

راهنمای مقدماتی SPSS

(قسمت اول - آمار توصیفی)

تهیه و تنظیم: مهدی اکبرزاده

با همکاری: سرکار خانم کرامتی

- سرکار خانم وحدت

مقدمه

با پیشرفت علوم و گسترش تکنولوژی، اهمیت استفاده از روش‌های آماری در علوم مختلف بیش از پیش مورد توجه قرار گرفته است و آموختن آمار کاربردی در هر رشته جزء ملزومات گردیده است. یک سری داده‌های آماری که در اولین وهله مطالعه بدست می‌آید، به شکلی است که برای یک تعبیر و تفسیر فوری غیرقابل استفاده است، هر چند که داده‌ها به صورت ساده‌ای باشند.

فرآیند آنالیز آماری کمک می‌کند تا پژوهشگر بتواند از داده‌های اولیه، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد.

اگر حجم داده‌ها بزرگ باشد، استفاده از روش‌های مختلف آنالیز آماری بسیار خسته کننده و مشکل خواهد بود، امروزه انواع نرم‌افزارهای مختلف آماری موجودند که قادرند انواع آنالیزهای آماری را انجام دهند، برخی از مهم‌ترین این نرم‌افزارها عبارتند از:

SPSS (۱)

S-PLUSE (۲)

MINITAB (۳)

SAS (۴)

در بین برنامه‌های فوق کامل‌ترین و دقیق‌ترین برنامه آماری، برنامه SAS می‌باشد. البته هر یک از برنامه‌های آماری فوق به دلایلی، در یک رشته به‌خصوص عمومی شده‌اند و هر کاربر با توجه به کاربرد آمار در رشته تحصیلی خود یکی را انتخاب می‌کند. برای مثال محققین در رشته کشاورزی از برنامه SAS بیشتر بهره می‌برند و یا افرادی که در رشته‌های مهندسی صنایع در حال تحقیق و پژوهش هستند از برنامه MINITAB برای انجام پروژه‌های صنعتی خود استفاده می‌کنند و ...

اما چرا SPSS؟

SPSS^۱ یکی از تواناترین و جامع‌ترین نرم‌افزارهای آماری است که با توجه به **سادگی کار** و سایر خصوصیات بارز آن امروزه پرکاربردترین نرم‌افزار آماری محسوب می‌شود. همان‌طور که از نام این نرم‌افزار مشخص است، از ابتدا این نرم‌افزار برای رشته علوم اجتماعی طراحی شده بود، اما به مرور زمان که نرم‌افزار بروز شده و با توجه به نیازهای مطرح‌شده در سایر علوم این نرم‌افزار کامل و کامل‌تر شده است و در حال حاضر کلیه محققین در رشته‌های علوم اجتماعی، علوم پزشکی، علوم تربیتی، روانشناسی، کشاورزی و ... در حال استفاده از این نرم‌افزار هستند و در این لحظه که شما تصمیم به فراگیری این نرم‌افزار آماری گرفته اید، تا ویرایش ۱۸ این نرم‌افزار در کشورمان به‌طور کامل موجود بوده و ما نیز با همین ویرایش کار خواهیم کرد.

آمار هنر بیرون کشیدن اطلاعات نهفته از داده های انبوه و سرگیجه آورست. آنچه مهم و مفید است اطلاعات است، اطلاعات همان داده های خام پردازش شده اند. اگر فکر می کنید که از داده ها به تنهایی کاری بر می آید، سخت در اشتباهید.

مهدی اکبرزاده

تابستان ۸۹

www.irmed.ir

تعریف متغیر و انواع آن

قبل از اینکه وارد مباحث مربوط به نرم افزار SPSS شویم بهتر است ابتدا با چند مفهوم از روش های تحقیق آشنا شویم. حواسمان باشد که با تمام این مفاهیم بعداً سروکار خواهیم داشت و باید این مفاهیم در ذهن ما کاملاً تداعی شده باشند.

متغیر: خصوصیتی از افراد مورد بررسی در جامعه گویند.

برای مثال "معدل دانشجویان دانشگاه علوم پزشکی سبزوار". در این مثال **جامعه آماری** موردنظر دانشجویان دانشگاه علوم پزشکی سبزوار بوده و معدل آنها به عنوان خصوصیتی از این جامعه **متغیر** مورد بررسی می باشد.

مثال هایی دیگر:


- ۱- تعداد کتاب های امانت گرفته شده از کتابخانه دانشگاه، **جامعه:** اعضاء کتابخانه. **متغیر:** تعداد کتاب امانت گرفته شده توسط هر عضو.
- ۲- نمره مربوط به EFQM کتابخانه های شهر سبزوار. **جامعه:** کتابخانه های شهر سبزوار. **متغیر:** نمره EFQM
- ۳- تعداد واحدهای گذرانده توسط دانشجویان دانشگاه. **جامعه:** دانشجویان **متغیر:** تعداد واحدهای گذرانده

به عنوان تمرین: مثال های دیگری از خودتان بنویسید و در آنها جامعه و متغیر را شناسایی کنید. 

انواع متغیرها:

- ۱- کمی (Scale):
 - پیوسته: (هر چیزی که قابل اندازه گیری است). مثل قد، وزن و ...
 - گسسته: هر چیزی که قابل شمارش باشد. مثل تعداد گلبول های قرمز. این نوع متغیرها در عمل به دو گونه ظاهر می شوند. ۱- با دامنه تغییرات زیاد ۲- با دامنه تغییرات کم

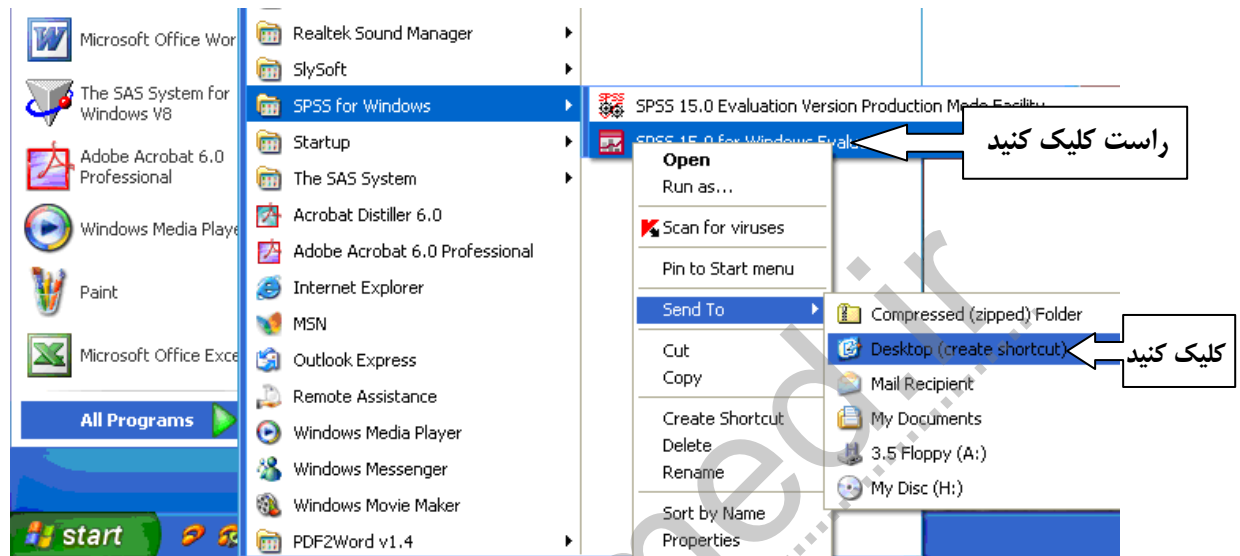
- ۲- کیفی (qualitative)
 - اسمی (Nominal): سطوح نسبت به هم ارجحیت ندارند مانند جنسیت یا گروه خونی.
 - رتبه ای یا ترتیبی (Ordinal): سطوح نسبت به هم ارجحیت یا ترتیب دارند. مانند میزان تحصیلات (سیکل - دیپلم - فوق دیپلم ...)، نمره شدت درد (صفر، یک، دو، و ...)
- دارای چند سطح (level) می باشند

به عنوان تمرین: در هر یک از مثال‌های زیر نوع متغیر مورد بررسی را مشخص کنید. 

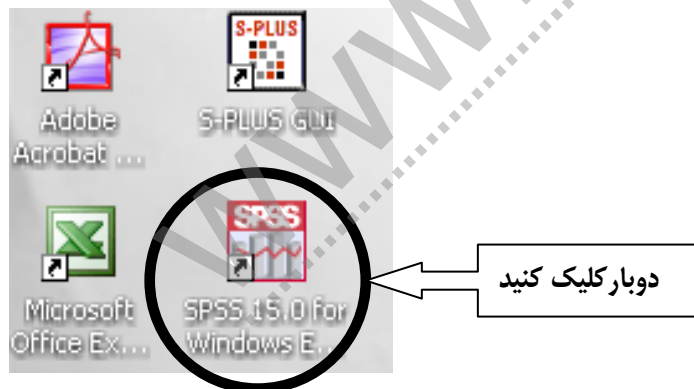
- ۱- سن
- ۲- جنسیت
- ۳- میزان تحصیلات
- ۴- گروه خونی
- ۵- سیگاری بودن یا نبودن
- ۶- تعداد دندان‌های شیری
- ۷- میزان درآمد
- ۸- میزان آگاهی دانشجویان از مقررات آموزشی
- ۹- رفتار اطلاع‌یابی دانشجویان از کتابخانه
- ۱۰- تعداد دانشجویان دانشگاه
- ۱۱- تعداد دانشجویان مردودی در یک رشته
- ۱۲- نوع ژن (موتاسیون)
- ۱۳- بیمار بودن یا نبودن فرد

شروع کار با SPSS و اعمال مقدماتی

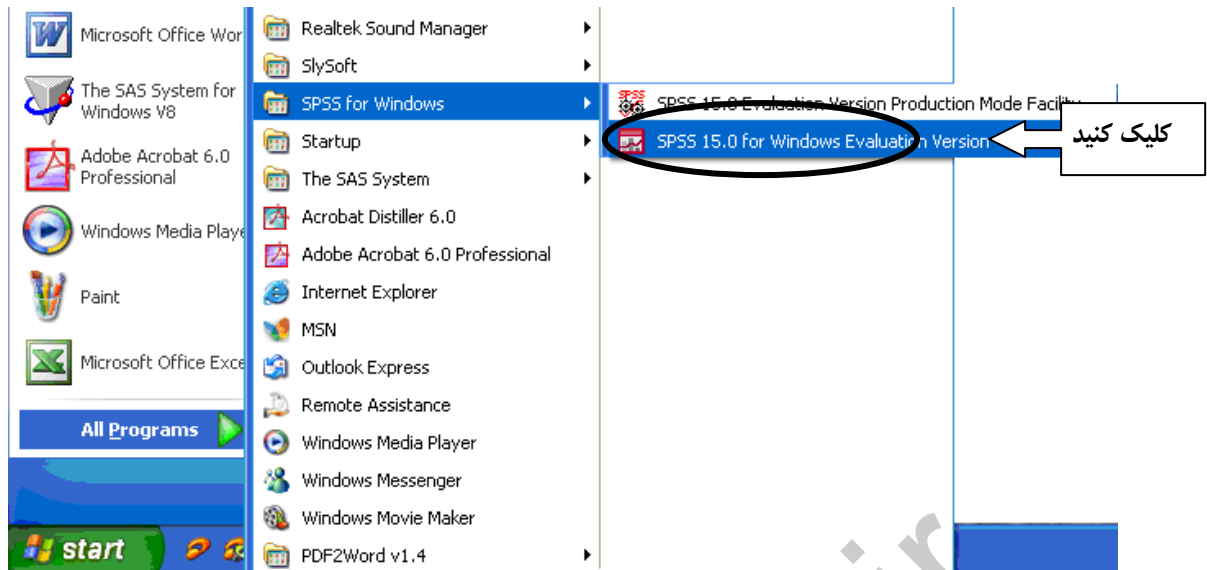
قبل از اینکه کار با این نرم افزار را شروع کنیم لازم است که ابتدا نحوه نصب این نرم افزار را فرا بگیریم. پس از اینکه نرم افزار را بر روی کامپیوتر نصب کردید، برای راحتی کار به طریق زیر یک میان‌بر از آنرا بر روی Desktop را ایجاد کنید:



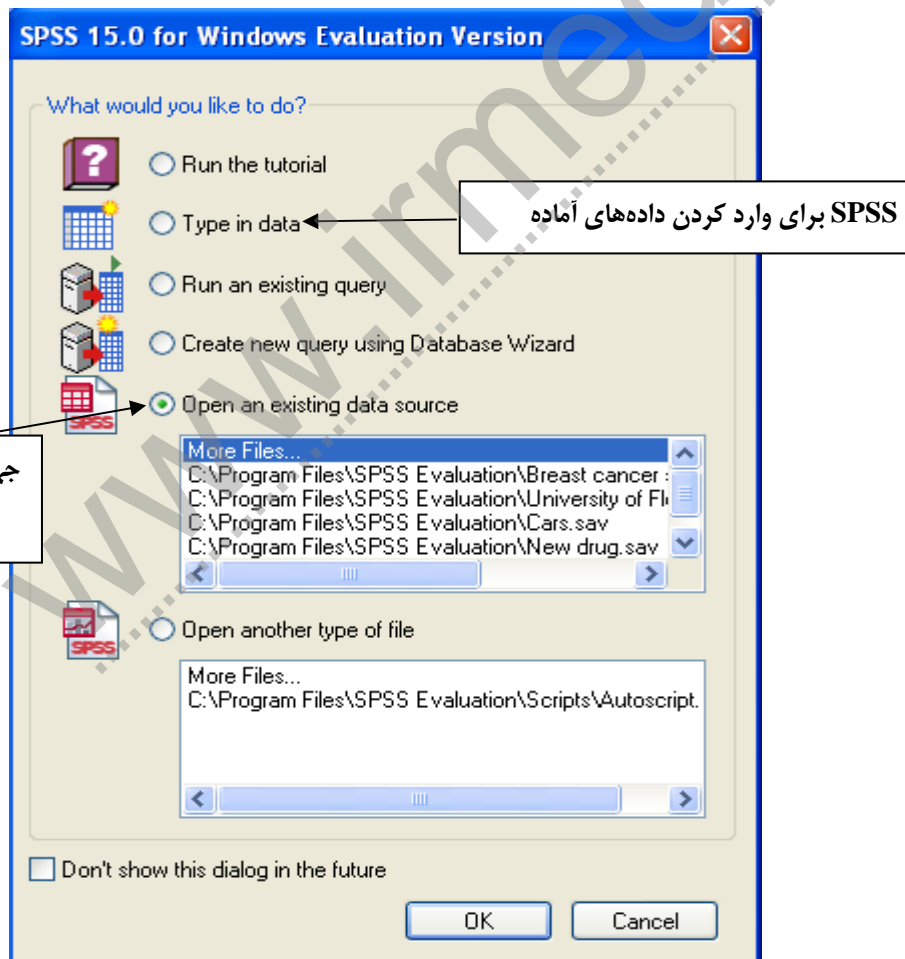
حال برای ورود به محیط نرم افزار SPSS در صورت وجود آیکون نرم افزار SPSS در Desktop روی آن Double-click کنید:



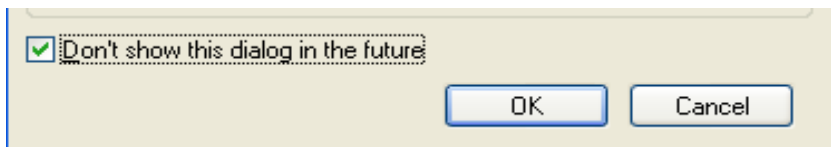
یا از منوی Start و به صورت زیر برنامه SPSS را فراخوانی کنید:



بعد از ورود به محیط SPSS احتمالاً با کادر زیر روبرو خواهید شد:

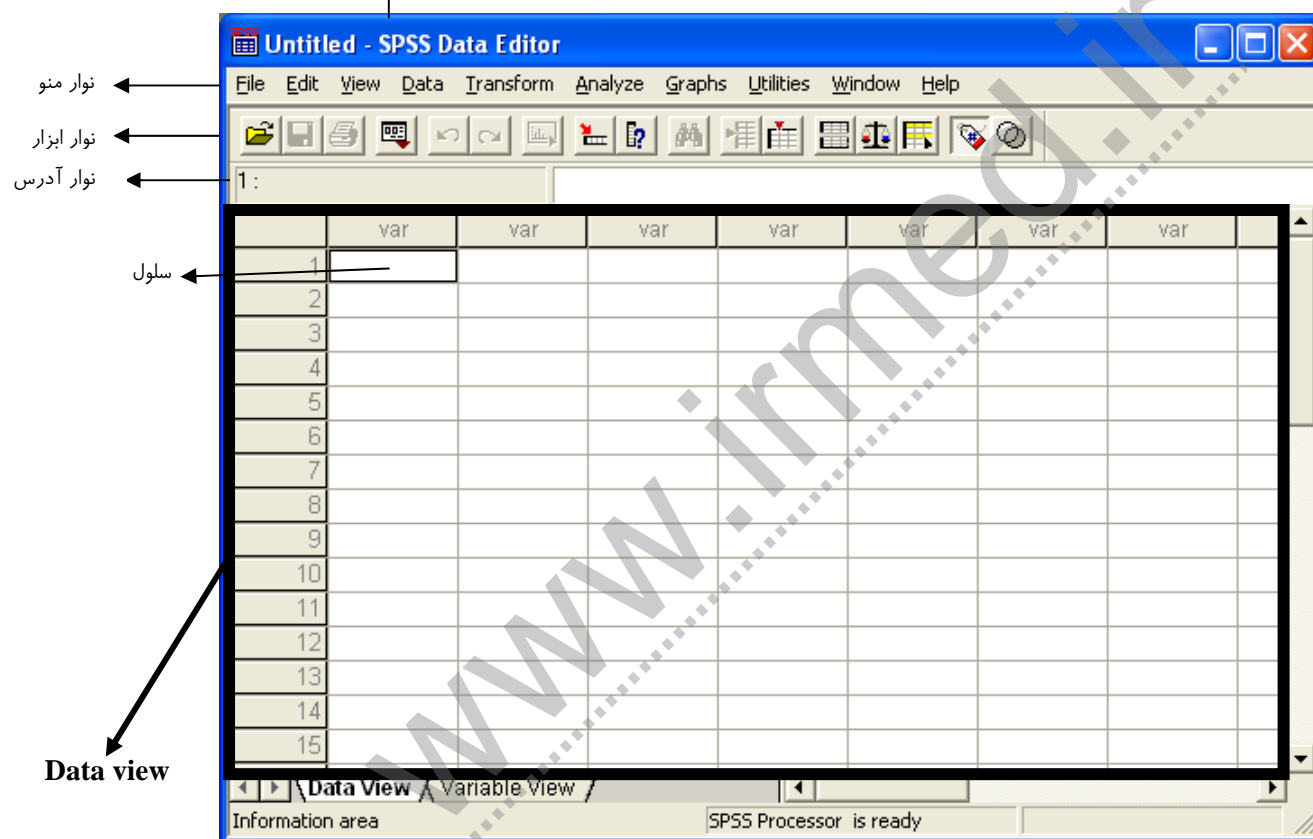


اگر در کادر زیر تیک بگذارید، در دفعات بعدی هنگام ورود به برنامه این سوال پرسیده نخواهد شد و به صورت پیش فرض یک صفحه کار خالی باز خواهد شد (این کار توصیه نمی‌شود).



با دادن پاسخ مناسب به کادر گفتگوی بالا وارد محیط SPSS خواهید شد که تصویر محیط اولیه SPSS را می‌توانید در زیر مشاهده کنید. که قسمت‌های مختلف نرم‌افزار در زیر نشان داده شده است.

نوار عنوان



بررسی اجمالی محیط کاری SPSS

قبل از آشنایی با نحوه وارد کردن داده‌ها لازم است که با محیطی از SPSS که ابتدا با آن برخورد خواهید خورد، بطور مختصر آشنا شوید:

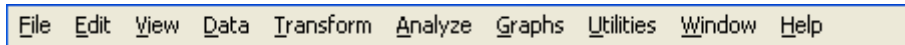
نوار عنوان:

برنامه SPSS نیز مانند سایر نرم‌افزارهای تحت Windows دارای نوار عنوان است که در آن نام فایل SPSS را مشاهده می‌کنید. برای مثال فایل داده‌های زیر به نام Employee data ذخیره شده است:



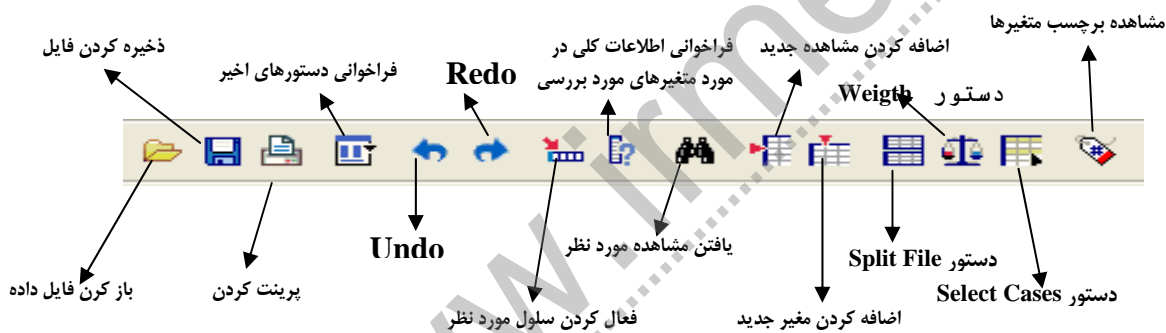
نوار منو:

همان طور که در شکل زیر مشاهده می کنید این نوار شامل منوهای زیادی می باشد که به مرور زمان با آن ها آشنا می شوید (مهم ترین منوهای این نوار برای شما، منوهای Data، Graph، Analyze و Transform می باشند):



نوار ابزار:

این نوار در زیر نوار منو قرار دارد. بر روی این نوار دکمه هایی تعبیه شده است که انجام دادن برخی اعمال را در SPSS راحت می کند. به مرور زمان با دکمه های این نوار آشنا خواهید شد. در شکل زیر دکمه های این نوار معرفی شده اند:



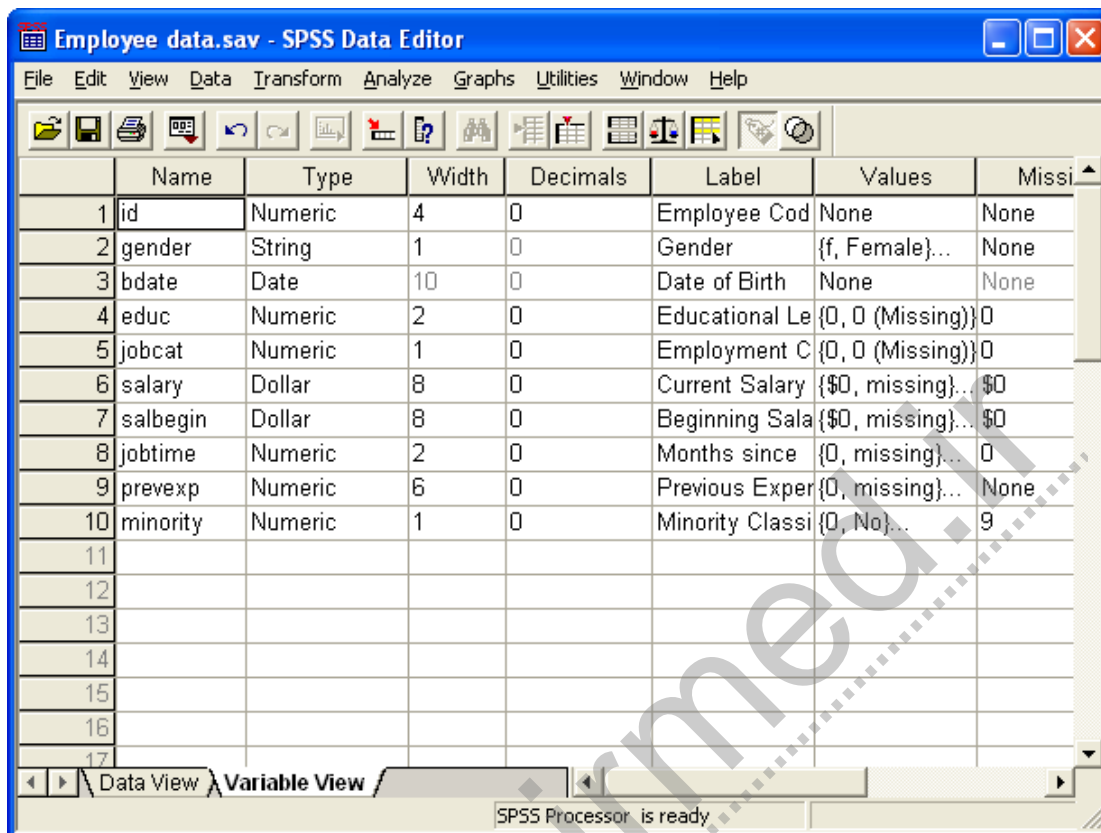
صفحه Data View

این صفحه مخصوص وارد کردن داده هاست. ما با این صفحه به دفعات زیاد بر خواهیم خورد:

	id	gender	bdate	educ	jobcat	salary	salbegin
1	1	Male	02/03/1952	15	Manager	\$57,000	\$27,000
2	2	Male	05/23/1958	16	Clerical	\$40,200	\$18,750
3	3	Female	07/26/1929	12	Clerical	\$21,450	\$12,000
4	4	Female	04/15/1947	8	Clerical	\$21,900	\$13,200
5	5	Male	02/09/1955	15	Clerical	\$45,000	\$21,000
6	6	Male	08/22/1958	15	Clerical	\$32,100	\$13,500
7	7	Male	04/26/1956	15	Clerical	\$36,000	\$18,750
8	8	Female	05/06/1966	12	Clerical	\$21,900	\$9,750
9	9	Female	01/23/1946	15	Clerical	\$27,900	\$12,750
10	10	Female	02/13/1946	12	Clerical	\$24,000	\$13,500
11	11	Female	02/07/1950	16	Clerical	\$30,300	\$16,500
12	12	Male	01/11/1966	8	Clerical	\$28,350	\$12,000
13	13	Male	07/17/1960	15	Clerical	\$27,750	\$14,250
14	14	Female	02/26/1949	15	Clerical	\$35,100	\$16,800
15	15	Male	08/29/1962	12	Clerical	\$27,300	\$13,500

صفحه Variable View

این صفحه نیز همان طور که از نامش پیداست مخصوص وارد کردن متغیرهاست:



نوار آدرس:

در این نوار می توانید آدرس سلول فعال در صفحه Data View مشاهده کنید. برای مثال در شکل زیر سلول فعال، مشاهده دوم از متغیر X2 است که عدد ۵ در آن تایپ شده است:

	id	gender	bdate	educ	jobcat
1	1	Male	02/03/1952	15	Manager
2	2	Male	05/23/1958	16	Clerical
3	3	Female	07/26/1929	12	Clerical
4	4	Female	04/15/1947	8	Clerical
5	5	Male	02/09/1955	15	Clerical



4 : educ	8
----------	---

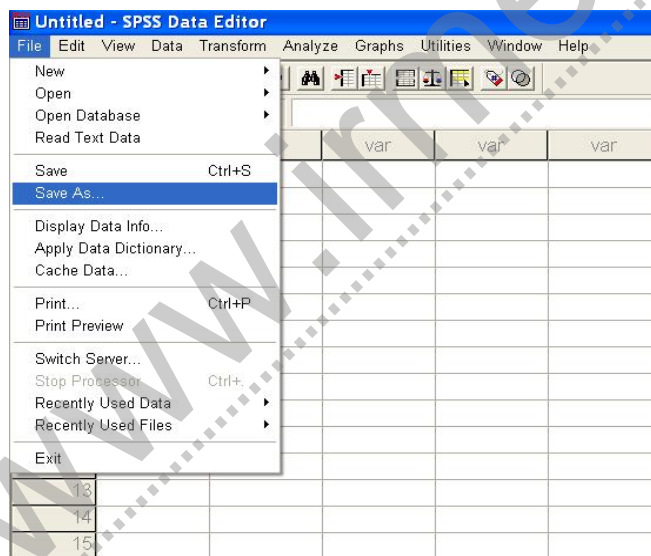
نوار وضعیت :

SPSS همواره وضعیت جاری خود را در کادری در پایین صفحه به نام **نوار وضعیت** به اطلاع شما می‌رساند، توجه کنید که زمانی برنامه به خوبی کار خواهد کرد که عبارت **SPSS Processor is ready** در این کادر نوشته شده باشد.

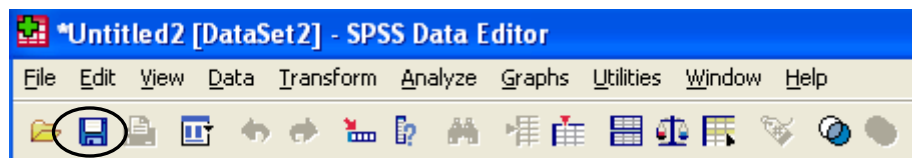


- ذخیره کردن و باز کردن فایلها در SPSS:

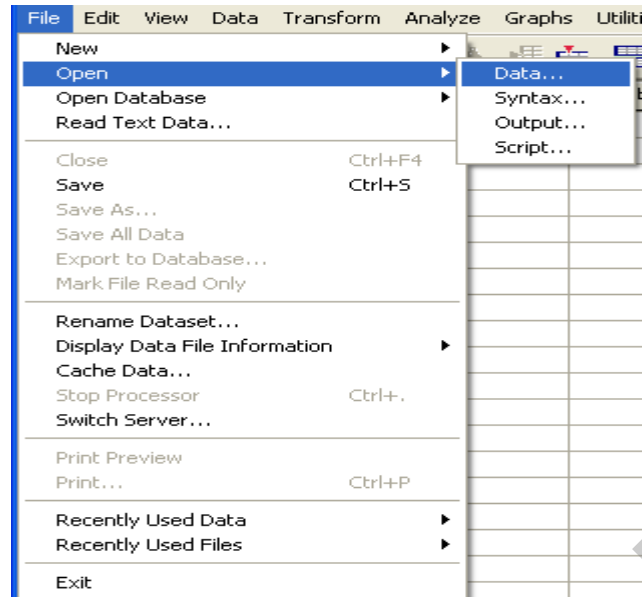
برای ذخیره کردن فایلها در SPSS همانند تمام برنامه های تحت ویندوز از منوی File استفاده می‌شود:



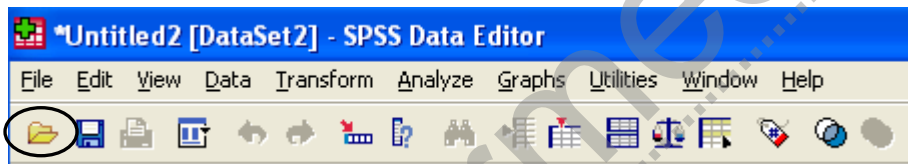
به جای مسیر فوق می‌توانید از کلید میانبر زیر استفاده کنید:



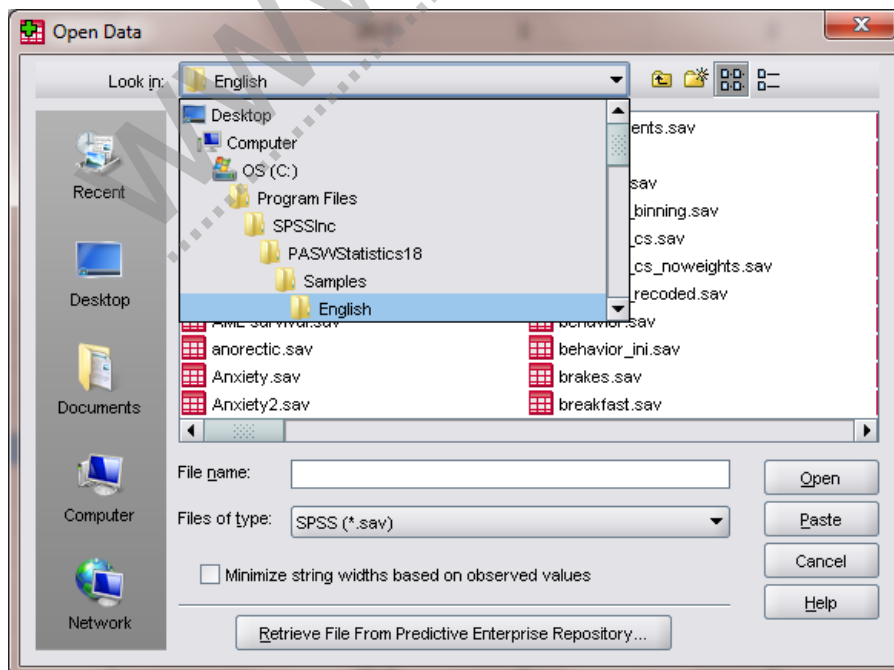
برای باز کردن یک فایل حاوی داده‌های SPSS از مسیر زیر استفاده کنید:



به جای مسیر فوق می‌توانید از کلید میانبر زیر استفاده کنید:



SPSS یکسری فایل‌های مثال دارد که شما می‌توانید برای تمرین از این فایل‌ها استفاده کنید. مسیر این فایل‌ها در شکل زیر نشان داده شده است.



نحوه وارد کردن داده ها در SPSS:

معمولاً داده‌هایی که در اختیار داریم عبارتند از مقادیر چندین متغیر برای یک واحد نمونه. اعداد اولین ستون سمت چپ، نماینده واحدهای نمونه ای هستند و در هر سطر مقادیر متغیرهای مورد بررسی برای هر واحد نمونه وارد خواهد شد.

به عنوان مثال داده‌های مربوط به ۲۰ نفر از دانشجویان از متغیرهای جنس (مرد و زن) و میزان تحصیلات (فوق دیپلم، لیسانس و فوق لیسانس) و معدل به صورت زیر است:

در این قسمت می‌خواهیم این داده‌ها را وارد SPSS کنیم

	id	gender	educ	average	var	var	var
1	1	male	14	11.50			
2	2	male	14	14.00			
3	3	female	18	16.00			
4	4	male	18	15.35			
5	5	male	16	14.14			
6	6	female	14	16.18			
7	7	male	16	17.00			
8	8	male	18	18.04			
9	9	male	14	14.54			
10	10	female	18	17.12			
11	11	female	14	16.12			
12	12	male	16	12.50			
13	13	male	16	13.60			
14	14	female	16	14.35			
15	15	female	14	14.35			
16	16	male	18	17.15			
17	17	female	18	18.00			
18	18	male	16	18.18			
19	19	male	14	15.16			
20	20	female	18	14.18			

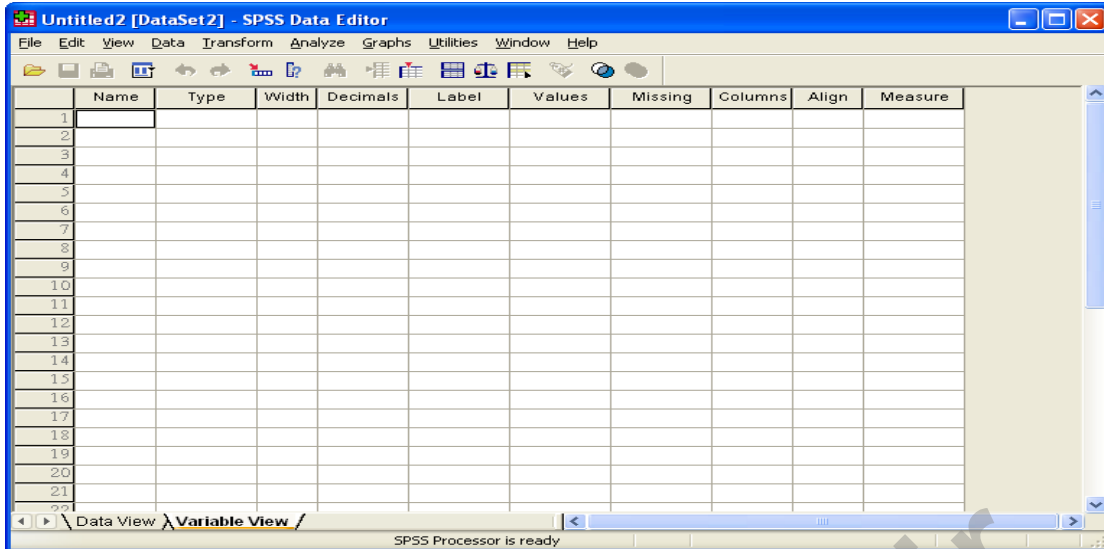
اما قبل از وارد کردن داده‌ها باید متغیرهای مختلفی که مورد بررسی قرار گرفته اند (جنسیت، میزان تحصیلات و معدل) در نرم افزار تعریف شوند.

تعریف متغیرها در SPSS:

برای تعریف متغیرها ابتدا روی عبارت Variable view در نوار زیر کلیک کنید:



محیط کار SPSS به صورت زیر تغییر خواهد کرد:



مشخصه های مختلف یک متغیر در SPSS :

Name: در این قسمت نام متغیر را وارد میکنیم. نام یک متغیر در SPSS حداکثر ۸ کاراکتر متشکل از ارقام و حروف است که شامل یکسری محدودیتها می باشد. برای مثال کاراکتر اول آن نباید رقم باشد، یا نباید در از کاراکترهای فاصله، @، #، \$، % و ... استفاده کرد. به عنوان مثال داریم:

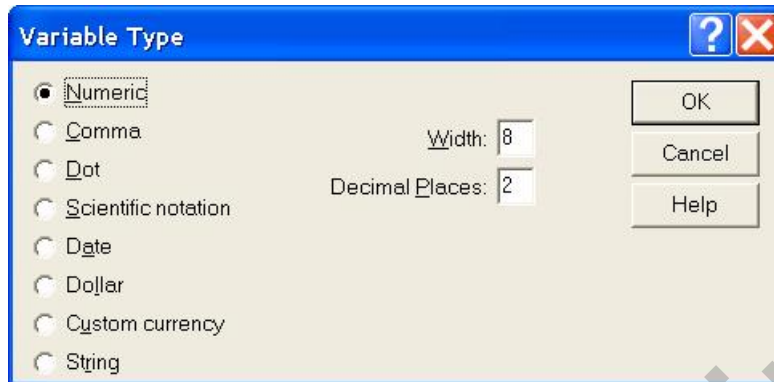
نامهای مجاز در SPSS	نامهای غیر مجاز در SPSS
Gender	2Gender
Educ	Mizane Tahselat
X2	2X

سعی کنید اسمی که انتخاب می کنید حتی الامکان بامفهوم باشد تا در مراحل بعدی، کارتان ساده تر شود. به عنوان مثال برای متغیرهای مثال ما نامهای gender، edu و aver در نظر می گیریم.

Type: روش تایپ اعداد مربوط به متغیر مردنظر را مشخص می کند. غیر از دو نوع Numeric (تایپ عددی مثل: قد، وزن، تعداد واحد، ...) و String (تایپ رشته ای مثل: اسم، گروه خونی، شغل و ...) سایر نوعها کاربرد چندانی ندارند و در موارد بسیار خاصی به کار می روند.

به طور کلی هرگاه خواستیم در data view مشاهدات مربوط به یک متغیر را به صورت عددی وارد کنیم گزینه Numeric را انتخاب می کنیم ولی اگر خواستیم مشاهدات مربوط به متغیری را به صورت حرفی

وارد کنیم گزینه String را انتخاب می کنیم. برای مثال برای تعریف متغیر جنسیت در SPSS اگر بخواهیم داده ها را به صورت ۱ و ۲ تعریف کنیم گزینه Numeric و اگر بخواهیم به صورت m و f (یا male و female) تعریف کنیم گزینه String را انتخاب می کنیم.



Decimals و Width: در این دو ستون تعداد کل ارقام عدد یا پهنای آن (Width) و تعداد رقم‌های اعشار (Decimals) نشان داده شده را می توان تعیین کرد.

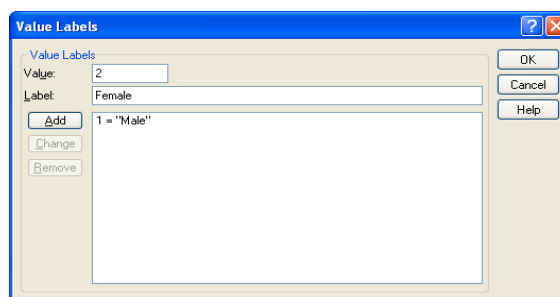
Label: همان طور که گفتیم اسم متغیرها در SPSS دارای محدودیت است (تا ۸ کاراکتر و ...). در صورتی که مایل باشیم میتوانیم در این قسمت یک عنوان (Label) برای متغیر قرار دهیم. به عنوان مثال اگر متغیر مورد بررسی تعداد افراد هر خانواده باشد می توانیم در قسمت Label تایپ کنیم:

Number of family members

➤ در صورتیکه برای متغیر، Label تعریف شود در خروجی‌های SPSS به جای نام متغیر این عنوان نمایش داده خواهد شد در غیر این صورت نام متغیر به کار خواهد رفت.

Value: در این قسمت اگر متغیر مورد بررسی دارای مقادیر مشخصی باشد، مقادیر آن و برچسب مقادیر را مشخص میکنیم. به عنوان مثال متغیر جنسیت دارای دو مقدار ۱ با برچسب "مرد" و ۲ با برچسب "زن" است.

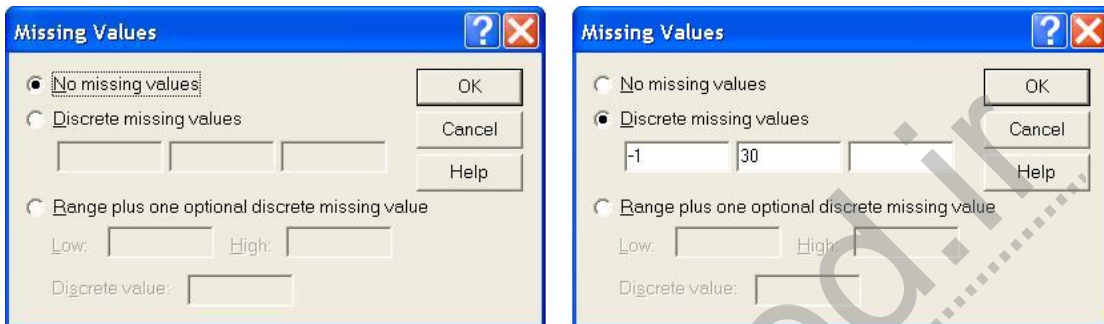
➤ برای اضافه کردن یک مقدار و برچسب مربوط به آن مطابق تصویر زیر مقدار و برچسب را در کادرهای مربوطه تایپ کرده و دکمه Add را کلیک کنید:



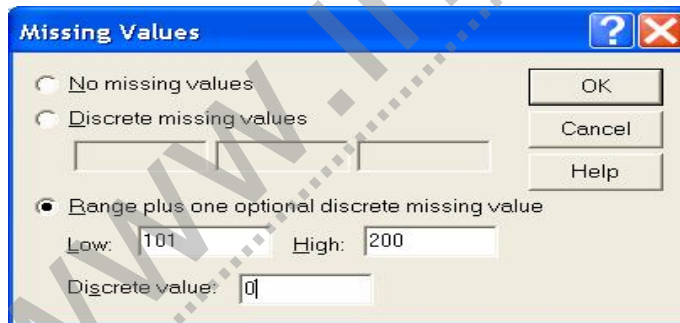
توصیه می شود که برای بهتر تعریف کردن متغیرهای کیفی، Type متغیرها را Numeric تعریف کنید و در قسمت Value برای هر یک از اعداد Label تعریف کنید.

Missing: در این قسمت نحوه معرفی داده های گمشده به SPSS تعیین میشود.

انتخاب گزینه اول به معنی عدم تعریف داده گمشده است. با انتخاب گزینه دوم میتوان چند مقدار مشخص را برای معرفی یک مقدار گمشده تعیین کرد. به عنوان مثال اگر متغیر مورد بررسی تعداد واحد باشد میتوان از مقادیری که خارج از دامنه تغییراتند چند مقدار برای معرفی مقادیر گمشده تعیین کرد. (مثلاً ۳۰ یا -۱) همچنین، در اکثر پرسشنامه ها اعدادی همانند ۹۹ یا ۹۹۹ برای این کار انتخاب می شوند.



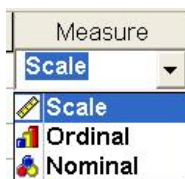
با انتخاب گزینه سوم میتوان یک بازه عددی را برای مشخص کردن مقادیر گمشده تعیین کرد. اگر متغیر مورد بررسی معدل دانشجویان باشد، میتوان اعداد بین ۲۰ تا ۱۰۰ را به عنوان داده گمشده تعریف کرد.



Columns: اندازه ستونی که محل وارد کردن داده های متغیر مورد بررسی است را در این قسمت وارد میکنیم.

Align: تعیین میکند که داده ها در سمت راست، چپ یا وسط خانه (cell) قرار گیرند.

Measure: در این قسمت مقیاس اندازه گیری داده ها را تعیین میکنیم:



Scale- (فاصله ای): اعداد حقیقی هستند که اعمال جبری روی آنها امکان پذیر است.

Ordinal- (رتبه ای): متغیری عددی است که مجاز به انجام عملیات جبری در مورد آنها نیستیم بلکه تنها ارجحیت یا برتری را نشان میدهد.

Nominal- (اسمی): فاقد هرگونه ارجحیت میباشد.

بعد از تعریف تمامی متغیرهای لازم با کلیک روی عبارت Data view به محیط وارد کردن داده ها برگشته و داده های مثال را وارد کنید.



www.irmed.ir

تحلیل آماری:

در اصطلاح عامیانه آمار به معنای ثبت و نمایش اطلاعات عددی در مورد یک موضوع، مثلاً ثبت نسبت سرباری، میزان تولد خام، درصد اشغال تخت، میزان آگاهی دانشجویان از مقررات آموزشی و غیره می‌باشد. ولی علم آمار امروزه دارای مفهومی بسیار وسیع‌تر از این کاربرد عامیانه است. مفاهیم عامیانه آمار زیر مجموعه‌ای از آمار مصطلح بین آماردانان است. از نقطه نظر علمی، آمار به مجموعه روش‌هایی برای جمع‌آوری، تنظیم و خلاصه کردن داده‌ها و انجام استنباط و نتیجه‌گیری بوسیله تجزیه و تحلیل آن‌ها، اطلاق می‌شود.

با بیان دیگر می‌توان گفت که آمار عبارت است از هنر و علم جمع‌آوری، تعبیر، تجزیه و تحلیل داده‌ها و استخراج تعمیم‌های منطقی در مورد پدیده‌های تحت بررسی.

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

آمار توصیفی

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

الف) خلاصه کردن و توصیف الگوی کلی

۱) فشرده کردن داده‌ها در قالب جدول‌های آماری

۲) نمایش آنها بوسیله نمودار

ب) محاسبه شاخصهای آماری

نقش آمار توصیفی در فرآیند تحلیل آماری بسیار مهم و حیاتی است. آمار توصیفی با خلاصه کردن داده‌ها، ویژگی‌های مهم آن‌را نمایان می‌سازد تا ایده‌های لازم را در ذهن پژوهشگر برای مرحله دوم تحلیل آماری (آمار استنباطی) ایجاد کند.

برای استفاده از مراحل مختلف آمار توصیفی می‌توان از چارت زیر بهره‌گیری کرد:



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

الف) خلاصه کردن و توصیف الگوی کلی:

یک مجموعه داده آماری شامل مجموعه‌ای از مقادیر یک یا چند متغیر است. متغیرها می‌توانند کمی یا کیفی (Categorical) باشند. متغیرهای عددی خود به دو دسته گسسته و پیوسته دسته بندی می‌شوند. این دسته بندی، روشهای آماری را که برای داده‌ها مناسب اند، مشخص می‌کند. یکی از روشهای خلاصه کردن و توصیف داده‌ها رسم یک نمودار آماری است. نوع نمودار مورد استفاده به نوع داده‌ها بستگی دارد و بسته به رسته‌ای بودن یا عددی بودن، نمودارهای مختلفی به کار برده می‌شود. جداول فراوانی هم بسته به نوع متغیر، متفاوت خواهند بود، لذا مراحل فوق را برای انواع مختلف متغیر، جداگانه بررسی خواهیم کرد.

الف - (۱) داده های کیفی (Categorical Data)

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

جدول فراوانی برای متغیرهای کیفی:

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

مثال: پزشکی گروه خونی ۲۰ نفر را مورد بررسی قرار داده است اطلاعات به صورت زیر است :

A,B,O,A,B, AB,B,O,A,B, A,B,O,AB,B, A,B,O,A,B

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

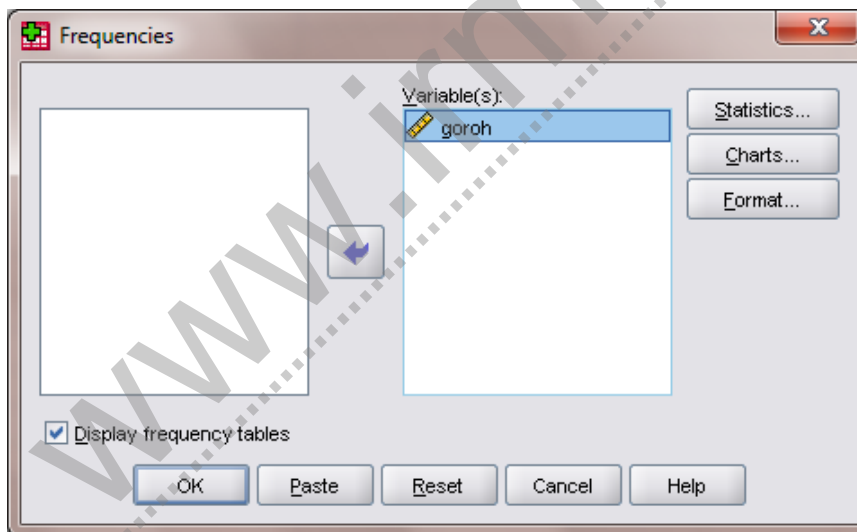
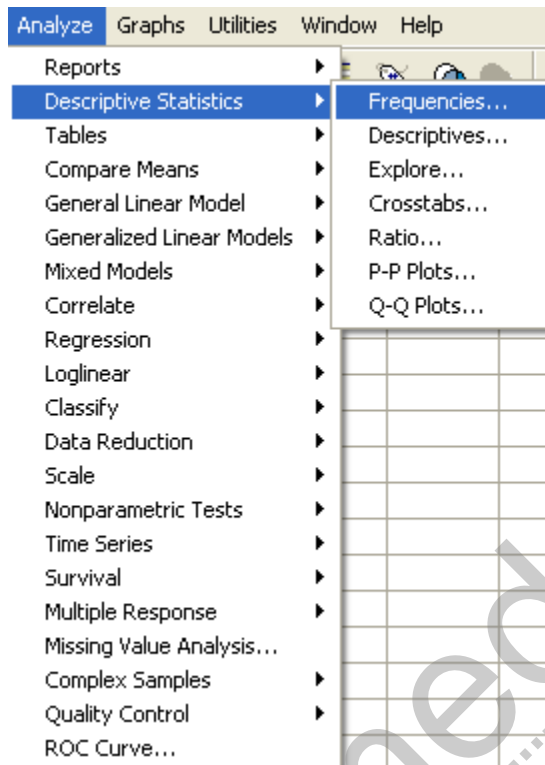
- چند نفر دارای گروه خونی A می باشند؟


- گروه های خونی A,B چند درصد از کل گروه های خونی را شامل می شوند؟

ابتدا داده ها را در محیط SPSS وارد می کنیم:


goroh
A
B
AB
A
B
AB
O
O
B
A
B
A
O
B
AB
AB
O
B
A
A

سپس برای رسم جداول فراوانی مسیر زیر را طی کنید تا کادر کناری باز شود:



متغیر مربوطه را انتخاب کرده روی دکمه  کلیک کنید

برای رسم جدول فراوانی، در کادر کنار عبارت Display frequency tables تیک بگذارید. و در

نهایت دکمه  را کلیک کنید. صفحه جداگانه ای تحت عنوان Output view باز خواهد شد که

خروجی های SPSS همواره در آن ظاهر خواهد شد. خروجی این مثال به صورت زیر خواهد بود:

Frequencies

Statistics

goroh		
N	Valid	20
	Missing	0

goroh					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	6	30.0	30.0	30.0
	B	8	40.0	40.0	70.0
	O	4	20.0	20.0	90.0
	AB	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

جدول اول تعداد داده های موجود (Valid) و داده های گمشده (Missing) را نشان می دهد. و جدول دوم جدول فراوانی متغیر است. ستون اول مقادیر متغیر، ستون دوم فراوانی هر مقدار، ستون سوم درصد فراوانی آن مقدار و ستون چهارم هم درصد فراوانی تجمعی می باشد. از روی این جدول سعی کنید به سوال های مطرح شده در صورت مثال را جواب دهید.

برای متغیرهای کیفی رتبه ای هم مراحل رسم جدول فراوانی به همین صورت خواهد بود.

📍 به عنوان تمرین: فایل Employee data را باز کنید و متغیرهای کیفی این فایل را با استفاده از جدول فراوانی توصیف کنید.

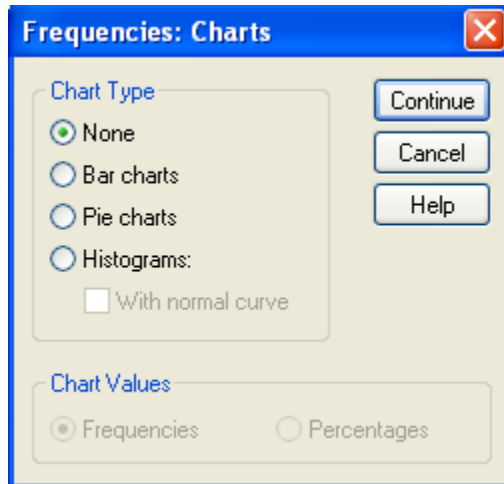
📍 به عنوان تمرین: فایل Breast cancer survival را باز کنید و متغیرهای کیفی این فایل را با استفاده از جدول فراوانی توصیف کنید.

- نمودارهای آماری برای متغیرهای کیفی:

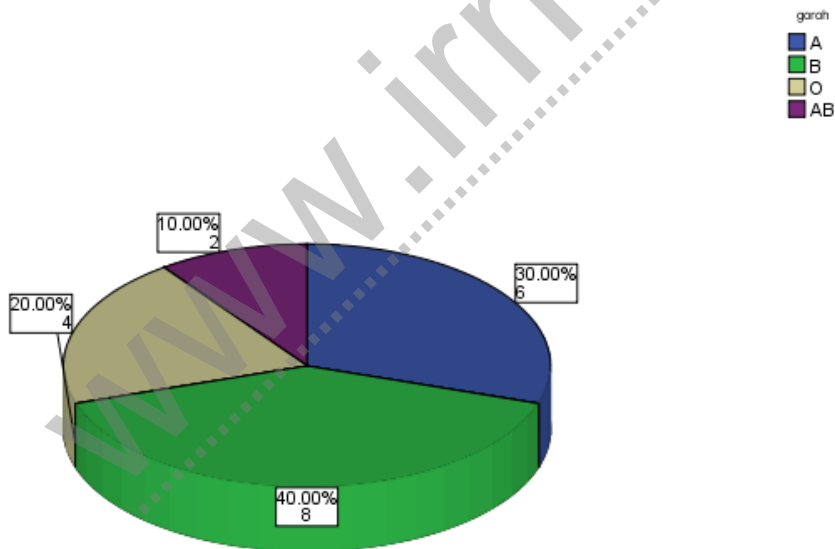
نمودارهای مناسب برای این نوع متغیرها عبارتند از: نمودار میله ای (Bar chart) و نمودار دایره ای (Pie chart).

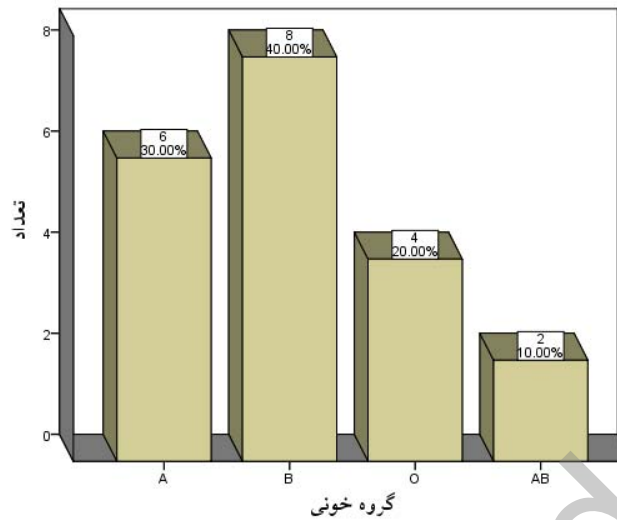
برای رسم این نمودارها می توان از دو راه زیر استفاده کرد:

۱. از مسیری که برای رسم جدول فراوانی طی کردیم رفته روی دکمه **Charts...** کلیک کنید تا کادر گفتگوی زیر باز شود:



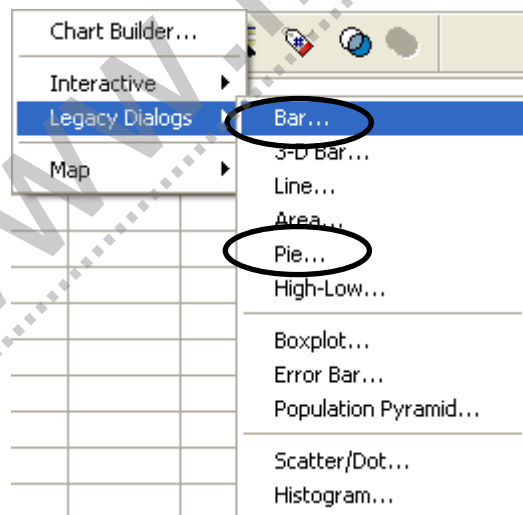
هر کدام از نمودارهای pie chart یا Bar chart را که مایل بودید انتخاب کرده و دکمه Continue را کلیک کنید. دکمه Ok را کلیک کنید. مسیر فوق را برای داده های مثال طی کنید و خروجی را ببینید:



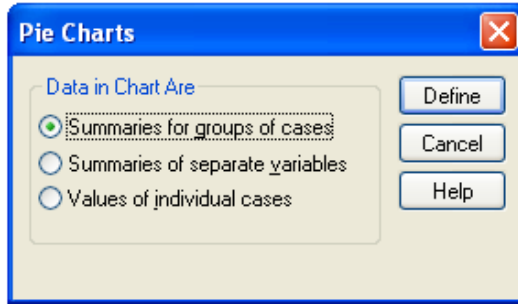
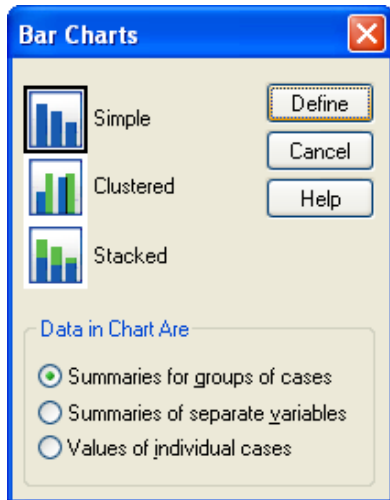


توجه کنید که نمودارهای دایره‌ای برای متغیرهای کیفی اسمی مناسب‌ترند و نمودارهای میله‌ای برای متغیرهای کیفی رتبه‌ای.

۲- راه دوم استفاده از منوی Graph است:



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



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

به عنوان تمرین: فایل Employee data را باز کنید و متغیرهای کیفی این فایل را با استفاده از نمودار ستونی و دایره‌ای توصیف کنید.

به عنوان تمرین: فایل Breast cancer survival را باز کنید و متغیرهای کیفی این فایل را با استفاده از نمودار ستونی و دایره‌ای توصیف کنید.

الف - ۲) داده های عددی (Numerical Data)

داده های عددی دو نوع اند: گسسته و پیوسته. مقادیر متغیرهای گسسته اعداد حاصل از شمارش می باشد برای مثال یک خانواده می تواند یک یا دو فرزند داشته باشد اما تعداد فرزندان خانواده نمی تواند عددی ما بین این دو باشد و در سمت مقابل، متغیرهای پیوسته فاقد واحدهای تفکیک پذیر می باشند. برای مثال وزن یک متغیر پیوسته است.

- متغیر عددی گسسته:

مقادیر یک متغیر گسسته با دامنه تغییرات کم، جدا از هم و معمولاً محدود است و برای رسم جدول فراوانی یک متغیر عددی گسسته همچون حالت متغیر کیفی عمل می کنیم. اما اگر تعداد مقادیر متفاوتی که یک متغیر گسسته با دامنه تغییرات زیاد می گیرد زیاد باشد، برای رسم جدول فراوانی، با آن مثل یک متغیر پیوسته رفتار خواهیم کرد.

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

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

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

-جدول فراوانی داده های فوق را با SPSS رسم کنید.

-نمودارهای میله ای و دایره ای را برای داده های فوق رسم کنید.

-مشخص کنید چند درصد از خانواده ها در طول زمستان ۶ قرص مصرف کرده اند؟ چند درصد حداکثر ۶ قرص مصرف کرده اند؟ چند درصد حداقل ۶ قرص مصرف کرده اند.

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

- متغیر عددی پیوسته:

-تبدیل داده های عددی پیوسته به گسسته

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

مثال: مدت زمان بیهوشی (بر حسب دقیقه) ۴۰ بیمار که به نزدیک ترین عدد صحیح گرد شده اند به قرار زیر است:

۵۲	۳۵	۲۴	۴۷	۳۶	۵۱	۳۴	۳۸	۴۶	۳۳
۴۷	۳۶	۳۸	۵۰	۴۷	۳۴	۴۱	۴۰	۴۲	۴۰
۲۶	۲۹	۳۰	۳۲	۳۰	۳۵	۳۷	۳۷	۴۱	۲۱
۳۱	۳۰	۲۶	۳۵	۴۵	۲۳	۴۳	۳۱	۳۴	۴۳

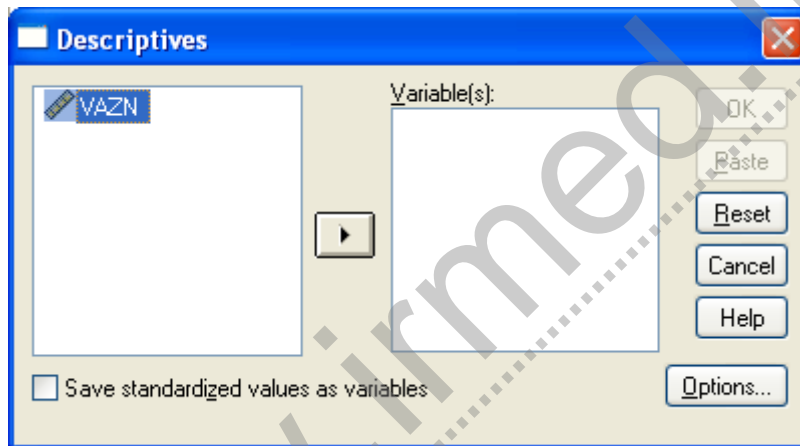
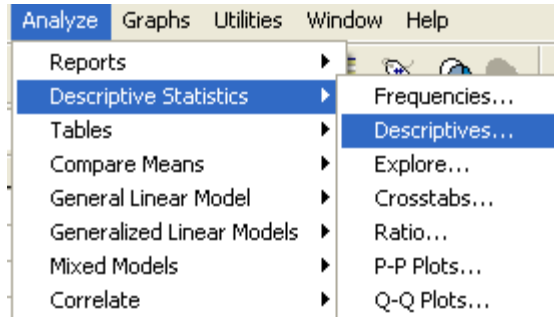
به عنوان تمرین: جدول فراوانی داده های فوق را رسم کنید. به چه نتیجه ای می رسید؟! 📊

هرگاه داده های ما پیوسته باشد، یکی از راه های توصیف این است که داده ها را به تعدادی رده با طول مساوی تقسیم می کنیم و در هر رده فراوانی داده ها را می شماریم برای بدست آوردن تعداد رده ها در رسم جدول فراوانی به ترتیب زیر عمل کنید:

۱. ابتدا حدود تغییرات داده ها را که از فرمول زیر بدست می آید محاسبه می کنیم:

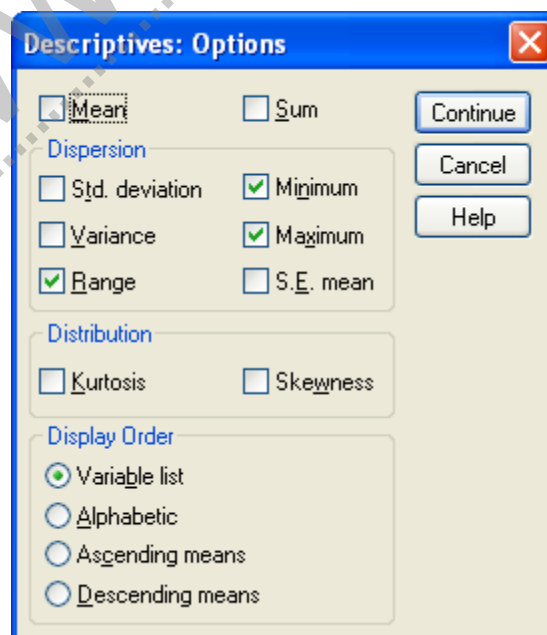
Range = Max - Min

برای محاسبه دامنه تغییرات، کمترین داده‌ها و بیشترین داده‌ها از مسیر زیر استفاده می‌کنیم:



روی Options... کلیک کنید تا کادر زیر ظاهر شود. گزینه‌های مورد نظر را تیک بگذارید. تا خروجی به

شکل زیر ظاهر شود:



Range, Maximum, Minimum را انتخاب کنید. دکمه Continue و سپس Ok را کلیک کنید.
 ۲. برای بدست آوردن تعداد رده ها یک قاعده عمومی وجود ندارد و معمولاً تعداد رده ها را بین ۵ تا ۲۵ رده
 اختیار می کنند. یک قاعده مفید استفاده از دستور استورگس^۲ است:

$$m=1+3.322 \log (n) \quad (n: \text{تعداد کل داده ها})$$

چون حاصل یک عدد اعشاری خواهد بود. آن را به بزرگترین عدد صحیح گرد می کنیم.
 در مثال بالا داریم:

$$m=1+3.322 \log (40) =6.322$$

پس تعداد طبقات را ۷ می گیریم.

۳. چون مدت زمان های بیهوشی به نزدیکترین عدد صحیح گرد شده اند بنابراین عدد ۳۵ در داده ها در واقع
 عددی بین ۳۴/۵ و ۳۵/۵ می باشد. عدد ۰/۵ را تغییر پذیری مقادیر داده ها می نامیم که در ساختن حدود
 طبقات مورد استفاده قرار می گیرد. طول هر طبقه را هم از فرمول زیر محاسبه می کنیم:

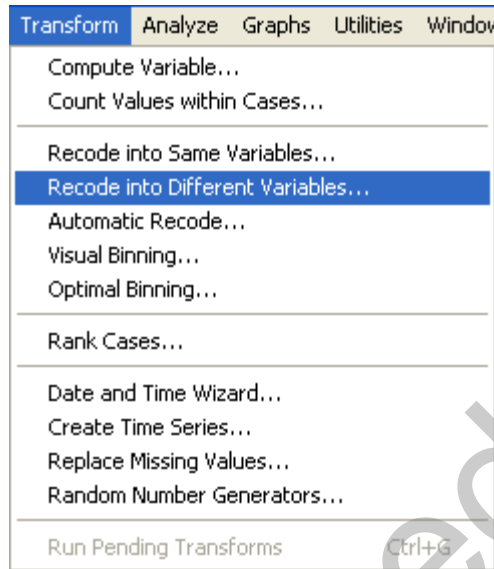
$$L=\text{Range}/m=4.574$$

پس طول هر طبقه را ۵ در نظر می گیریم. حال باید ۷ طبقه به طول ۵ بسازیم. طبقات مورد نظر به صورت
 زیر خواهد بود:

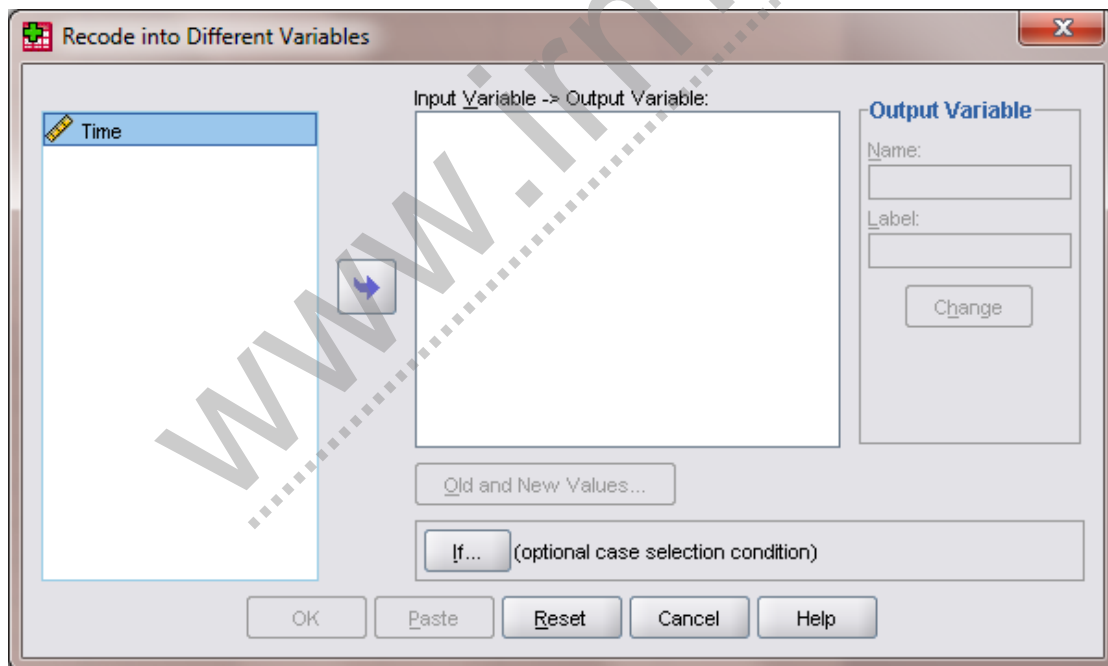
20.5-25.5	→ 1
25.5-30.5	→ 2
30.5-35.5	→ 3
35.5-40.5	→ 4
40.5-45.5	→ 5
45.5-50.5	→ 6
50.5-55.5	→ 7


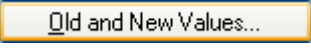
انتخاب حدود طبقات به صورت فوق باعث می شود که هر عدد دقیقاً در یک دسته قرار گیرد.

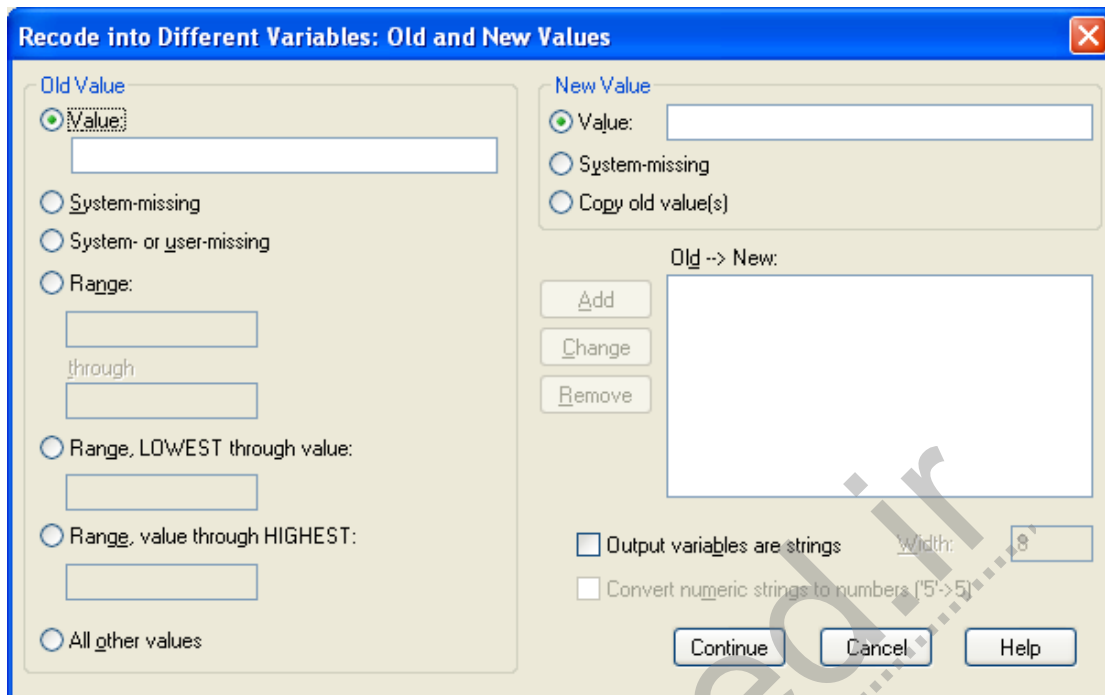
۴. حال متغیر جدیدی تعریف می کنیم که با توجه به واقع شدن داده در یکی از فواصل هفت گانه بالا یکی از مقادیر ۱ تا ۷ را بپذیرد. برای تعریف این متغیر مسیر زیر را طی کنید:



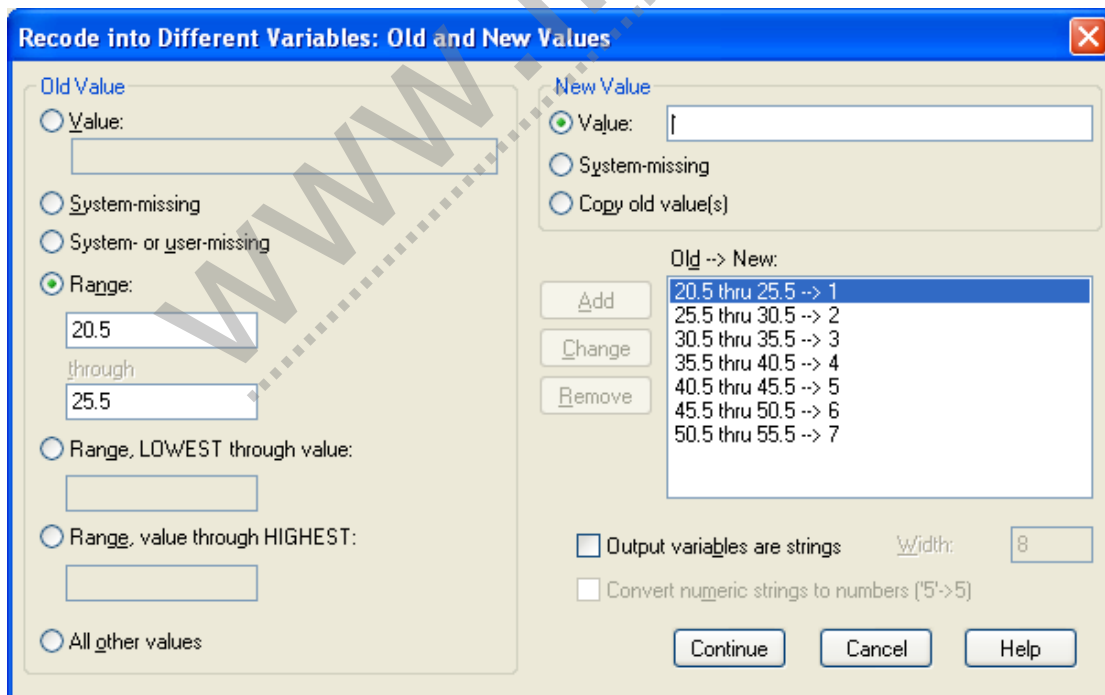
کادر گفتگوی زیر ظاهر می شود:



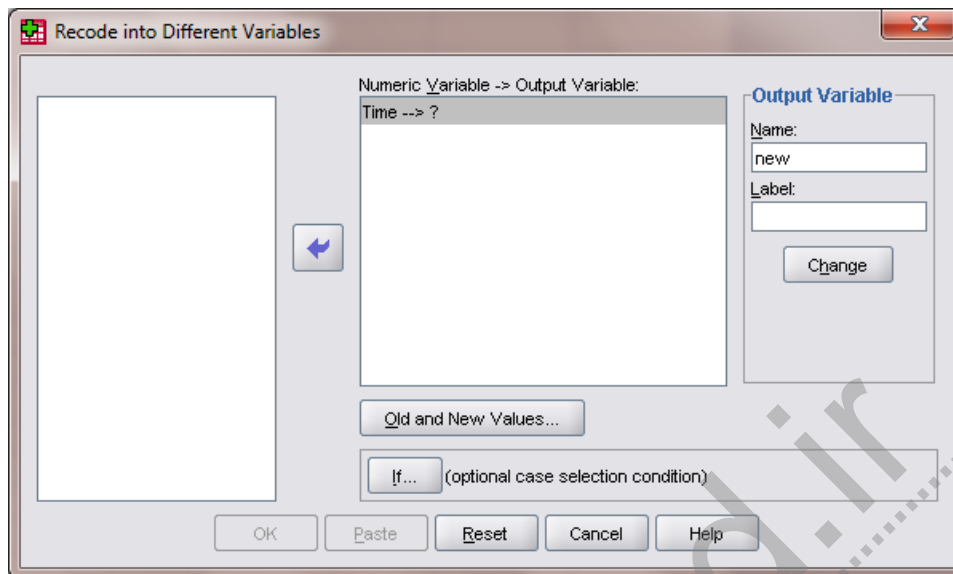
متغیر مورد نظر را انتخاب کرده دکمه  را فشار دهید. سپس روی  کلیک کنید کادر گفتگوی زیر باز می شود که در آن مقادیر متغیر جدید را با توجه به مقادیر متغیر اولیه مشخص می کنیم:



همانند تصویر زیر در سمت چپ کادر گفت و گوی بالا حدود طبقات را وارد کرده، در سمت راست کادر گفت و گو مقدار متغیر جدید (شماره طبقه) را وارد کنید. سپس دکمه Add را کلیک کنید. این کار را برای تمام طبقات به ترتیب انجام دهید.



پس از تعریف مقادیر دکمه Continue را کلیک کنید. مطابق تصویر زیر در قسمت Output variable نام متغیر جدید را تایپ کرده دکمه change را کلیک کنید.

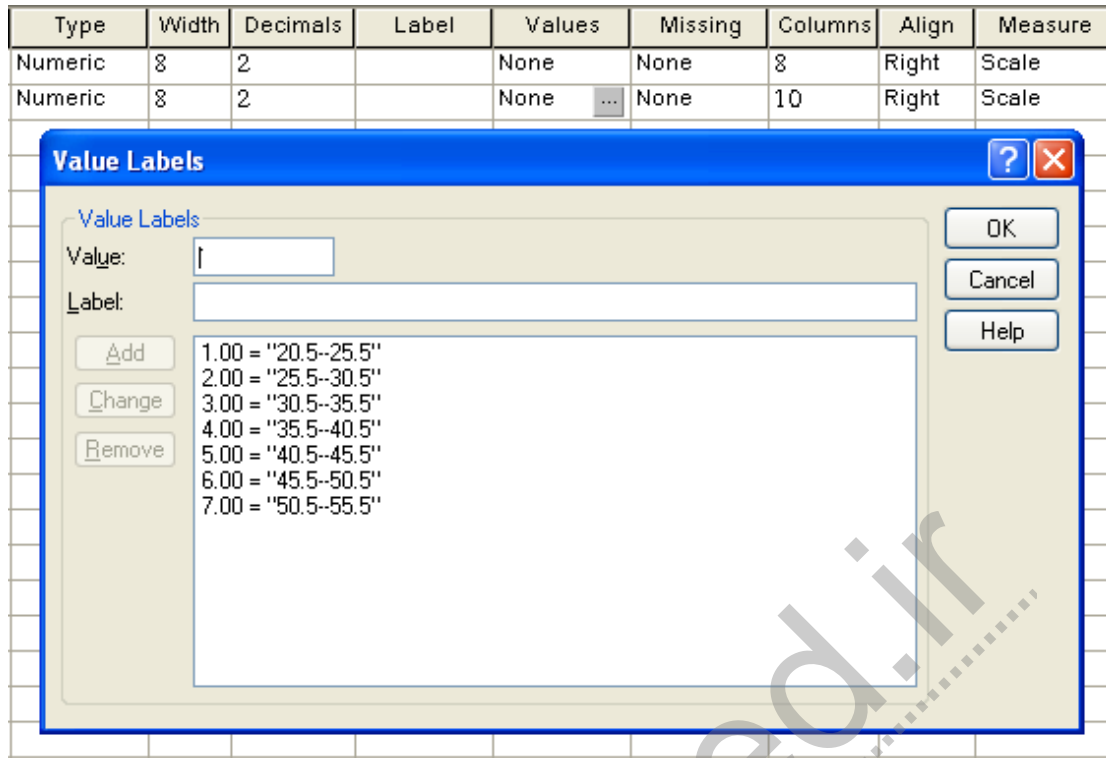


در پایان دکمه ok را کلیک کرده و قسمتی از نتیجه کار را به صورت زیر مشاهده کنید:

	Time	new
1	52.00	7.00
2	47.00	6.00
3	26.00	2.00
4	31.00	3.00

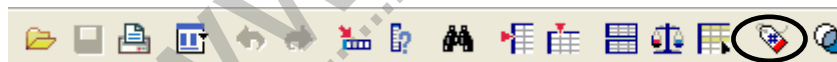
مشاهده می کنید که متغیر جدید با مقادیر ۱ تا ۷ ساخته شده است

۵. به قسمت variable view بروید و در قسمت value برای مقادیر متغیر جدید، بر چسب‌هایی به صورت زیر تعریف کنید:

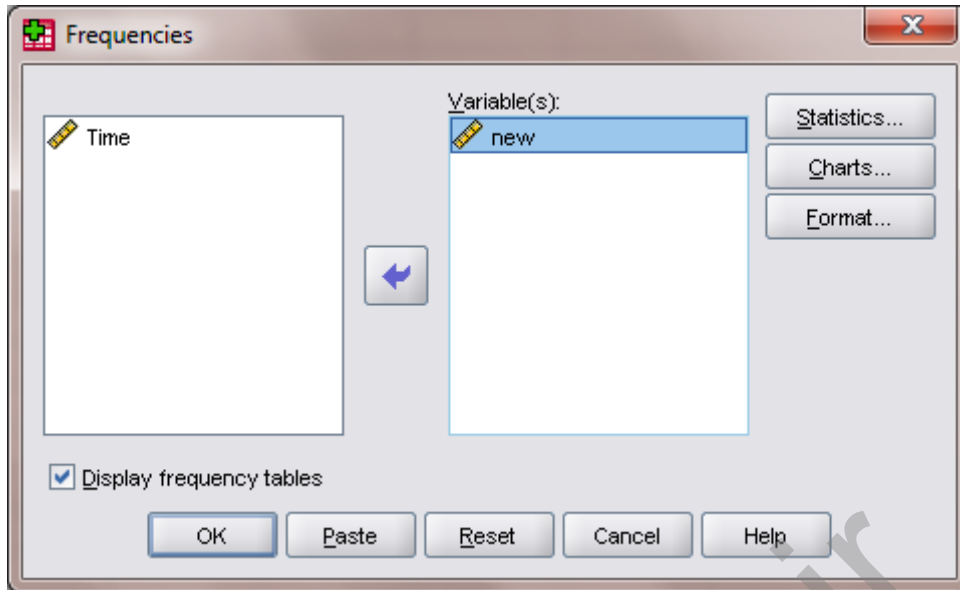


به قسمت Data view برگردید.

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



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



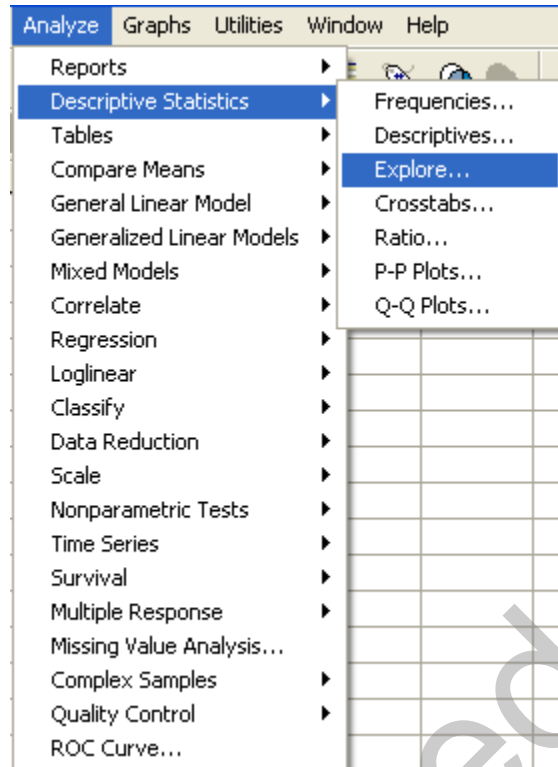
جدول فراوانی داده های مثال زمان بیهوشی به صورت زیر خواهد بود:

new

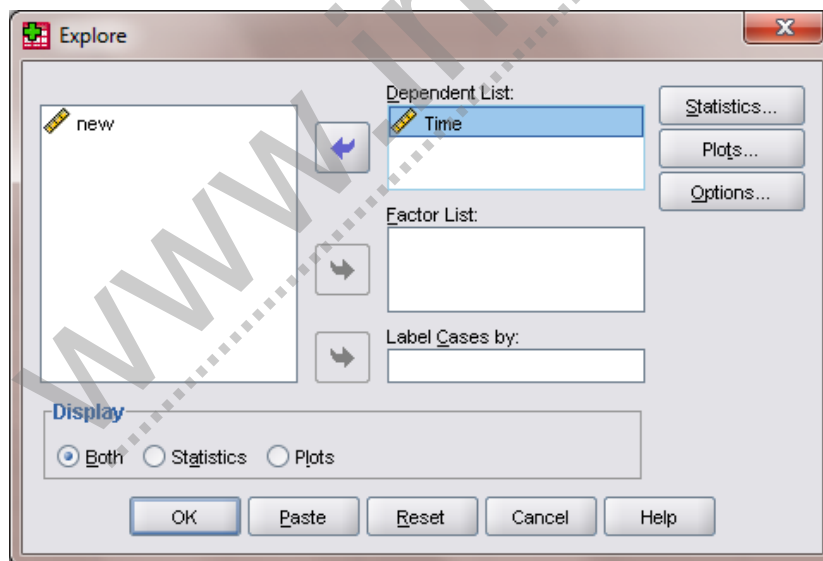
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 20.5--25.5	3	7.5	7.5	7.5
25.5--30.5	6	15.0	15.0	22.5
30.5--35.5	10	25.0	25.0	47.5
35.5--40.5	8	20.0	20.0	67.5
40.5--45.5	6	15.0	15.0	82.5
45.5--50.5	5	12.5	12.5	95.0
50.5--55.5	2	5.0	5.0	100.0
Total	40	100.0	100.0	

- رسم نمودار برای داده های کمی پیوسته:

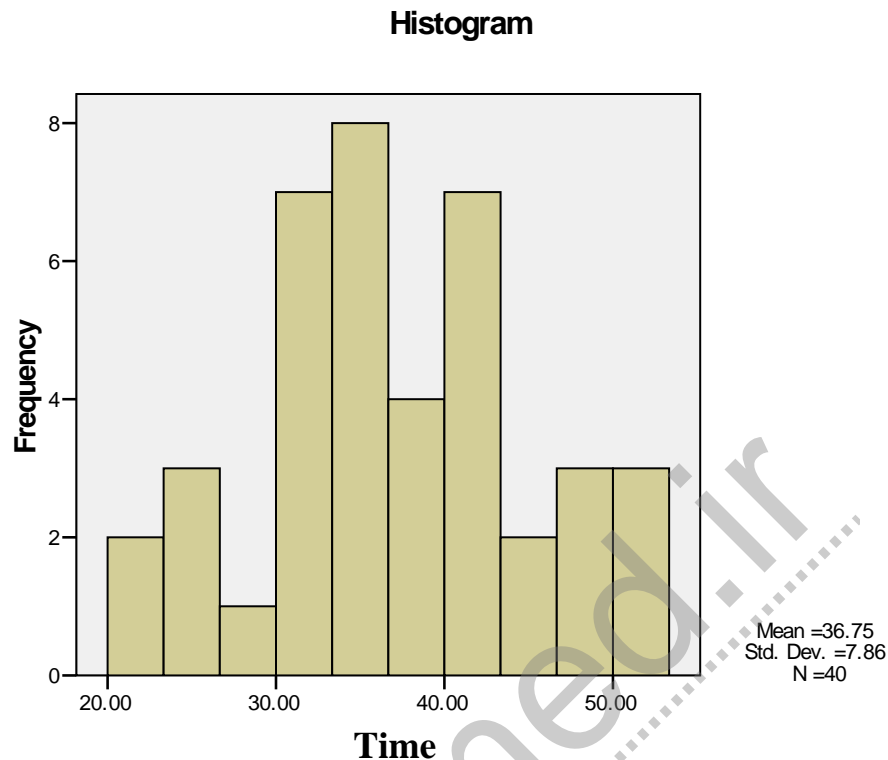
برای داده های پیوسته نمودارهای مختلفی به کار می روند که هر کدام کاربردهای خاص خود را دارد. ارجح ترین این نمودارها Histogram است. برای رسم هیستوگرام داده های پیوسته مسیر زیر را طی کنید:



کادر گفتگوی زیر ظاهر خواهد شد:



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



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

ب) محاسبه شاخصهای آماری:

مرحله اول آمار توصیفی یعنی تشکیل جداول فراوانی و رسم نمودارهای آماری در SPSS بیان شد. مرحله دوم در آمار توصیفی، خلاصه کردن داده ها در قالب اعدادی است که موسوم به **شاخصهای آماری** هستند.

شاخصهای آماری به دو دسته تقسیم می شوند: شاخصهائی که گرایش به مرکز یا مرکزیت داده ها را اندازه می گیرد (شاخصهای مرکزی) و شاخصهائی که برای اندازه گیری تغییرپذیری داده ها به کار می رود (شاخصهای پراکندگی)

-شاخص‌های مرکزی:

شاخص‌های مرکزی مهم عبارتند از:

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

میانه (Median) و چندکها: میانه به داده وسطی داده‌ها اطلاق می‌شود و در داده‌های کم تعداد یک شاخص پرکاربرد و کار آمد است.

میانه داده ای است که تقریباً نصف داده‌ها از آن کمتر و نصف داده‌ها از آن بیشترند.

تعریف چندکها هم معادل میانه است، چندک مرتبه p ، مقداری است که تقریباً $100p$ درصد داده‌ها از آن کمتری مساوی آن و $100(1-p)$ درصد داده‌ها از آن بیش‌ترند. ساده‌ترین نوع چندکها، چارکها (Quartiles) و دهکها هستند.

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

- چارک دوم: معادل میانه است.

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

- دهک اول: مقداری است که یک دهم داده‌ها از آن کمتر یا مساوی با آن هستند. سایر دهک‌ها هم به همین صورت تعریف می‌شوند.

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

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

-شاخص‌های پراکندگی:

غیر از شاخص‌هایی که گرایش داده‌ها را به یک مقدار مرکزی نشان می‌دهد، علاقه مند به شاخص‌هایی هستیم که به نوعی میزان پراکندگی داده‌ها را بیان کنند. مهم‌ترین شاخص‌های آماری پراکندگی عبارتند از:

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

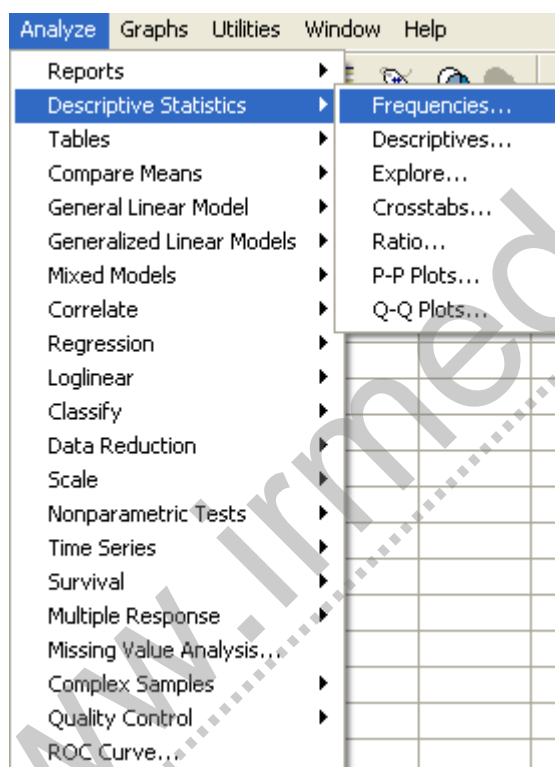
واریانس (Variance): میانگین مربعات تفاضل داده ها از میانگین را واریانس گویند.

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

انحراف استاندارد میانگین (Standard Error of Mean): جذر حاصل تقسیم واریانس بر تعداد داده

ها را انحراف استاندارد میانگین گویند.

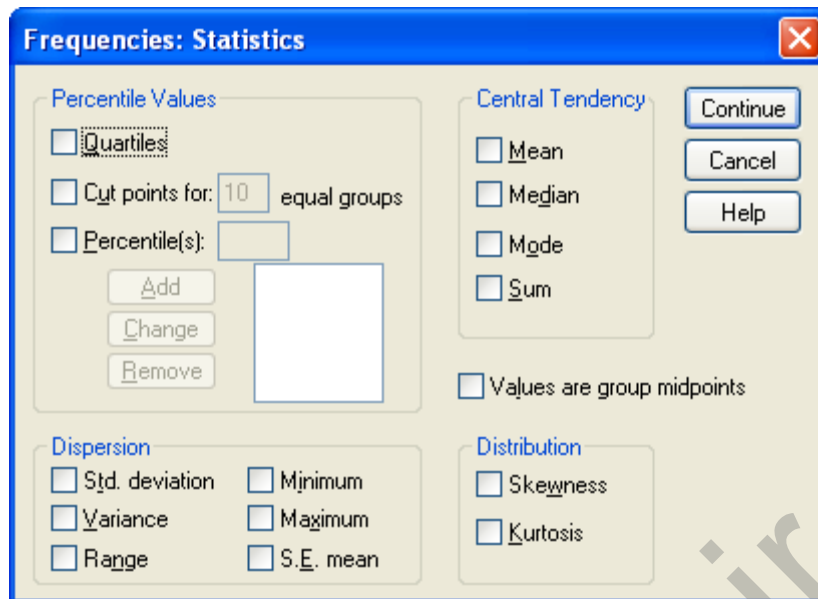
برای محاسبه شاخصهای بالا، ابتدا مسیر زیر را طی کنید:



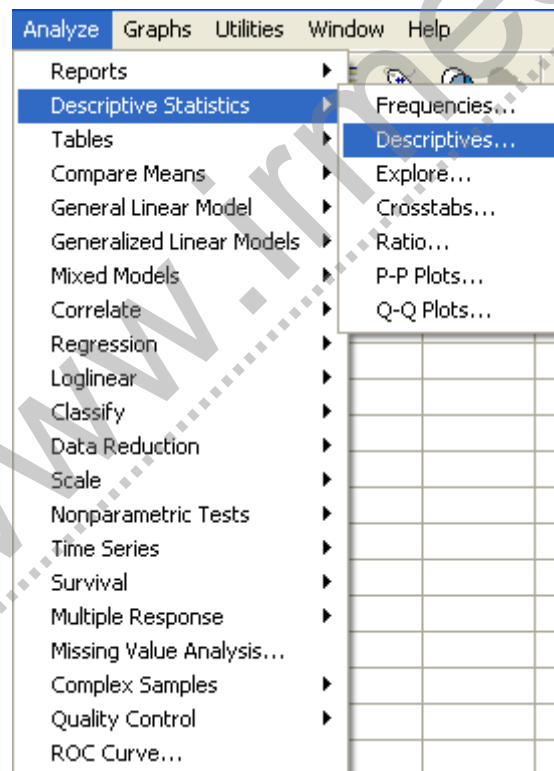
روی دکمه **Statistics...** کلیک کنید. کادر زیر باز خواهد شد. شاخصهای مرکزی را میتوانید از کادر

Central Tendency و شاخصهای پراکندگی را از کادر Dispersion انتخاب کنید.

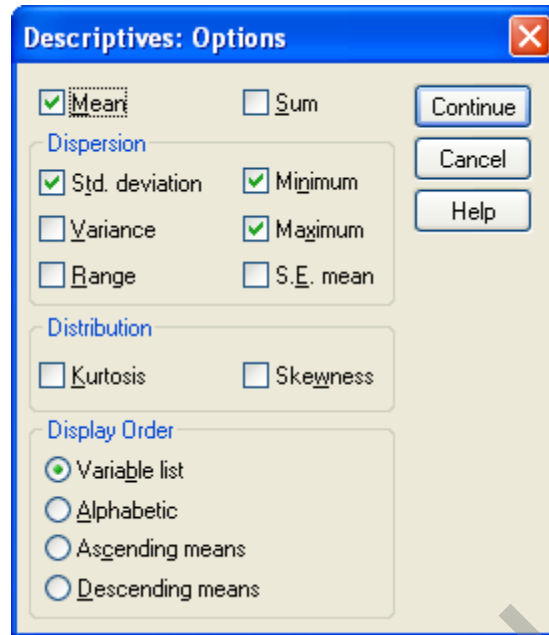
پس از انتخاب شاخصهای مورد نظر دکمه Continue و سپس OK را کلیک کنید.



و یا میتوان از مسیر زیر استفاده کرد:

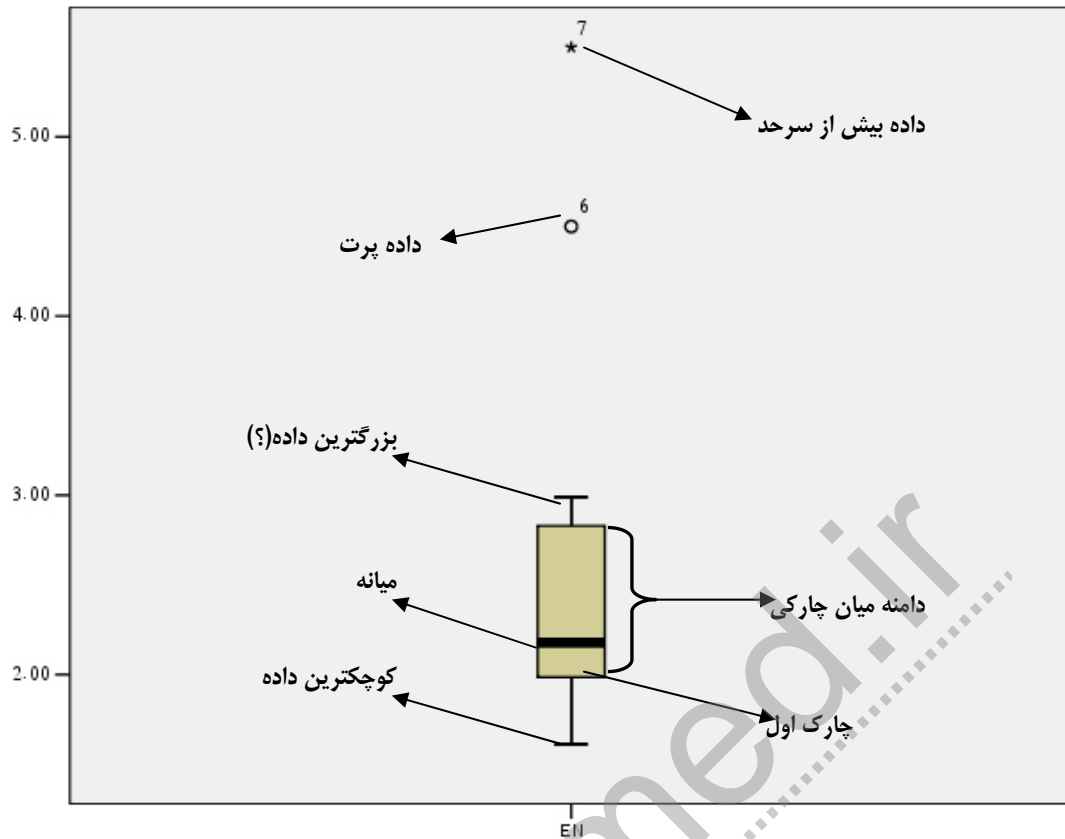


روی دکمه Option... کلیک کنید و شاخصهای موردنظرتان را انتخاب کنید:

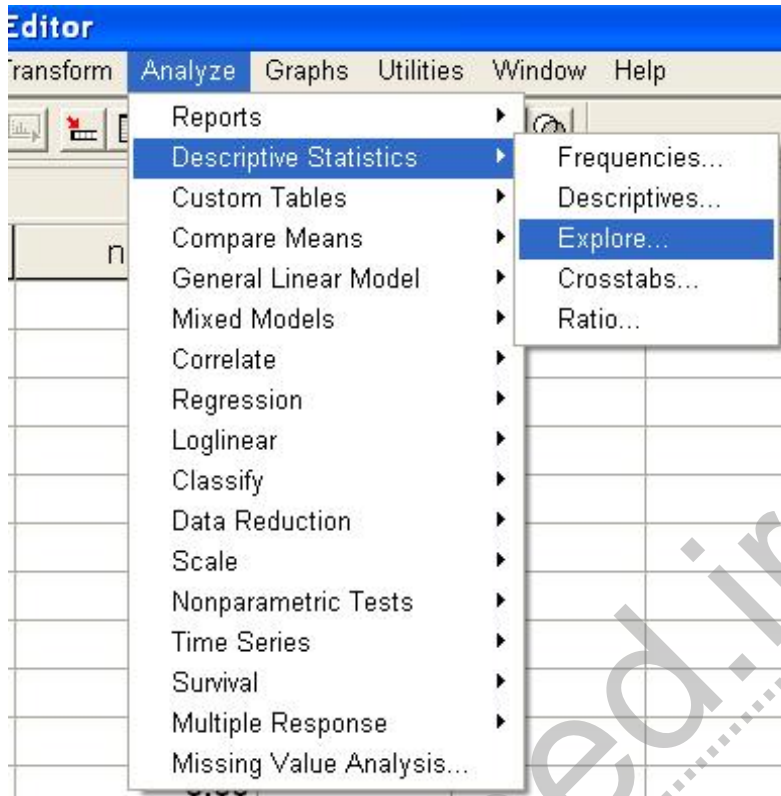


نمودار Box plot

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

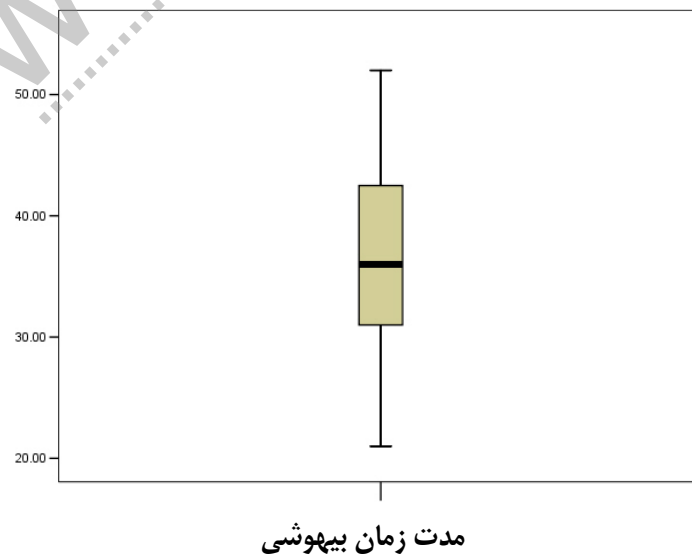


فاصله میان چارک اول و چارک سوم را دامنه میان چارکی (IQR) می گویند. اگر داده‌ای بیش از ۱.۵ برابر دامنه میان چارکی از چارک اول یا چارک سوم فاصله داشته باشد، داده پرت محسوب می‌شود و به صورت یک نقطه خارج از نمودار نشان داده می‌شود و نمودار بدون احتساب آن رسم می‌شود. برای رسم Box plot مسیر زیر را طی کنید:



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

نمودار جعبه‌ای داده‌های مثال مدت زمان بیهوشی در زیر رسم شده است. نحوه تقارن داده ها، میزان پراکندگی و وجود یا عدم وجود داده های پرت را بخوبی میتوان در نمودار مشاهده کرد:



به عنوان تمرین: فایل Employee data را باز کنید و متغیرهای کمی این فایل را با استفاده از شاخص‌های آماری و نمودارهای مناسب توصیف کنید.

به عنوان تمرین: فایل Breast cancer survival را باز کنید و متغیرهای کمی این فایل را با استفاده از شاخص‌های آماری و نمودارهای مناسب توصیف کنید.

چند نمودار مهم و کاربردی دیگر:

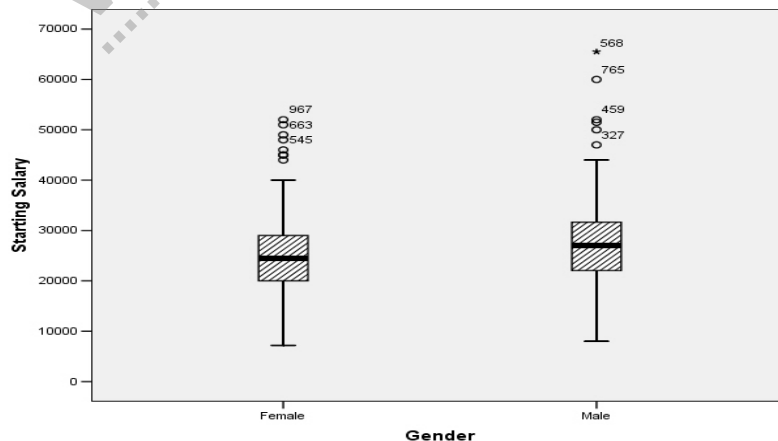
علاوه بر نمودارهای فوق که با هم بررسی کردیم، چند نمودار دیگر را نیز در زیر بررسی می‌کنیم. این نمودارها زمانی مورد استفاده قرار می‌گیرند که ما با دو یا سه متغیر سروکار داشته باشیم، به طوری که دقیقاً یکی از آنها از نوع عددی و دیگر متغیر(ها) کیفی باشند. برای مثال می‌خواهیم حقوق (متغیر عددی) زنان و مردان (متغیر کیفی) را در یک شغل بخصوص مقایسه کنیم. به عبارتی می‌خواهیم ببینیم حقوق مردان بیشتر است یا زنان؟ در این قسمت شما را با این نمودارها آشنا کرده و نحوه رسم آنها را نیز در کلاس با هم بررسی می‌کنیم.

تذکر: برای حل این مثالها فایل داده‌ها را با نام **University of Florida graduate salaries** را از پوشه **SPSSSEVAL** باز کنید.

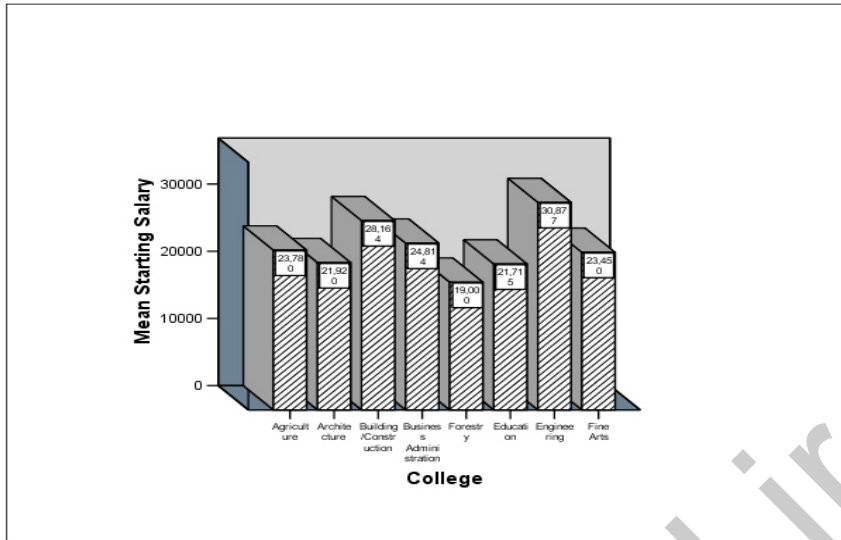
۱) زمانی که تنها یک متغیر عددی و یک متغیر رشته‌ای داشته باشیم:

می‌خواهیم حقوق مردان و زنان را با هم مقایسه کنیم. برای این مقایسه می‌توانیم از دو نمودار زیر استفاده کنیم:

الف) نمودار Box plot (برای دو جامعه):



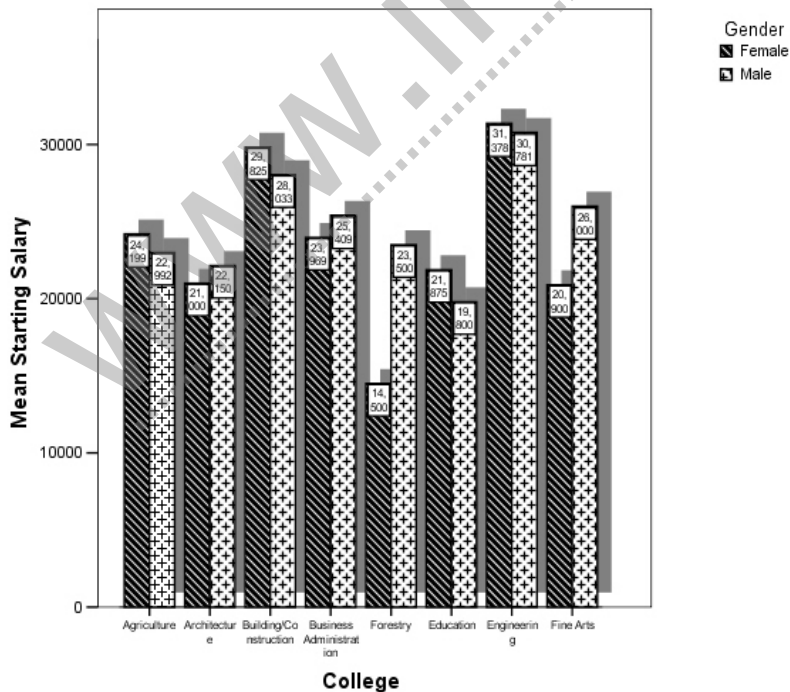
ب) نمودار ستونی (برای مقایسه دو جامعه) :



۲) زمانی که تنها یک متغیر عددی و دو متغیر کیفی داشته باشیم:

الف) فرض کنید می‌خواهیم حقوق زنان و مردان را در هر دانشکده با هم مقایسه کنیم. در این حالت از

نمودار ستونی خوشه‌ای استفاده می‌کنیم:



ب) حال فرض کنید نمودار بالا را بخواهیم بصورت جدا برای زن و مرد رسم کنیم:

