آمار توصیفی
به کمک نرم افزار SPSS

دانشگاه علوم پزشکی جندی شاپور اهواز

تهیه کننده: مریم سیدطبیب

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 داده های اولیه، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد.اگر حجم داده ها 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 ، برنامه های کاربردی در زمینه تجزیه وتحلیلهای آماری است ؛ نرم افزاری آماری با قبلیتهای انجام توصیفی زیبا و گویا از اطلاعات، شامل رسم نمودارها و چارت های گوناگون و محاسبات مربوط به میانگین، انحراف معیار واریانس، میانه و غیره.

کلمه **SPSS** مخفف **Statistical package for social science** (نرم افزار آماری برای علوم اجتماعی) می باشد. این نرم‌افزار که یکی از نرم‌افزارهای تخصصی آمار است، بیشتر به بحث‌های آماری در حیطهٔ علوم اجتماعی، روانشناسی و علوم رفتاری و ... می پردازد.

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
- تکنیک‌های تجزیه و تحلیل سری‌های زمانی
- 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

❖ *مریم سیدطبیب، دانشگاه علوم پزشکی جندی شاپور اهواز*

آمار علم و عمل توسعه دانش انسانی از طریق استفاده از داده های تجربی است. آمار مطالعه لذت بخشی است در باب این موضوع که چگونه می توان جهان ناشناخته ای را با گشودن چند دریچه به روی آن توصیف کرد . با پرداختن به آمار لذت فکر کردن به یک شیوه کاملا جدید را کشف خواهیم کرد.

زمانی به استفاده از علم آمار رو می آوریم که:

۱) بخواهیم داده ها را به صورت یک مجموعه، خلاصه کرده و توصیف نماییم.
۲) بخواهیم اطمینان دهیم که در شرایطی که یک پروژه دقیقاً به همان شکل تکرار شود، همان یافته ها بدست خواهد آمد.

اهداف اصلی آمار عبارتند از:

۱) انجام استنباط درباره جامعه از طریق تجزیه و تحلیل اطلاعات موجود در داده های نمونه ای.
۲) سنجش میزان عدم حتمیتی که در این استنباط ها وجود دارد. علمی که برای رسیدن به هدف های فوق اهمیت دارد، عبارتست از طرح ریزی فرآیند و دامنه نمونه گیری به طور یکه مشاهدات مبنایی برای استخراج استنباط های معتبر تشکیل دهند.

انتخاب یک آزمون آماری :

انتخاب یک آزمون آماری وابسته به موارد زیر است:
۱) پرسش پژوهش
۲) نقشه یا طرح پژوهش
۳) نوع داده های جمع آوری شده

برای انجام یک آزمون آماری با مفاهیم بسیاری روبه رو هستیم. در زیر به بیان تعدادی از این مفاهیم پرداخته و در ادامه چند آزمون آماری متداول را مورد بررسی قرار می دهیم.

یک مفهوم مهم در این رابطه ***نمونه*** (Sample) می باشد. یک نمونه مجموعه ای از مشاهدات (که اغلب به شکل تصادفی انتخاب می شوند) است که می تواند در یک مجموعه مرجع یا ***جامعه*** (Population) از مشاهدات احتمالی انجام شود. در قیاس با قرعه کشی، می توان یک جامعه را مانند اعداد فرض کرد که در گردونه قرار دارند و در این حالت اعدادی که انتخاب می شوند به عنوان نمونه در نظر گرفت. نمونه های گرفته شده از یک جامعه ثابت، می توانند از حیث ویژگیهایشان بسیار متفاوت باشند. بنابراین یک نمونه تصادفی الزاماً نماینده واقعی از جامعه مرجع نیست و ممکن است ویژگیهای بسیار متفاوتی (مانند میانگین و انحراف معیار) ازآن داشته باشد. مقادیر مربوط به ویژگیهای یک نمونه (میانگین، انحراف معیار و غیره) را ***آماره*** (Statistics) و ویژگیهای معادل آنها در جامعه اصلی را ***پارامتر*** (Parameters) می نامند. در تحقیقات ، ما همیشه به دنبال ویژگیهای جامعه اصلی (پارامترها) و نه ویژگیهای نمونه (آماره) هستیم.

استنباط آماری :

استنباط آماری که در واقع یک نوع نتیجه گیری کلی از جزء به کل است، در معرض آزمایش و خطاست. یک جنبه از استنباط آماری محاسبه برآوردهایی (Estimates) از پارمترهای جامعه مانند میانگین جامعه از طریق آماره های نمونه ها مانند میانگین نمونه است.

فرضیه (**hypothesis**) :

فرضیه آماری نقطه آغاز آزمون فرض است. فرضیه آماری یک بیان مقداری در باره پارامترهای جامعه است و اصولا بدون داشتن فرضیه آماری امکان انجام یک آزمون دشوار است.فرضیه آماری به دو دسته فرض صفر (H0) و فرض خلاف (H1) بیان می‌شود . اغلب فرضیه بیانگر این مطلب است که یک ارتباط علیتی بین دو متغیر وجود دارد به شکلی که میزان یکی (متغیر مستقل یا Independent) تا حدودی تعیین کننده دیگری متغیر وابسته یا (Dependent) است.

متغیر (**Variable**) : ویژگی یا خاصیت یک فرد، شئ و یا موقعیت است که شامل یک سری از مقادیر با دسته بندیهای متناسب است. قد، وزن، گروه خونی و جنس نمونه هایی از متغیر هستند.

انواع متغیر:

متغیرهای ***کمی*** مانند قد، وزن یا سن درجه بندی می شوند و به همین دلیل قابل اندازه گیری می باشند.

متغیرهای ***کیفی*** مانند جنس، گروه خونی یا ملیت فقط دارای نوع هستندو قابل بیان با استفاده از واحد خاصی نیستند.

واریانس (**variance**) :

به میزان پراکندگی یک مجموعه از داده ها حول میانگین آن داده ها ، واریانس می گویند. بیان علمی تر از این واژه واریانس؛گستردگی توزیع احتمال و فاصله متغیر تصادفی از مقدار میانگین می باشد.

داده (**data**) : داده‌عبارتست از نمایش ذخیره شده اشیاء فیزیکی ، چیزهای مجرد، بوده‌ها (واقعیات) ،رویدادها یا موجودیتهای دیگر قابل مشاهده که در تصمیم سازی بکار می‌آید .

انواع داده :

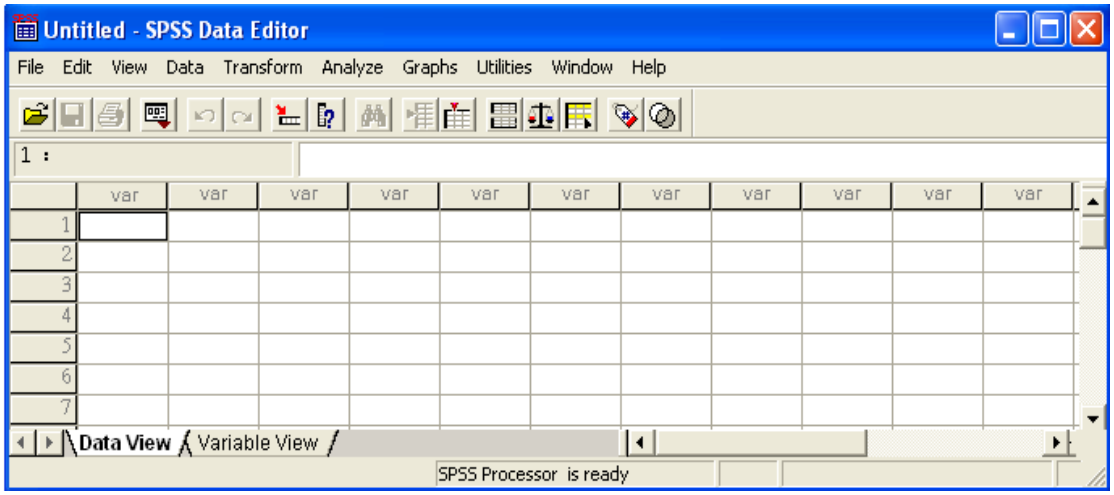
۱) داده های ***کمی*** (فاصله ای) Interval اعدادی هستند که بیانگر کمیت به صورت واحدهای عددی و بر اساس یک مقیاس مستقل است .قد و وزن مثالهای بارز دادههای کمی هستند.

۲) دادههای ***رتبه ای*** Ordinal مشتمل بر رتبه ها، تعلق داشتن به گروههای رتبه بندی شده یا اطلاعات ترتیبی است. به عنوان مثال اگر دو داور به یک مجموعه ۱۰ تایی از نقاشی رتبه یک (برای بهترین) تا رتبه ۱۰ (برای بدترین) بدهند. مجموعه داده ها مشتمل بر ۱۰ جفت رتبه خواهد بود. که هر جفت برای یک نقاشی است.

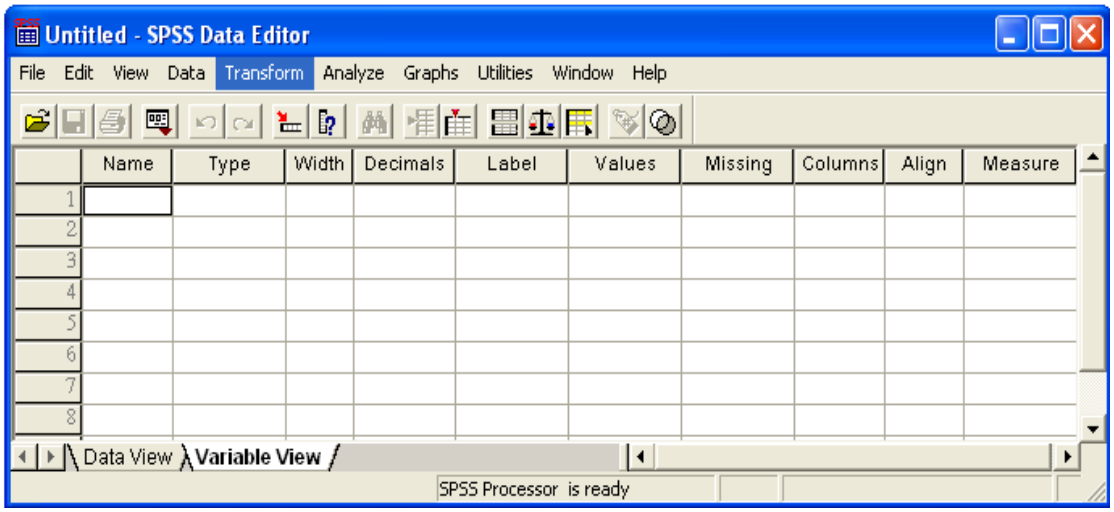
۳) دادههای ***اسمی*** (nomial) که مربوط به متغیر یا خواص کیفی مانند جنس یا گروه خونی است و بیانگر عضویت در یک گروه خاص می باشد.

وقتی برنامه ***Spss*** را باز می کنیم پنجره ای به نـام ***Untitled- SPSS Editor*** (ویرایشـگر ***SPSS*** – بـدون نـام) نشان داده می شود که شامل دو پنجره مختلف به نامهای زیر می باشد:

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



2- ***Variable view*** (نمای متغیر): برای تعریف کردن متغیرها استفاده می شود.



در حالت کلی، روش معمول این است که ابتدا بـا اسـتفاده از پنجـره ***Variable view*** متغیرهـا را تعریـف کـرده و سپس با استفاده از پنجره ***Data view***، اعداد را به SPSS وارد کنیم. توضیحات را با ذکر یک مثال ارائه می دهیم.

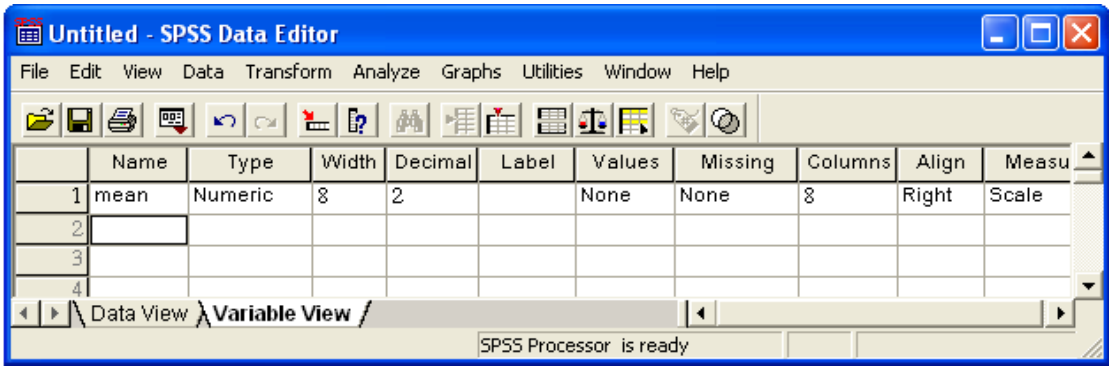
مثال: معدل دیپلم(کمی) و نـوع دیـپلم(کیفـی) ۱۵ نفـر از دانشـجویان سـال اول دانشـگاه در چنـد رشـته مختلـف دانشگاهی به صورت زیر می باشد. به منظور استفاده از این اطلاعات در تجزیه و تحلیلهـا، اطلاعـات (داده هـا) را بـه صورت زیر به نرم افزار وارد می کنیم.

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

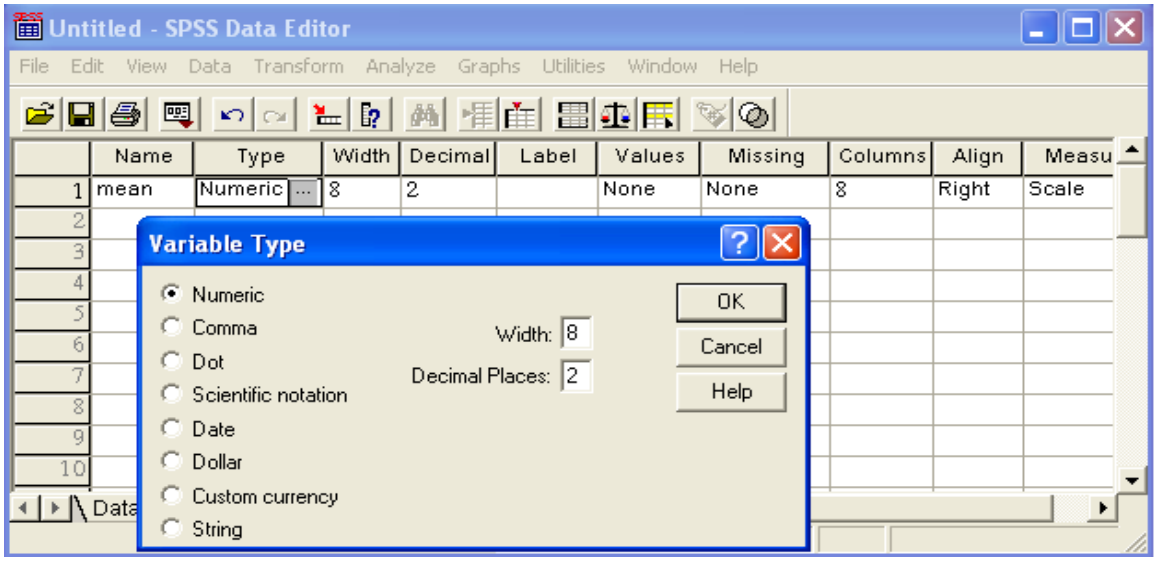
کد ۱ = دیپلم ریاضی کد ۲ = دیپلم تجربی کد ۳ = دیپلم انسانی

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

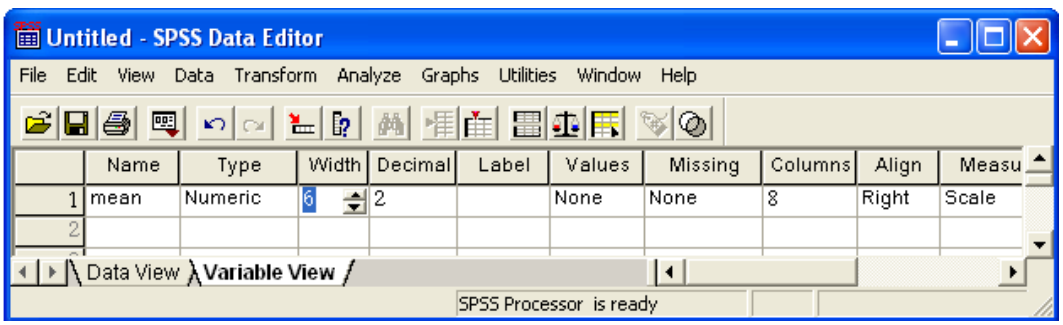
❖ در ستون ***Name*** ابتدا نامی برای متغیر انتخاب می کنیم (در اینجا داده های مربوط به معدل دانشجویان)،بعد از وارد کردن نام، سایر ستونها با پیش فرضهایی که نرم افزار طراحی کرده است به صورت زیر نمایش داده می شوند که ما می توانیم به دلخواه و با توجه به نوع داده، تغییراتی در ستونها ایجاد کنیم.



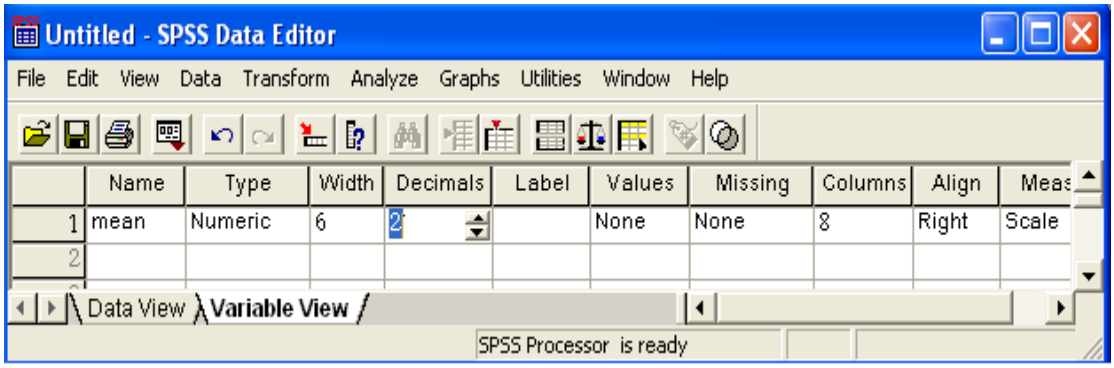
❖ در مرحله بعد، درستون ***Type*** نوع متغیر را تعیین می کنیم. در این ستون روی مربع کوچک خاکستری رنگ کلیک کرده تا پنجره ***Variable Type*** باز شود. و از بین گزینه ها نوع داده مناسب را برای متغیر مورد نظر تعیین می کنیم. با توجه به اینکه معدل دانشجویان عددی می باشد همان گزینه ***Numeric*** (داده عددی)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



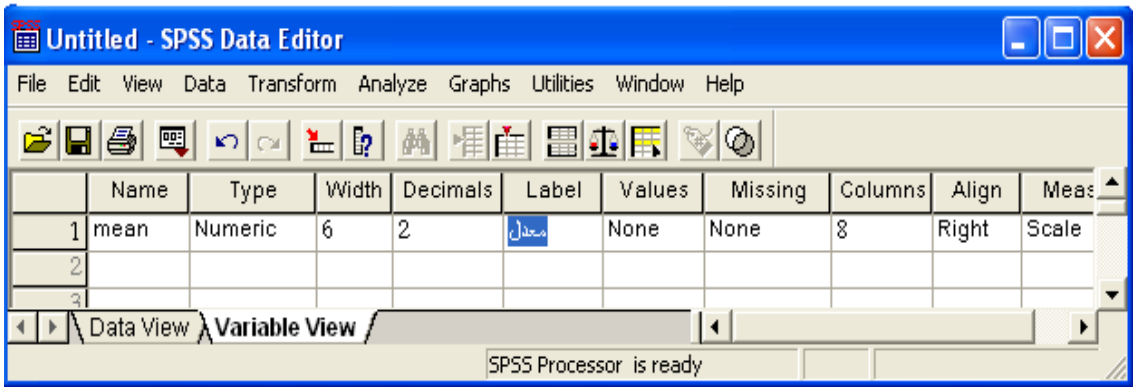
❖ از ستون بعدی (**Width**) برای تغییر دادن پهنای متغیر استفاده می شود. با کلیک روی ستون، دو پیکان کوچک بالا و پایین نشان داده می شود که می توانیم با بالا و پایین کردن، پهنای مورد نظر را تغییر دهیم.



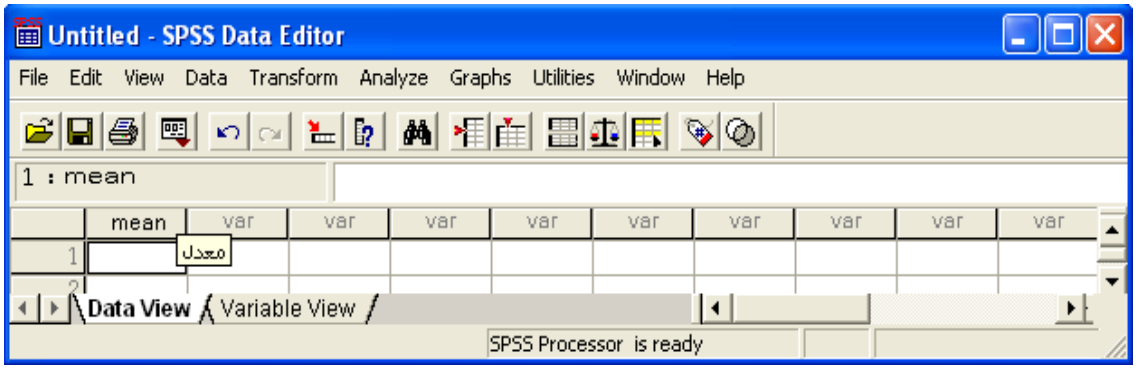
❖ از ستون **Decimals** برای تعیین تعداد ارقام اعشار داده های مورد نظر استفاده می شود. مانند روش قبل روی ستون کلیک کرده و تعداد ارقام اعشار را تعیین می کنیم. برای این داده ها با توجه به اینکه معدل یک فرد می تواند به طور مثال ۱۸،۲۵ باشد به همین دلیل ستون مورد نظر را به همان صورت پیش فرض نگه می داریم. به این معنی که در ستون معدل در پنجره (**Data View**) اعداد مربوط به معدل دانشجویان دو رقم اعشار نشان داده می شوند.



❖ 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.

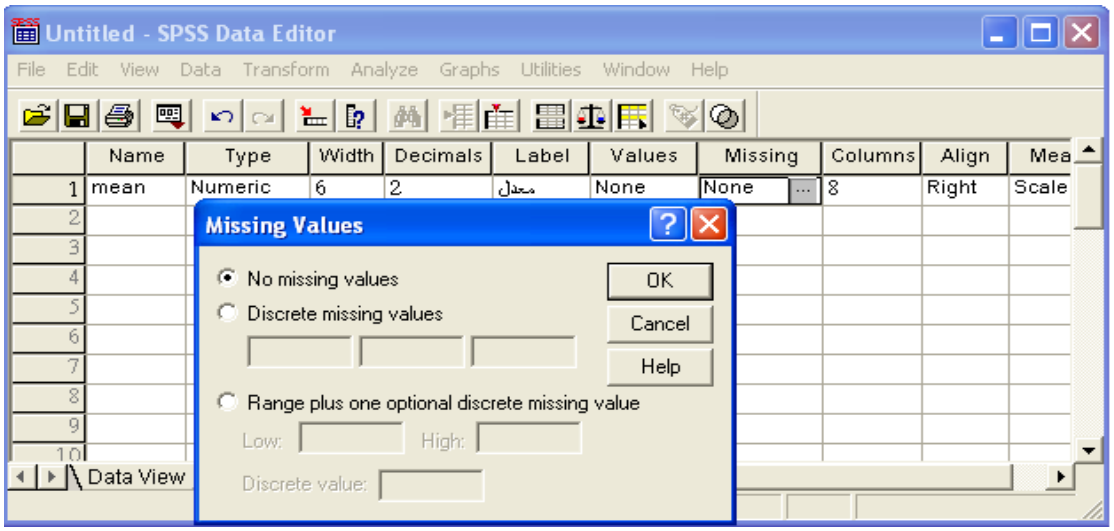


با این کار وقتی در پنجره Data view ، ماوس را روی عنوان Mean نگه داریم عنوان تایپ شده در Lable مشاهده می شود.



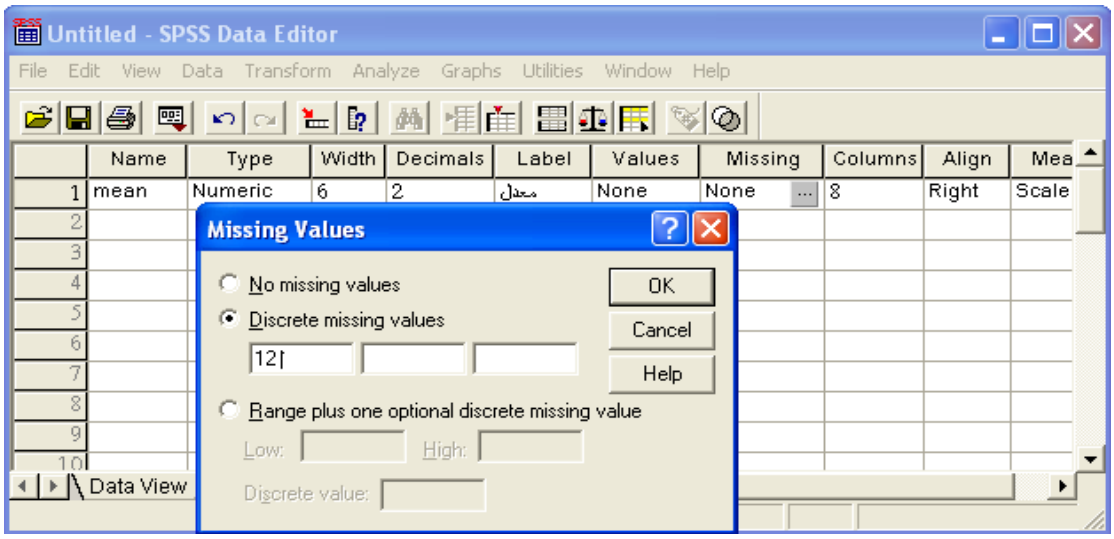
❖ ستون Values برای متغیرهای گروه بندی مورد استفاده قرار می گیرد. چون معدل دیپلم دانشجویان متغیر گروه بندی نمی باشد، اطلاعاتی در این قسمت اضافه نمی کنیم.

❖ ستون بعدی (***Missing***) مربوط به داده های گمشده می باشد. در این قسمت با کلیک بر روی ستون مورد نظر پنجره ***Missing Values*** باز می شود.



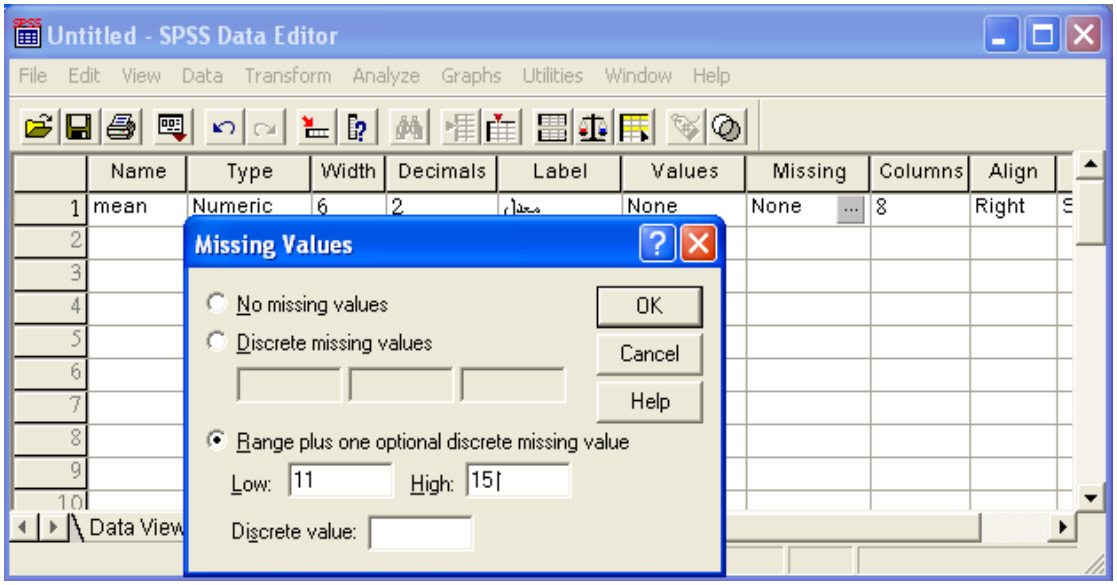
زمانی که در بین اطلاعات جمع آوری شده داده گمشده ای وجود نداشته باشد پیش فرض***No missing values*** به همان حالت خود باقی می ماند. اما اگر داده گمشده وجود داشته باشد، برای مشخص کردن آن باید گزینه ***Discrete missing values*** را فعال کرد و شماره سطر مربوط به داده مورد نظر را در مستطیلهای زیر وارد کرد.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

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.

❖ *مریم سیدطبیب، دانشگاه علوم پزشکی جندی شاپور اهواز*

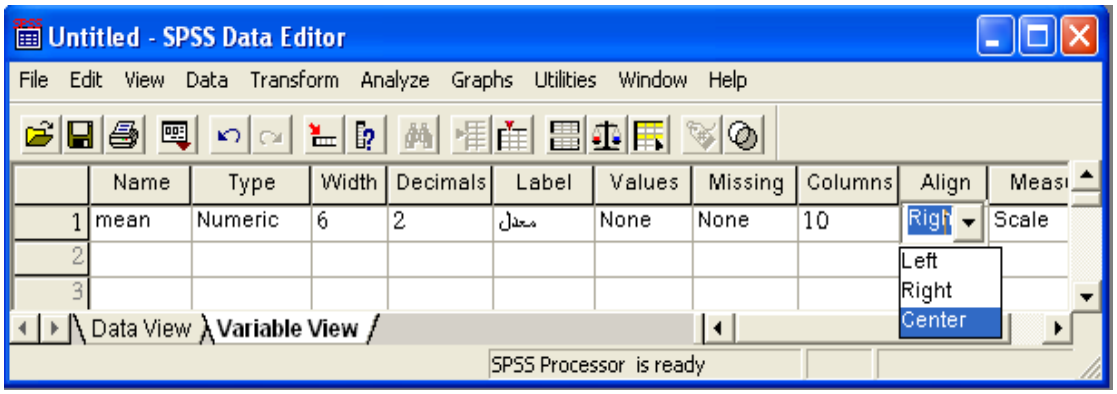


و اگر داده های گمشده به صورتی بودند که یک سری از آنها پشت سر هم و یکی از آنها جدا بود، شماره داده گمشده جدا را در قسمت ***Discrete value*** اضافه می کنیم.

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*** می باشد؛ که به مانند ستون **Decimals** می توان تغییراتی در آن ایجاد کرد.

- تراز کردن دادهها در ستون ***Align*** قابل انجام شدن می باشد.



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.

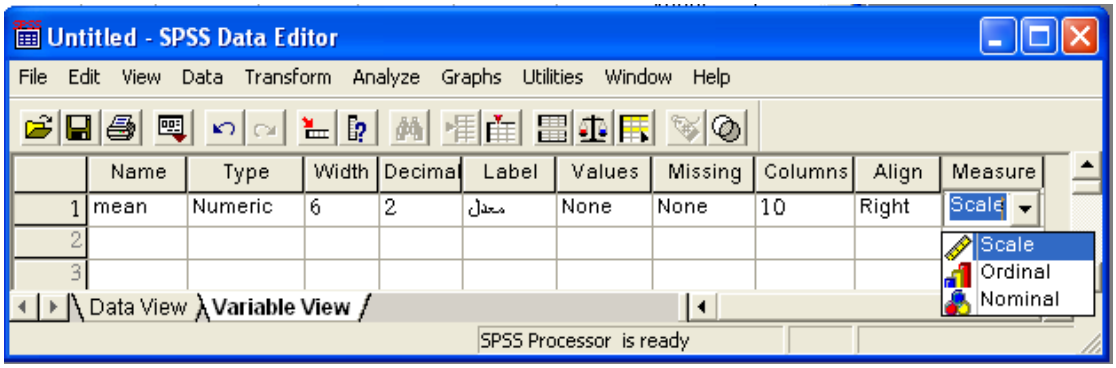
- آخرین ستون مورد بررسی در پنجره ***Variable view*** مربوط به ستون مقیاس اندازه گیری متغیرها (***Measure***) می باشد. که شامل سه نوع زیر است:

Scale = داده های فاصله ای و نسبتی

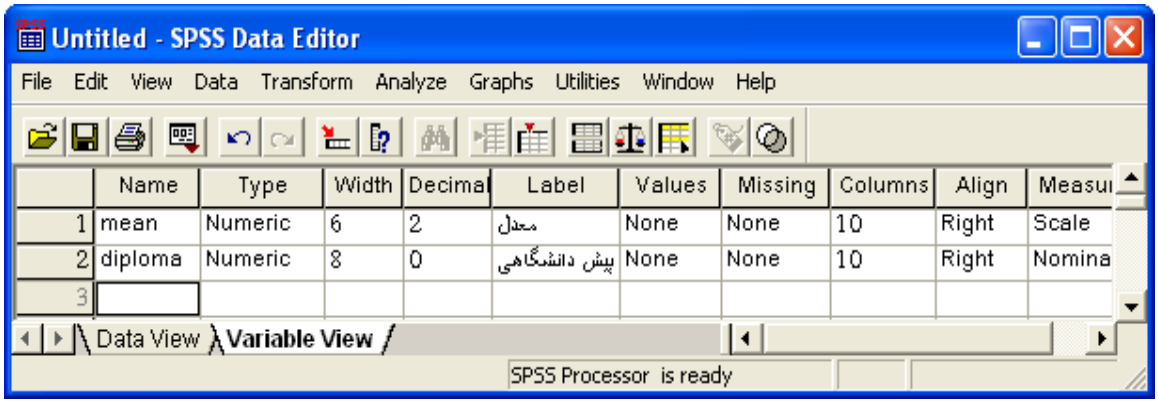
Ordinal = داده های رتبه ای

Nominal = داده های اسمی

داده های مربوط به معدل دانشجویان از نوع ***Scale*** می باشند.

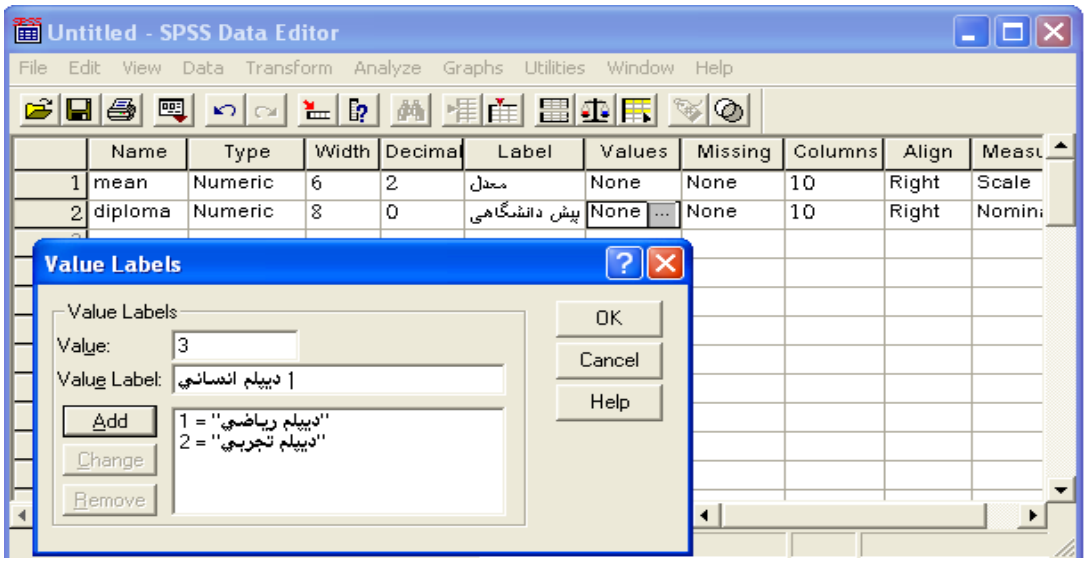


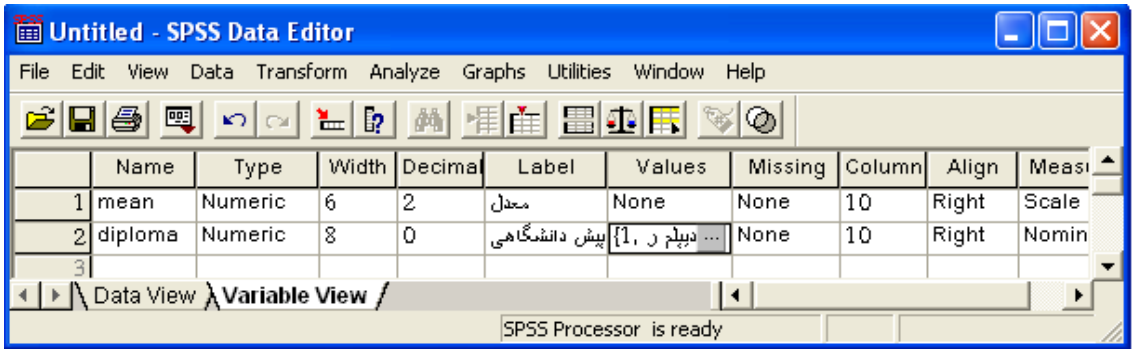
به همین ترتیب متغیر نوع دیپلم را در پنجره ***Variable view*** تعریف کرده و متناسب با نوع داده، ستونهای مختلف را تنظیم می کنیم.



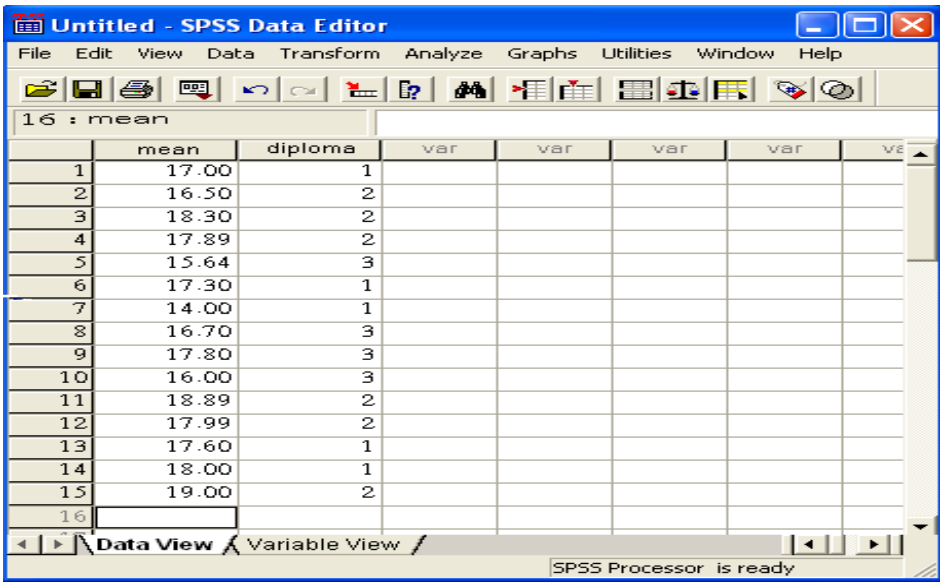
با توجه به اینکه نوع دیپلم داده اسمی می باشد، در این قسمت می توانیم در ستون ***Values*** هر کدام از کدهای دیپلم را با یک برچسب نشان داد.

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.



استفاده از آزمونهای آماری آزمایشاتی که دارای دو گروه مقایسه هستند را می توانیم بوسیله آزمونt (***T Test***) مورد تجزیه و تحلیل قرار دهیم(جزوه شماره۱). 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 می دهد.

☒ **روشهای تحلیل آماری داده را میتوان به دو گروه تقسیم کرد:**

روشهای اکتشافی و روشهای تائیدی.

در *روشهای اکتشافی* ***(آمار توصیفی)***، از محاسبات ساده و نمودارهای آسان برای خلاصه کردن داده استفاده میشود تا آنچه در داده نهان است آشکار شود.

در *روشهای تائیدی* ***(آمار استنباطی)***، با بکار گیری مفاهیمی از نظریه احتمال سعی می کنیم برای سوال های معینی پاسخ هایی بیابیم. نظریه احتمال درتصمیم گیری ها نقش مهمی دارد، زیرا شیوه هایی برای اندازه گیری، بیان و تحلیل عدم حتمییتهای وابسته به وقایع آینده ارائه میکند .

در این جزوه سعی بر این داریم که بخش آمار توصیفی را به کمک نرم افزار SPSS مورد بررسی قرار دهیم .

☒ **داده ها اغلب به صورت انبوهی از اعداد ارائه می شوند و به خودی خود خام هستند. برای آنکه بتوان آنها را پخته کرد و حقایق را درباره آنها جویا شد باید:**

الف: آنها را در جدولهایی تنظیم کرد.(جدول فراوانی)

ب: از روی جدولها نمودارهایی رسم نمود.(نمودار آماری)

ج: آنها را در یک یا چند عدد خلاصه کرد.(شاخص آماری)

جداول آماری

توصیف داده ها :

توزیع فراوانی تعدادی از داده ها عبارت است از:

جدول مرتب شده مقادیر آن دادهها که تکرار وقوع هر داده در آن مشخص شده است؛ به عبارت ساده تر توزیع فراوانی جدولی خلاصه شده از داده های جمع آوری شده جامعه آماری است.

برای بررسی و تجزیه و تحلیل داده ها، ابتدا آنها را طبقه بندی کرده و در یک جدول تنظیم می کنیم. این نوع جداول را جدولهای آماری گویند.

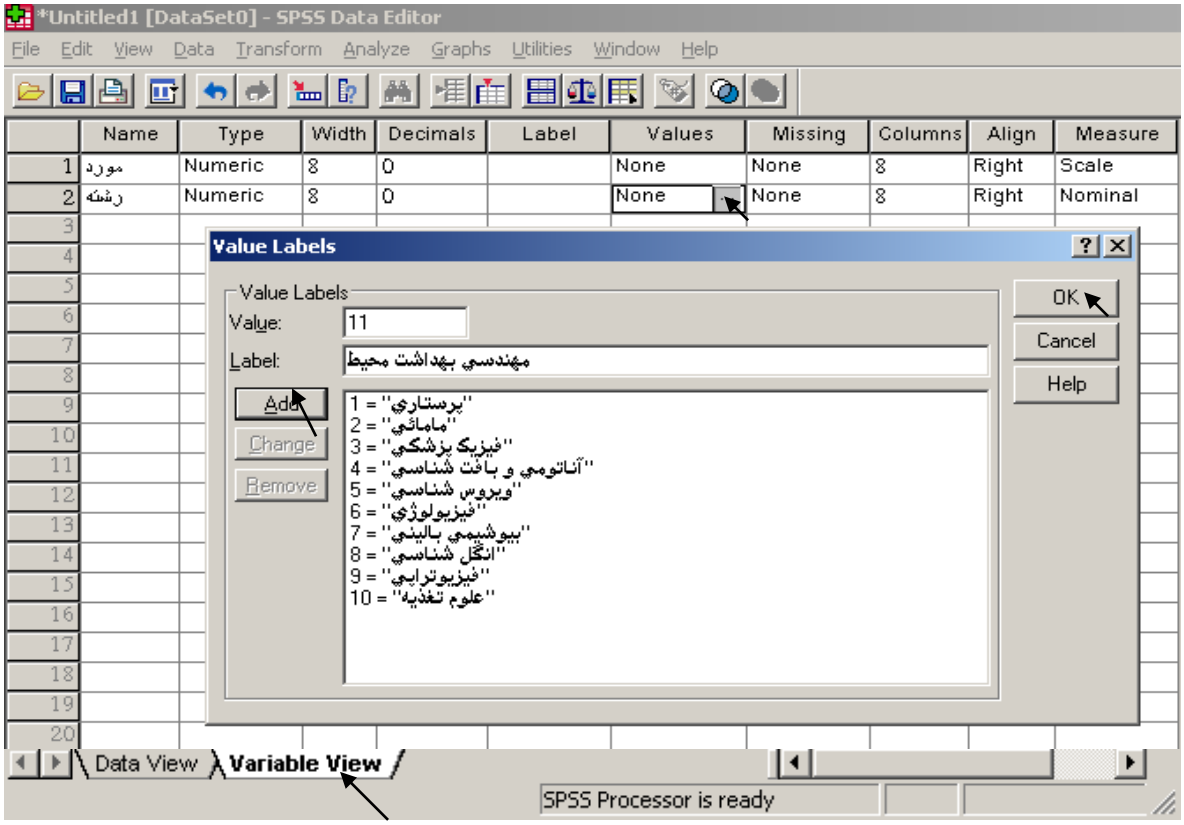
متداولترین جداول آماری، جدول‌های توزیع فراوانی (Frequencies) هستند.

مثال۱- داده های وارد شده در نرم افزار SPSS مربوط به رشته تحصیلی ۵۱ نفر از دانشجویان ورودی کارشناسی ارشد دانشگاه علوم پزشکی اهواز می باشند که هر رشته با یک کد بیان شده است. جدول فراوانی مربوط به تعداد دانشجویان و رشته های آنها را تشکیل دهید.

برای ورود اطلاعات در نرم افزار ابتدا در قسمت Variable View دو متغیر یکی به نام مورد و دیگری به نام رشته تحصیلی تعریف می کنیم و در سطر مربوط به متغیر رشته بر روی گزینه Values کلیک کرده تا پنجره مورد نظر باز شود و برای هر کدام از کدهای مربوط، رشته مورد نظر را تعریف می کنیم.

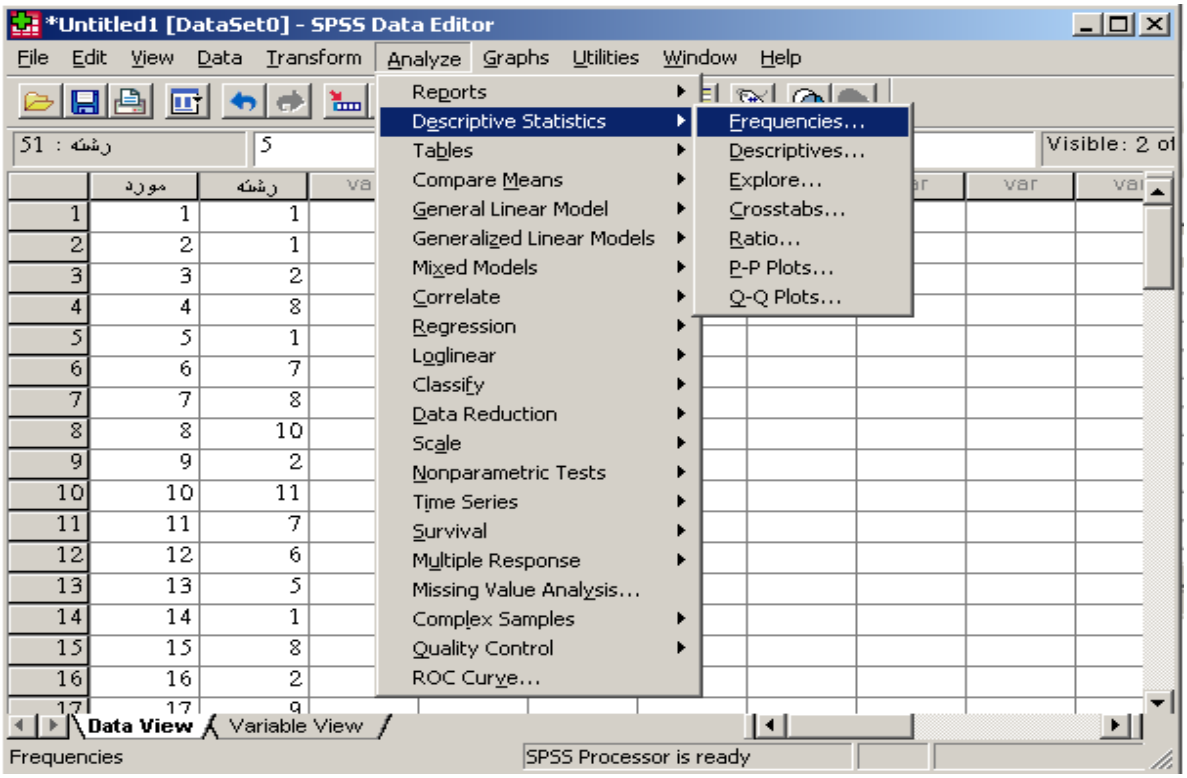
۱- پرستاری ۲- مامائی ۳- فیزیک پزشکی ۴- آناتومی و بافت شناسی ۵- ویروس شناسی ۶- فیزیولوژی
۷- بیوشیمی بالینی ۸- انگل شناسی ۹- فیزیوتراپی ۱۰- علوم تغذیه ۱۱- مهندسی بهداشت محیط

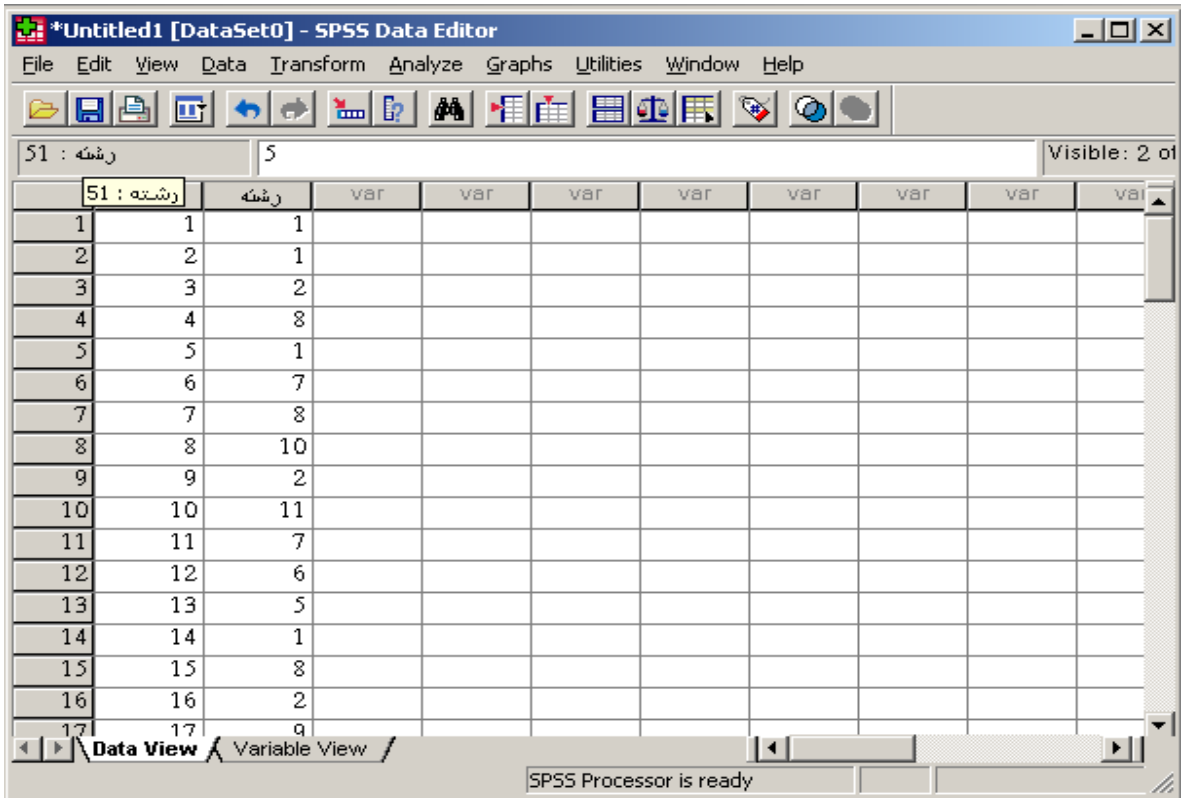
داده های مربوط به مثالهای مطرح شده در این جزوه در SPSS 15 وارد شده است.

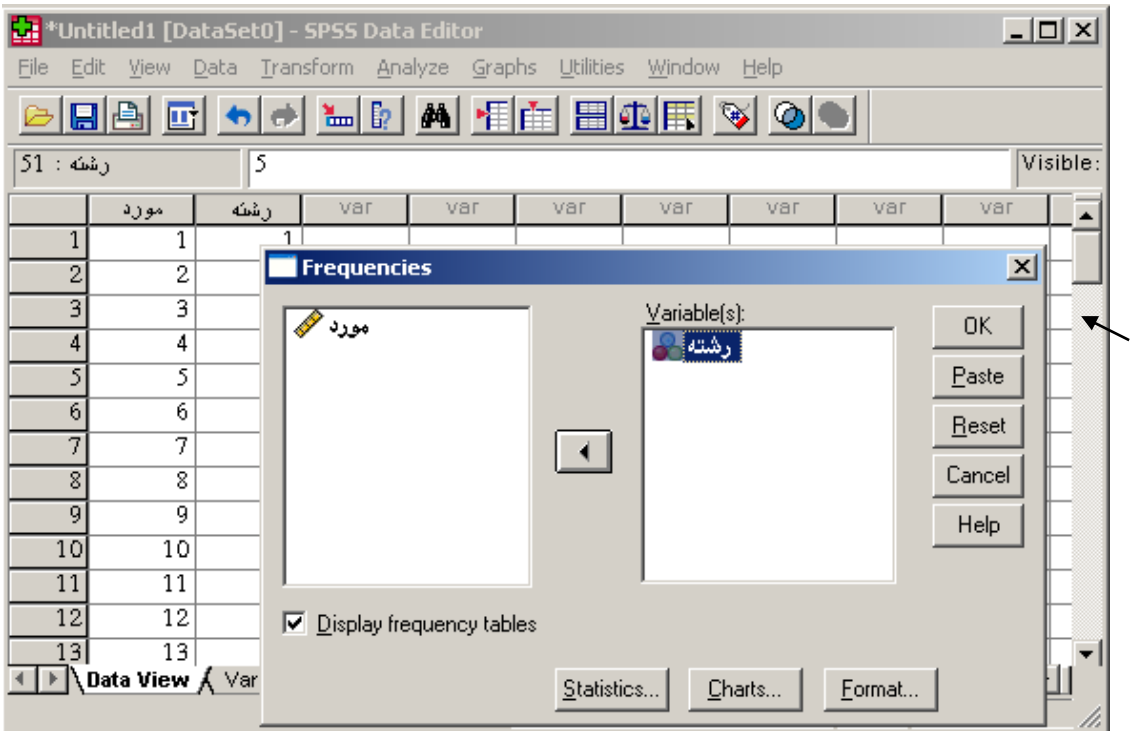


با کلیک کردن بر روی گزینه Add، رشته مهندسی بهداشت به جمع بقیه رشته ها افزوده می شود و در ادامه بر روی گزینه Ok کلیک کرده تا در خروجی نرم افزار به جای کدها، اسم رشته ها جایگزین شود.

سپس در قسمت DataView داده های خود را وارد می کنیم.


*Untitled1 [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
Data View
Variable View
Frequencies
SPSS Processor is ready


*Untitled1 [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Data View
Variable View
SPSS Processor is ready



بعد از انتقال متغیر رشته از پنجره سمت چپ به پنجره سمت راست و کلیک کردن بر روی کلمه Ok خروجی زیر

Statistics

رشته

N	Valid	51
	Missing	0

رشته

		فراوانی	درصد		فراوانی تجمعی
		Frequency	Percent	Valid Percent	Percent
Valid	پرستاري	8	15,7	15,7	15,7
	مامائي	8	15,7	15,7	31,4
	فيزيک پزشکي	4	7,8	7,8	39,2
	آناتومي و بافت شناسي	4	7,8	7,8	47,1
	ويروس شناسي	3	5,9	5,9	52,9
	فيزيولوژي	4	7,8	7,8	60,8
	بيوشيمي باليني	4	7,8	7,8	68,6
	انگل شناسي	4	7,8	7,8	76,5
	فيزيوتراپي	4	7,8	7,8	84,3
	علوم تغذيه	4	7,8	7,8	92,2
	مهندسي بهداشت محيط	4	7,8	7,8	100,0
	Total	51	100,0	100,0	

جمع کل

حاصل می شود.

اگر اسامی رشته ها برای کدهای ۱ تا ۱۱ تعریف نشود در خروجی به جای اسم رشته، کد آن آورده می شود. لازم به ذکر است که اعداد وارد شده در ستون رشته تحصیلی صرفاً به صورت کد می باشند و ارزش عددی ندارند. به طور کل وقتی داده های ما از نوع کیفی(اسمی یا رتبه ای) می باشند از این نوع جدول برای تفسیر داده استفاده می کنیم.

شاخصهای آماری

هنگامی که داده ها را به صورت خام و یا به صورت توزیع فراوانی نشان دادیم، نمی توان یک عقیده کلی درباره آنها ارائه نمود. لذا نیاز به خصوصیات و شاخصهایی است که نماینده کل داده ها باشند.

بنابراین ضروری است که هر توزیع را بوسیله چند شاخص محدود که آنها را پارامترهای جامعه می نامیم، معین کنیم.

به هر ویژگی جامعه پارامتر می گویند.

☒ **پارامتر کمیتی است که از اطلاعات مورد مطالعه در جامعه مورد بررسی بدست می آید و با در دست داشتن آن می توان خصوصیات جامعه را مشخص نمود.**

پارامترهای جامعه را به چهار گروه تقسیم می کنند:

۱- گروه اول پارامترهای وضعیتی هستند که به طور خلاصه حد متوسط یا میانگین نامیده می شوند، از این پارامترها برای پی بردن به مرکز توزیع استفاده می شود.

۲- گروه دوم پارامترهای مشخص کننده پراکندگی هستند که مشخص کننده نوع پراکنندگی و توزیع صفت در جامعه می باشند و آنها را شاخصهای معرف پراکنندگی می نامند و از آنها برای پی بردن به پراکندگی توزیع استفاده می شود.

۳- پارامترهای مشخص کننده چولگی که اندازه توزیع ها را در مقایسه با توزیع نرمال مشخص می نماید، این پارامترها نشان می دهند که مشاهدات جامعه به کدام سو تمایل دارند.

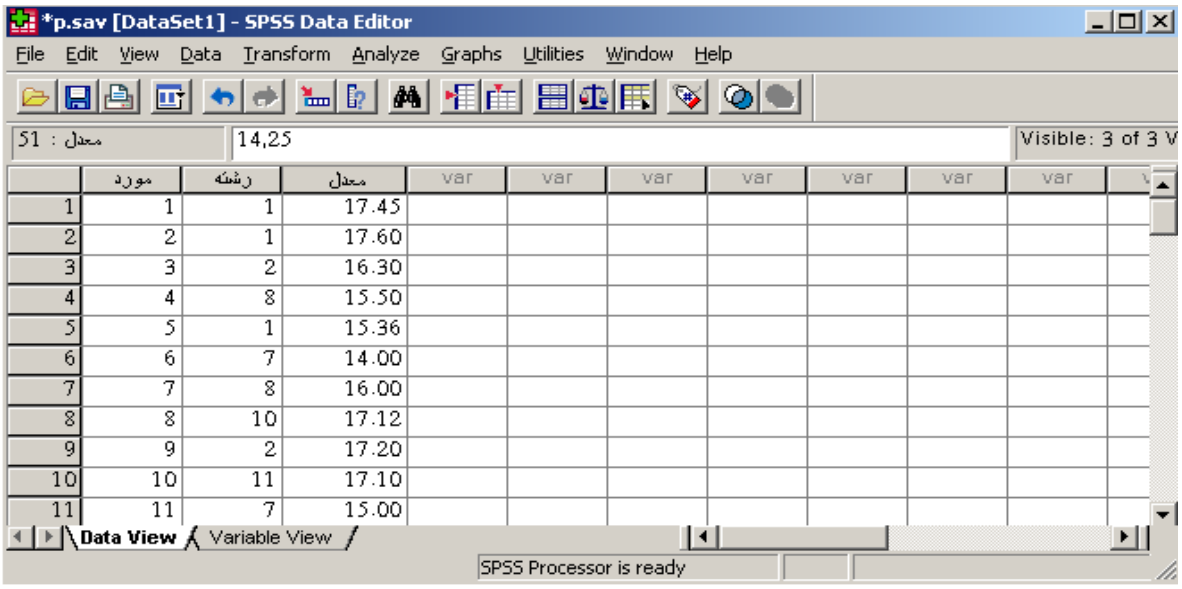
۴- گروه چهارم از پارامترهای مشخص کننده توزیع، ضریب کشیدگی است که بلندی توزیعها را در مقایسه با توزیع نرمال معین می کند.

مثال۲- به عنوان مثالی دیگر اگر در کنار رشته قبولی این ۵۱ دانشجو معدل دوره کارشناسی آنها نیز ذکر شده باشد. حال برای توصیف داده های خود مراحل زیر را انجام می دهیم. به مانند قبل در پنجره Variable View متغیر معدل را تعریف می کنیم. با توجه به اینکه معدل یک متغیر کمی و از نوع پیوسته می باشد در قسمت مقیاس از مقیاس Scale که خود به صورت پیش فرض تعریف شده است استفاده می کنیم.

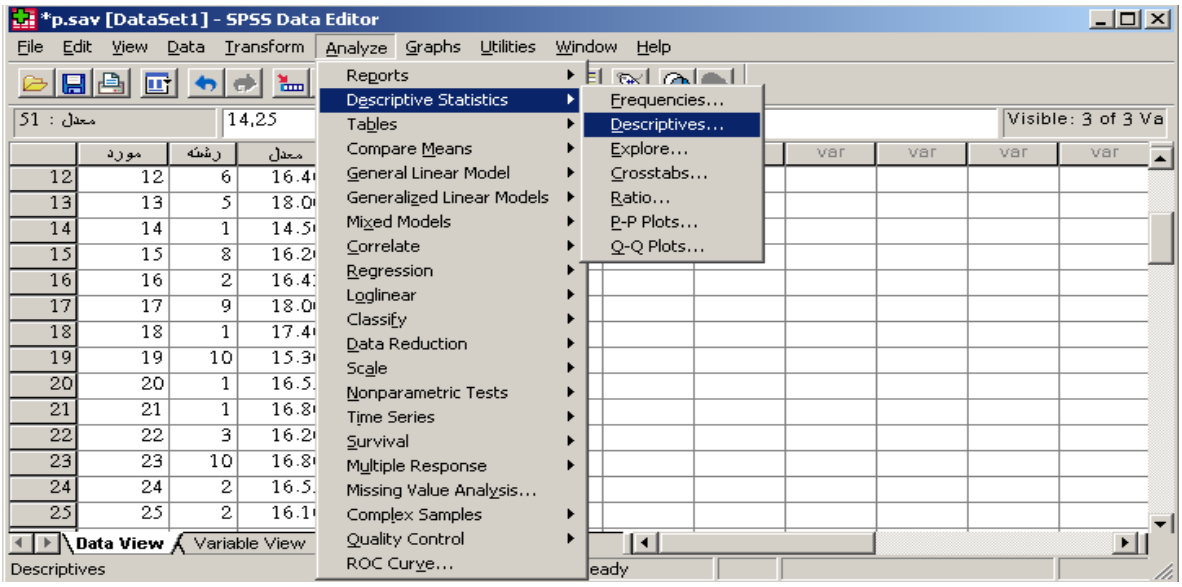
برای تفسیر داده های پیوسته از شاخص های دیگری استفاده می کنیم که مهمترین آنها شاخصهای مرکزی (میانگین، میانه و ....) و شاخصهای پراکندگی (واریانس، انحراف استاندارد و....) می باشند.

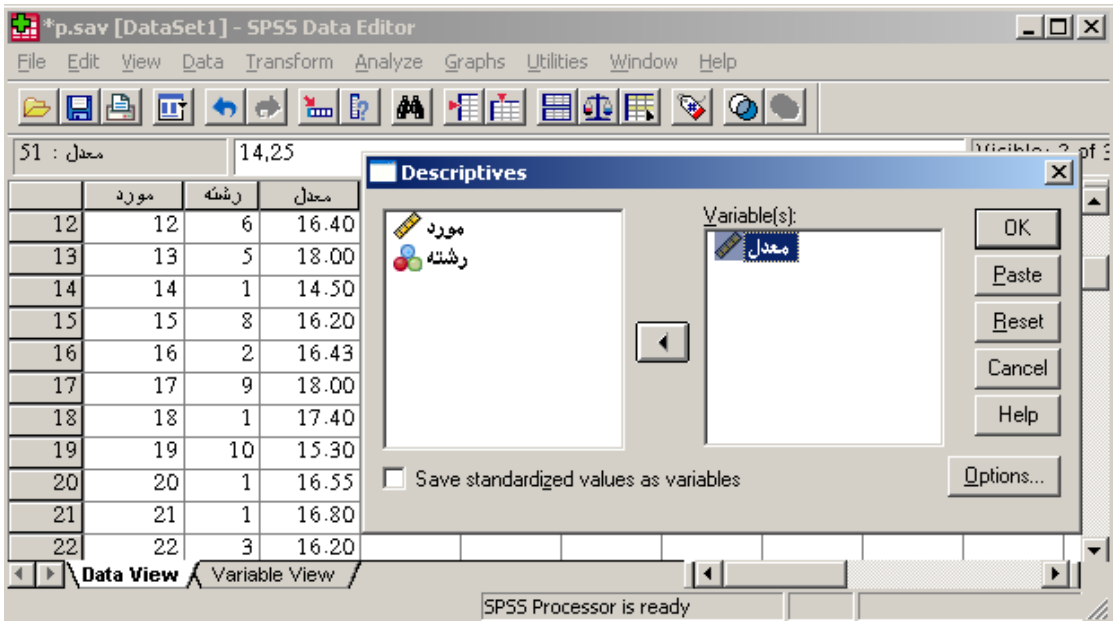
در این قسمت می خواهیم بدون در نظر گرفت رشته تحصیلی به طور کلی اطلاعاتی درباره متغیر معدل ۵۱ دانشجو بدست آوریم.

داده های استفاده شده در مثالها، صرفاً به عنوان نمونه می باشند و مستند نمی باشند.

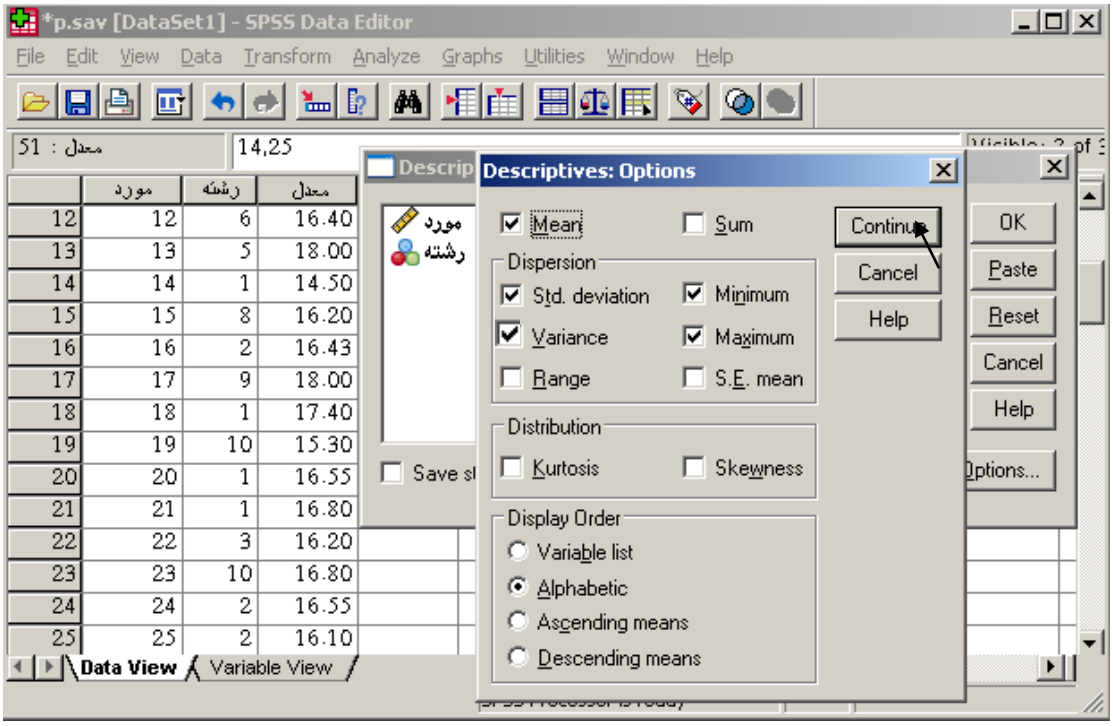


سپس برای توصیف داده های خود مراحل زیر را در نرم افزار Spss انجام می دهیم.





قبل از اینکه بر روی گزینه Ok کلیک کنیم، بر روی گزینه Option کلیک کرده و در پنجره مورد نظر اطلاعاتی را که در مورد داده ها مورد نیاز بوده، مشخص می کنیم.



در ادامه بر روی گزینه Continue و بعد از آن بر روی گزینه Okکلیک کرده تا خروجی مورد نظر ظاهر شود.

Descriptive Statistics

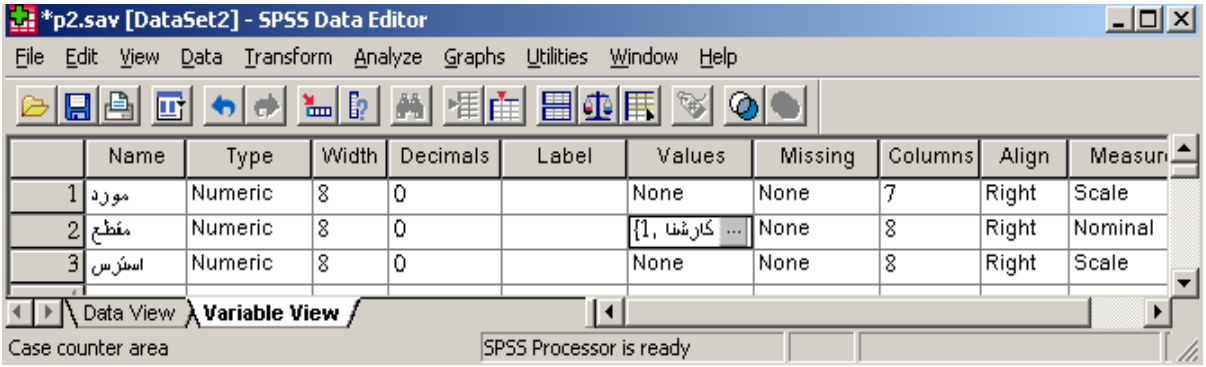
واریانس	انحراف استاندارد	میانگین

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
معدل	51	14.00	18.20	16.3622	1.08786	1,183
Valid N (listwise)	51					

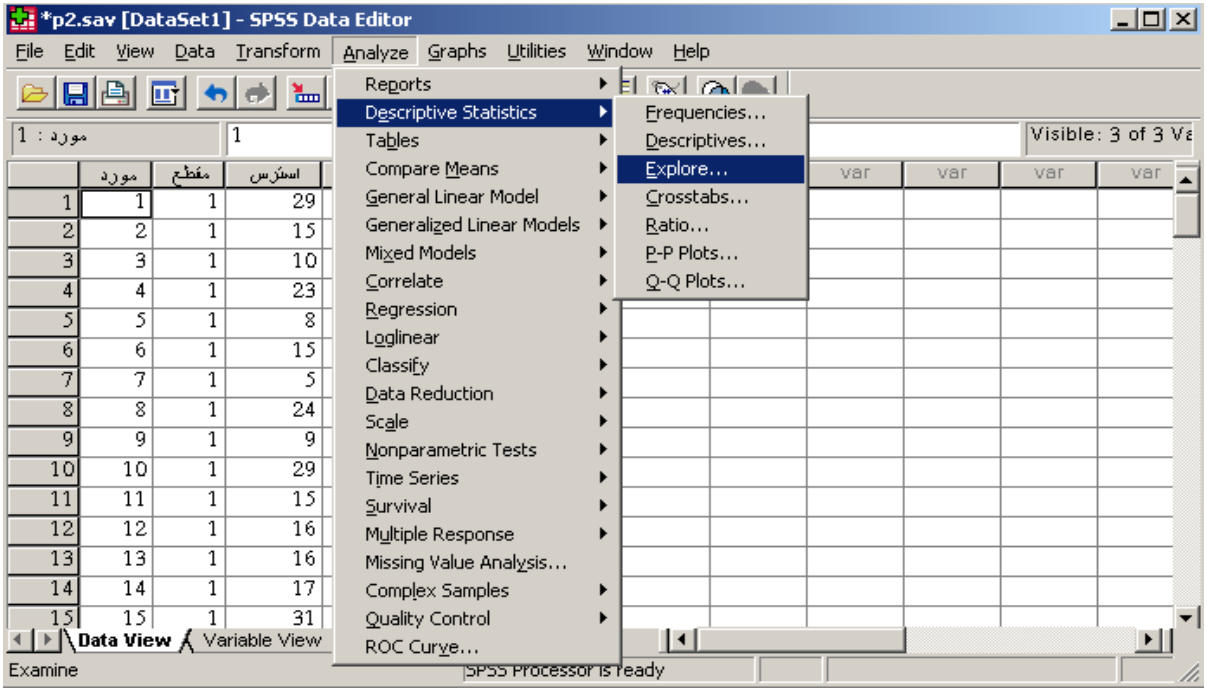
مثال۳- داده های پیش رو مربوط به میزان استرس (نمره استرس) ۹۰ نفر از دانشجویان در سه مقطع کارشناسی، کارشناسی ارشد و دکتری می باشند. می خواهیم آماره های توصیفی میزان استرس(عددی که برای میزان استرس هر فرد بیان می شود در اصل نمره ایست که فرد محقق از مجموعه ای پرسشها برای هر فرد بدست می آورد) را برای سه گروه به تفکیک بدست آوریم. با توجه به اینکه داده های از نوع کمی می باشند مانند قبل سعی می کنیم از برخی شاخصها، برای توصیف داده ها استفاده کنیم.

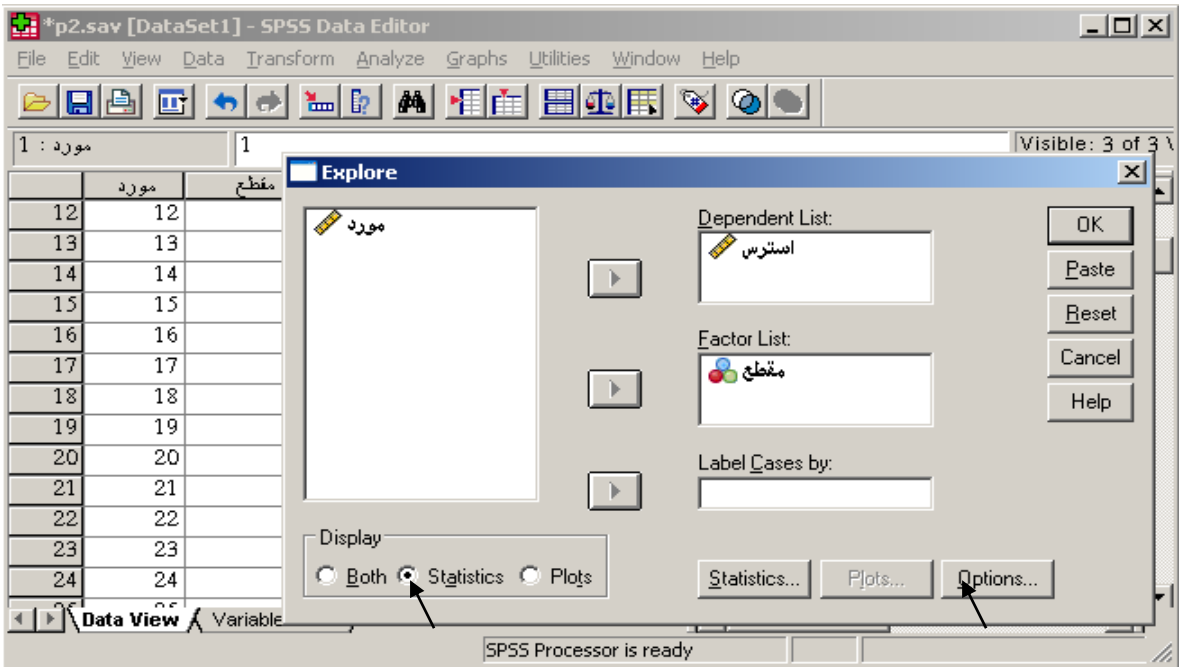
در مثال ۲ همانگونه که مشاهده شد، بدون در نظر گرفتن رشته قبولی میانگین را برای معدل تمامی دانشجویان محاسبه کردیم. در این قسمت می خواهیم میانگین و شاخصهای دیگر متغیر استرس را برای هر مقطع بطور جداگانه بدست آوریم.

برای این کار در پنجره Variable View سه متغیر (مورد، مقطع تحصیلی و نمره استرس) دانشجویان را تعریف می کنیم. در قسمت مربوط به Measure برای متغیر مقطع تحصیلی مقیاس Nominal و برای متغیر استرس از مقیاس Scale استفاده می کنیم. تعداد فراوانی دانشجویان هر مقطع می تواند یکسان یا با هم متفاوت باشند.
۱- کارشناسی (۴۰ نفر) ۲- کارشناسی ارشد (۳۰ نفر) ۳- دکتری (۲۰ نفر)



سپس داده های مورد نظر را در پنجره Data View وارد نموده و به صورت زیر عمل می کنیم.





بعد از اینکه دو متغیر استرس و مقطع را از پنجره سمت چپ به پنجره های سمت راست منتقل کردیم، در قسمت Display گزینه Statistics را فعال می کنیم و سپس بر روی گزینه Ok کلیک کرده تا خروجی مورد نظر بدست آید.

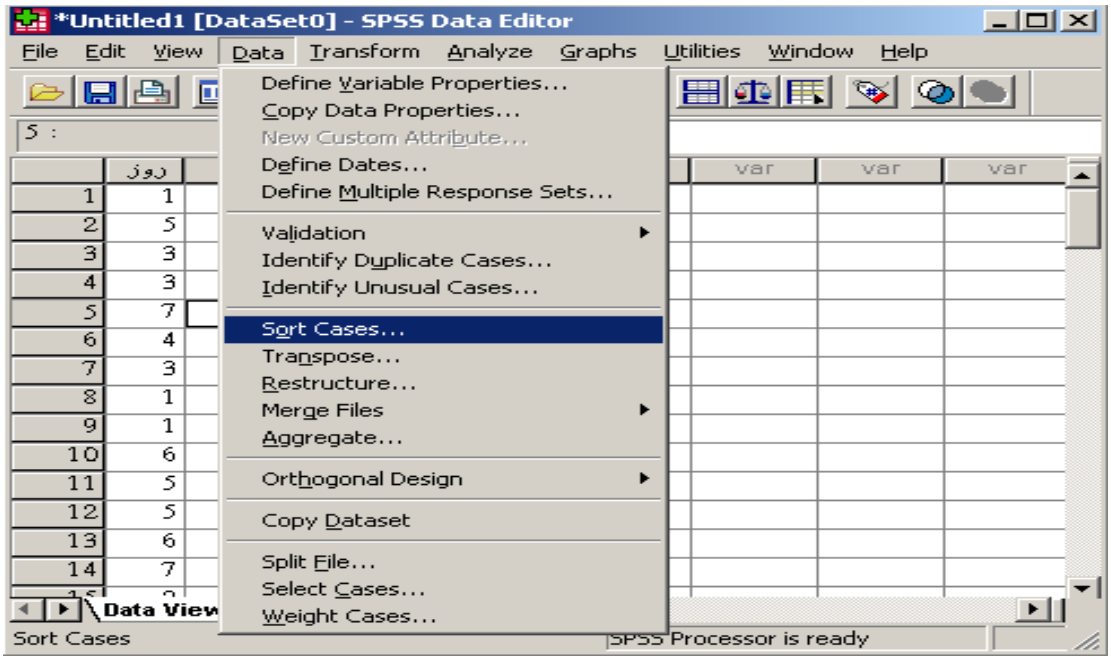
Descriptives

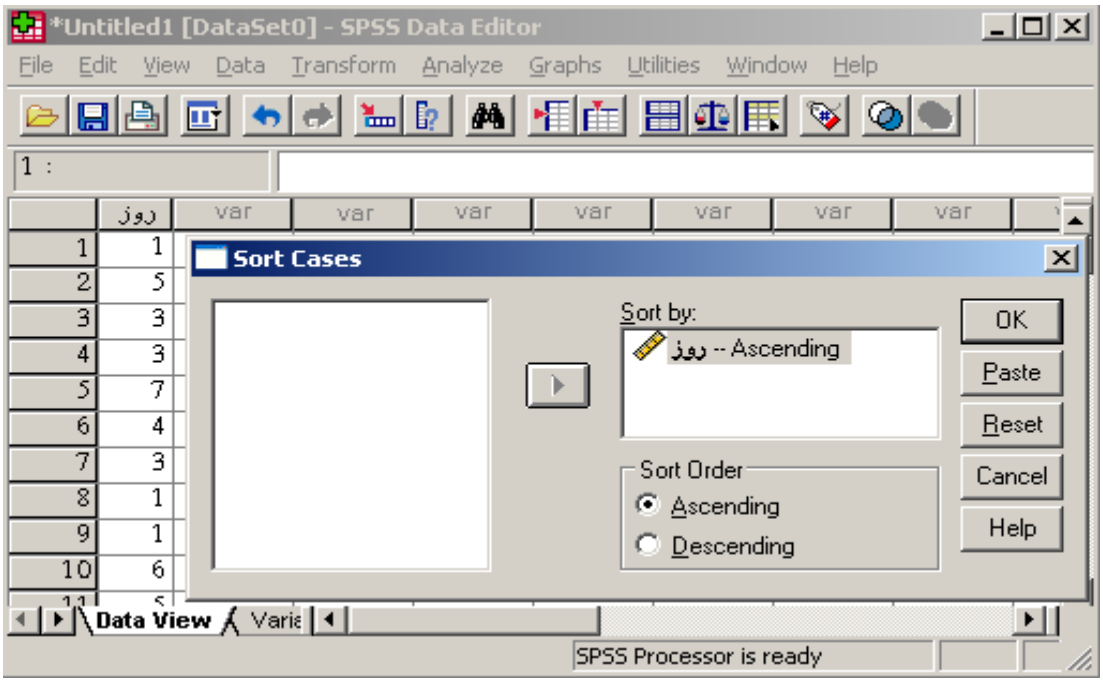
	مقطع		Statistic	Std. Error
استرس	كارشناسي	Mean	15.25	1.082
		Median	15.00	
		Variance	46.808	
		Std. Deviation	6.842	
		Minimum	3	
		Maximum	31	
		Range	28	
		Skewness	.594	.374
		Kurtosis	.173	.733
	كارشناسي ارشد	Mean	15.67	1.169
		Median	15.00	
		Variance	40.989	
		Std. Deviation	6.402	
		Minimum	4	
		Maximum	30	
		Range	26	
		Skewness	.436	.427
		Kurtosis	-.037	.833
	دكتري	Mean	15.90	1.294
		Median	15.50	
		Variance	33.463	
		Std. Deviation	5.785	
		Minimum	3	
		Maximum	26	
		Range	23	
		Skewness	-.411	.512
		Kurtosis	-.179	.992

❖ مرتب کردن داده ها از کوچک به بزرگ یا بر عکس (Sort Cases)

هنگامی که تصمیم داشته باشیم داده های وارد شده برای یک متغیر را در پنجره Data View به ترتیب صعودی یا نزولی مرتب کنیم باید مراحل زیر را انجام دهیم.

مثال۴- تعداد روزهای عدم حضور ۵۰ دانشجو در کلاس های درس در یک ترم تحصیلی به صورت زیر داده شده است. مانند روشهای قبل متغیر روز را در قسمت Variable View تعریف می کنیم و داده های مورد نظر را در پنجره Data View وارد می کنیم. سپس برای تنظیم داده ها طریق زیر را انجام می دهیم.





بعد از انتقال متغیر روز از پنجره سمت چپ به پنجره سمت راست قبل از اینکه بر روی گزینه Ok کلیک کنیم، می توانیم تصمیم گیری کنیم که می خواهیم داده ها را به صورت صعودی(Ascending) یا

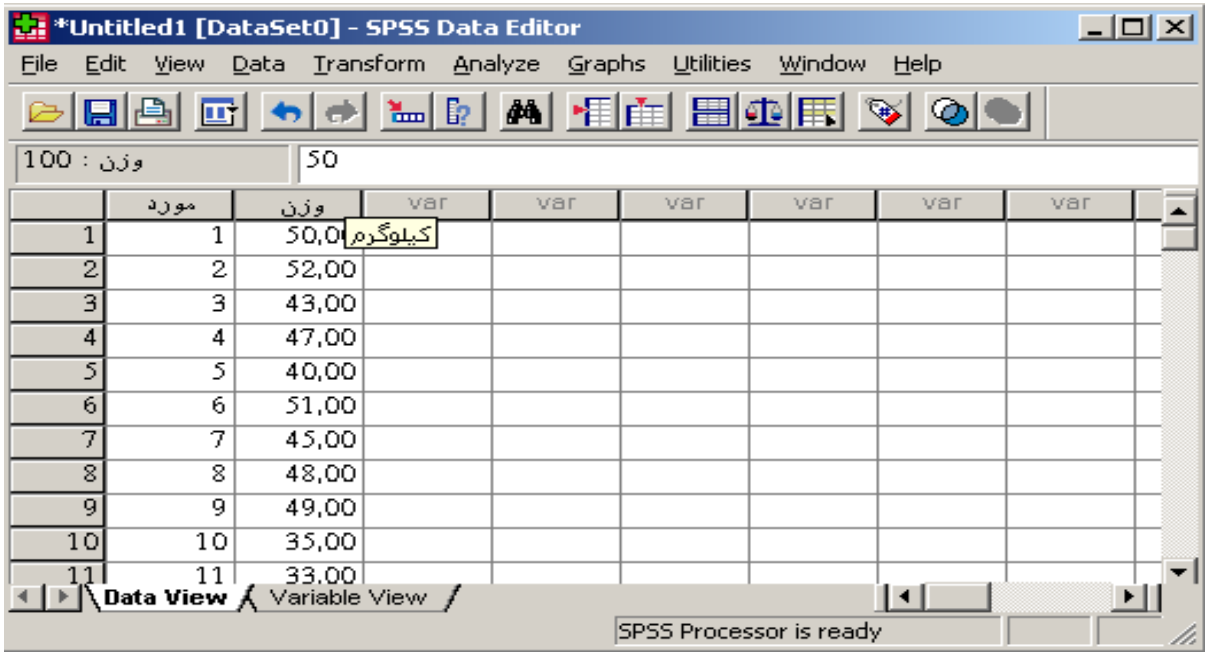
نزولی(Descending) مرتب کنیم. اگر چند متغیر داشته باشیم با Sort کردن یکی از متغیرها، سایر متغیرها نیز متناسب با آن تغییر می کنند.

❖ کدگذاری کردن داده ها (Recode into Different or Same Variables)

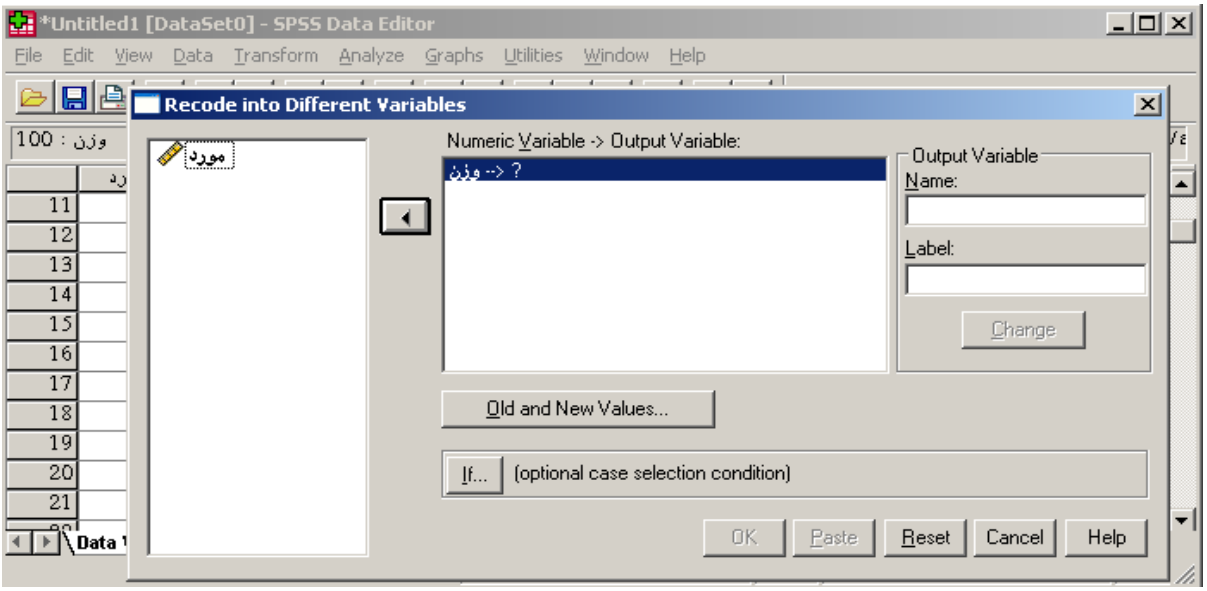
گاهی وقتها به دلیل حجم زیاد داده تصمیم به دسته بندی آنها می گیریم. همچنین ممکن است تصمیم داشته باشیم بعضی از داده های خود را با تعدادی داده دیگر عوض کنیم. برای این منظور از Recode کردن داده استفاده می شود.

مثال ۵- به عنوان نمونه ای برای حالت اول، فرض کنیم که داده های درون پنجره Data View مربوط به وزن ۱۰۰ دانش آموز سوم راهنمایی می باشد. می خواهیم این دانش آموزان را به سه دسته لاغر، متوسط و چاق دسته بندی کنیم تا به طور مثال بتوانیم نموداری یا جدولی برای فراوانیهای هر گروه داشته باشیم. برای این منظور محدوده وزن برای هر گروه را مشخص می کنیم و به صورت زیر عمل می کنیم.

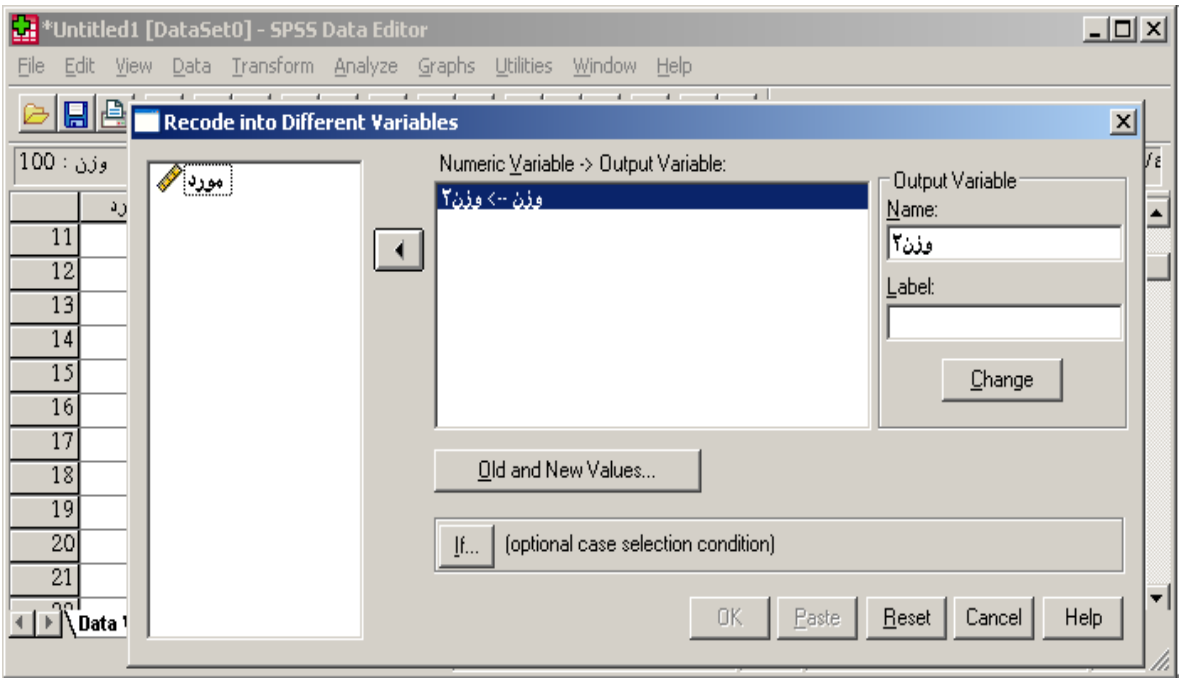
به مانند قبل متغیر وزن را در قسمت Variable View تعریف می کنیم و داده های مورد نظر را در پنجره Data View وارد می کنیم.



اگر به جای Recode into Different Variables از گزینـه بـالایی آن یعنـی Recode into Same Variables استفاده کنیم داده های جدید را در همان ستون متغیر اصلی جایگزین می کند و این کـار باعـث مـی شـود کـه داده های اصلی ما حذف شوند. برای اینکه داده های اصلی را نیز داشته باشیم گزینه پایینی را انتخاب مـی کنـیم تـا نـرم افزار داده های جدید را در ستون جدیدی که خودمان نامگذاری می کنیم قرار دهد.

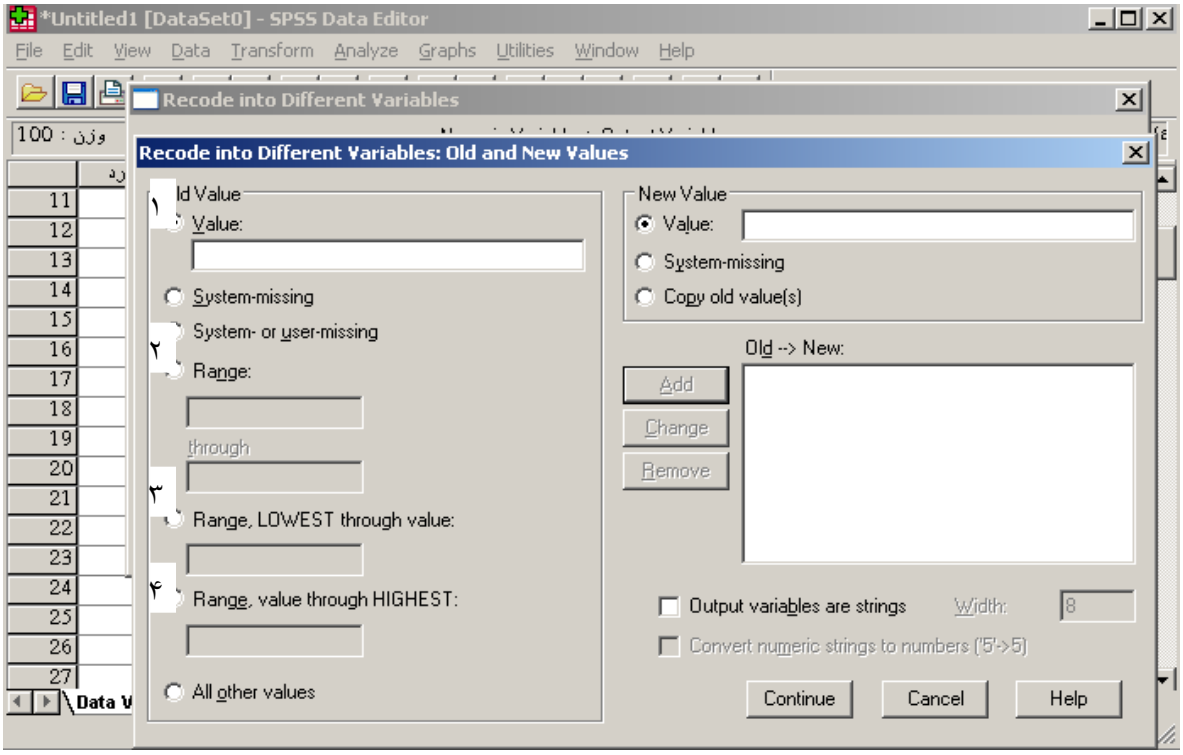


متغیر وزن را از پنجره سمت چپ به پنجره سمت راست منتقل می کنیم. روبروی کلمه وزن علامت سوال دیـده مـی شود که در اصل از ما سوال می کند که چه نامی را برای ستون جدیدی که داده های کدگـذاری شـده در آن قـرار می گیرد انتخاب می کنیم. نام دلخواه خـود را در گوشـه سـمت راسـت Output Variable در مسـتطیل Name قرار می دهیم. سپس گزینه Change فعال می شود و با کلیک بر روی گزینه Change نام وزن را به نـام دلخـواه خود برای ستون جدید تغییر می دهیم. در این مثال نام وزن را به وزن۲ تغییر می دهـیم کـه در پنجـره ایـن تغییـر



نمایان شده است.

در مرحله بعد برای انجام کد گذاری بر روی گزینه Old and New Values کلیک کرده تـا پنجـره مـورد نظـر بـاز شود.



پنجره مورد نظر شامل دو قسمت Old Value و New Value می باشد.

اگر بخواهیم داده خود را با یک عدد جدید جایگزین کنیم در سمت چپ (۱) در مسـتطیل Value عـدد مـورد نظـر را وارد کرده و در سمت راست در مستطیل Value عددی را که قرار است جایگزین شود وارد می کنیم. سـپس سـه گزینه Add، Change و Remove فعال می شوند و با انتخاب گزینه Add تغییر خود را در پنجره

Old ⟶ New

تثبیت می کنیم.

اگر بخواهیم برای فاصله ای از داده ها، عدد خاصـی را تعریـف کنـیم در قسـمت چـپ بـر روی گزینـه Range (۲) کلیک کرده سپس دامنه مورد نظر خود را به بیان دیگر مرز پایین و بالای فاصله خود را مشخص می کنـیم و مجـدد، سمت راست در مستطیل Value عددی را که قرار است جایگزین شود وارد می کنیم و Add می کنیم.

اگر فاصله اعدادی که می خواهیم تعریف کنیم یـک طرفـه باشـد بـرای مقـادیر کمتـر از (<) از گزینـه (۳) و بـرای مقادیر بیشتر از (>) از گزینه (۴) استفاده می کنیم.

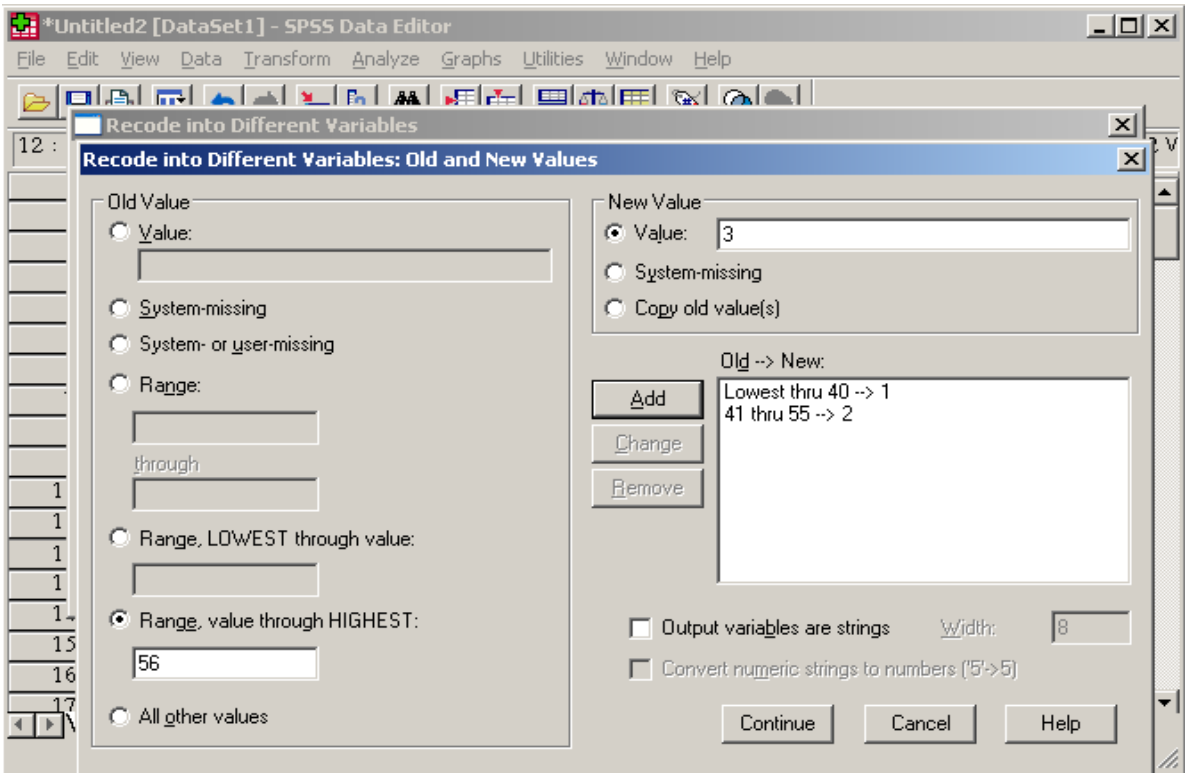
برای این مثال اگر فرض کنیم قد این دانش آموزان تقریباً در یـک سـطح باشـد بنـابراین وزن آنهـا را بـه صـورت زیـر دسته بندی می کنیم.

کمتر یا مساوی با۴۰ ⟵ ۱

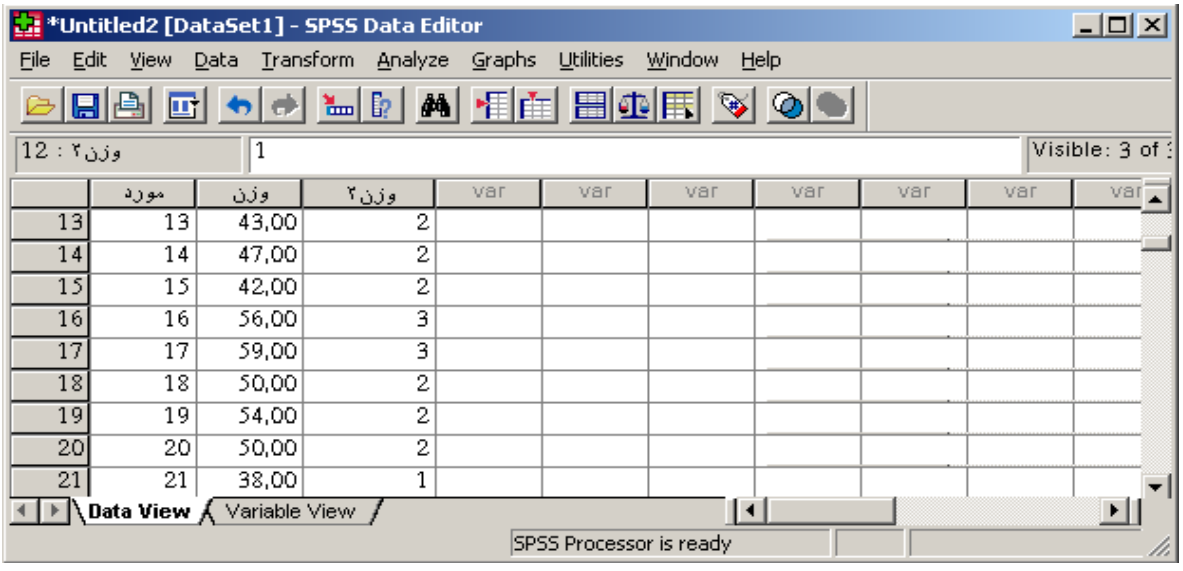
از ۴۱ تا ۵۵ ⟵ ۲

بیشتر یا مساوی با ۵۶ ⟵ ۳

هر زمان که در انجام تغییرات اشـتباهی صـورت گرفـت مـی تـوانیم بـا اسـتفاده از گزینـه Remove یـا Change تغییرات خود را اصلاح کنیم .



سپس بر روی گزینـه Add ، در ادامـه Continue و بعـد از آن Ok کلیـک کـرده و در پنجـره مربـوط بـه Data ستون مورد نظر را مشاهده می کنیم.



نمودارهای آماری

گام دیگری در جهت تلخیص داده ها، نمایش آنها به صورت نمودارهای آماری است.

- برای نمایش توزیعهای فراوانی، اغلب از نمودار استفاده می شود. نمودارها به ما کمک می کنند که تصویر توزیع به سهولت دیده شود.
- روشهای نمایش داده ها علاوه بر استفاده های نظری برای استفاده های علمی و کاربردی نیز مورد توجه هستند.
- نمودار های آماری در حقیقت ماهیت داده ها را در کمترین زمان ممکن و با ساده ترین بیان به ما نشان می دهند.

روش نمایش داده ها با توجه به نوع مقیاس داده ها انتخاب می شوند:

چنانچه مقیاس داده ها از نوع فاصله ای و نسبی(نسبتی) باشد، از نمودارهای کمی ***Numerical Chart*** استفاده می شود. و اگر داده ها از مقیاس اسمی یا رتبه ای پیروی کنند به کمک نمودارهای توصیفی **Chart Descriptive** قابل نمایش می باشند.

متداولترین نمودارهای آماری عبارتند از:

- **نمودارهای دایره ای**
- **نمودارهای میله ای**
- **نمودارهای مستطیلی (هیستوگرام)**
- **چند بر فراوانی**

در این جزوه به طور خلاصه دو نمودار میله ای و هیستوگرام را با ذکر مثال توضیح می دهیم.

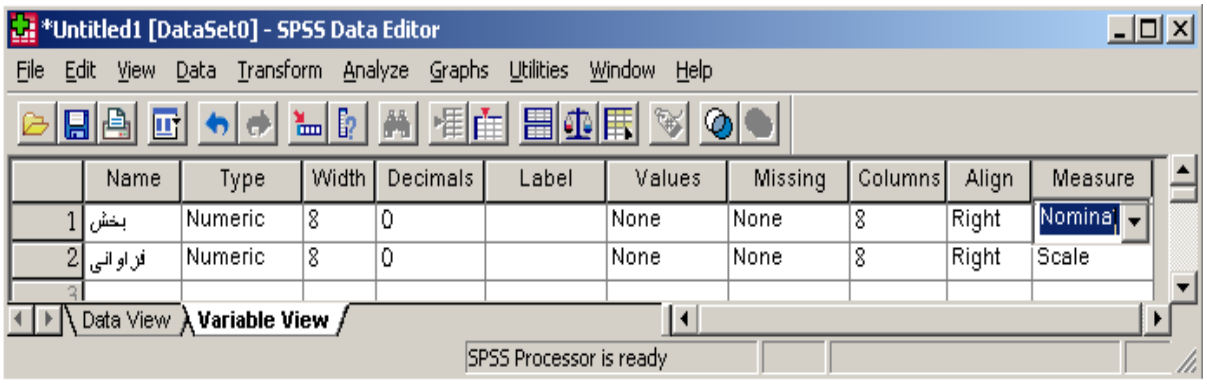
نمودار میله ای(Bar Chart)

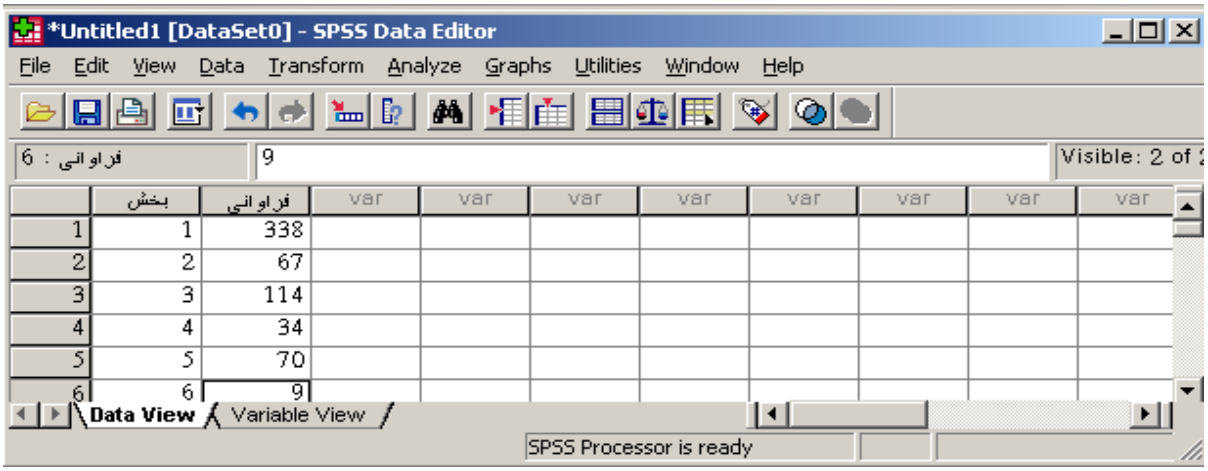
نمودار میله ای برای متغیرهای کیفی و کمی گسسته به کار می رود. در این نمودار بر محور افقی گونه های متغیر(چنانچه کیفی باشد) و یا مقادیر متغیر (چنانچه کمی باشد) درج می شوند. محور عمودی یا محور فراوانی نیز باید متناسب با بزرگی و کوچکی بیشترین فراوانی در طبقات، مقیاس بندی شود.

مثال ۶- داده های مورد نظر مربوط به تعداد بیماران بستری شده در ۶ بخش یکی از بیمارستانهای اهواز در مدت زمان یک ماه می باشد. برای داده های زیر نمودار مناسب را رسم کنید.

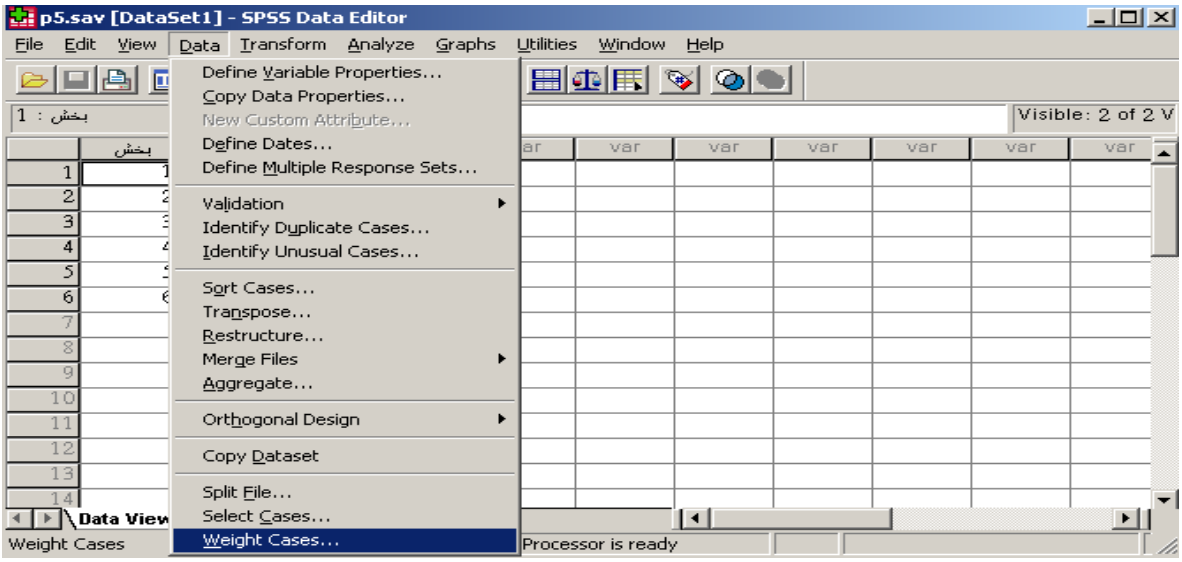
داخلی ۳۳۸ نفر، اطفال ۶۷ نفر، زنان و زایمان۱۱۴ نفر، روانپزشکی۳۴ نفر، CCU ۷۰ نفر، ICU ۹ نفر

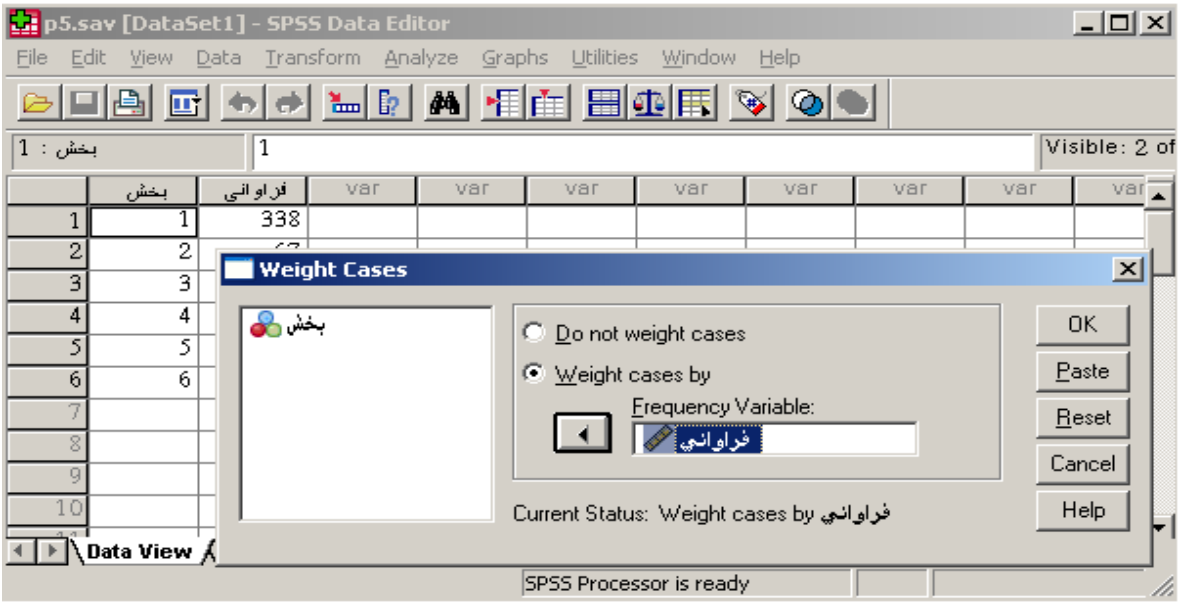
برای رسم نمودار در نرم افزار Spss در پنجره Variable View دو متغیر، یکی بخش و دیگری فراوانی را که نشاندهنده فراوانی هر یک از بخشهای بیمارستانی می باشد، تعریف می کنیم.



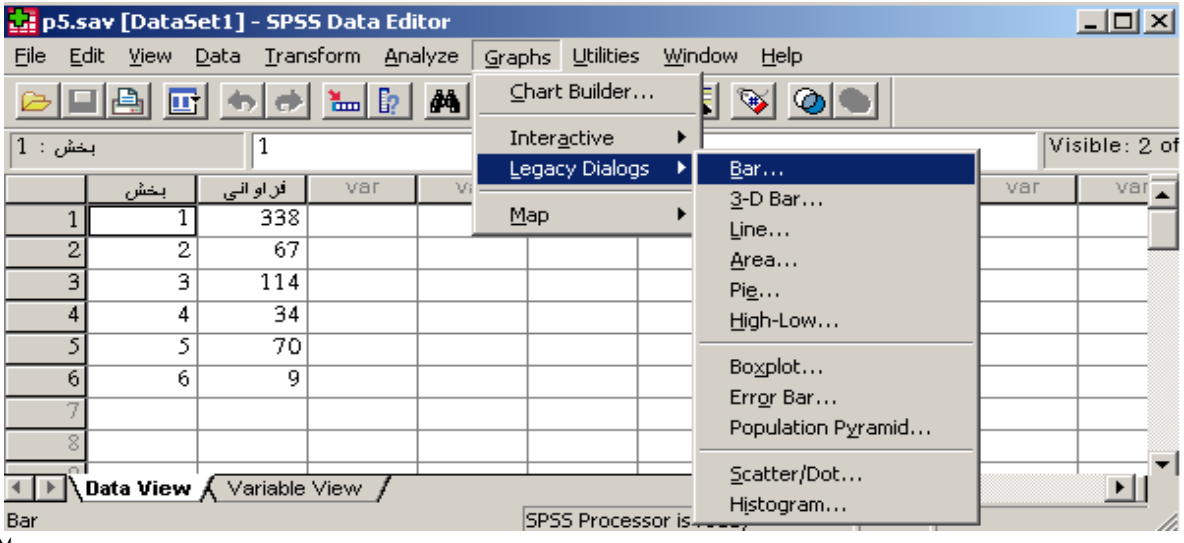


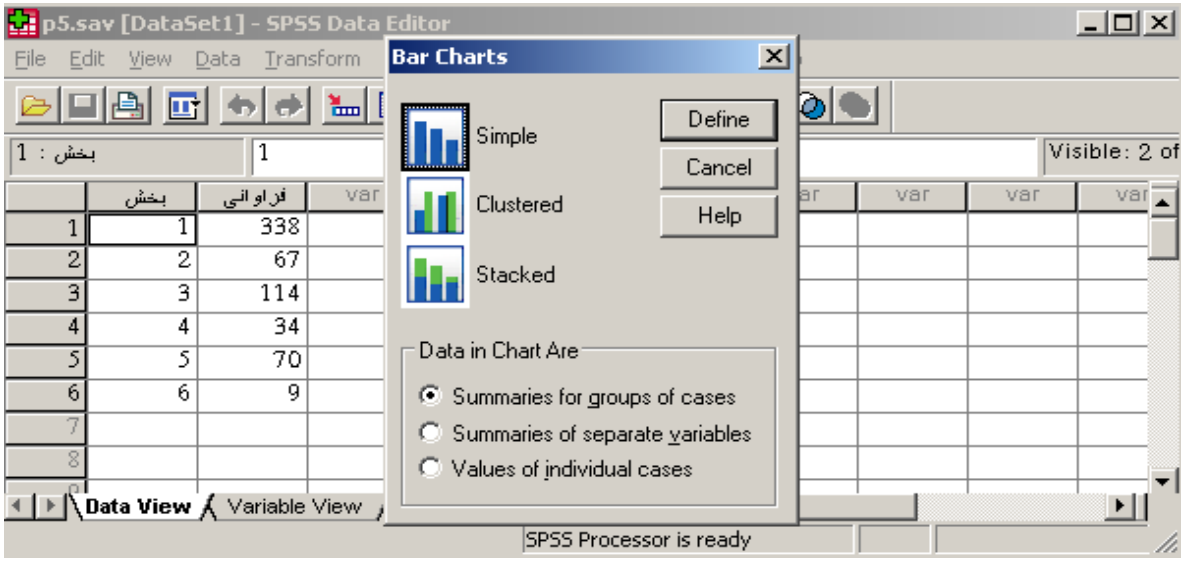
اعداد وارد شده درستون بخش به صورت کد می باشند و ارزش عددی ندارند. بنابراین می توانیم هر یک از این کد ها را با نام بخش مربوط به خود در قسمت Valuesدر پنجره Variable View جایگزین کنیم(مانند مثال ۱). همچنین اعداد مربوط به ستون فراوانی نیز به تنهایی ارزش عددی ندارند بلکه هر کدام از این اعداد مربوط به فراوانی بخش مورد نظر می باشند. بنابراین باید ستون بخش را بوسیله ستون فراوانی وزن دار کنیم. با این کار در خروجیها نشان می دهیم که فراوانی هر بخش به چه میزان می باشد.



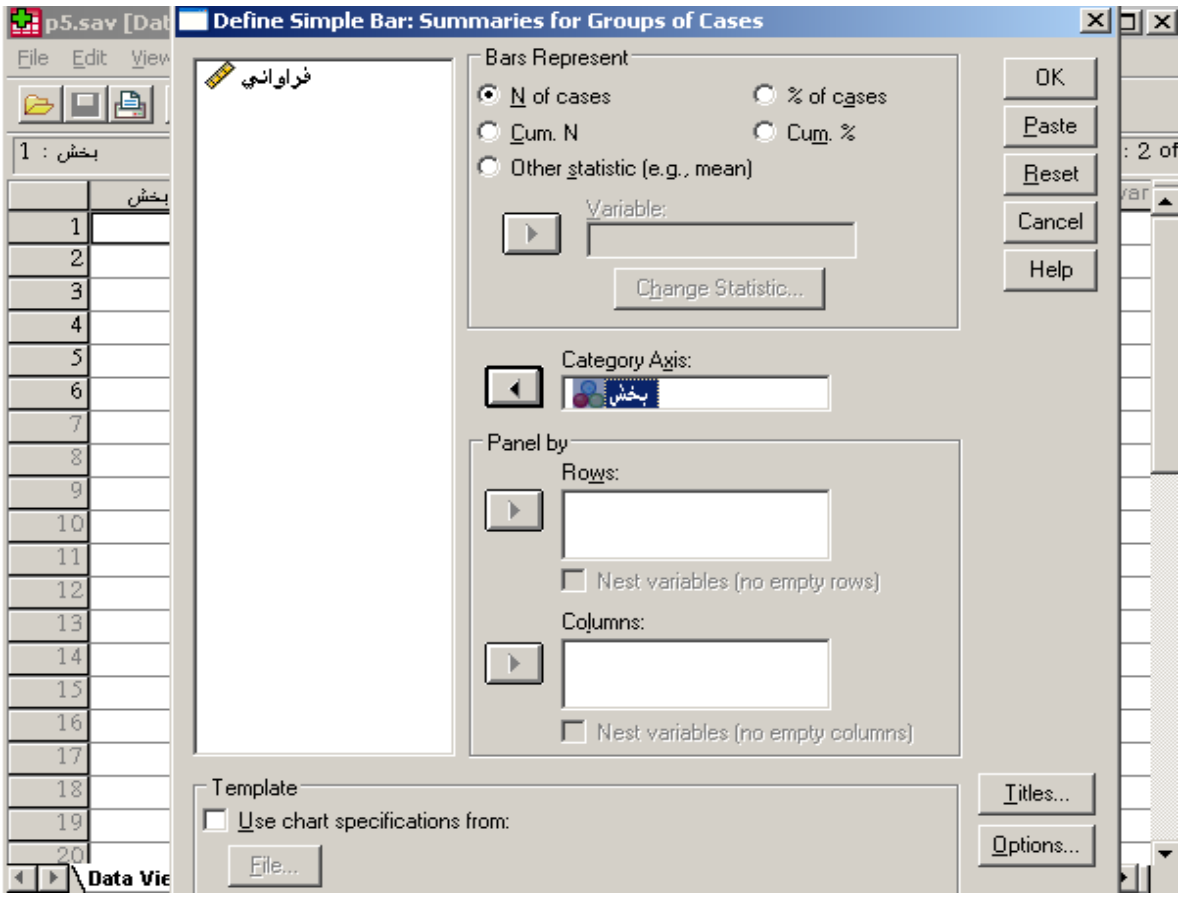


سپس گزینه Ok را فعال می کنیم و نرم افزار، ستون بخش را به وسیله ستون فراوانی وزن دار می کند. در ادامه برای رسم نمودار به صورت زیر عمل می کنیم.

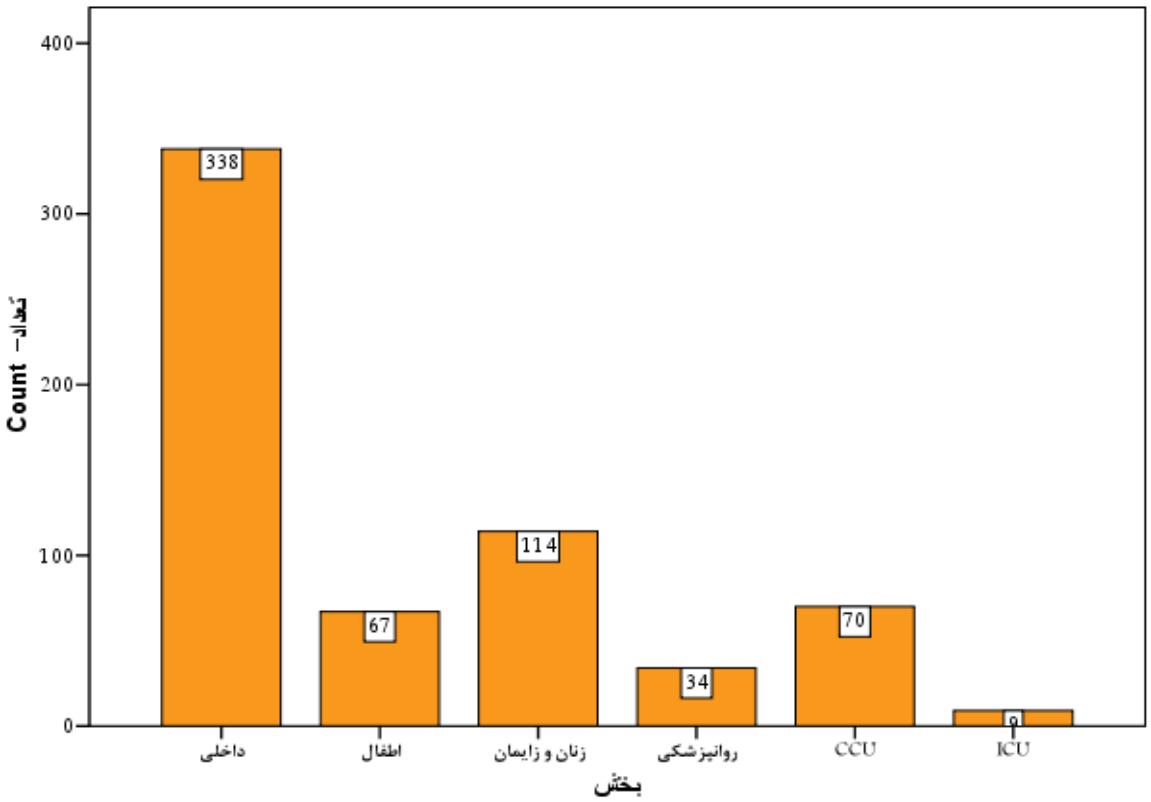


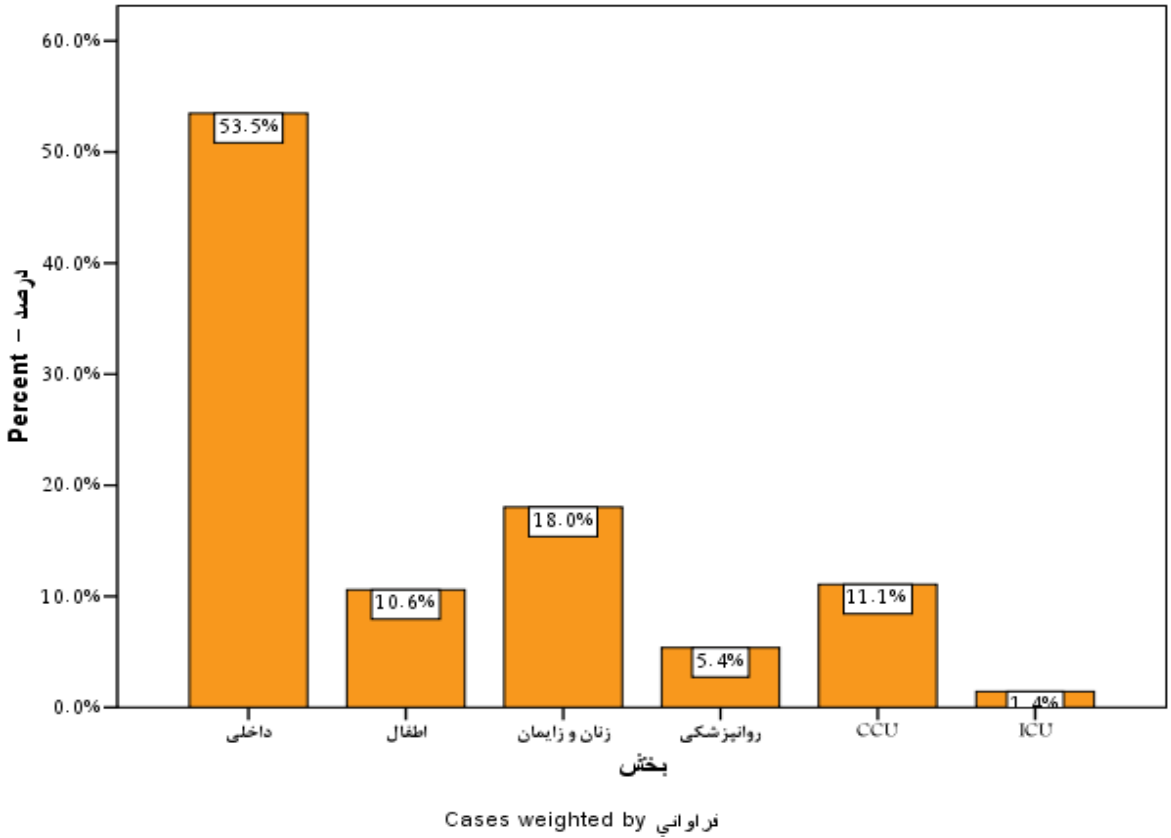


سپس در پنجره Bar Chart بر روی نمودار با فرمت Simple کلیک کرده و در ادامه بر روی گزینه Define کلیک می کنیم.



متغیر بخش را از پنجره سمت چپ به پنجره Category Axis منتقل کرده، در ادامه به طور مثال اگر بخواهیم داده ها بر حسب تعداد باشند درقسمت بالا (Bars Represent)، N of Cases را و یا اگر بر حسب فراوانی نسبی باشند % of Cases را انتخاب می کنیم. این مسیر را دو بار تکرار می کنیم تا هر دو نمودار رسم شوند.


400
300
200
100
0
Count - تعداد
338
67
114
34
70
9
داخلی
اطفال
زنان و زایمان
روانپزشکی
CCU
ICU
بخش
Cases weighted by فراواني


60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Percent - درصد
53.5%
10.6%
18.0%
5.4%
11.1%
1.4%
داخلی
اطفال
زنان و زایمان
روانپزشکی
CCU
ICU
بخش
Cases weighted by فراواني

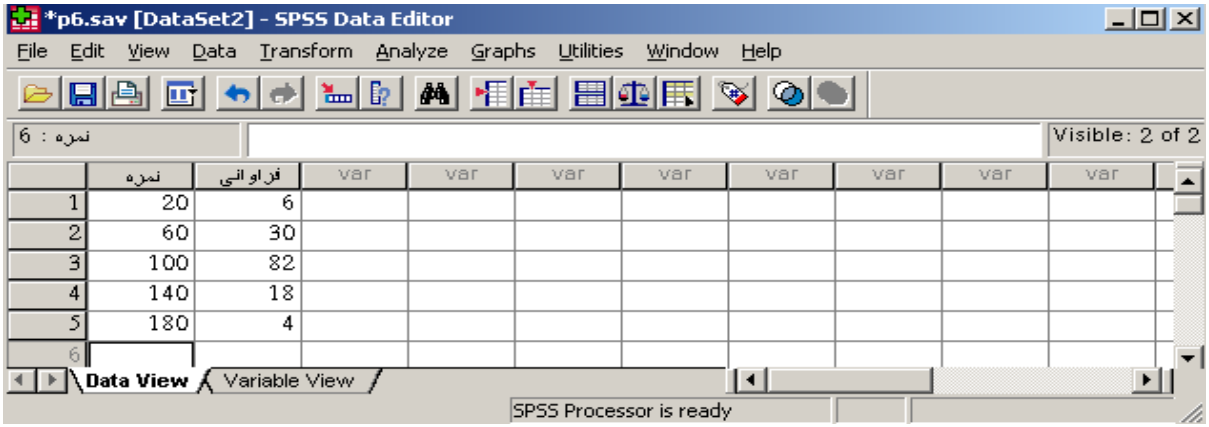
نمودار مستطیلی (Histogram)

برای نشان دادن شکل و مرکز یک توزیع و نحوه پراکندگی داده‌ها به کار می رود. معمولاً برای متغیرهای کمی پیوسته مناسب می باشد.

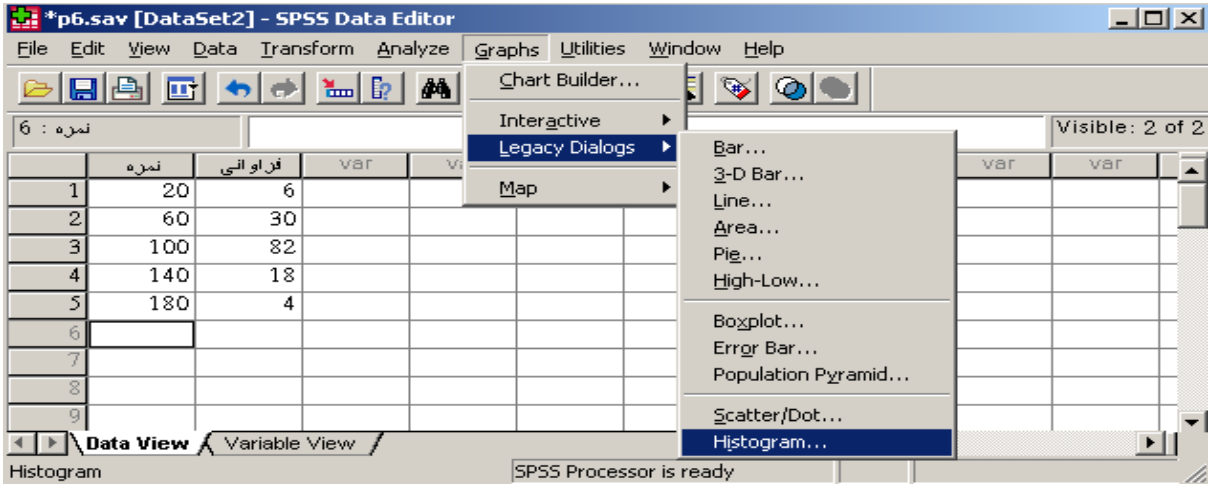
مثال ۷- داده های زیر مربوط به نمره ارزشیابی ۱۴۰ نفر از کارمندان دانشگاه علوم پزشکی جندی شاپور اهواز می باشند. با توجه به نوع داده ها نمودار مناسب را برای داده ها رسم کنید.

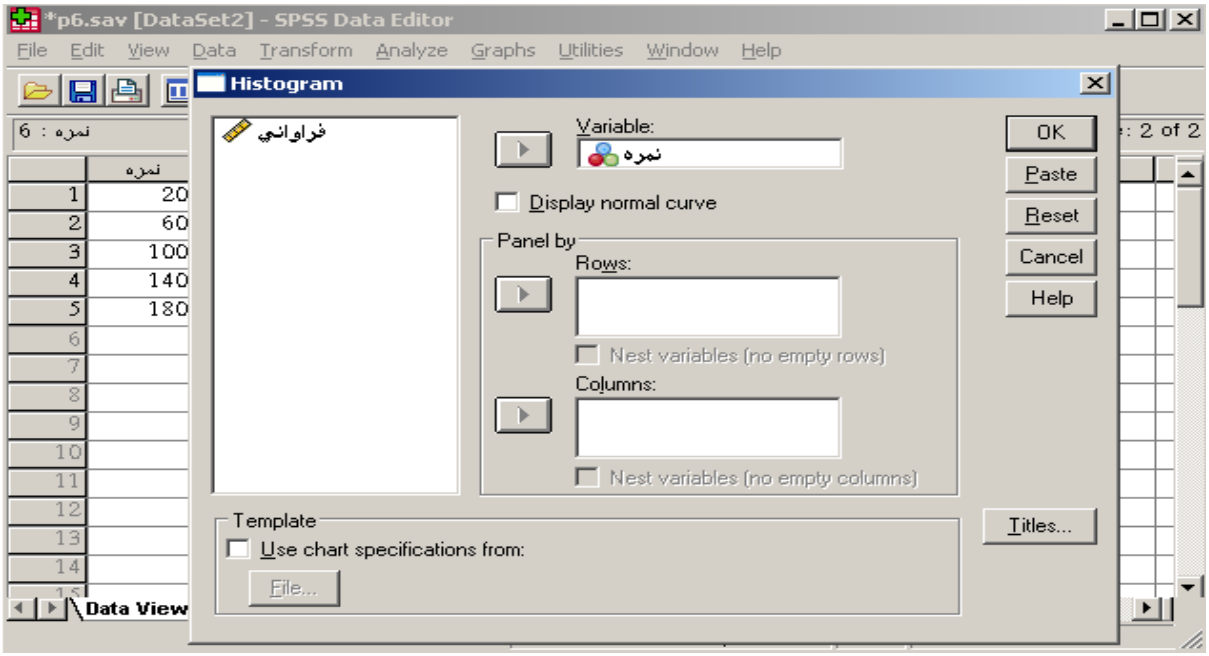
با توجه به اینکه داده های ما از نوع کمی پیوسته می باشند، ابتدا ۱۴۰ داده (نمره) را در جدولی طبقه بندی می کنیم و سپس مانند روش قبل نمودار را رسم می کنیم.

طبقات: (نمره ارزشیابی)	فراوانی
۰-۴۰	۶
۴۰-۸۰	۳۰
۸۰-۱۲۰	۸۲
۱۲۰-۱۶۰	۱۸
۱۶۰-۲۰۰	۴

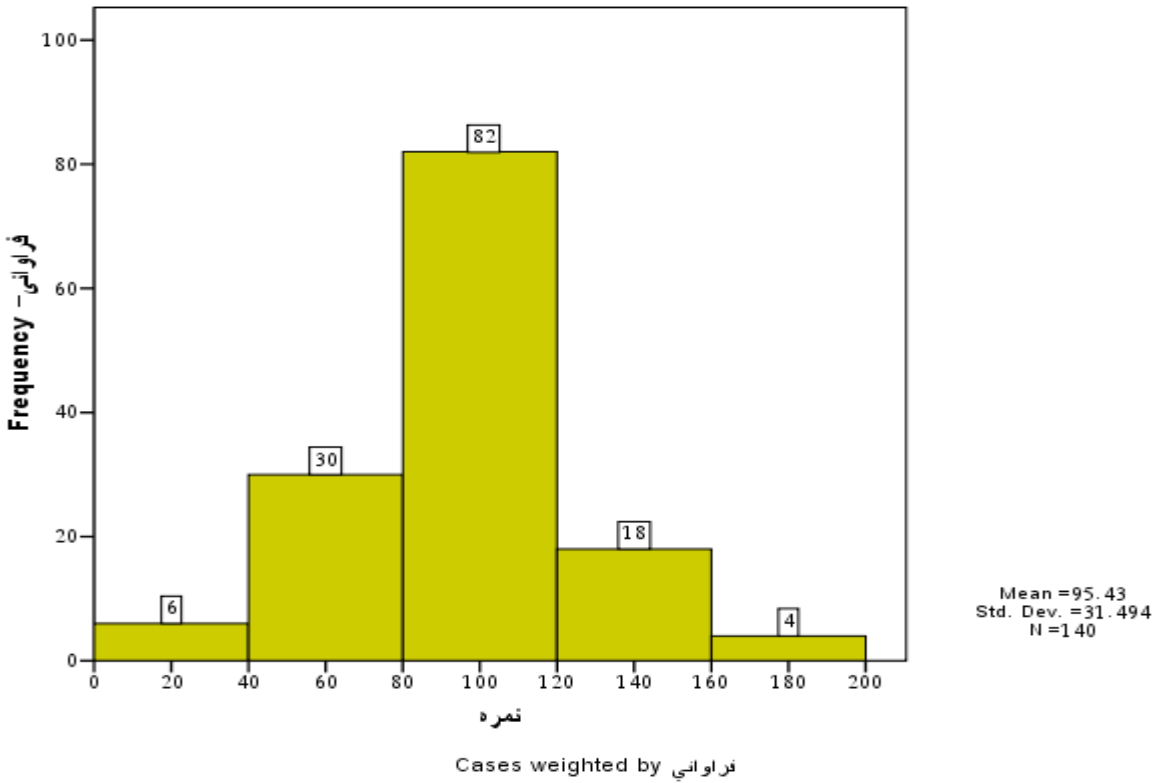


در ستون مربوط به نمره می توانیم به جای مرز وسط طبقات از کدهای ۱ تا ۵ استفاده کنیم. چون در این نوع نمودار هیستوگرام طول طبقات یکسان می باشد. در ادامه برای رسم نمودار مسیر زیر را طی می کنیم.





پس از منتقل کردن متغیر نمره از پنجره چپ به پنجره Variable بر روی Ok کلیک کرده تا نمودار ظاهر شود.



با توجه به اینکه نمودار هیستوگرام پراکندگی و تمرکز داده های کمی پیوسته را نشان می دهد نتیجه می گیریم که بیشترین پراکندگی نمرات ارزشیابی پرسنل در فاصله ۸۰ تا ۱۲۰ بوده است.

بعد از رسم نمودار و ظاهر شدن در خروجی Spss می توانیم با دو بار کلیک کردن بر روی نمودار و باز شدن پنجره Chart Editor به ویرایش نمودار مانند تغییر رنگ نمودار، ظاهر کردن فراوانی ها بر روی ستونها ، تنظیم اعداد روی محور

X ها و Y ها و ..... پرداخت.

چند نمونه آزمونهای متداول آماری :

آزمون T به طور متداول جهت تعیین معنی داری تفاوت بین دو میانگین به کار می رود.

۱) ***آزمون t*** : برای مقایسه میانگینهای ***دو نمونه مستقل*** از داده های کمی، اغلب آزمون t با نمونه های مستقل مناسب است.

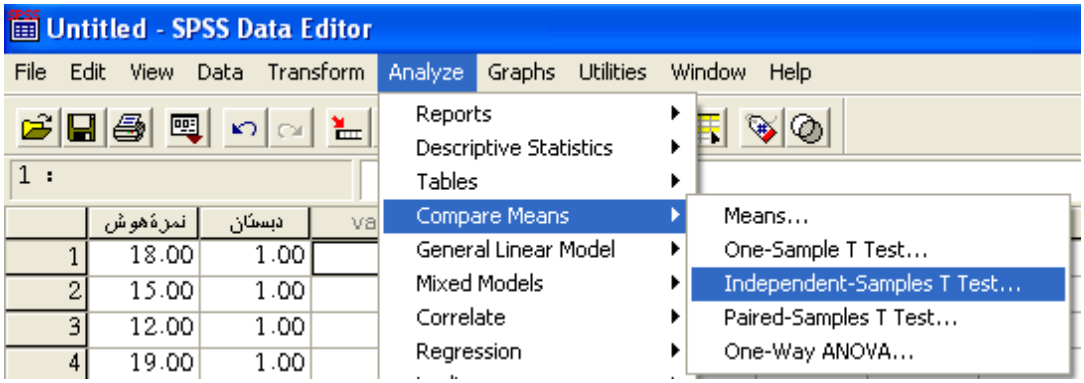
به طور مثال بررسی مقایسه میزان ضریب هوشی بین دو گروه از دانش آموزان دختر مقطع اول دو دبستان جداگانه را در نظر بگیرید. چون میزان ضریب هوشی هر کدام از این گروها تاثیری در میزان ضریب هوشی گروه دیگر ندارد، از آزمون t با نمونه های مستقل استفاده می شود.

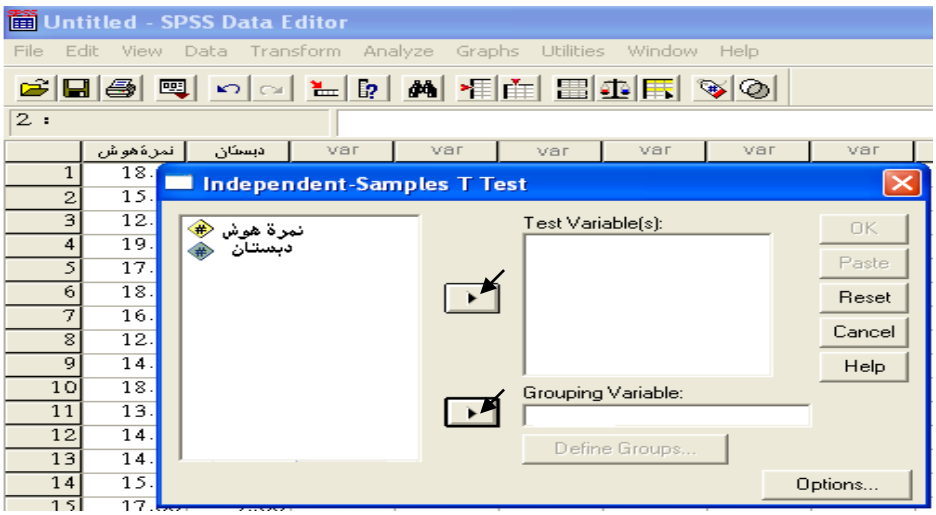
برای انجام آزمون، داده های مربوط به هر گروه را (به روش توضیح داده شده در بالا) در ستونهای جداگانه وارد کرده و به انجام آزمون مورد نظر می پردازیم.

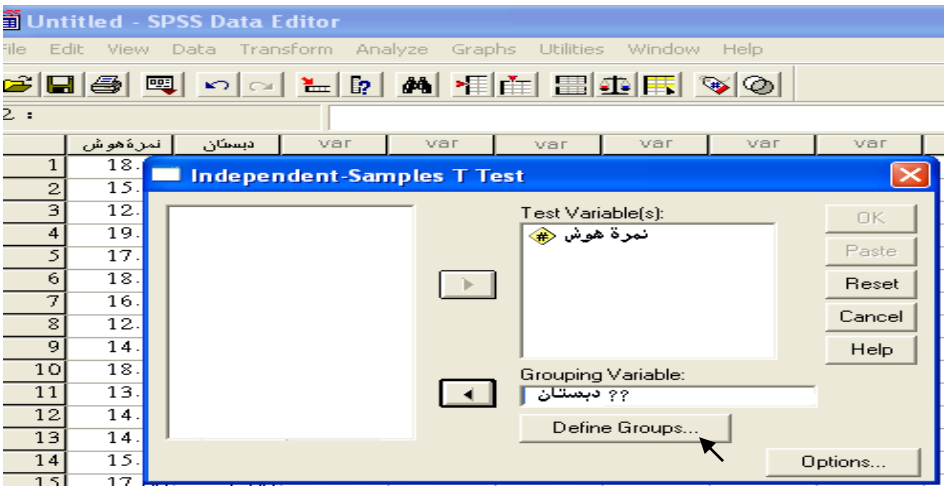
به هر کدام از دبستانها یک کد جداگانه می دهیم .

دبستان اول = (۱)

دبستان دوم = (۲)







خروجی آزمون :

Group Statistics

	دبستان	N تعداد	Mean میانگین	Std. Deviation انحراف استاندارد	Std. Error Mean میانگین انحراف استاندارد
نمرةهوش	1.00	10	15.9000	2.55821	.80898
	2.00	10	12.6000	2.87518	.90921

در خروجی جدول زیر قبل از آزمون برابری میانگینها برابری واریانس ها مورد بررسی قرار گرفته است. ابتدا فرض برابری واریانسها و همچمین فرض عدم برابری واریانسها ارائه شده است. اگر (p>.05) باشد، فرض یکنواختی واریانسها زیر سوأل نیست و از آزمون t با واریانسهای برابر استفاده می شود. اگر (p<.05)، فرض یکنواختی واریانسها زیر سوأل است و آزمون t نرمال با واریانسهای برابر باید با آزمون t بر اساس برآوردهای واریانس جداگانه جایگزین شود.

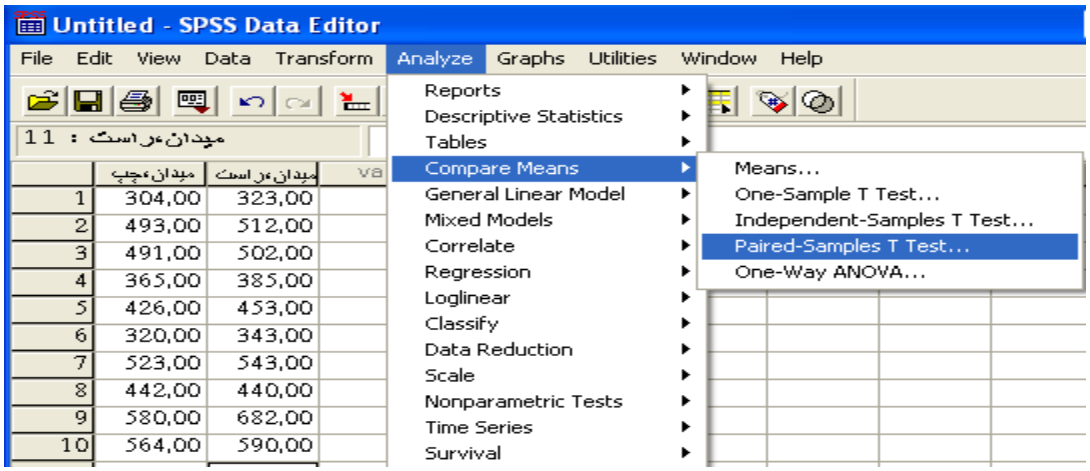
Independent Samples Test آزمون نمونه هاي مستقل

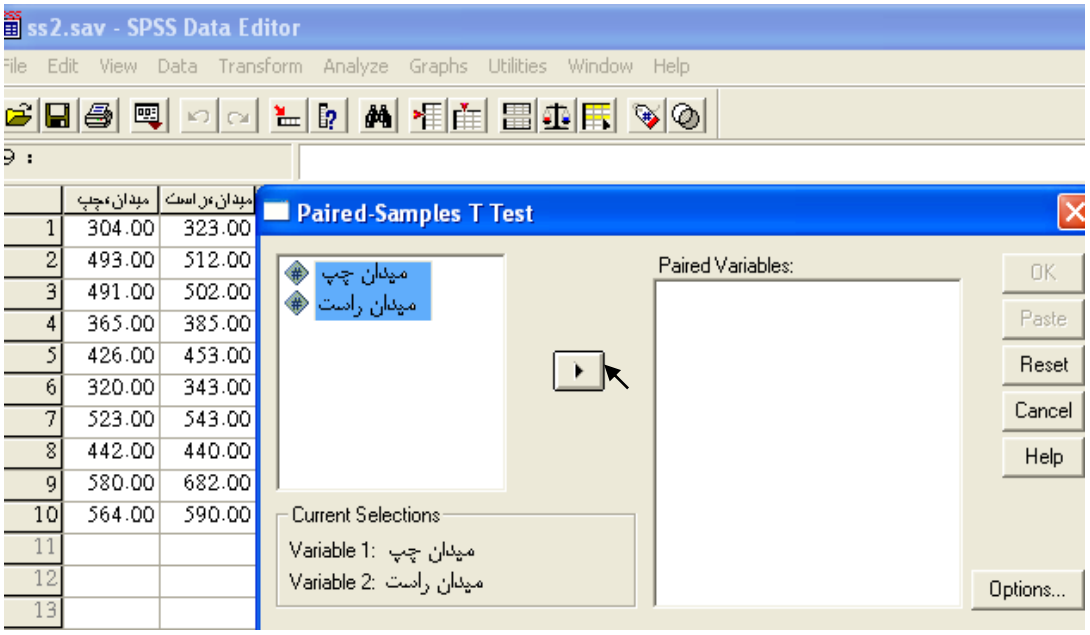
		Levene's Test for Equality of Variances آزمون فرض يكنواختي واريانسها		t-test for Equality of Means آزمونt براي برابري ميانگينها					
								95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
نمرةهوش	Equal variances assume	.434	.518	2.712	18	.014	3.30000	1.21701	.74316	5.8568
	Equal variances not assumed			2.712	17.8	.014	3.30000	1.21701	.74068	5.8593

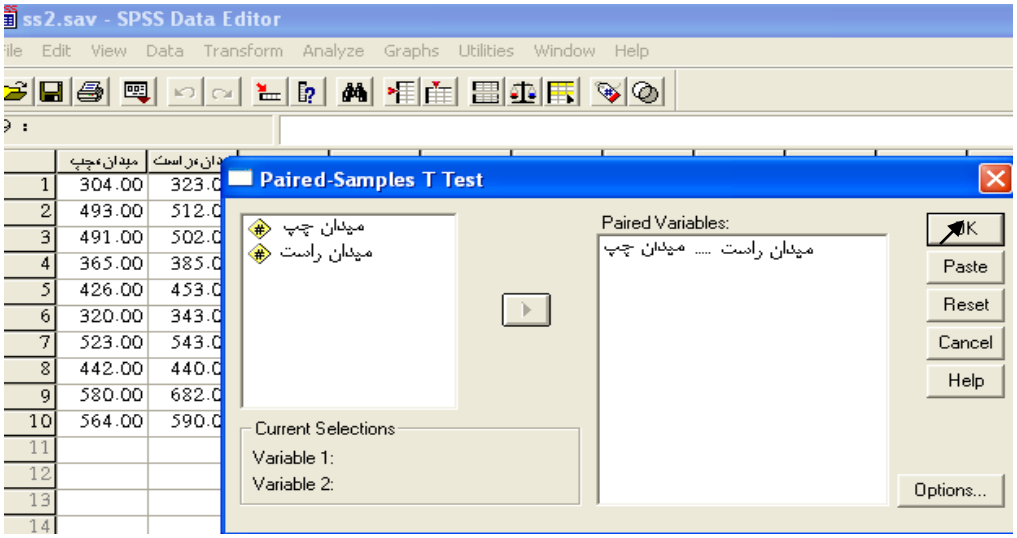
۲) ***آزمون t*** : برای مقایسه میانگینهای ***یک مجموعه از داده های وابسته*** ، از آزمون t با داده های جفت استفاده می شود.

به طور مثال در مورد چگونگی تقسیم عملکرد قشر مغز، فرض کنید ۵۰ شرکت کننده هر یک به یک هدف مثلث شکل و مربع شکل که مساحت آنها برابر است، تیراندازی می کنند. در این حالت دو نمونه از امتیازات به دست می آید . این مطالعه سبب ایجاد دو نمونه وابسته از داده ها می شود. به این معنی که افراد تواناتر نسبت به افراد ضعیفتر در هر دو قسمت آزمایش احتمالاً امتیاز بیشتری کسب می نمایند. بنابراین انتظار می رود دو نمونه دارای همبستگی باشند.

برای مثالی دیگر در بررسی چگونگی تقسیم عملکرد قشر مغز، شرکت کننده به یک نقطه مرکزی در صفحه کامپیوتر نگاه می کند و از وی خواسته می شود هنگامی که کلماتی را که در هر طرف از نقطه ظاهر می شود، تشخیص داد، دکمه ای را فشار بدهد. به منظور اطمینان از اینکه کلمه مورد نظر را به درستی تشخیص داده است، از شرکت کننده خواسته می شود که آن را تایپ کند. فرضیه آزمایش آن است که لغاتی که در میدان بینایی راست ظاهر می شوند، سریعتر از آنهایی که در میدان بینایی چپ ظاهر می شوند، تشخیص داده می شود. زیرا اولی توسط نیمکره چپ مغزی پردازش می شود که تصور می شود در اطلاعات کلامی قویتر است. در این مثال که تنها دو حالت وجود دارد، آزمایش ایجاد داده های جفت می کند.







خروجی آزمون :

Paired Samples Statistics آماره هاي نمونه هاي جفت(دو تايي)

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	میدان چپ	450.8000	10	97.08507	30.70100
	میدان راست	477.3000	10	112.09128	35.44638

Paired Samples Correlations همبستگي نمونه هاي جفت

		N	Correlation	Sig.
Pair 1	میدان چپ & میدان راست	10	.975	.000

Paired Samples Test آزمون نمونه هاي جفت

		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	میدان چپ - میدان راست	-26.50	27.81387	-46.397	-6.60316	-3.01	9	.015

آزمونهای رگرسیونی

مجموعه ای از روشها که به عنوان ***روشهای رگرسیون*** (Regression Method) شناخته می شوند، وجود دارند که از ***وجود رابطه بین دو متغیر*** برای پیش بینی کردن مقدار یک متغیر (وابسته) از روی متغیر (مستقل) خبر می دهند .

در ***رگرسیون*** هدف آن است که بعضی از مشخصات از روی آگاهی در مورد سایر مشخصات برآورد یا پیش بینی شود که این کار از طریق ساختن معادله رگرسیون صورت می گیرد.

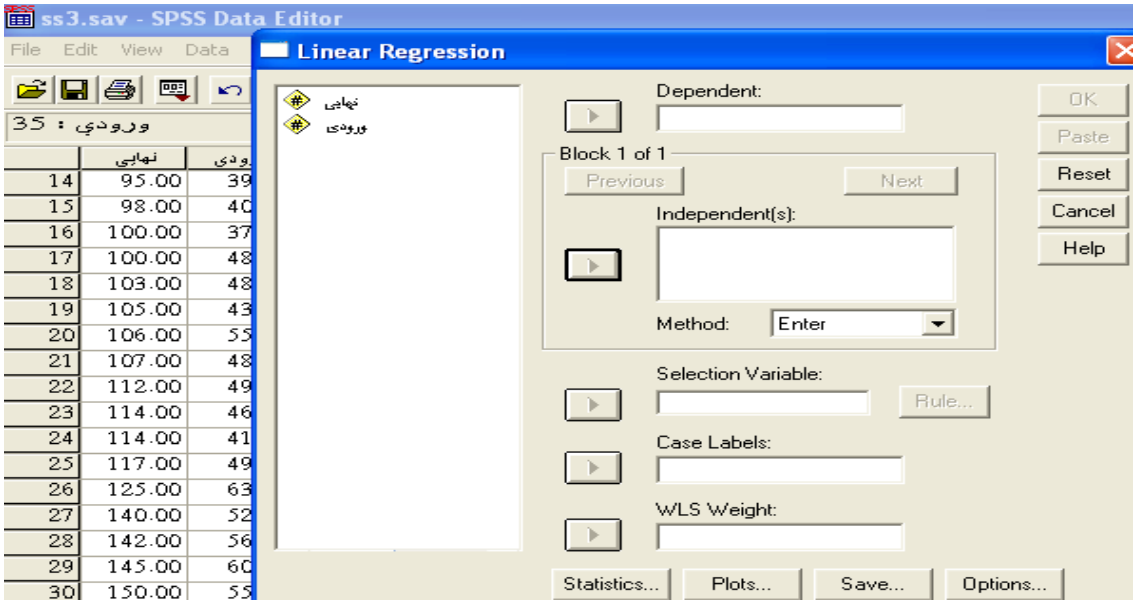
در رگرسیون ***دو متغیره ساده*** (Simple two variable) مقادیر یک متغیر (متغیر وابسته یا y) از روی مقادیر متغیر دیگری (متغیر مستقل یا x) به کمک یک معادله خطی (خط مستقیم) برآورد می شود.

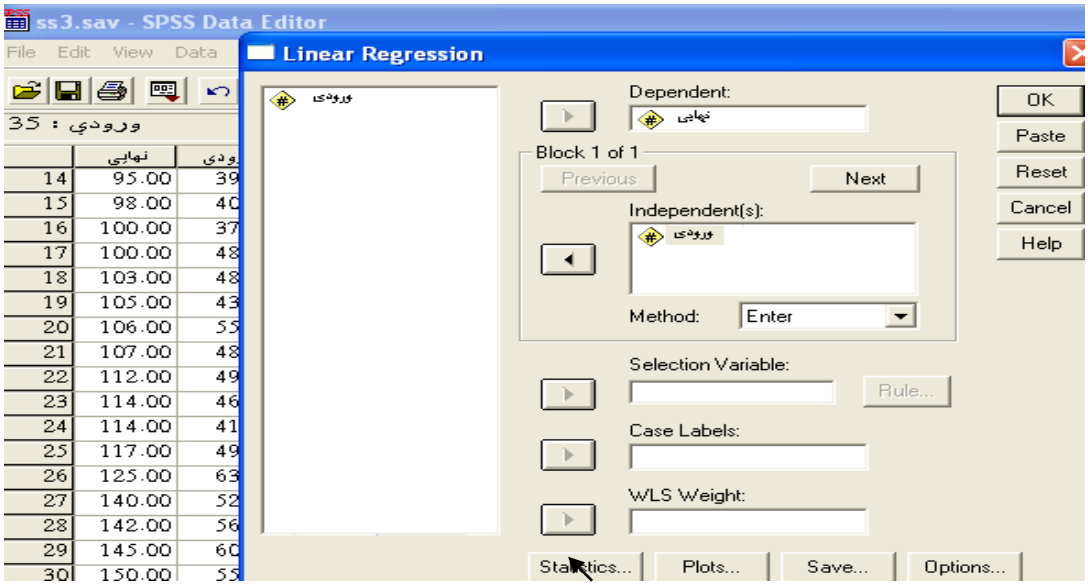
برای روشن شدن موضوع مورد بحث، مسئولان یک دانشگاه را در نظر بگیرید . مسئولان در مورد کارآیی روشهایی که برای انتخاب دانشجو استفاده می شود حساسیت زیادی دارند. به این صورت که چقدر بین عملکرد فرد در امتحان ورودی با عملکرد وی در امتحان نهایی رابطه نزدیکی وجود دارد. همچنین تا چه میزان می توان عملکرد دانشجویان را در دانشگاه از روی نمرات آنها در آزمونهای آماری پیش بینی کرد.

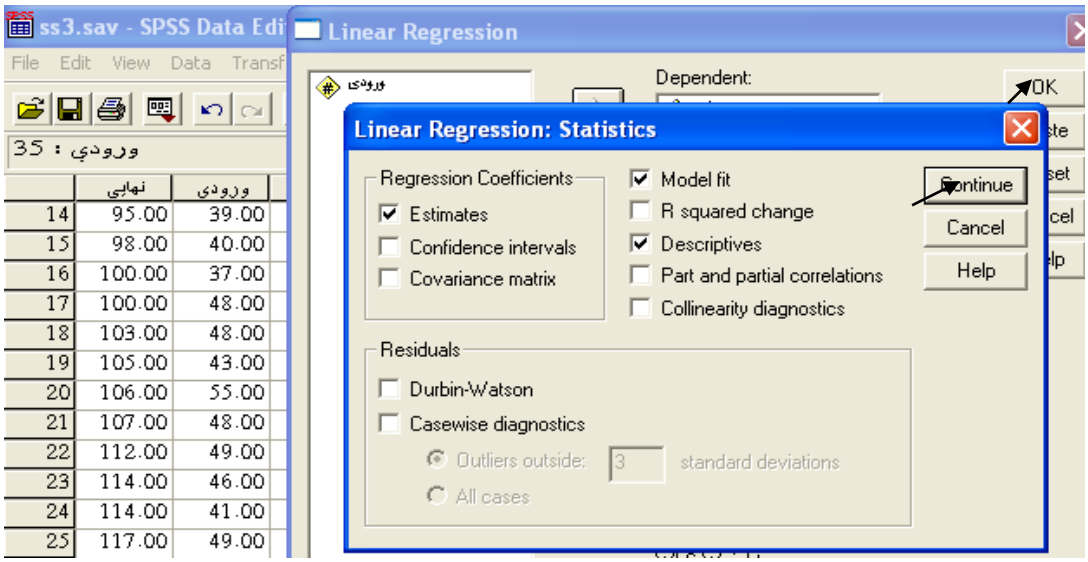
Y= نمرات امتحان نهایی دانشگاه

X= نمرات امتحان ورودی

$Y=aX+b$




ss3.sav - SPSS Data Editor
File Edit View Data
35 : ورودي
نهايي
ورودي
14 95.00 39
15 98.00 40
16 100.00 37
17 100.00 48
18 103.00 48
19 105.00 43
20 106.00 55
21 107.00 48
22 112.00 49
23 114.00 46
24 114.00 41
25 117.00 49
26 125.00 63
27 140.00 52
28 142.00 56
29 145.00 60
30 150.00 55
Linear Regression
ورودی
Dependent:
نهایی
Block 1 of 1
Previous
Next
Independent(s):
ورودی
Method: Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
OK
Paste
Reset
Cancel
Help


ss3.sav - SPSS Data Edi
File Edit View Data Transf
35 : ورودي
نهايي
ورودي
14 95.00 39.00
15 98.00 40.00
16 100.00 37.00
17 100.00 48.00
18 103.00 48.00
19 105.00 43.00
20 106.00 55.00
21 107.00 48.00
22 112.00 49.00
23 114.00 46.00
24 114.00 41.00
25 117.00 49.00
Linear Regression
ورودی
Dependent:
OK
Linear Regression: Statistics
Regression Coefficients
Estimates
Confidence intervals
Covariance matrix
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Continue
Cancel
Help
Residuals
Durbin-Watson
Casewise diagnostics
Outliers outside: 3 standard deviations
All cases

Descriptive Statistics آماره هاي توصيفي

	Mean ميانگين	Std. Deviation انحراف استاندارد	N
نهايي	105,5294	35,81284	34
ورودي	47,3235	7,42935	34

Correlations همبستگي

		نهايي	ورودي
Pearson Correlation ضرائب پيرسون	نهايي	1,000	,671
	ورودي	,671	1,000
Sig. (1-tailed)	نهايي	.	,000
	ورودي	,000	.
N	نهايي	34	34
	ورودي	34	34

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19054,121	1	19054,121	26,202	,000[a]
	Residual	23270,350	32	727,198		
	Total	42324,471	33			

a. Predictors: (Constant), ورودي

b. Dependent Variable: نهايي

Coefficients[a] ضرائب

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-47,531	30,257		-1,571	,126
	ورودي	3,234	,632	,671	5,119	,000

a. نهايي: متغير وابسته

آزمون مربع کای برای رابطه بین دو متغیر کیفی

آزمون X^2 برای بررسی مستقل بودن دو متغیر تصادفی مثلاً مستقل بودن قد و وزن نوزاد ، همچنین برای همتوزیع بودن دو متغیر تصادفی به طور مثال همتوزیع بودن وزن نوزاد ایرانی و نوزاد چینی به کار می برند.

معمولاً برای بررسی استقلال یا عدم استقلال از دو ویژگی یا بیشتر از جمعیتی از این آزمون استفاده می کنیم.

این آزمون را با ارائه مثالی توضیح می دهیم: در این آزمون دو ویژگی مورد نظر یکی جنسیت و دیگری نوع انتخاب شیء توسط کودکان می باشد.

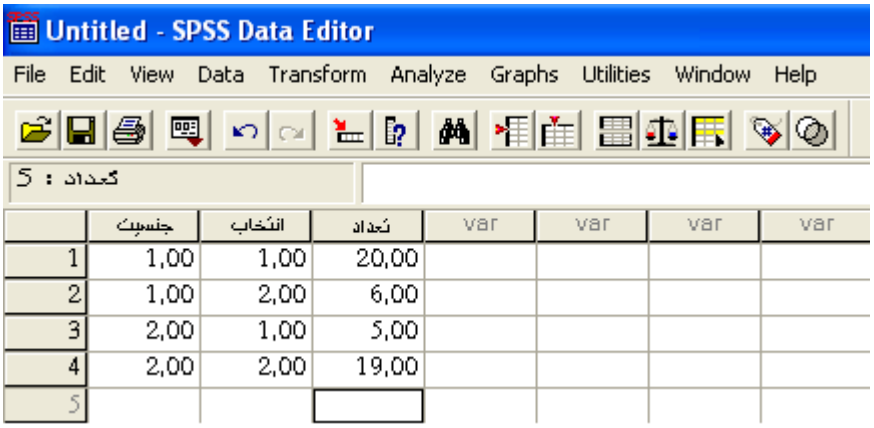
آزمایشی بر روی انتخابهای کودکان

فرض کنید محققی به مشاهده تعدادی کودک می پردازد که وارد یک اتاق می شوند و انتخاب هر یک از کودکان از بین دو شیء را ثابت می‌کند. وی می خواهد بداند که آیا تمایل پسرها و دخترها در انتخاب اشیاء متفاوت، با هم فرق می کند. در این آزمایش دو متغیر جنسیت و انتخاب شیء مورد توجه است. در اصطلاح آماری، این محقق می پرسد که آیا این متغیرها رابطه ای با یکدیگر دارند؟ آیا دخترها نسبت به پسرها یکی از اشیاء را بیشتر انتخاب می کنند و آیا پسرها شیء دیگر را بیشتر از دخترها انتخاب می‌کنند؟ جدول (۲×۲)

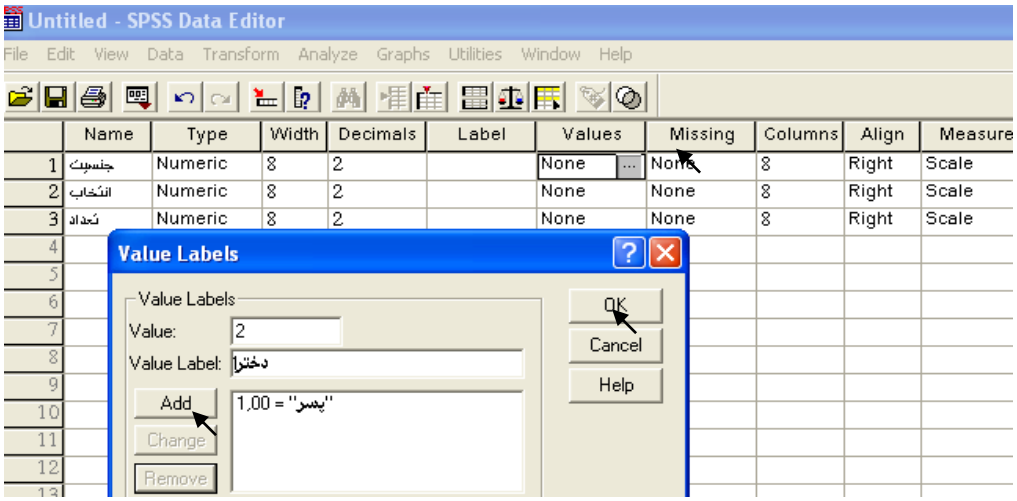
انتخاب‌های ۵۰ کودک از میان دو شیء

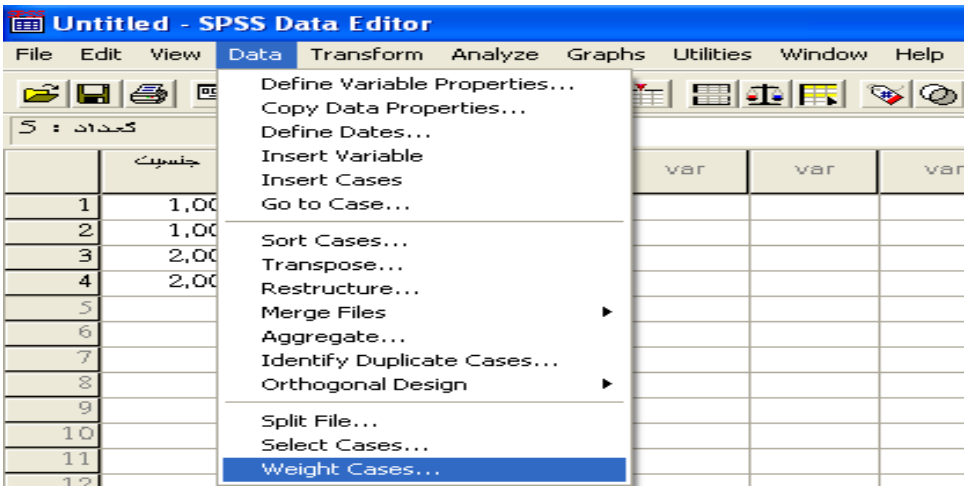
شیء	پسرها(۱)	دخترها(۲)
A (۱)	۲۰	۵
B (۲)	۶	۱۹

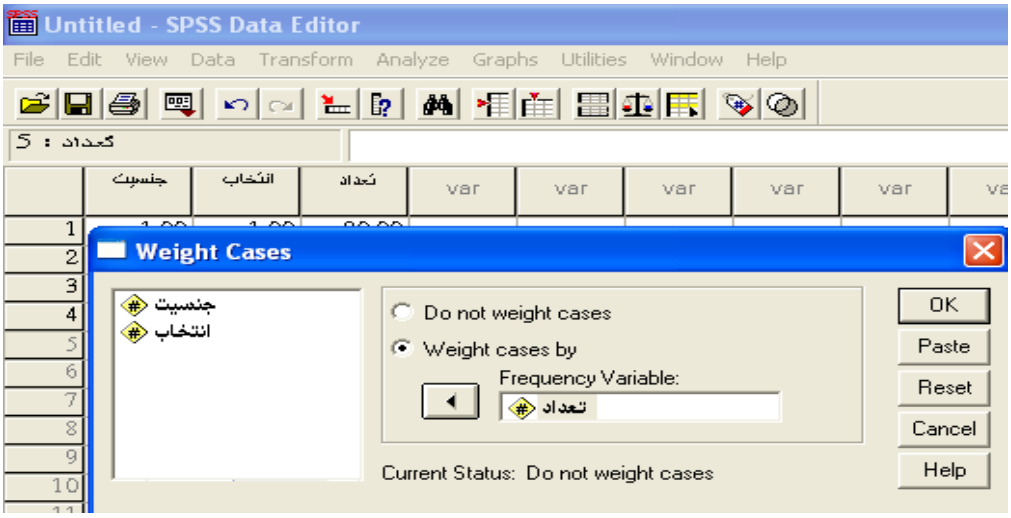
در این بررسی به جای وارد کردن ۵۰ داده مورد نظر داده ها را به صورت خلاصه مطابق شکل زیر وارد کرده و به هر خانه جدول بالا وزن معادل آنها را می دهیم .(تعداد)

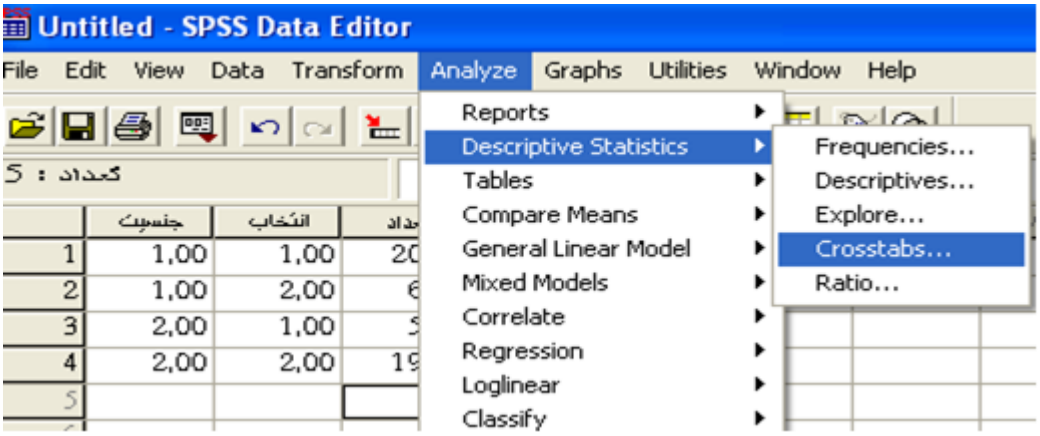


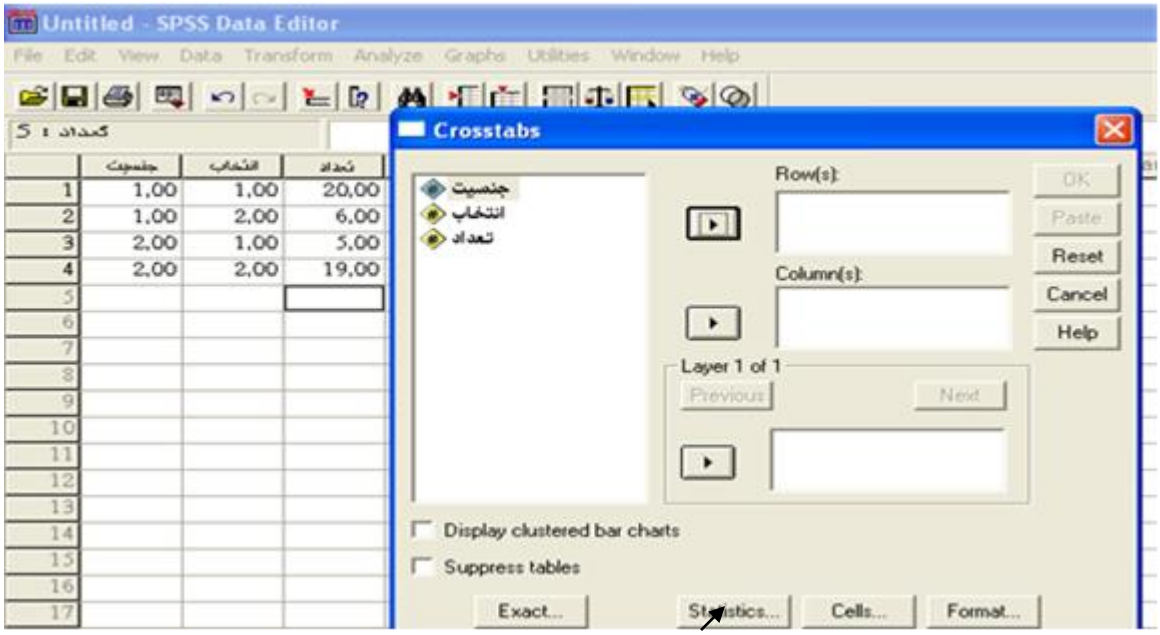
در این مرحله برای انتخاب کردن معادل کد های ۱ و ۲ در ستونهای جنسیت و انتخاب (از نظر پسر . دختـر بـودن و همچنین شیء Aو B) بر روی Variable View در پایین صفحه کلیـک کـرده تـا پنجـره زیـر بـاز شـود و سـپس مطابق زیر، معادل کدهای ۱ و ۲ را برای دو ستون مورد نظر تعریف می کنیم .برای ستون انتخاب هم به روش مشـابه این مراحل را انجام می دهیم.

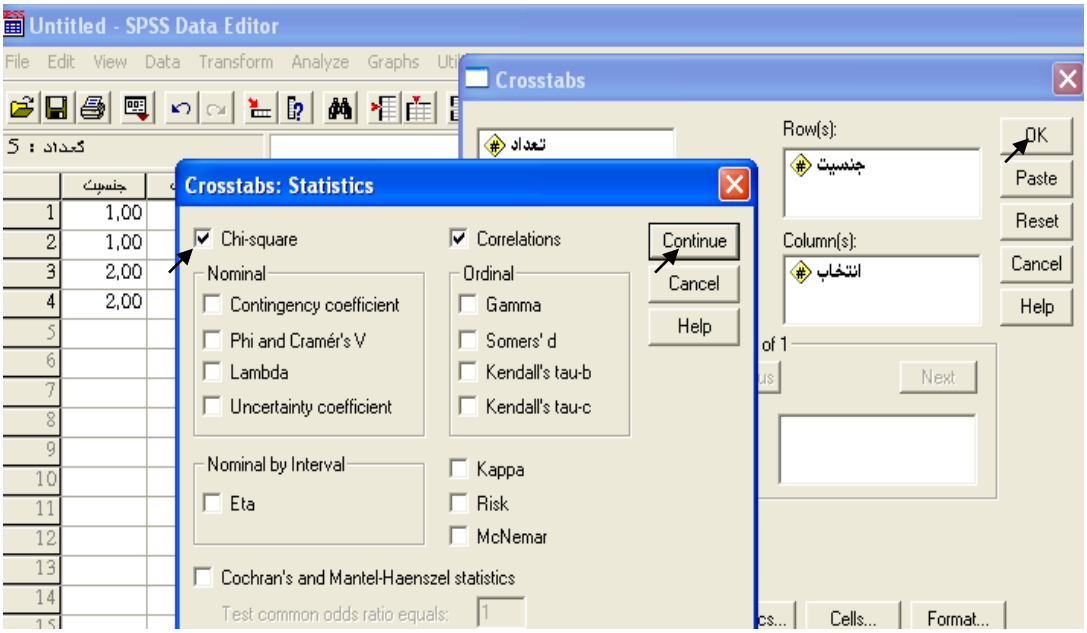





Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
5 : تعداد
جنسیت
انتخاب
تعداد
var
Weight Cases
جنسیت
انتخاب
Do not weight cases
Weight cases by
Frequency Variable:
تعداد
OK
Paste
Reset
Cancel
Help
Current Status: Do not weight cases


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
5 : تعداد
جنسیت
انتخاب
1 1,00 1,00
2 1,00 2,00
3 2,00 1,00
4 2,00 2,00
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Crosstabs
جنسیت
انتخاب
تعداد
Row(s):
Column(s):
OK
Paste
Reset
Cancel
Help
Layer 1 of 1
Previous
Next
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...


Untitled - SPSS Data Editor
Crosstabs
Crosstabs: Statistics
Chi-square
Correlations
Continue
Cancel
Help
Nominal
Contingency coefficient
Phi and Cramér's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals:
Row(s):
Column(s):
OK
Paste
Reset
Cancel
Help
Next
Cells...
Format...

Crosstabulation جنسیت * انتخاب

Count

		انتخاب		
		شيء A	شيء B	Total
جنسیت	پسر	20	6	26
	دختر	5	19	24
Total		25	25	50

آزمون کای اسکور Chi-Square Tests

	Value	درجه آزادی df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	15,705[b]	1	,000		
Continuity Correction[a]	13,542	1	,000		
Likelihood Ratio	16,661	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	15,391	1	,000		
N of Valid Cases	50				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,00.

آنالیز واریانس:

روشی که برای مقایسه بیش از دو تیمار به کار می رود تجزیه و تحلیل واریانس(***ANOVA***) نامیده می شود. 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 می گیرد.

هدف از آزمون بررسی زیر می باشد:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k \\ H_1: \mu_i \neq \mu_j \quad \text{حداقل یکی از } \mu_i \text{ با سایر آنها تفاوت داشته باشد} \end{cases}$$

روش آزمون آنالیز واریانس را با انجام یک مثال بیان می کنیم.

برای این کار ابتدا دو فرض را با هم مقایسه می کنیم. اگر فرض H_o پذیرفته شود که تجزیه و تحلیل به پایان می رسد و نشاندهنده این موضوع می باشد که میان تیمارهای (میانگینها) گروههای تفاوتی وجود ندارد.
اما اگر فرض H_o رد شود نشاندهنده اختلاف میان تیمارها می باشد و باید بدنبال اختلافها بگردیم.

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

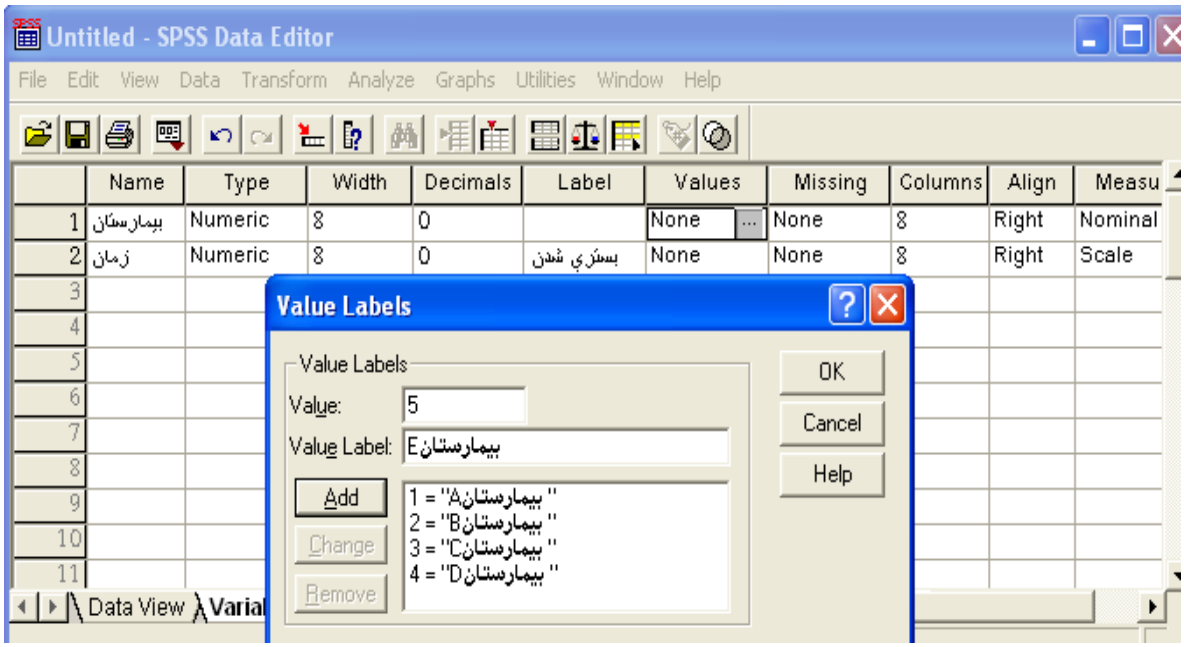
تیمار= متوسط زمان بستری شدن بیماران

تعداد / تیمار	1	2	3	4	5	6	7	8	9	10	μ_i	$\bar{y}_i$
بیمارستان A	۷	۷	۸	6	7	5	8	7	6	۸	μ_1	
بیمارستان B	8	8	8	8	7	7	6	6	6	5	μ_2	
بیمارستان C	7	5	5	5	4	7	4	4	5	5	μ_3	
بیمارستانD	8	9	9	11	6	10	11	11	10	12	μ_4	
بیمارستانE	4	9	6	4	4	4	5	5	4	6	μ_5	

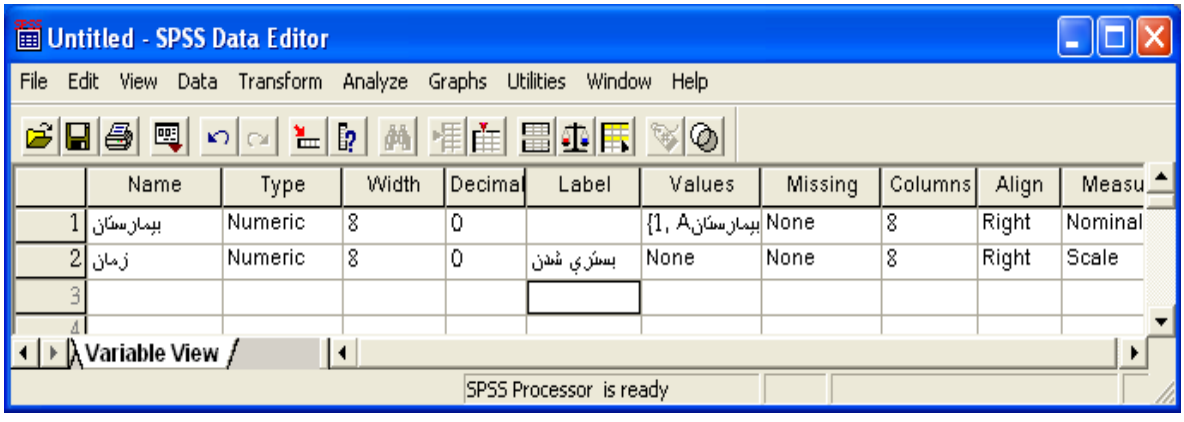
جدول ۱

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

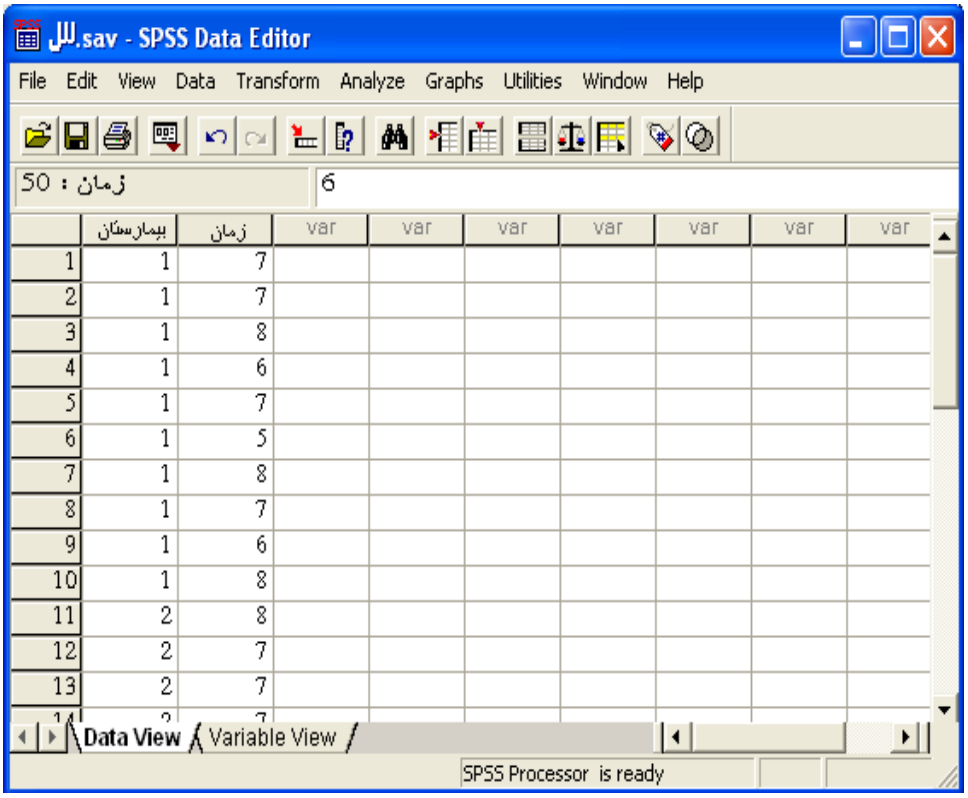
در این قسمت برای متغیر نوع بیمارستان در ستون ***Values*** نوع بیمارستانها را تعریف می کنیم.
(در پنجره ***Data View*** به جای اسامی بیمارستان از کدهای (۱ و ۲ و.....) استفاده می کنیم).



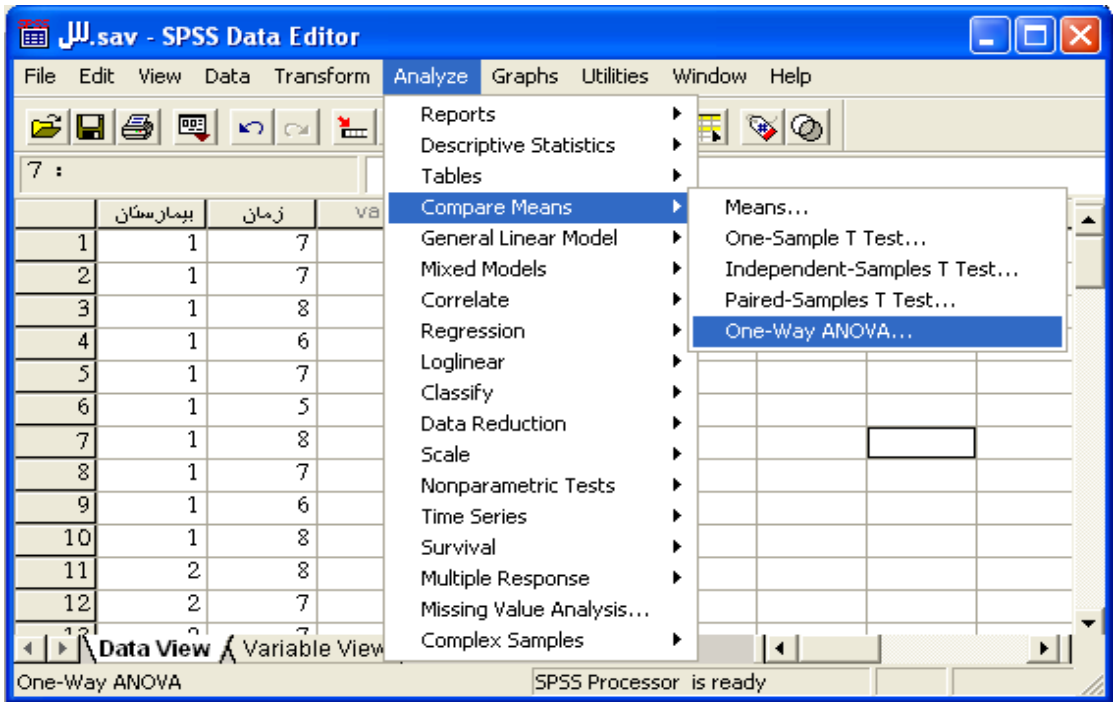
و در پایان با کلیک کردن بر روی ***ok*** پنجره ***Variable View*** به صورت زیر در می آید.



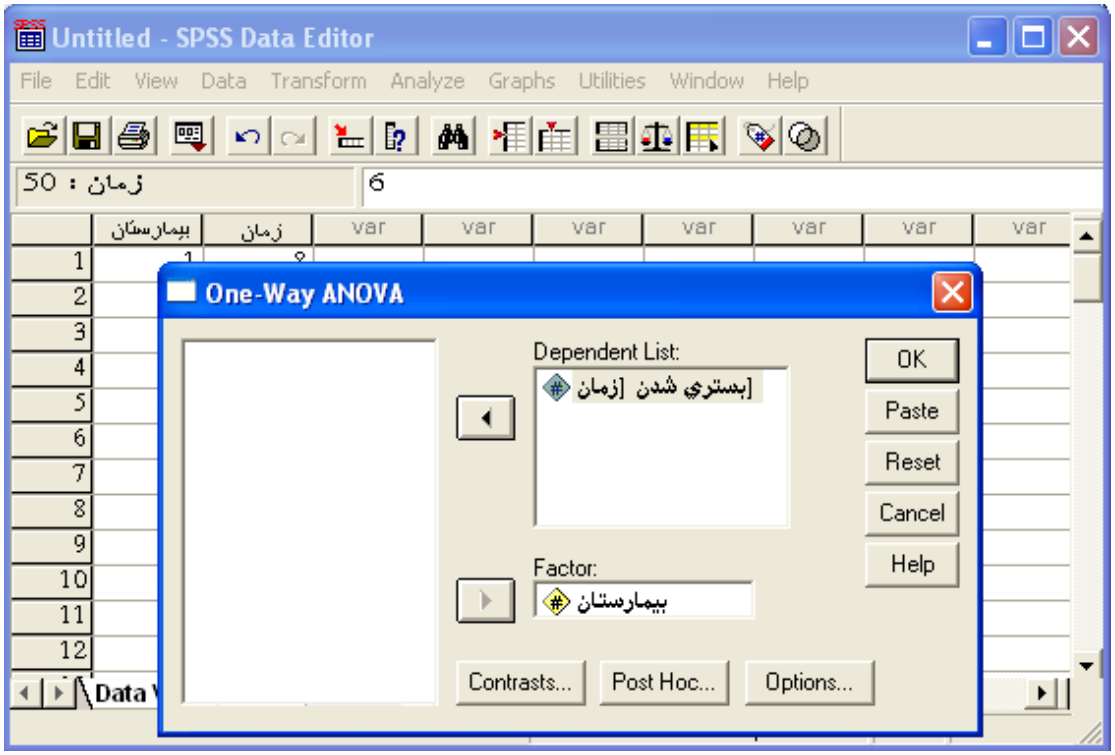
در پنجره ***Data View*** داده ها را به صورت زیر وارد می کنیم. در ستون بیمارستان کد مربوط به بیمارستانها را (۱-۵) وارد کرده و جلوی هر کد در ستون زمان، مدت زمان بستری شدن بیماران بیمارستانهای مختلف را طبق جدول ۱ وارد می کنیم.



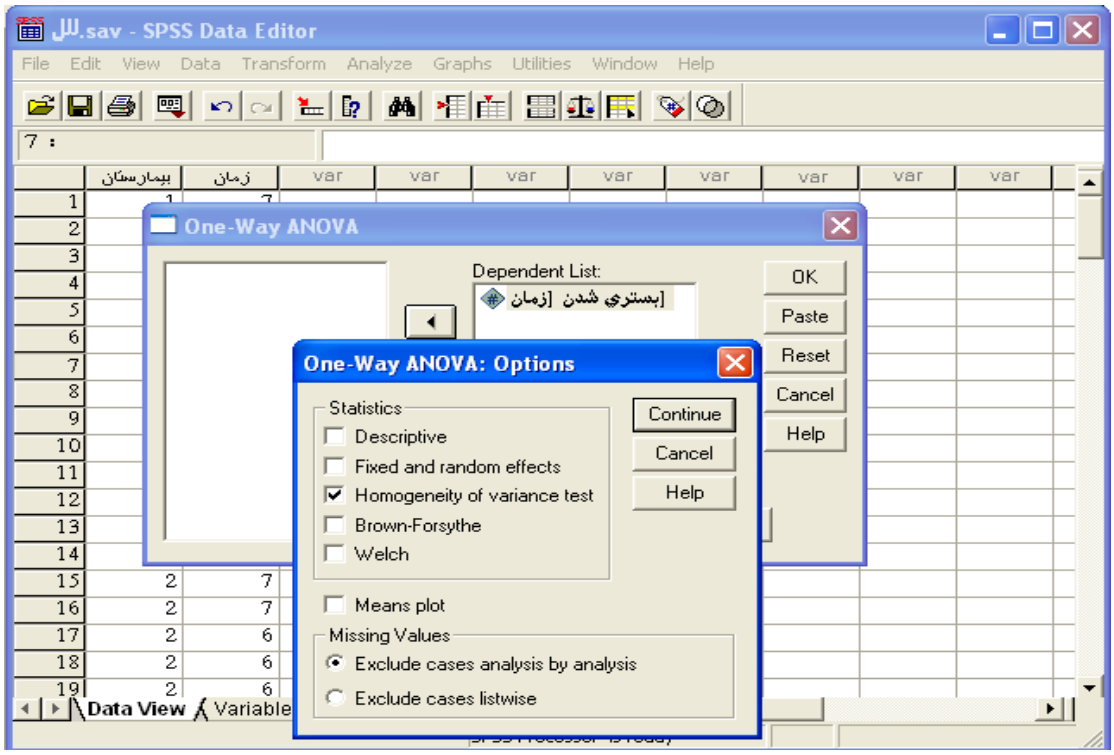
برای انجام آزمون به صورت زیر عمل می کنیم:



سپس با کلیک کردن بر روی دو متغیر و با استفاده از مثلث های سیاه کوچک آنها را به مستطیلهای سمت راست منتقل می کنیم .



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*** را فعـال کـرده ، در ادامـه ابتدا بر روی کلمه ***Continue*** و سپس ***Ok*** کلیک می کنیم تا خروجیهای زیر بدست آیند.

آزمون همگنی واریانسها (جدول ۲)

Test of Homogeneity of Variances

بستري شدن

Levene Statistic	df1	df2	Sig.
1.344	4	45	.269

آزمون برابری میانگینها (جدول ۳)

ANOVA

بستري شدن

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	141.680	4	35.420	20.646	.000
Within Groups	77.200	45	1.716		
Total	218.880	49			

نتایج بدست آمده از جدول ۲ نشان می دهد که در آزمون مقایسه بین واریانسهای ***مدت زمان بستری شدن*** ۵ بیمارستان اختلاف معنی داری از نظر آماری وجود ندارد (P-value>0.05) .اما نتایج بدست آمده از جدول ۳ نشان می دهد که میان میانگینهای ***مدت زمان بستری شدن*** ۵ بیمارستان اختلاف معنی داری وجود دارد. (P-value <0.05).

به دلیل اینکه داده ها نشان می دهند که میانگینهای ۵ بیمارستان با هم تفاوت معنی داری دارند در نتیجه به دنبال اختلافها می باشیم.

بدین منظور مسیر بالا را دوباره تکرار می کنیم 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 می توانیم 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:

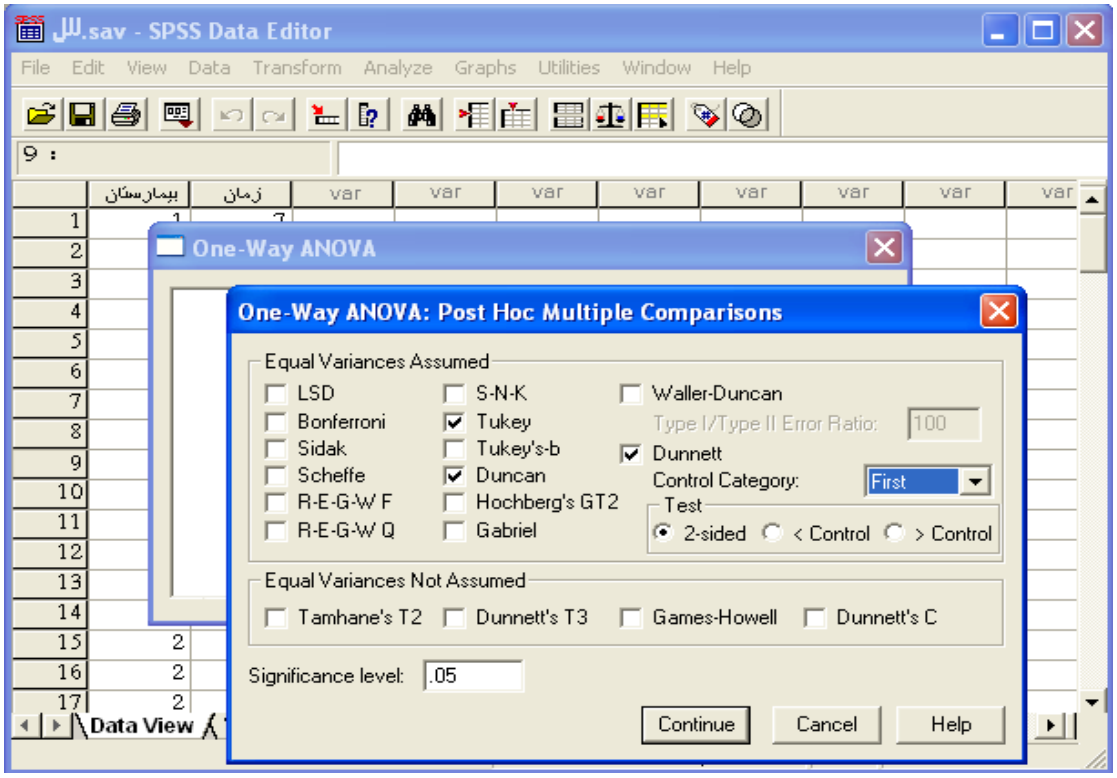
(Equal Variances Assumed)

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_0 که برابری واریانسها را مطرح می کند رد نشده شده است از آزمونهای بالایی استفاده می کنیم.

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



در پنجره بالا در حالتی که واریانس گروهها 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 می گیریم تا سایر گروهها را با آن بسنجند . این گروه می توان گروه اول(***First***) یا گروه آخر (***Last***) باشد. انتخاب هر کدام از این گروهها به عنوان گروه اول یا آخر در نتایج تغییری ایجاد نمی کند. برای این کار با فعال کردن آزمون ***Dunnett*** گزینه ***Control Category*** فعال شده و در مربع روبروی آن گروه کنترل را انتخاب می کنیم. در مرحله بعد با کلیک بر روی ***Continue*** به پنجره قبلی می رویم و با کلیک بر روی ***Ok*** می توانیم خروجیهای مورد نظر را مشاهده کنیم.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: بستري شدن

Dunnett t (2-sided)[a]

(I) بیمارستان	(J) بیمارستان	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Bبیمارستان	Aبیمارستان	.000	.603	1.000	-1.53	1.53
Cبیمارستان	Aبیمارستان	-1.800*	.603	.016	-3.33	-.27
Dبیمارستان	Aبیمارستان	2.800*	.603	.000	1.27	4.33
Eبیمارستان	Aبیمارستان	-1.800*	.603	.016	-3.33	-.27

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

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Multiple Comparisons

Dependent Variable: بستري شدن

	(I) بيمارستان	(J) بيمارستان	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	بيمارستانA	بيمارستانB	.000	.603	1.000	-1.71	1.71
		بيمارستانC	1.800*	.603	.035	.09	3.51
		بيمارستانD	-2.800*	.603	.000	-4.51	-1.09
		بيمارستانE	1.800*	.603	.035	.09	3.51
	بيمارستانB	بيمارستانA	.000	.603	1.000	-1.71	1.71
		بيمارستانC	1.800*	.603	.035	.09	3.51
		بيمارستانD	-2.800*	.603	.000	-4.51	-1.09
		بيمارستانE	1.800*	.603	.035	.09	3.51
	بيمارستانC	بيمارستانA	-1.800*	.603	.035	-3.51	-.09
		بيمارستانB	-1.800*	.603	.035	-3.51	-.09
		بيمارستانD	-4.600*	.603	.000	-6.31	-2.89
		بيمارستانE	.000	.603	1.000	-1.71	1.71
	بيمارستانD	بيمارستانA	2.800*	.603	.000	1.09	4.51
		بيمارستانB	2.800*	.603	.000	1.09	4.51
		بيمارستانC	4.600*	.603	.000	2.89	6.31
		بيمارستانE	4.600*	.603	.000	2.89	6.31
	بيمارستانE	بيمارستانA	-1.800*	.603	.035	-3.51	-.09
		بيمارستانB	-1.800*	.603	.035	-3.51	-.09
		بيمارستانC	.000	.603	1.000	-1.71	1.71
		بيمارستانD	-4.600*	.603	.000	-6.31	-2.89

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

Homogeneous Subsets

بستري شدن

	بيمارستان	N	Subset for alpha = .05		
			1	2	3
Tukey HSD[a]	بيمارستانC	10	5.10		
	بيمارستانE	10	5.10		
	بيمارستانA	10		6.90	
	بيمارستانB	10		6.90	
	بيمارستانD	10			9.70
	Sig.		1.000	1.000	1.000
Duncan[a]	بيمارستانC	10	5.10		
	بيمارستانE	10	5.10		
	بيمارستانA	10		6.90	
	بيمارستانB	10		6.90	
	بيمارستانD	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 یـا عـدم برابـری میانگینهـا از دو سـتون 95% Confidence Interval و Sig استفاده می کنیم. عدد مشاهده شده در ستون Sig معرف P-Valuse بدست آمده در آزمونها می باشـد و چون آزمونها در سطح ۰،۰۵ مورد بررسی قرار می گیرند مبنای پذیرش یا عدم پذیرش آنها یعنـی قبـول یـا رد فـرض اولیه H_o مقایسه با مقدار ۰،۰۵ می باشد.

در آزمونهای یک دامنه اگر Sig<0.05 فرض اولیه H_o رد مـی شـود و اگـر Sig>0.05 فـرض اولیـه H_o رد نمـی 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) اسـت کـه نمـایش دهنـده 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: فاصله اطمینان $= 1-\alpha$
بنابراین وقتی مقدار α یعنی سطح معنی داری برابر ۰،۰۵ باشد فاصله اطمینان ۰،۹۵ می شود.

این فاصله اطمینان یک حد پایین(L) و یک حد بالا(U) دارد و اگر عدد صفر را شـامل شـود(L<0<U) نشـاندهنده این است که فرض H_o یعنی برابری میانگینها رد نمی شود و این هم ارز Sig>0.05 می باشد و اگر این بـازه شـامل صفر نباشد هم ارز این است که Sig<0.05 و بیان می کند که فرض H_o یعنی برابری میانگینها رد می شود.

در خروجی مربوط به آزمون Dunnett مشاهده می کنیم: بیمارستان A که به عنـوان گـروه کنتـرل در نظـر گرفتـه شد با تک تک بیمارستانها از نظر میانگین مدت زمان بستری شدن مقایسه آماری شد و نتایج در جـدول آورده شـده اسـت. در سـتون مربـوط بـه Sig کـه مقـادیر P-Valuse را بـرای هـر ازمـون جداگانـه نشـان مـی دهـد، هـر جـا Sig<0.025 باشد نشان می دهد که فرض برابری دو میانگین رد شده است. به بیان دیگر بین میانگین مـدت زمـان بستری شدن برای دو بیمارستان تفاوت معنی داری وجود دارد. علت استفاده از Sig<0.025 بـه جـای Sig<0.05 در این است که آزمون Dunnett یک ازمون دو دامنه است(2-Side).
نتایج نشان می دهند که میانگین بیمارستان A با بیمارستان B تفـاوت نـدارد(Sig>0.025). امـا میـان میـانگین بیمارستان A با میانگین سایر بیمارستانها تفاوت معنی داری وجود دارد(Sig<0.025).

درخروجی مربوط به آزمون Tukey تک تک بیمارستانها با هم مقایسه می شوند. در این آازمون به علت یک دامنـه بودن هر کجا Sig<0.05 فرض برابری میانگینها رد می شود(به بیان دیگر بین میانگین مدت زمـان بسـتری شـدن دو بیمارستان از نظر آماری تفاوت معنی داری وجود دارد.)

در خروجـی مربـوط بـه زیرمجموعـه هـای همگـن (Homogeneous Subsets) آزمونهـای تـوکی و دانکـن میانگینهایی که با هم تفاوت ندارند را در یک زیر گروه قرار می دهند.
به طور معمول نتایج بدست آمده از آزمون ها هر کدام تأییدی بر نتیجه آزمون دیگر است. البته بـا توجـه بـه درجـه دقت آزمونها در بعضی مواقع ممکن است که نتیجه بدست آمده در یک آزمون با نتیجه بدست آمده در آزمـون دیگـر متفاوت باشد.

آزمون ضریب همبستگی

همبستگی (Correlation)

در بسیاری از موارد در انجام مطالعات تحقیقاتی به دنبال بررسی رابطه دو متغیر تصادفی می باشیم که هیچ کدام را نمی توان به عنوان علت برای دیگری انتخاب کرد. در اصل برای بررسی میزان هماهنگی میان دو متغیر به دنبال شاخصهایی می گردیم که در اصل دو ویژگی زیر را داشته باشند:

۱- به واحد دو جامعه وابسته نباشد

۲- کراندار باشد

به طور مثال در تحولات اقصادی به دنبال رابطه میان تقاضای نفت خام در برابر تقاضای طلا می باشیم. یا در مطالعه تحولات اجتماعی به دنبال رابطه درآمد سرپرست خانواده و میزان تحصیل فرزندان می باشیم و مثالهایی از این دست.....

مجموعه اطلاعات (داده های) موجود در انجام یک آزمون همبستگی که شامل اندازه های بدست آمده از دو متغیر **X** و **Y** می باشند را می توان به صورت یک نمونه تصادفی دو متغیره $(X_1, Y_1), \ldots\ldots\ldots, (X_n, Y_n)$ بیان کرد.

مطالعه رابطه بین متغیرها بوسیله ((تحلیل همبستگی)) (Corroletion Analysis) انجام می شود. که بیانگر وجود یک رابطه خطی بین دو متغیر می باشد.

فرمول ضریب همبستگی به صورت زیر می باشد

$$r = \frac{Cov(X,Y)}{S_x S_y} = \frac{\sum\sum x_i y_i - n\bar{x}\bar{y}}{S_x S_y}$$

با توجه به مقدار r در حاتهای مختلف تفسیرهای گوناگونی از رابطه X و Y خواهیم داشت.

$-1<r<1$

حالتهای مختلف برای r

۱- $r = 1$ در این حالت همبستگی کامل و مستقیم گوییم. با افزایش مقدار x مقدار y به طور قطعی زیاد می شود.

۲- $r = -1$ در این حالت همبستگی را کامل و معکوس گوییم. با افزایش مقدار x مقدار y کاهش می یابد.

۳- $-1 < r < 0$ همبستگی ناقص و معکوس است. با افزایش مقدار x مقدار y به طور نسبی کاهش می یابد.

۴- $۰ < r < ۱$ همبستگی ناقص و مستقیم است. با افزایش مقدار x مقدار y به طور نسبی افزایش می یابد.

۵- $r = 0$
- ۱- رابطه خطی وجود ندارد. (به طور مثال رابطه ممکن است از نوع درجه دو باشد)
- ۲- شیب خط صفر می باشد.

انواع ضریب همبستگی با توجه به نوع متغیرهای مورد مطالعه

۱- پیرسن : در این روش متغیر X و Y هر دو پیوسته می باشند. (در پنجره ***Variable View*** در ستون مربوط به مقیاسها(***Measure***) داده ها باید از نوع ***Scale*** انتخاب شوند.)

۲- کندال : در این روش هر دو متغیر **X** و **Y** باید به صورت طبقه بندی شده باشد یعنی(***Ordinal,Nominal***) (در پنجره Variable Viewدر ستون مربوطه باید داده ها از نوع Ordinal انتخاب شوند.)

۳- اسپیرمن : در این روش متغیر **X** گسسته و متغیر **Y**پیوسته می باشد.
(متغیر X در پنجره ***Variable View***در ستون مربوطه باید از نوع ***Ordinal*** و متغیر Y از نوع ***Scale*** انتخاب شوند.)

در آزمونهای بالا متغیر **X** را به عنوان متغیر مستقل و متغیر **Y** را به عنوان متغیر وابسته در نظر می گیریم.

– نکته:. در بعضی مواقع می توانیم با طبقه بندی داده های پیوسته آنها را به صورت طبقه بندی شده در آوریم .

به طور مثال اگر بخواهیم تأثیر درآمد خانواده را بر روی معدل فرزندان بررسی کنیم به دلیل اینکه درآمد و معدل هر دو متغیر پیوسته می باشند باید از ضریی همبستگی پیرسن استفاده کنیم. اما می توانیم با تقسیم بندی درآمد به طور مثال به سه گروه کم درآمد، متوسط و پر درآمد از ضریب همبستگی اسپیرمن استفاده کنیم.

در آزمونهای همبستگی یک طرفه فرضهای زیر بررسی می شوند.

$$1\begin{cases} H_o: \ r = o \\ H_1: \ r > o \end{cases} \qquad 2\begin{cases} H_o: \ r = o \\ H_1: \ r < o \end{cases}$$

جهت نامساوی با توجه به برآورد نمونه ای r تعیین می شوند. اگر r مثبت باشد آزمون ۱ و اگر منفی باشد از آزمون ۲ استفاده می کنیم.

مثال

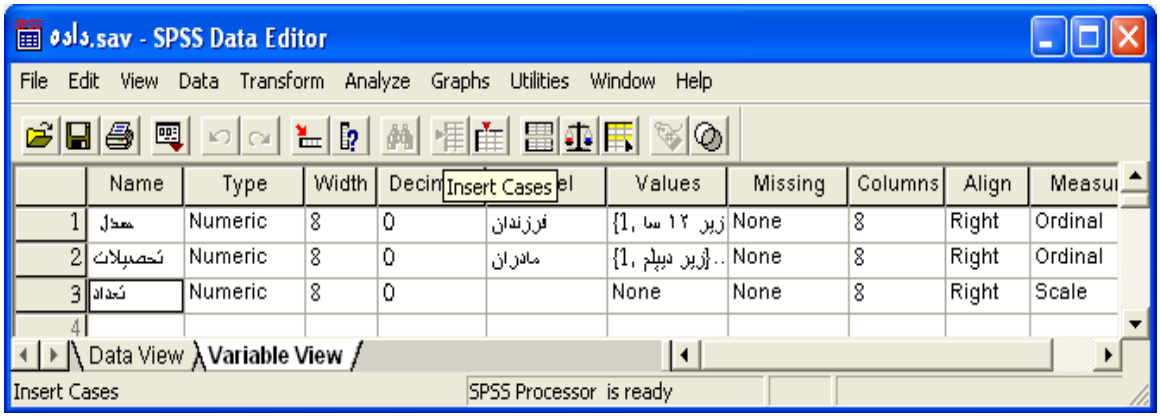
در جدول زیر میزان معدل دانش آموزان و میزان تحصیلات آنها آمده است. هدف تعیین میزان همبستگی و نوع ارتباط معدل با میزان تحصیلات مادر می باشد

.

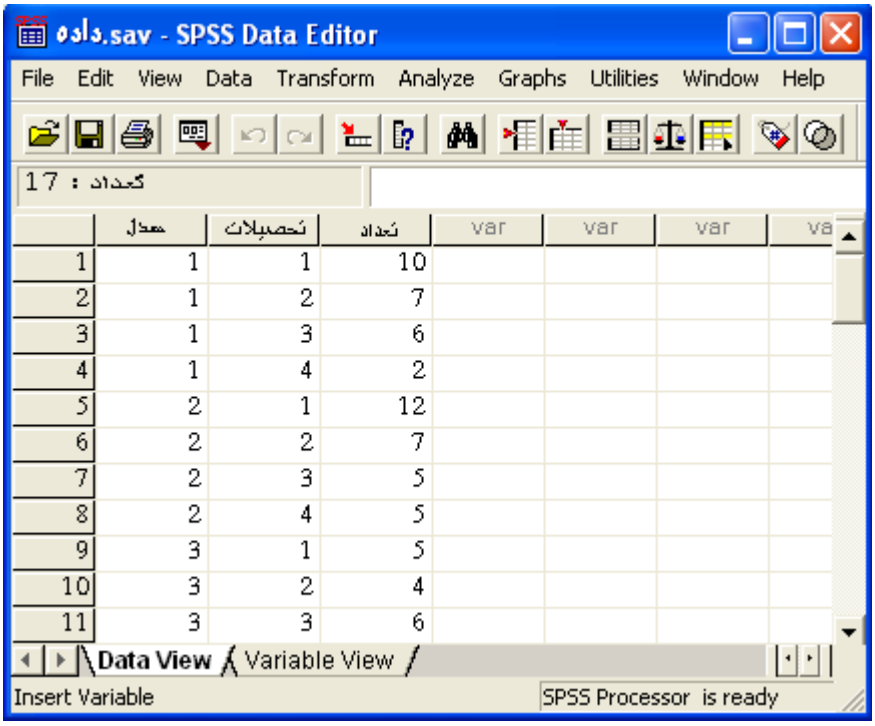
تحصیلات / معدل	(۱) زیردیپلم	(۲) دیپلم	(۳) فوق دیپلم	(۴) لیسانس وبالاتر
(۱) زیر ۱۲	۱۰	۷	۶	۲
(۲) ۱۲-۱۵	۱۲	۷	۵	۵
(۳) ۱۵-۱۷	۵	۴	۶	۱۰
(۴) ۱۷-۲۰	۵	۷	۱۰	۱۲

وقتی در نرم افزار Spss داده ها را به صورت رتبه ای مشخص کردیم می توانیم در تجزیه و تحلیلها تفسیر راحتری از خروجیها داشته باشیم.

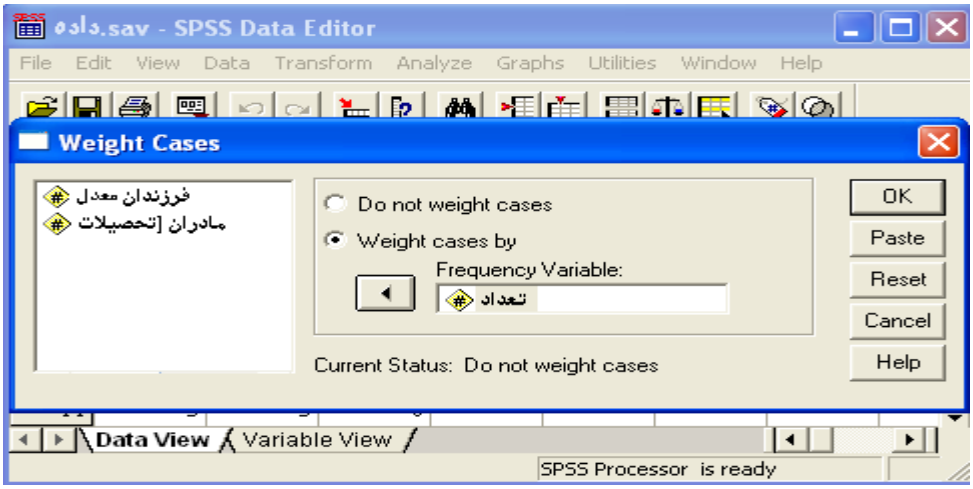
به مانند مثال قبلی در پنجره ***Variable View*** دو متغیر معدل و تحصیلات را وارد می کنیم و ستونها را متناسب با نوع متغیرها تنظیم می کنیم. در ستون ***Values*** با توجه به کدهایی که در جدول بالا داده شده است متغیرها را معرفی می کنیم.

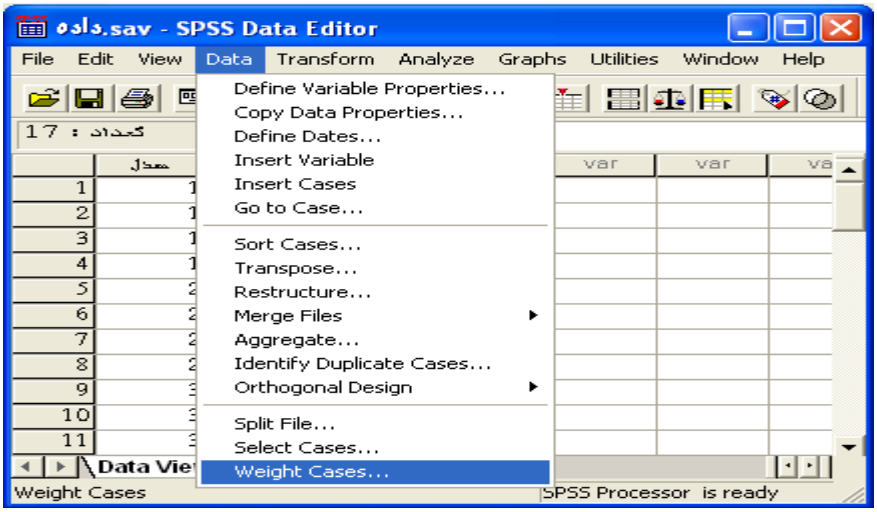


سپس داده را به صورت زیر در پنجره ***Data View*** وارد می کنیم

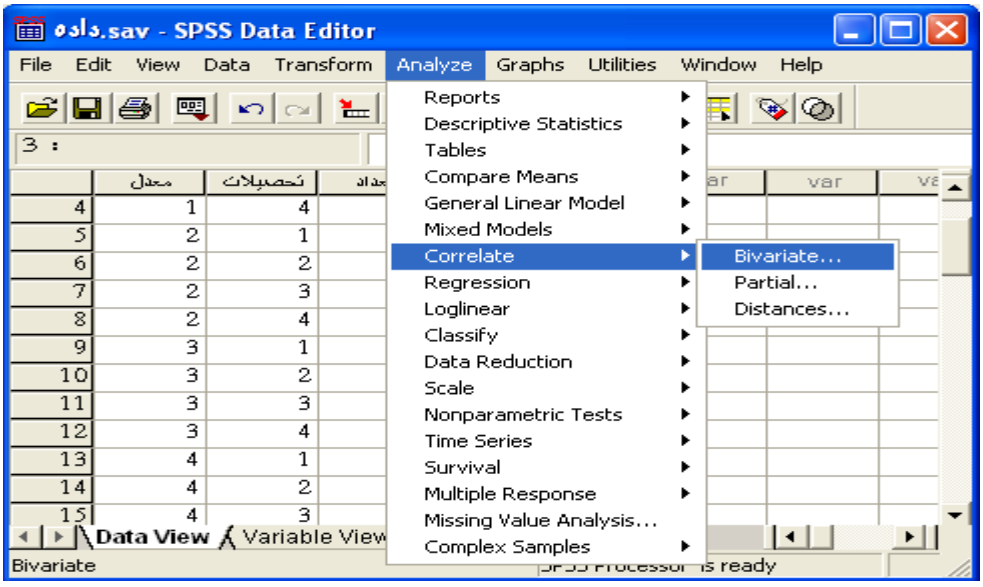


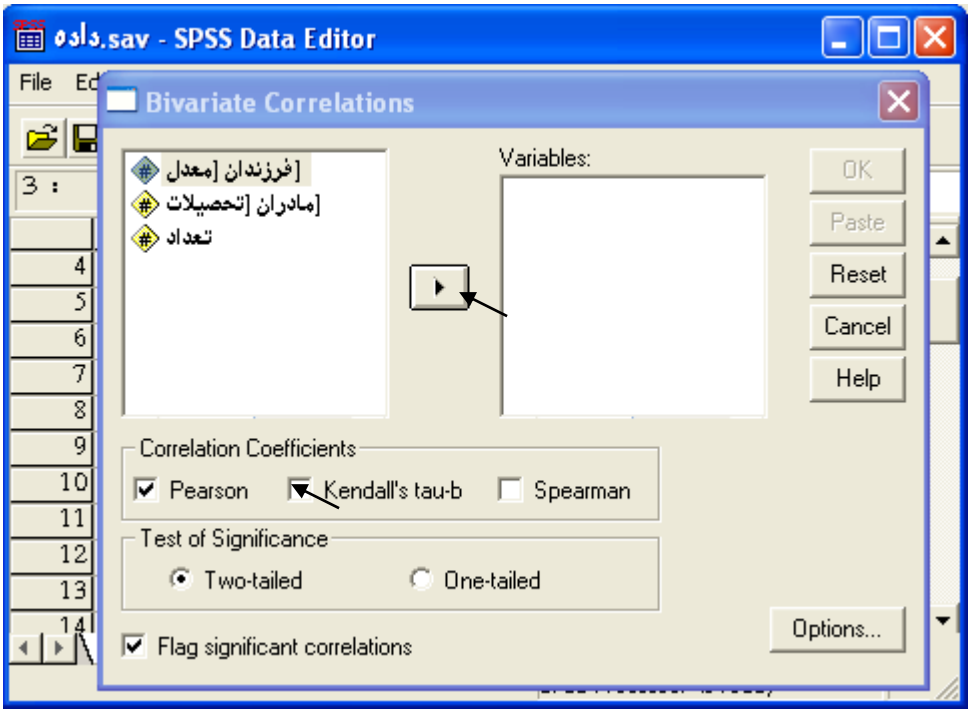
برای اینکه مشخص کنیم ستون تعداد معرف فراوانی هایی هست که به دو متغیر سن و تحصیلات نسبت داده شده است و و از نوع اندازه گیری شده نمی باشند باید دو ستون تحصیلات و سن توسط ستون تعداد وزن دار شود. برای این کار به صورت زیر عمل می کنیم.





با کلیک بر روی کلمه Ok داده ها بوسیله ستون تعداد، وزن دار می شوند. سپس مسیر زیر را انتخاب می کنیم Bivariate Correlation باز شود.





با بردن دو متغیر سن فرزندان و تحصیلات مادران به مربع سمت راستی (***Variable***) و انتخاب ضریب همبستگی متناسب با داده ها، در این سوال (فعال کردن گزینه ***Kendall's tau-b*** در قسمت ***Correlation Coefficients***) و انتخاب نوع آزمون (یک طرفه بودن(یک دامنه) یا دو طرفه بودن(دو دامنه) در قسمت ***Test of significance***) با کلیک کردن بر روی ***Ok*** خروجیهای مربوط به آزمون را به صورت زیر مشاهده کرد.

Correlations

			فرزندان	مادران
Kendall's tau_b	فرزندان	Correlation Coefficien	1.000	.260**
		Sig. (2-tailed)	.	.001
		N	113	113
	مادران	Correlation Coefficien	.260**	1.000
		Sig. (2-tailed)	.001	.
		N	113	113

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

مقدار ***P-Values*** بدست آمده برابر است ۰،۰۰۱ می باشد که با توجه به دو دامنه بودن آزمون (***2-tailed***) با ۰،۰۲۵ مقایسه می شود و چون $P<0.025$ می باشد نتیجه می شود که یک نوع رابطه بین معدل دانش آموزان و تحصیلات مادران آنها وجود دارد و چون $r = 0.260$ بدست آمده است و $0<r<1$ می باشد نتیجه می گیریم که همبستگی از نوع مستقیم و ناقص می باشد و اینگونه تفسیر می شود که با افزایش تحصیلات مادران (با رفتن از کد ۱ به سمت کد ۴) معدل دانش آموزان نیز (با رفتن از کد ۱ به سمت کد ۴) افزایش پیدا می کند و نشان دهنده نقش موثر و مثبت تحصیل مادر بر بالا بردن کیفیت تحصیلی فرزندانشان می باشد. در این مثال اگر تک تک معدلها را به همراه نوع مدرک مادر در اختیار داشتیم، در یک ستون نوع مدرک و در ستون دیگر داده های پیوسته معدل را وارد می کردیم و برای انجام آزمون از ضریب همبستگی اسپیرمن استفاده می کردیم.