



بسمه تعالی

تہیہ کنندہ :
واحد توسعه تحقیقات بالینی بیمارستان کارگر نژاد ومتینی



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 و تحلیل های آماری است؛ نرم افزاری آماری با قابلیت های 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 نرم افزار که یکی از نرم افزارهای تخصصی آمار است، بیشتر به بحث های آماری در حیطه علوم اجتماعی، روانشناسی و علوم رفتاری و ... می پردازد.

قابلیت های نرم افزار SPSS به شرح زیر است:

- تهیه خلاصه های آماری مانند گرافها، جداول، آماره ها و ...
- انواع توابع ریاضی مانند قدر مطلق، تابع علامت، لگاریتم، توابع مثلثاتی و ...
- تهیه انواع جداول سفارشی مانند جداول فراوانی، فراوانی جمعی، درصد فراوانی و ...
- انواع توزیع های آماری شامل توزیع های گسسته و پیوسته
- تهیه انواع طرح های آماری
- انجام آنالیز واریانس یکطرفه، دوطرفه، چندطرفه و آنالیز کوواریانس
- تکنیک های تجزیه و تحلیل سری های زمانی
- ایجاد داده های تصادفی و پیوسته
- محاسبه انواع آماره های توصیفی
- انواع آزمون های مرتبط با مقایسه میانگین بین دو یا چند جامعه مستقل و وابسته
- قابلیت مبادله اطلاعات با نرم افزارهای دیگر
- برآزش انواع مختلف رگرسیون



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 کاملاً جدید را کشف خواهیم کرد.

زمانی به استفاده از علم آمار رو می آوریم که:

- 1) بخواهیم داده ها را به صورت یک مجموعه، خلاصه کرده و توصیف نماییم.
- 2) 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:

- 1) انجام استنباط درباره جامعه از طریق تجزیه و تحلیل اطلاعات موجود در داده های نمونه ای.
- 2) 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 طوریکه مشاهدات مبنایی برای استخراج استنباط های معتبر تشکیل دهند.

انتخاب یک آزمون آماری :

انتخاب یک آزمون آماری وابسته به موارد زیر است:

- 1) پرسش پژوهش
- 2) نقشه یا طرح پژوهش
- 3) 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 یا 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) از آن



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 پارامترهای 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 اصولاً بدون داشتن فرضیه آماری امکان انجام یک آزمون دشوار است. فرضیه آماری به دو دسته فرض صفر (H_0) و فرض خلاف (H_1) بیان می شود. 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، شیء 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 هستند و 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) : داده عبارتست از نمایش ذخیره شده اشیاء فیزیکی، چیزهای مجرد، بوده ها (واقعیات)، رویدادها یا موجودیهای دیگر قابل مشاهده که در تصمیم سازی بکار می آید.

انواع داده :

1) 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.

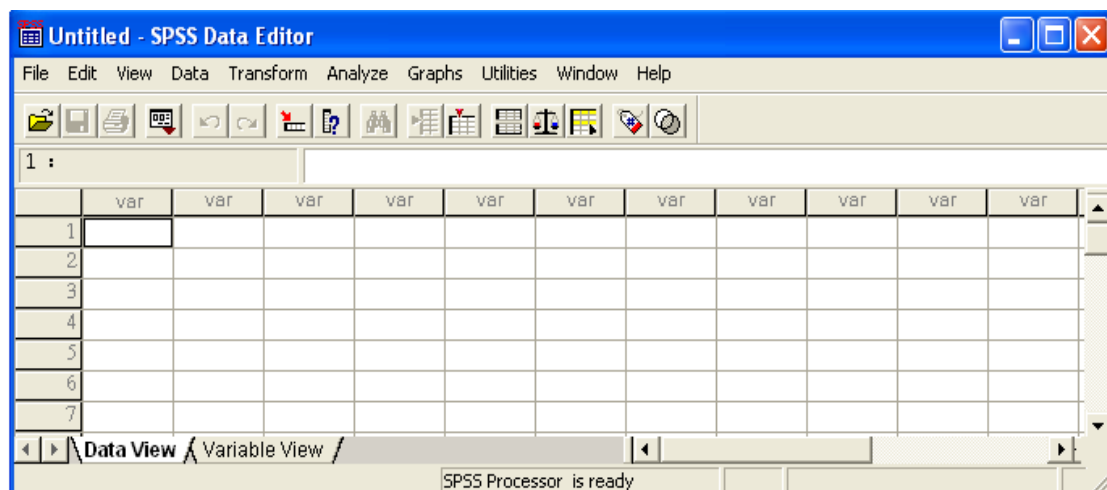


2) داده‌های رتبه‌ای Ordinal مشتمل بر رتبه‌ها، تعلق داشتن به گروه‌های رتبه‌بندی شده یا اطلاعات ترتیبی است. به عنوان مثال اگر دو داور به یک مجموعه 10 تایی از نقاشی رتبه‌بندی (برای بهترین) تا رتبه 10 (برای بدترین) بدهند. مجموعه داده‌ها مشتمل بر 10 جفت رتبه خواهد بود. که هر جفت برای یک نقاشی است.

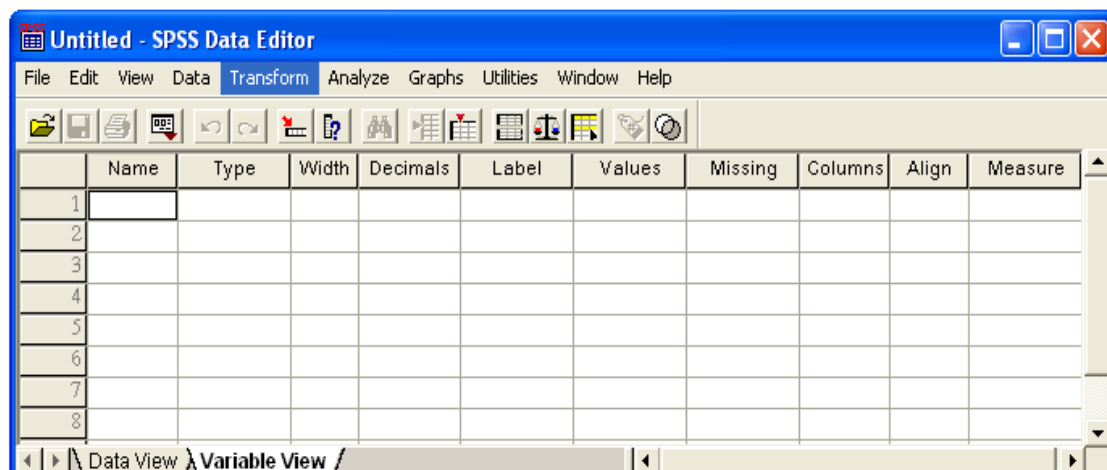
3) داده‌های اسمی (nomial) که مربوط به متغیر یا خواص کیفی مانند جنس یا گروه خونی است و بیانگر عضویت در یک گروه خاص می‌باشد.

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

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



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





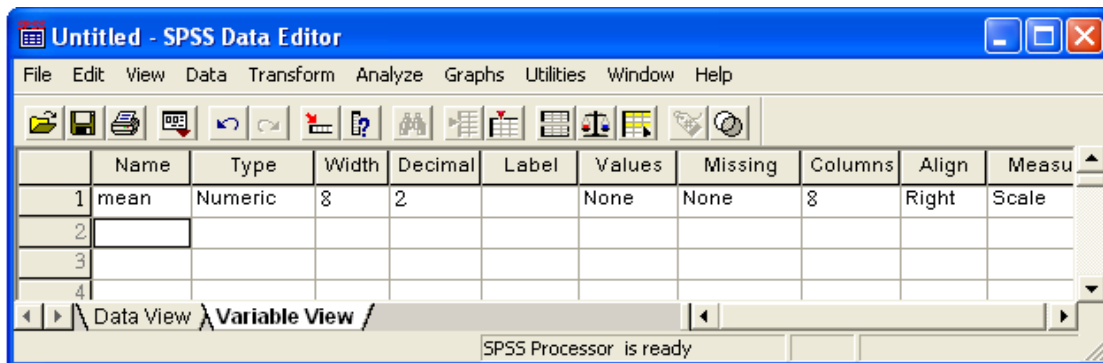
در حالت کلی، روش معمول این است که ابتدا با استفاده از پنجره **Variable view** متغیرها را تعریف کرده و سپس با استفاده از پنجره **Data view**، اعداد را به SPSS وارد کنیم. توضیحات را با ذکر یک مثال ارائه می دهیم.

مثال: معدل دیپلم (کمی) و نوع دیپلم (کیفی) 15 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 (داده ها) را به صورت زیر به نرم افزار وارد می کنیم.

کد 1 = دیپلم ریاضی کد 2 = دیپلم تجربی
 کد 3 = دیپلم انسانی

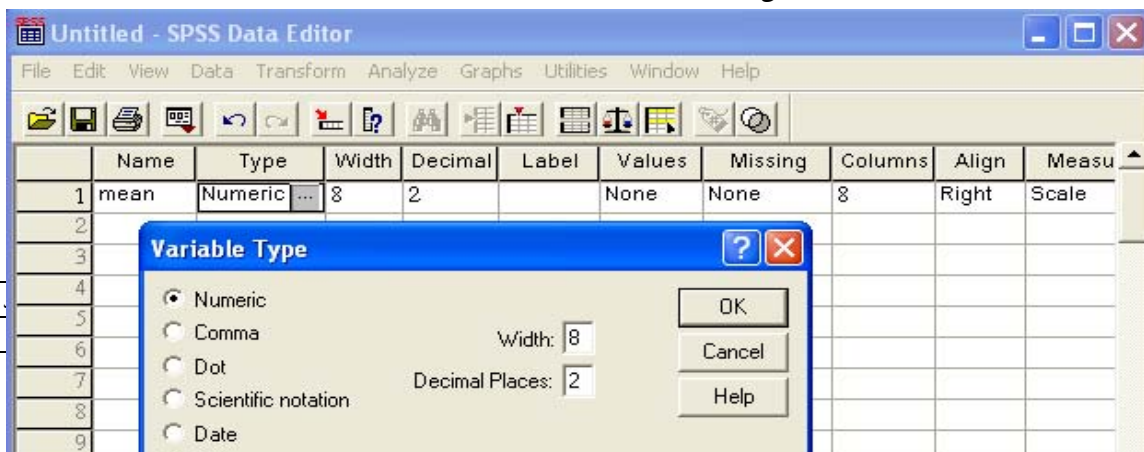
برای این کار بعد از وارد شدن به نرم افزار SPSS به صورت زیر عمل می کنیم:

❖ 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 دخواه 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

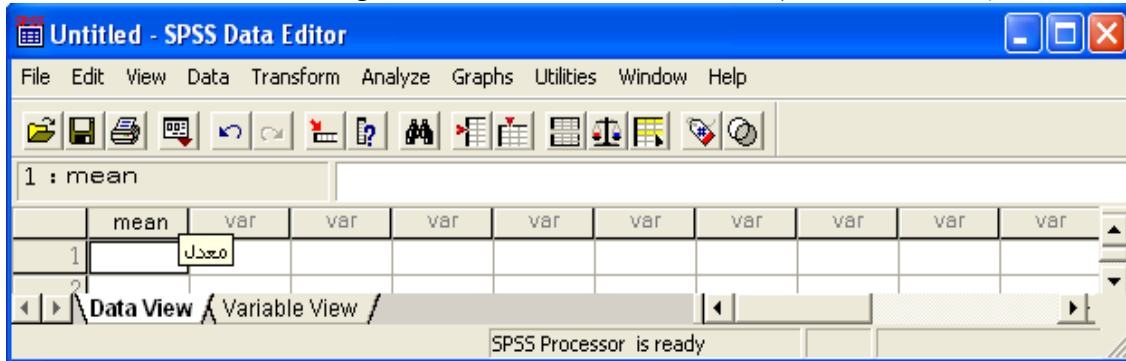


معدل

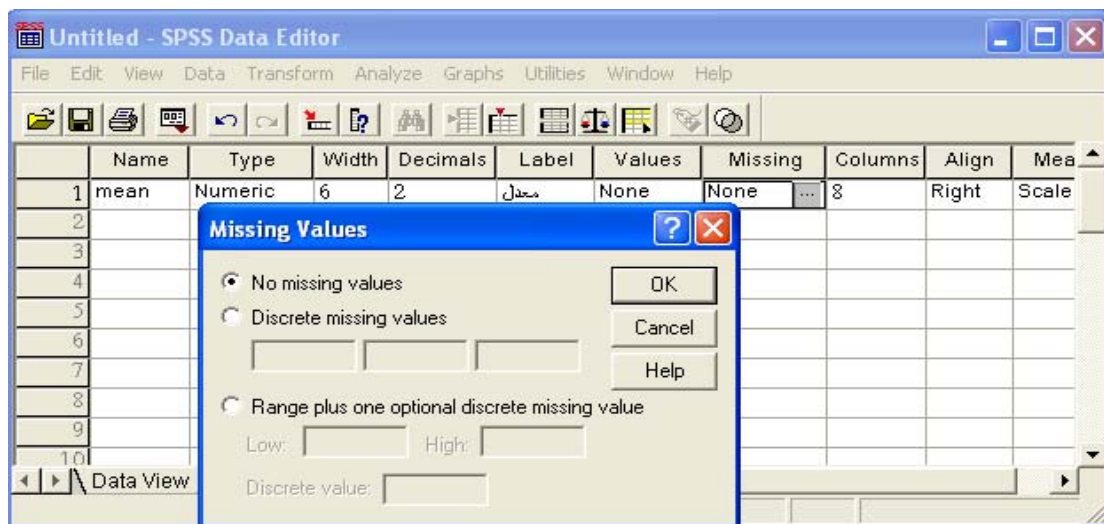
1	2
18	19



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



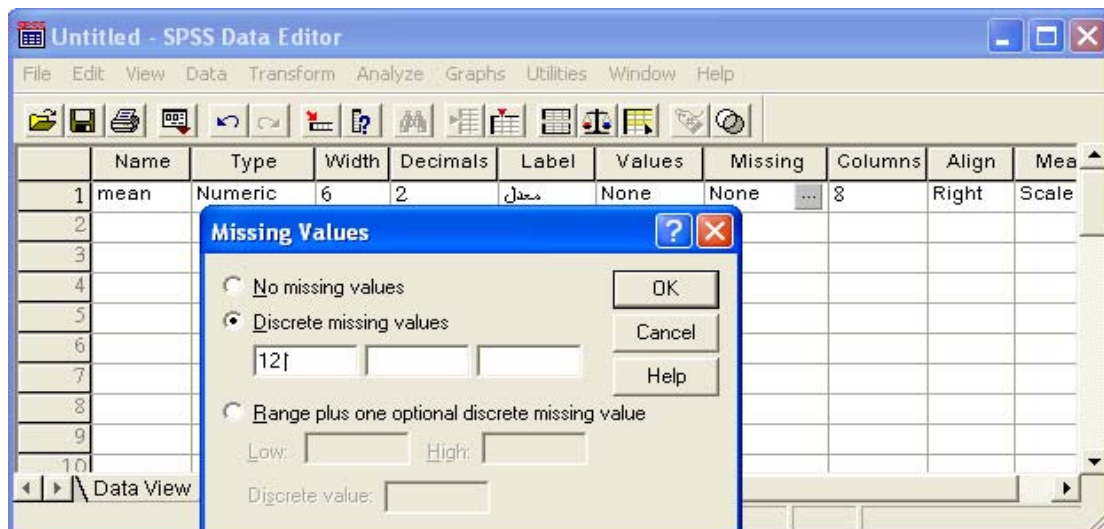
- ❖ ستون Values برای متغیرهای گروه بندی مورد استفاده قرار می گیرد. چون معدل دیپلم دانشجویان متغیر گروه بندی نمی باشد، اطلاعاتی در این قسمت اضافه نمی کنیم.
- ❖ ستون بعدی (Missing) مربوط به داده های گمشده می باشد. در این قسمت با کلیک بر روی ستون مورد نظر پنجره Missing Values باز می شود.



زمانی که در بین اطلاعات جمع آوری شده داده گمشده ای وجود نداشته باشد پیش فرض *No missing values* 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 15 دانشجو در دسترس نباشد، به صورت زیر عمل می کنیم:

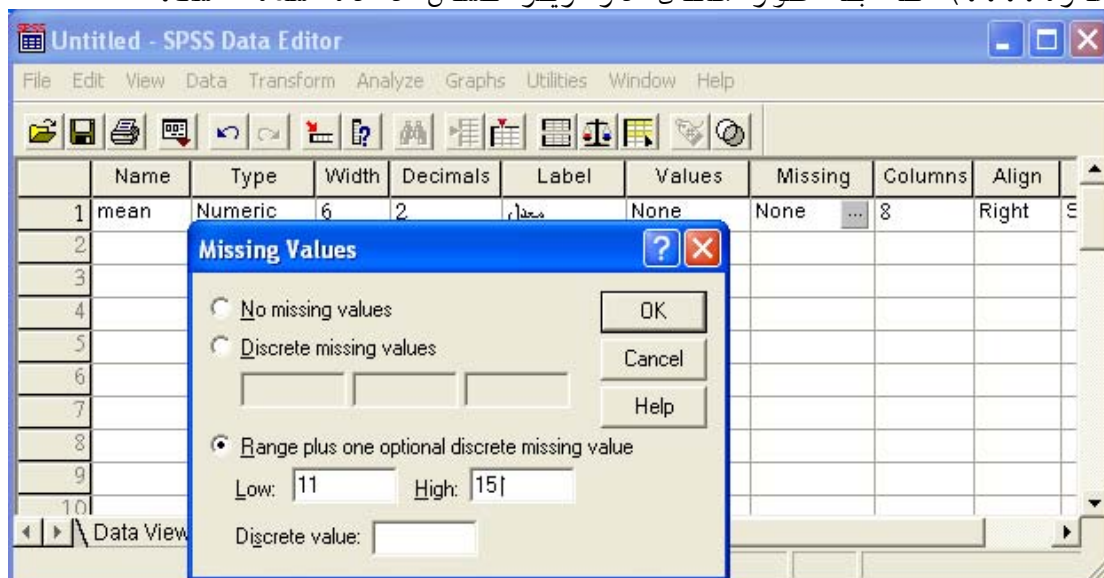


اگر عدد، مربوط به خانه شماره 12 که بیان کننده معدل دانشجوی دوازدهم است، باشد و به بیان دیگر داده گمشده باشد باید شماره 12 را به صورت زیر در مربع مورد نظر وارد کرد. به همین ترتیب اگر داده گمشده دیگری داشتیم، شماره های آنها را در



مستطیلهای بعدی وارد می کنیم. (برای حداکثر 3 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.

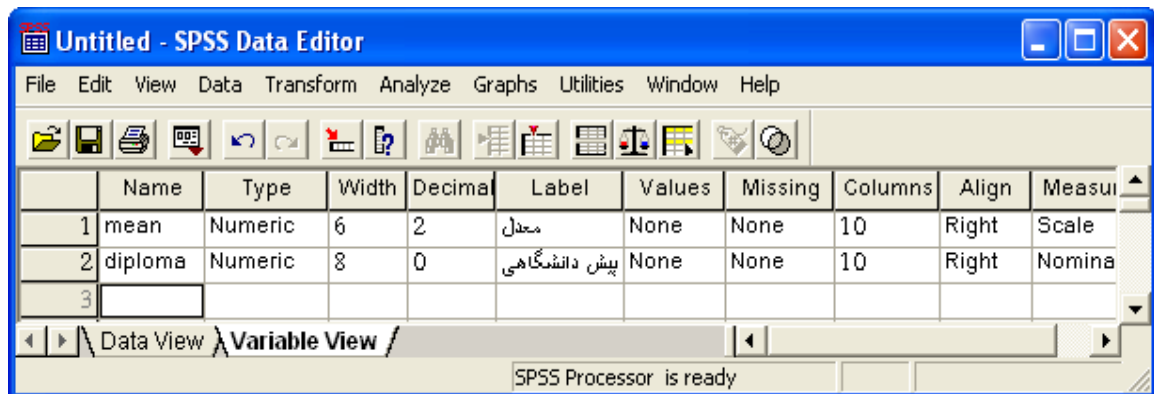


و اگر داده های گمشده به صورتی بودند که یک سری از آنها پشت سر هم و یکی از آنها جدا بود، شماره داده گمشده جدا را در قسمت **Discrete value** اضافه می کنیم.

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

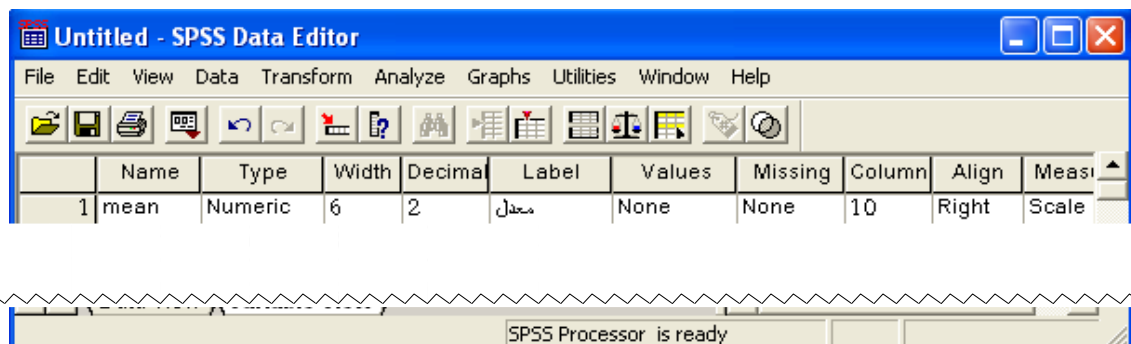


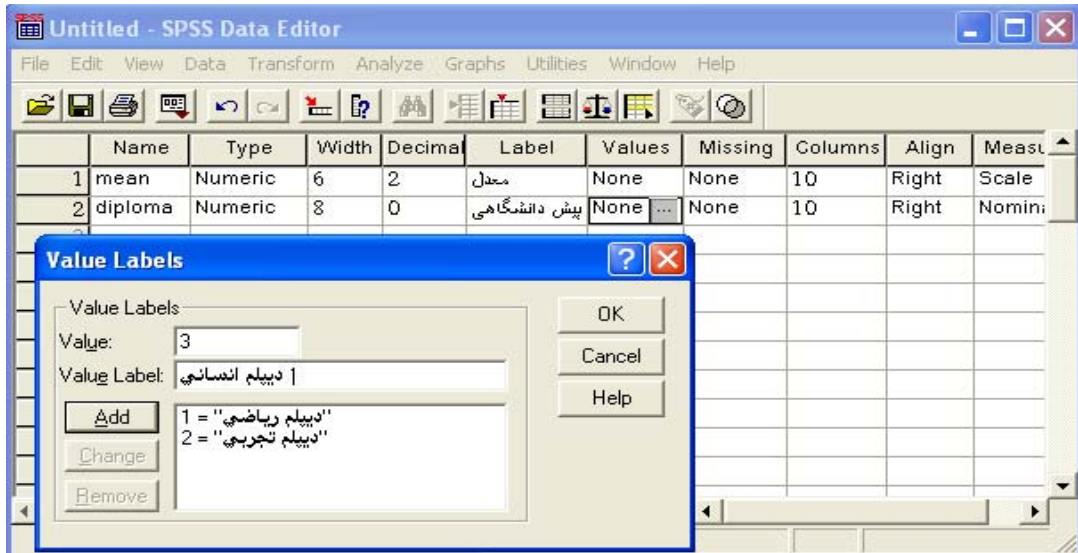
به همین ترتیب متغیر نوع دیپلم را در پنجره **Variable view** تعریف کرده و متناسب با نوع داده، ستونهای مختلف را تنظیم می کنیم.



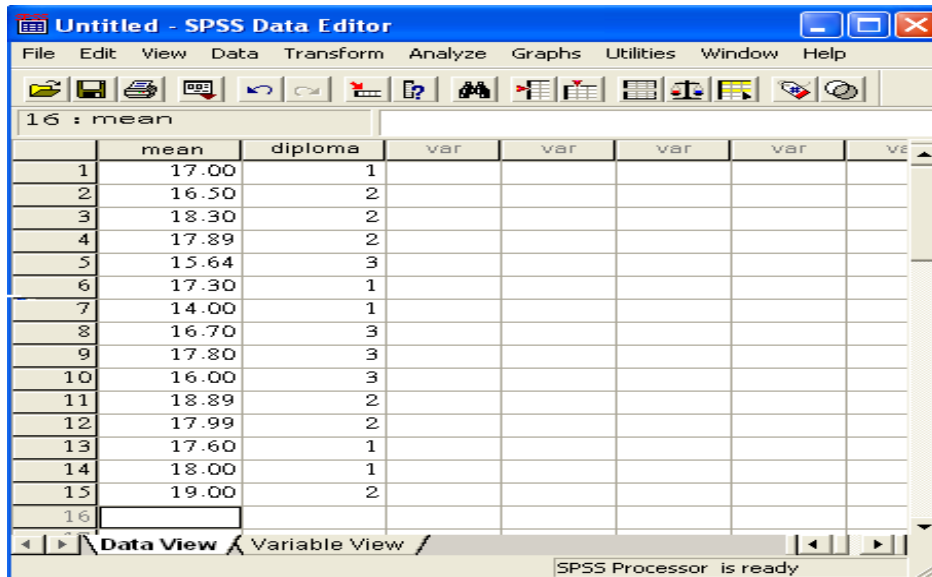
با توجه به اینکه نوع دیپلم داده اسمی می باشد، در این قسمت می توانیم در ستون **Values** هر کدام از کدهای دیپلم را با یک برجسب نشان داد.

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 1 معادل آن یعنی دیپلم ریاضی مشاهده می شود.





سپس در پنجره *Data view* اعداد را به



صورت زیر وارد می نمایم.

استفاده از آزمونهای آماری آزمایشاتی که دارای دو گروه مقایسه هستند را می توانیم بوسیله آزمون t (T Test) مورد تجزیه و تحلیل قرار دهیم (جزوه شماره 1). 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. 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 آزمون، اختلاف میان میانگینهای کلیه تیمارهای موجود در آزمایش، مورد بررسی قرار می گیرد.

هدف از آزمون بررسی زیر می باشد:

$$\left\{ H_0: \mu_1 = \mu_2 = \dots = \mu_k \right.$$



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

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_0 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_0 رد شود نشاندهنده اختلاف میان تیمارها می باشد و باید دنبال اختلافها بگردیم.

مثال: متوسط زمان بستری شدن بیماران در ماه شهریور برای یک بیماری خاص در 5 بیمارستان به صورت زیر می باشد؛ بررسی کنید که آیا میان متوسط زمان بستری شدن (به روز) بیماران 5 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

جدول 1

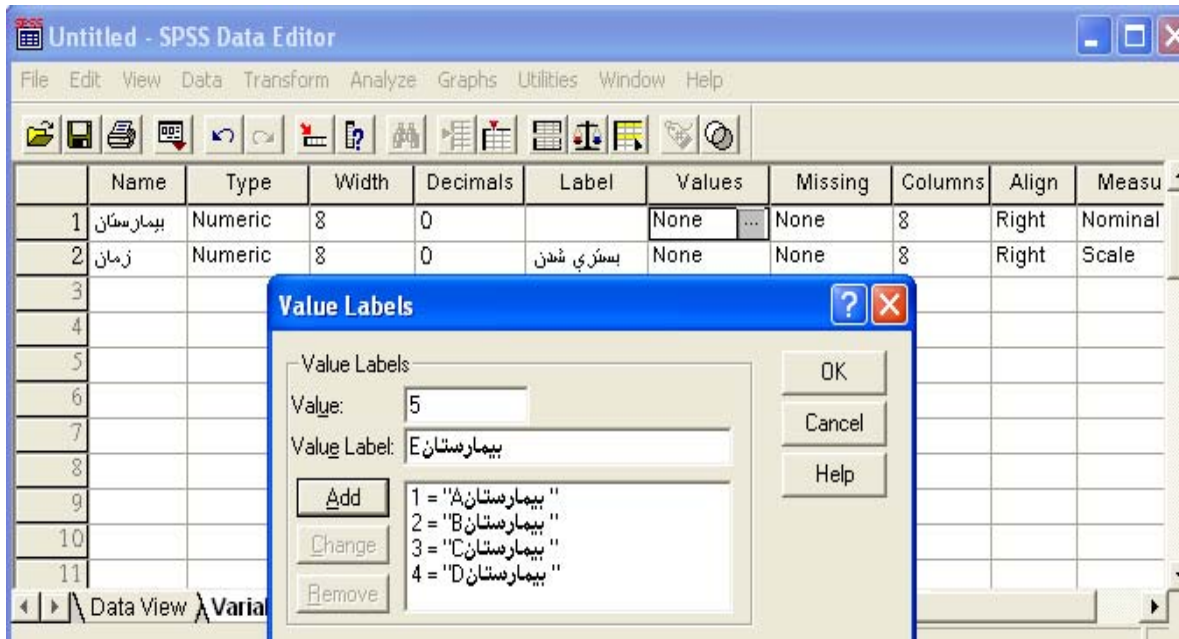
برای انجام آزمون در نرم افزار ابتدا در پنجره **Variable View** دو متغیر به نام نوع بیمارستان و دیگری زمان بستری شدن بیمار تعریف می کنیم و سپس مانند توضیحاتی که در آغاز گفته شد ستونهای مورد نظر را متناسب با نوع متغیر تنظیم می کنیم.

تعداد تیمار	1	2	3	4	5	6	7	8	9	μ_i	\bar{y}_i
بیمارستان A	7	7	8	6	7	5	8	7	6	μ_1	
بیمارستان B	8	8	8	8	7	7	6	6	6	μ_2	
بیمارستان C	7	5	5	5	4	7	4	4	5	μ_3	
بیمارستان D	8	9	9	11	6	10	11	11	10	μ_4	
بیمارستان E	4	9	6	4	4	4	5	5	4	μ_5	

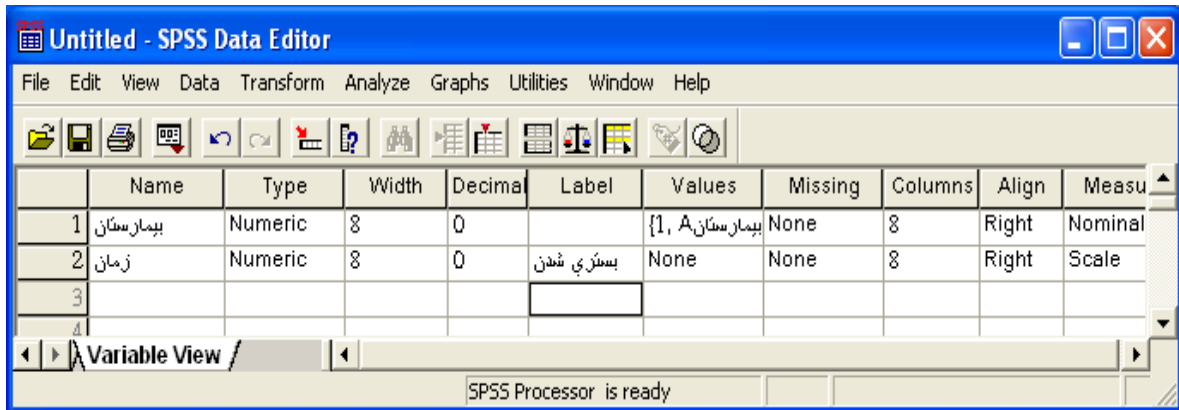
در این قسمت برای متغیر نوع بیمارستان در ستون **Values** نوع بیمارستانها را تعریف می کنیم.



در پنجره **Data View** به جای اسامی بیمارستان از کدهای 1 و 2 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 (1-5) وارد کرده و جلوی هر کد در ستون زمان، مدت زمان بستری شدن بیمارانی بیمارستانهای مختلف را طبق جدول 1 وارد می کنیم.



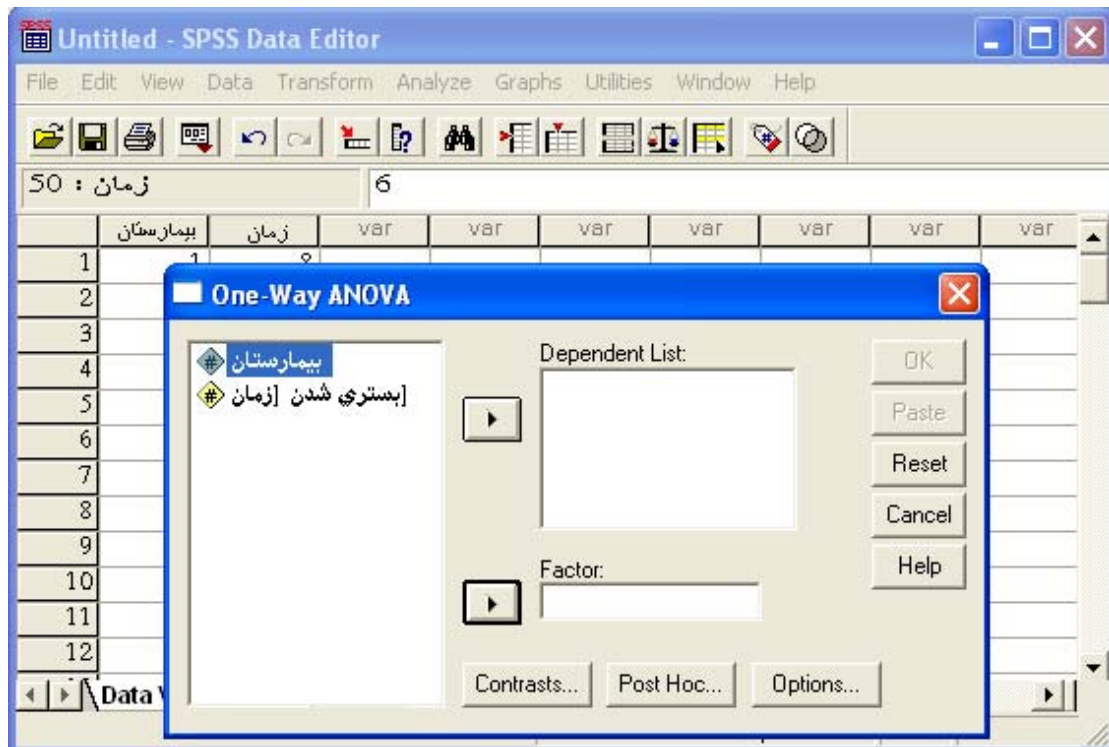
	بیمارستان	زمان	var	var	var	var	var	var	var
1	1	7							
2	1	7							
3	1	8							
4	1	6							
5	1	7							
6	1	5							
7	1	8							
8	1	7							
9	1	6							
10	1	8							
11	2	8							
12	2	7							
13	2	7							

برای انجام آزمون به صورت زیر عمل می کنیم:

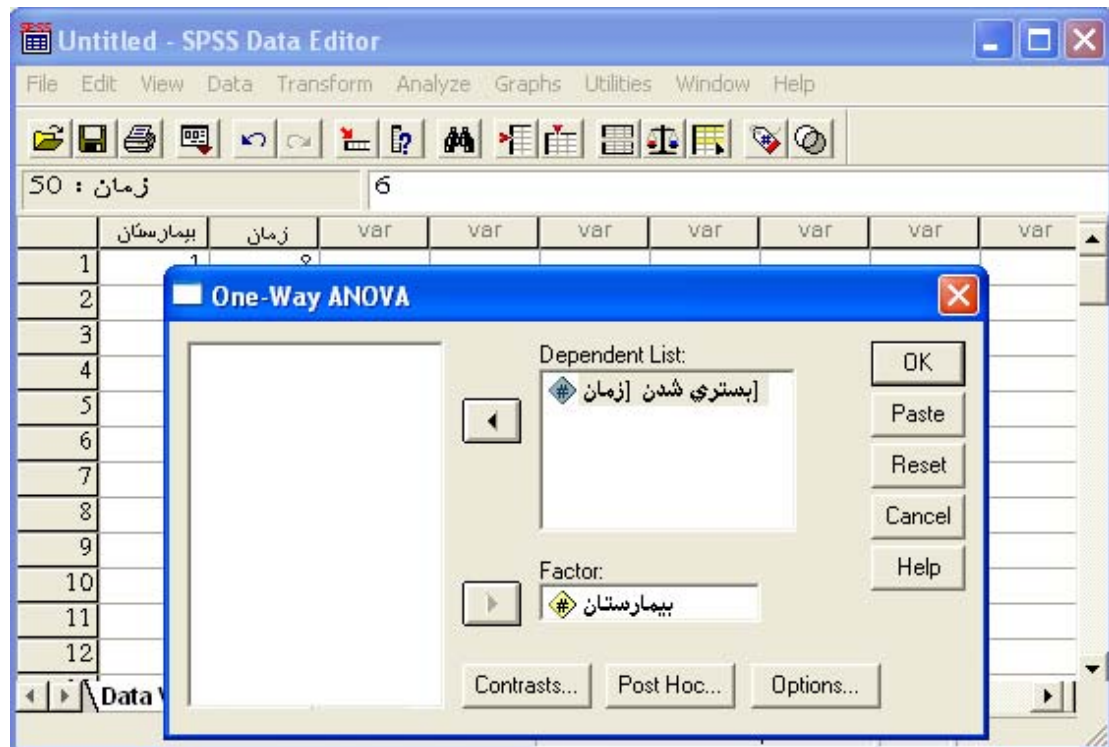
7 :

	بیمارستان	زمان	va
1	1	7	
2	1	7	
3	1	8	
4	1	6	
5	1	7	
6	1	5	
7	1	8	
8	1	7	
9	1	6	
10	1	8	
11	2	8	
12	2	7	

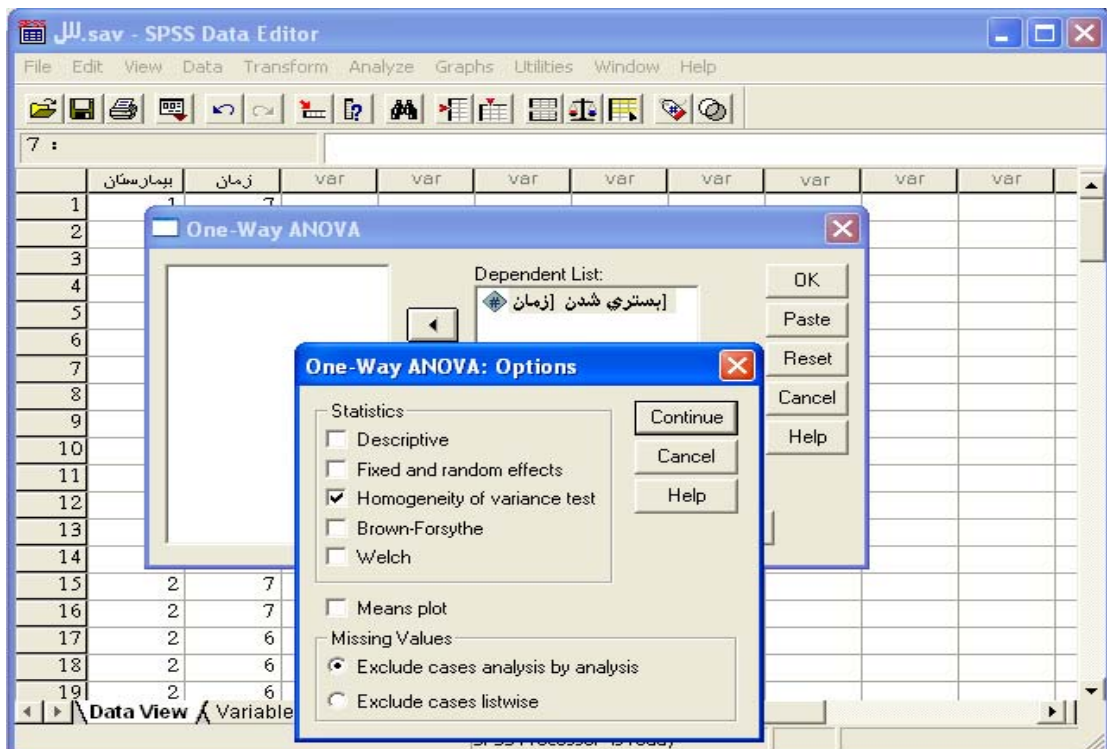
- Reports
- Descriptive Statistics
- Tables
- Compare Means**
 - Means...
 - One-Sample T Test...
 - Independent-Samples T Test...
 - Paired-Samples T Test...
 - One-Way ANOVA...**
- General Linear Model
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Data Reduction
- Scale
- Nonparametric Tests
- Time Series
- Survival
- Multiple Response
- Missing Value Analysis...
- Complex Samples



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



قبل از اینکه بر روی کلمه **ok** کلیک کنیم برای بررسی اینکه آیا بین واریانس های (مدت زمان بستری شدن بیماران) 5 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 فعال کرده ، در ادامه ابتدا بر روی کلمه **Continue** و سپس **Ok** کلیک می کنیم تا خروجیهای زیر بدست آیند.

آزمون همگنی واریانسها (جدول 2)

Test of Homogeneity of Variances ANOVA

	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			

آزمون برابری میانگینها (جدول 3)

نتایج بدست آمده از جدول 2 نشان می دهد که در آزمون مقایسه بین واریانسهای **مدت زمان بستری شدن** 5 بیمارستان اختلاف معنی داری از نظر آماری وجود ندارد ($P\text{-value} > 0.05$). اما نتایج بدست آمده از جدول 3 نشان می دهد که میان میانگینهای **مدت زمان بستری شدن** 5 بیمارستان اختلاف معنی داری وجود دارد. ($P\text{-value} < 0.05$).



به دلیل اینکه داده ها نشان می دهند که میانگینهای 5 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:

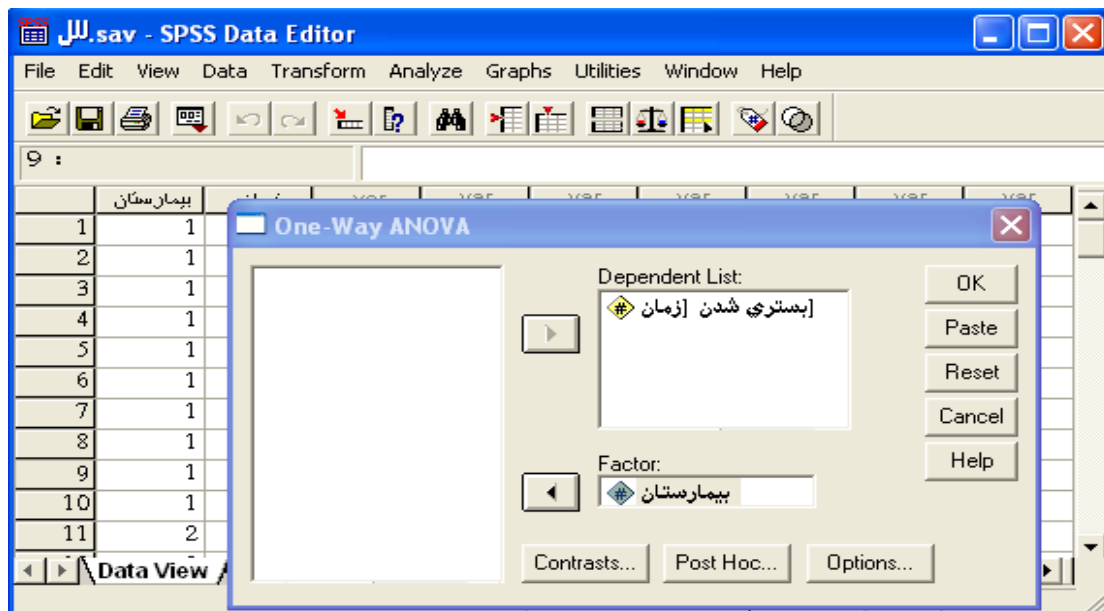
(Equal Variances Assumed)

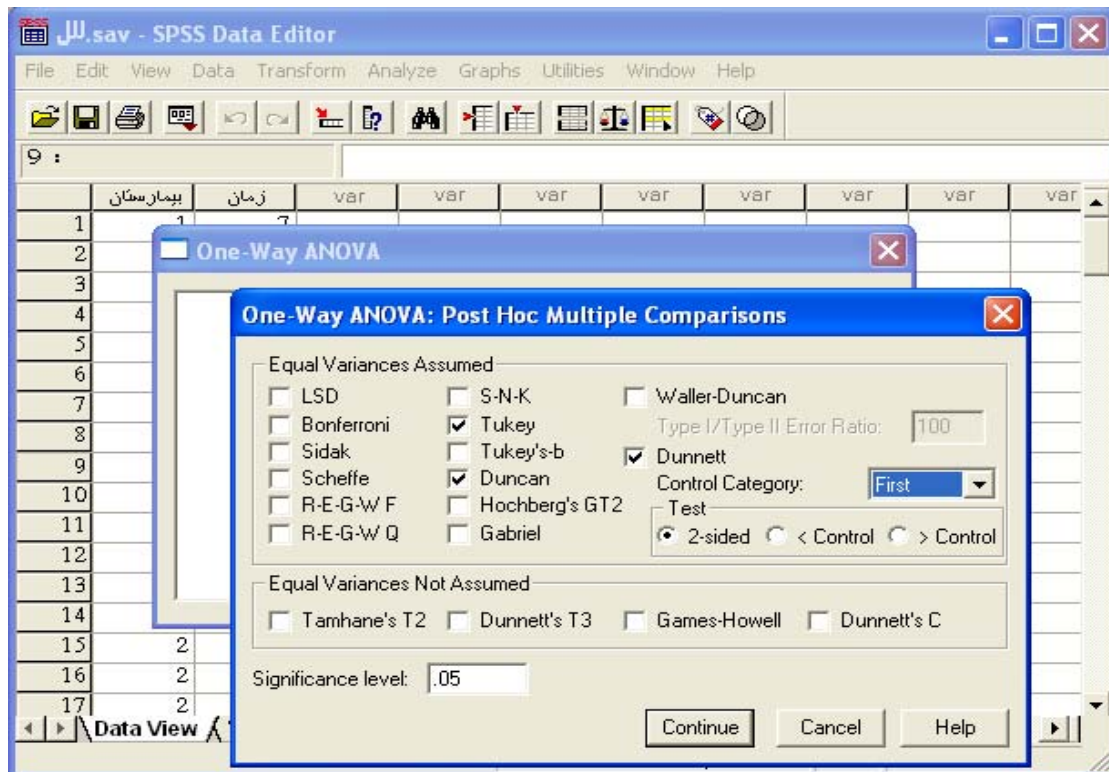
قسمت پایین مربوط به آزمونهای مورد استفاده در حالتی که واریانس جوامع متفاوت باشند:

(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 (*Duncan*، *Tukey*، *Dunnnett*) را برای تجزیه و تحلیل آماری انتخاب می کنیم.

در مورد آزمون *Dunnnett* این نکته را باید مورد توجه قرار داد که از میان یکی از گروهها (5 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 *Dunnnett* 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:

Dunnnett t (2-sided)^a

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
B	A	.000	.603	1.000	-1.53	1.53
C	A	-1.800*	.603	.016	-3.33	-.27
D	A	2.800*	.603	.000	1.27	4.33
E	A	-1.800*	.603	.016	-3.33	-.27

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

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

Multiple Comparisons

Dependent Variable:

	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	A	B	.000	.603	1.000	-1.71	1.71
		C	1.800*	.603	.035	.09	3.51
		D	-2.800*	.603	.000	-4.51	-1.09
		E	1.800*	.603	.035	.09	3.51
	B	A	.000	.603	1.000	-1.71	1.71
		C	1.800*	.603	.035	.09	3.51
		D	-2.800*	.603	.000	-4.51	-1.09
		E	1.800*	.603	.035	.09	3.51
	C	A	-1.800*	.603	.035	-3.51	-.09
		B	-1.800*	.603	.035	-3.51	-.09
		D	-4.600*	.603	.000	-6.31	-2.89
		E	.000	.603	1.000	-1.71	1.71
	D	A	2.800*	.603	.000	1.09	4.51
		B	2.800*	.603	.000	1.09	4.51
		C	4.600*	.603	.000	2.89	6.31
		E	4.600*	.603	.000	2.89	6.31
	E	A	-1.800*	.603	.035	-3.51	-.09
		B	-1.800*	.603	.035	-3.51	-.09
		C	.000	.603	1.000	-1.71	1.71
		D	-4.600*	.603	.000	-6.31	-2.89

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

Homogeneous Subsets

	N	Subset for alpha = .05			
		1	2	3	
Tukey HSD ^a	C	10	5.10		
	E	10	5.10		
	A	10		6.90	
	B	10		6.90	
	D	10			9.70
	Sig.		1.000	1.000	1.000
Duncan ^a	C	10	5.10		
	E	10	5.10		
	A	10		6.90	
	B	10		6.90	
	D	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-Value** بدست آمده در آزمونها می باشد و چون آزمونها در سطح **0.05** مورد بررسی قرار می گیرند مبنای پذیرش یا عدم پذیرش آنها یعنی قبول یا رد فرض اولیه H_0 مقایسه با مقدار **0.05** می باشد.

در آزمونهای یک دامنه اگر $Sig < 0.05$ فرض اولیه H_0 رد می شود و اگر $Sig > 0.05$ فرض اولیه H_0 رد نمی شود. اما در آزمونهای دو دامنه به جای **0.05** از **0.025** استفاده می شود.

هم ارز با ستون **Sig** ستون مربوط به فاصله اطمینان (**95% Confidence Interval**) است که نمایش دهنده یک فاصله اطمینان **95 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 **0.05** باشد فاصله اطمینان **0.95** می شود.

این فاصله اطمینان یک حد پایین (**L**) و یک حد بالا (**U**) دارد و اگر عدد صفر را شامل شود ($L < 0 < U$) نشاندهنده این است که فرض H_0 یعنی برابری میانگینها رد نمی شود و این هم ارز $Sig > 0.05$ می باشد و اگر این بازه شامل صفر نباشد هم ارز این است که $Sig < 0.05$ و بیان می کند که فرض H_0 یعنی برابری میانگینها رد می شود.

در خروجی مربوط به آزمون **Dunnett** مشاهده می کنیم: 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-Value** را برای هر آزمون جداگانه نشان می دهد، هر جا $Sig < 0.025$ باشد نشان می دهد 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 $Sig < 0.025$ به جای $Sig < 0.05$ در این است که آزمون **Dunnett** یک آزمون دو دامنه است (**2-Side**).

نتایج نشان می دهند که میانگین بیمارستان **A** با بیمارستان **B** تفاوت ندارد ($Sig > 0.025$). اما میان میانگین بیمارستان **A** با میانگین سایر بیمارستانها تفاوت معنی داری وجود دارد ($Sig < 0.025$).

در خروجی مربوط به آزمون **Tukey** تک تک بیمارستانها با هم مقایسه می شوند. در این آزمون به علت یک دامنه بودن هر کجا $Sig < 0.05$ فرض برابری میانگینها رد می شود (به بیان دیگر بین



میانگین مدت زمان بستری شدن دو بیمارستان از نظر آماری تفاوت معنی داری 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 می دهند. به طور معمول نتایج بدست آمده از آزمون ها 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.

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:

1- به واحد دو جامعه وابسته نباشد

2- کراندار باشد

به طور مثال در تحولات اقتصادی 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_n, Y_n) ، بیان کرد.

مطالعه رابطه بین متغیرها بوسیله ((تحلیل همبستگی)) (Analysis Corroletion) 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 حتهای مختلف تفسیرهای گوناگونی از رابطه X و Y خواهیم داشت.

$$-1 < r < 1$$

حالتهای مختلف برای r

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



2- $r = -1$ در این حالت همبستگی را کامل و معکوس گوییم. با افزایش مقدار X مقدار Y کاهش می یابد.

3- $-1 < r < 0$ همبستگی ناقص و معکوس است. با افزایش مقدار X مقدار Y به طور نسبی کاهش می یابد.

4- $0 < r < 1$ همبستگی ناقص و مستقیم است. با افزایش مقدار X مقدار Y به طور نسبی افزایش می یابد.

5- $r = 0$ رابطه خطی وجود ندارد. (به طور مثال رابطه ممکن است از نوع درجه دو باشد)
 -2 شیب خط صفر می باشد.

انواع ضریب همبستگی با توجه به نوع متغیرهای مورد مطالعه

1- 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.)

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

3- اسپرمن: 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 ضریب 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 اسپرمن استفاده کنیم.

در آزمونهای همبستگی یک طرفه فرضهای زیر بررسی می شوند.

$$1 \left\{ \begin{array}{l} H_0: r = 0 \\ H_1: r < 0 \end{array} \right. \quad 2 \left\{ \begin{array}{l} H_0: r = 0 \\ H_1: r > 0 \end{array} \right.$$



$$H_1: r > 0$$

$$H_1: r < 0$$

جهت نامساوی با توجه به برآورد نمونه ای r تعیین می شوند. اگر r مثبت باشد آزمون 1 و اگر منفی باشد از آزمون 2 استفاده می کنیم.

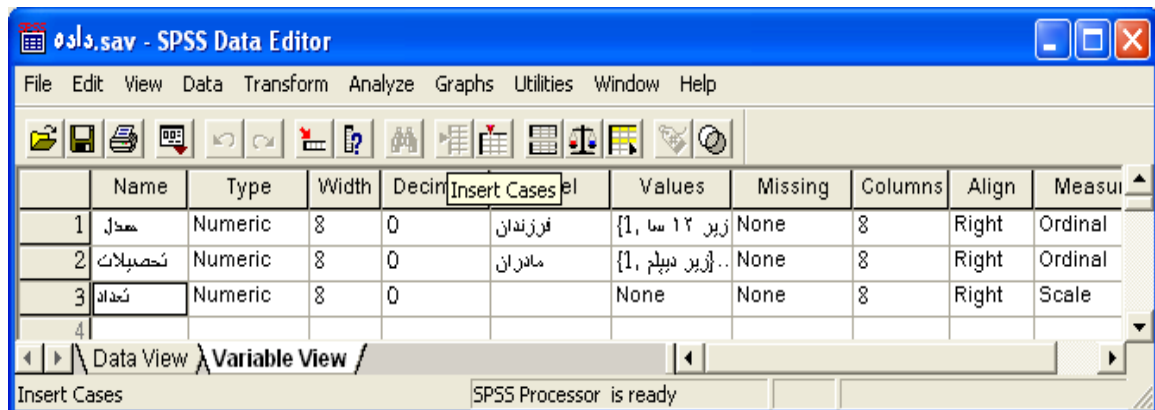
مثال

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	(1) زیردیپلم	(2) دیپلم	(3) فوق دیپلم	(4) لیسانس وبالاتر
(1) زیر 12	10	7	6	2
(2) 12-15	12	7	5	5
(3) 15-17	5	4	6	10
(4) 17-20	5	7	10	12

وقتی در نرم افزار Spss داده ها را به صورت رتبه ای مشخص کردیم می توانیم در تجزیه و تحلیلها تفسیر راحتتری 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



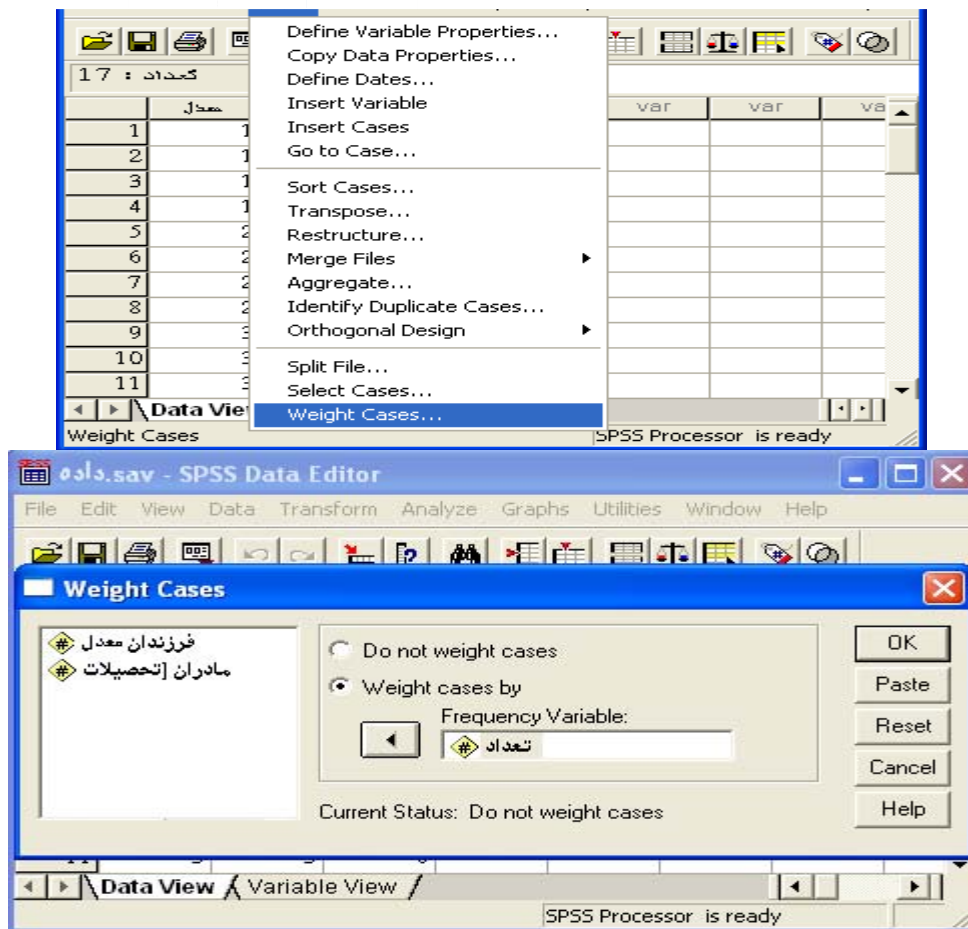
بالا داده شده است متغیرها را معرفی می کنیم.



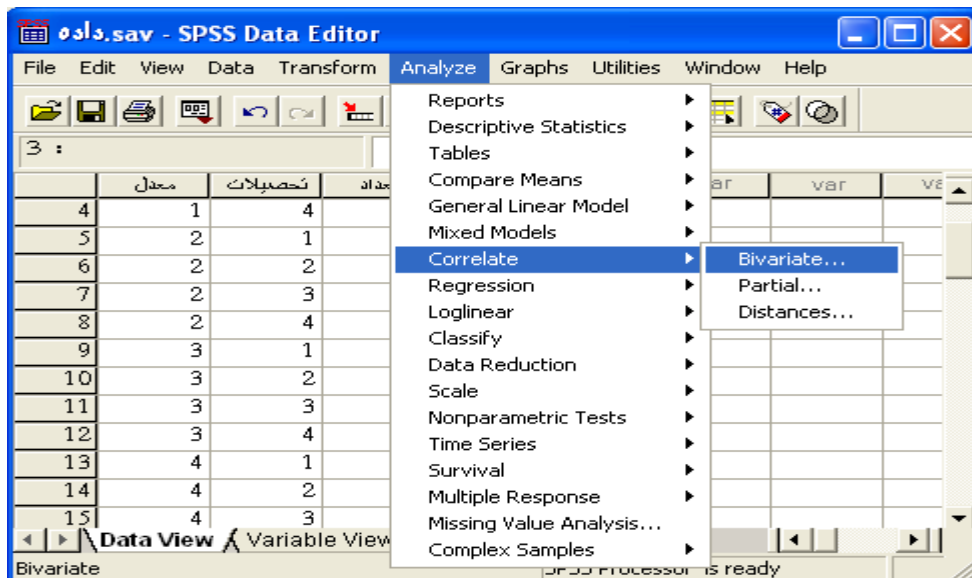
سپس داده را به صورت زیر در پنجره *Data View* وارد می کنیم

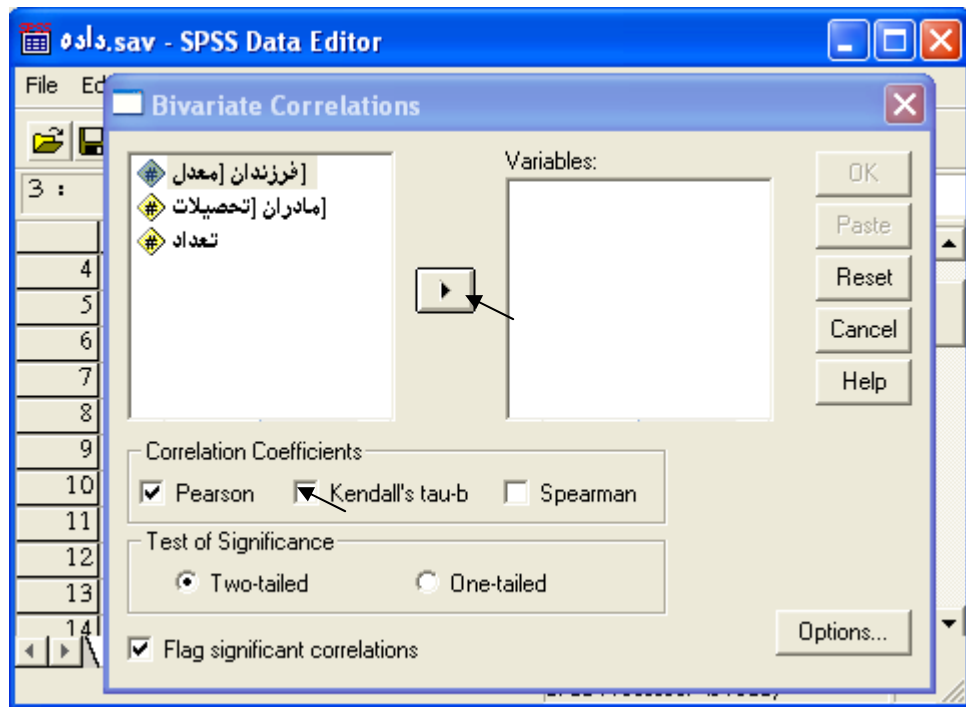
	حداکثر	تحصیلات	تعداد	var	var	var	va
1	1	1	10				
2	1	2	7				
3	1	3	6				
4	1	4	2				
5	2	1	12				
6	2	2	7				
7	2	3	5				
8	2	4	5				
9	3	1	5				
10	3	2	4				
11	3	3	6				

برای اینکه مشخص کنیم ستون تعداد معرف فراوانی هایی هست که به دو متغیر سن و تحصیلات نسبت داده شده است 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 وزند 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

Kendall's tau_b	Correlation Coefficient	1.000	.260*
	Sig. (2-tailed)	.	.001
	N	113	113
	Correlation Coefficient	.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 0.001 می باشد که با توجه به دو دامنه بودن آزمون (*2-tailed*) با 0.025 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 1 به سمت کد 4) معدل دانش آموزان نیز (با رفتن از کد 1 به سمت کد 4) افزایش پیدا می کند و



نشان دهنده نقش موثر و مثبت تحصیل مادر بر بالا بردن کیفیت تحصیلی فرزندان 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 اسپرمن استفاده می کردیم.