مخدوش کننده‌ای مانند سن، جنس، تغذیه، شیوه زندگی و عوامل ژنتیکی برای هر فرد ثابت باقی مانده و در عمل تنها اثر درمان مشاهده می‌گردد. این در حالیست که در آزمون t-test ما میانگین دو نمونه مستقل را با هم مقایسه می‌نماییم که در این حالت ممکن است این دو گروه از نظر عواملی که برشمردیم با هم متفاوت بوده و هنگام مقایسه میانگین دو گروه نتایج قابل اعتمادی بدست نیاوریم. بنابراین در شرایطی که این امکان وجود دارد که هر فرد کنترل خودش منظور شود بهتر است ما از روش فوق به منظور طرح‌ریزی آزمایش خود استفاده نماییم. در هر حال ذکر این نکته ضروری است که جور سازی عملی پر هزینه و وقت گیر بوده و در مواردی عملاً انجام آن غیر ممکن است و ما باید بر اساس دو نمونه مستقل آزمایش را طرح‌ریزی نماییم. به منظور انجام آزمون t زوج شده از فرمول زیر استفاده می‌نماییم:

$$t_{obs} = \frac{|\bar{d}|}{s_d / \sqrt{n}}$$

در فرمول فوق d تفاوت مقدار قبل و بعد می‌باشد (یا برعکس) که میانگین این تفاوتها $\bar{d}$ خواهد بود. همچنین s_d انحراف معیار تفاوت‌های مشاهده شده می‌باشد که از فرمول زیر محاسبه می‌گردد.

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \qquad s_d^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}$$

اگر مقدار $t_{obs} > t_{\alpha/2}^{n-1}$ آنگاه فرض برابری میانگین‌ها قبل و بعد از مداخله رد خواهد شد. در این مقایسه-n 1 درجه آزادی توزیع t نام دارد.

بعد–قبل(d_i)	وزن بعد ازدرمان	وزن قبل از درمان
۴	۸۱	۸۵
۲	۷۲	۷۴
۱۰	۷۹	۸۹
۸	۷۰	۷۸
۰	۷۷	۷۷
۵	۷۵	۸۰
۵	۷۰	۷۵

بر اساس اطلاعات فوق $\bar{d} = ۴/۸۶$ و $s_d = ۳/۳۹$ میباشد که بر این اساس $t_{obs} = ۳/۷۹$ بدست خواهد آمد:

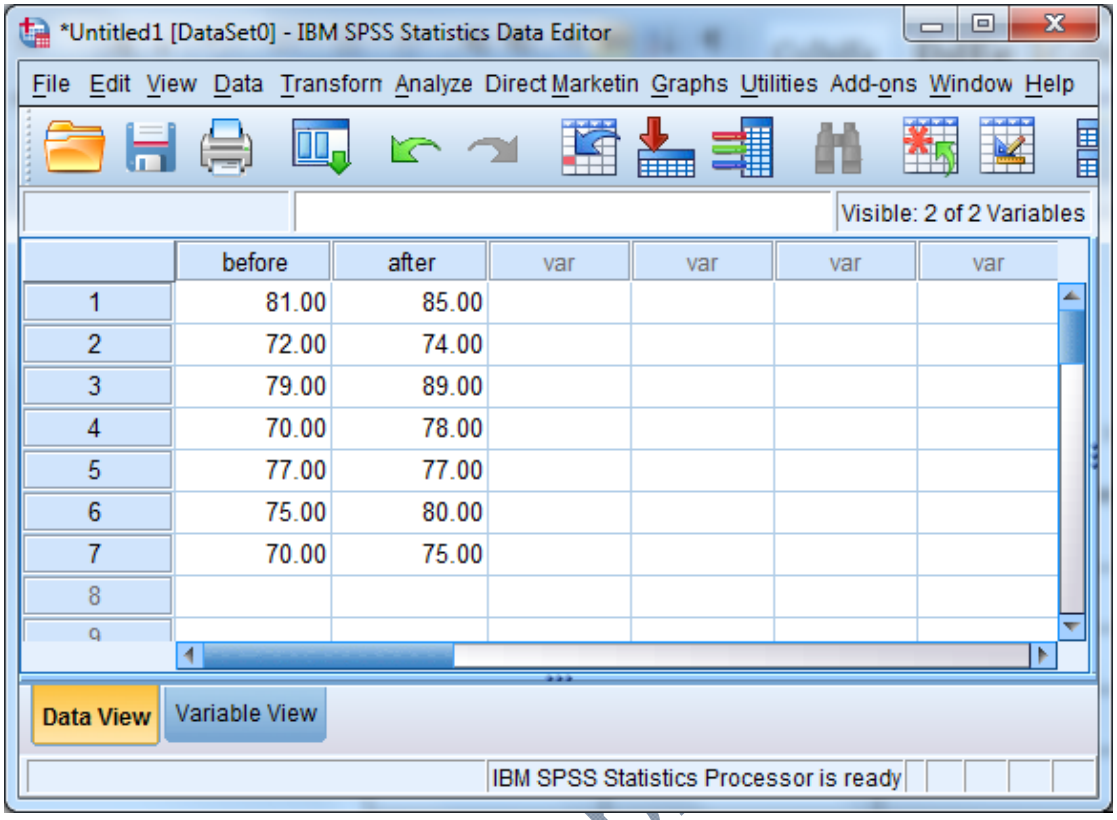
با توجه به آنکه مقدار t_{obs} از مقدار جدول بیشتر است در نتیجه دلایل کافی برای پذیرش فرض صفر وجود ندارد. $t_{obs} = \frac{|\bar{d}|}{s_d/\sqrt{n}} = \frac{۴/۸۶}{۳/۳۹/\sqrt{۷}} = ۳/۷۹$

$$t_{obs} = ۳/۷۹ > t_{\alpha/2}^{n-1} = t_{.۰۲۵}^{۶} = ۲/۳۰۶$$

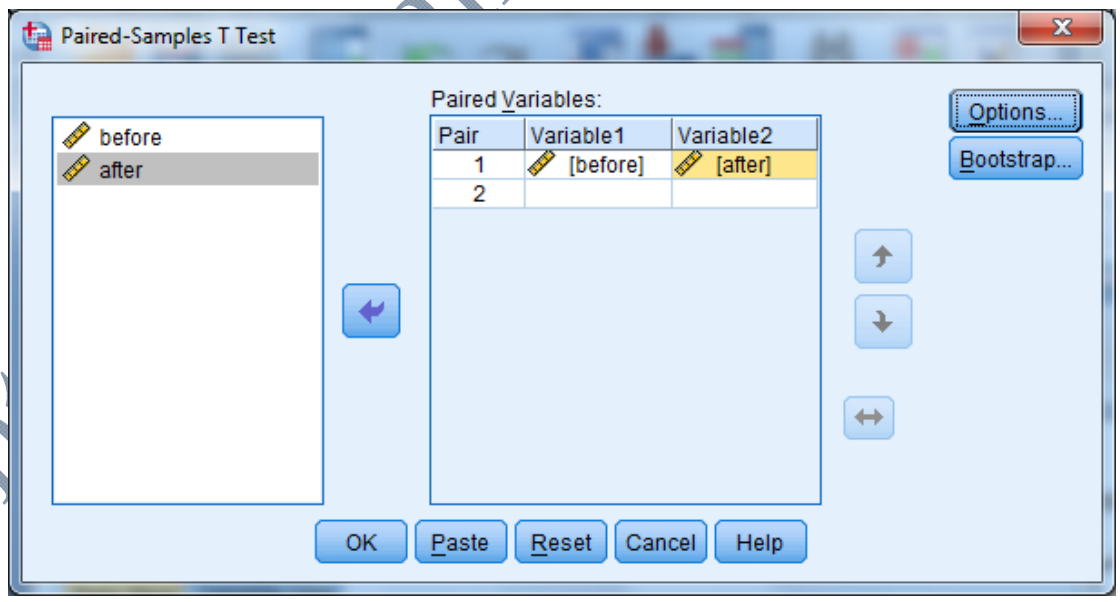
در ادامه چگونگی انجام آزمون بالا را در نرم افزار بیان خواهیم کرد.

روش ورود اطلاعات در نرم افزار SPSS

در این روش باید اطلاعات قبل و بعد در دو ستون متفاوت و در مقابل هم همانند آنچه که در جدول بالا مشاهده شد در نرم افزار spss، همانند شکل زیر وارد شوند.



سپس مسیر Analyze / Compare means / paired sample t-test را انتخاب کرده تا کادر زیر باز شود.



در این قسمت دو متغیر موجود در سمت راست (before & after)را به قسمت paired variable در سمت چپ منتقل نموده و دستور OK را اجرا نمایید.

خروجی این آزمون به صورت زیر خواهد بود.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	before	74.8571	7	4.37526	1.65369
	after	79.7143	7	5.46852	2.06691

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	before & after	7	.785	.036

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	before - after	-4.85714	3.38765	1.28041	-7.99020	-1.72409	-3.793	6	.009

با توجه به مقادیر جدول فوق نتیجه آزمون را تفسیر کنید؟

تمرین(): یک پژوهشگر روان‌شناختی ۱۰ جفت بیمار دیابتی را به دو گروه واگذار کرد. هر عضو از جفت بیمار دیابتی به گروه یوگا یا گروه مراقبه متعالی واگذار شدند. آنها بر اساس مقادیر AIC هموگلوبین جفت شدند. بعد از ٦ ماه، AIC هموگلوبین آنها تعیین شد. نتایج در جدول زیر داده شده است. در سطح اطمینان ۹۵ درصد تاثیر دو گروه یوگا و مراقبه متعالی را بر مقادیر AIC هموگلوبین بررسی کنید؟

گروه یوگا	۶/۴	۶/۶	۶/۸	۷	۷/۳	۶/۶	۷	۷/۵	۷/۹	۶/۸
گروه مراقبه متعالی	۶/۲	۶/۵	۶/۹	۶/۸	۷/۴	۶/۳	۷/۲	۷/۳	۷/۲	۶/۹

تمرین(): نمونه‌ای از بیماران دیابتی، قند خون صبحگاهی آنها اندازه‌گیری شد. آنها از رژیم غذایی خاصی به مدت یک‌ماه استفاده کردند و دوباره قند خون صبحگاهی آنها اندازه‌گیری شد. اعتقاد بر این است که رژیم غذایی حساسیت بدن به انسولین را افزایش می‌دهد. آزمون را انجام داده و نتیجه را تفسیر کنید.

قبل رژیم	۱۲۰	۱۲۵	۱۱۵	۱۳۵	۱۴۰	۱۳۲	۱۵۰	۱۲۸	۱۱۵	۱۲۶
بعد از رژیم	۱۱۵	۱۲۳	۱۲۰	۱۳۰	۱۲۵	۱۲۳	۱۳۷	۱۳۰	۱۲۵	۱۴۰

تمرین(): ظرفیت تنفس پنج نفر که به تصادف انتخاب شده بودند، قبل و بعد از معالجه‌ی معینی، اندازه‌گیری شده و داده‌های زیر بدست آمدند. آزمون برابر ظرفیت تنفس افراد را قبل و بعد از معالجه در سطح اطمینان ۹۹٪ انجام دهید.

شخص		۱	۲	۳	۴	۵
ظرفیت	قبل (x)	۲۷۵۰	۲۳۶۰	۲۹۵۰	۲۸۳۰	۲۲۵۰

تنفس	بعد (y)	۲۸۵۰	۲۳۸۰	۲۹۳۰	۲۸۶۰	۲۳۲۰

آزمون مقایسه میانگین ۳ یا بیشتر از سه گروه مستقل (ANOVA):

با استفاده از آزمون‌های آماری آزمایشاتی که دارای دو گروه مقایسه هستند را می‌توانیم بوسیله آزمون t (T-Test) مورد تجزیه و تحلیل قرار دهیم. اما اگر آزمایشی شامل بیش از دو گروه باشد، باید بین هر دو گروه از آنها با استفاده از آزمون t تعداد زیادی مقایسات دو گانه صورت گیرد که این امر علاوه بر افزایش تعداد مقایسات، امکان اینکه اختلاف بین تیمار به طور تصادفی (معنی دار) باشد را نیز افزایش می‌دهد.

روشی که برای مقایسه بیش از دو تیمار به کار می‌رود تجزیه و تحلیل واریانس(Analysis of Variance) نامیده می‌شود. از مزایای استفاده از این آزمون این است که تنها با انجام یک بار آزمودن، اختلاف میان میانگین‌های کلیه تیمارهای موجود در آزمایش، مورد بررسی قرار می‌گیرد.

هدف از آزمون بررسی زیر می باشد:

$H_0: \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k$

$H_1: \mu_i \neq \mu_j$ حداقل یکی از μ_i با سایر آنها تفاوت داشته باشد

این روش بر مبنای آزمون F انجام خواهد شد. آزمون مقایسه میانگین سه گروه یا بیشتر بر مبنای جدول زیر که جدول تحلیل واریانس نام دارد میباشد.

منبع	مجموع توانهای دوم	درجه آزادی	MS(Mean square)	F
بین گروهها	$SSB = \sum_{j=1}^{k} n_j(\bar{y}_i - \bar{y})^2$	K-1	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSE}$
درون گروهها	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_i)^2$	N-K	$MSE = \frac{SSE}{N-K}$	
کل	$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$	N-1		

اگر در این جدول مقدار F مشاهده شده بزرگتر از مقدار بحرانی جدول F باشد فرض صفر رد خواهد شد. یعنی اگر $F_{obs} > F_{\alpha}(k-1, N-k)$ باشد آنگاه فرض صفر رد می‌شود. البته در نرم افزار SPSS علاوه بر مقدار آماره F احتمال معنی داری یا همان P-value نیز گزارش می‌گردد. درچنین شرایطی P-value کوچکتر از α (0.05 یا 0.01) معنی دار در نظر گرفته خواهد شد. در هر حال در هنگام استفاده از جدول

ANOVA باید به این نکته مهم توجه داشته باشیم که اگر $F_{obs} > F_{\alpha}(k-1, N-k)$ یا $P-value < 0.05$ باشد ما تنها نتیجه می‌گیریم که حداقل دو گروه از گروه‌های مورد مطالعه از نظر آماری تفاوت معنی‌داری را نشان می‌دهند اما نمی‌توانیم تشخیص دهیم که این گروه‌هایی که از نظر آماری متفاوتند کدام گروه‌ها خواهند بود. در چنین شرایطی ما نیاز داریم که از آزمون‌های تعقیبی(پسین) به منظور شناسایی این تفاوت‌ها استفاده نماییم.

آزمون های پسین(Post Hoc)

بر اساس توضیحات بالا اگر جدول ANOVA نشان دهد که تفاوت معنی‌داری بین میانگین گروه‌های مورد مطالعه وجود دارد آنگاه علاقه‌مند هستیم که بدانیم کدام گروه‌ها با هم تفاوت معنی داری دارند. انتخاب آزمون‌های پسین بستگی به نظر پژوهشگر دارد. مهمترین این آزمون‌ها در جدول زیر نمایش داده شده‌اند.

آزمون	توان	خطا
LSD (least significant difference)	↑	↑
نیومن-کولز		
دانکن(Duncan)		
توکی(Tukey)		
شفه (Scheffe)		

همانطور که در جدول فوق مشاهده می شود، این آزمونها بر اساس توان و خطایشان مورد استفاده قرار می-گیرند.

توان: احتمال کشف تفاوت‌ها وقتیکه واقعا متفاوت هستند.

خطا: احتمال کشف تفاوت‌ها وقتی واقعا متفاوت نیستند.

بنابراین اگر شما پژوهشگری هستید که به کوچکترین تفاوت‌ها حساس هستید آزمون LSD را انتخاب کنید زیرا این آزمون به کوچکترین تفاوتها حساس است (توان آن بالاتر از سایر آزمون‌های پسین می‌باشد). اما بدانید که بر اساس این آزمون ممکن است گروه‌هایی که واقعاً متفاوت نیستند با این آزمون معنی‌دار نشان داده شوند (به این مفهوم که علی رغم آنکه این آزمون توان بالاتری نسبت به سایر آزمون‌ها دارد اما خطای آن هم بیشتر است). اما آزمون شفه از همه این آزمون‌ها محافظه کارانه تر است به این مفهوم که اگر شما پژوهشگری هستید که تنها به دنبال گروه هایی هستید که واقعا متفاوتند(به عبارتی محافظه کار هستید

و نمی‌خواهید هر تفاوتی را معنی‌دار بیان کنید) از آزمون شفه استفاده نمایید. این آزمون تفاوت‌های معنی‌دار کمتری را نشان می‌دهد (یعنی توان آن نسبت به سایر آزمون‌ها کمتر است) اما خطای آن هم کمتر است. به عبارتی اگر تفاوتی را نشان می‌دهد احتمالا این تفاوت واقعاً وجود داشته که با این آزمون مشخص شده است. به عبارتی این آزمون هر تفاوتی را معنی‌دار نشان نمی‌دهد.

روش ورود داده‌ها برای انجام آزمون ANOVA در نرم افزار SPSS

برای این کار ابتدا دو فرض را با هم مقایسه می‌کنیم. اگر فرض H_o پذیرفته شود که تجزیه و تحلیل به پایان می‌رسد و نشان دهنده این موضوع می‌باشد که میان تیمارهای (میانگین‌های) گروه‌های مورد بررسی تفاوتی وجود ندارد.

اما اگر فرض H_o رد شود نشان دهنده اختلاف میان تیمارها می‌باشد و باید بدنبال اختلاف‌ها بگردیم.

مثال(۸): متوسط زمان بستری شدن بیماران در شهریور ماه برای یک بیماری خاص در ۵ بیمارستان به صورت زیر است. بررسی کنید که آیا میان متوسط زمان بستری شدن(به روز) بیماران ۵ بیمارستان تفاوت معنی‌داری وجود دارد یا نه؟ در صورت وجود اختلاف نشان دهید که میان کدام بیمارستان‌ها در این زمینه تفاوت وجود دارد.

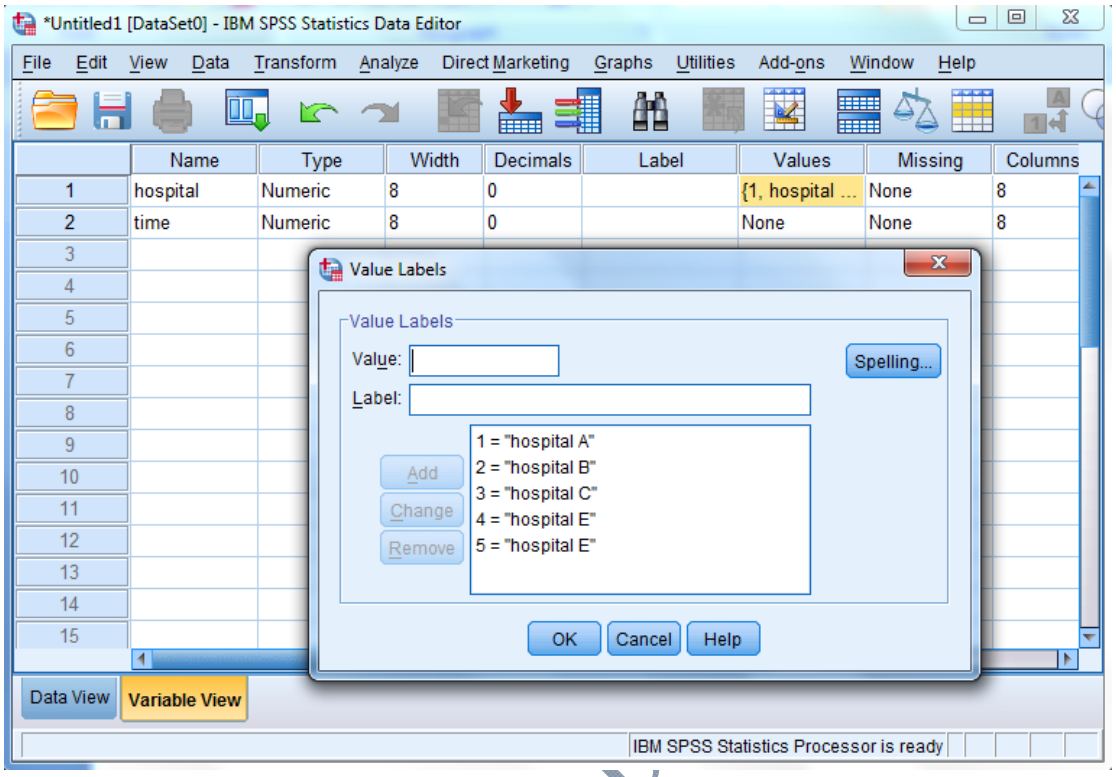
تیمار= متوسط زمان بستری شدن بیماران

نوع بیمارستان	متوسط زمان بستری شدن(به روز)									
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
بیمارستان A	۷	۷	۸	۶	۷	۵	۸	۷	۶	۸
بیمارستان B	۸	۸	۸	۸	۷	۷	۶	۶	۶	۵
بیمارستان C	۷	۵	۵	۵	۴	۷	۴	۴	۵	۵
بیمارستانD	۸	۹	۹	۱۱	۶	۱۰	۱۱	۱۱	۱۰	۱۲
بیمارستانE	۴	۹	۶	۴	۴	۴	۵	۵	۴	۶

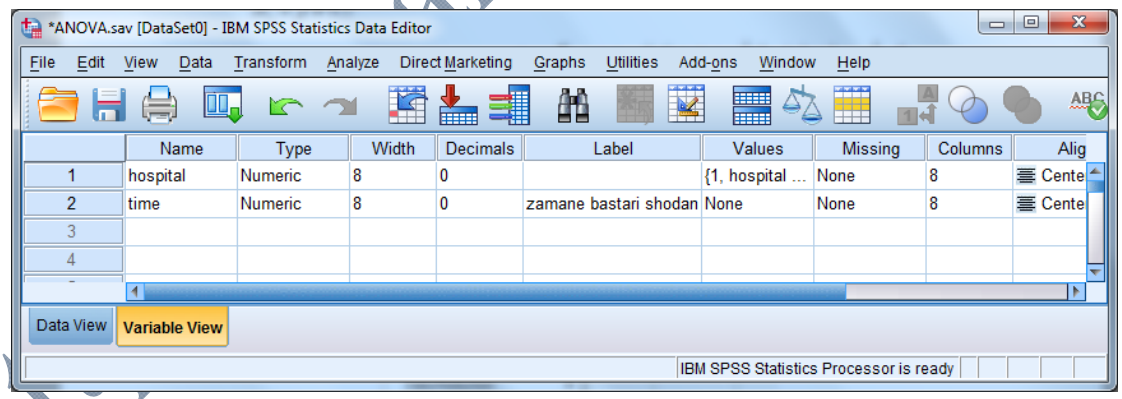
برای انجام آزمون در نرم افزار ابتدا در پنجره Variable View دو متغیر به نام نوع بیمارستان(hospital) و دیگری زمان بستری شدن بیمار(time) تعریف می‌کنیم و سپس مانند توضیحاتی که در آغاز گفته شد ستون‌های مورد نظر را متناسب با نوع متغیر تنظیم می‌کنیم.

در این قسمت برای متغیر نوع بیمارستان در ستون Values نوع بیمارستان‌ها را تعریف می‌کنیم.

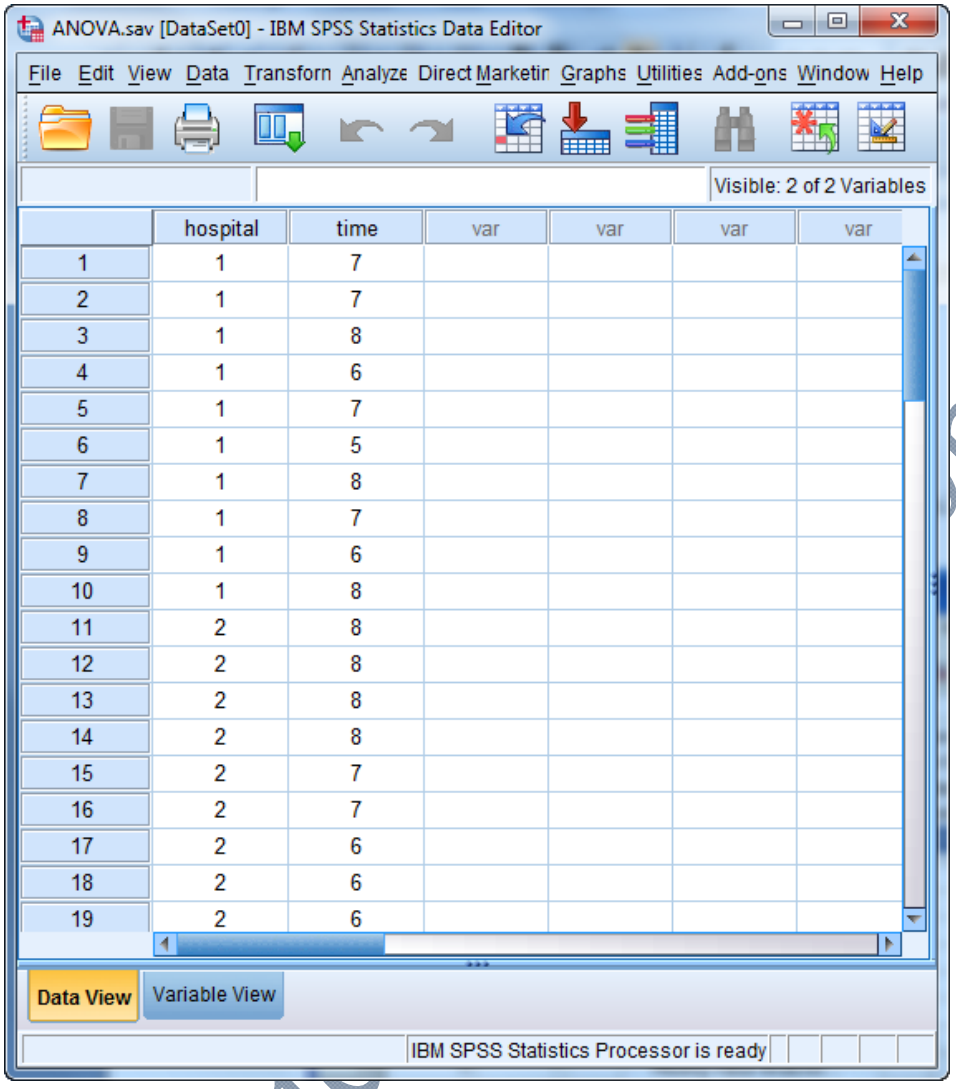
(در پنجره Data View به جای اسامی بیمارستان از کدهای (۱ و ۲ و.....) استفاده می‌کنیم).



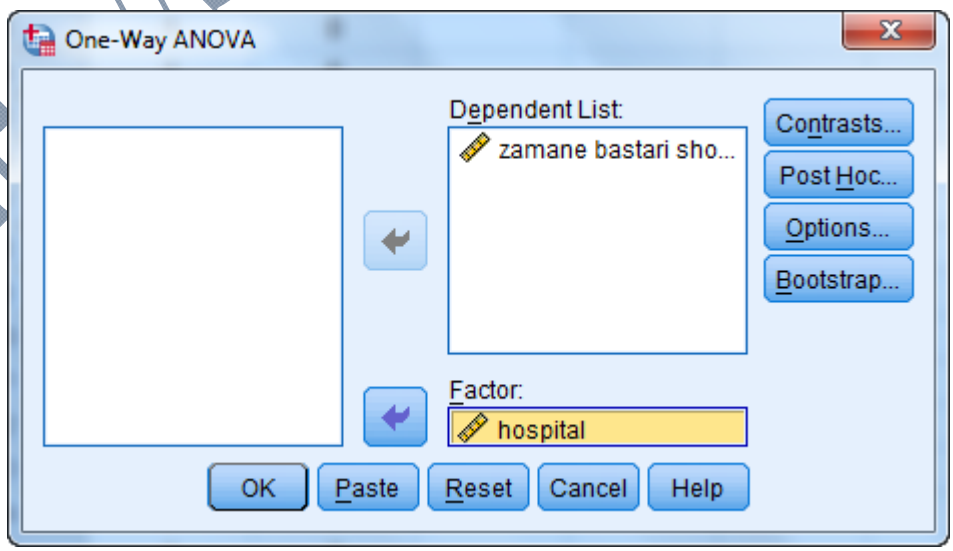
و در پایان با کلیک کردن بر روی ok پنجره Variable View به صورت زیر در می‌آید.



در پنجره Data View داده‌ها را به صورت زیر وارد می‌کنیم. در ستون بیمارستان کد مربوط به بیمارستانها را (۱ تا ۵) وارد کرده و جلوی هر کد در ستون زمان، مدت زمان بستری شدن بیماران بیمارستان‌های مختلف را طبق جدول داده‌ها وارد می‌کنیم.

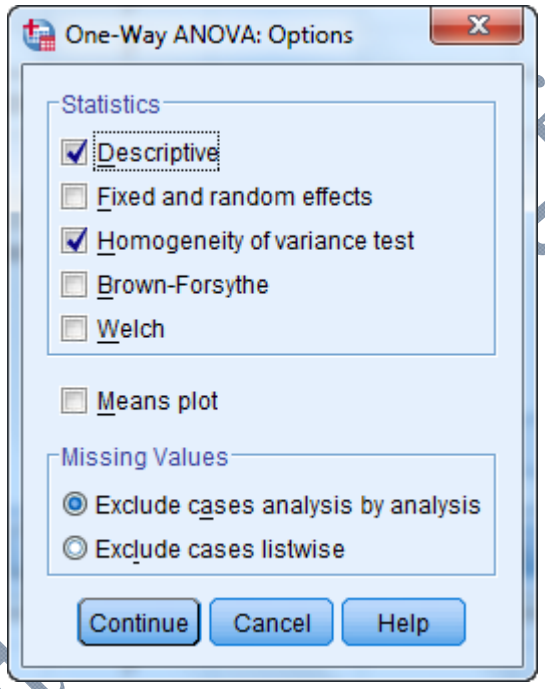


برای انجام آزمون مسیر Analyze / Compare means / One-Way ANOVA.... را انتخاب تا کادر بعدی باز شود.



سپس متغیر time را به کادر Dependent List و متغیر hospital را به کادر Factor انتقال دهید.

قبل از اینکه بر روی کلمه ok کلیک کنیم برای بررسی اینکه آیا بین واریانسهای (مدت زمان بستری شدن بیماران) ۵ بیمارستان تفاوت وجود دارد یا خیر بر روی کلمه option کلیک می کنیم تا پنجره زیر باز شود. در پنجره One-Way ANOVA:Option گزینه Homogeneity of variance test و Discriptive را فعال کرده، در ادامه ابتدا بر روی کلمه Continue و سپس Ok کلیک می‌کنیم تا خروجی های زیر بدست آیند.



Descriptives(1)

zamane bastari shodan

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
hospital A	10	6.90	.994	.314	6.19	7.61	5	8
hospital B	10	6.90	1.101	.348	6.11	7.69	5	8
hospital C	10	5.10	1.101	.348	4.31	5.89	4	7
hospital E	10	9.70	1.767	.559	8.44	10.96	6	12
hospital E	10	5.10	1.595	.504	3.96	6.24	4	9
Total	50	6.74	2.136	.302	6.13	7.35	4	12

Test of Homogeneity of Variances(2)

zamane bastari shodan

Levene Statistic	df1	df2	Sig.
1.031	4	45	.402

ANOVA(3)
zamane bastari shodan

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	141.920	4	35.480	19.542	.000
Within Groups	81.700	45	1.816		
Total	223.620	49			

جدول(۱) آماره‌های توصیفی مدت زمان بستری شدن را به تفکیک ۵ بیمارستان به ما می‌دهد. نتایج بدست آمده از جدول (۲) نشان می‌دهد که در آزمون مقایسه بین واریانس‌های مدت زمان بستری شدن ۵ بیمارستان اختلاف معنی‌داری از نظر آماری وجود ندارد (P-value>0.05). اما نتایج بدست آمده از جدول (۳) نشان می‌دهد که میان میانگین‌های مدت زمان بستری شدن ۵ بیمارستان اختلاف معنی‌داری وجود دارد.(P-value <0.05).

به دلیل اینکه داده‌ها نشان می‌دهند که میانگین‌های ۵ بیمارستان با هم تفاوت معنی‌داری دارند در نتیجه به دنبال اختلاف‌ها می‌باشیم.

بدین منظور مسیر بالا را دوباره تکرار می‌کنیم و به جای کلیک بر روی گزینه Option گزینه Post Hoc را انتخاب می کنیم تا پنجره One- Way ANOVA: Post Hoc Multiple Comparisons باز شود.

در این پنجره انواع آزمون‌هایی را که می‌توانیم برای مقایسه میانگین‌ها مورد استفاده قرار دهیم آورده شده است.

این پنجره از دو بخش تقسیم شده است.

قسمت بالا مربوط به آزمون‌های مورد استفاده در حالتی که واریانس جوامع تفاوتی نداشته باشند: (Equal Variances Assumed)

قسمت پایین مربوط به آزمون‌های مورد استفاده در حالتی که واریانس جوامع متفاوت باشند: (Equal Variances Not Assumed)

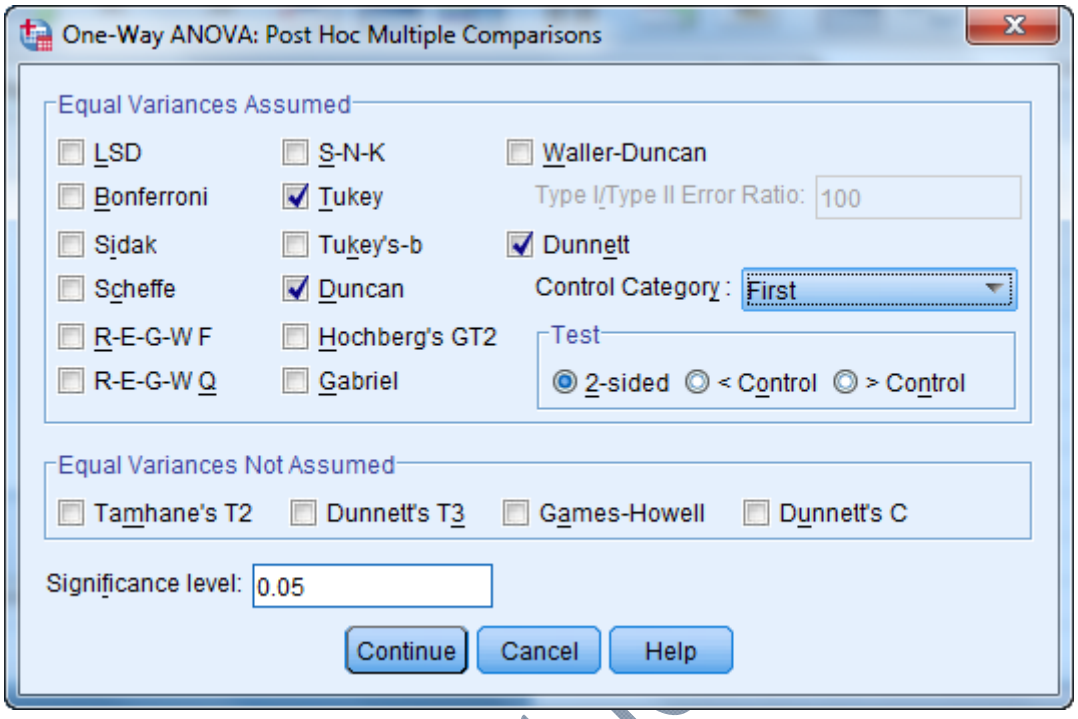
در این مثال چون فرض همگنی واریانس‌ها پذیرفته شد، به بیان دیگر فرض H_o که برابری واریانس‌ها را مطرح می‌کند رد نشده شده است از آزمون‌های بالایی استفاده می‌کنیم.

در این قسمت چند مورد از مهمترین آزمونها را مورد بررسی قرار می دهیم.

مسیر انتخابی به صورت زیر خواهد بود:

Analyze / Compare means / One-Way ANOVA / Post Hoc…

پس از انتخاب مسیر بالا کادر زیر باز خواهد شد.



در پنجره بالا در حالتی که واریانس گروه ها با هم اختلاف معنی داری ندارند به طور نمونه سه آزمون متداول (Dunnett ،Duncan،Tukey) را برای تجزیه و تحلیل آماری انتخاب می کنیم.

در مورد آزمون Dunnett این نکته را باید مورد توجه قرار داد که از میان یکی از گروه‌ها (۵ بیمارستان) یکی را به عنوان گروه کنترل(شاهد) در نظر می‌گیریم تا سایر گروه‌ها را با آن بسنجند. این گروه می‌تواند گروه اول(First) یا گروه آخر(Last) باشد. انتخاب هر کدام از این گروه‌ها به عنوان گروه اول یا آخر در نتایج تغییری ایجاد نمی‌کند. برای این کار با فعال کردن آزمون Dunnett گزینه Control Category فعال شده و در مربع روبروی آن گروه کنترل را انتخاب می‌کنیم. در مرحله بعد با کلیک بر روی Continue به پنجره قبلی می‌رویم و با کلیک بر روی Ok می توانیم خروجی‌های مورد نظر را مشاهده کنیم.

Multiple Comparisons							
Dependent Variable: zamane bastari shodan							
	(I) hospital	(J) hospital	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	hospital A	hospital B	.000	.603	1.000	-1.71	1.71
		hospital C	1.800*	.603	.035	.09	3.51
		hospital E	-2.800*	.603	.000	-4.51	-1.09
		hospital E	1.800*	.603	.035	.09	3.51
	hospital B	hospital A	.000	.603	1.000	-1.71	1.71
		hospital C	1.800*	.603	.035	.09	3.51
		hospital E	-2.800*	.603	.000	-4.51	-1.09
		hospital E	1.800*	.603	.035	.09	3.51
	hospital C	hospital A	-1.800*	.603	.035	-3.51	-.09
		hospital B	-1.800*	.603	.035	-3.51	-.09
		hospital E	-4.600*	.603	.000	-6.31	-2.89
		hospital E	.000	.603	1.000	-1.71	1.71
	hospital E	hospital A	2.800*	.603	.000	1.09	4.51
		hospital B	2.800*	.603	.000	1.09	4.51
		hospital C	4.600*	.603	.000	2.89	6.31
		hospital E	4.600*	.603	.000	2.89	6.31
	hospital E	hospital A	-1.800*	.603	.035	-3.51	-.09
		hospital B	-1.800*	.603	.035	-3.51	-.09
		hospital C	.000	.603	1.000	-1.71	1.71
		hospital E	-4.600*	.603	.000	-6.31	-2.89
Dunnett t (2-sided)[b]	hospital B	hospital A	.000	.603	1.000	-1.53	1.53
	hospital C	hospital A	-1.800*	.603	.016	-3.33	-.27
	hospital E	hospital A	2.800*	.603	.000	1.27	4.33
	hospital E	hospital A	-1.800*	.603	.016	-3.33	-.27
*. The mean difference is significant at the 0.05 level.							
b. Dunnett t-tests treat one group as a control, and compare all other groups against it.							

Homogeneous Subsets

zamane bastari shodan					
	hospital	N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD[a]	hospital C	10	5.10		
	hospital E	10	5.10		
	hospital A	10		6.90	
	hospital B	10		6.90	
	hospital E	10			9.70
	Sig.		1.000	1.000	1.000
Duncan[a]	hospital C	10	5.10		
	hospital E	10	5.10		
	hospital A	10		6.90	
	hospital B	10		6.90	
	hospital E	10			9.70
	Sig.		1.000	1.000	1.000
Means for groups in homogeneous subsets are displayed.					
a. Uses Harmonic Mean Sample Size = 10.000.					

برای نتیجه گرفتن درباره وضعیت برابری یا عدم برابری میانگین‌ها از دو ستون 95% Confidence Interval و Sig استفاده می کنیم. عدد مشاهده شده در ستون Sig معرف P-Valuse بدست آمده در آزمون‌ها می‌باشد و چون آزمون‌ها در سطح ۰/۰۵ مورد بررسی قرار می‌گیرند مبنای پذیرش یا عدم پذیرش آنها یعنی قبول یا رد فرض اولیه H_o مقایسه با مقدار ۰/۰۵ است.
در آزمون‌های یک دامنه اگر $Sig <$ ۰/۰۵ فرض اولیه H_o رد می شود و اگر $Sig >$ ۰/۰۵ فرض اولیه H_o رد نمی شود. اما در آزمونهای دو دامنه به جای ۰/۰۵ از ۰/۰۲۵ استفاده می شود.

هم ارز با ستون Sig ستون مربوط به فاصله اطمینان(95% Confidence Interval) است که نمایش دهنده یک فاصله اطمینان ۹۵ درصدی و همچنین تأییدی بر نتایج بدست آمده در ستون Sig می باشد.
اگر p-values بدست آمده را با α نشان دهیم رابطه زیر برقرار است:

$$P(\text{فاصله اطمینان}) = ۱ - \alpha$$

بنابراین وقتی مقدار α یعنی سطح معنی‌داری برابر ۰/۰۵ باشد فاصله اطمینان ۰/۹۵ می‌شود.
این فاصله اطمینان یک حد پایین(L) و یک حد بالا(U) دارد و اگر عدد صفر را شامل شود(L<0<U) نشان دهنده این است که فرض H_o یعنی برابری میانگین‌ها رد نمی‌شود و این هم ارز $Sig >$ ۰/۰۵ می‌باشد و اگر این بازه شامل صفر نباشد هم ارز این است که $Sig <$ ۰/۰۵ و بیان می‌کند که فرض H_o یعنی برابری میانگین‌ها رد می‌شود.
در خروجی مربوط به آزمون Dunnett مشاهده می‌کنیم که؛ بیمارستان A که به عنوان گروه کنترل در نظر گرفته شد با تک تک بیمارستان‌ها از نظر میانگین مدت زمان بستری شدن مقایسه آماری شد و نتایج در جدول آورده شده است. در ستون مربوط به Sig که مقادیر P-Valuse را برای هر ازمون جداگانه نشان می‌دهد، هر جا $Sig <$ ۰/۰۲۵باشد نشان می‌دهد که فرض برابری دو میانگین رد شده است. به بیان دیگر بین میانگین مدت زمان بستری شدن برای دو بیمارستان تفاوت معنی داری وجود دارد. علت استفاده از $Sig <$ ۰/۰۲۵به جای $Sig <$ ۰/۰۵در این است که آزمون Dunnett یک ازمون دو دامنه است(2-Side).
نتایج نشان می‌دهند که میانگین بیمارستان A با بیمارستان B تفاوت ندارد($Sig >$ ۰/۰۲۵). اما میان میانگین بیمارستان A با میانگین سایر بیمارستان‌ها تفاوت معنی‌داری وجود دارد($Sig <$ ۰/۰۲۵).

در خروجی مربوط به آزمون Tukey تک تک بیمارستان‌ها با هم مقایسه می‌شوند. در این آزمون به علت یک دامنه بودن هر کجا $Sig <$ ۰/۰۵ فرض برابری میانگین‌ها رد می‌شود (به بیان دیگر بین میانگین مدت زمان بستری شدن دو بیمارستان از نظر آماری تفاوت معنی‌داری وجود دارد.)

در خروجی مربوط به زیرمجموعه های همگن (Homogeneous Subsets) آزمونهای توکی و دانکن میانگین‌هایی که با هم تفاوت ندارند را در یک زیر گروه قرار می‌دهند.

به طور معمول نتایج بدست آمده از آزمون‌ها هر کدام تأییدی بر نتیجه آزمون دیگر است. البته با توجه به درجه دقت آزمون‌ها در بعضی مواقع ممکن است که نتیجه بدست آمده در یک آزمون با نتیجه بدست آمده در آزمون دیگر متفاوت باشد.

فصل چهارم:

آزمون‌های ناپارامتری

آزمون‌های آماری پارامتری(Parametric Tests):

آزمون‌های آماری پارامتری را می‌توان موثرترین نوع آزمون‌ها دانست. برای استفاده از آزمون‌های پارامتری، پیش فرض‌هایی لازم است که باید در اجرا رعایت کنیم. این پیش‌فرض ها به چگونگی توزیع نمرات داده-ها و نوع مقیاس مورد استفاده بستگی دارد و به صورت زیر هستند:

۱- هر یک از نمونه‌ها مستقل از هم هستند. یعنی انتخاب یک نمونه، به انتخاب هیچ نمونه دیگری وابسته نیست.

۲- واریانس نمونه‌ها برابر یا تقریباً برابر است. این مطلب زمانیکه حجم نمونه کم است از اهمیت خاصی برخوردار است.

۳- داده‌ها در سطح سنجش فاصله‌ای و نسبی می‌باشند(یعنی کمی هستند). داده‌های اسمی و ترتیبی برای استفاده از آزمون‌های پارامتری مناسب نیستند.

٤- توزیع داده‌ها در جامعه، نرمال و یا نزدیک به نرمال است.

۵- تمامی آزمون‌های آماری پارامتری، میانگین یک یا چند متغیر را در یک گروه یا بیش‌تر مقایسه می‌کنند. زمانیکه مفروضات فوق برای یک مجموعه از داده برقرار نباشد از آزمون‌های ناپارامتری استفاده خواهیم کرد.

آزمون‌های آماری ناپارامتری(Non-Parametric Tests):

برای استفاده از آزمون‌های ناپارامتری، رعایت پیش فرض‌های زیر در خصوص نمره داده‌ها و نوع مقیاس مورد استفاده لازم است:

۱- نرمال بودن توزیع جامعه‌ای که نمونه از ان انتخاب شده معلوم نمی‌باشد.

۲- داده‌ها در سطح سنجش اسمی و ترتیبی هستند(داده‌های کیفی). داده‌های فاصله‌ای و نسبی برای استفاده از آزمون‌های ناپارامتری مناسب نیستند.

۳- تمامی آزمون‌های ناپارامتری، میانه یک یا چند متغیر را در یک گروه یا بیشتر مقایسه می‌کند.

در SPSS انواع گوناگونی از آزمون‌های ناپارامتری برای تحلیل‌های تک نمونه‌ای، دونمونه‌ای و چند نمونه‌ای وجود دارد. اینها عموماً دارای پارامتری معادل هستند. در جدول زیر آزمون‌های پارامتری و معادل آنها آورده شده است.

تعداد گروه‌های مورد مقایسه	وضعیت گروه‌ها	آزمون پارامتری	آزمون ناپارامتری
یک گروه	----	آزمون t تک نمونه‌ای	آزمون دوجمله‌ای آزمون تصادفی بودن آزمون χ^2 یک متغیره آزمون کلموگروف-اسمیرنف یک متغیره
دو گروه	مستقل	آزمون t مستقل	من-ویتنی
	وابسته	آزمون t وابسته	آزمون علامت ویلکاکسون آزمون مک نمار
سه گروه و بیشتر	مستقل	ANOVA	آزمون میانه آزمون کروسکال والیس
	وابسته	-----	آزمون کوکران آزمون فریدمن آزمون رتبه‌های دبلیو کندال

در ادامه به بررسی آزمون‌های بالا خواهیم پرداخت:

آزمون دوجمله‌ای/ نسبت(Binomial test):

در برخی از فرضیه‌ها، با متغیرهایی سروکار داریم که دو وجهی هستند. یعنی دو مقوله یا طبقه دارند و درصدد هستیم تا نسبت این مقوله‌ها و طبقات را با همدیگر و با توجه به یک نسبت فرضی مقایسه کنیم. برای آزمون این فرضیه‌ها در SPSS از آزمون ناپارامتری دوجمله‌ای استفاده می‌شود. آزمون دوجمله‌ای مشخص می‌کند که آیا نسبت مشاهده شده با نسبت مفروض فرق دارد یا خیر؟

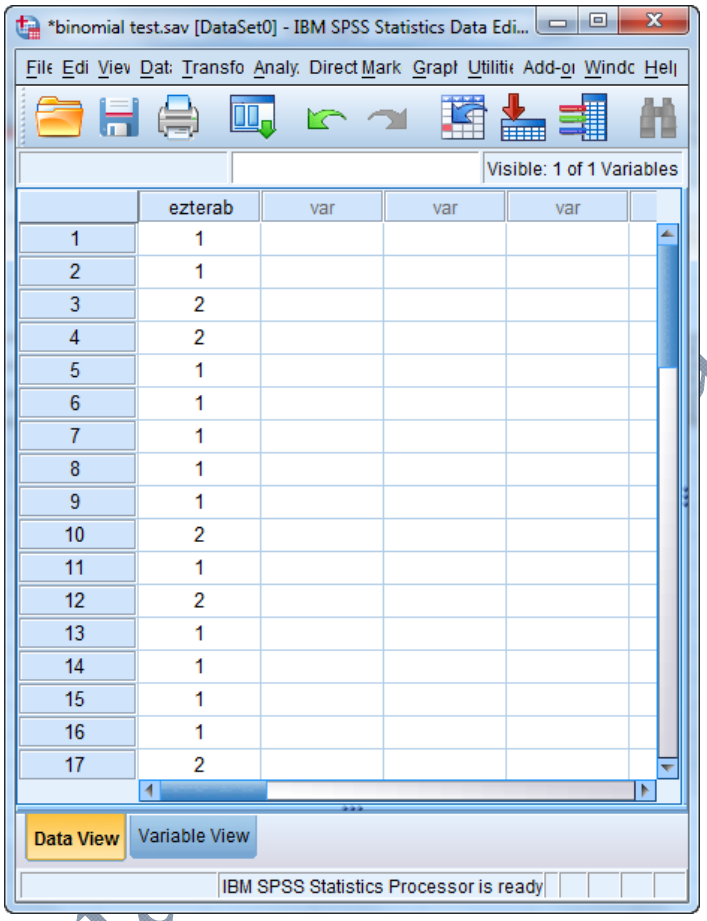
مثال(۹): پژوهشگری در یکی از تحقیقات خود پیرامون میزان اضطراب شغلی دانشجویان، در یکی از مفروضات خود آورده است که نسبت بالایی(۷۰ درصد) از دانشجویان دارای اضطراب شغلی بالایی هستند. پژوهشگر برای بررسی فرضیه خود تعداد ۳۰ دانشجو را به تصادف انتخاب و پس از بررسی آنها را در دو طبقه اضطراب شغلی بالا(با کد ۱) و اضطراب شغلی پایین(با کد ۲) قرار داده است. نتایج در جدول زیر داده شده است. ادعای پژوهشگر را در سطح ۰/۰۵ آزمون کنید؟

۱	۱	۲	۲	۱	۱	۱	۱	۱	۲
۱	۲	۱	۱	۱	۱	۲	۱	۱	۱

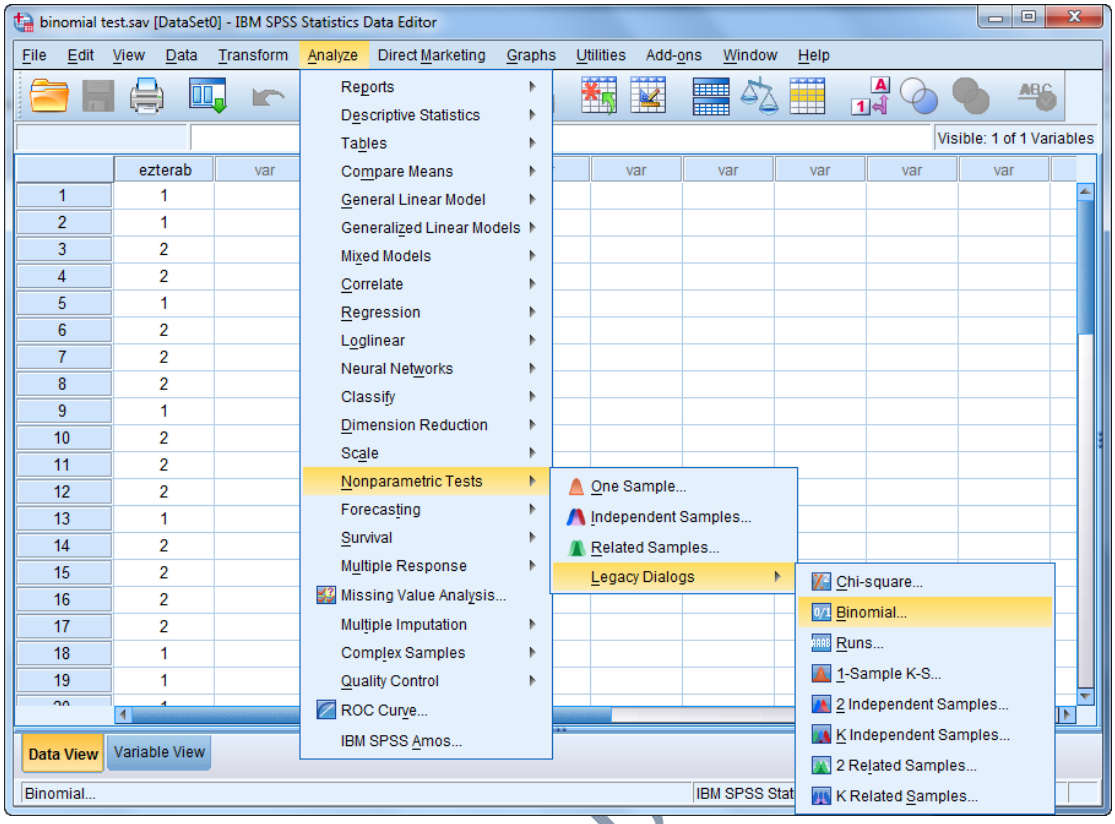
۱	۲	۲	۱	۲	۱	۱	۱	۱	۱

روش ورود اطلاعات در نرم افزار SPSS

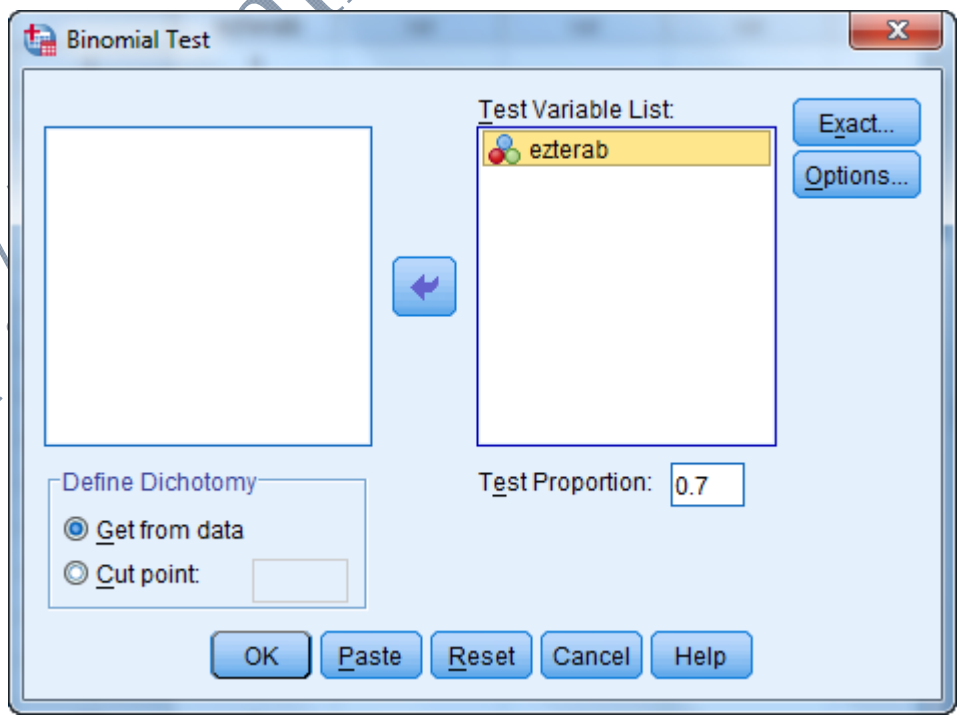
ابتدا تمامی داده‌ها را در یک ستون تحت عنوان ezterab همانند شکل زیر وارد می‌کنیم.



مسیر Analyze / Nonparametric Tests /Legacy Dialogs / Binomial… را انتخاب کنید تا کادر زیر باز شود.



متغیر مورد آزمون(ezterab) را وارد کادر Test Variable List می‌کنیم. در قسمت Test Proportion مقدار نسبت مورد آزمون که در این مثال برابر با ۰/۷ است را می‌نویسیم.



دکمه Ok را می‌زنیم. خروجی به صورت زیر خواهد بود.

NPar Tests

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
ezterab	Group 1	ezterabe shoghli bala	23	.8	.7	.432
	Group 2	ezterabe shoghli paeen	7	.2		
	Total		30	1.0		

همانطور که از خروجی بالا مشخص است، نسبت مشاهده شده برای افراد با اضطراب شغلی بالا برابر با ۰/۸ است. همچنین مقدار .Sig برابر با ۰/٤۳۲ شده که بزرگتر از مقدار ۰/۰۵ است. بنابراین می‌توان در سطح اطمینان ۹۵ درصد پذیرفت که نسبت بالایی از دانشجویان دارای اضطراب شغلی بالایی هستند.

تمرین(): وزن ۲٤ دانشجو که به تصادف از بین دانشجویان یک دانشگاه انتخاب شده‌اند، در جدول زیر داده شده است. آیا می‌توان گفت که؛ ٦۰ درصد از دانشجویان این دانشگاه کمتر از ٦۵ کیلوگرم وزن دارند؟

۵۹/۶	۶۹/۲	۵۸/۳۱	۶۰/۶	۷۵	۶۳/۹	۷۴/۵	۶۵/۳۸
۷۲/۸	۶۶/۶	۶۷	۸۴	۶۹/۳	۷۸/۹	۶۷/۸	۶۲/۷
۷۰/۸	۵۹/۸	۶۵/۳	۶۵	۷۰/۳	۶۳/۸	۶۴	۵۸/۵

آزمون تصادفی بودن(Runs):

در بسیاری از آنالیزهای آماری یکی از فرض‌های اساسی "تصادفی بودن مشاهدات" است. برای مثال در بررسی مانده‌های یک مدل آماری، برقراری فرض تصادفی بودن (استقلال مشاهدات) حیاتی است. انحراف از فرض تصادفی بودن می‌تواند به دلیل وجود روندهای افزایشی یا کاهشی، رفتارهای دوره‌ای یا افزایش تغییرپذیری و برخی علل دیگر باشد.

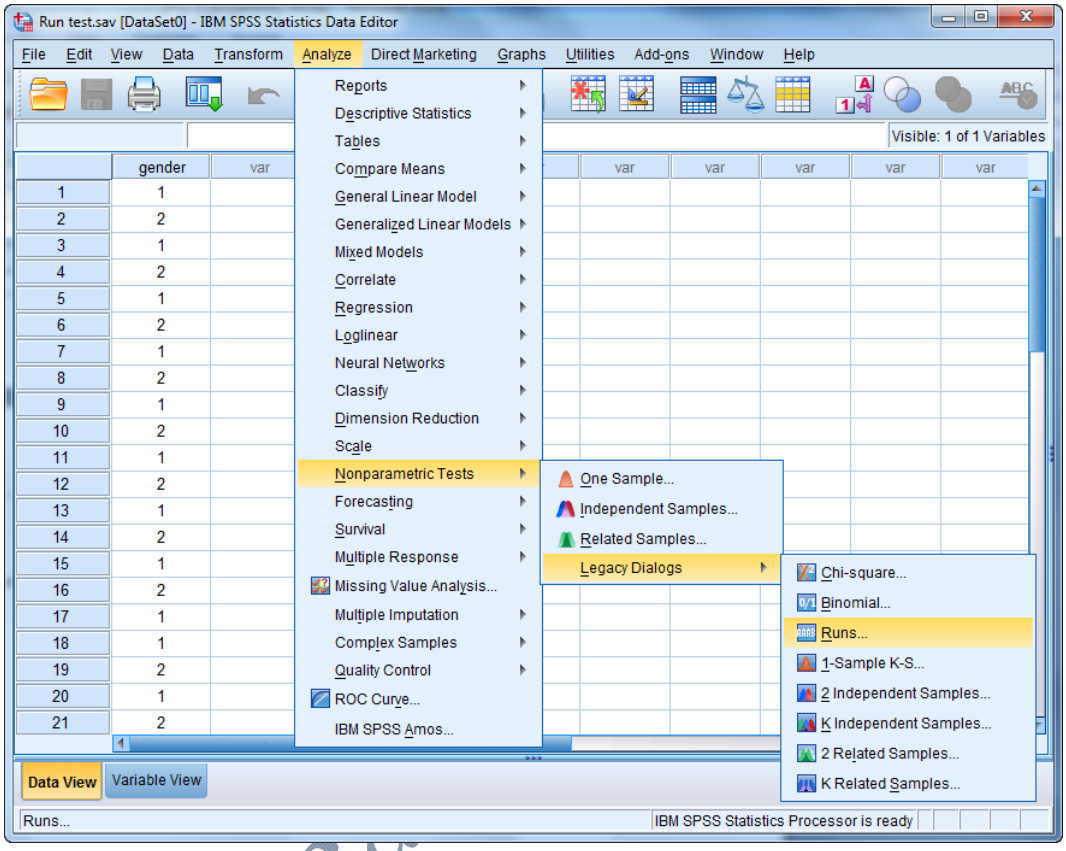
مثال(۱۰): پژوهشگری در یک تحقیق می‌خواهد بداند که آیاد ترتیب مراجعه دانشجویان دختر(کد۱) و پسر(کد ۲) به کتابخانه دانشگاه برای دریافت کتاب در طول یک روز، به یک اندازه و تصادفی است یا خیر؟ برای این منظور وی جنسیت ۳۰ مراجعه‌کننده به کتابخانه را در یک روز به ترتیب به صورت زیر مشخص می‌نماید. آزمون وی را در سطح ۰/۰۵ بررسی کنید؟

۲	۱	۲	۱	۲	۱	۲	۱	۲	۱
۱	۲	۱	۱	۲	۱	۲	۱	۲	۱
۱	۱	۲	۱	۱	۲	۱	۲	۱	۲

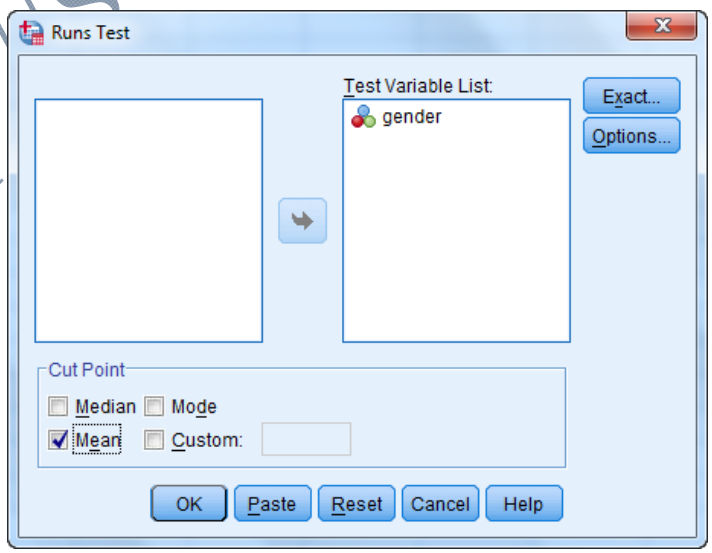
روش ورود اطلاعات در نرم افزار SPSS

داده ها را در یک ستون تحت عنوان متغیر gender وارد می‌کنیم.

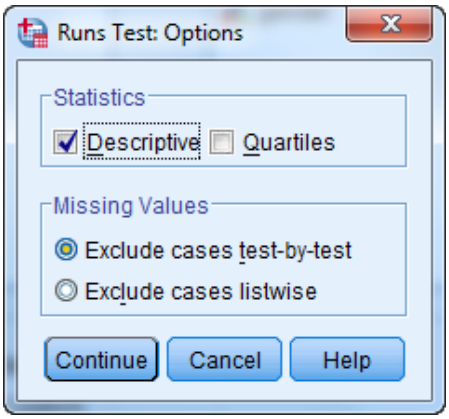
سپس مسیر Analyze / Nonparametric Tests /Legacy Dialogs / Runs ... را انتخاب کنید.



در کادر زیر متغیر gender را وارد کادر Test Variable List کنید. در کادر Cut Point حالت Median را غیر فعال و حالت Mean را انتخاب کنید.



بر روی گزینه Options... کلیک کرده تا کادر زیر باز شود. در کادر باز شده گزینه Descriptive را علامت‌دار کنید.



سپس روی دکمه‌های Continue و Ok کلیک کنید تا خروجی به صورت زیر مشاهده شود.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
gender	30	1.43	.504	1	2

Runs Test

	gender
Test Value[a]	1.43
Cases < Test Value	17
Cases >= Test Value	13
Total Cases	30
Number of Runs	27
Z	4.076
Asymp. Sig. (2-tailed)	.000

a. Mean

همانطور که از جداول بالا مشخص است، سطح معناداری آزمون برابر با صفر و کوچکتر از مقدار ۰/۰۵ است. بنابراین در سطح اطمینان ۹۵ درصد فرض صفر پژوهش یعنی تصادفی بودن تعداد مراجعات دختران و پسران پذیرفته نمی‌شود. با توجه به اینکه برای دختران کد(۱) و برای پسران کد(۲) انتخاب شده است. بر اساس نتایج جدول نیز مشخص است که تعداد مراجعه کنندگان دختر(۱۷نفر) بیشتر از تعداد مراجعه کنندگان پسر(۱۳ نفر) است. بنابراین می‌توان چنین نتیجه گرفت که دانشجویان دختر بیشتر از دانشجویان پسر به کتابخانه مراجعه می‌کنند.

آزمون χ^2 تک متغیره/ نیکوئی برازش(Chi-Square):

آزمون χ^2 تک متغیره که به همراه آزمون کلموگروف- اسمیرنوف تک نمونه‌ای به آزمون نیکوئی برازش[1] معروف است، یکی از پرکاربردترین آزمون‌ها در رشته‌های مدیریت و علوم تربیتی است. این آزمون برای فرضیه‌هایی به کار می‌رود که در آن محقق از یک متغیر ترتیبی برای تنظیم فرضیه استفاده کرده است. هدف این آزمون مقایسه فراوانی‌های مشاهده شده با فراوانی‌های مورد انتظار، به ویژه از طریق مقادیر و فراوانی-های باقیمانده است. به عبارت دیگر آزمون کای اسکوئر از طریق مقایسه دو فراوانی مشاهده شده و مورد انتظار به آزمون فرضیه می‌پردازد.

آزمون کای اسکوئر (χ^2) یک متغیره دو معادل به صورت زیر دارد:

۱- **آزمون پارامتری t تک نمونه‌ای(One-Sample T test)**: که برای فرضیه‌های تفاوتی با متغیرهای در سطح سنجش فاصله‌ای/نسبی به کار می‌رود.

۲- **آزمون ناپارامتری دوجمله‌ای(Binomial test)**: که برای فرضیه‌های تفاوتی با متغیرهای در سطح سنجش اسمی دو وجهی به کار می‌رود.

از جمله پیش‌فرض‌های استفاده از توزیع χ^2 تک متغیره به قرار زیر است:

۱- مقیاس داده‌ها ترتیبی باشد.

۲- فراوانی مورد انتظار تمامی طبقات باید حداقل (۱) باشد.

۳- بیش از ۲۰ درصد فراوانی‌های مورد انتظار خانه‌های جدول نباید کوچکتر از عدد ۵ باشد.

مثال(۱۱): پژوهشگری در تحقیقی با عنوان "بررسی عوامل موثر بر نشاط در بین شهروندان تهرانی" فرضیه خود را به صورت زیر بیان کرده است: "به نظر می‌رسد که، شهروندان تهرانی از نشاط پایینی برخوردار باشند". پژوهشگر برای بررسی فرضیه خود تعداد ۱۰۰ نفر از شهروندان تهرانی را به تصادف انتخاب و به وسیله یک پرسشنامه که شامل سوالاتی با طیف لیکرت ۵ تایی(خیلی کم(۱)، کم(۲)، متوسط(۳)، زیاد(٤)، خیلی زیاد(۵)) بوده است. میزان نشاط آنها را طبق جدول زیر به دست آورده است. فرضیه پژوهشگر را بررسی کنید؟

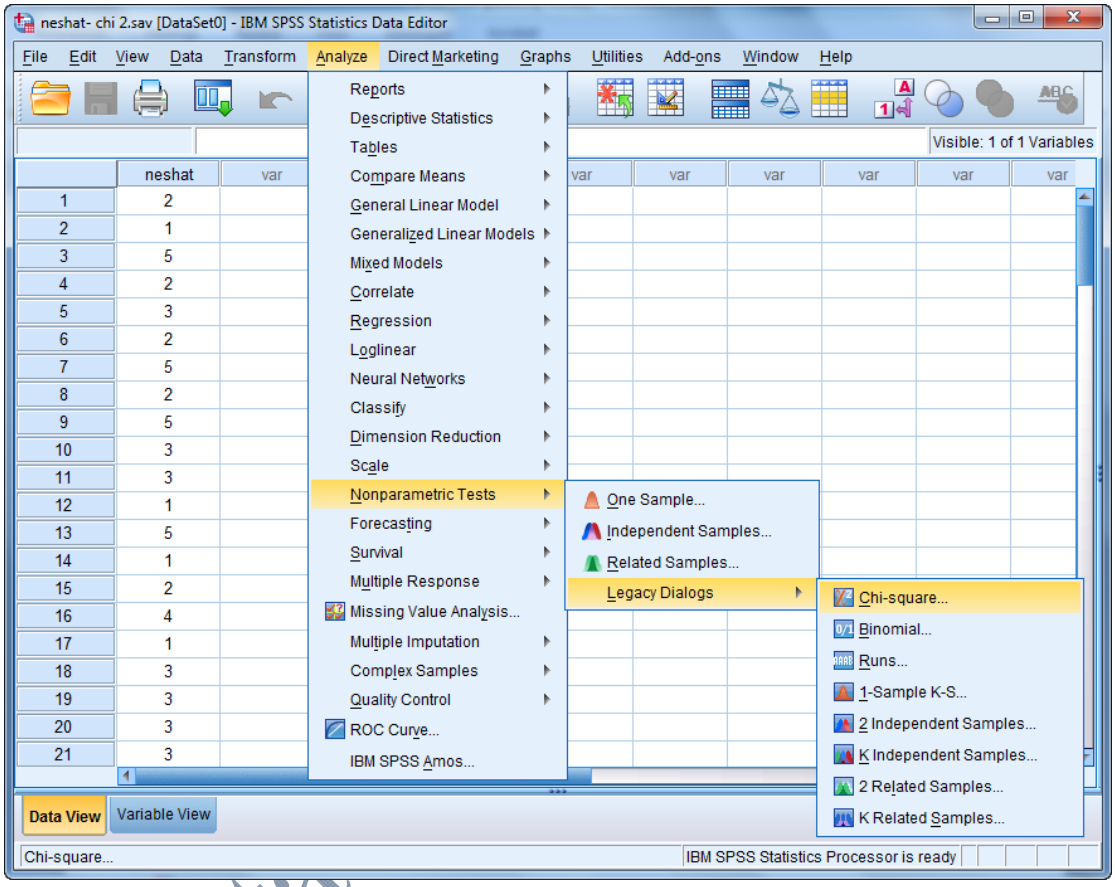
۲	۱	۱	۲	۳	۲	۱	۲	۱	۳	۳	۱	۵	۱	۲	۴	۱	۳	۳	۳
۳	۳	۲	۴	۱	۳	۲	۲	۴	۵	۲	۵	۳	۱	۳	۲	۱	۲	۱	۵
۵	۳	۳	۱	۴	۵	۳	۴	۱	۱	۱	۴	۲	۱	۵	۵	۵	۱	۱	۳
۲	۴	۳	۴	۵	۴	۲	۲	۱	۵	۱	۳	۱	۱	۱	۱	۳	۱	۵	۳
۱	۴	۱	۲	۵	۳	۴	۵	۳	۱	۱	۴	۳	۱	۲	۱	۲	۱	۵	۵

[1] Goodness of fit

روش ورود اطلاعات در نرم افزار SPSS

ابتدا داده‌ها را در یک ستون تحت عنوان متغیر neshat وارد خواهیم کرد.

سپس مسیر Analyze / Nonparametric Tests /Legacy Dialogs / Chi-Square… را انتخاب کنید.

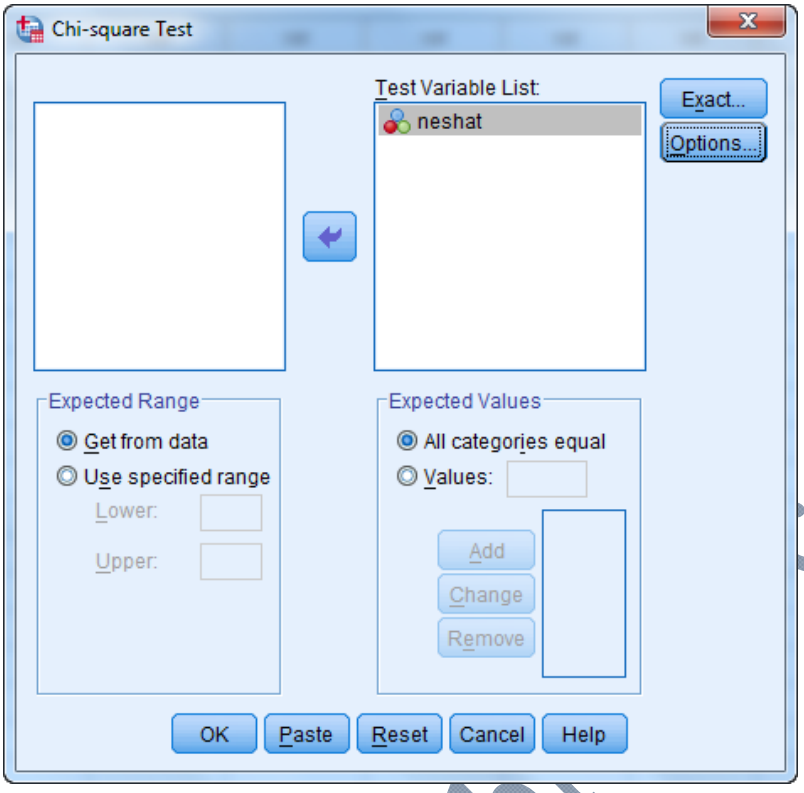


در کادر محاوره‌ای باز شده در قسمت Expected Values فراوانی‌های مورد انتظار را به دو طریق می‌توانیم وارد نماییم.

۱- **حالت برابر(All categories equal)**: در این حالت خود نرم افزار فراوانی‌های هر طبقه را برابر در نظر می‌گیرد. در این روش فراوانی مورد انتظار برای هر طبقه، از تقسیم تعداد کل فراوانی‌ها $(N = ۱۰۰)$ بر تعداد طبقات متغیر $(k = ۵)$ به دست می‌آید.

۲- **حالت نابرابر(Values)**: در این حالت خودمان بر اساس یک مبنای نظری می‌توانیم فراوانی‌های مورد انتظار یک طبقه را به صورت دستی و نابرابر تعریف کنیم. یعنی فراوانی‌های تمامی طبقات را که انتظار داریم فراوانی‌های انها بیشتر و یا کمتر از بقیه باشد، مشخص کنیم.

در این مثال ما فراوانی تمامی طبقات را یکسان در نظر گرفته‌ایم. همانند شکل زیر:



سپس بر روی گزینه Ok کلیک کنید تا خروجی به صورت زیر نمایش داده شود.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
neshat	100	2.74	1.474	1	5

Chi-Square Test

Frequencies

neshat

	Observed N	Expected N	Residual
kheili kam	32	20.0	12.0
kam	18	20.0	-2.0
motevaset	22	20.0	2.0
ziad	12	20.0	-8.0
kheili ziad	16	20.0	-4.0
Total	100		

Test Statistics

	neshat
Chi-Square	11.600[a]
df	4
Asymp. Sig.	.021

با توجه به خروجی بالا در مورد فرضیه پژوهشگر بحث کنید؟

آزمون کلموگروف-اسمیرنف تک نمونه‌ای (One-Sample Kolmogorov-Smirnov):

همانطور که پیش‌تر بیان شد، فرمول χ^2 تک متغیره وقتی صادق است که؛ فراوانی‌های مورد انتظار بیش از ۲۰ درصد خانه‌های جدول کمتر از ۵ نباشد. بنابراین در صورتیکه فراوانی‌های مورد انتظار بیش از ۲۰ درصد خانه‌های جدول کمتر از ۵ باشد، دیگر فرمول χ^2 صادق نیست. این حالت اغلب زمانی رخ می‌دهد که حجم نمونه کمتر از ۵۰ باشد. در چنین حالتی به جای آزمون χ^2 تک متغیره از آزمون کلموگروف-اسمیرنف استفاده می‌شود.

مثال(۱۲): برای بررسی رضایت مشتریان یک بانک از نحوه برخورد کارکنان، تعداد ۵۰ نمونه از مراجعه کنندگان انتخاب و بر اساس یک نظر سنجی میزان رضایت آنها به صورت زیر بر اساس یک طیف لیکرت ۵ گزینه‌ای(خیلی کم(۱)، کم(۲)، متوسط(۳)، زیاد(٤)، خیلی زیاد(٥)) به دست آمده است. فرضیه تحقیق که به صورت زیر تعریف شده است را بررسی کنید؟

فرضیه تحقیق: "به نظر می‌رسد که مشتریان بانک از نحوه برخورد کارکنان رضایت دارند."

۵	۱	۳	۳	۵	۲	۵	۲	۳	۲
۳	۵	۲	۵	۴	۲	۲	۳	۱	۴
۲	۴	۱	۱	۱	۴	۳	۵	۴	۱
۱	۳	۱	۵	۱	۲	۲	۴	۵	۴
۳	۴	۱	۱	۳	۵	۴	۳	۵	۲

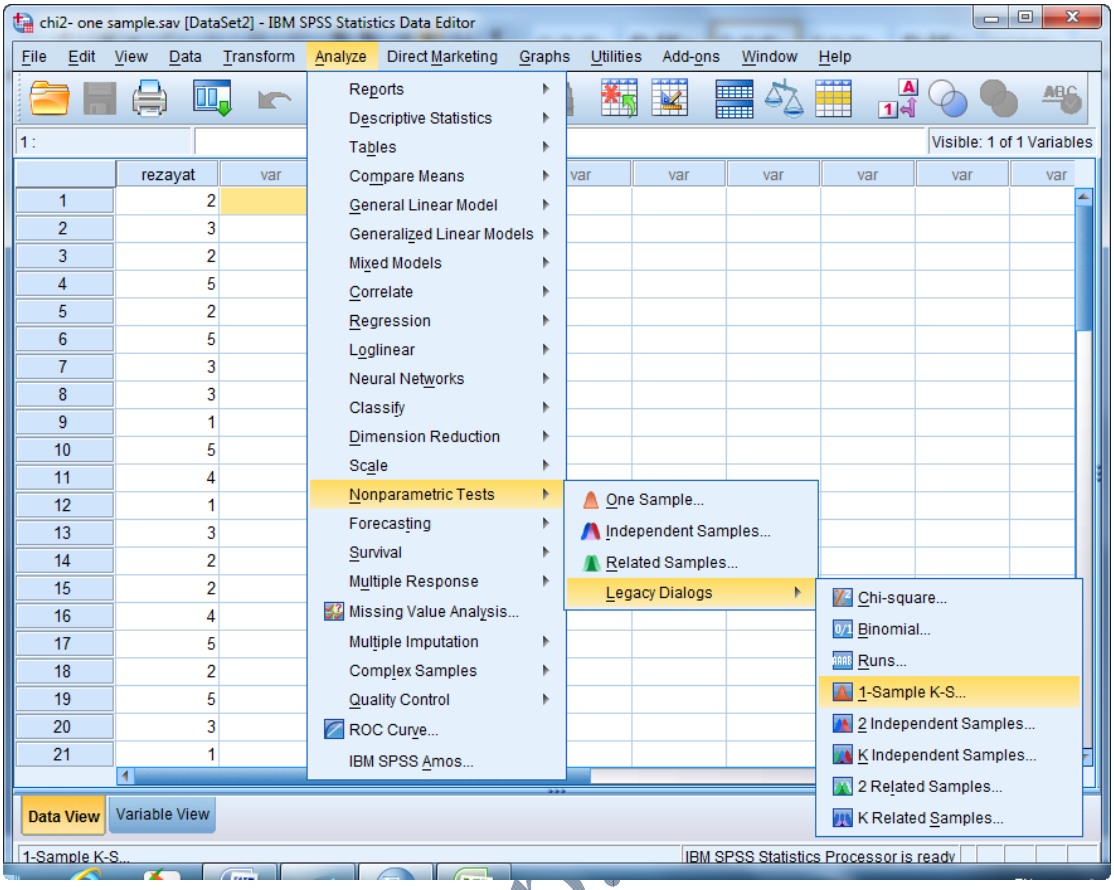
روش ورود اطلاعات در نرم افزار SPSS

ابتدا داده‌ها را در یک ستون تحت عنوان متغیر rezayat وارد خواهیم کرد.

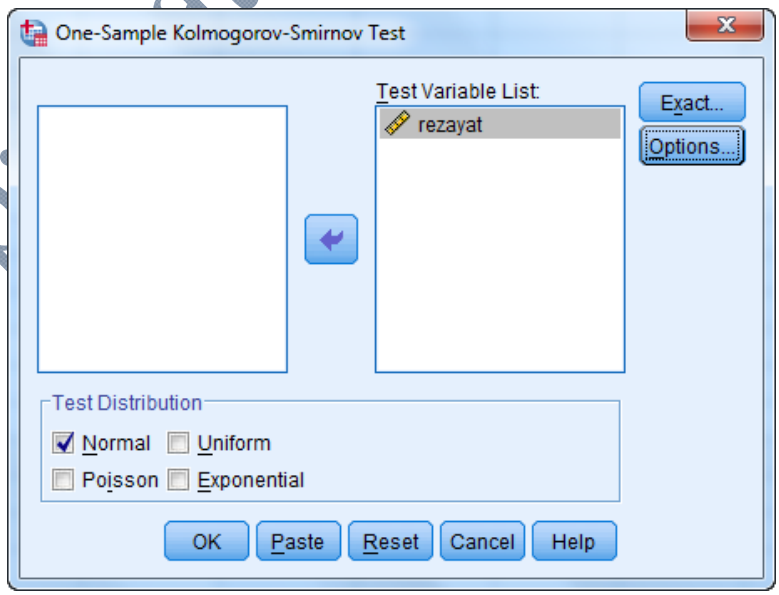
سپس مسیر

Analyze / Nonparametric Tests /Legacy Dialogs / 1-Sample K-S

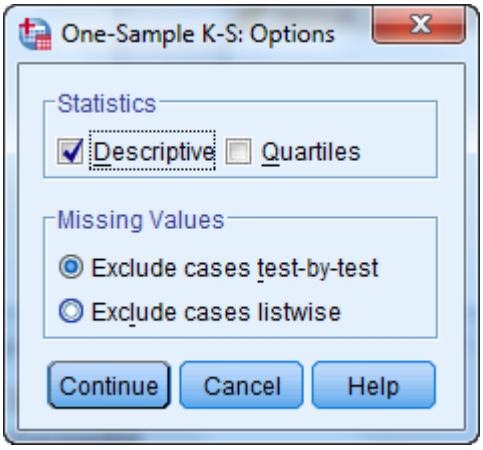
را انتخاب کنید.



در کادر محاوره‌ای باز شده متغیر rezayat را به کادر Test Variable List منتقل کنید. در قسمت Test Distribution گزینه Normal را انتخاب کنید.



بر روی دکمه Options.. کلیک کرده و گزینه Descriptive را انتخاب کنید.



سپس بر روی گزینه های Continue و Ok کلیک کنید تا خروجی زیر حاصل شود.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
rezayat	50	2.94	1.449	1	5

One-Sample Kolmogorov-Smirnov Test

		rezayat
N		50
Normal Parameters[a,b]	Mean	2.94
	Std. Deviation	1.449
Most Extreme Differences	Absolute	.162
	Positive	.162
	Negative	-.148
Kolmogorov-Smirnov Z		1.144
Asymp. Sig. (2-tailed)		.146

a. Test distribution is Normal.
b. Calculated from data.

همانطور که از داده‌های جدول بالا مشخص است سطح معناداری برابر با ۰/۱٤٦ و بزرگتر از مقدار ۰/۰۵ است. در نتیجه داده‌ها از توزیع نرمال(توزیع بهنجار، متقارن) برخوردار هستند. در نتیجه فرض پژوهش تایید نمی‌گردد.

آزمون من- ویتنی(Mann-Whitney):

هر گاه دو نمونه مستقل از جامع ای مفروض باشد و داده‌های مربوط به این دو نمونه مستقل رتبه ای و یا داد های کمی غیر نرمال باشند و هدف از آزمون، مقایسه یک متغیر بر روی این دو نمونه مستقل باشد از آزمون من-ویتنی استفاده می‌گردد. این آزمون یکی از قوی‌ترین آزمون‌های ناپارامتری و جانشینی برای آزمون t با دو نمونه مستقل است.

در این آزمون فرضیات صفر و یک به صورت زیر تعریف می گردد:

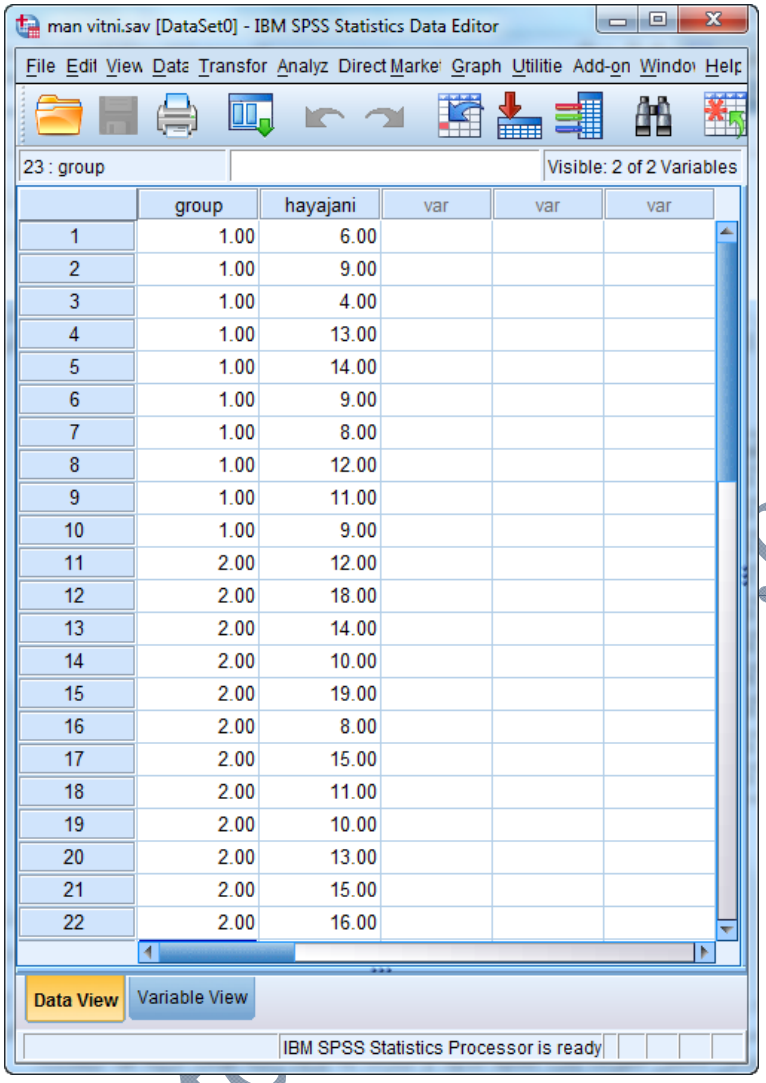
$$\begin{cases} H_0: \text{تفاوتی بین دو گروه وجود ندارد.} \\ H_1: \text{بین دو گروه تفاوت وجود دارد.} \end{cases}$$

مثال(۱۳): برای بررسی تاثیر والدین بر میزان هیجان‌پذیری کودکان، ۱۲ کودک از خانواده‌های دو والـدی و ۱۰ کودک از خانواده‌های تک والدی انتخاب و نمرات هیجان‌پذیری آنها محاسبه و در جدول زیر ثبت شده است. آزمون تاثیر والدین بر میزان هیجان‌پذیری کودکان را انجام دهید؟

تک والدی	۶	۹	۴	۱۳	۱۴	۹	۸	۱۲	۱۱	۹	
دو والدی	۱۲	۱۸	۱۴	۱۰	۱۹	۸	۱۵	۱۱	۱۰	۱۳	۱۵	۱۶

روش ورود اطلاعات در نرم افزار SPSS

برای ورود داده‌ها به نرم افزار، همانند آزمون مقایسه میانگین‌های دو گروه مستقل عمل خواهیم کرد. دو متغیر به نام‌های group و hayajan تشکیل خواهیم داد که در ان متغیر group نشان دهنده تک والدی یا دو والدی بودن است که به ترتیب با کدهای (۱) و (۲) مشخص می‌شود و متغیر hayajani نیز نمره هیجان‌پذیری افراد است.

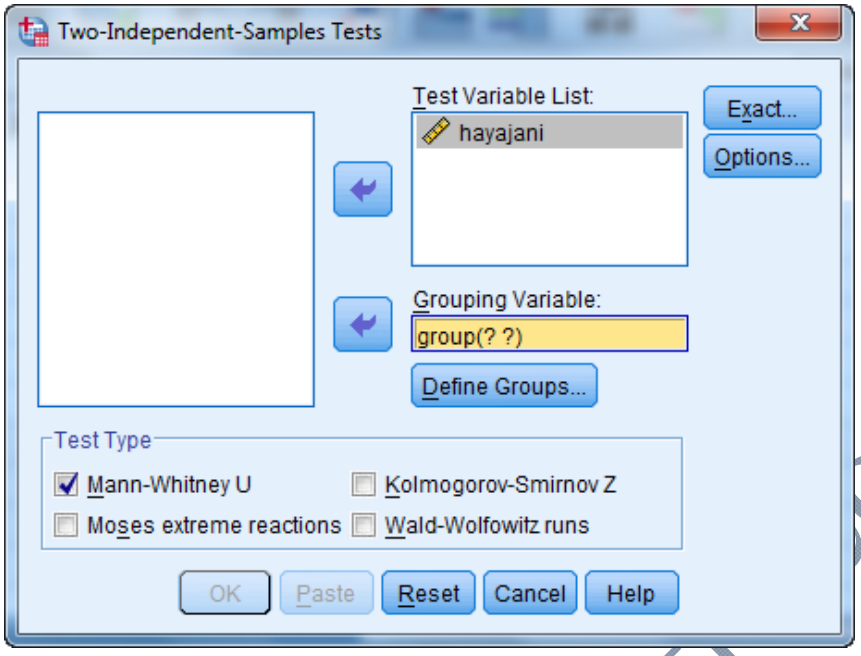


سپس مسیر:

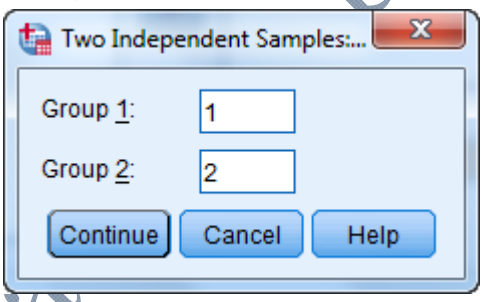
Analyze / Nonparametric Tests / Legacy Dialogs / 2 Independent Samples

را انتخاب تا کادر زیر باز شود.

متغیر hayajni را به کادر Test Variable List و متغیر group را به کـادر Grouping Variable منتقل می‌کنیم.



سپس رو گزینه Define Groups... کیک کرده تا کادر زیر باز شود. مطابق شکل، کدهای انتخاب شده برای دو گروهی که قرار است با هم مقایسه شوند را در دو کادر می‌نویسیم.



دکمه Continue و Ok را می‌زنیم. خروجی به صورت زیر خواهد بود.

Ranks

	group	N	Mean Rank	Sum of Ranks
hayajani	1 valedi	10	7.85	78.50
	2 valedi	12	14.54	174.50
	Total	22		

Test Statistics[a]

	hayajani
Mann-Whitney U	23.500
Wilcoxon W	78.500
Z	-2.414
Asymp. Sig. (2-tailed)	.016
Exact Sig. [2*(1-tailed Sig.)]	.014[b]

a. Grouping Variable: group
b. Not corrected for ties.

همانطور که از خروجی بالا مشخص است، مقدار آماره آزمون من ویتنی برابر با ۲۳/۵ است. همچنین Sig=0.016 است. که کوچکتر از مقدار ۰/۰۲۵ (آزمون دو دامنه است)می‌باشد. در نتیجه در سطح اطمینان ۹۵٪ نمره هیجان‌پذیری کودکان در دو گروه با هم برابر نیست. با توجه به مقادیر ستون Mean Rank مشخص است که نمره هیجان‌پذیری در کودکان خانواده‌های دو والدی بالاتر است.

آزمون‌های علامت(Sign) و ویلکاکسون(Wilcoxon):

این آزمون‌ها معادل آزمون پارامتری t زوجی هستند. برای مقایسه میانگین‌های دو گروه وابسته، زمانیکه پیش فرض‌های استفاده از آزمون‌های پارامتری برقرار نباشد از آزمون‌های ناپارامتری معادل آن استفاده خواهد شد. البته بین آزمون‌های علامت و ویلکاکسون تفاوت‌هایی وجود دارد. آزمون علامت نسبت به داده‌های پرت حساس نیست و این در حالی است که آزمون ویلکاکسون نسبت به داده‌های پرت حساس می‌باشد.

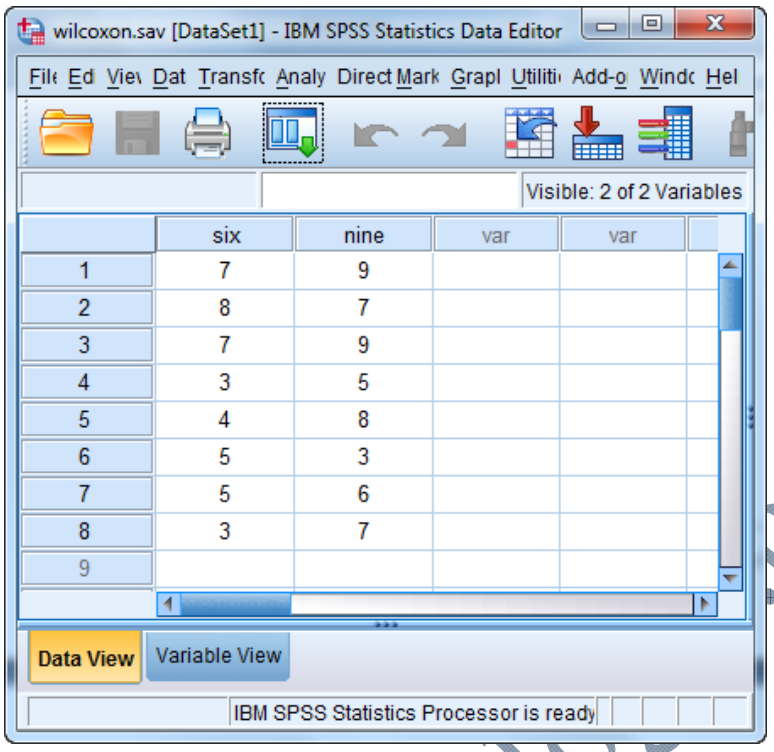
آزمون علامت را هنگامی به کار می‌بریم که ارزشیابی متغیر مورد مطالعه با روش‌های عادی یا روش‌های اعمال شده قابل اندازه‌گیری نباشد. از آزمون ویلکاکسون هم زمانیکه داده‌ها رتبه‌ای باشند استفاده خواهد شد.

مثال(۱٤): تعداد ۱۰ نوزاد انتخاب و تعداد تماس‌های چشمی آنها در سنین ٦ و ۹ ماهگی به صورت جدول زیر یادداشت شده است. آزمون وجود تفاوت در تعداد تماس‌های چشمی در سنین مختلف را انجام دهید؟

شماره نوزاد	۱	۲	۳	۴	۵	۶	۷	۸
۶ ماهگی	۷	۸	۷	۳	۴	۵	۵	۳
۹ ماهگی	۹	۷	۹	۵	۸	۳	۶	۷

روش ورود اطلاعات در نرم افزار SPSS

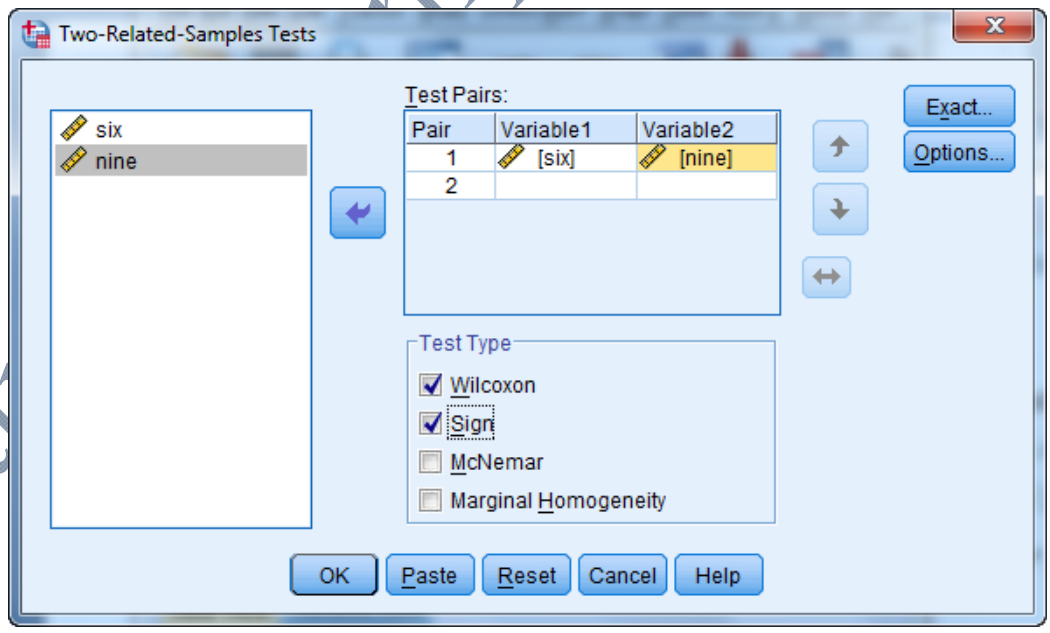
برای ورود داده‌ها به محیط SPSS همانند آزمون t زوجی عمل خواهیم کرد. دو متغیر را در دو ستون جداگانه همانند شکل زیر وارد خواهیم کرد.



سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / 2 Related Samples

را انتخاب تا کادر زیر باز شود.



متغیرهای six و nine را به کادر Test Pairs منتقل کرده و در قسمت Test Type گزینه‌های wilcoxon و Sign(آزمون علامت) را تیک میزنیم. سپس دکمه Ok را می‌زنیم. خروجی به صورت زیر خواهد بود.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
nine - six	Negative Ranks	2[a]	3.00	6.00
	Positive Ranks	6[b]	5.00	30.00
	Ties	0[c]		
	Total	8		

a. nine < six
b. nine > six
c. nine = six

Test Statistics[a]

	nine - six
Z	-1.706[b]
Asymp. Sig. (2-tailed)	.088

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

Sign Test

Frequencies

		N
nine - six	Negative Differences[a]	2
	Positive Differences[b]	6
	Ties[c]	0
	Total	8

a. nine < six
b. nine > six
c. nine = six

Test Statistics[a]

	nine - six
Exact Sig. (2-tailed)	.289[b]

a. Sign Test
b. Binomial distribution used.

آزمون مک‌نمار(McNemar):

آزمون مک نمار معادل ناپارامتری آزمون t دو نمونه‌ای وابسته است، در شرایطی که نوع متغیر مورد بررسی اسمی باشد. هنگامی که موقعیت استفاده از آزمون t دو نمونه‌ای وابسته فراهم نباشد و بخواهیم تأثیر یک فعالیت را روی رفتار آزمون شوندگان مشخص کنیم و میزان تغییر نظر آزمون شوندگان را قبل و بعد از مداخله مشخص نماییم، از آزمون مک نمار استفاده می‌کنیم. برای نمونه می‌توان به بررسی نظرات افراد در مورد مشارکت در انتخابات ریاست جمهوری پیش از سخنرانی در آن مورد و پس از سخنرانی، اشاره نمود. پس اجرای آزمون مک نمار مستلزم وجود دو مجموعه از داده‌های دو مقوله‌ای است. این داده‌ها می‌توانند

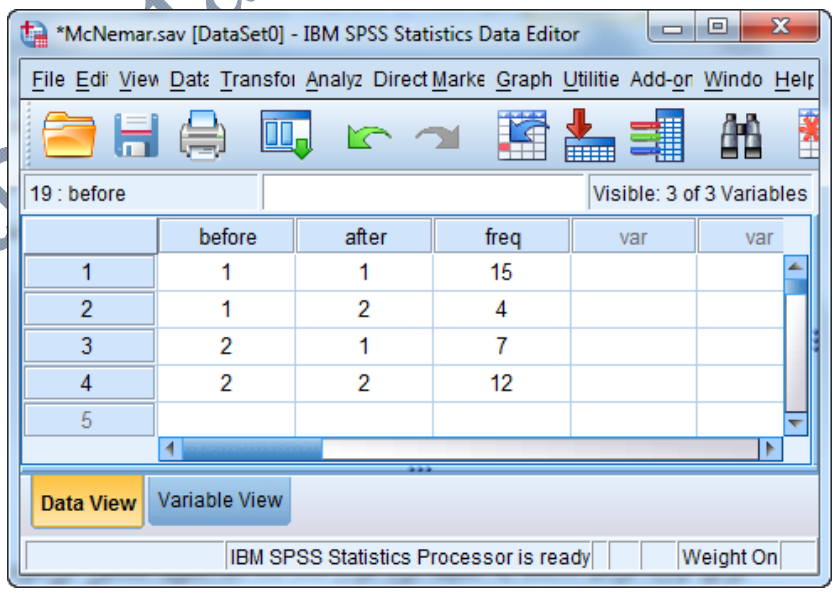
به طور ذاتی دو مقوله‌ای باشند نظیر «بلی یا خیر» و یا به صورت طبقه‌ای، رتبه‌ای یا فاصله‌ای باشند که پژوهشگر آن‌ها را به صورت دو مقوله‌ای کدبندی می‌کند نظیر گرایش سیاسی که به صورت فاصله‌ای اندازه گیری شود، اما پژوهشگر آن‌ها را به صورت «مثبت یا منفی» مقوله‌بندی نماید.

مثال(۱۵): نظر ۳۸ نفر از دانشجویان در مورد اجرای یک برنامه اقتصادی اخذ شد. سپس در جلسه‌ای درباره مضرات و مزایای اجرای این برنامه توضیحاتی به آنان داده شد. در مرحله بعد نظر آنان را دوباره در مورد اجرای برنامه اقتصادی جویا شدیم. نتایج در جدول زیر داه شده است. آزمون تاثیر جلسه توجیهی را بر نظرات دانشجویان بررسی کنید؟

		بعد جلسه	
		موافق	مخالف
قبل جلسه	مخالف	۴	۱۵
	موافق	۱۲	۷

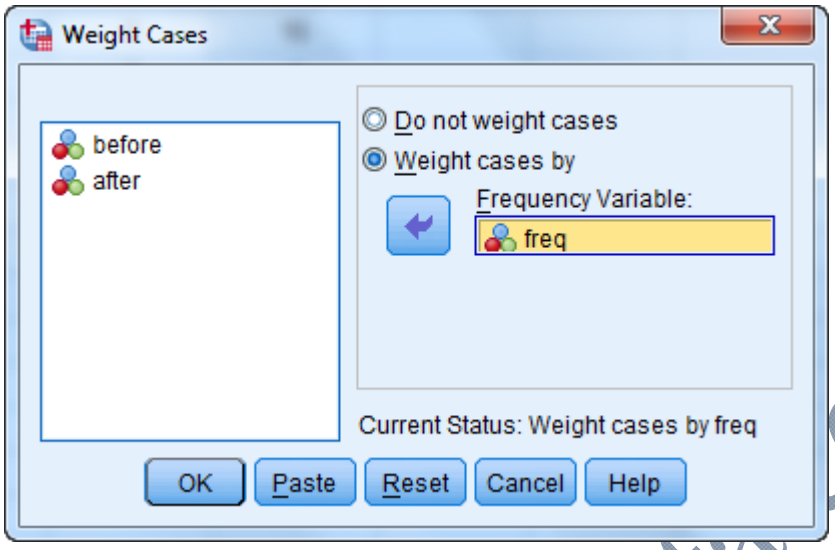
روش ورود اطلاعات در نرم افزار SPSS

برای وارد کردن داده‌ها به محیط نرم افزار SPSS سه ستون تشکیل می‌شود. ستون اول با نام before وضعیت نظرات دانشجویان را قبل از جلسه بیان می‌کند که شامل دو وضعیت مخالف با کد(۱) و موافق با کد(۲) است. برای هر کدام از این وضعیت‌های دو وضعیت نیز بعد از جلسه وجود دارد. که در ستون دوم و تحت عنوان after آورده شده است. ستون یوم با نام freq است که فراوانی‌های به دست امده در جدول بالا است. داده‌های وارد شده به صورت زیر خواهند بود:



حال برای اینکه ستون فراوانی را به متغیرها نسبت دهیم همانطور که در فصل اول بیان شد. مسیر

Data / Weight Cases… را انتخاب تا کادر زیر باز شود.



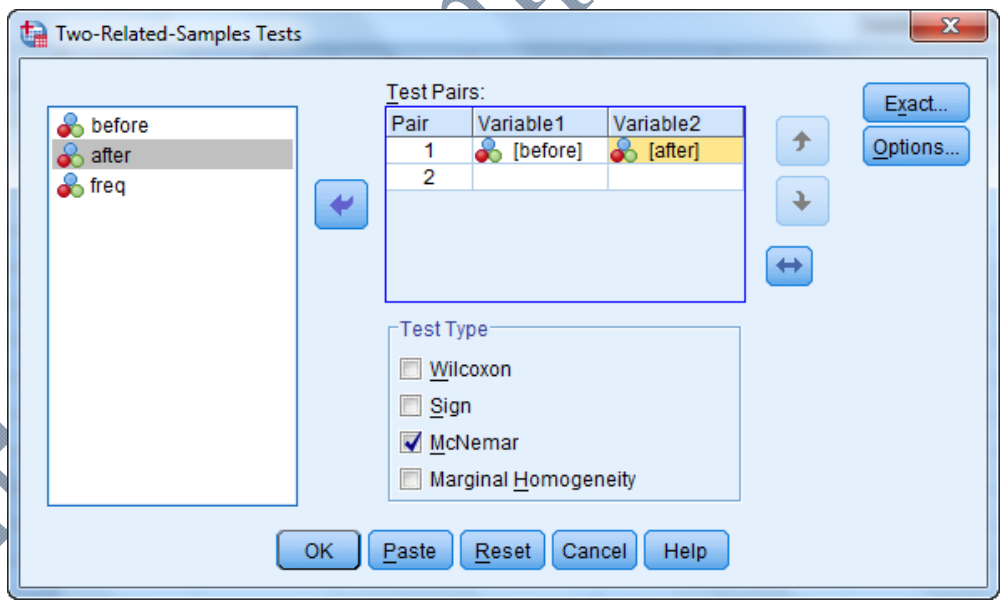
گزینه Weight cases by را انتخاب و متغیر freq را به کادر Frequency Variable انتقال می‌دهیم.

سپس Ok را می‌زنیم.

سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / 2 Related Samples

را انتخاب تا کادر زیر باز شود.



متغیرهای before و after را به کادر Test Pairs منتقل کرده و در قسمت Test Type گزینه McNemar را تیک میزنیم. سپس دکمه Ok را می‌زنیم. خروجی به صورت زیر خواهد بود.

McNemar Test

Crosstabs
before & after

before	after	
	disagree	agree
disagree	15	4
agree	7	12

Test Statistics[a]

	before & after
N	38
Exact Sig. (2-tailed)	.549[b]

a. McNemar Test
b. Binomial distribution used.

خروجی بالا را تفسیر کنید؟

آزمون‌های ناپارامتری برای K گروه مستقل:

آزمون‌های ناپارامتری برای چند گروه مستقل شامل دو آزمون زیر می‌باشد که هر کدام برای مقیاس خاصی به کار می‌روند:

۱- آزمون میانه(Median Test)

۲- آزمون کروسکال- والیس(Kruskal Wallis Test)

آزمون میانه(Median Test):

آزمون میانه در میان آزمون‌های پارامتری معادل آزمون‌های t، Z و F و در میان آزمون‌های ناپارامتری معادل آزمون کروسکال والیس است، منتهی با توان کمتر. هدف از آزمون میانه این است که آیا مقادیر میانه یک متغیر وابسته در بین دو یا چند نمونه مستقل متفاوت است یا خیر؟

آزمون میانه هم برای گروه‌های مستقل و هم وابسته کاربرد دارد و لزومی ندارد که حتماً حجم گروه‌های نمونه با یکدیگر برابر باشند.

آزمون میانه زمانی به کار می‌رود که دو یا چند گروه از میان دو یا چند جامعه مستقل با توزیع‌های یکسان انتخاب شده‌اند. در این آزمون مقیاس اندازه‌گیری حداقل ترتیبی است و بین داده‌ها نباید هم رتبه وجود داشته باشد.

مثال(۱۶): برای مقایسه میزان رضایت مشتریان سه بانک دولتی، نمونه‌ای تصادفی به حجم ۱۵ از مشتریان هر بانک انتخاب و میزان رضایت آنها از خدمات بانکی را در یک طیف صفر تا ۲۰ تایی جویا شدیم. نتایج در جدول زیر داده شده است. آزمون یکسان بودن میزان رضایت از خدمات بانکی ۳ بانک مورد بررسی را انجام دهید؟

بانک A	۱۶	۱۵	۱۸	۱۳	۱۳	۱۰	۱۶	۱۷	۱۷	۱۴	۱۹	۸	۱۷	۱۶	۱۳
بانک B	۱۵	۱۱	۱۸	۱۶	۱۲	۱۴	۲۰	۱۲	۱۰	۱۴	۱۶	۱۱	۱۶	۸	۱۷
بانک C	۱۵	۹	۱۸	۱۰	۲۰	۱۲	۱۷	۱۷	۹	۸	۱۷	۱۵	۸	۱۶	۱۷

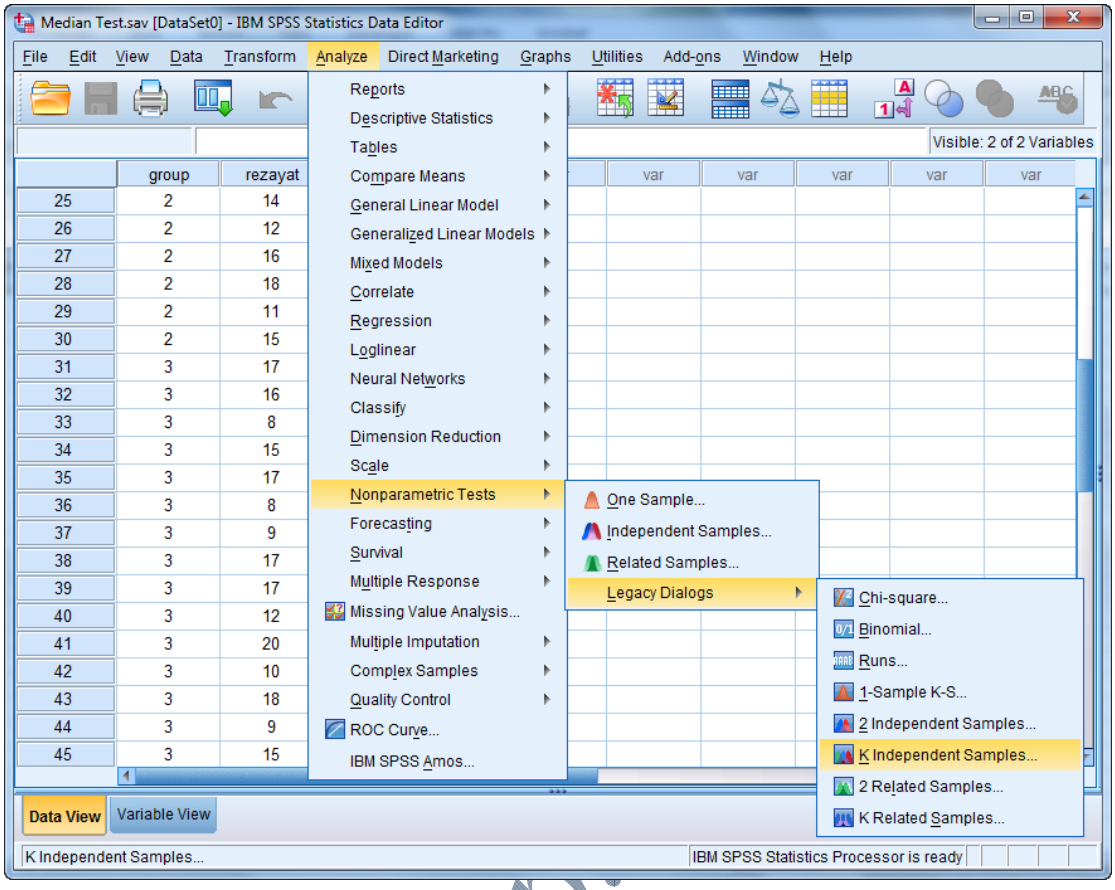
روش ورود اطلاعات در نرم افزار SPSS

ابتدا دو ستون تحت عنوان group و rezayat ایجاد می‌کنیم. متغیر group که نشان دهنده نوع بانک‌ها است(بانکA (کد۱)، بانکB (کد۲)، بانکC (کد۳))و متغیر rezayat نیز که نمرات میزان رضایت مشتریان از عملکرد بانک‌ها است.

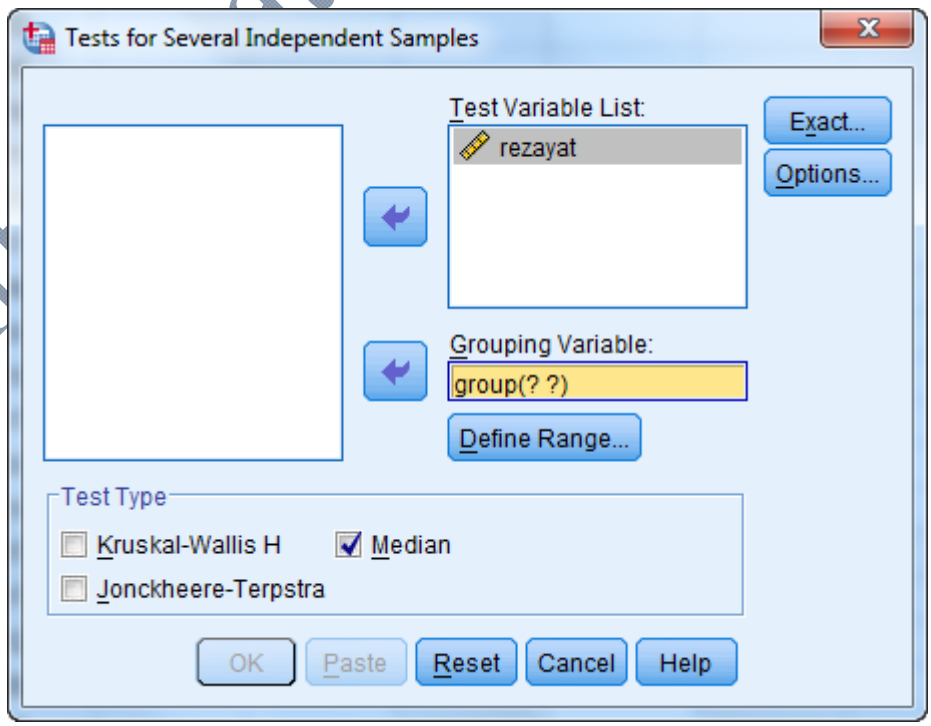
سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / K Independent Samples

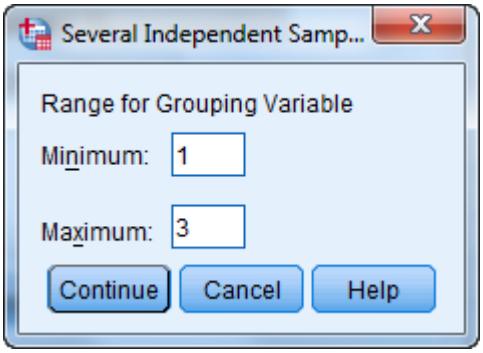
را انتخاب کنید.



در کادر محاوره باز شده متغیر rezayat را وارد کادر Test Variable List کنید و متغیر گروه‌بندی group را وارد کادر Grouping Variable نمایید. سپس در کادر Test Type آزمون Median را انتخاب کنید.



روز دکمه Define Range... کلیک کنید تا پنجره زیر باز شود. در این پنجره، عدد ۱ را در کادر Minimum و عدد ۳ را در کادر Maximum تایپ کنید.



سپس رو دکمه Continue و Ok کلیک کنید تا خروجی زیر حاصل شود.

NPar Tests

Median Test

Frequencies

		group		
		Bank A	Bank B	Bank C
rezayat	> Median	8	6	7
	<= Median	7	9	8

Test Statistics[a]

	rezayat
N	45
Median	15.00
Chi-Square	.536[b]
df	2
Asymp. Sig.	.765

a. Grouping Variable: group

b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 7.0.

در تفسیر نتایج آزمون میانه، علاوه بر تعیین معنی‌داری تفاوت یا عدم تفاوت میزان رضایت مشتریان از عملکرد ۳ بانک مورد مطالعه، می‌توان پی برد که این میزان رضایت در کدام بانک بیشتر و در کدام بانک کمتر است. برای این منظور از نتایج جدول اول(تحت عنوان Frequencies) استفاده خواهد شد. البته در این مثال با توجه به اینکه سطح معناداری برابر با ۰/۷۶۵ و بزگتر از مقدار ۰/۰۵ شده است. در نتیجه تفاوت معناداری بین میزان رضایت مشتریان از عملکرد ۳ بانک مورد بررسی وجود ندارد.

آزمون کروسکال–والیس(Kruskal-WallisTest):

این آزمون معادل ناپارامتری آزمون تحلیل واریانس یک‌طرفه(ANOVA) است و همانند این آزمون زمانی به کار می‌رود که تعداد گروه‌ها بیش از ۲ باشد. مقیاس اندازه‌گیری در کروسکال والیس حداقل باید ترتیبی باشد. فرضیه‌ها در این آزمون بدون جهت هستند و فقط تفاوت را نشان می‌دهند و جهت بزرگتر یا کوچکتر بودن بودن گروه‌ها را از جهت میانگین نشان نمی‌دهد.

تذکر: آزمون کروسکال–والیس همانند آزمون آنالیز واریانس به بررسی برابری میانگین‌ها در جوامع مختلف می‌پردازد. البته با این تفاوت که آزمون ناپارامتری کروسکال والیس فاقد آزمون پسین بوده و برای کشف تفاوت‌های دو به دو باید از آزمون من–ویتنی استفاده نماییم.

مثال(۱۷): در یک تحقیق برای بررسی و مقایسه میزان میزان امید به آینده مردم یک منطقه در بین گروه‌های مختلف تحصیلی(بیسواد، تحصیلات ابتدایی، راهنمایی و دبیرستان، تحصیلات دانشگاهی)، نمونه‌ای به حجم ۱۵ از هر گروه انتخاب و مقدار امید به آینده را بر اساس یک طیف صفر تا ۲۰ به صورت زیر اندازه گرفته است. آزمون یکسان بودن امید به آینده را در بین گروه‌های مختلف بررسی کنید؟

بیسواد	۹	۴	۷	۱۴	۱۰	۱۴	۶	۱۵	۸	۶	۱۲	۱۳	۶	۱۵	۱۴
تحصیلات ابتدایی	۱۰	۱۰	۱۱	۱۲	۱۱	۱۵	۱۴	۷	۹	۱۱	۷	۹	۱۳	۱۶	۱۸
راهنمایی و دبیرستان	۱۱	۱۲	۱۰	۹	۱۳	۱۱	۱۴	۱۰	۱۰	۱۴	۱۳	۱۷	۱۵	۲۰	۹
تحصیلات دانشگاهی	۱۷	۱۴	۱۳	۱۶	۱۴	۱۶	۱۵	۱۶	۱۱	۱۵	۱۵	۱۱	۱۷	۱۳	۱۳

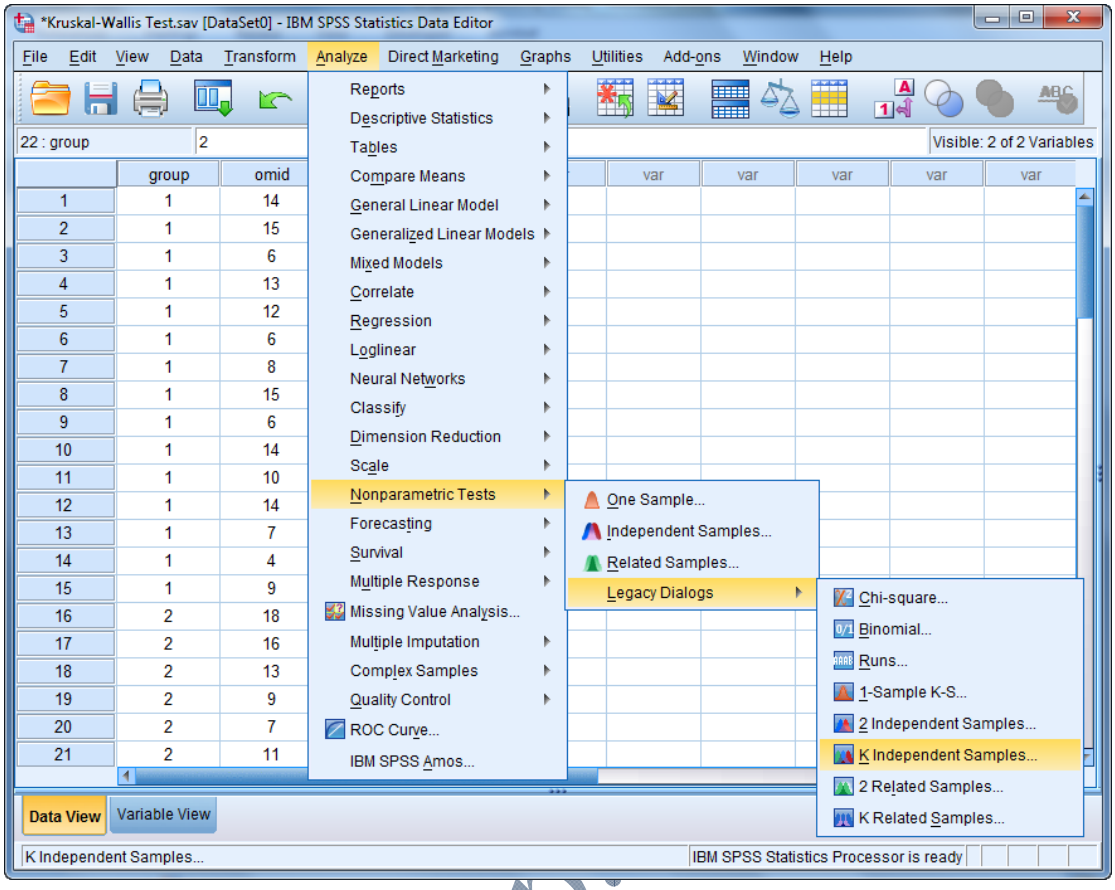
روش ورود اطلاعات در نرم افزار SPSS

روش ورود داده‌ها به نرم افزار دقیقاً همانند روش ANOVA است. ابتدا دو ستون تحت عنوان group و omid ایجاد می‌کنیم. متغیر group که نشان دهنده گروه‌های تحصیلی است(بیسواد(۱)، تحصیلات ابتدایی(۲)، راهنمایی و دبیرستان(۳) و تحصیلات دانشگاهی(٤)). و متغیر omid نیز که نمرات میزان امید به آینده افراد مورد بررسی است.

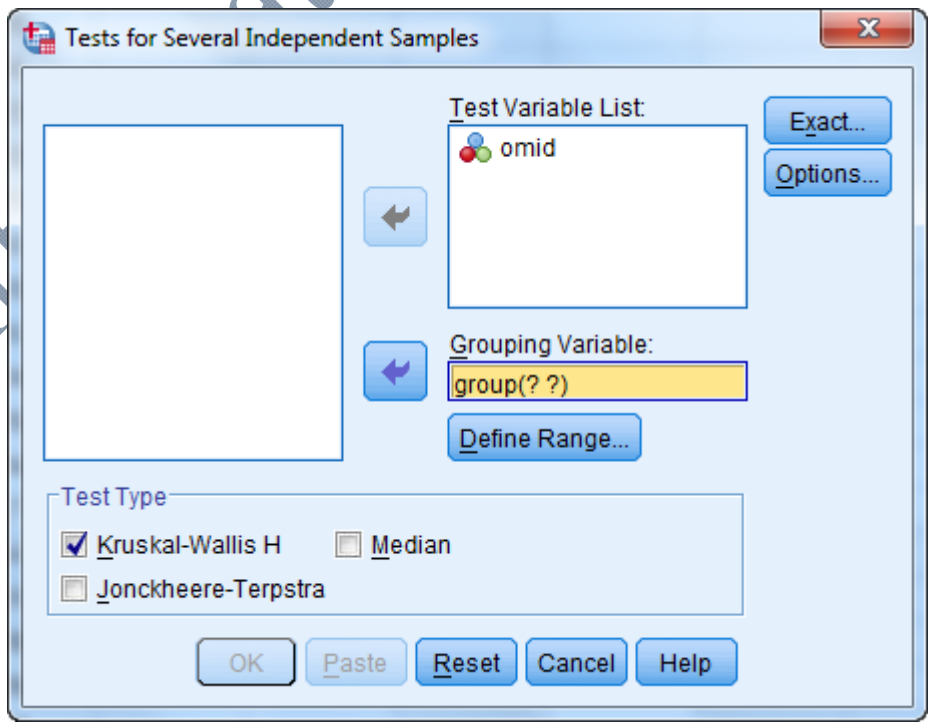
سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / K Independent Samples

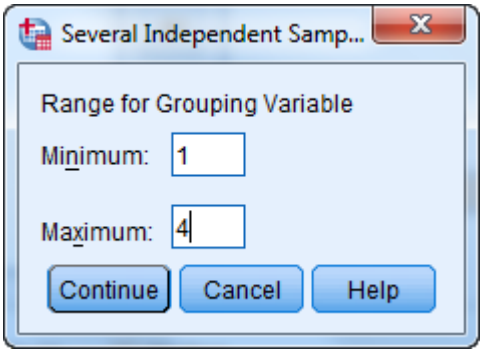
را انتخاب کنید.



در کادر محاوره‌ای باز شده، متغیر omid را به کادر Test Variable List منتقل کنید. متغیر گروه‌بندی group را وارد کادر Grouping Variable نمایید. سپس در کادر Test Type آزمون Kruskal-Wallis H را انتخاب کنید.



روز دکمه Define Range... کلیک کنید تا پنجره زیر باز شود. در این پنجره، عدد ۱ را در کادر Minimum و عدد ٤ را در کادر Maximum تایپ کنید.



سپس رو دکمه Continue و Ok کلیک کنید تا خروجی زیر حاصل شود.

NPar Tests
Kruskal-Wallis Test

Ranks

	group	N	Mean Rank
omid	bisavad	15	22.00
	ebtedaii	15	26.53
	rahnamaii & dabirestan	15	30.50
	daneshgahi	15	42.97
	Total	60	

Test Statistics[a,b]

	omid
Chi-Square	12.082
df	3
Asymp. Sig.	.007

a. Kruskal Wallis Test
b. Grouping Variable: group

در جدول اول تعداد اعضای هر گروه و میانگین رتبه محاسبه شده برای هر گروه داده شده است. با توجه به خروجی جدول دوم سطح معناداری آزمون برابر با ۰/۰۰۷ و کوچکتر از مقدار ۰/۰۵ است. در نتیجه تفاوت معنی‌داری بین گروه‌های مختلف تحصیلی در میزان امید به آینده وجود دارد.

همانطور که قبلاً نیز بیان شد، فرضیه‌ها در این آزمون بدون جهت است. یعنی فقط تفاوت را نشان می‌دهد. برای بررسی تفاوت بین گروه‌ها باید از مقایسات زوجی توسط آزمون من–ویتنی استفاده شود. این کار را برای گروه‌های تحصیلی مثال بالا انجام دهید.

آزمون‌های ناپارامتری برای K گروه وابسته:

آزمون‌های ناپارامتری برای چند گروه وابسته شامل ۳ آزمون زیر می‌باشد که هر کدام برای مقیاس خاصی به کار می‌روند:

۱- آزمون کوکران(Cochran Test)

۲- آزمون فریدمن(Freidman Test)

۳- آزمون رتبه‌های دبلیو کندال (Kendall's W Ranks Test)

آزمون کوکران(Cochran Test):

همانطور که پیش‌تر بیان شد، آزمون مک‌نمار برای دو گروه نمونه مربوط به هم بود. حال اگر تعداد گروه‌ها بیش‌تر از ۲ باشد، از آزمون کوکران استفاده می‌شود. این آزمون گویای این است که آیا تفاوت فراوانی‌ها یا نسبت‌ها در گروه‌های مختلف معنی‌دار است یا خیر؟

در واقع، آزمون کوکران تعمیم یافته آزمون مک‌نمار است.این آزمون برای مقایسه بیش از دو گروه که وابسته باشند و مقیاس آنها اسمی باشند، به کار می‌روند. همچنین در این آزمون، همانند آزمون مک‌نمار، جواب‌ها باید دوتایی باشند. همچنین در این آزمون به جای چند سوال می‌توان یک سوال را در موقعیت-های مختلف ارزیابی کرد.

مثال(۱۸): فرضیه‌ای به صورت زیر تعریف شده است:

"به نظر می‌رسد که؛ با نزدیک شدن به انتخابات، تمایل مردم به شرکت در انتخابات بیشتر می-شود."

برای بررسی فرضیه فوق تعداد ۲۰ نفر از شهروندان به طور تصادفی انتخاب و از آنها در مورد تمایل(کد ۱) یا عدم تمایل(کد ۰) آنها برای شرکت در انتخابات در ٤ دوره زمانی مختلف، پرسش شده است.

آذر ماه	۱	۱	۰	۰	۰	۰	۰	۰	۱	۰	۱	۰	۰	۰	۰	۰	۰	۱	۰	۱
دی ماه	۱	۰	۰	۰	۱	۰	۱	۰	۰	۱	۰	۱	۱	۰	۰	۱	۰	۱	۱	۱
بهمن ماه	۱	۱	۰	۱	۰	۱	۰	۱	۱	۰	۱	۱	۰	۰	۱	۰	۰	۱	۱	۱
اسفند ماه	۱	۱	۱	۱	۱	۰	۱	۰	۱	۱	۱	۰	۱	۰	۱	۱	۱	۱	۱	۱

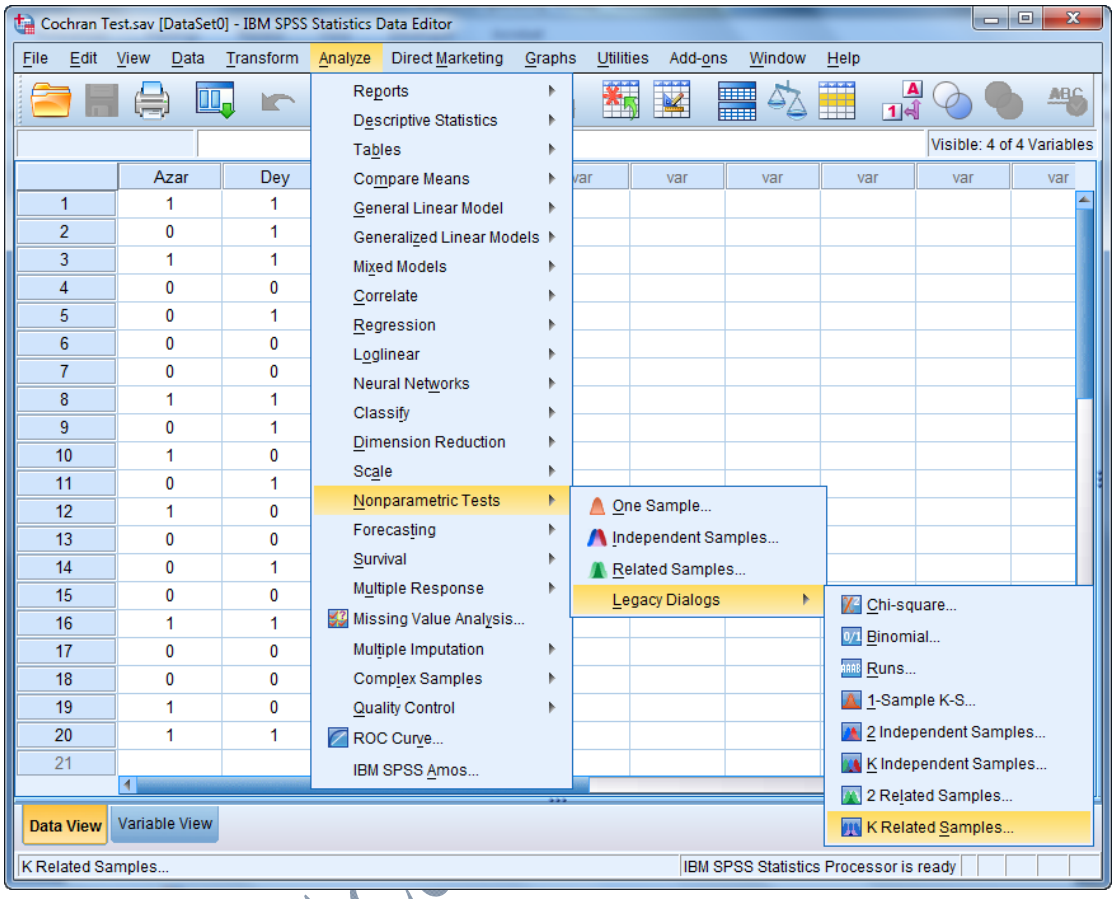
روش ورود اطلاعات در نرم افزار SPSS

چهار ستون تحت عناوین Azar، Dey، Bahman و Esfand ایجاد می‌کنیم.

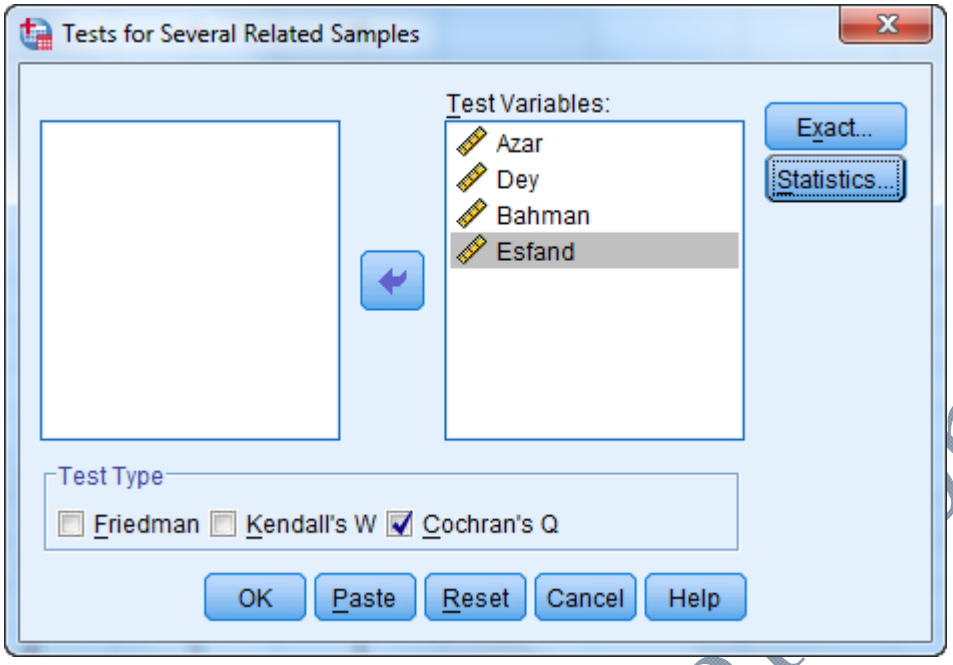
سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / K Related Samples

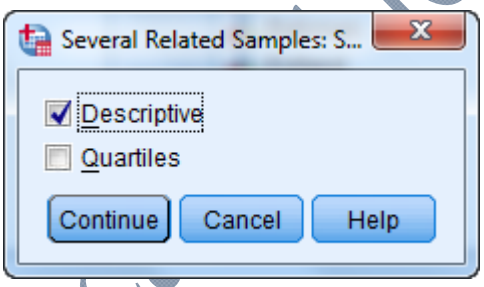
را انتخاب کنید.



در کادر محاوره‌ای باز شده هر چهار متغیر Azar، Dey، Bahman و Esfand را انتخاب و به کادر Test Variables منتقل کنید. سپس از کادر Test Type آزمون Cochran's Q را تیک بزنید.



سپس گزینه ..Statistics را انتخاب و گزینه Desriptive را تیک بزنید.



سپس گزینه‌های Continue و Ok را انتخاب کرده تا خروجی زیر نمایش داده شود.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Azar	20	.30	.470	0	1
Dey	20	.50	.513	0	1
Bahman	20	.60	.503	0	1
Esfand	20	.80	.410	0	1

Cochran Test

Frequencies

	Value	
	0	1
Azar	14	6
Dey	10	10
Bahman	8	12
Esfand	4	16

Test Statistics

N	20
Cochran's Q	11.143[a]
df	3
Asymp. Sig.	.011

a. 1 is treated as a success.

با توجه به اینکه سطح معناداری برابر با ۰/۰۱۱ و کوچکتر از مقدار ۰/۰۵ شده است. در نتیجه تفاوت نسبت شرکت مردم در انتخاب در دوره‌های مختلف معنادار است. با توجه به ستون میانگین جدول Descriptive Statistics مشخص می‌شود که، نسبت مشارک مردم در آذر ماه در کمترین میزان و در اسفند ماه در بیشترین میزان قرار دارد.

آزمون فریدمن(Friedman):

این آزمون برای مقایسه چند گروه از نظر میانگین رتبه آن‌هاست . معلوم می‌کند که آیا این گروه‌ها می‌توانند از یک جامعه باشند یا خیر؟ مقیاس در این آزمون باید حداقل رتبه‌ای باشد. پس آزمون فریدمن برای تحلیل واریانس داده‌های ناپارامتری از طریق رتبه‌بندی، و همچنین مقایسه میانگین رتبه‌بندی گروه‌های مختلف به کار می‌رود.

مثال(۱۹): فرضیه‌ای به صورت زیر تعریف شده است:

"به نظر می‌رسد که میزان رضایت شهروندان از عملکرد شهرداری در طول ۴ سال متفاوت است"

برای بررسی فرضیه فوق، تعداد ۲۰ نفر از شهروندان انتخاب و میزان رضایت آنها از عملکرد شهرداری در طی ٤ سال متفاوت توسط یک طیف ۵ گزینه‌ای(خیلی کم(۱)، کم(۲)، متوسط(۳)، زیاد(٤) و خیلی زیاد(۵)) مورد سنجش قرار گرفت. نتایج در جدول زیر داده شده است. درستی فرضیه را بررسی کنید؟

سال اول	۳	۱	۱	۳	۲	۳	۳	۱	۱	۳	۲	۲	۳	۲	۳	۱	۱	۱	۳	۲
سال دوم	۳	۱	۴	۲	۴	۱	۴	۴	۳	۳	۳	۱	۱	۲	۳	۳	۱	۳	۴	۳
سال سوم	۴	۳	۲	۳	۲	۴	۴	۵	۳	۵	۲	۲	۵	۵	۲	۵	۴	۵	۱	۲
سال چهارم	۴	۵	۲	۲	۵	۳	۲	۴	۳	۵	۴	۴	۴	۲	۴	۵	۴	۴	۵	۲

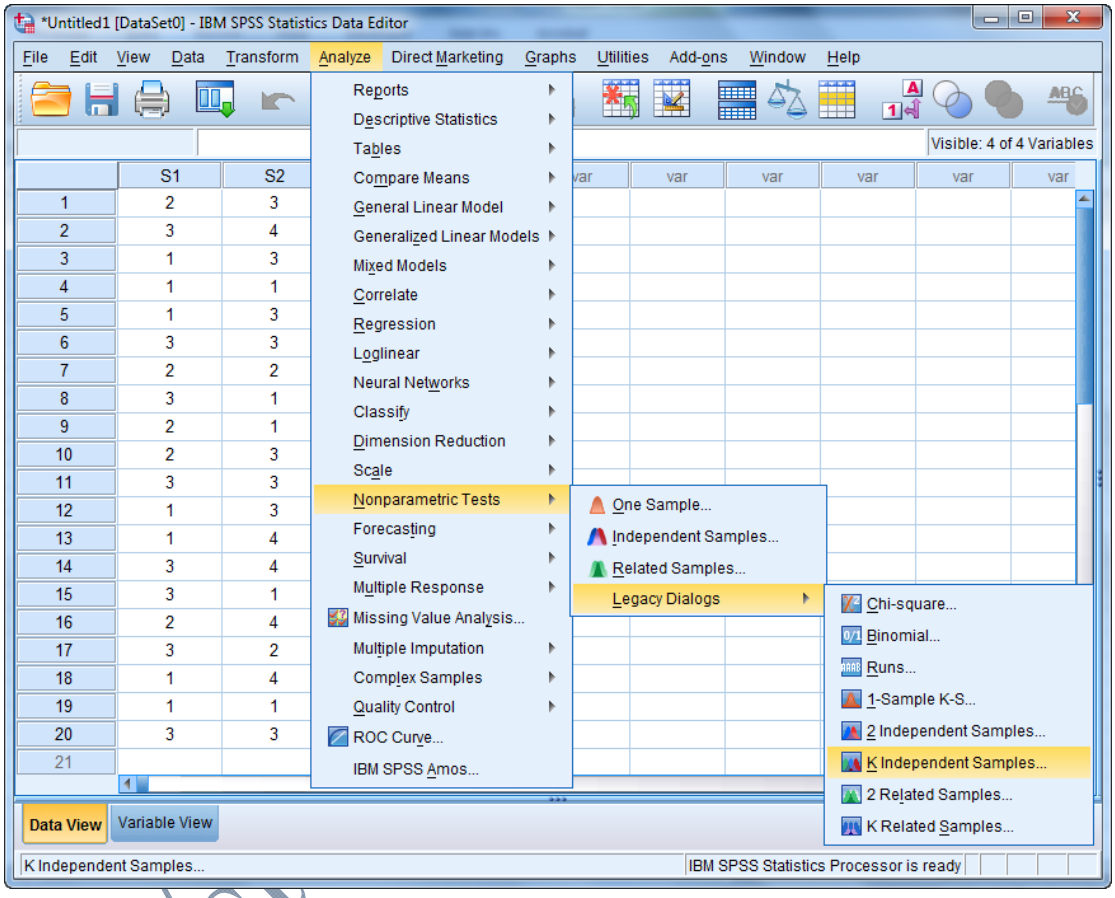
روش ورود اطلاعات در نرم افزار SPSS

چهار ستون تحت عناوین S1، S2، S3 و S4 برای ٤ سال مختلف ایجاد می‌کنیم.

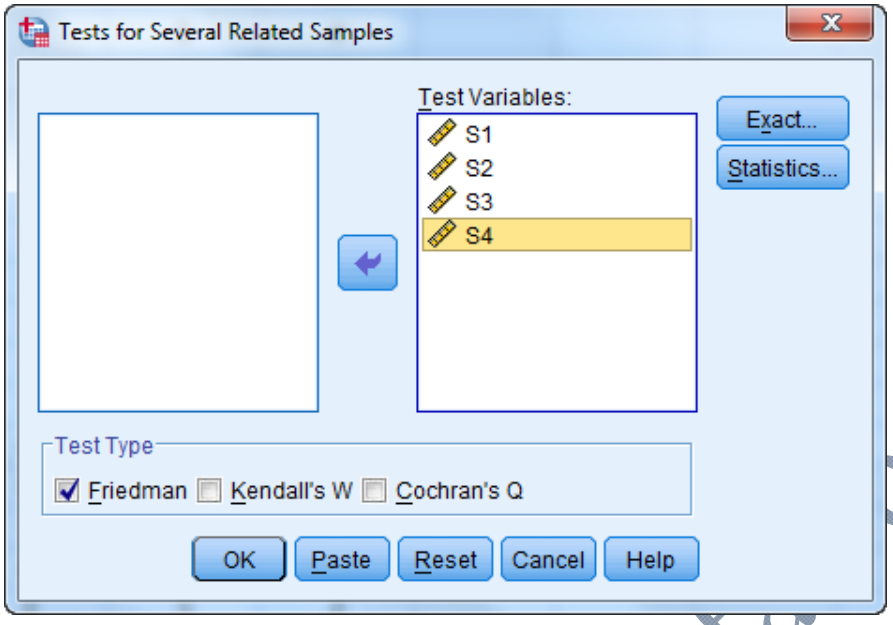
سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / K Related Samples

را انتخاب کنید.



در کادر محاوره‌ای باز شده هر چهار متغیر S1، S2، S3 و S4 را انتخاب و به کادر Test Variables منتقل کنید. سپس از کادر Test Type آزمون Friedman را تیک بزنید.



سپس گزینه Ok را انتخاب کرده تا خروجی زیر نمایش داده شود.

NPar Tests

Friedman Test

Ranks

	Mean Rank
S1	1.75
S2	2.25
S3	2.95
S4	3.05

Test Statistics[a]

N	20
Chi-Square	15.953
df	3
Asymp. Sig.	.001

a. Friedman Test

با توجه به اینکه سطح معناداری برابر با ۰/۰۰۱ و کوچکتر از مقدار ۰/۰۵ شده است. در نتیجه تفاوت میزان رضایت مردم از عملکرد شهرداری در سال‌های مختلف معنادار است. با توجه به ستون Mean Rank در جدول اول مشخص می‌شود که بیشترین میزان رضایت در سال چهارم با میانگین رتبه ۳/۰۵ و کمترین میزان رضایت در سال اول با میانگین رتبه ۱/۷۵ است.

آزمون رتبه‌های دبلیو کندال(Kendall's W Ranks):

آزمون رتبه‌های دبلیو کندال، که شکل نرمال شده آزمون فریدمن است، به عنوان یک ضریب توافق، به سنجش میزان توافق رتبه‌بندی‌ها در بین پاسخگویان می‌پردازد. در این آزمون هر پاسخگو به عنوان یک قضاوت کننده یا رتبه دهنده و هر گویه یا سوال نیز به عنوان یک متغیر تلقی می‌شود. در ادامه برای هر یک از این متغیرها، میانگین رتبه‌ها محاسبه می‌شود. این آزمون با مقایسه میانگین رتبه‌ها در بین متغیرها، تفاوت این میانگین‌ها را بررسی می‌کند. مقدار آزمون رتبه‌های دبلیو کندال بین (۰) تا (۱) نوسان دارد که در آن، مقادیر نزدیک به (۰) نشان از توافق کمتر و مقادیر نزدیک به (۱) نشان از توافق بیشتر بین پاسخگویان در خصوص متغیرهای مورد نظر دارند.

مثال(۲۰): فرضیه‌ای به صورت زیر تعریف شده است:

"به نظر می‌رسد که؛ رتبه‌بندی شهروندان از عملکرد دولت در حوزه‌های مختلف اقتصادی، اجتماعی، فرهنگی و سیاسی متفاوت است."

برای بررسی فرضیه فوق، تعداد ۲۰ نفر از شهروندان انتخاب و میزان ارزیابی آنها از عملکرد دولت در ٤ حوزه اقتصادی، اجتماعی، فرهنگی و سیاسی، توسط یک طیف ۵ گزینه‌ای(خیلی کم(۱)، کم(۲)، متوسط(۳)، زیاد(٤) و خیلی زیاد(۵)) مورد سنجش قرار گرفت. نتایج در جدول زیر داده شده است. درستی فرضیه را بررسی کنید؟

اقتصادی	۳	۱	۱	۳	۲	۳	۳	۱	۲	۳	۲	۲	۳	۲	۳	۱	۱	۱	۳	۲
اجتماعی	۳	۱	۴	۲	۴	۱	۲	۴	۳	۳	۲	۱	۱	۲	۳	۲	۱	۳	۴	۳
فرهنگی	۴	۳	۲	۳	۲	۳	۴	۱	۳	۵	۲	۲	۲	۲	۲	۴	۴	۲	۱	۲
سیاسی	۴	۴	۵	۵	۳	۳	۲	۴	۳	۲	۴	۴	۴	۲	۳	۱	۴	۲	۵	۲

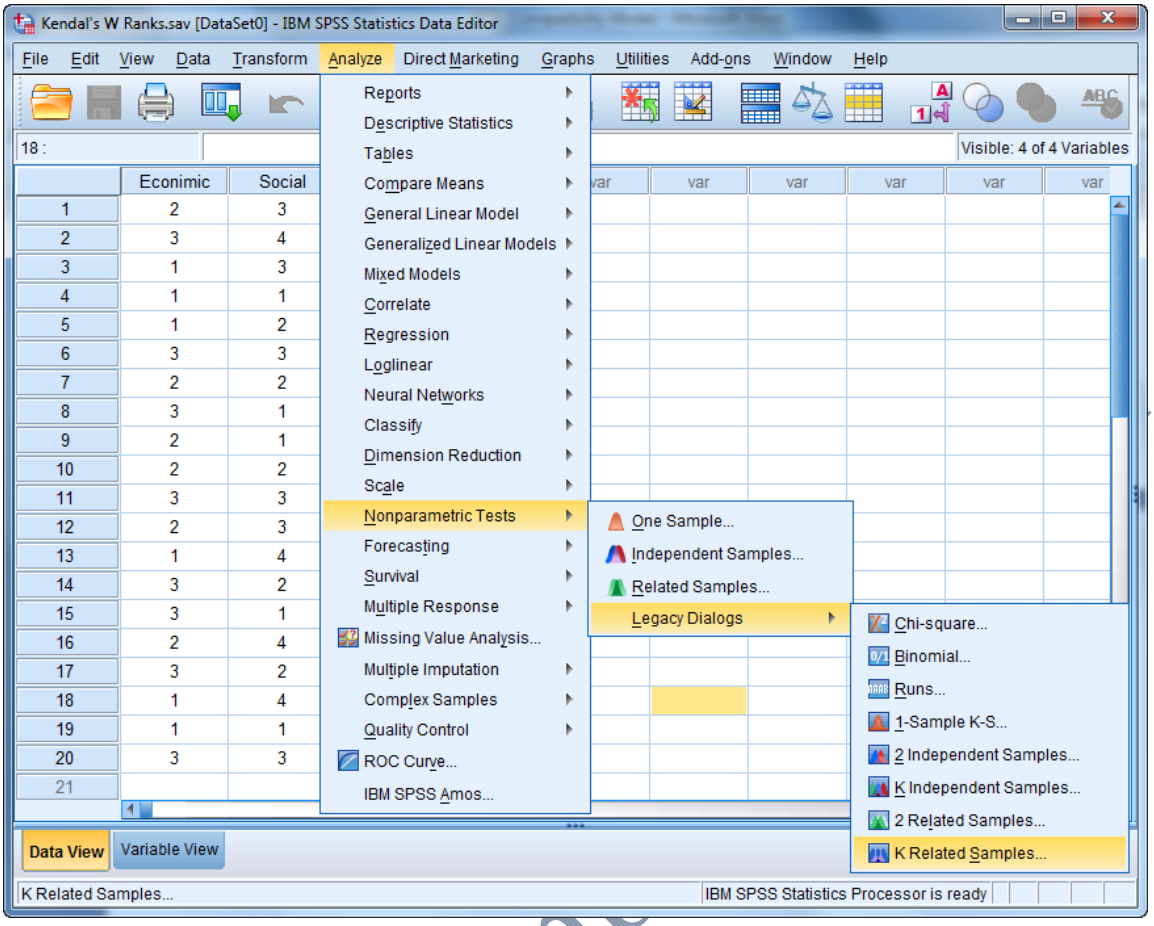
روش ورود اطلاعات در نرم افزار SPSS

چهار ستون تحت عناوین Econimic، Social، Cultural و Political برای داده های ٤ حوزه مختلف مورد بررسی، ایجاد می‌کنیم.

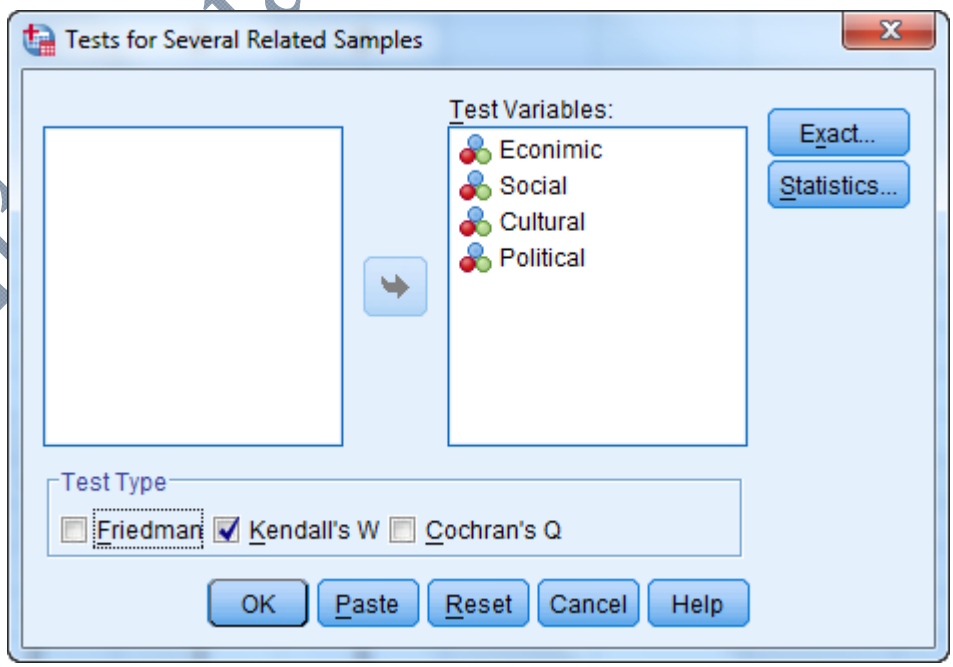
سپس مسیر:

Analyze / Nonparametric Tests / Legacy Dialogs / K Related Samples

را انتخاب کنید.



در کادر محاوره‌ای باز شده هر چهار متغیر Econimic، Social، Cultural و Political را انتخاب و به کادر Variables Test منتقل کنید. سپس از کادر Type Test آزمون Lendall's W را تیک بزنید.



سپس روی گزینه Ok کلیک کنید تا خروجی زیر نمایش داده شود.

NPar Tests

Kendall's W Test

Ranks

	Mean Rank
Econimic	2.00
Social	2.38
Cultural	2.55
Political	3.08

Test Statistics

N	20
Kendall's W[a]	.154
Chi-Square	9.212
df	3
Asymp. Sig.	.027

a. Kendall's Coefficient of Concordance

همانطور که از جداول بالا مشخص است، سطح معناداری کوچکتر از مقدار ۰/۰۵ است. در نتیجه ارزیابی شهروندان از عملکرد دولت در ٤ حوزه مورد بررسی یکسان نیست. با توجه به مقادیر ستون Mean Rank مشخص می‌شود که؛ از نظر شهروندان دولت بیشترین موفقیت را در زمینه سیاسی و کمترین موفقیت را در زمینه اقتصادی داشته است.

فصل پنجم:
آزمون‌های همبستگی

همبستگی‌های دو متغیره شامل دو نوع همبستگی فاصله‌ای(با استفاده از آزمون Pearson) و ترتیبی(با استفاده از آزمون Kendall's tau-b و Spearman) می‌باشد. این نوع همبستگی‌ها برای تعیین شدت و و جهت پیوند بین دو متغیر فاصله‌ای یا نسبی، هر دو ترتیبی و یک ترتیبی و یکی فاصله‌ای مفید می‌باشند.

ملاحظات مهم در تفسیر نتایج ضرایب همبستگی:

۱- چنانچه سطح معناداری آزمون کوچکتر از مقدار ۰/۰۵ باشد بدین معنی است که؛ همبستگی بین دو متغیر معنادار است و این دو متغیر با یکدیگر ارتباط دارند. اما اگر سطح معناداری بزرگتر از مقدار ۰/۰۵ باشد، در آن صورت همبستگی بین دو متغیر معنادار نبوده و این دو متغیر ارتباط خطی با یکدیگر ندارند.

۲- مقادیر تمامی ضرایب همبستگی بین ۱- تا ۱+ در نوسان است، که علامت آن، نشانگر جهت رابطه(مثبت یا منفی) است.

۳- همبستگی هر متغیر با خودش برابر با یک است. به همین خاطر در جدول ماتریس همبستگی‌ها، در کنار رابطه هر متغیر با خودش، عدد یک آمده است.

٤- ضریب همبستگی در مقابل مقادیر(داده‌های) دورافتاده بسیار حساس است. چنانچه یک مقدار با سایر مقادیر موجود در یک مجموعه داده تفاوت زیادی داشته باشد، در آن صورت این مقدار می‌توان دتاثیر بسیار زیادی در میزان ضریب همبستگی ایجاد کند.

۵- در هنگام محاسبه ضریب همبستگی با دستور Analyze / Correlate / Binariate در کادر Bivariate correlations یک کادر فرعی به نام Correlations coefficients وجود دارد که در آن به صورت پیش فرض، ضریب همبستگی Pearson انتخاب شده است. اگر هر دو متغیر مورد نظر ما به صورت مقیاس فاصله‌ای یا نسبی بودند، آن را تغییر نمی‌دهیم. اما اگر یکی از آن دو و یا هر دو در سطح سنجش ترتیبی باشند، برای سنجش شدت همبستگی از آماره‌های ضریب همبستگی Kendall's tau-b و Spearman استفاده می‌شود.

البته آماره Spearman در مواردی به کار می‌رود که رتبه‌ها تکرار ناپذیرند. لذا برای گروه‌های کوچک(مثلاً تحقیقات روان‌شناسی) مناسب است. اما برای رتبه‌های تکرارپذیر(مثل علوم اجتماعی) از آماره Kendall استفاده می‌شود.

٦- در اجرای دستور همبستگی‌ها، در کادر فرعی Test of significance گزینه Two-tailes به طور پیش فرض انتخاب شده است که این گزینه برای فرضیه‌های فاقد جهت می‌باشد. مانند فرضیه: "بین

احساس محرومیت در افراد و رضایت آن‌ها از زندگی رابطه وجود دارد". در حالیکه فرضیه "بین احساس محرومیت در افراد و رضایت آن‌ها از زندگی رابطه منفی و معکوس وجود دارد" به صورت جهت دار است.

بنابراین در کادر Test of significance، هنگامی که فرضیه ما فاقد جهت است، از گزینه Two-tails(دو دامنه‌ای) و در غیر این صورت از گزینه One-tailes(یک دامنه‌ای) استفاده می‌کنیم.

نحوه تفسیر دامنه ضریب همبستگی:

آماردانان و تحلیل‌گران مختلف دامنه‌های متفاوتی را برای تفسیر ضریب همبستگی در نظر گرفته‌اند. در زیر به مهمترین این دسته‌بندی‌ها که کاربرد بیشتری دارد اشاره شده است:

مقدار	نحوهٔ داوری
۰ -۰/۲۵	همبستگی مستقیم- ضعیف
۰/۲۵ -۰/۵	همبستگی مستقیم – نسبتا قوی
۰/۵ -۰/۷۵	همبستگی مستقیم- شدید
۰/۷۵ -۱	همبستگی مستقیم- بسیار شدید
۰	همبستگی وجود ندارد
-۰/۲۵ - ۰	همبستگی معکوس – ضعیف
-۰/۵ - -۰/۲۵	همبستگی معکوس – نسبتا شدید
-۰/۷۵ - -۰/۵	همبستگی معکوس – شدید
-۱ - -۰/۷۵	همبستگی معکوس - بسیار شدید

ضریب همبستگی پیرسون(Pearson Correlation Cofficient):

ضریب همبستگی پیرسون یک ضریب همبستگی پارامتری است که برای محاسبه درجه و میزان ارتباط خطی بین دو متغیر در سطح فاصله‌ای و نسبی به کار می‌رود.

پیش فرض‌های استفاده از ضریب همبستگی پیرسون:

۱- هر دو متغیر در سطح سنجش فاصله‌ای/ نسبی باشند.

۲- توزیع داده‌ها نرمال باشد.

۳- رابطه بین دو متغیر خطی باشد.(رابطه خطی رابطه‌ای است که نمودار پراکنش آن به صورت خط باشد.)

مثال(۲۱): پزشکی فرضیه‌ای را به صورت زیر مطرح کرده است:

"بین وزن و فشار خون افراد رابطه وجود دارد"

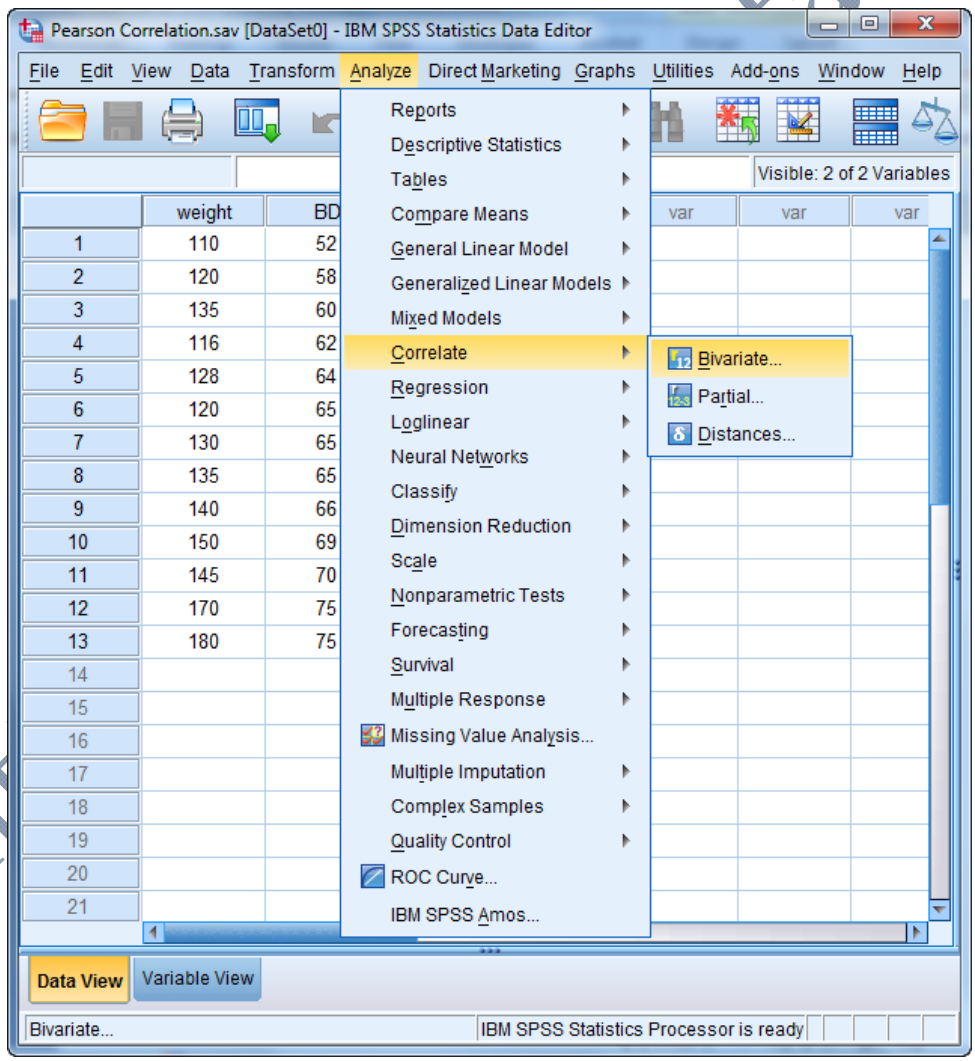
برای بررسی ادعای پزشک، یک نمونه ۱۳ تایی انتخاب و وزن و فشار خون آنها را به صورت جدول زیر یادداشت کرده‌ایم. ادعای پزشک را بررسی کنید؟

فشار خون	۱۱۰	۱۲۰	۱۳۵	۱۱۶	۱۲۸	۱۲۰	۱۳۰	۱۳۵	۱۴۰	۱۵۰	۱۴۵	۱۷۰	۱۸۰
وزن	۵۲	۵۸	۶۰	۶۲	۶۴	۶۵	۶۵	۶۵	۶۶	۶۹	۷۰	۷۵	۷۵

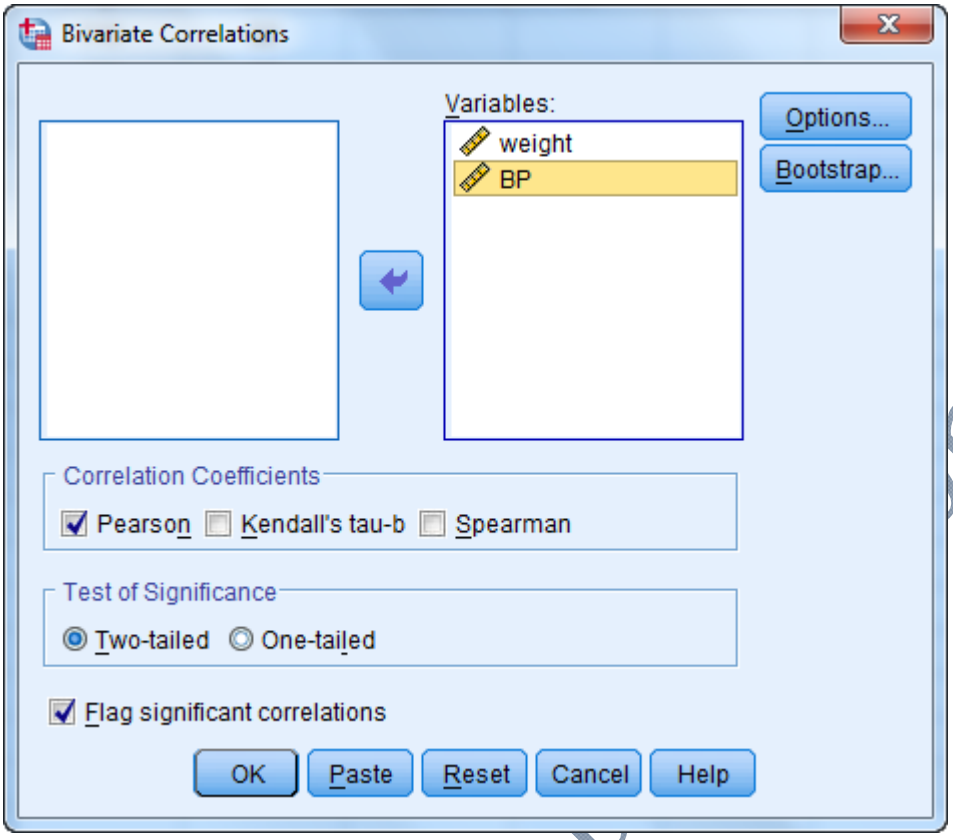
روش ورود اطلاعات در نرم افزار SPSS

دو ستون تحت عناوین وزن(weight) و فشار خون(BP) ایجاد می‌کنیم.

سپس مسیر: Analyze / Correlate / Bivariate را انتخاب کنید.



در کادر باز شده دو متغیر Weight و BP را به کادر Variables منتقل کنید. سپس از کادر Correlation Cofficients گزینه Pearson را انتخاب کنید.



بر روی دکمه Ok کلیک کرده تا خروجی زیر نمایش داده شود.

Correlations

Correlations

		weight	BP
weight	Pearson Correlation	1	.889**
	Sig. (2-tailed)		.000
	N	13	13
BP	Pearson Correlation	.889**	1
	Sig. (2-tailed)	.000	
	N	13	13

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

خروجی جدول بالا را تفسیر کنید؟

ضریب همبستگی اسپیرمن(Spearman) و تاو-بی کندال(Kendall's tau-b):

ضریب همبستگی اسپیرمن که به ضریب همبستگی رتبه‌ای اسپیرمن معروف است، یک ضریب همبستگی ناپارامتری بر اساس رتبه است که میزان همبستگی بین دو متغیر در سطح ترتیبی را اندازه می‌گیرد. در واقع ضریب همبستگی اسپیرمن معادل ناپارامتری ضریب همبستگی پیرسون است.

ضریب همبستگی کندال به یک ضریب متقارن معروف است . برای اندازه‌گیری شدت همبستگی بین دو متغیر ترتیبی و یا یک متغیر ترتیبی و دیگری فاصله‌ای به کار می‌رود.

مفهوم این ضریب همبستگی و کاربرد آن در مثال‌های پژوهشی عیناً همانند ضریب همبستگی اسپیرمن است، با این تفاوت که، اگر تعداد طبقات رتبه‌ها زیاد نباشد، این ضریب بر ضریب همبستگی اسپیرمن برتری دارد.

برای مثال اگر میزان مسئولیت‌پذیری و اعتماد به نفس یک گروه از آزمودنی‌ها را با یک طیف لیکرت ۵ یا ۷ درجه‌ای از خیلی زیاد تا خیلی کم اندازه‌گیری کنیم، ضریب تاو-بی کندال برای بررسی ارتباط آنها مناسب است.

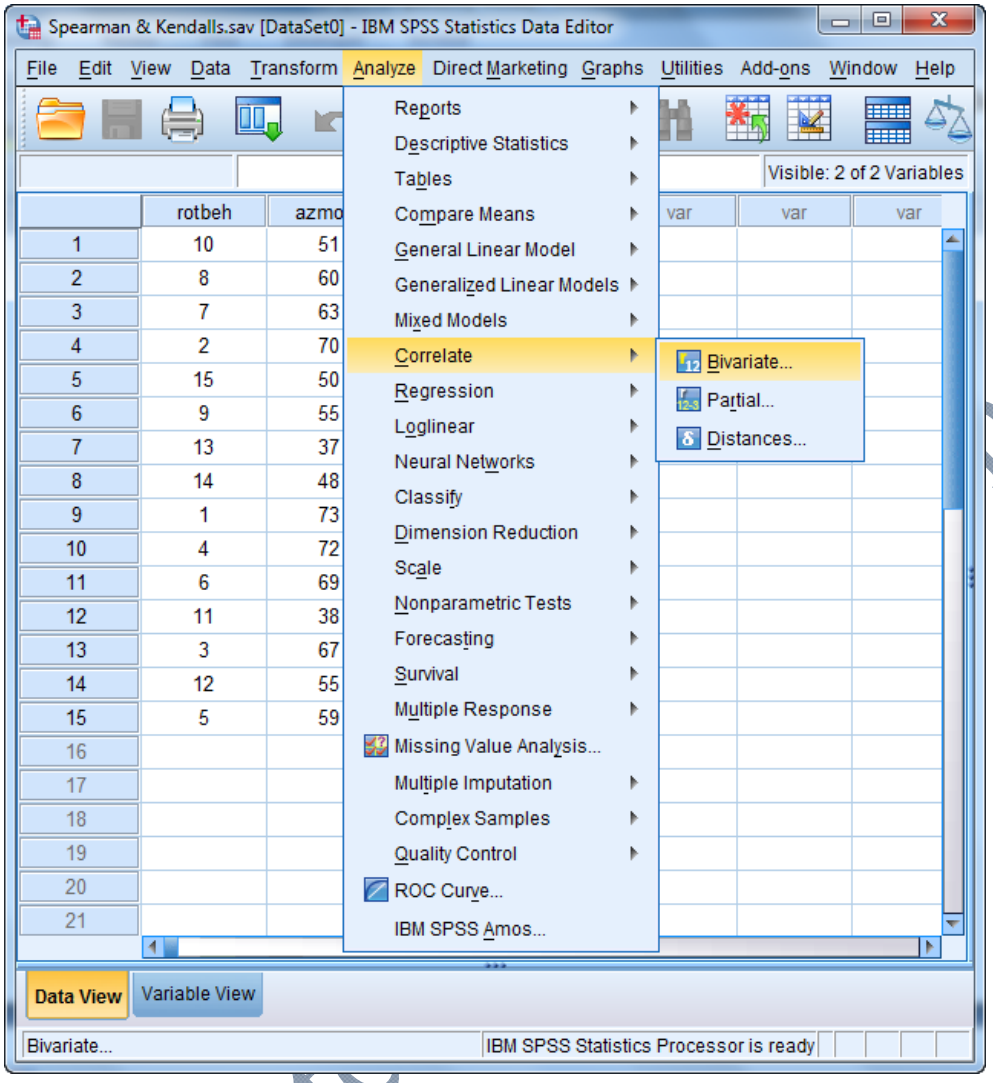
مثال(۲۲): مصاحبه کننده‌ای که مسئول استخدام تعداد زیادی ماشین نویس است می‌خواهد قدرت رابطهٔ بین رتبه‌های داده شده بر مبنای یک مصاحبه و نمرات یک آزمون استعداد را تعیین کند. داده‌های ۱۵ متقاضی در زیر داده شده است. همبستگی بین دو متغیر را محاسبه و آزمون معناداری ضریب همبستگی به دست آمده را تعیین کنید؟

رتبه مصاحبه	۵	۱۲	۳	۱۱	۶	۴	۱	۱۴	۱۳	۹	۱۵	۲	۷	۸	۱۰
نمره آزمون	۵۹	۵۵	۶۷	۳۸	۶۹	۷۲	۷۳	۴۸	۳۷	۵۵	۵۰	۷۰	۶۳	۶۰	۵۱

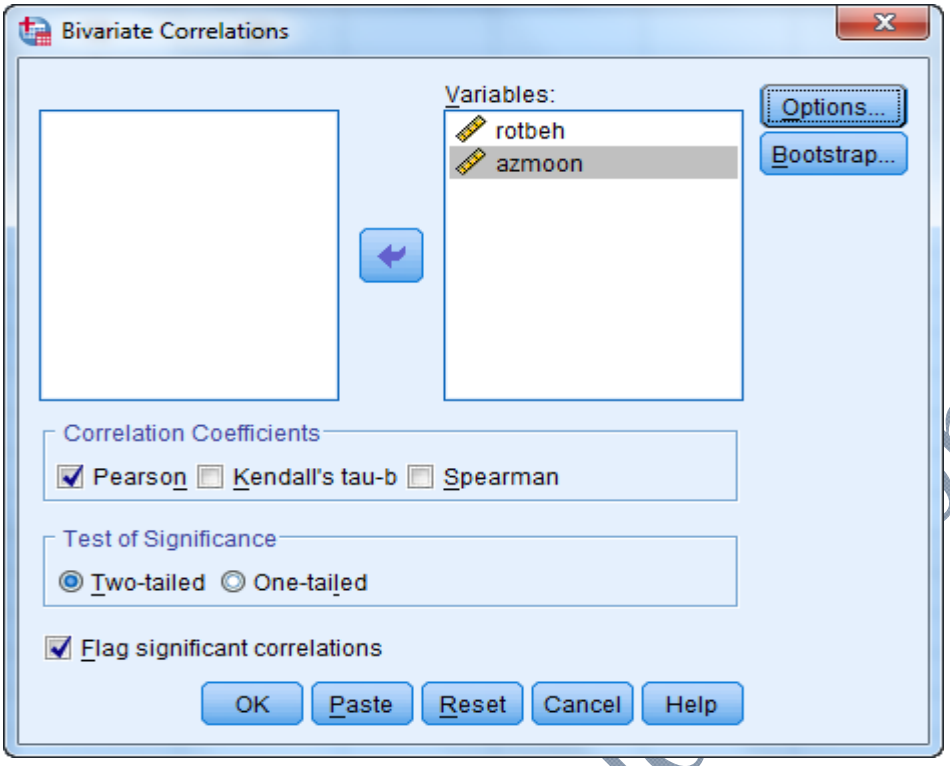
روش ورود اطلاعات در نرم افزار SPSS

دو ستون تحت عناوین رتبه مصاحبه(rotbeh) و نمره آزمون(azmoon) ایجاد می‌کنیم.

سپس مسیر: Analyze / Correlate / Bivariate را انتخاب کنید.



در کادر باز شده دو متغیر rotbeh و azmoon را به کادر Variables منتقل کنید. سپس از کادر Correlation Cofficients گزینه‌های Spearman و Kendall's tau-b را انتخاب کنید.



سپس بر روی گزینه Ok کلیک کنید تا خروجی نمایش داده شود.

Nonparametric Correlations

Correlations

			rotbeh	azmoon
Kendall's tau_b	rotbeh	Correlation Coefficient	1.000	-.746**
		Sig. (2-tailed)	.	.000
		N	15	15
	azmoon	Correlation Coefficient	-.746**	1.000
		Sig. (2-tailed)	.000	.
		N	15	15
Spearman's rho	rotbeh	Correlation Coefficient	1.000	-.903**
		Sig. (2-tailed)	.	.000
		N	15	15
	azmoon	Correlation Coefficient	-.903**	1.000
		Sig. (2-tailed)	.000	.
		N	15	15

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

- خروجی بالا را تفسیر کنید؟
- نتیجه کدام همبستگی در مثال بالامعتبرتر است؟ چرا؟

همبستگی جزئی/ تفکیکی(Partial Correlation):

به دلیل پیچیده بودن روابط بین متغیرها، اغلب اوقات برخی از روابط از دید محقق پنهان می‌ماند و ممکن است در تحلیل در نظر گرفته نشوند. به عنوان مثال ممکن است که یک محقق بخواهد که رابطه بین میزان نشاط و میزان تحصیلات(به سال) را بررسی کند. اما پژوهشگر بر این عقیده است که در این بین میزان درآمد افراد نیز که با توجه به میزان تحصیلات آنان متفاوت است، بر میزان نشاط آنها تاثیر می‌گذارد. حال برای یافتن رابطه بین میزان نشاط و میزان تحصیلات با حذف اثر میزان درآمد، از همبستگی جزئی استفاده خواهیم کرد.

همبستگی جزئی نوعی همبستگی است که ضمن محاسبه میزان رابطه خطی بین دو متغیر، اثر سایر متغیرها را کنترل می‌کند. این ضریب، میزان همبستگی بین یک متغیر مستقل و یک متغیر وابسته را، پس از حذف میزان همبستگی این دو متغیر با یک یا چند متغیر مستقل دیگر، نشان می‌دهد.

پیش فرض‌ها:

۱- تمامی متغیرها باید متقارن و دارای مقیاس کمی(فاصله‌ای/ نسبی) باشند. یعنی هم دو متغیر اصلی و هم متغیرهای کنترل باید کمی باشند.

۲- متغیرهای مورد آزمون باید رابطه خطی با هم داشته باشند.

۳- همبستگی جزئی تنها برای مدل‌های کوچک مفید است. یعنی مدل‌هایی که ۳ یا ٤ متغیر را در بر می‌گیرند.

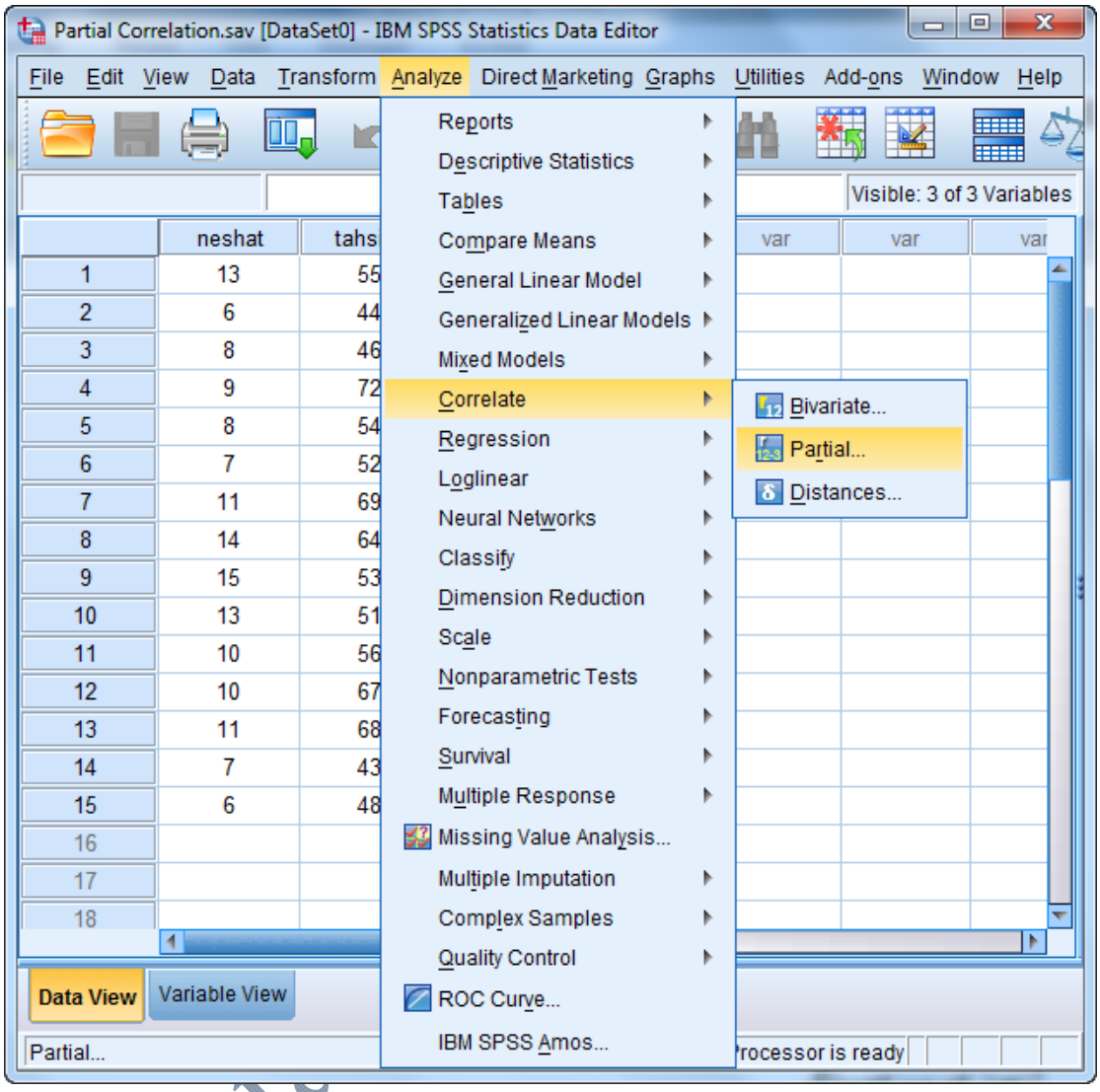
مثال(۲۳): پژوهشگری قصد دارد تا رابطه بین میزان نشاط و میزان تحصیلات(به سال) را بررسی نماید. با این فرض که میزان درآمد افراد بر رابطه بین میزان نشاط و میزان تحصیلات تاثیرگذار است. پژوهشگر قصد حذف اثر درآمد را بر این رابطه دارد. بدین منظور تعداد ۱۵ نفر را به تصادف انتخاب و اطلاعات زیر حاصل شده است.

میزان نشاط	۶	۷	۱۱	۱۰	۱۰	۱۳	۱۵	۱۴	۱۱	۷	۸	۹	۸	۶	۱۳
میزان تحصیلات	۴۸	۴۳	۶۸	۶۷	۵۶	۵۱	۵۳	۶۴	۶۹	۵۲	۵۴	۷۲	۴۶	۴۴	۵۵
درآمد(هزار تومان)	۸۱۶	۶۹۵	۳۲۴	۸۷۹	۷۷۶	۶۹۸	۳۸۳	۶۷۵	۳۰۸	۸۷۹	۳۸۶	۴۳۱	۷۱۰	۴۶۹	۸۵۳

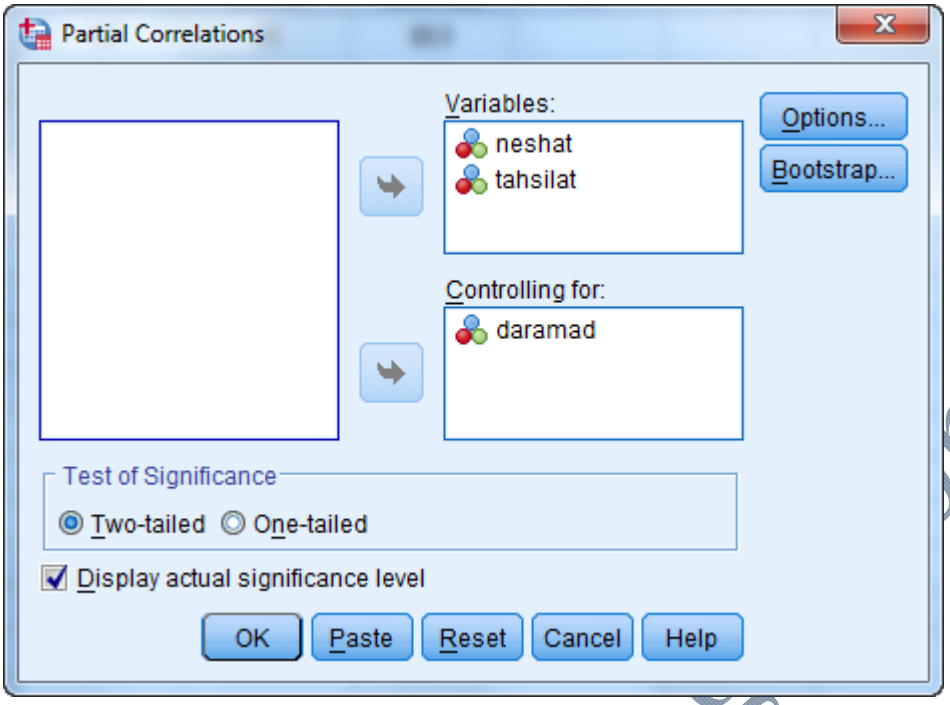
روش ورود اطلاعات در نرم افزار SPSS

سه ستون تحت عناوین میزان نشاط(neshat)، میزان تحصیلات (tahsilat) و درآمد(daramad) ایجاد می‌کنیم.

سپس مسیر: Analyze / Correlate / Bivariate را انتخاب کنید.



در کادر باز شده دو متغیر neshat و tahsilat را به کادر Variables و متغیر daramad را به کادر Controlling for: منتقل کنید.



سپس بر روی گزینه Ok کلیک کنید تا خروجی زیر حاصل شود.

Partial Corr

Correlations

Control Variables			neshat	tahsilat
daramad	neshat	Correlation	1.000	.391
		Significance (2-tailed)	.	.167
		df	0	12
	tahsilat	Correlation	.391	1.000
		Significance (2-tailed)	.167	.
		df	12	0

خروجی بالا را تفسیر کنید؟

همبستگی بین متغیرهای اسمی:

در این بخش با سوالاتی بدین صورت روبرو هستیم:

- آیا رابطه‌ای 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؟

در چنین مسائلی که متغیرهای مورد بررسی اسمی بوده و قابل اندازه‌گیری نبوده و تنها شمارش پذیرند. برای بررسی رابطه یا استقلا بین متغیرها از آزمون خی‌دو/ مجذور کی/ کای اسکوئر (Chi-Sauare) استفاده خواهیم کرد.

جداول توافقی:

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

مثال(۲٤): داده‌های زیر نتایج یک بررسی بر روی ۳۰ نفر است که جنسیت و مصرف سیگار پرسیده شده است:

جنسیت	**زن**	**زن**	**زن**	**مرد**	**مرد**	**مرد**	**زن**	**مرد**	**مرد**	**زن**	**مرد**	**زن**	**زن**	**زن**	**مرد**
مصرف سیگار	**خیر**	**خیر**	**بلی**	**بلی**	**بلی**	**بلی**	**خیر**	**خیر**	**خیر**	**خیر**	**خیر**	**خیر**	**بلی**	**خیر**	**بلی**
جنسیت	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**مرد**	**زن**	**زن**	**مرد**	**زن**	**مرد**
مصرف سیگار	**بلی**	**خیر**	**خیر**	**خیر**	**بلی**	**بلی**	**بلی**	**خیر**	**خیر**	**بلی**	**خیر**	**بلی**	**خیر**	**خیر**	**بلی**

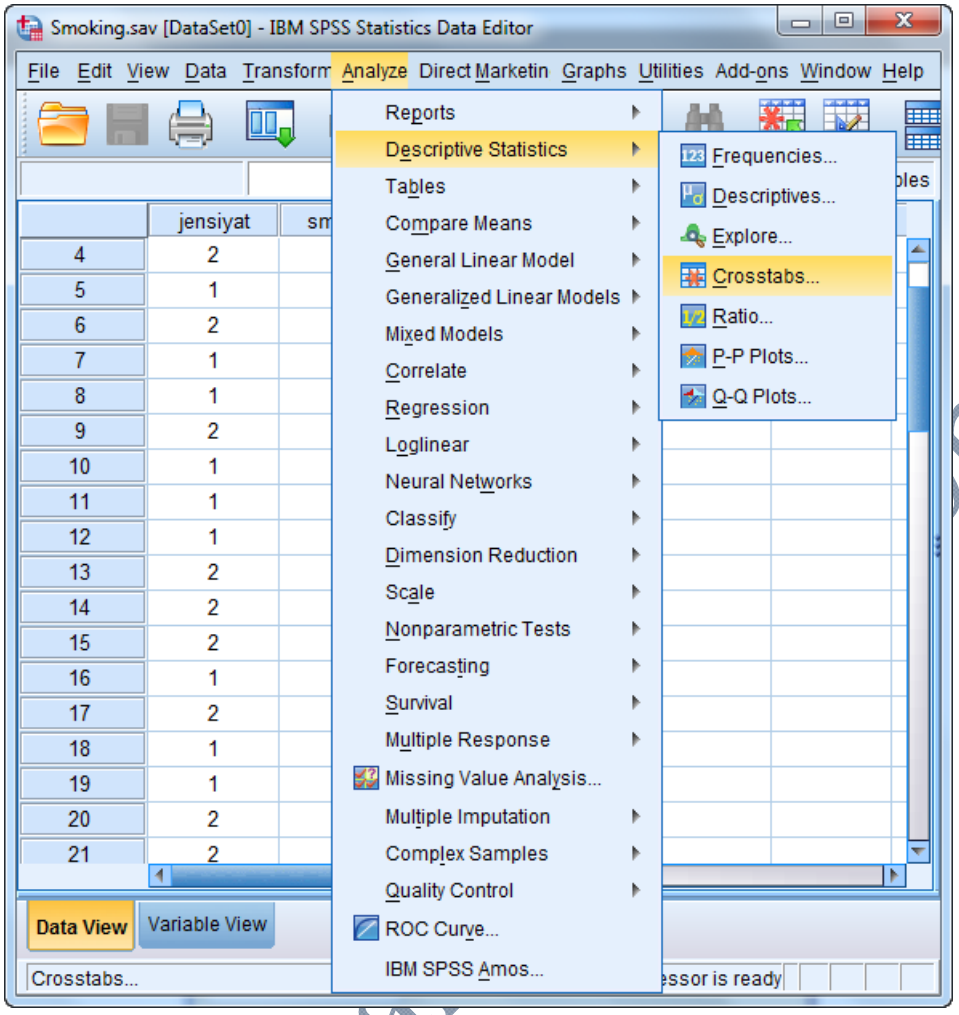
روش ورود اطلاعات در نرم افزار SPSS

دو ستون تحت عناوین جنسیت(jensiyat) و مصرف سیگار(smoking) ایجاد می‌کنیم. برای ورود اطلاعات به نرم افزار کدهای زیر را برای متغیرها در نظر می‌گیریم:

جنسیت: مرد(کد ۱)، زن(کد ۲)

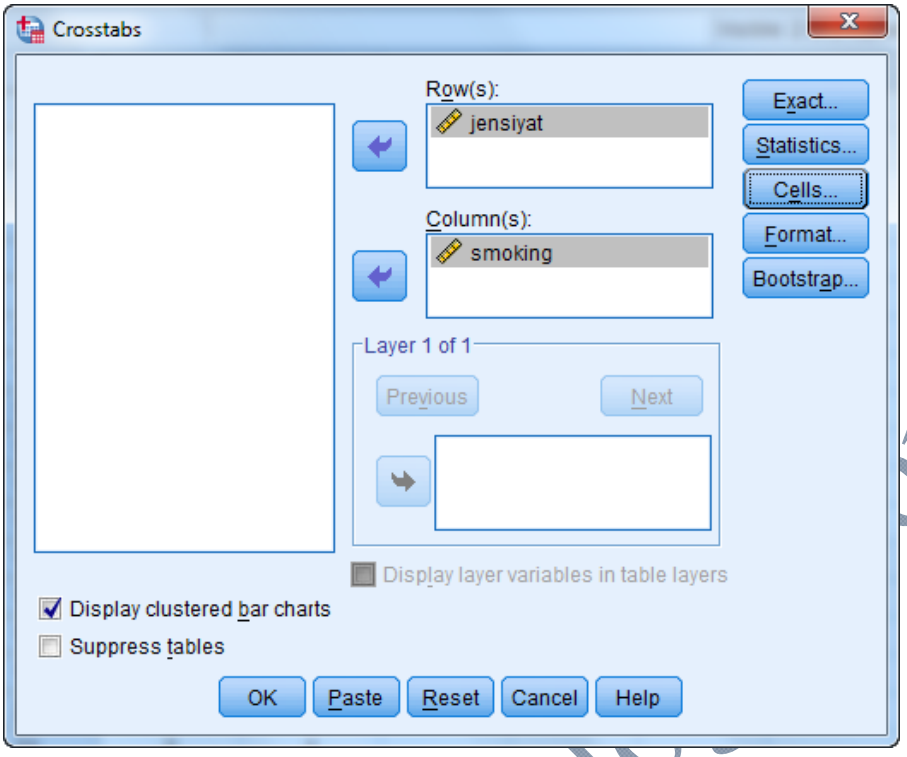
مصرف سیگار: بلی(کد ۱)، خیر(کد ۰)

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

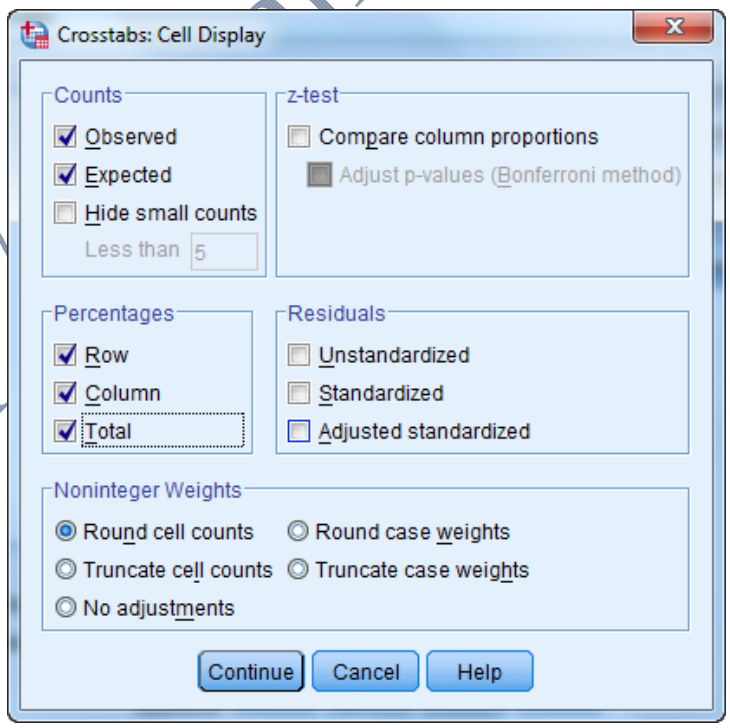


تا کادر زیر باز شود.

مطابق شکل متغیر jensiyat را به کادر Row(s) و متغیر smoking را به کادر Column(s) منتقل کنید. سپس گزینه Display Clustered bar Charts را تیک بزنید تا نمودار خوش‌ای جنسیت و مصرف سیگار را نمایش دهد.



برای نمایش مقادیر مورد انتظار و درصدهای هر سلول روی گزینه Cells... کرده و مطابق شکل زیر موارد مورد نظر را تیک بزنید.



سپس دکمه Continue و Ok را بزنید تا خروجی به صورت زیر حاصل شود.

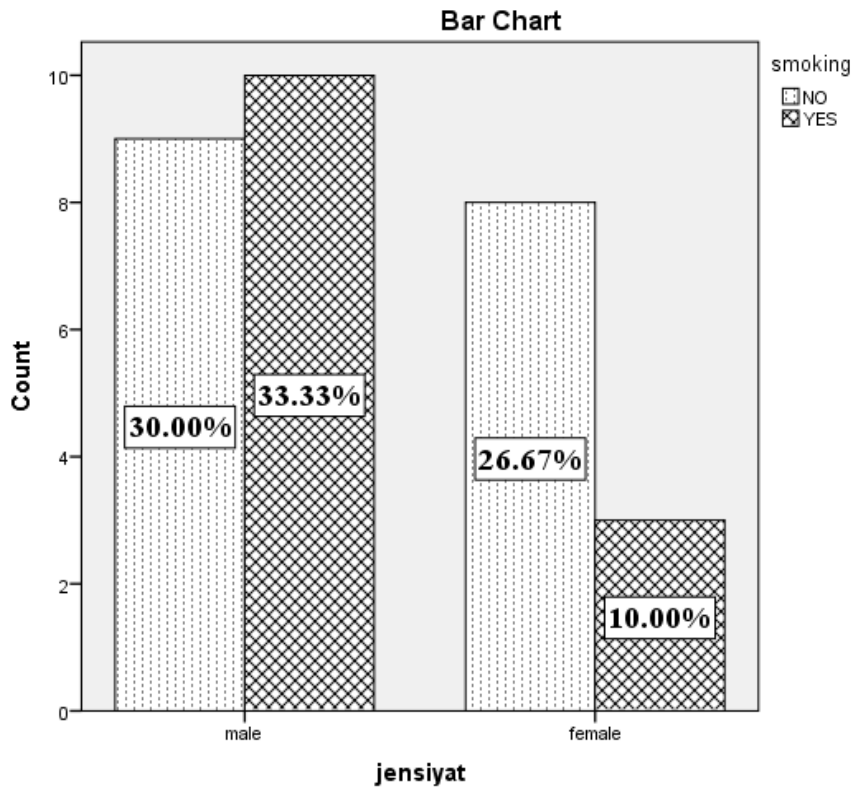
Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
jensiyat * smoking	30	100.0%	0	0.0%	30	100.0%

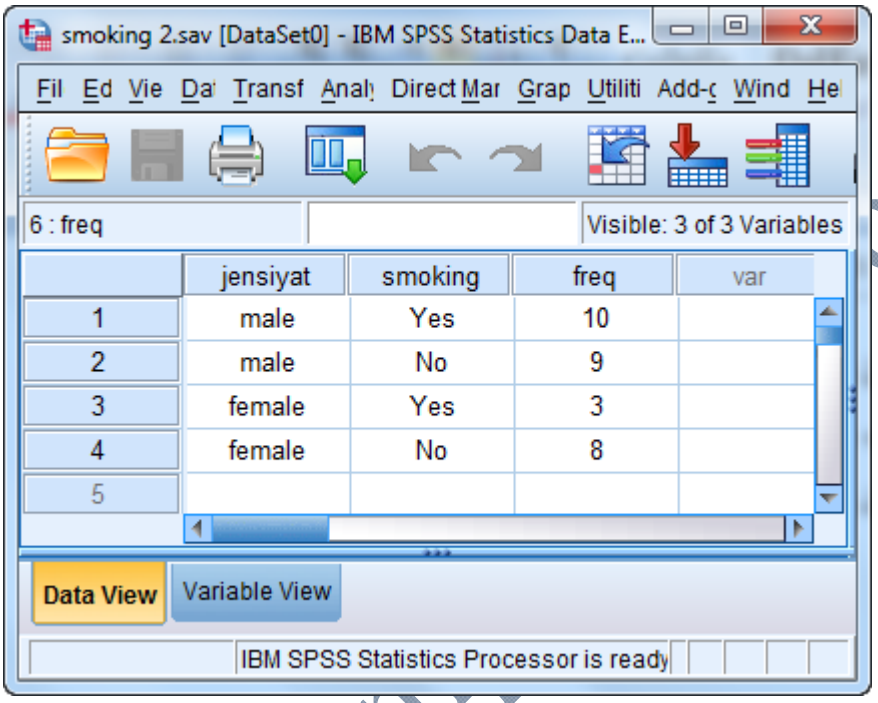
jensiyat * smoking Crosstabulation

			smoking		Total
			NO	YES	
jensiyat	male	Count	9	10	19
		Expected Count	10.8	8.2	19.0
		% within jensiyat	47.4%	52.6%	100.0%
		% within smoking	52.9%	76.9%	63.3%
		% of Total	30.0%	33.3%	63.3%
	female	Count	8	3	11
		Expected Count	6.2	4.8	11.0
		% within jensiyat	72.7%	27.3%	100.0%
		% within smoking	47.1%	23.1%	36.7%
		% of Total	26.7%	10.0%	36.7%
Total		Count	17	13	30
		Expected Count	17.0	13.0	30.0
		% within jensiyat	56.7%	43.3%	100.0%
		% within smoking	100.0%	100.0%	100.0%
		% of Total	56.7%	43.3%	100.0%



رسم جدول فراوانی بدون داشتن داده‌های خام:

مثال(۲۵): 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. داده‌ها را به ترتیب زیر در SPSS وارد می‌کنیم:

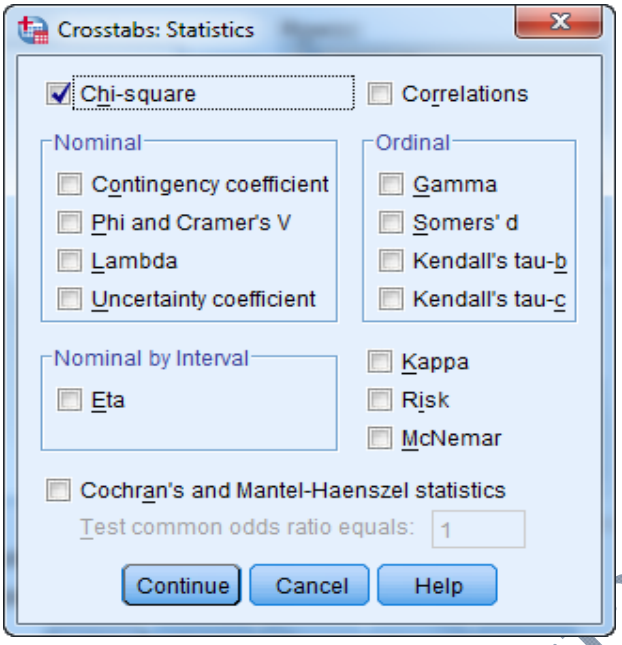


سپس به داده ها بر حسب متغیر freq وزن می‌دهیم و مراحل رسم جدول توافقی را طی می‌کنیم.

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 رابطه‌ای 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 فرد به سیگار ندارد. ولی در صورتیکه دو متغیر مستقل نباشند، با هم رابطه دارند. برای آزمون استقلال دو متغیر از آزمون خی‌دو استفاده خواهیم کرد.

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



نتیجه آزمون استقلال برای داده‌های 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	1.824[a]	1	.177		
Continuity Correction[b]	.938	1	.333		
Likelihood Ratio	1.876	1	.171		
Fisher's Exact Test				.259	.167
N of Valid Cases	30				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.77.
b. Computed only for a 2x2 table

نتایج آزمون خی دو پیرسون در سطر اول آمده است. فرض صفر در این آزمون استقلال متغیرهاست و در صورتیکه فرض صفر رد شود، نشان دهنده وجود رابطه بین دو متغیر خواهد بود.

در این مثال سطح معناداری آزمون خی دو پیرسون برابر با ۰/۱۷۷ و بزرگتر از مقدار ۰/۰۵ شده است. در نتیجه فرض صفر پژوهش یعنی استقلال بین جنسیت و مصرف سیگار رد نمی‌شود. و این بدان معنی است که؛ جنسیت افراد تاثیری بر مصرف سیگار آنها ندارد.

فصل ششم:

رگرسیون

مقدمه

رگرسیون خطی یکی از تکنیک‌های پیچیده آماری برای داده‌هایی است که معمولاً در سطح سنجش فاصله-ای می‌باشند. رگرسیون خطی به دو صورت رگرسیون خطی ساده و رگرسیون خطی چند متغیره مطرح می-گردد.

رگرسیون خطی ساده به پیش‌بینی مقدار یک متغیر وابسته بر اساس یک متغیر مستقل می‌پردازد. اما رگرسیون چند متغیره روشی است برای بررسی تاثیر جمعی و فردی دو یا چند متغیر مستقل (X_i) در تغییرات یک متغیر وابسته (Y). از انجا که وظیفه اصلی علم، پیش بینی و تبیین پدیده‌هاست. بنابراین در تحقیقاتی که بر پیش‌بینی یا تبیین ناظرند، تحلیل رگرسیون می‌تواند نقش بارزی ایفا کند. در تحقیقات پیش-بینی، عمدتاً کاربرد علمی مورد تاکید است. در این نوع تحثیقات محقق می‌کوشد تا بر اساس اطلاع از یک یا چند متغیر مستقل، به یک معادله رگرسیونی دست یابد و از ان برای پیش‌بینی مقادیر متغیر وابسته استفاده کند.

روش‌های رگرسیون خطی

برای ورود متغیرهای رگرسیونی به مدل ۵ روش وجود دارد که محقق بسته به هدف تحلیل خود می‌تواند از یکی از این ۵ روش استفاده کند که معمولاً نتایج این ۵ روش مشابه یکدیگرند. در زیر به تفکیک به شرح ماهیت هر روش پرداخته شده می‌شود:

۱- روش همزمان(Enter Method)

در این روش کلیه متغیرهای مستقل به طور همزمان وارد مدل می‌شوند تا تاثیر کلیه متغیرهای مهم و غیر مهم بر متغیر وابسته مشخص گردد. یک از مشکلات روش همزمان این است که چون تمامی متغیرها بدون توجه به ضریب همبستگی‌شان با متغیر وابسته وارد معادله می‌شوند، بنابراین احتمالاً متغیرهایی هم که حضورشان در معادله معنی دار نیست در آن باقی می‌مانند که در اثر این حضر نابجا، مقادیر F و R^2 کاهش می‌یابد.

۲- روش گام به گام(Stepwise Method)

این روش همانند روش Forward متغیرها را یک به یک وارد مدل می‌کند. در این روش ورود متغیرها به مدل را تا زمانی انجام می‌دهیم که معنی داری متغیر به ۹۵ درصد برد. یعنی سطح خطا ۵ درصد گردد.

فرق این روش با روش Forward در این است که در روش Forward متغیرهای وارد شده در تحلیل در معادله باقی می‌ماندند. ولی در روش Stepwise با ورود متغیر جدید، متغیرهایی که قبلاً وارد مدل شده اند از نو آزموده می‌شوند تا مشخص شود که آیا هنوز هم حضور انها در مدل به موفقیت کمک میکند یا خیر؟ بنابریان امکان دارد برخی از متغیرهایی که در مراحل قبلی وارد مدل شده بودن در مراحل بعدی حذف شوند.

۳-روش حذف(Remove Method)

با این روش می‌توان متغیرهای یک بلوک را از مدل رگرسیونی حذف کرد. بنابراین این روش را نمی‌توان به عنوان روش اولین بلوک به کار برد. زیرا متغیرها بایستی در یکی از بلوک های قبلی وارد مدل شده و سپس در بلوک های بعدی با انتخاب این روش آن ها را حذف نمود. روش حذف کاملاً مانند روش Enter است، اما کاربرد چندانی در رگرسیون چند متغیره ندارد.

٤-روش پس رونده(Backward Method)

در این روش همانند روش Enter ابتدا کلیه متغیرهای مستقل وارد معادله شده و اثر کلیه متغیرها بر روی متغیر وابسته شنجیده می‌شود. اما برخلاف روش Enter، در این روش، به مرو متغیرهای ضعیف‌تر و کم اثرتر یک پس از دیگری از معادله خارج شده و در نهایت این امر تا زمانی ادامه می‌یابد که خطای ازمون معناداری به ۱۰ درصد برسد.

۵-روش پیش‌رونده(Forward Method)

این روش، ابتدا همبستگی ساده بین هر یک از متغیرهای مستقل را با متغیر وابسته محاسبه می‌کند. سپس، متغیر مستقلی که بیشترین همبستگی را با متغیر وابسته دارد و به عبارتی بیشترین مقدار واریانس آن را تبیین می‌کند، وارد تحلیل می‌شود. دومین متغیری که وارد مدل می‌شود، متغیری است که پس از تفکیک متغیر اول، بیش‌ترین ضریب همبستگی را با متغیر وابسته داشته باشد. این روش تا زمانی ادامه پیدا می‌کند که خطای آزمون به ۵ درصد برسد.

رگرسیون خطی ساده/ دو متغیره (Simple Linear Regression)

رگرسیون خطی ساده زمانی مورد استفاده قرار می‌گیرد که یک متغیر وابسته (y) و یک متغیر مستقل (x) داریم.

معادله خط رگرسیون در این حالت به صورت زیر است:

$$y = \beta_0 + \beta_1 x_i$$

که در آن β_0 مقدار ثابت و β_1 شیب خط رگرسیونی و یا ضریب x است.

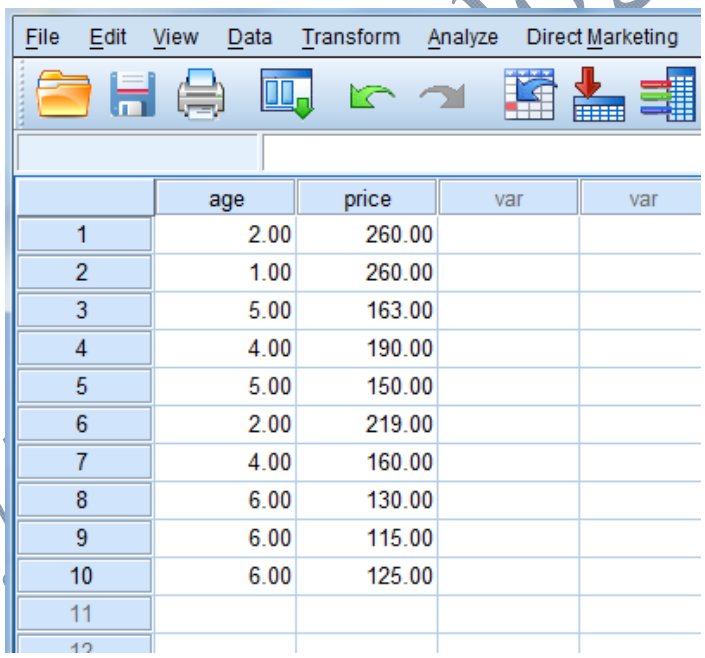
مثال(۲۶): در جدول زیر داده های مربوط به قیمت ۱۰ ناوچه رزمی که ۱ تا ۶ سال از تولید آنها گذشته است، اورده شده است.

سن	۲	۱	۵	٤	۵	۲	٤	٦	٦	٦
قیمت فروش	۲٦۰	۲٦۰	۱٦۳	۱۹۰	۱۵۰	۲۱۹	۱٦۰	۱۳۰	۱۱۵	۱۲۵

برای بررسی رابطه بین سن و قیمت فروش ناوچه از روش رگرسیون خطی ساده استفاده خواهد شد.

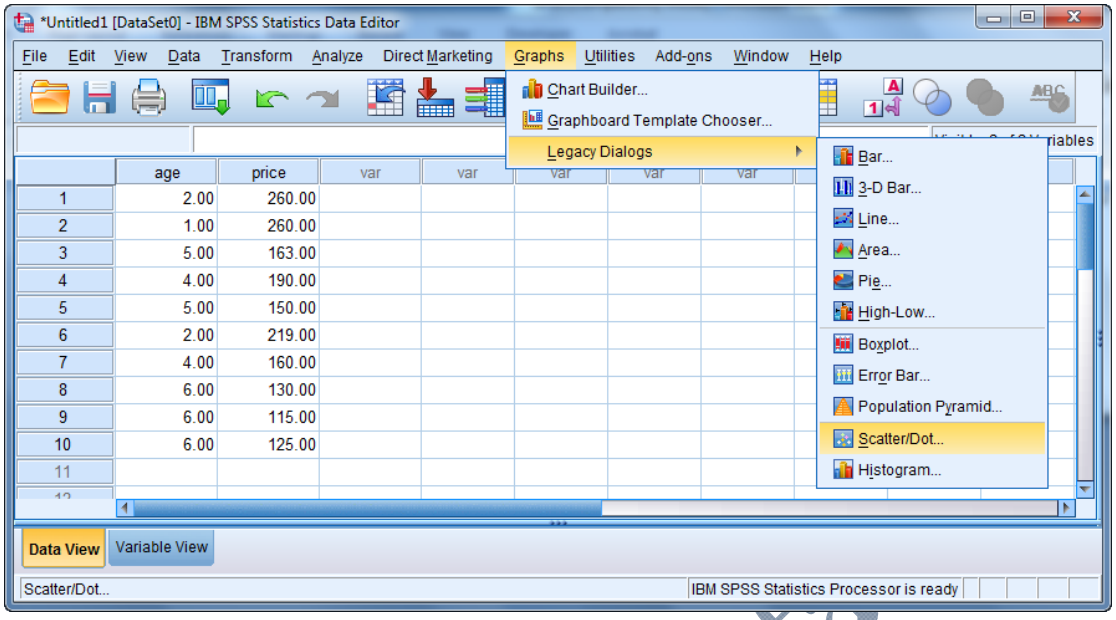
روش ورود اطلاعات در نرم افزار SPSS

دو ستون تحت عناوین سن(age) و قیمت فروش(price) ایجاد می‌کنیم. همانند شکل زیر:

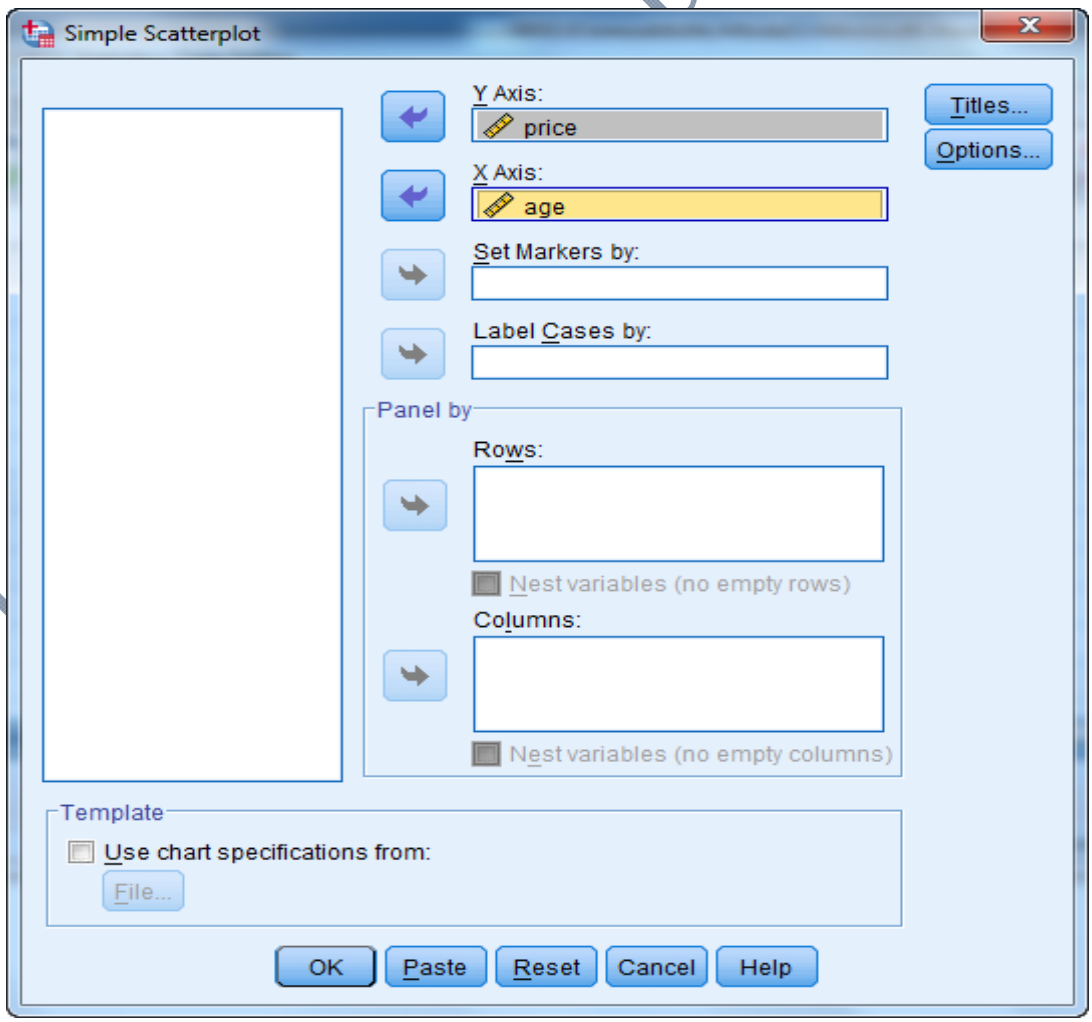


	age	price	var	var
1	2.00	260.00		
2	1.00	260.00		
3	5.00	163.00		
4	4.00	190.00		
5	5.00	150.00		
6	2.00	219.00		
7	4.00	160.00		
8	6.00	130.00		
9	6.00	115.00		
10	6.00	125.00		
11				

وجود یک رابطه خطی بین دو متغیر age و price ابتدا به وسیله نمودار پراکنش(scatterplot) مورد بررسی قرار می‌گیرد. همانند مسیر زیر:

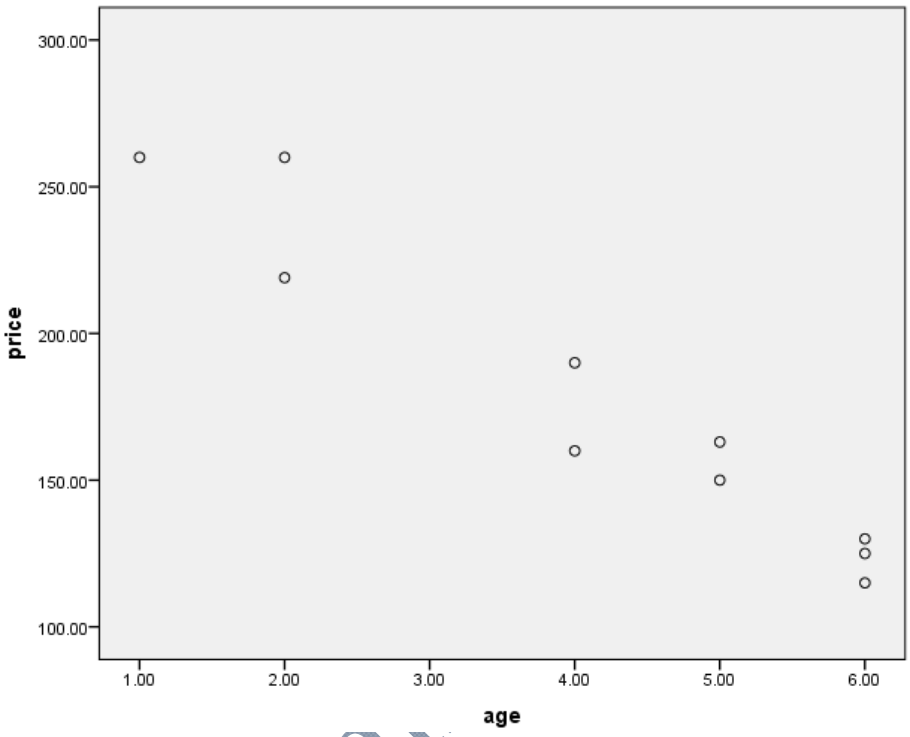


سپس در صفحه باز شده گزینه Simple Scatter را همانند شکل زیر انتخاب و سپس گزینه Define را کلیک می‌کنیم.



در کادر باز شده متغیرها را به کادرهای نمایش داده شده، همانند شکل بالا منتقل خواهیم کرد. سپس دکمه OK را کلیک میکنیم.

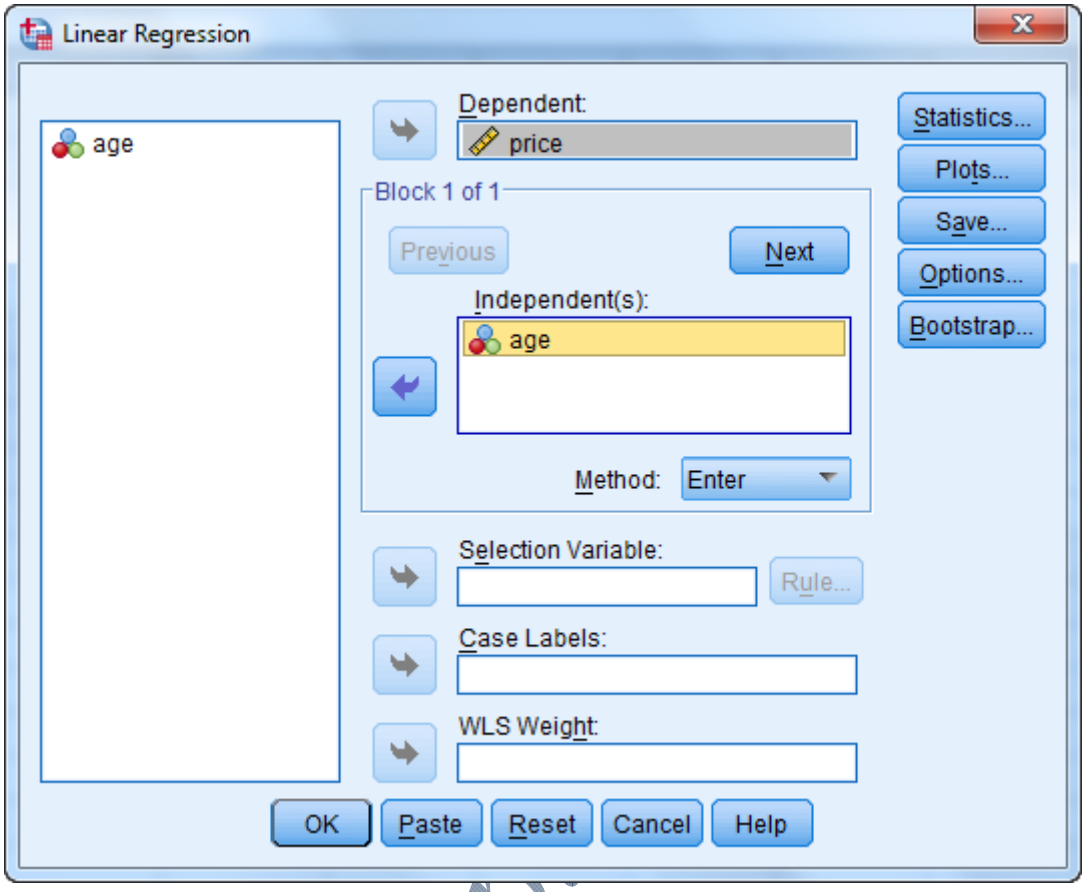
خروجی نرم افزار به صورت شکل زیر است. همانگونه که از نمودار پراکنش داده ها مشخص است، به نظر می‌رسد که یک رابطه خطی معکوس(منفی) بین متغیرها وجود داشته باشد.



در ادامه برای برازش یک مدل خطی رگرسیون ساده به دو متغیر age و price مسیر زیر را دنبال می‌کنیم.

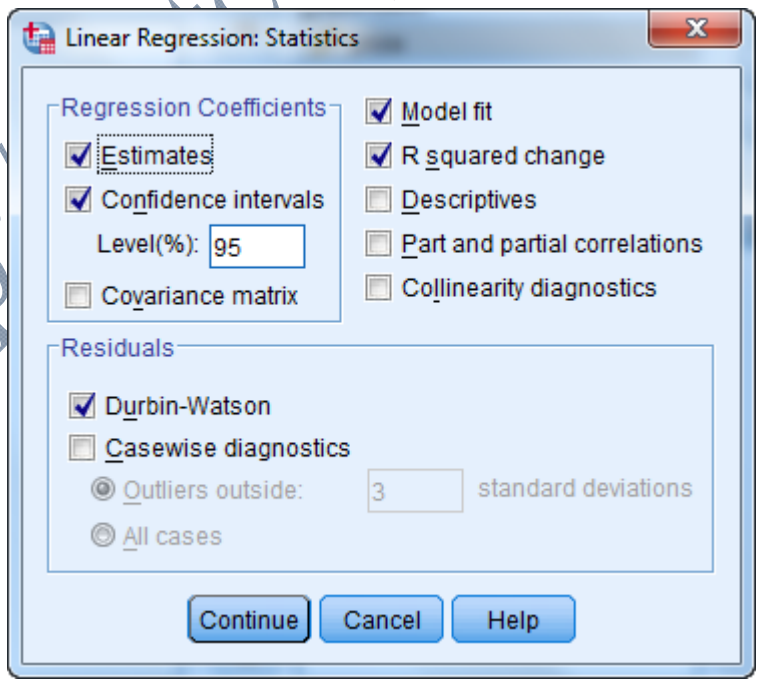
Analyze / Regession / Linear…

در کادر باز شده متغیر Price را کادرDependent و متغیرage را به کادر Independent(s) انتقال می‌دهیم.



سپس روی گزینه Statistics کلیلک کرده و گزینه های مورد نظر را همانند شکل زیر انتخاب می‌کنیم.

سپس روی گزینه Continue و OK کلیک می‌کنیم.



Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	age[b]	.	Enter

a. Dependent Variable: price

b. All requested variables entered.

اولین جدول خروجی تحت عنوان(Variables Entered/Removed[a]) می‌باشد که در آن تعداد مدل، متغیرهای وارد شده و متغیرهای خارج شده از مدل و همچنین روش مورد استفاده که در این قسمت به صورت پیش فرض روش Enter بوده است را نمایش می‌دهد.

دومین جدول خروجی تحت عنوان(Model Summary) در واقع جدول آماره های مربوط به برازش مدل می‌باشد.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.968[a]	.937	.929	14.24653	.937	118.533	1	8	.000	1.451

a. Predictors: (Constant), age

b. Dependent Variable: price

۱- **Model**: تعداد مدل تشکیل شده را نشان می‌دهد که در این مثال برابر با یک است.

۲- **R:** به ضریب همبستگی چندگانه معروف است و میزان همبستگی چندگانه بین متغیرهای مستقل و وابسته را نشان می‌دهد.مقدار این همبستگی بین صفر تا ۱ نوسان دارد. هر چه مقدار این همبستگی به (۱) نزدیک تر باشد نشان از همبستگی قوی بین متغیرهای مستقل و متغیر وابسته دارد. در این مثال مقدار R برابر با ۰/۹۶۸ است که مقدار قابل قبول بالایی را نشان می‌دهد.

۳- **R Square:** به مجذور ضریب همبستگی چندگانه یا ضریب تعیین معروف است و با علامت R^2 نوشته می‌شود. مقدار این ضریب نیز بین (۰) تا (۱) در نوسان و هر چه به مقدار (۱) نزدیک‌تر باشد نشان از آن دارد که متغیرهای مستقل توانسته‌اند که میزان زیادی از متغیر وابسته را تبیین کنند و برعکس. در این مثال مقدار این ضریب برابر با ۰/۹۳۷ است که نشان می‌دهد که؛ ۹۳/۷ درصد واریانس متغیر Price توسط متغیر age پیش بینی می‌شود.

٤- **Adjusted R Square:** اشکال ضریب تعیین این است که میزان موفقیت مدل را بیش از اندازه برآورد می‌کند و کمتر تعداد متغیرهای مستقل و همین‌طور حجم نمونه را در نظر می‌گیرد. همچنین، ضریب تعیین تعداد درجات آزادی را به حساب نمی‌آورد. از این رو برخی آماردانان

ترجیح می‌دهند تا از شاخص دیگری به نام ضریب تعیین تعدیل شده استفاده کنند. ضریب تعیین تعدیل شده که آن را با R^2_{adj} نیز نمایش می‌دهند مقدار ضریب تعیین را به منظور انعکاس بیشتر میزان نیکوئی برازش مدل تصحیح می‌کند. برای تفسیر ضریب تعیین معمولاً از این مقدار استفاده می‌شود. چون در این ضریب، مقدار ضریب تعیین با درجات آزادی تعدیل شده است. در این مثال مقدار ضریب تعیین تعدیل شده برابر با ۰/۹۲۹ است.

۵- **Durbin-Watson:** یکی از مفروضه‌های اساسی در تحلیل رگرسیون چندگانه، استقلال متغیرهای مستقل و یا به عبارت دیگر عدم ارتباط نمره‌های خطای متغیرهای مستقل با یکدیگر است. که این مفروضه توسط آزمون دوربین واتسون بررسی می‌شود. آزمون دوربین-واتسون که نتیجه ان در آخرین ستون سمت راست امده است این مفروضه را آزمون می‌کند. به طور سرانگشتی می‌توان گفت که اگر مقدار آماره‌ی این آزمون بین ۱/۵ تا ۲/۵ قرار داشته باشد، می‌توان استقلال مشاهدات را پذیرفت و آزمون را ذیل کرد. مقدار آماره این آزمون در پژوهش حاضر ۱/٤٥۱ است که با توجه به حساس بودن آزمون به حجم نمونه و تعداد متغیرهای مستقل می‌توان استقلال مشاهدات را در این مثال پذیرفت.

جدول سوم تحت عنوان(ANOVA) که در زیر داده شده است، نتایج تحلیل واریانس مدل رگرسیونی برازش داده شده را نشان می‌دهد.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24057.891	1	24057.891	118.533	.000[b]
	Residual	1623.709	8	202.964		
	Total	25681.600	9			

a. Dependent Variable: price

b. Predictors: (Constant), age

در این جدول، منبع تغییرات متغیر وابسته در دو منبع رگرسیون(Regression) و باقیمانده (Residual) نشان داده شده و برای هر یک از این منابع، مجموع مجذورات، درجه آزادی، میانگین مجذورات آمده است.

۱- منبع رگرسیون: اطلاعات مربوط به میزان تغییرات متغیر وابسته را که در نتیجه مدل تحقیق است، نشان می‌دهد.

۲- منبع باقیانده: اطلاعات مربوط به میزان تغییرات متغیر وابسته را که خارج از مدل تحقیق است، نشان می‌دهد.

بنابراین هر چه مقدار مجموع مجذورات باقیمانده کوچک‌تر از مجموع مجذورات رگرسیون باشد، نشان دهنده قدرت تبیین‌گری بالای مدل در توضیح تغییرات متغیر وابسته است.

برای بررسی نیکویی برازش مدل رگرسیونی برازش داده شده به مدل، از مقدار Sig ستون اخر استفاده خواهیم کرد. همانطور که از جدول بالا مشخص است، مقدار سطح معناداری برابر با صفر می‌باشد که در سطح ۱٪ معنادار است. و این نشان می‌دهد که متغیر مستقل از قدرت تبیین بالای برخوردار بوده و می‌تواند درصد بالایی از واریانس متغیر وابسته را تبیین نماید. به عبارت دیگر مدل رگرسیونی تحقیق مدل خوبی است و به کمک ان قادریم تا تغیرات متغیر وابسته را به کمک متغیر مستقل مورد نظر تبیین کنیم.

جدول بعدی تحت عنوان (Coefficients) نتایج مربوط به ضریب تاثیر رگرسیونی متغیر مستقل age بر متغیر وابسته price را نشان می‌دهد.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	291.602	11.433		25.506	.000	265.238	317.966
	age	-27.903	2.563	-.968	-10.887	.000	-33.813	-21.993

a. Dependent Variable: price

اولین مقدار جدول عدد ثابت(Constant) است که همان عرض از مبدا است و میزان تغییران متغیر وابسته را بدون دخالت متغیر(های) مستقل نشان می‌دهد(یعین زمانیکه متغیر(های) مستقل صفر باشند).

ضرایب رگرسیونی دو دسته اند:

۱) **ضرایب رگرسیونی استاندارد نشده**(B): ضرایب مربوط به مدل رگرسیونی برآورد شده می‌باشند. که در این مثال برابر با مقدار ۲۷/۹۰۳- است.

۲) **ضرایب رگرسیونی استاندارد شده** $(Beta = \beta)$: از آنجا که در تحلیل رگرسیون، مقیاس اغلب متغیرهای مستقل از واحدهای متفاوتی تشکیل شده است، بنابراین به راحتی نمی‌توان به مقایسه سهم هر متغیر مستقل در تبیین تغییرات یا واریانس متغیر وابسته پرداخت. از همین رو، ضرایب رگرسیونی استاندارد شده به ما کمک می‌کنند تا سهم نسبی هر متغیر مستقل را در تبیین تغییرات متغیر وابسته مشخص کنیم.یعنی هرچه مقدار ضریب بتای یک متغیر بیشتر باشد، نقش آن در پیش بینی تغییرات متغیر وابسته بیشتر خواهد بود. از این رو به محققین پیشنهاد می‌شود که در تفسیر نتایج تاثیر

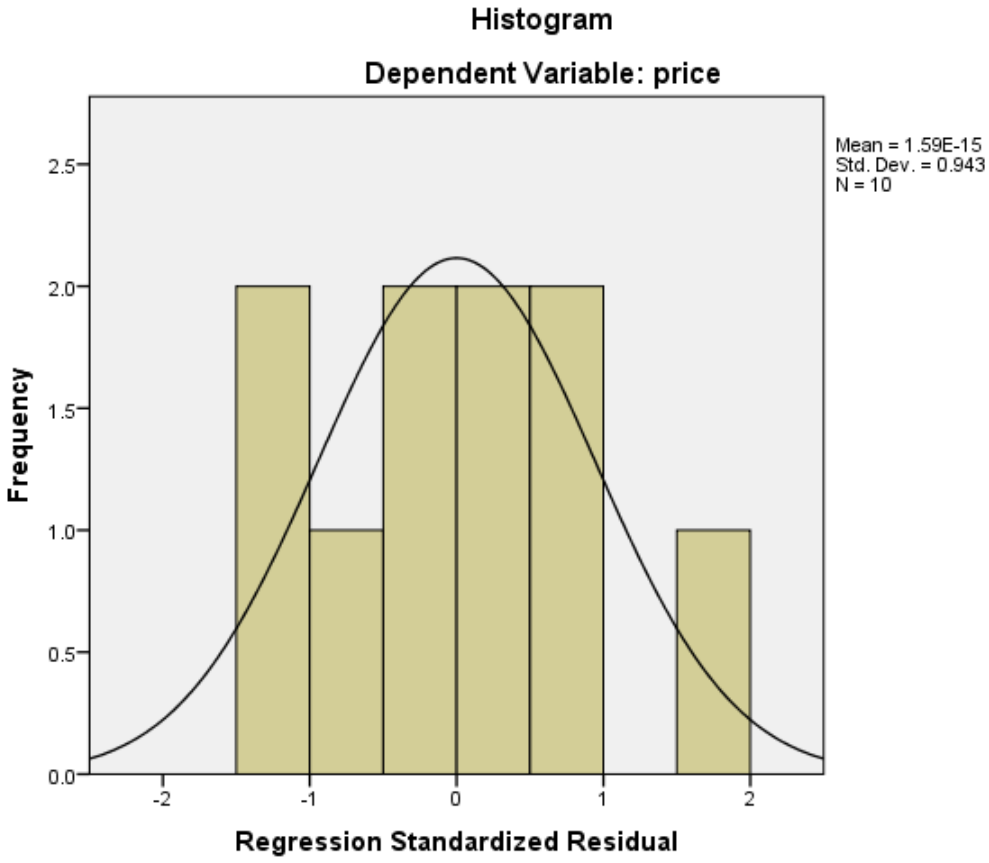
رگرسیونی بر اساس ضرایب آن، به ضرایب رگرسیونی استاندارد شده اشاره شود تا ضرایب رگرسیونی استاندارد نشده.

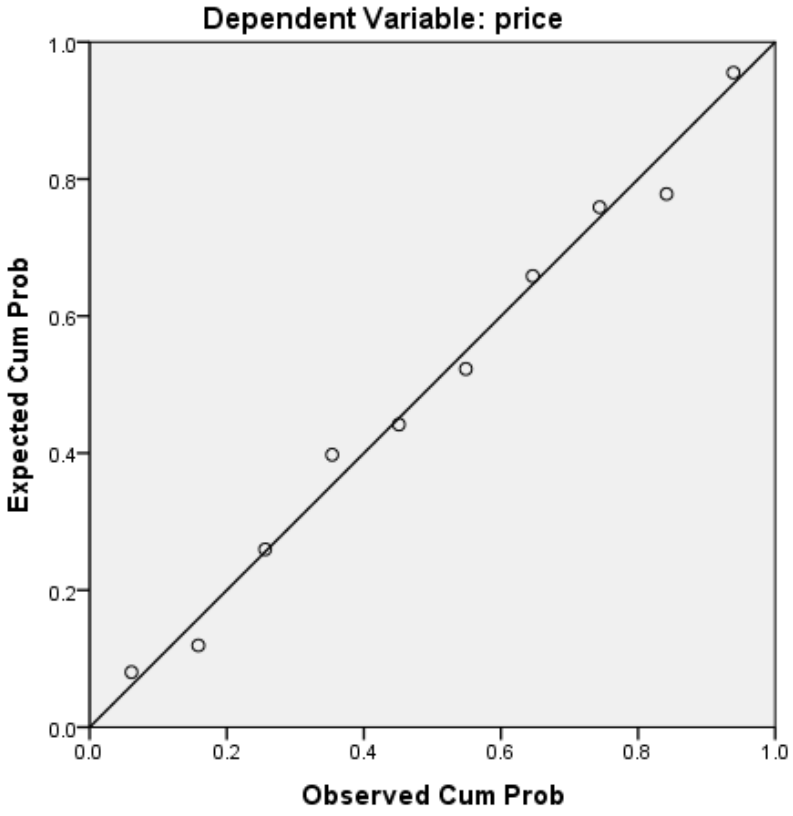
البته در این مثال چون تنها یک متغیر مستقل(age) داریم، عملاً امکان مقایسه سهم نسبی متغیرها در مدل وجود ندارد. اما در تحلیل رگرسیون چند متغیره میتوان از ضریب تاثیر رگرسیونی استاندارد شده برای مقایسه سهم نسبی هر متغیر مستقل در مدل استفاده کرد.

توجه داشته باشیم که ضریب رگرسیونی استاندارد شده بر اساس مقادیر انحراف استاندارد سنجیده می‌شود که در تفسیر آن باید رعایت کنیم. به عنوان مثال، ضریب بتای (۰/۹۶۸-) نشان می‌دهد که تغییر یک انحراف استاندارد در متغیر age، باعث تغییر (۰/۹۶۸-) انحراف استاندارد در متغیر price می‌شود. اما در ضریب رگرسیون استاندارد نشده می‌گوئیم که؛ تغییر یک واحد در متغیر age باعث تغییر (۲۷/۹۰۳-) در متغیر price می‌شود.

آماره t و سطح معناداری(Sig.) اهمیت نسبی حضور هر متغیر مستقل در مدل را نشان می‌دهد. در این مثال مقدار sig.=0.000 شده است که کوچکتر از مقدار ۰/۰۱ می‌باشد. بنابراین ضریب رگرسیونی متغیر age در سطح خطای ۱٪ معنادار شده است.

همچنین در ادامه کار برای بررسی فرضیه نرمال بودن خطاها می توان از نمودار هیستوگرام و نمودار Q-Q Plot استفاده کرد. این دو نمودار برای مثال بالا در زیر داده شده است. همانطور که از هر دو نمودار زیر مشخص است، فرض نرمال بودن باقیمانده ها تایید می‌شود.


Histogram
Dependent Variable: price
Mean = 1.59E-15
Std. Dev. = 0.943
N = 10
Frequency
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: price
Expected Cum Prob
Observed Cum Prob

رگرسیون خطی چندگانه/چند متغیره (Multiple Linear Regression):

با استفاده از رگرسیون چند متغیره، محقق می‌تواند رابطه خطی موجود بین مجموعه‌ای از متغیرهای مستقل با یک متغیر وابسته را به شیوه‌ای مطالعه نماید که در آن، روابط موجود فی‌مابین متغیرهای مستقل نیز مورد ملاحظه قرار گیرد. وظیفه رگرسیون این است که به تبیین واریانس متغیر وابسته کمک کند و این وظیفه تا حدودی از طریق براورد مشارکت متغیرها (دو یا چند متغیر مستقل) در این واریانس به انجام می‌رسد. تحلیل رگرسیون چندمتغیره برای مطالعه تاثیرات چند متغیر مستقل در متغیر وابسته کاملاً مناسب است.

در مجموع، هدف اصلی از کاربرد رگرسیون چند متغیره آن است که ترکیبی خطی از متغیرهای مستقل را به گونه‌ای ایجاد کند که حداکثر همبستگی را با متغیر وابسته نشان دهد. در نتیجه، از این ترکیب خطی می‌توان در جهت پیش‌بینی مقادیر متغیر وابسته استفاده نمود و اهمیت هر یک از متغیرهای مستقل را در پیش‌بینی مورد نظر ارزیابی نمود.

در رگرسیون چند متغیره، مقادیر یک متغیر(متغیر وابسته یا Y) از روی مقادیر دو یا چند متغیر دیگر(متغیرهای مستقل $X_1, X_2, ..., X_K$) براورد می‌شوند. این کار از طریق ساختن یک معادله خطی به شکل عمومی زیر انجام می‌شود:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$

در این معادله پارامتر β_0 مقدار عرض از مبدا یا مقدار ثابت و پارامترهای $(\beta_i : i = 1,2,...,k)$ ضرایب رگرسیونی هستند. این معادله به عنوان معادله رگرسیون خطی چندگانه Y بر اساس متغیرهای $X_1, X_2, ..., X_K$ شناخته می‌شود.

مثال(۲۷): پژوهشگری در یک مطالعه قصد دارد تا تاثیر سه متغیر مستقل حافظه، هوش عملی و هوش کلامی را بر متغیر وابسته پیشرفت تحصیلی بررسی کند. برای این کار وی تعداد ۳۰ دانش آموز را به تصادف انتخاب و متغیرهای بالا در آنها اندازه می‌گیرد.

هوش کلامی	130	128	129	127	126	125	120	119	118	116	115	113	117	120	112
هوش عملی	117	112	123	112	110	114	112	123	133	123	130	127	121	113	122
حافظه	30	27	28	27	25	23	21	19	18	17	15	12	12	14	15
پیشرفت تحصیلی	30	28	27	26	26	25	21	20	20	19	16	14	14	20	14
هوش کلامی	130	122	127	126	123	125	120	119	118	120	115	124	125	123	119
هوش عملی	117	111	123	112	110	114	112	123	133	123	130	123	112	110	123
حافظه	30	27	28	27	25	23	21	19	18	17	15	28	27	25	19
پیشرفت تحصیلی	27	28	27	30	21	25	12	20	20	19	16	27	26	21	20

برای بررسی رابطه بین سن و قیمت فروش ناوچه از روش رگرسیون خطی چندمتغیره استفاده خواهد شد.

روش ورود اطلاعات در نرم افزار SPSS

چهار ستون تحت عناوین هوش کلامی(hosh_kalami)، هوش عملی(hosh_amali)، حافظه(hafezeh) و پیشرفت تحصیلی(Pishraft) ایجاد می‌کنیم. همانند شکل زیر:

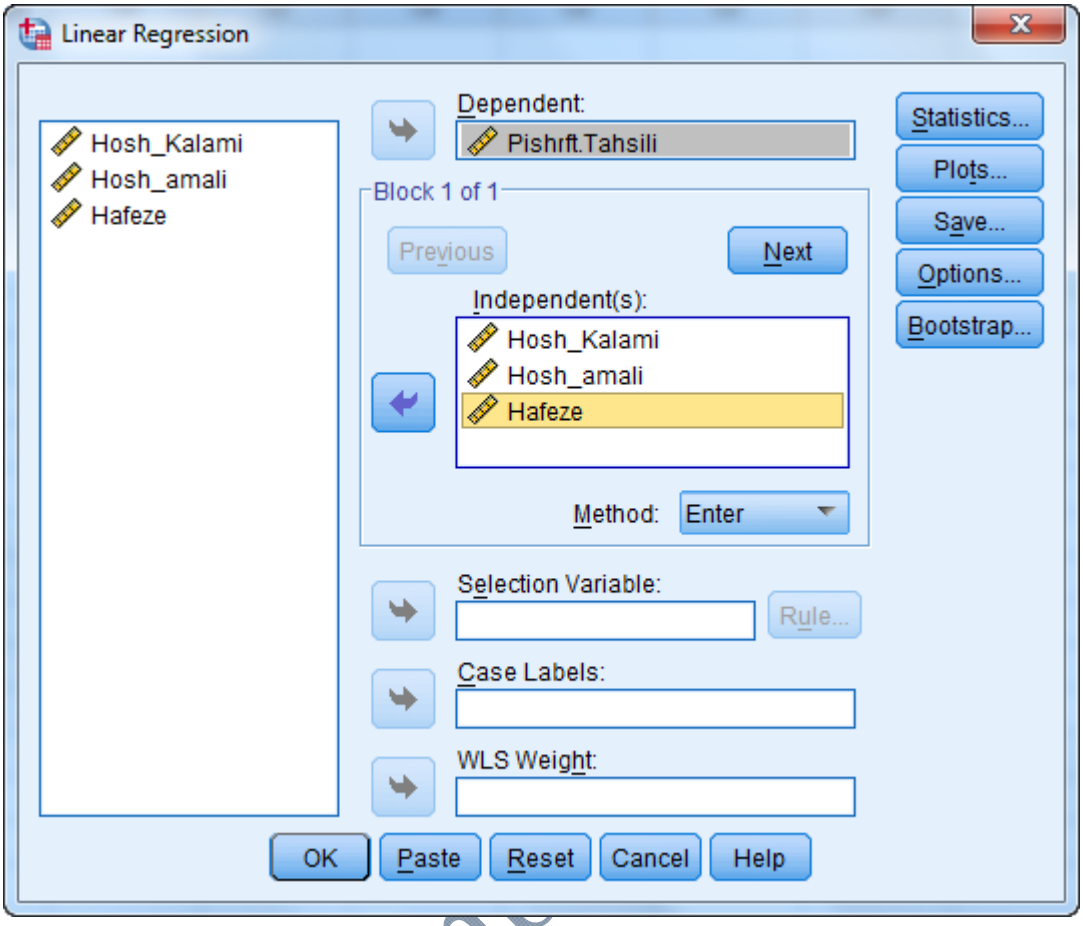
	Hosh_Kalami	Hosh_amali	Hafeze	Pishrft.Tahsili
1	130.00	117.00	30.00	30.00
2	128.00	112.00	27.00	28.00
3	129.00	123.00	28.00	27.00
4	127.00	112.00	27.00	26.00
5	126.00	110.00	25.00	26.00
6	125.00	114.00	23.00	25.00
7	120.00	112.00	21.00	21.00
8	119.00	123.00	19.00	20.00
9	118.00	133.00	18.00	20.00
10	116.00	123.00	17.00	19.00
11	115.00	130.00	15.00	16.00
12	113.00	127.00	12.00	14.00
13	117.00	121.00	12.00	14.00
14	120.00	113.00	14.00	20.00
15	112.00	122.00	15.00	14.00
16	130.00	117.00	30.00	27.00
17	122.00	111.00	27.00	28.00
18	127.00	123.00	28.00	27.00
19	126.00	112.00	27.00	30.00
20	123.00	110.00	25.00	21.00
21	125.00	114.00	23.00	25.00
22	120.00	112.00	21.00	12.00
23	119.00	123.00	19.00	20.00

ata View | Variable View

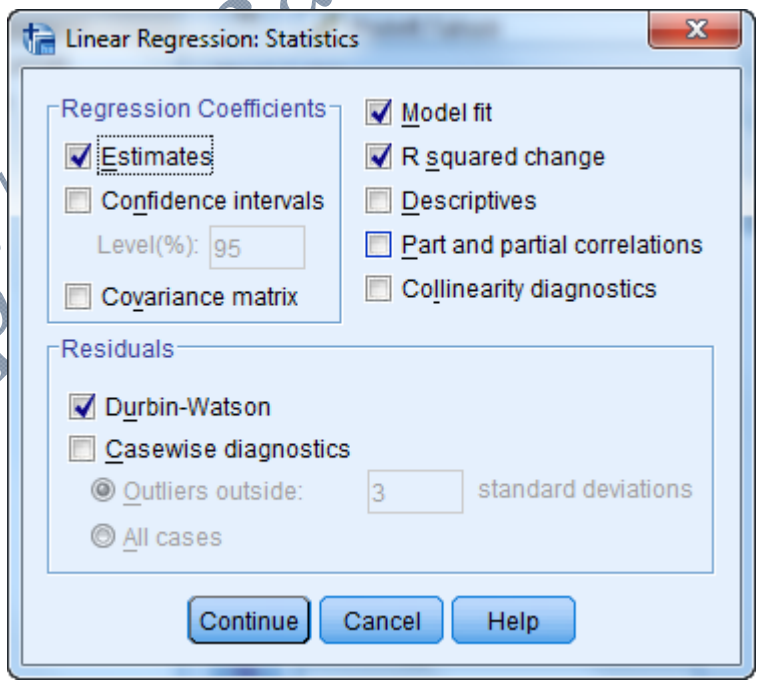
در ادامه برای برازش یک مدل خطی رگرسیون چند متغیره مسیر زیر را دنبال می‌کنیم.

Analyze / Regession / Linear…

در کادر باز شده متغیر Pishraft را کادرDependent و متغیرهای هوش کلامی(hosh_kalami)، هوش عملی(hosh_amali) و حافظه(hafezeh) را به کادر Independent(s) انتقال می‌دهیم.



سپس روی گزینه Statistics کلیلک کرده و گزینه های مورد نظر را همانند شکل زیر انتخاب می‌کنیم.



سپس روی گزینه Continue و OK کلیک می‌کنیم.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Hafeze, Hosh_amali, Hosh_Kalami[b]	.	Enter

a. Dependent Variable: Pishrft.Tahsili
b. All requested variables entered.

اولین جدول خروجی تحت عنوان(Variables Entered/Removed[a]) می‌باشد که در آن تعداد مدل، متغیرهای وارد شده و متغیرهای خارج شده از مدل و همچنین روش مورد استفاده که در این قسمت به صورت پیش فرض روش Enter بوده است را نمایش می‌دهد.

دومین جدول خروجی تحت عنوان(Model Summary) در واقع جدول آماره های مربوط به برازش مدل می‌باشد.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.902[a]	.813	.792	2.35981	.813	37.722	3	26	.000	2.640

a. Predictors: (Constant), Hafeze, Hosh_amali, Hosh_Kalami
b. Dependent Variable: Pishrft.Tahsili

مقدار R Square برابر با ۰/۸۱۳ بوده و نشان دهنده این است که ۸۱/۳ درصد واریانس متغیر پیشرفت تحصیلی توسط سه متغیر مستقل تبیین می‌شود که این مقدار قابل توجهی است. همچنین مقدار اماره دوربین واتسون برابر با ۲/٦٤ شده است که اندکی از مقدار ۲/۵۰ بالاتر است و لی با توجه به حجم نمونه این مقدار از آماره قابل قبول می‌باشد.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	630.180	3	210.060	37.722	.000[b]
	Residual	144.786	26	5.569		
	Total	774.967	29			

a. Dependent Variable: Pishrft.Tahsili
b. Predictors: (Constant), Hafeze, Hosh_amali, Hosh_Kalami

جدول سوم، جدول نیکویی برازش برای مدل رگرسیونی (ANOVA) است. با توجه به مقدار .Sig که برابر با صفر شده است نتیجه می‌شود که مدل از برازش قابل قبولی برخوردار است.

جدول بعدی خروجی ضرایب رگرسیونی است که در قسمت رگرسیون سساده به طور کامل توضیح داده شده است.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-59.880	26.436		-2.265	.032
	Hosh_Kalami	.522	.219	.510	2.389	.024
	Hosh_amali	.075	.073	.106	1.031	.312
	Hafeze	.431	.194	.467	2.221	.035

a. Dependent Variable: Pishrft.Tahsili

همانطور که از داده های جدول بالا مشخص است، ضرایب غیر استاندارد و ضرایب استاندارد برای سه متغیر مستقل و مقدار ثابت آورده شده است. برای هر ضریب آزمونی به عنوان آزمون معناداری انجام می-شود. با توجه به آزمون معناداری ضرایب نتیجه می‌شود که متغیر هوش عملی تاثیر معناداری بر پیشرفت تحصیلی ندارد. همچنین برای مقایسه میزان تاثیر متغیرهای مستقل بر متغیر وابسته پیشرفت تحصیلی از ضرایب استاندارد استفاده می‌شود. بر اساس این ضرایب، هوش کلامی با ضریب استاندارد ۰/۵۱۰ بیشترین تاثیر را بر متغیر پیشرفت تحصیلی دارد.

نمودار هیستوگرام و نمودار Q-Q Plot برای بررسی نرمال بودن باقیمانده ها در زیر داده شده است.