کارگاه روش های آماری در تحلیل داده ها

دانشگاه علوم پزشکی خراسان شمالی

نمایی از باغات روستای رویین

آبان ماه ۱۳۹۵

کارگاه مقدماتی

محاسبات آماری بوسیله SPSS

درس اول

مقدمه

SPSS نرم افزار تجزیه و تحلیل آماری و مدیریت داده ها می باشد. این نرم افزار در ابتدا برای پژوهشگران علوم انسانی ابداع شد (SPSS مخفف Statistical Package for the Social Sciences است). در حال حاضر با گسترش چشمگیر SPSS، این نرم افزار یکی از کاربردی ترین نرم افزارهای آماری شده است. گرچه قابلیت های تجزیه و تحلیل اطلاعات در SPSS در بسیاری از جهات از نرم افزارهای قدرتمند آماری همچون SAS و S-Plus کمتر می باشد اما سادگی نصب و تسهیلاتی که برای ورود اطلاعات و تجزیه و تحلیل آنها در SPSS وجود دارد باعث استفاده بیشتر این نرم افزار در عرصه کاربرد شده است.

هدف اصلی یک نرم افزار آماری، بکارگیری روش های آماری برای تجزیه و تحلیل اطلاعات می باشد. در کنار این هدف مهم، تولید انواع نمودار به عنوان ابزاری برای انتقال آسان و سریع اطلاعات در این نرم افزارها مورد توجه است. با این وجود برای رسم نمودارهای آماری مناسب، می توان از نرم افزارهای گزارش گیری همچنون Excell که قابلیت ویژه ای در تولید انواع نمودارها دارند استفاده نمود.

شروع کار با SPSS

چنانچه نرم افزار SPSS روی ویندوز نصب شده باشد با دوبار کلیک (فشار دگمه سمت چپ موس) روی آیکون (نشانه) SPSS، پنجره ای به نام «ویرایشگر داده ها» (SPSS Data Ediotr)، ظاهر می گردد. اگر آیکون SPSS روی Desktop وجود نداشته باشد با اجرای مسیر زیر SPSS فراخوانی می شود.

Start→Programs→SPSS Inc

ویرایشگر داده ها (SPSS Data Editor)، پنجره اصلی در SPSS بوده که کلیه عملیات ورود اطلاعات و اجرای فرامین مربوط به کارهای آماری و تولید نمودار در آن انجام می گیرد. علاوه بر این پنجره،

پنجره های "خروجی" (Output) "دستوری" (Syntax) و "ویرایشگر نمودار" (Chart Editor) در SPSS وجود دارند. نتیجه اجرای کلیه فرامین در پنجره خروجی ظاهر می شود.

پنجره ویرایشگر داده ها (SPSS Data Editor)

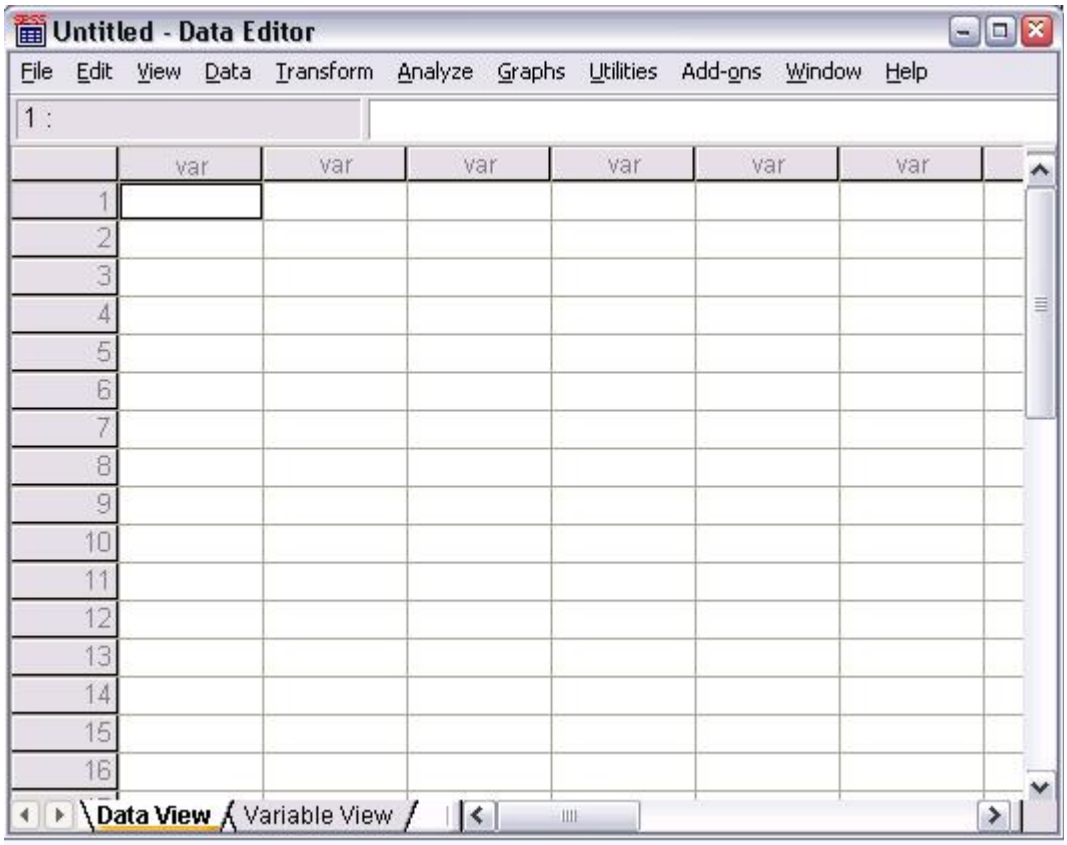
در نوار دستور واقع در قسمت بالای پنجره ویرایشگر داده ها، منوهای File، Edit، View، Data، Transfrom، Analyz، Graphs، Utilities ، Window،add-ons ، Help قرار دارند. منو های Data، Transfrom، Analyz، Graphs منوهای اختصاصی SPSS می باشند. دیگر منوها که در سایر نرم افزارهای تحت ویندوز نیز قابل مشاهده هستند را منوهای عمومی گویند.

- در منوی File، فرمان های مربوط به ایجاد فایل جدید داده ها، فراخوانی یک فایل داده اطلاعات یا خروجی، ذخیره، چاپ و فرمان مربوط ه ورود اطلاعات از سایر نرم افزارهای آماری و گزارش گیری وجود دارد.
- در منوی Edit، فرمان های مربوط به ویرایش اطلاعات (از جمله کپی، جایگزینی و نمایش یک داده خاص) وجود دارد.
- در منوی View، فرمان های مربوط به نمایش اطلاعات وجود دارد.
- در منوی Data، می توان از تسهیلاتی که SPSS برای ورود اطلاعات در اختیار کاربر قرار می دهد استفاده نمود.
- در منوی Transform، فرمان های مربوط به تغییر مقدار یا مقادیر یک متغیر، محاسبه متغیر جدید بر اساس تابعی از متغیرهای موجود، کدبندی متغیر کمکی و یا تغییر کد متغیر کیفی وجود دارد.
- در منوی Analyze، فرمان های لازم برای اجرای کلیه کارهای آماری وجود دارد.
- در منوی Graphs، فرمان های مربوط به تولید نمودارهای گوناگون از جمله انواع نمودارهای ستونی، دایره ای، هیستوگرام، جعبه ای و پراکنش وجود دارد.
- در منوی add-ons برخی از نرم افزار های مرتبط با SPSS در صورت وجود می توان اجرا کرد.

تعریف متغیرها و ورود اطلاعات

پنجره ویرایشگر داده ها شامل دو بخش به نام های "نمایش داده ها" (Data view) و "نمایش متغیرها" (Variable view) می باشد که در هر زمان تنها یکی از آنها در معرض نمایش است. بخش نمایش داده ها، صفحه گسترده ای است که جهت ورود اطلاعات بکار می رود. در بخش نمایش داده ها، مشاهدات هر

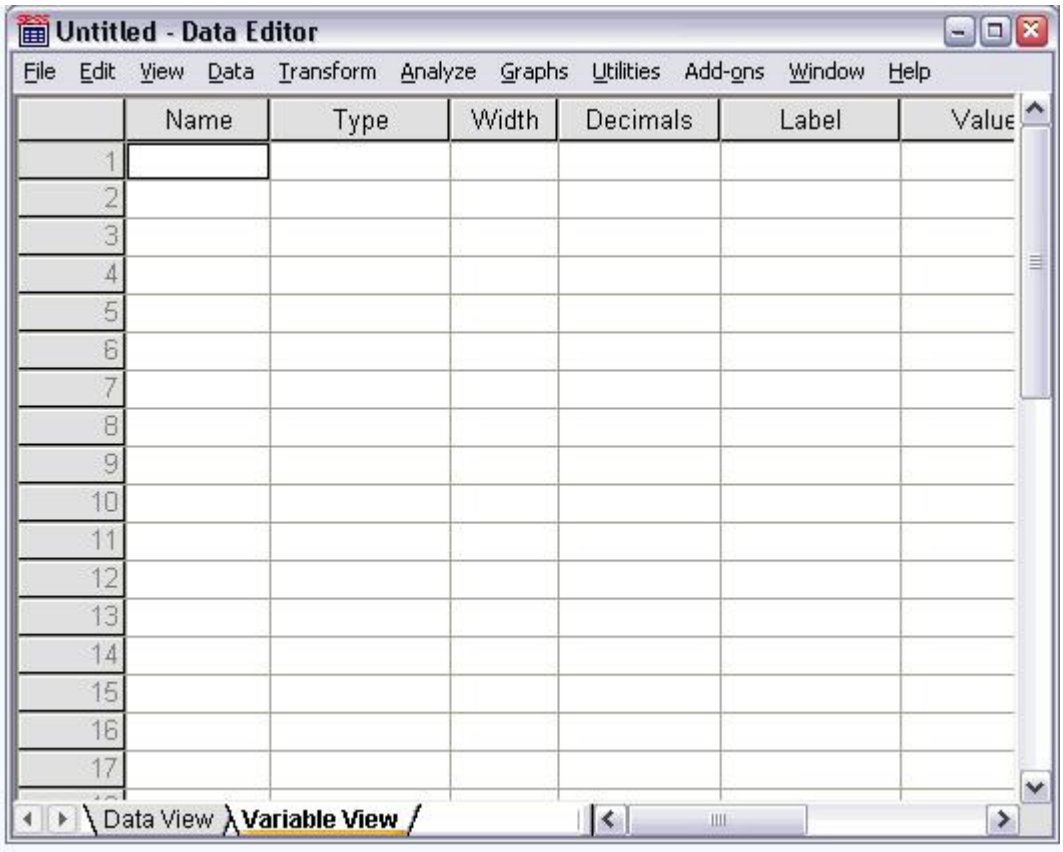
مورد (هر آزمودنی) در یک سطر و مشاهدات هر متغیر در یک ستون قرار می گیرد. نام متغیرها در سر ستون ها و شماره آزمودنی ها در اولین ستون سمت چپ قرار دارند. با وجود این بهتر است در هر تحلیلی، ستون اول نیز به شماره پرسشنامه یا شماره آزمودنی اختصاص یابد. این ستون معمولا با نام ID نام گذاری می گـردد. (شـکل ذیل)



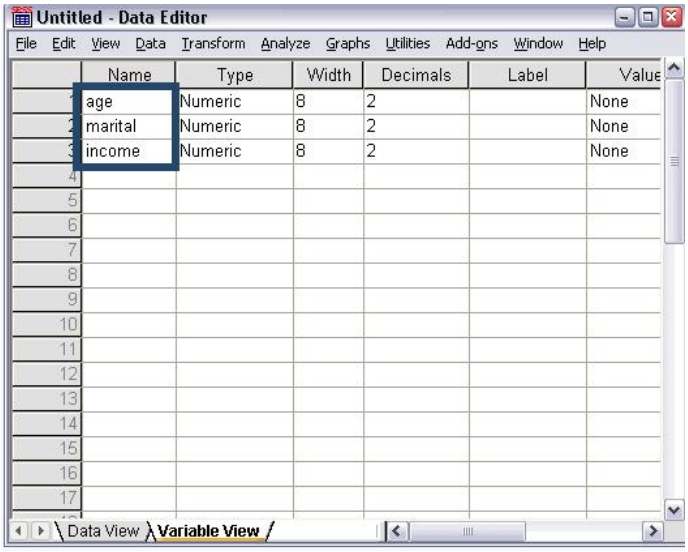
نمایش متغیرها (Variable View)

در بخش نمایش متغیرها،(Variable view) کارهای مربوط به نام گذاری و تعیین مشخصـات متغیرهـا انجام می گیرد.

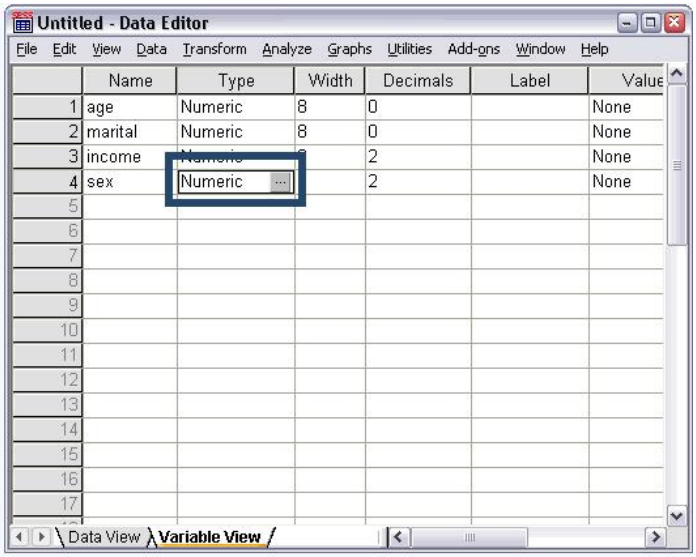
با کلیک کردن روی Variable View به صفحه ای می رویم که در آن صفحه، در ستون اول شماره متغیر و در ستون های بعدی ، نام و مشخصات مربوط به هر متغیر وارد می گردد.



- در قسمت Name، نام متغیر وارد می شود. توجه کنید که نام متغیر در SPSS ۱۱.۵حداکثر از ۸ کاراکتر (حرف) تشکیل شده است برای نام متغیر نمی توان از فضای خالی استفاده کرد. بعضی از کارکترهای موجود در صفحه کلید (مثل ؟ و !) قابل استفاده در نام متغیر نیستند.

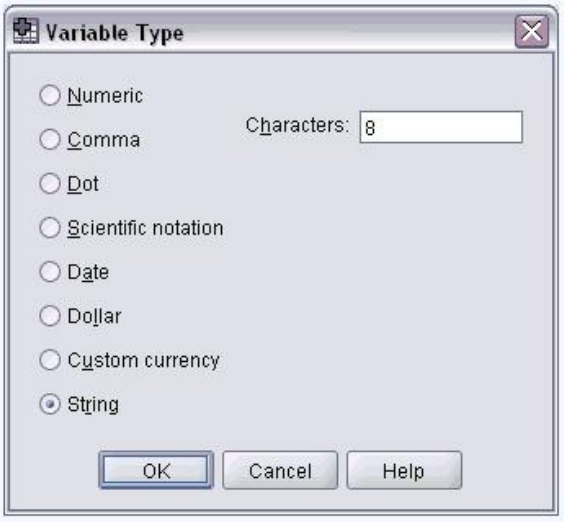


در قسمت Type، لازم است نوع متغیر معلوم شود. پیش فرض SPSS این است که متغیر از نوع عددی (Numeric) است. با کلیک در قسمت Type منوی شامل پیشنهادهای دیگر برای نوع متغیر ظاهر می شود.

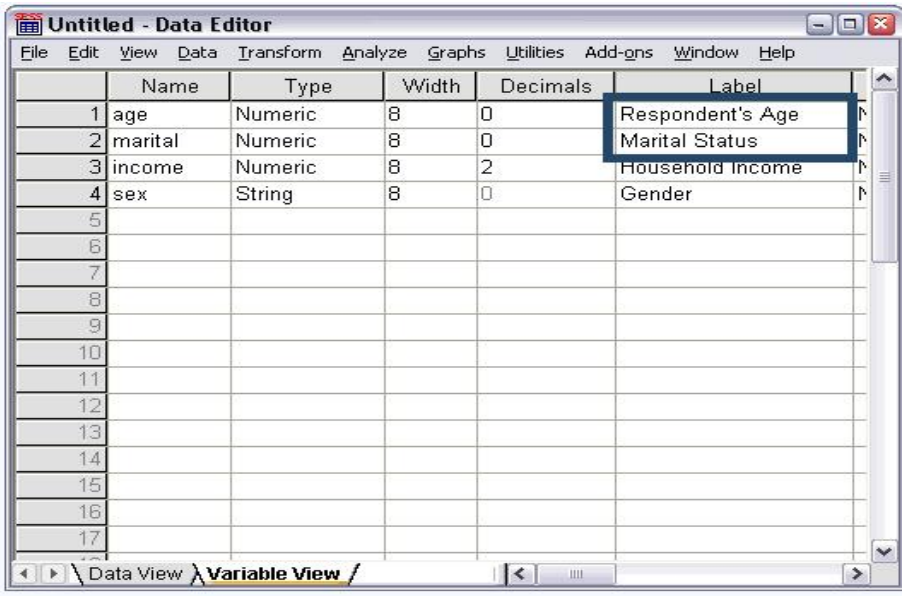


در حالت کلی چنانچه متغیر، نام آزمودنیها (اسامی افراد یا آزمودنی ها) باشد نوع متغیر رشته ای (String) بوده و در سایر موارد می توان نوع عددی را به کار برد. با این حال گزینه های دیگری که در عمل کاربرد کمی دارند پیشنهاد شده است. این گزینه ها عبارتند از گزینه Common، که برای نمایش اعداد چند رقمی که هر سه رقم آنها با یک کاما جدا شود بکار می رود. گزینه Dot، که برای نمایش اعداد چند رقمی که هر سه رقم آنها با یک نقطه جدا شود استفاده می شود. نماد علمی (Scientific notation)، که در این صورت رقم به صورت مضربی از ۱۰ به نمایش در می آید. تاریخی (Date)، که برای نمای اطلاعات

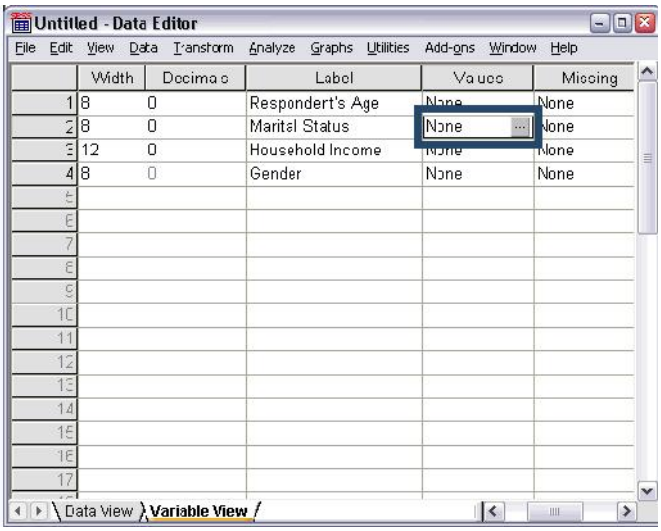
زمانی بکار می رود. دلاری (Dollar)، که برای نمایش اطلاعات هزینه ای بر حسب دلار بکار می رود. سفارشی (Custom Carrancy)، که برای ثبت اطلاعات خاص بکار می رود. رشته ای (String)، که برای ثبت اطلاعات کاراکتری همچون اسامی بکار می رود.

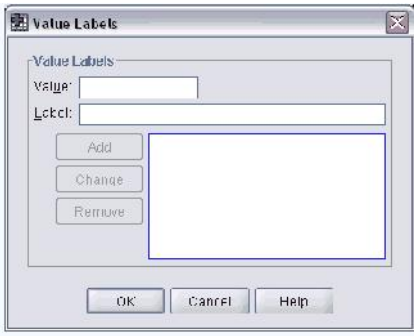


- در بخش اعشاری (Decimal)، تعداد رقم های اعشار متغیر عددی مشخص می گردد. پیش فرض SPSS عدد دو می باشد.
- بخش برچسب (Lable)، چنانچه بخواهیم نام کلی متغیر یا توضیحی در مورد آن ثبت شود، با کلیک در قسمت Lable این کار انجام می گیرد. توجه شود که در صورت درج اطلاعاتی در بخش برچسب، این اطلاعات در خروجی به عنوان نام متغیر ظاهر می شود.

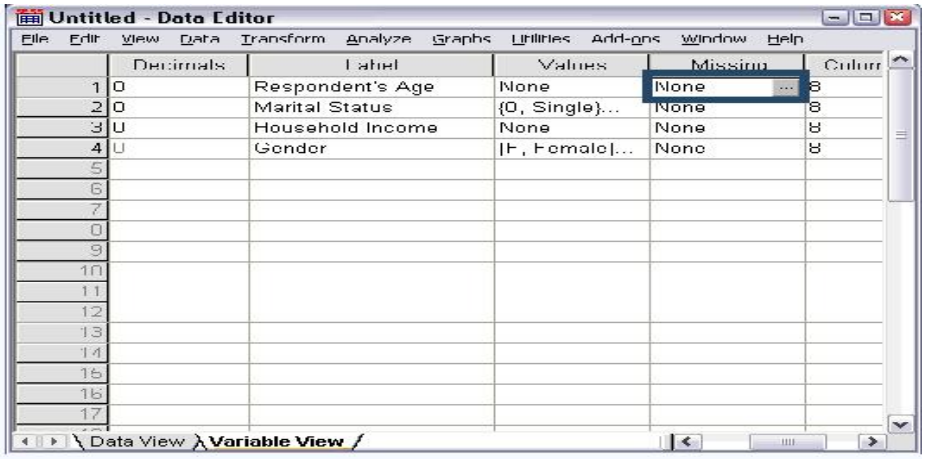


– در بخش مقدار (Value)، برچسب کدهای عددی بکار گرفته شده برای متغیرهای کیفی یا اسمی تعریف می گردند. با کلیک در مربع سمت راست پنجره ای ظاهر می شود که کدها را در قسمت Valueو نام کد را در lable تعریف می کنیم ,سپس add می نماییم. در این صورت در خروجی، برچسب های متغیر ها به جای کدهای عددی گزارش می شوند.





-گم شده (Missing) : اگر لازم باشد مقدار یا مقادیر خاصی از متغیر مورد بررسی به عنوان داده گم شده در نظر گرفته شود آن مقادیر در این قسمت پیشنهاد می شوند.



– برای انجام این کار با کلیک کردن در قسمت گمشده، منوی ظاهر می شود که می توان با فعال نمودن گزینه Discrete missing سه مقدار خاص و با فعال نمودن گزینه Range of missing value دامنه ای از مقادیر (برای متغیرهای پیوسته) و با فعال نمودن گزینه Range pluse one discrete missing value ترکیبی از دو انتخاب فوق را به عنوان داده گمشده پیشنهاد داد.

توجه کنید که مقادیر گمشده در تجزیه و تحلیل اطلاعات مورد استفاده قرار نمی گیرند. علاوه بر این در هنگام ورود اطلاعات، چنانچه مقدار یک متغیر ثبت نشود سیستم SPSS آن را با نقطه نشان خواهد داد که به این نوع گمشده، گمشده سیستمی (System Missing) گویند.

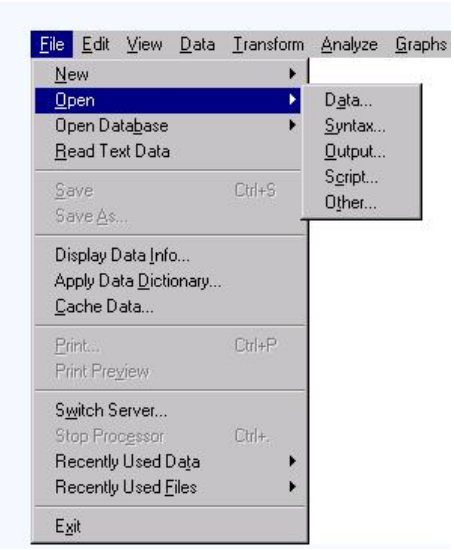
– در بخش ستون (Column)، پهنای هر ستون مشخص می گردد.

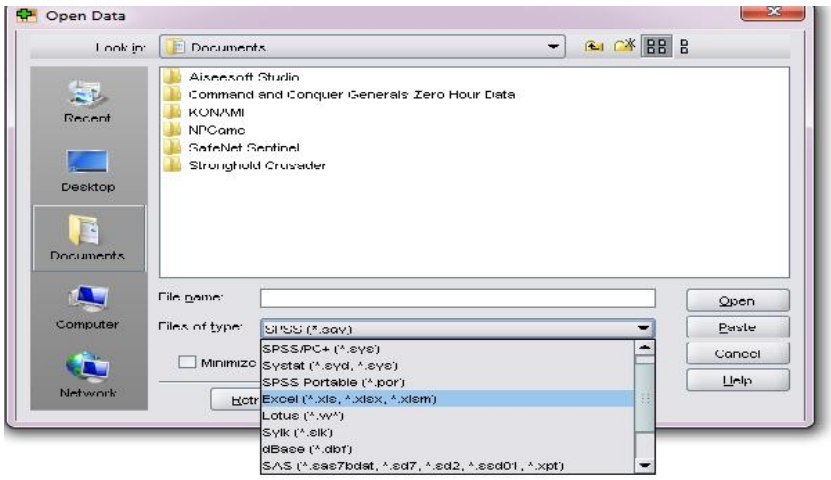
– با کلیک در بخش هم ترازی (Alignment) می توان نحوه قرار گرفتن اطلاعات در هر ستون را که می تواند راست، مرکز یا سمت چپ هر ستون باشد پیشنهاد نمود.

با کلیک در بخش اندازه گیری (Measure)، می توان مقیاس اندازه گیری متغیر را پیشنهاد داد. پیش فرض SPSS، نسبی (Scale) بودن مقیاس اندازه گیری متغیرها است. علاوه بر این می توان مقیاس های رتبه ای (Ordinal) و اسمی (Nominal) را به ترتیب برای متغیرهای کیفی رتبه ای و اسمی پیشنهاد نمود.توصیه می شود مقیاس ها مطابق با پیش فرض SPSS باشد و تغییر داده نشود. مگر اینکه داده ها از نرم افزار های دیگری همچون Excel فراخوانی شده باشد .بهتر است برای متغیر های کیفی که به صورت عددی کد داده شده است و متغیر های کمی ، همان (Scale) باشد.

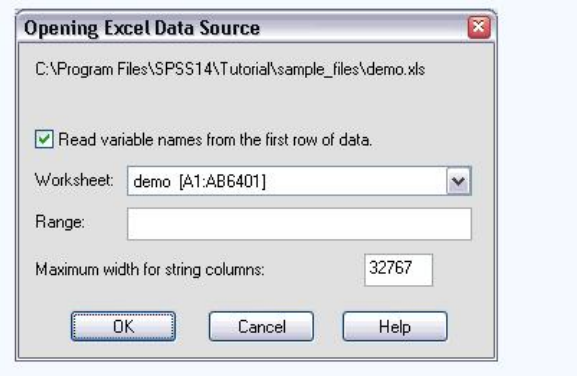
خواندن سایر فایل ها:

چنانچه داده ها با فرمت دیگری بجز SPSS داشتیم براحتی توسط این نرم افزار خوانده می شود. برای این منظور از منوی فایل به open رفته و روی data کلیک می کنیم . پس از تعیین نوع فایل ، مسیر را انتخاب و فایل را Open می نماییم.





با توجه به نوع فایل پنجره های متفاوتی ظاهر می شود . به عنوان مثال اگر فایل EXCEL را باز نماییم پنجره ذیل باز می شود که میتوان محل data را از نظر sheet مربوطه معرفی و سپس OK نماییم.



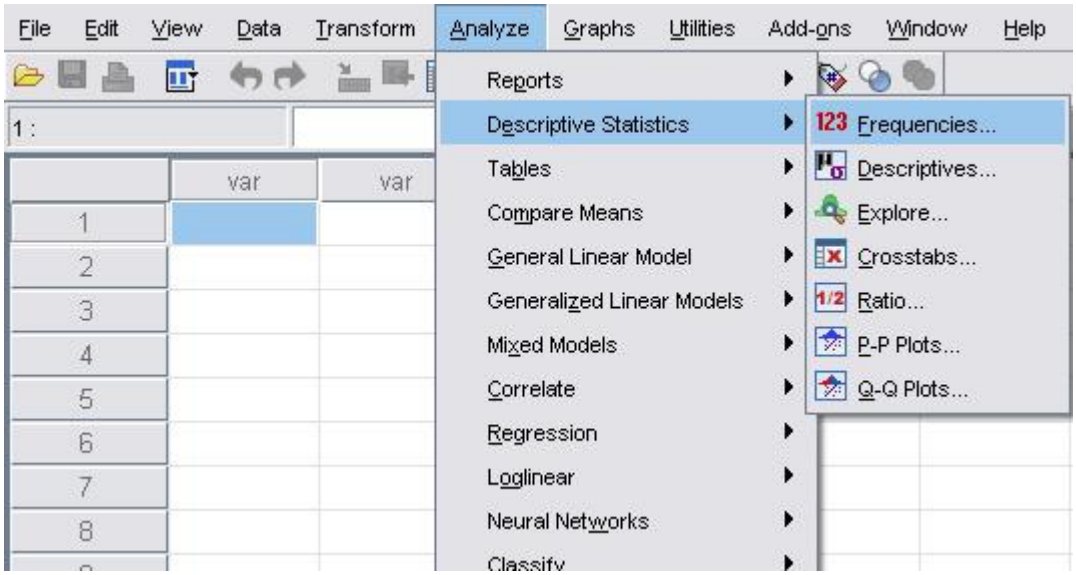
خواندن فایل ها از بانک های اطلاعاتی همچون SQL server و Access و سایر بانک ها نیز برای این نرم افزار میسر است که بدلیل استفاده کمتر از آن از توضیح دادن آن صرفنظر میشود .

توصیف داده ها

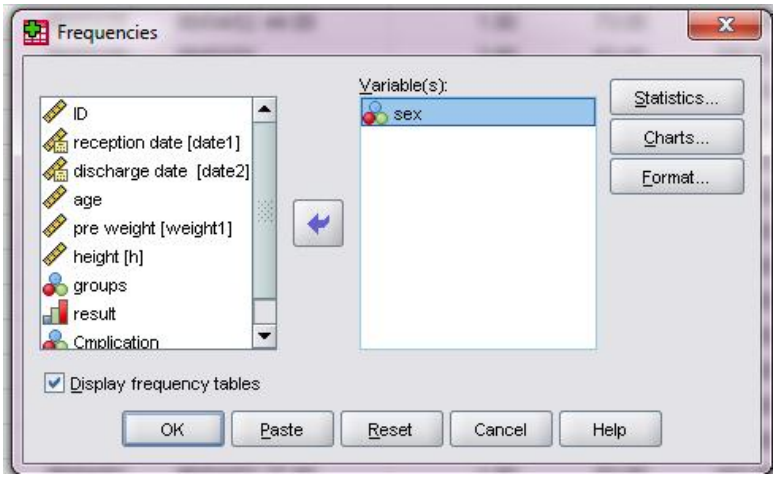
جدول فراوانی:

برای توصیف متغیرهای کیفی (مثل جنس، شغل، تحصیلات، شدت بیماری و ...) می توان جدول توزیع فراوانی هر کدام را که شامل فراوانی (Frequency)، درصد فراوانی نسبی (Percent)، درصد فراوانی نسبی معتبر (Valid percent)، و فراوانی تجمعی (Cumulative percent)، می باشد را محاسبه نمود.

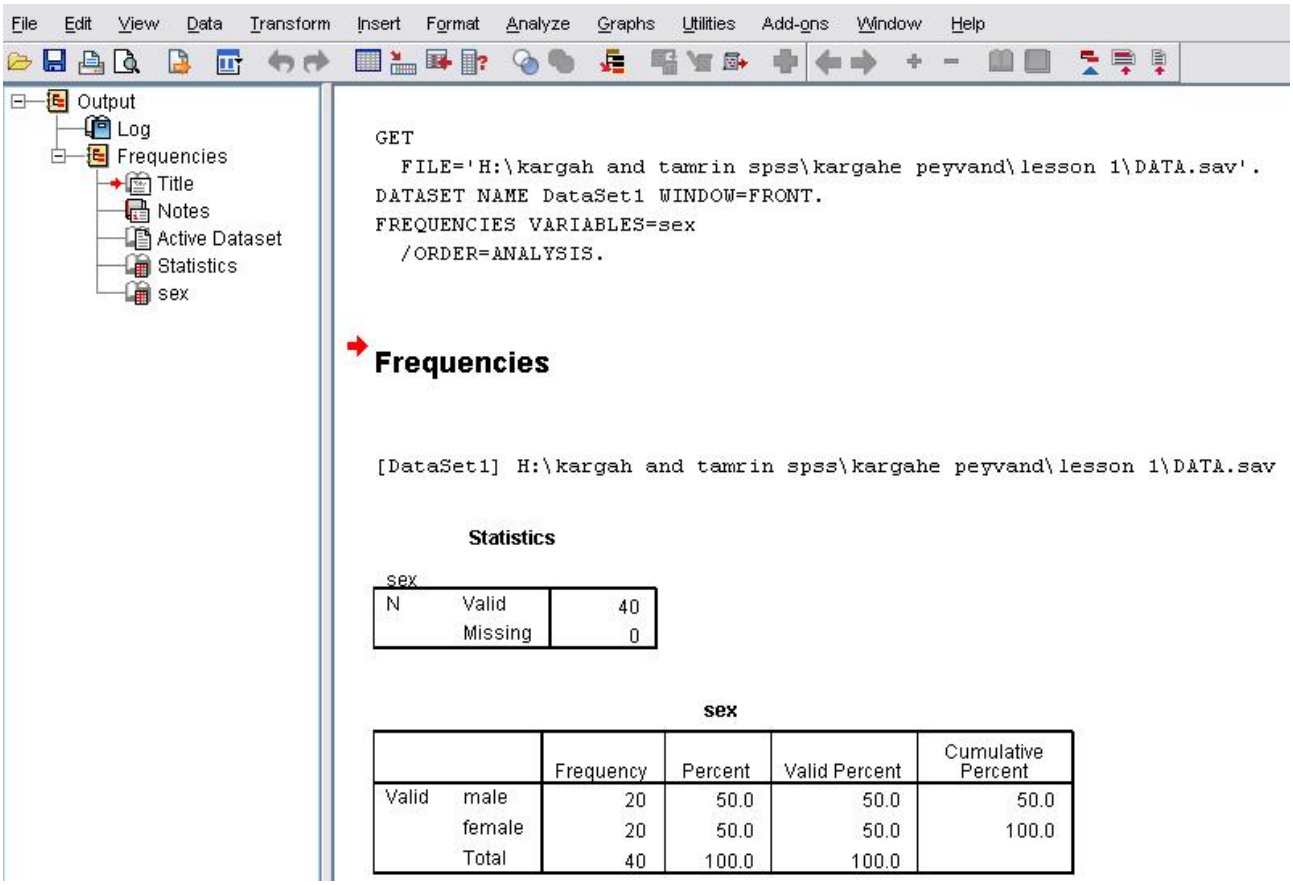
برای انجام این کار در SPSS لازم است مسیر زیر را انتخاب نمود.



با این کار منوی Frequency ظاهر گشته پس از انتخاب متغیرهای مورد نظر از لیست متغیرها OK را کلیک می نمائیم.

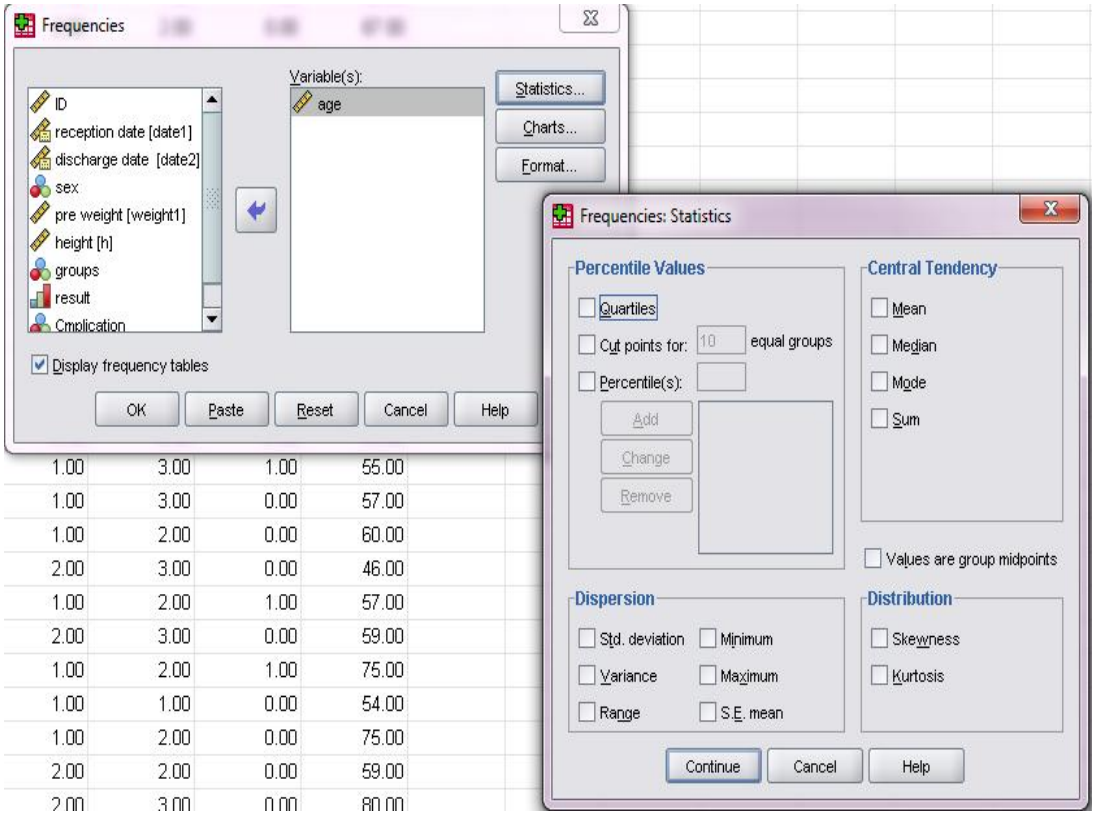


وخروجی به صورت ذیل می باشد.



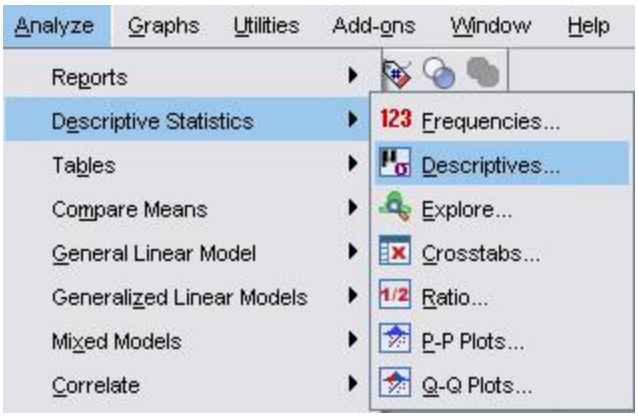
تذکر: گاهی اوقات برخی مقادیر برای یک متغیر نامعلوم است یا اصطلاحا دارای مقادیر گمشده (Missing Value) می باشیم، در این صورت از (Valid percent) برای تعیین درصد فراوانی استفاده می کنیم که درصد موارد معلوم را محاسبه می کند.

برای محاسبه شاخص های آماری می توان از مسیر فوق استفاده نمود و با انتخاب گزینه Statistics در این منو شاخص های آماری همچون میانگین، میانه، صدک ها، چندک ها و کشیدگی را نیز بدست آورد. برای محاسبه شاخص ها برای متغیر کمی، از این منو کمتر استفاده می شود.

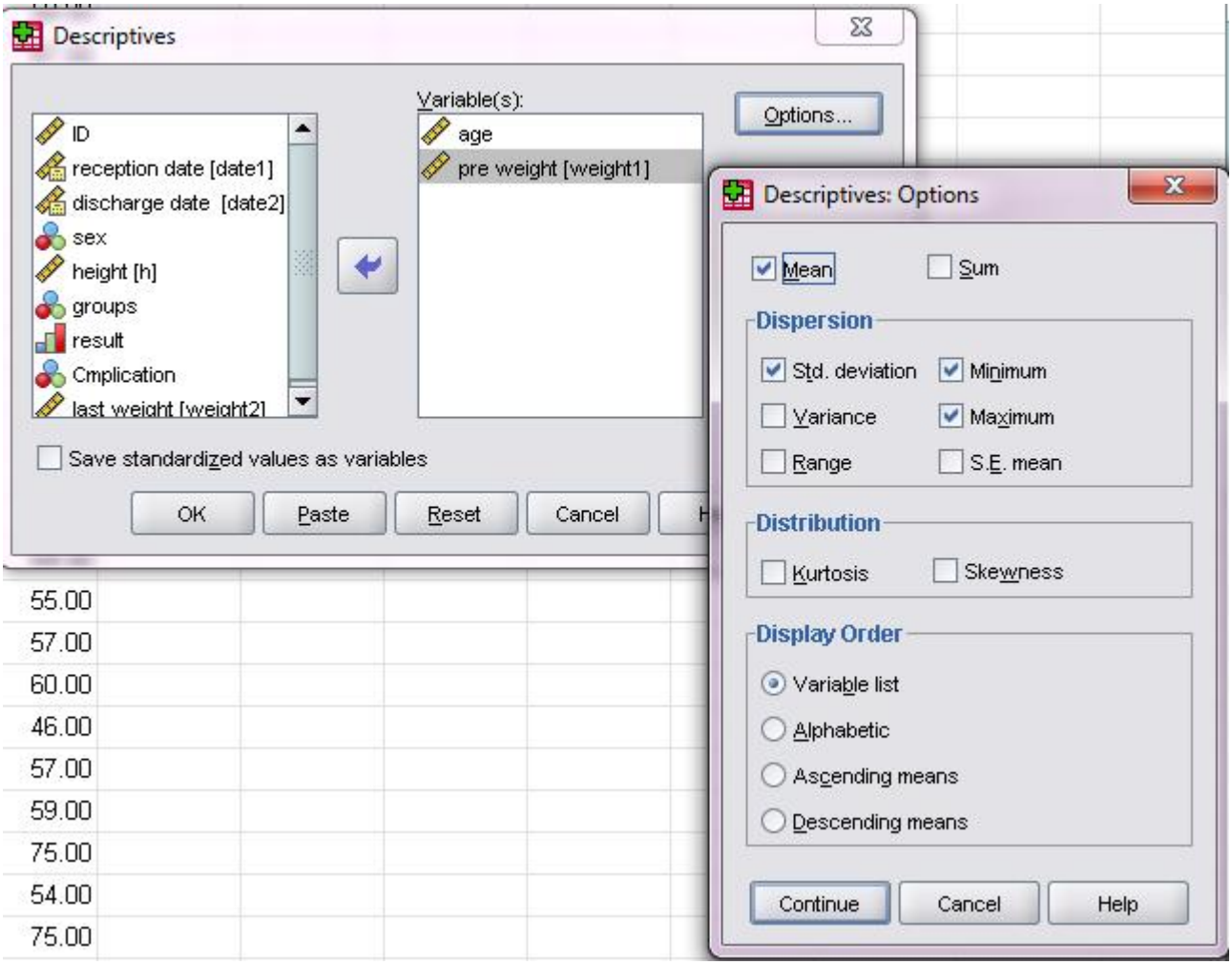


توصیف متغیرهای کمی

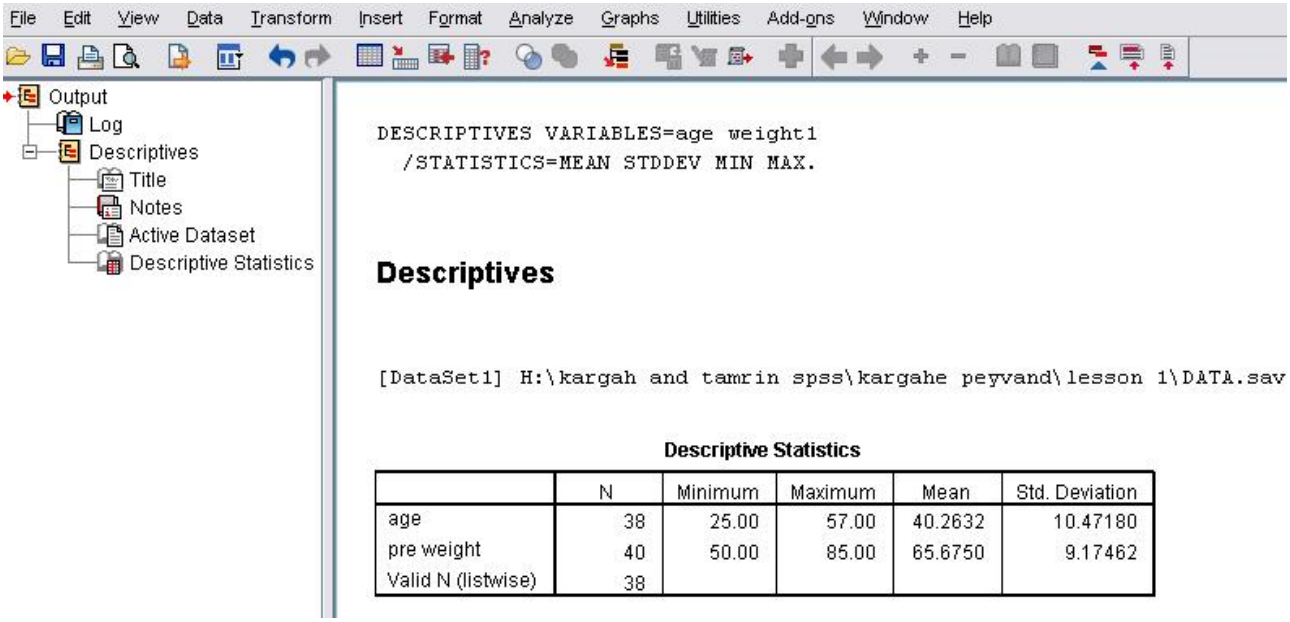
برای متغیرهای کمی (سن، وزن، قد، BMI، فشار خون سیستولیک و ...) می توان یکی از شاخص های متمایل به مرکز (Central Tendency) یعنی میانگین (Mean)، و شاخص های پراکندگی (Dispersion) شامل: دامنه (Range)، یا واریانس (Variance)، انحراف معیار (Standard Deviation) و خطای معیار (Standard Deviation of Mean) یا (S.E.Mean) را محاسبه کرد. همچنین در این منو می توان شاخص هایی چون حداکثر (Maximum)، حداقل (Minimum)، کشیدگی توزیع (Kurtusis) و چولگی توزیع (Skwness) را بدست آورد. برای محاسبه مقادیر فوق مسیر زیر انتخاب می شود.



با انتخاب مسیر فوق، منو ظاهر شده و متغیرهای مورد نیاز را از لیست متغیرها انتخاب مـی نمـائیم. بـا انتخاب OK برای متغیر یا متغیرهای انتخاب شده تعداد، حداقل، حداکثر، میانگین و انحراف معیـار ظـاهر مـی شود برای محاسبه سایر شاخص ها از گزینه Option در منوی Descriptive استفاده می نمائیم.

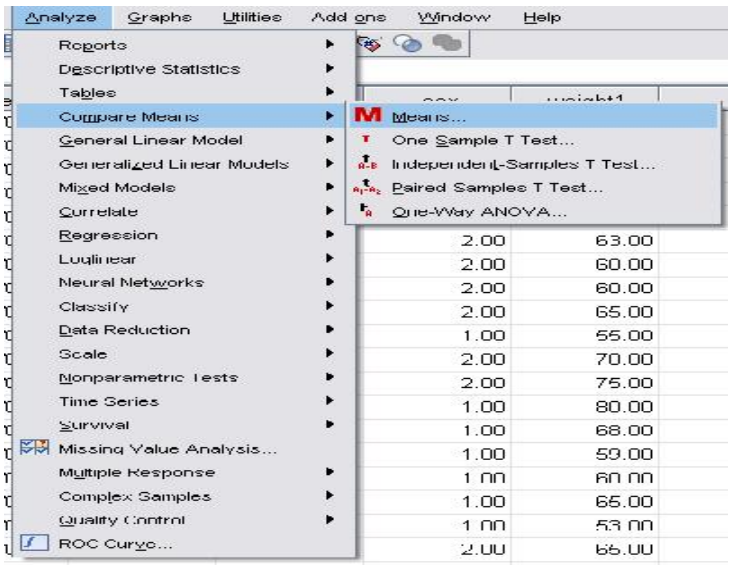


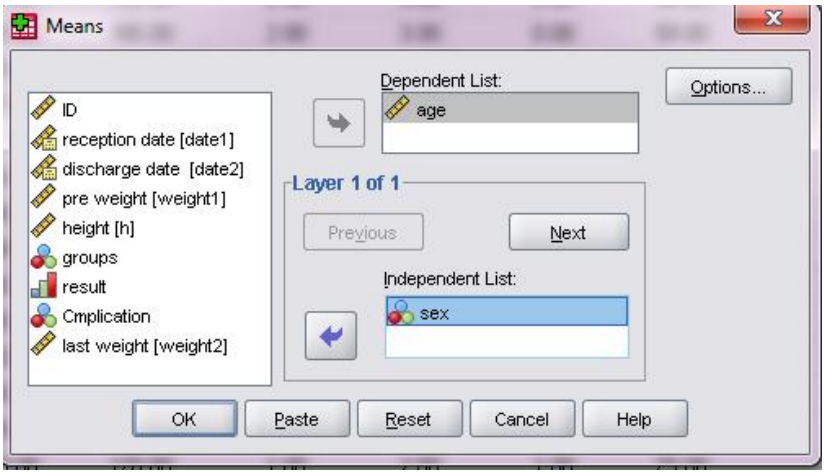
پس از کلیک کردن روی continue و سپس OK خروجی به صورت ذیل ظاهر می شود. کـه در آن مقـادیر درخواست شده ، شامل تعداد، دامنه ،حداقل، حداکثر، میانگین و انحراف معیار آمده است.



به دست آوردن شاخص های آماری برحسب گروه ها:

اگر بخواهیم شاخص های آماری را برای زیر گروه های جمعیـت و لایـه هـای مختلـف بـه دست آوریم ، از منوی analyze به compare mean می رویم .





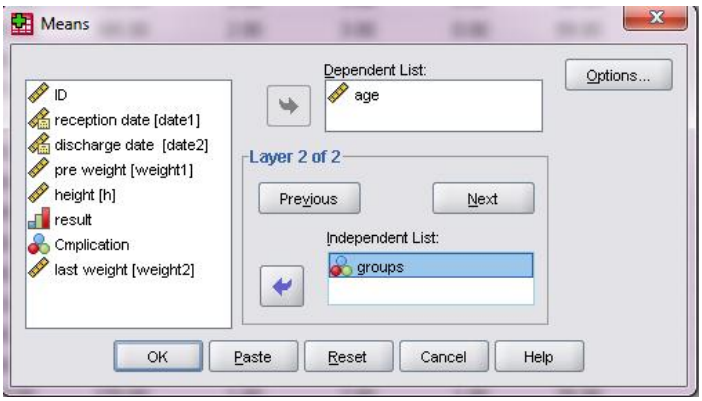
و خروجی به صورت ذیل می شود.

Report

age

sex	Mean	N	Std. Deviation
male	٤١.٩٤٧٤	١٩	١٠.٩٣٦٥٥
female	٣٨.٥٧٨٩	١٩	٩.٩٩٠٦٤
Total	٤٠.٢٦٣٢	٣٨	١٠.٤٧١٨٠

و در لایه های دیگرنیز می توان شاخص ها را به دست آورد مثلا متغیر sex را با کلیک روی next به داخل برده و متغیر بعدی مثل گروه را به سمت راست منتقل می کنیم .



و خروجی به صورت ذیل در می آید.

Report

age

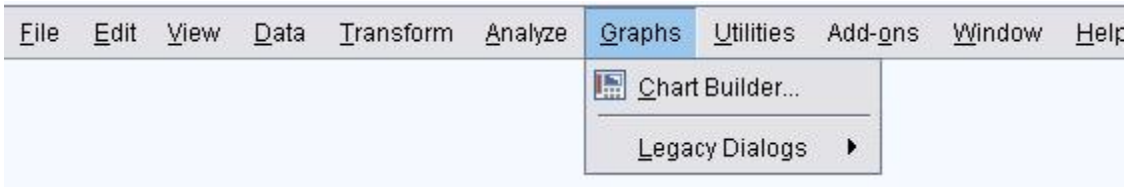
sex	groups	Mean	N	Std. Deviation
male	a	٤٤.١٨١٨	١١	٧.٧٣٠٦٩
	b	٣٨.٨٧٥٠	٨	١٤.٢٦٧٢٢
	Total	٤١.٩٤٧٤	١٩	١٠.٩٣٦٥٥
female	a	٤٠.٥٧١٤	٧	١٠.٢٤٤٦٣
	b	٣٧.٤١٦٧	١٢	١٠.١٠٣٦٣
	Total	٣٨.٥٧٨٩	١٩	٩.٩٩٠٦٤
Total	a	٤٢.٧٧٧٨	١٨	٨.٦٨٧٧٦
	b	٣٨.٠٠٠٠	٢٠	١١.٦٠٣٠٨
	Total	٤٠.٢٦٣٢	٣٨	١٠.٤٧١٨٠

رسم نمودار

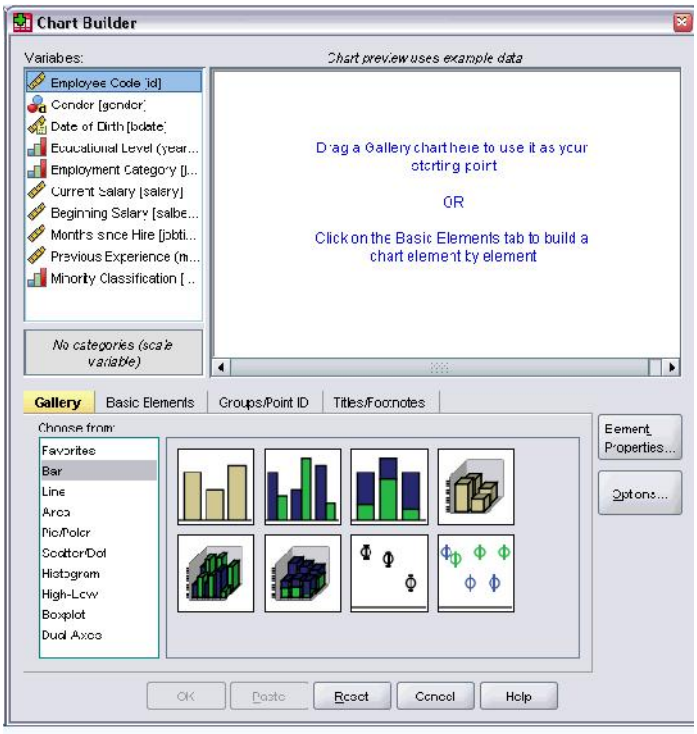
همانطور که قبلا گفته شد، نرم افزار SPSS جهت تجزیه و تحلیل داده ها بکار برده می شود اما می توان برخی نمودارها را نیز توسط آن رسم نمود. برای رسم نمودار در منوی Graph انواع نمودارها قرار داده شده است.

نمودار ستونی (Bar Chart):

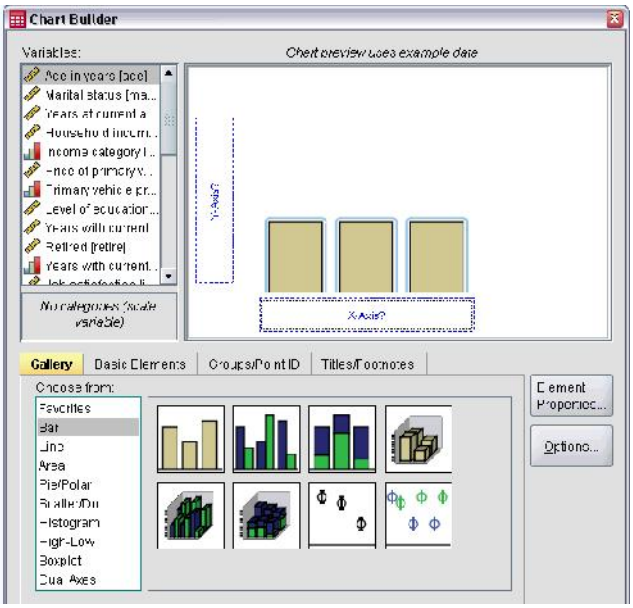
یکی از نمودارهایی است که برای متغیرهای کیفی بکار گرفته می شود.از منوی Graph روی Chart builder کلیک می کنیم.



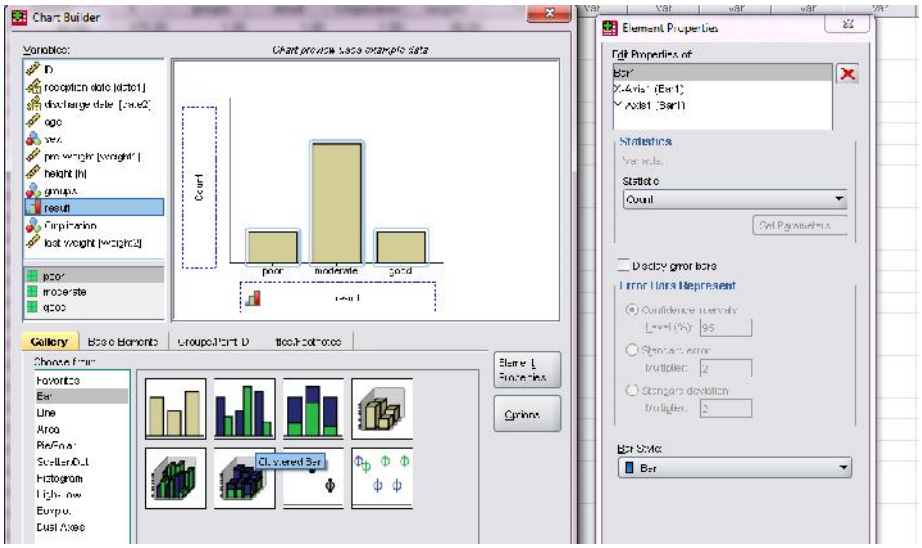
پنجره ذیل ظاهر می شود.



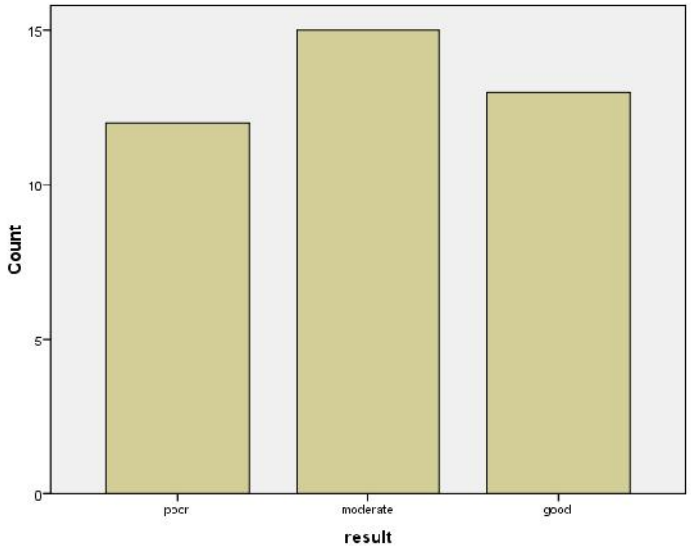
با انجام کلیک راست روی نمودار مورد نظر و انتخاب copy chart to convas پنجره ذیل ظاهر می شود



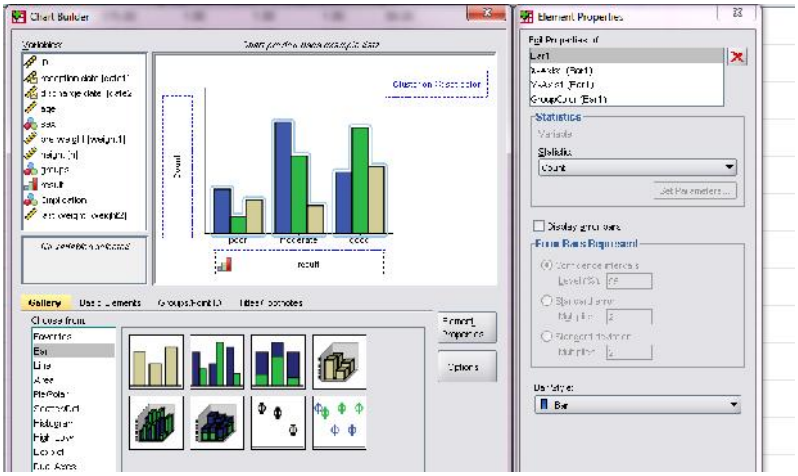
با انتخاب متغیر و انتقال به متغیر X نمودار ستونی رسم می شود.



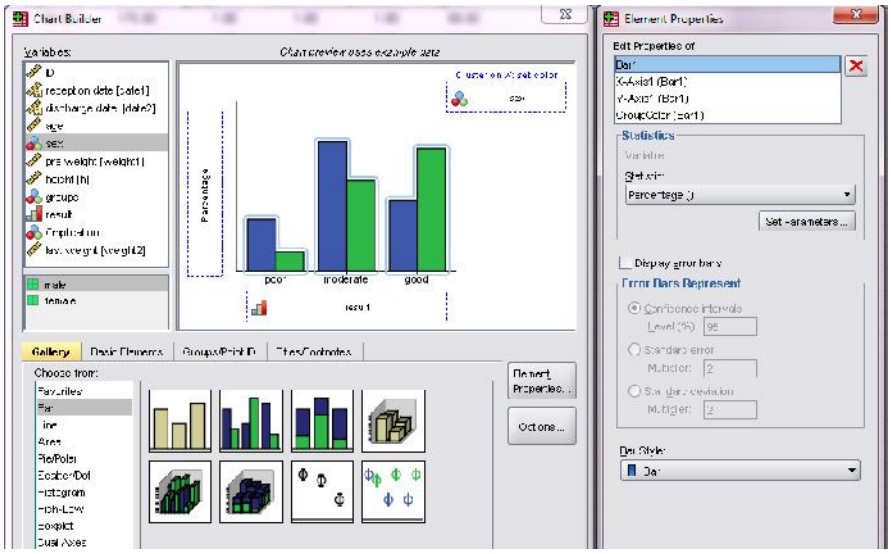
خروجی به صورت ذیل خواهد بود.



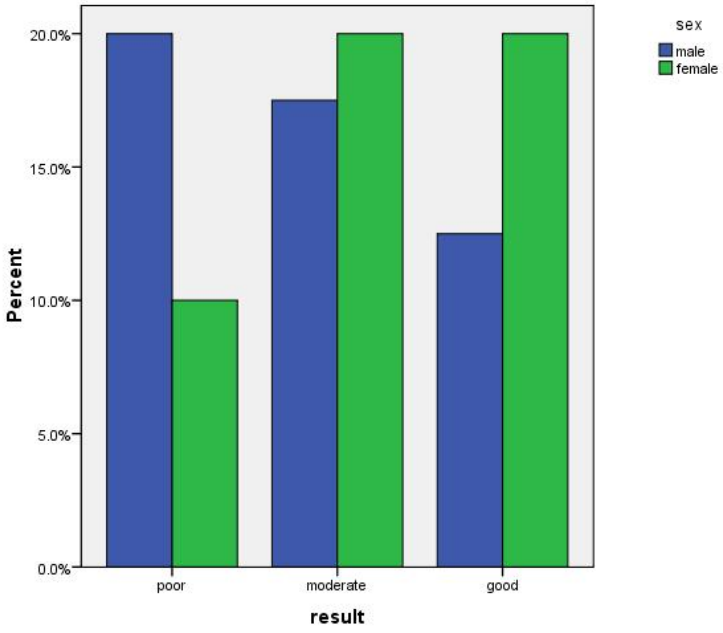
در Bar Chart می توان نمودار دو گروه را کنار هم قرار داد تا امکان مقایسه فراهم شود



یک متغیر در محور X ها و متغیر دیگر را در قسمت cluster (گوشه سمت راست بـالا) قـرار مـی دهیم .



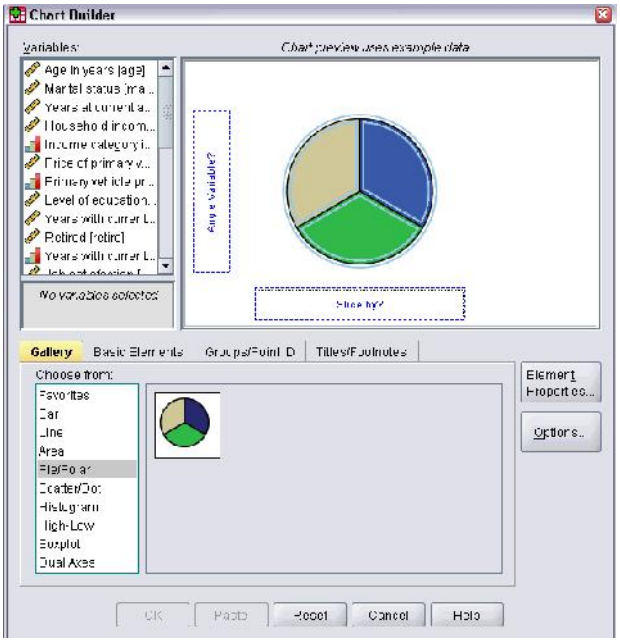
خروجی به شکل ذیل می باشد.

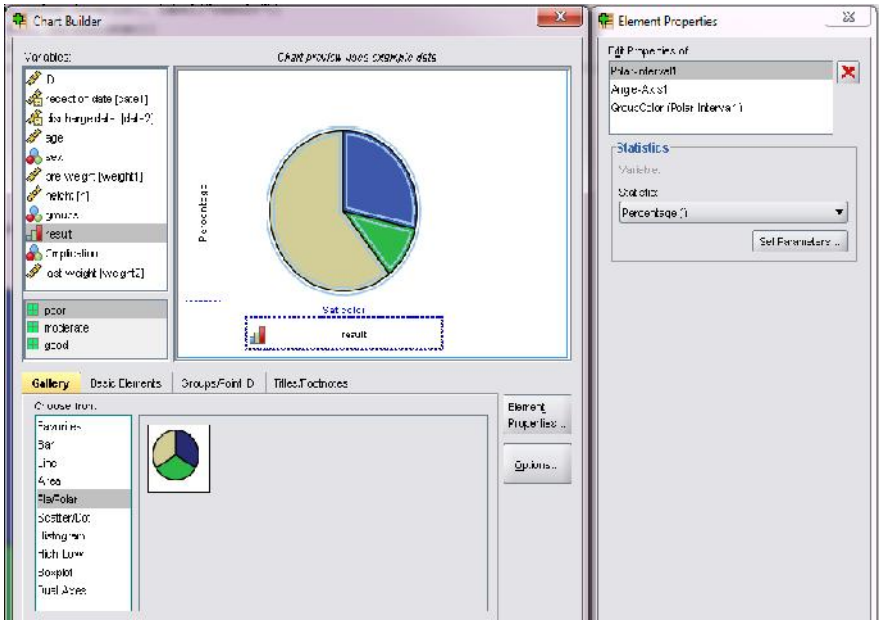


نمودار دایره ای (Pie Chart):

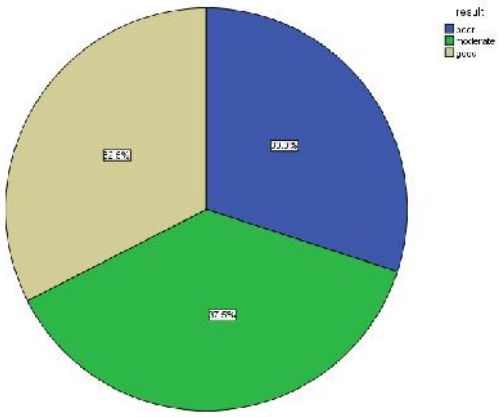
معمولا برای متغیرهای کیفی استفاده می شود. از قسمت Pie به

پنجره مربوطه وارد می شویم.



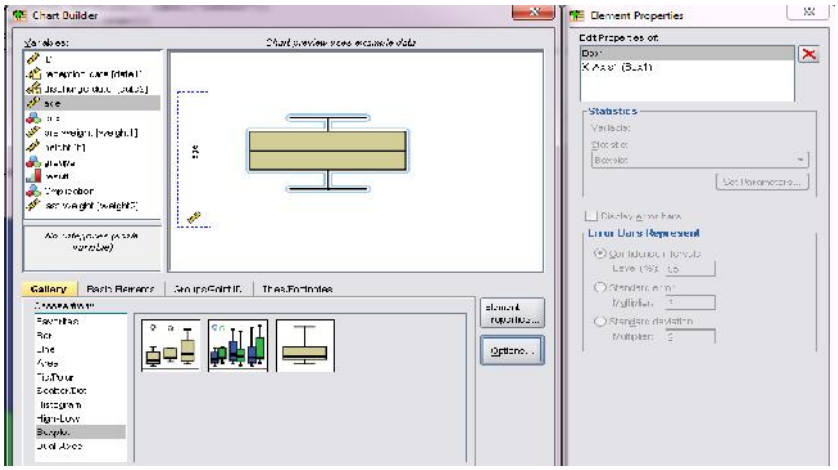


و نتیجه به صورت ذیل می باشد.

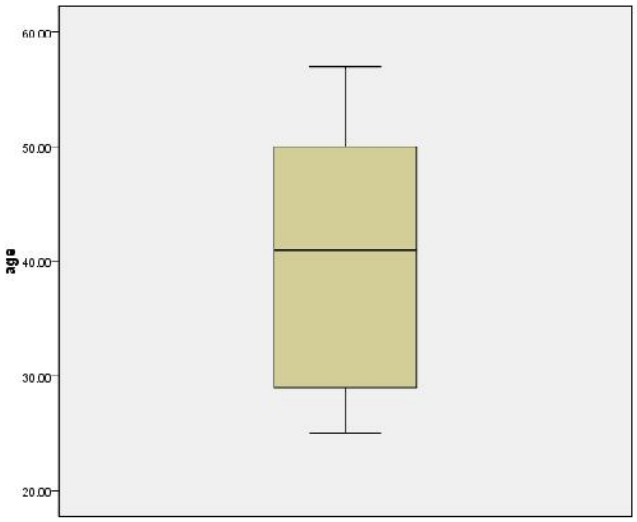


نمودار جعبه ای (Box plot):

این نمودار اغلب در مواردی رسم می گردد که متغیر دارای توزیع نرمال نیســت.در نمــودار حداقل ،حداکثر، میانه ،چارک اول (صدک ۲۵) و چارک سوم (صدک۷۵) نشان داده می شود.

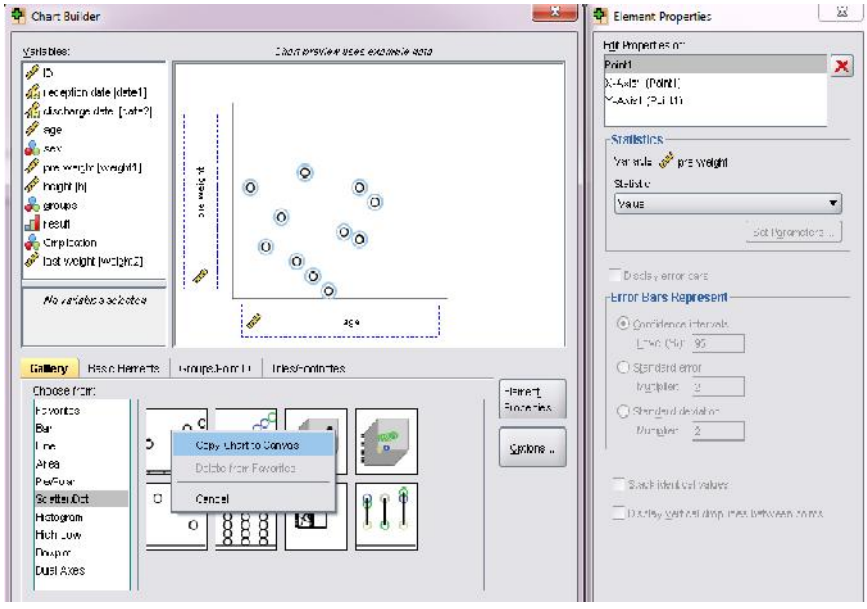


این نمودار را برای سن افراد رسم کرده ایم.

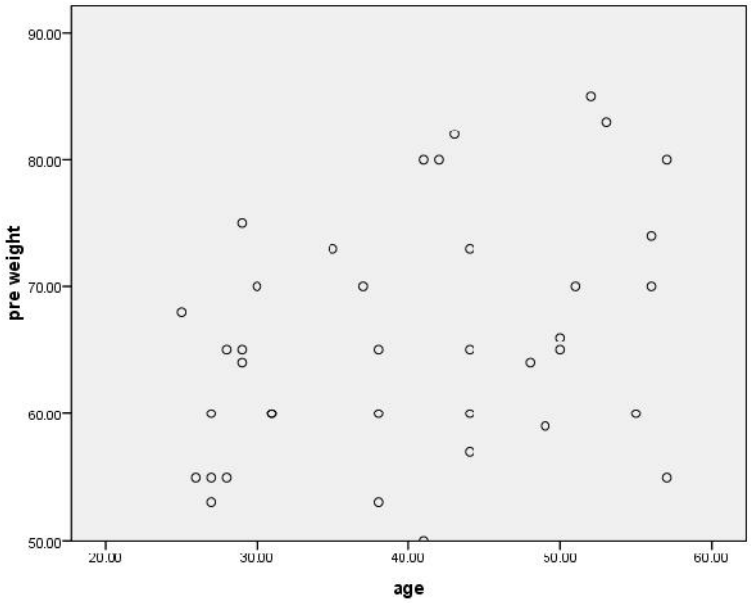


نمودار پراکنش :

این نمودار برای نشان دادن همبستگی بین دو متغیر کمی رسم می شود. برای رسم نمودار پراکنش، گزینه Scatter از منوی Graphs را انتخاب می نماییم.



در قسمت خروجی نمودار ذیل به دست می آید.

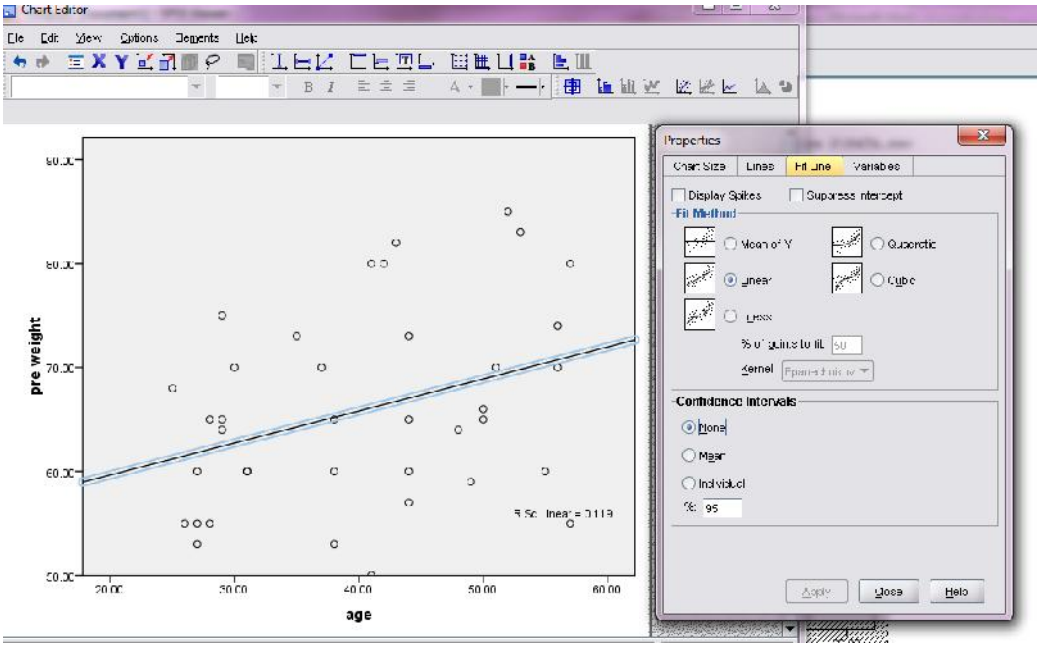


سایر نمودارها همچون نمودار مستطیلی (هیستوگرام)، نمودار چندضلعی و ... را می توان رسم نمود.

ویرایشگر نمودار:

برای ویرایش نمودن نمودارها، با انجام دو بار کلیک روی نمودار به قسمت ویرایشگر نمودار می رویم. تغییر رنگ، سه بعدی نمودن، تغییرات در فونت، تغییرات در عنوان در این قسمت انجام می شود. سپس ویرایشگر نمودار را می بندیم ☒ تغییرات انجام شده در فایل Output قرار می گیرد.

به عنوان مثال برای رسم خط رگرسیون ، پس از دو بار کلیک روی Scatter plot پنجره زیر ظاهر می شود.



تذکر: توصیه می شود رسم نمودار با استفاده از نرم افزارهای دیگر مثل Excell صورت پذیرد.

مدیریت داده ها (Data Management):

۱- کدگذاری مجدد (recode)

۲- محاسبات (ایجاد متغیر مرکب)

۳- فیلتر کردن (select, split)

۴- مباحث پیشرفته

- محاسبات متغیرهای تاریخ
- شمارش تعداد سوال های پاسخ داده شده
- ادغام فایل ها (merge files)-

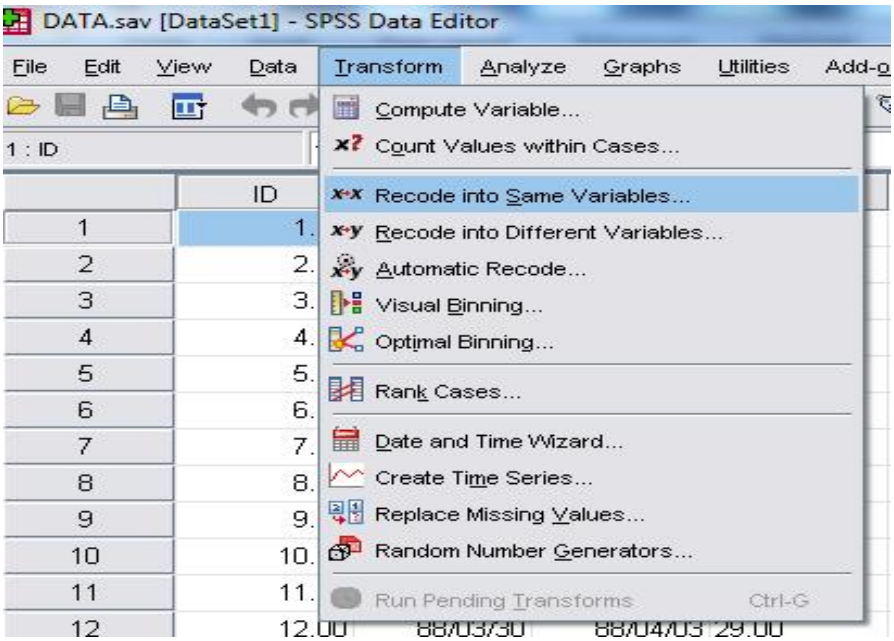
۱=کدگذاری مجدد :

در بسیاری از موارد لازم است که متغیرهای کمی دسته بندی شوند، بعنوان مثال دسته بندی افراد به گروه های سنی ۱۰-۰ سال ۲۰-۱۱ و ... یا ممکن است بخواهیم دو گروه شغلی کارگر و کشاورز را در هم ادغام کنیم.برای این منظور از مطالب ذیل استفاده می نماییم.

ابتدا به قسمت ویرایشگر داده ها می رویم، از نوار منو، روی کلید Transform رفته، آن را کلیک می کنیم، سپس Recode را انتخاب می کنیم، دو گزینه پیش رو داریم.

الف : recode in to same variable

ب) recode in to different variable

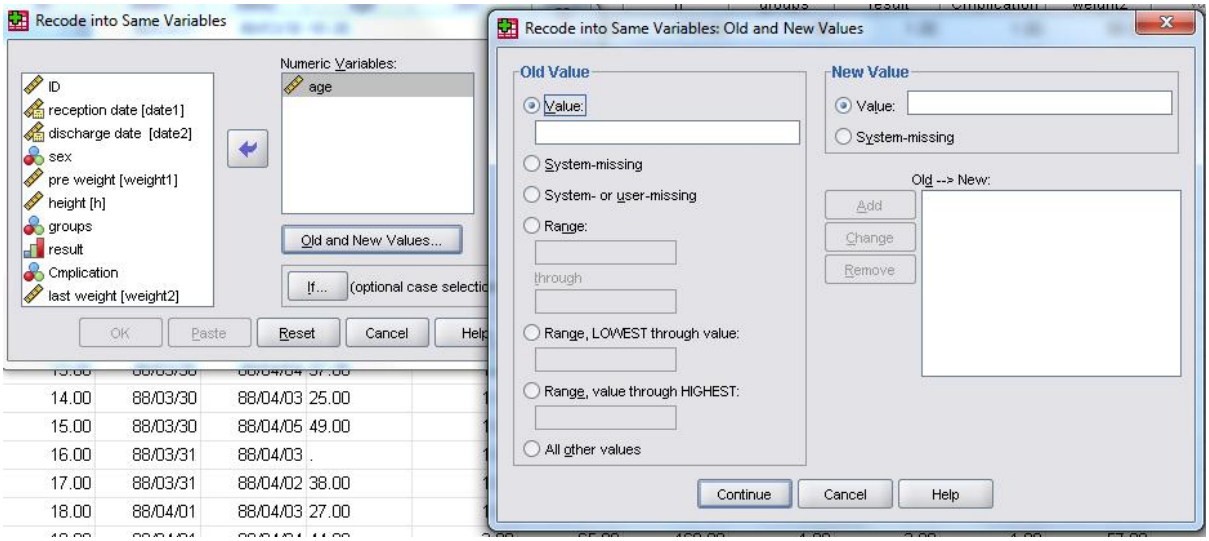


الف) اگر بخواهیم نتیجه دسته بندی روی همان متغیر جایگزین شود(Same Variable). تذکر مهم : با این عمل داده های متغیر قبلی از بین میرود.

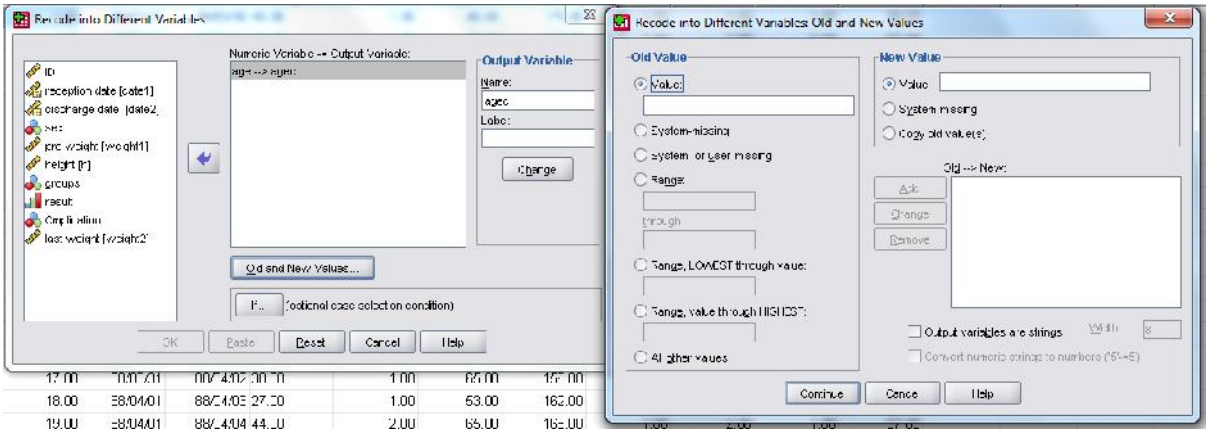
ب) اگر بخواهیم نتیجه دسته بندی بصورت یک متغیر دیگر و در ستون جدیدی در ویرایشگر داده ها ثبت شود.(متغیر اول حفظ شده تغییری نمی کند).

الف- روی Same Variable کلیک می کنیم، پنجره ای ظاهر می شود، متغیری که قرار است دسته بندی شود از لیست متغیرها انتخاب و آن را به سمت راست می آوریم، کلید Old and New را انتخاب نموده، پنجره جدیدی ظاهر می شود، قسمت سمت چپ مربوط به Old و سمت راست مربوط به New می باشد. در قسمت Old یکی از گزینه های مورد نظر را انتخاب می کنیم مثلاً Range اولی را فعال می کنیم (با کلیک کردن در دایره توخالی کنار آن) اگر بخواهیم سنین ۰ تا ۱۰ را در یک گروه قرار دهیم عدد صفر را در مستطیل اول و ۱۰ را در مستطیل دوم قرار می دهیم، سپس به قسمت بالای صفحه سمت راست رفته و کد ۱ را برای آن فاصله بر می گزینیم. کلید Add را زده تا تعریف انجام شود، سپس به قسمت Old بر می گردیم و فاصله ۲۰-۱۱ را قرار داده و در قسمت New کد ۲ را برای این فاصله انتخاب و Add می کنیم و به همین ترتیب ادامه داده تا دسته بندی پایان یابد. با انجام Continue به صفحه قبل برگشته و Ok را می زنیم. در این صورت متغیر قبلی به

کدهای ۱، ۲ و ۳ ... تبدیل مـی شـود. حـال بایسـتی روی Variable View در پـائین سـمت چـپ ویرایشگر داده ها رفته و این کدها را برای متغیر تعریف کنیم کد ۱ یعنی ۱۰-۰ و کد ۲ یعنـی ۲۰-۱۱ و ...



ب- Different Variable را بر می گزینیم در پنجره ای که ظاهر می شود متغیر مورد نظر را انتخاب و به پنجره سمت راست می بریم سپس در مستطیلی که ذیل نام متغیر خروجی (Output Variable Name) نام جدیدی می گذاریم این نام از قانون اسم متغیرها پیروی می کند (یعنی نباید بین حروف جای خالی باشد یا از علائم نباید استفاده کرد، بیشتر از ۸ کاراکتر نباشد و ...)پش از انتخاب نام متغیر جدید ، کلید change را کلیک می کنیم تا در نوار آبی شکل ذیل قـرار گیـرد و سـپس بـه قسمت Old and New Variable رفته و از این قسمت به بعد مانند قسمت الف ادامه می دهیم.

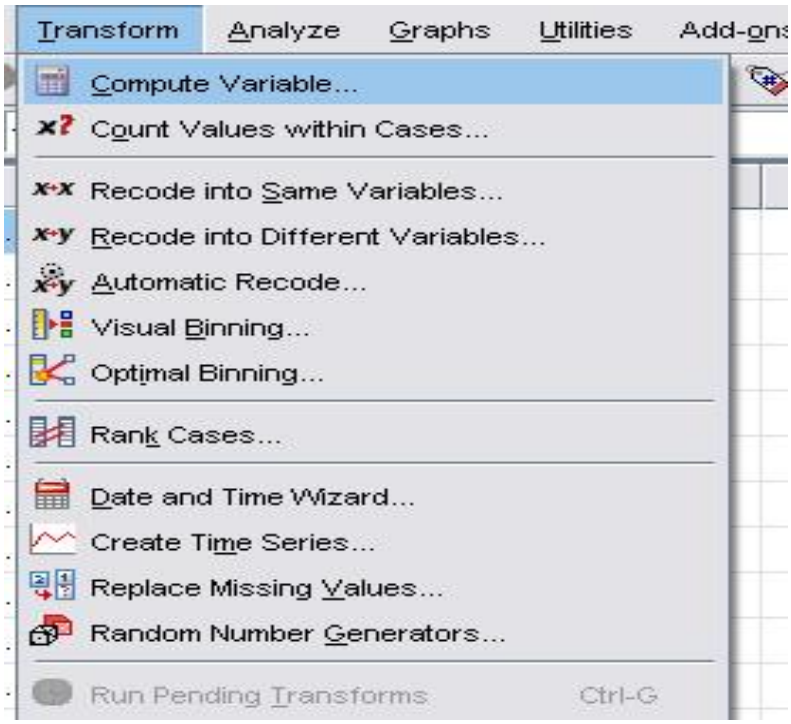


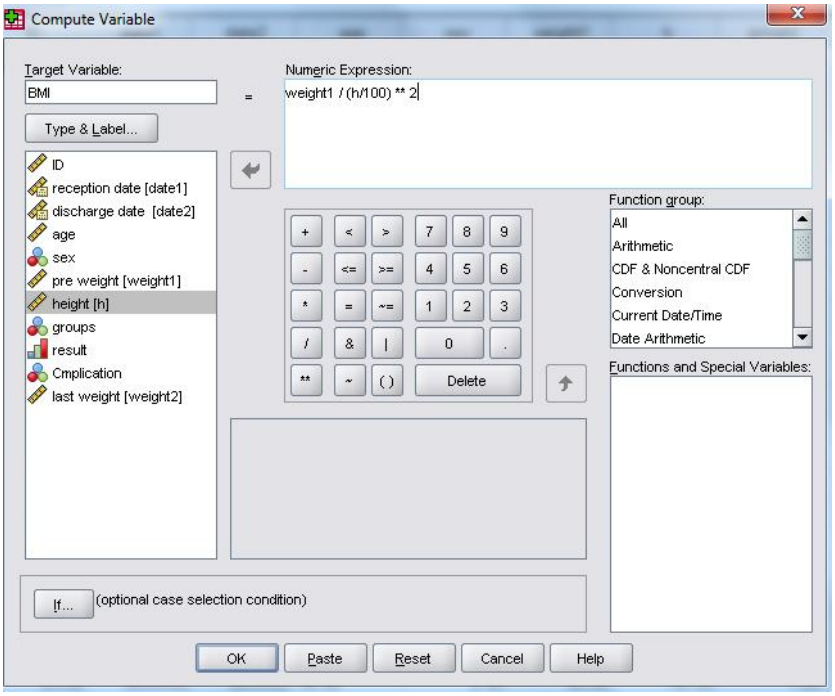
۲- محاسبات

ایجاد متغیرهای مرکب

اگر بخواهیم از روی دو یا چند متغیر یک اندیکس خاص (مثلا BMI) را تولید کنیم به محاسبات نیاز داریم. یا بخواهیم مجموع نمراتی که هر فرد از چند سئوال بدست می آورد را با هم جمع کنیم از قسمت محاسبات استفاده می نمائیم. به ویرایشگر داده ها رفته و کلید Transform را کلیک می کنیم در اولین سطر عبارت Compute را برمی گزینیم، نامی را برای متغیری که می خواهیم تولید کنیم انتخاب کنیم (مثلا Score یا هر اسم دیگر) سپس فرمول مربوطه را با استفاده از لیست متغیرها و کلیدهای محاسباتی که در همین پنجره وجود دارد می نویسیم و سپس کلید Ok را می فشاریم.

مثال: اگر قد بر حسب سانتیمتر و وزن بر حسب کیلوگرم باشد، BMI که حاصل تقسیم وزن به کیلوگرم بر مربع قد (بر حسب متر) چنین محاسبه می شود.





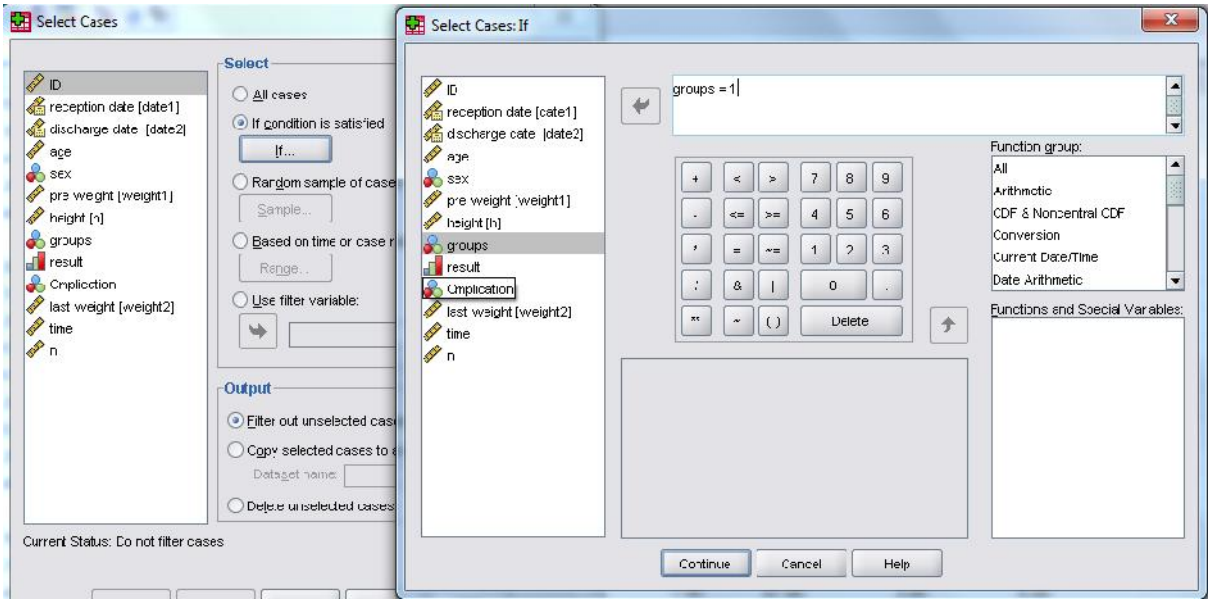
انجام محاسبات بدین ترتیب است که ابتدا از داخلی ترین پرانتز محاسبه آغاز می شود سپس توان و پس از آن ضرب و تقسیم و بعد جمع و تفریق. با انجام یک (Frequency) می توان از صحت محاسبات مطلع شد.

۳- فیلتر کردن (انتخاب گروه خاص برای تحلیل داده ها)

اگر بخواهیم شاخص های آماری یا توزیع فراوانی را برای گروه هایی خاص انجام دهیم. مثلا بخواهیم در افرادی که سن بیشتر از ۴۰ سال دارند، میانگین و انحراف معیار BMI را بدست آوریم، یا فقط در افراد مورد بخواهیم برخی شاخص ها را محاسبه کنیم .برای این منظور در ویرایشگر داده ها، از منوی Data وارد می شویم سپس Select case را کلیک می کنیم، پنجره ای ظاهر می شود، در قسمت فوقانی پنجره If را فعال می کنیم و سپس شرط مربوطه را می نویسیم مثلا Age>۴۰ یا group=۱ سپس با Ok کردن، بطور اتوماتیک فعالیت های بعد از آن صرفا برای افراد بالاتر از ۴۰ سال یا برای افرادی که کد متغیر گروه آنها ۱ می باشد، خواهد بود.

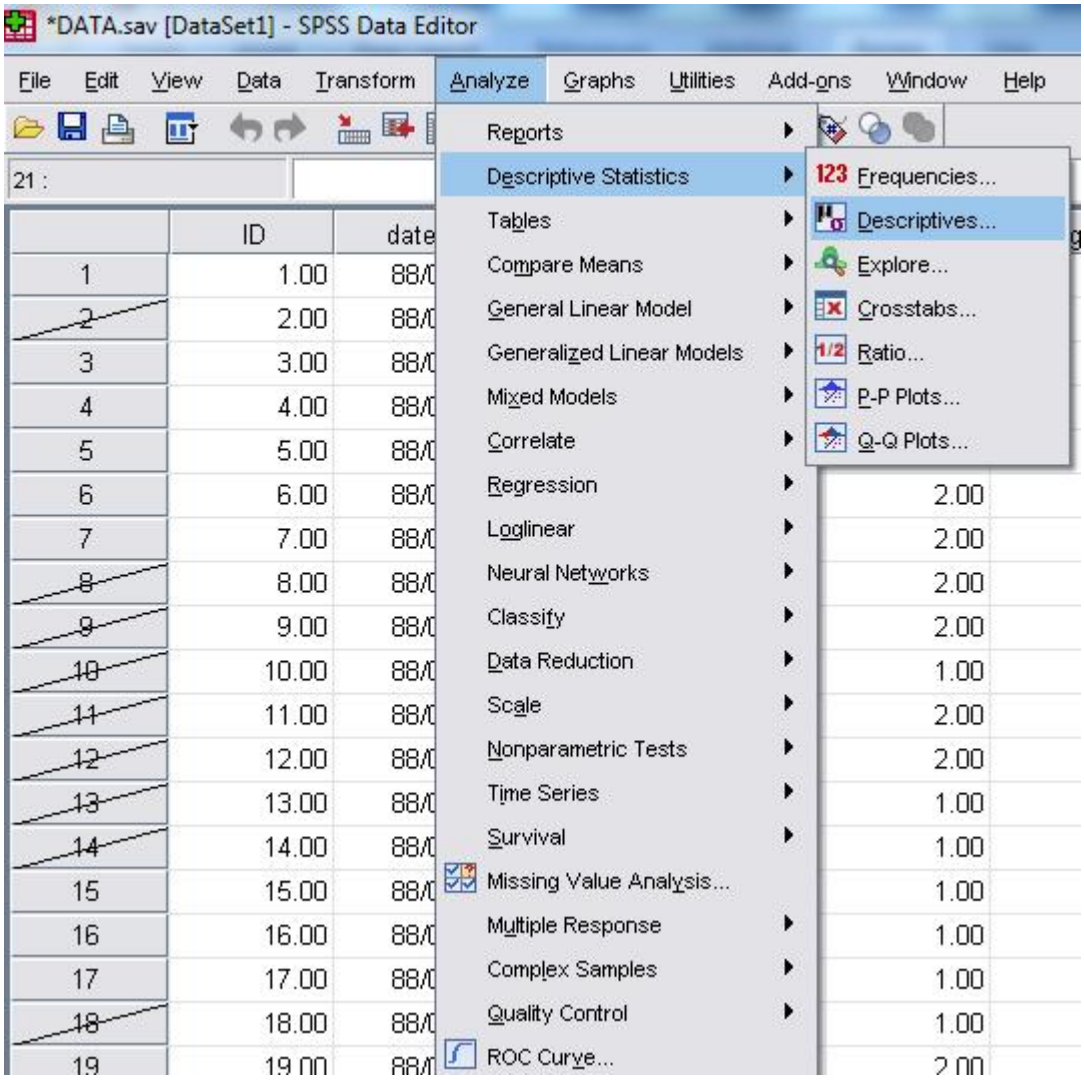
تذکر۱ : اگر بخواهیم محاسباتی را راجع به افرادی که هم در گروه ۱ هستند و هم age>۴۰ بین آنها عبارت & را قرار می دهیم .

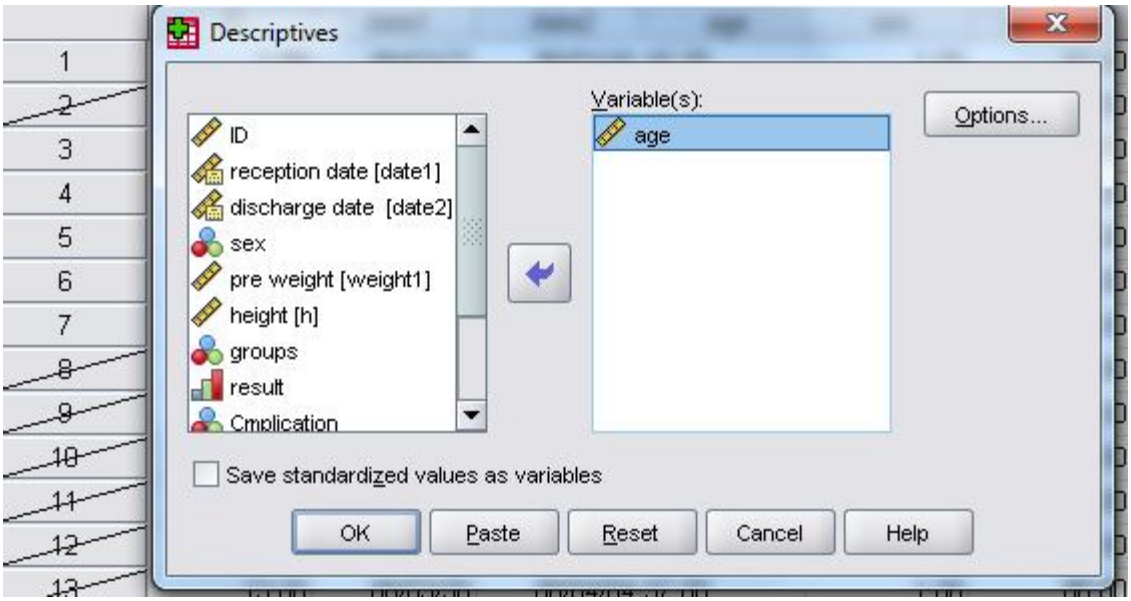
تذکر ۲ : پس از محاسبات لازم بر روی این گروه خاص ، با رفتن مجدد به منوی Data و انجام Select case قسمت All Case را فعال کرده و از آن به بعد مجددا فعالیت ها روی کل افراد صورت خواهد پذیرفت.(این قسمت ممکن است فراموش شود).



مثال: میانگین و انحراف معیار سن را برای افرادی که در گروه ۱ هستند ،بدست آورید.

پس از انجام مراحل فوق از منوی اصلی analyzeرا انتخاب ومسیر ذیل را ادامه می دهیم.





با ok نمودن خروجی فقط برای افرادی محاسبه می شود که در گروه ۱ هستند.

Descriptives

```
[DataSet1] G:\kargah ha\kargah and tamrin spss\spss worksho
```

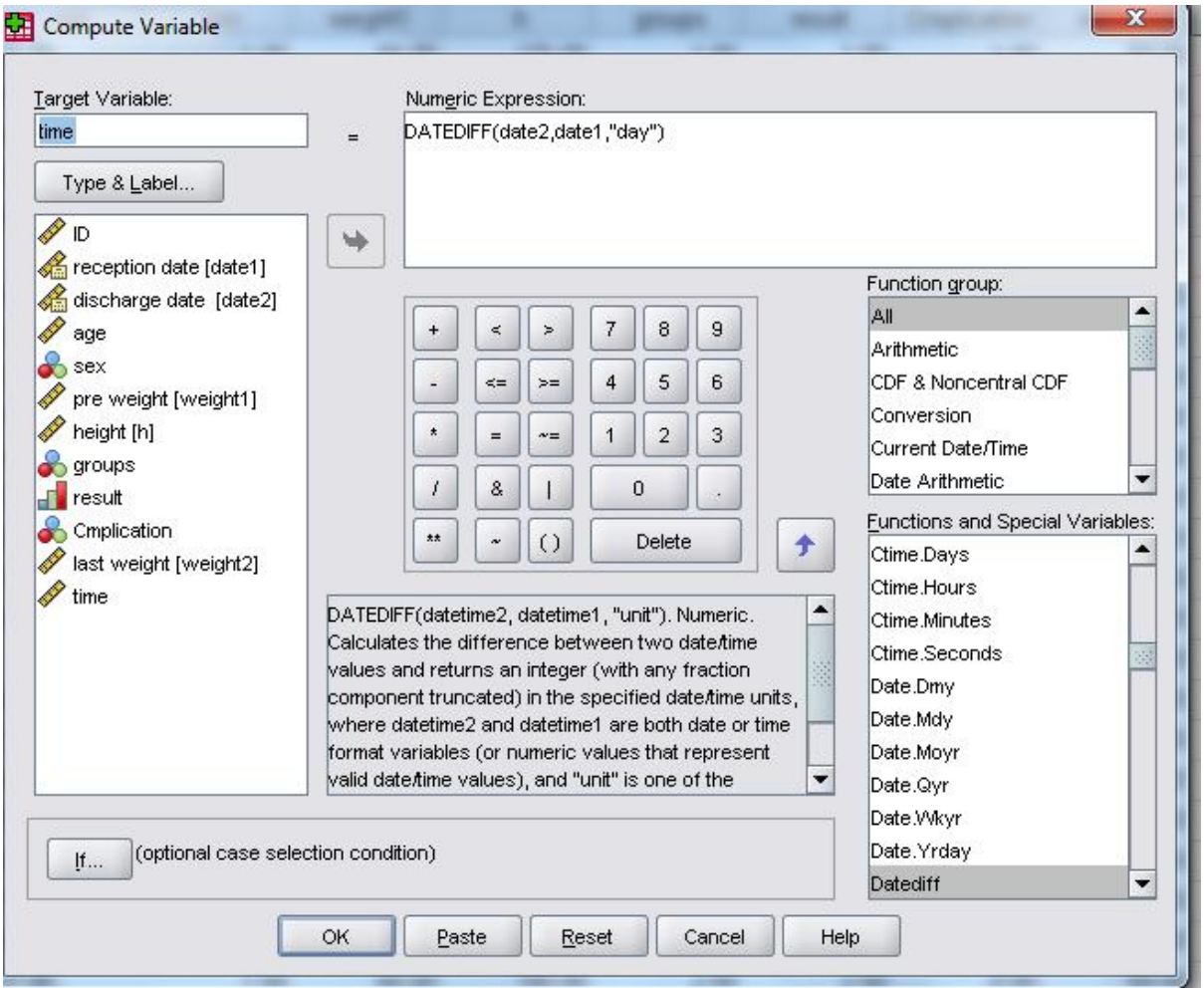
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age	18	28.00	56.00	42.7778	8.68776
Valid N (listwise)	18				

۴– مباحث پیشرفته :

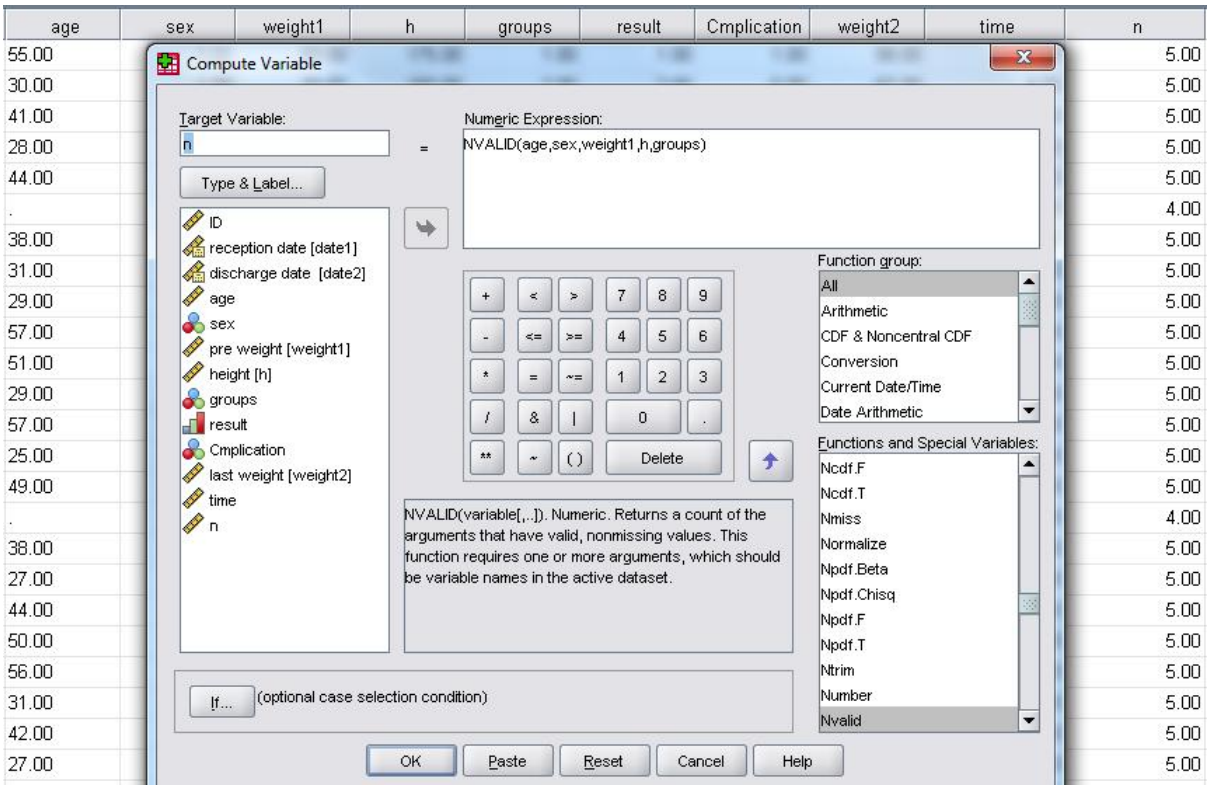
عملیلت روی متغیر های تاریخ :

تفاضل : اگر بخواهیم طول مدت بستری را به دست آوریم کافی است که تاریخ ترخیص را منهای تاریخ پذیرش نماییم . فرمت تاریخ شمسی به صورت yy/mm/dd باشد .



شمارش تعداد پاسخ های داده شده برای هر فرد :

گاهی اوقات لازم است تعداد پاسخ هایی که به یک پرسشنامه داده شده است شمارش شود . به این دلیل که اگر مثلا به بیش از ۱۰ سوال را پاسخ نداده باشد آن فرد از مطالعه حذف شود . در اینگونه موارد از دستور Nvalid استفاده می شود . در منوی translation گزینه computرا انتخاب و متغیر جدید را در target variableگذاریم و در سمت راست پس از انتخاب nvalid و قرار دادن متغیر ها در آن تعداد متغیر هایی که هر فرد پاسخ داده ، در متغیری که تعریف کردیم نوشته می شود .

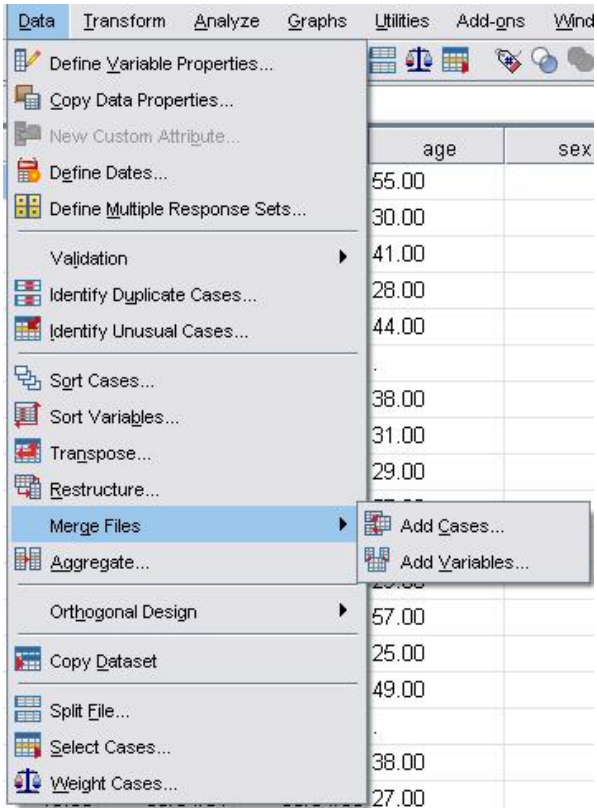


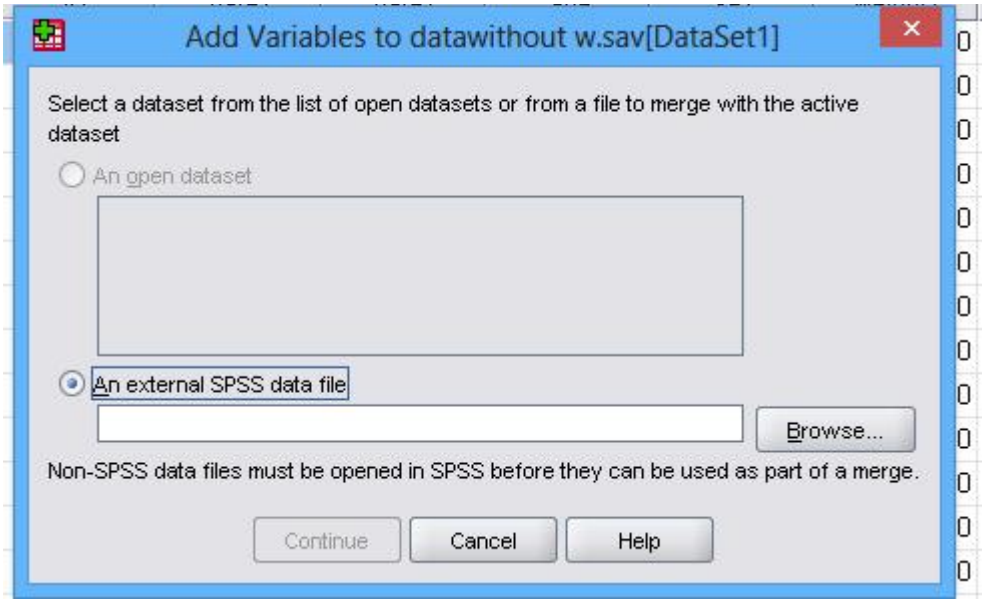
ادغام کردن فایل ها :

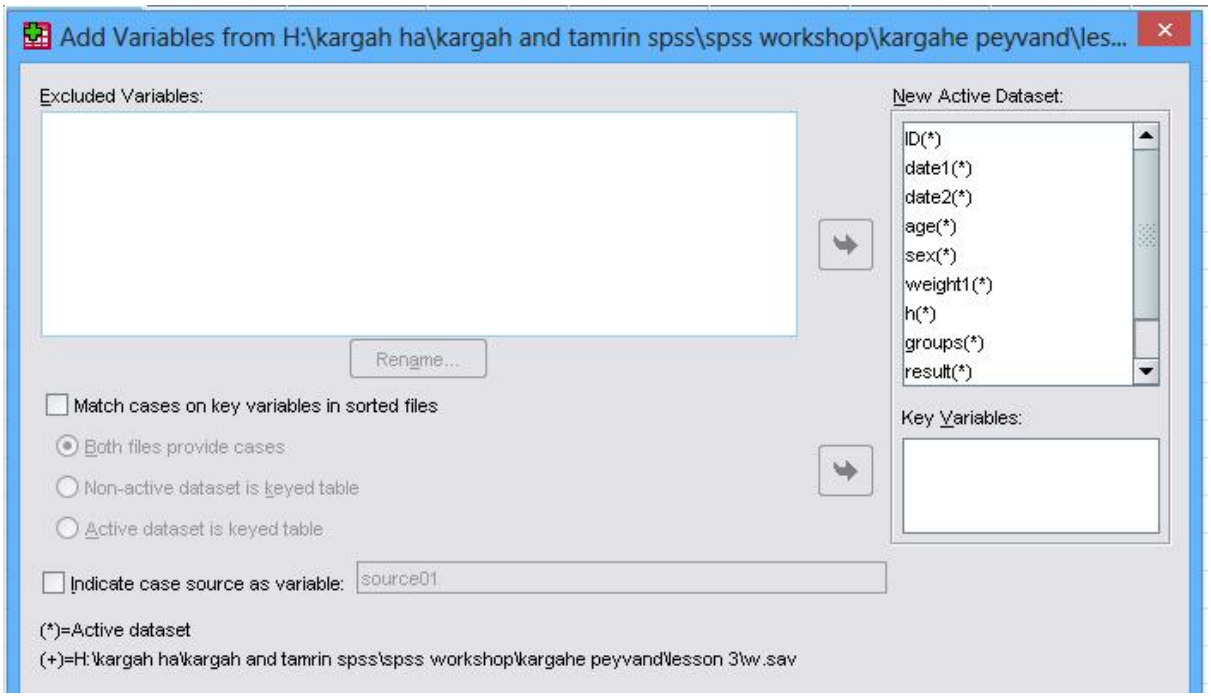
اگر بخواهیم به یک فایل داده یک فایل دیگری را الحاق کنیم چند حالت ممکن است پیش بیاید :

۱- متغیر ها در هر دو فایل یکسان هستند و بخواهیم افراد یک فایل را به دیگری اضافه کنیم.
۲- افراد یکسان هستند ولی متغیر ها متفاوت هستند و بخواهیم متغیر ها ی یک فایل را به دیگری اضافه نماییم .
۳- هم متغیر ها و هم افراد متفاوت باشند

برای انجام وضعیت اول از نوار منو Data را انتخاب و سپس به mergefile میرویم و add cases را برمی گزینیم


Data
Transform
Analyze
Graphs
Utilities
Add-ons
Wind
Define Variable Properties...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Sort Cases...
Sort Variables...
Transpose...
Restructure...
Merge Files
Aggregate...
Orthogonal Design
Copy Dataset
Split File...
Select Cases...
Weight Cases...
Add Cases...
Add Variables...
age
sex
55.00
30.00
41.00
28.00
44.00
38.00
31.00
29.00
57.00
25.00
49.00
38.00
27.00

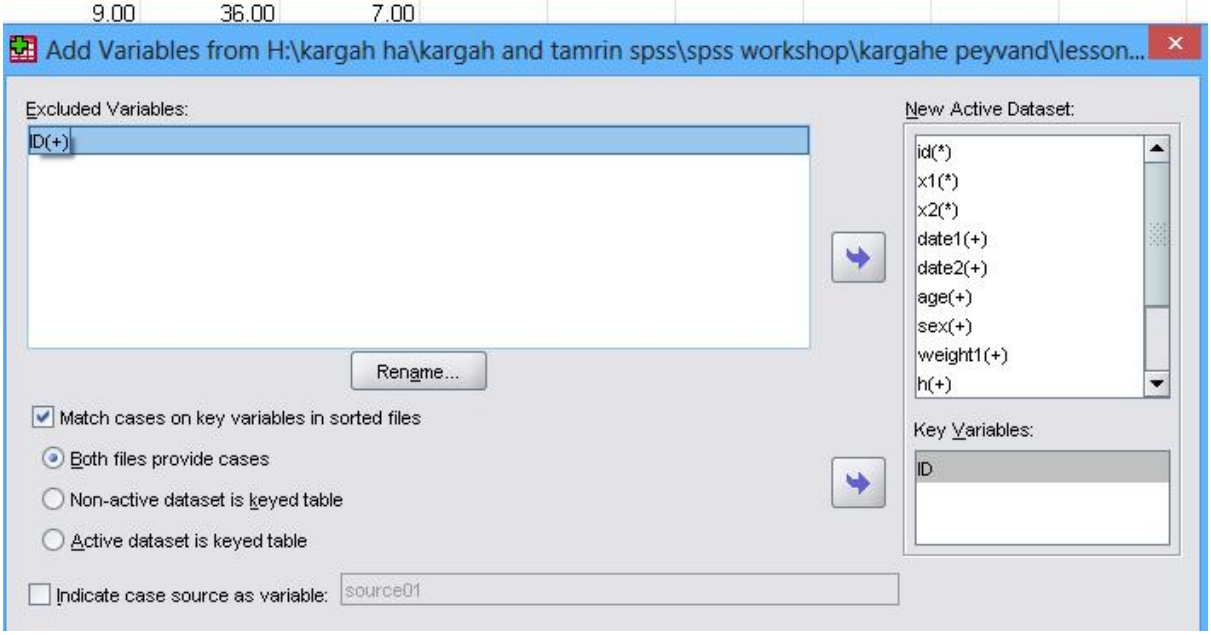

Add Variables to datawithout w.sav[DataSet1]
Select a dataset from the list of open datasets or from a file to merge with the active dataset
An open dataset
An external SPSS data file
Browse...
Non-SPSS data files must be opened in SPSS before they can be used as part of a merge.
Continue
Cancel
Help



برای اضافه کردن متغیر جدید بایستی add variable را انتخاب کنیم .

برای انتخاب داده های یک سری از افراد از بانک داده ها بایستی یک متغیر مشترک بین آن ها باشد . مثلا شماره ردیف آن ها . ابتدا دادهای هر دو فایل را نسبت به آن متغیر بصورت صعو دی یا نزولی مرتب میکنیم سپس از add variable مـی رویـم و متغیـر مشـترک را انتخاب و mach cases را علامت می زنیم و متغیر مشترک را به key variable میبریم و ok

id	x1	x2	var	var	var	var	var	var
7.00	23.00	2.00						
9.00	36.00	7.00						



آزمون فرضیه :

مقدمه :

به طور کلی هدف از آزمون های آماری یک تصمیم گیری در مورد پارامتر جامعه است که این تصمیم می تواند درست یا نادرست باشد. هرگاه بخواهیم یک ادعا درباره پارامتر مجهول جامعه را از طریق تایید آن به وسیله اطلاعات حاصل از نمونه نشان دهیم . (مثلا میانگین فشارخون سیستولیک در جامعه ای با عدد ۱۳۰ اختلاف دارد). ادعای مورد نظر را به عنوان فرضیه یک (فرض جانشین) (Alternative Hypothesis)یا H_1و حالت نفی آن ادعا (میانگین فشارخون سیستولیک در جامعه ای برابر عدد ۱۳۰ می باشد) را به عنوان فرضیه صفر H_0 (Null Hypothesis)در نظر می گیریم. با توجه به مطالعات گوناگون، فرضیه ها و آزمون های گوناگونی خواهیم داشت. به عنوان مثال در حالت آزمون برابری میانگین دو جامعه مستقل فرضیه جانشین (فرضیه ادعا) این است که میانگین دو جامعه برابر نیستند در مقابل اینکه میانگین های دو جامعه برابر هستند (فرضیه صفر).

$$\begin{cases} H_0 : «_1 = «_2 \\ H_1 : «_1 \neq «_2 \end{cases}$$

در حالت معمول اساس را بر درستی فرضیه صفر قرار می دهیم مگر اینکه شواهد کافی با استفاده از نمونه خلاف آن را قویا نشان دهد که در این صورت گوییم نتیجه معنی دار بدست آمده است. آزمون فرضیه تعیین روشی است که بر اساس اطلاعات بدست آمده از نمونه بتوان نسبت به درستی یا نادرستی فرضیه صفر تصمیم گیری نمود. واضح است که در تصمیم گیری درباره رد فرضیه صفر با دو نوع خطا مواجه هستیم. ممکن است فرضیه صفر واقعا درست باشد اما نتیجه آزمون منجر به رد آن باشد که این خطا را خطای نوع اول می نامیم. همچنین ممکن است فرضیه صفر را رد نکنیم در حالی که فرضیه صفر واقعا غلط باشد. که آن را خطای نوع دوم می نامیم.

واقعیت / نتیجه آزمون	فرضیه صفر درست است	فرضیه صفر درست نیست
رد فرضیه صفر	خطای نوع اول	-
رد نکردن فرضیه صفر	-	خطای نوع دوم

احتمال ارتکاب خطای نوع اول را با α نشان داده و آن را سطح معنی داری آزمون گوئیم. همچنین احتمال ارتکاب خطای نوع دوم را با β نشان می دهیم و مقدار (β-۱) را توان آزمون می گوییم.

واضح است که آزمون فرضیه باید به گونه ای باشد که در آن احتمال ارتکاب خطاهای نوع اول و دوم تا حد ممکن حداقل باشند. یک راه حل برای این منظور افزایش حجم نمونه است. اما چنانچه در انتخاب نمونه محدودیت داشته باشیم لازم است از روش آزمونی استفاده نماییم که با توجه به اهمیت بیشتر خطای نوع اول، احتمال ارتکاب این خطا را ثابت گرفته و همزمان احتمال خطای نوع دوم را حداقل نماید.

در یک تحلیل آماری رد کردن فرضیه صفر به عنوان "نتیجه معنی دار Significance" نامیده می شود. بنابراین یک نتیجه معنی دار نتیجه ای است که علی القاعده بر حسب شانس و تصادف رخ نداده باشد اگر چه سطح معنی دار (α) باید مشخص باشد. مقادیر متداول برای سطح معنی دار آزمون (α) عبارتند از : ۰/۰۱، ۰/۰۵ و ۰/۱. معمولا وقتی که واژه "نتیجه معنی" دار استفاده می شود، منظور از آن رد کردن فرضیه صفر در سطح خطای در نظر گرفته شده است (مثلا ۰/۰۵ یا ۰/۰۱) است. یک نتیجه "غیر معنی دار" یعنی اینکه فرضیه صفر رد نشده است.

در عمل فرض می نماییم که فرضیه صفر صحیح است، پس از آن نمونه ای از آن جامعه بطور تصادفی انتخاب می کنیم و احتمال آنکه این نمونه ها مطابق فرضیه صفر باشد را محاسبه میکنیم . اگر احتمال آن کم بود (کمتر از سطح معنی داری ، که معمولا ۰.۰۵ است)می گوییم فرضیه صفر درست نیست و آن را رد می کنیم . این مقدار در نرم افزارهای آماری با عبارت "مقدار احتمالی P-Value" یا عبارت "میزان معنی داری Significance" و در SPSS با .Sig مشخص می شود. در یک قاعده کلی فرضیه صفر را رد می کنیم هر گاه P-Value یا Significance آزمون کمتر از احتمال خطای نوع اول گردد که در این صورت گوئیم آزمون در سطح α معنی دار است.(معمولا ۰.۰۵ و بندرت ۰.۰۱ یا ۰.۱ نیز درنظر گرفته می شود)

بررسی فرضیه نرمال بودن مشاهدات

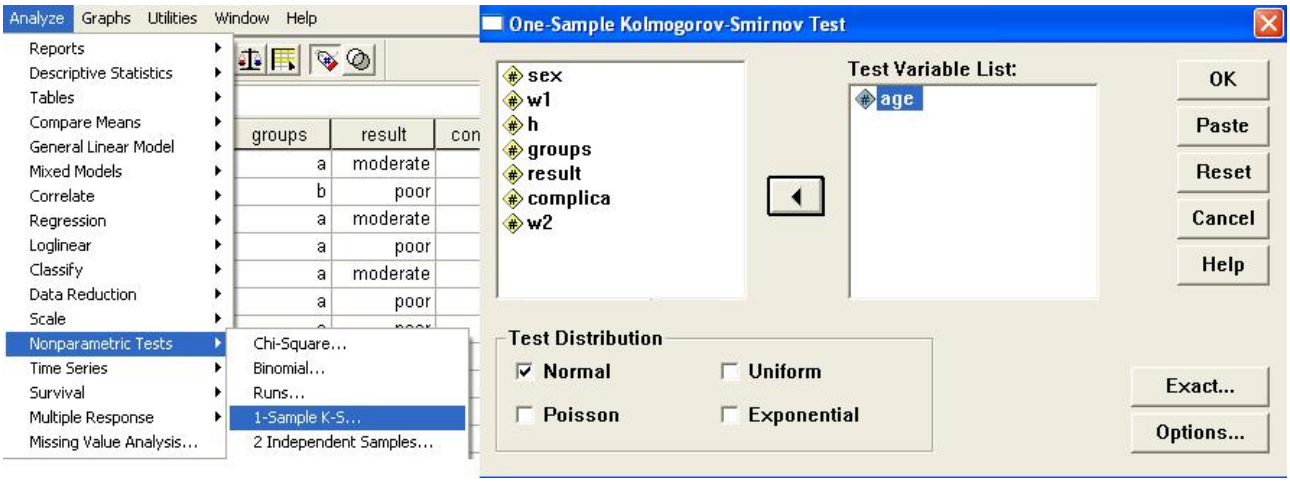
یکی از فرضیات اصلی برای اکثر آزمون های آماری، نرمال بودن توزیع مشاهدات است. اجرای آزمون نرمال بودن استفاده از آزمون کلموگرف-اسمیرنوف(Kolmogorov-smirnov) است. این آزمون یک آزمون ناپارامتری برای بررسی توزیع مشاهدات است. پیش فرض SPSS، انجام آزمون برای بررسی توزیع نرمال است. میزان تقریبی معنی داری آزمون (Asymptotic Significance) در انتهای خروجی آزمون کلوموگروف-اسمیرونوف می آید که با مقایسه آن با α می توان آزمون را در سطح معنی داری α انجام و نسبت به نرمال بودن توزیع مشاهدات تصمیم گیری نمود.

اگرα=۰/۰۵ در نظر بگیریم چنانچه P-value>۰/۰۵ باشد، می توان فرض کـرد ، توزیـع مشـاهدات نرمـال است.

مثال : آیا می توان فرض کرد توزیع سن در جامعه ای که از آن نمونه انتخاب کرده ایم نرمال است ؟

$(H_0 : x \to normal)$

مسیر ذیل را از منوی اصلی از قسمت analyze انتخاب می کنیم.



با OK کردن خروجی به صورت ذیل ظاهر می شود.

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		AGE
N		38
Normal Parameters a,b	Mean	40.2632
	Std. Deviation	10.47180
Most Extreme Differences	Absolute	.154
	Positive	.154
	Negative	-.087
Kolmogorov-Smirnov Z		.949
Asymp. Sig. (2-tailed)		.329

a. Test distribution is Normal.

b. Calculated from data.

بزرگتر از ۰/۰۵ پس نرمال است

آزمون مقایسه میانگین یک جامعه با یک عدد ثابت

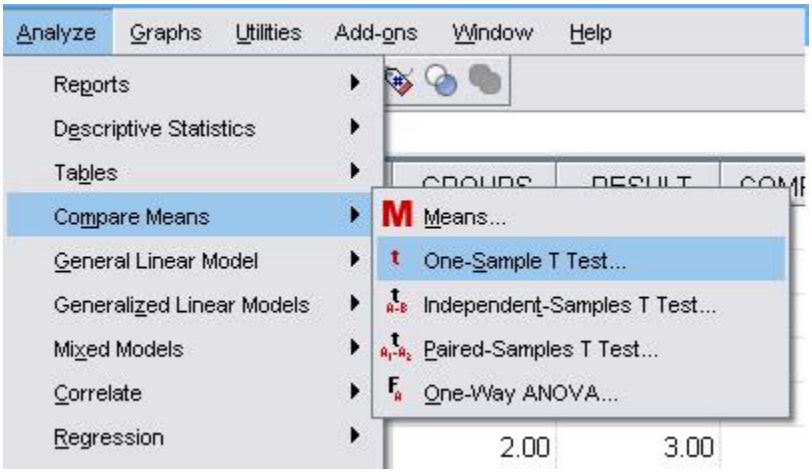
هدف از انجام این آزمون بررسی برابر بودن میانگین یک متغیر کمی جامعه با یک مقدار مشخص می باشد ($H_0: \mu = \mu_0$).

مثال: آیا میانگین سن در افراد جامعه برابر با عدد ۵۰ می باشد($H_0: \mu = ۵۰$).

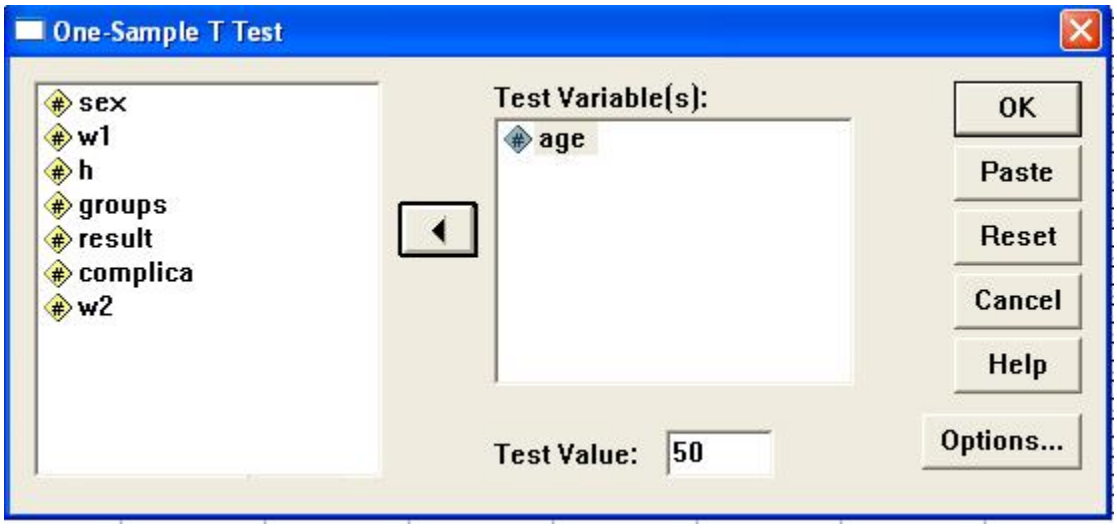
پیش فرض انجام آزمون این است که متغیر مورد مطالعه دارای توزیع نرمال است(توسط کلموگروف – اسمیرنوف آزمون می شود که قبلا گفته شد).

نمونه ای از جامعه انتخاب می کنیم و میانگین و انحراف معیار سن را در نمونه محاسبه می کنیم . اختلاف میانگینی که از نمونه به دست آمده است را با عدد ۵۰ را به دست می آوریم.این اختلاف را استاندارد می کنیم و آن را t می نامیم . اختلاف استاندارد شده ی زیاد منتج به رد فرضیه صفر می شود .

برای اجرای آزمون مقایسه میانگین با یک عدد ثابت مسیر زیر را در SPSS انتخاب می کنیم.



با این کار منوی One Sample T Test ظاهر می گردد. در این منو لازم است متغیر مورد مطالعه را از لیست متغیرها انتخاب و وارد قسمت Test Variable نمود. آنگاه مقدار μ_0 را وارد گزینه Test Value می کنیم.(در مثال بالا عدد ۵۰) با مشاهده مقدار Sig. در جدول خروجی آزمون و مقایسه آن با α، نسبت به درستی یا رد فرض صفر تصمیم گیری می نماییم. پیش فرض SPSS برای مقدار μ_0 عدد صفر می باشد یعنی چنانچه مقدار μ_0 را وارد Test Value ننمایم، SPSS آزمون برابری میانگین متغیر مورد مطالعه را با عدد صفر انجام می دهد.



و خروجی به صورت :

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
AGE	38	40,2632	10,47180	1,69875

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
AGE	-5,732	37	,000	-9,7368	-13,1788	-6,2948

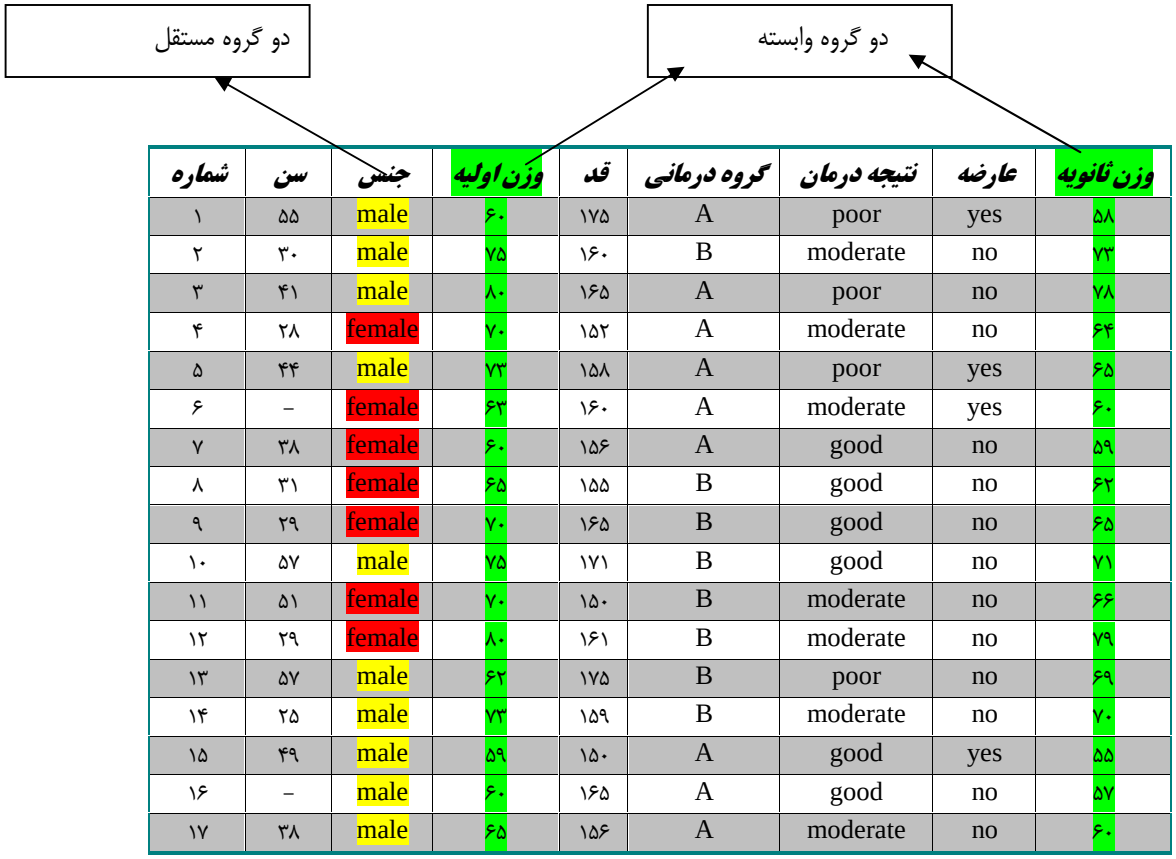
p-value<۰.۰۵

مقدار sig.(۲-taild) که همان p-value می باشد کمتر از ۰/۰۵ است لذا میانگین سن جامعه ای که از آن نمونه گرفته ایم با ۵۰ اختلاف معنی داری دارد.

مقایسه های بین دو گروه:

اگر بخواهیم دو گروه را با هم مقایسه کنیم ممکن است این دو گروه مستقل از هم باشند یـا وابسـته . دو گروه وقتی مستقل هستند که افراد آن با هم متفاوت باشد مثلا گروه ۱ را زنان تشـکیل دهنـد و گـروه ۲ را

مردان . ولی دو گروه وابسته به دو گروهی می گویند که افراد همان افراد گروه ۱ هستند ولی در شرایط زمانی یا مکانی دیگر . مثل گروه ۱ افراد قبل از یک مداخله و گروه ۲ همان افراد بعد از مداخلـه . در داده هـای ذیـل sex دو گروه مستقل هستند ولیوزن اولیه و وزن ثانویه دو گروه وابسته می باشند .

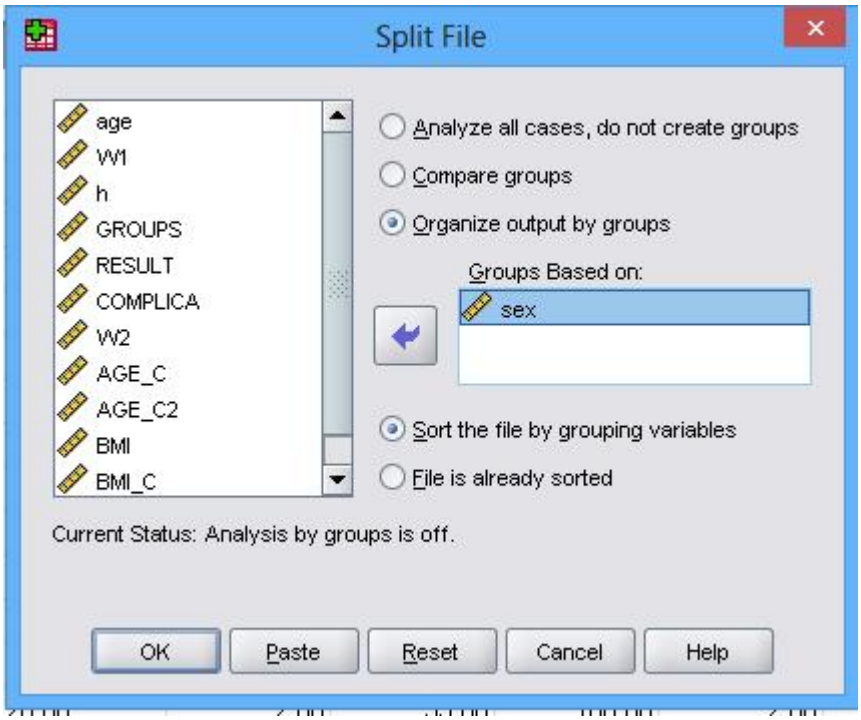


شماره	سن	جنس	وزن اولیه	قد	گروه درمانی	نتیجه درمان	عارضه	وزن ثانویه
۱	۵۵	male	۶۰	۱۷۵	A	poor	yes	۵۸
۲	۳۰	male	۷۵	۱۶۰	B	moderate	no	۷۳
۳	۴۱	male	۸۰	۱۶۵	A	poor	no	۷۸
۴	۲۸	female	۷۰	۱۵۲	A	moderate	no	۶۴
۵	۴۴	male	۷۳	۱۵۸	A	poor	yes	۶۵
۶	–	female	۶۳	۱۶۰	A	moderate	yes	۶۰
۷	۳۸	female	۶۰	۱۵۶	A	good	no	۵۹
۸	۳۱	female	۶۵	۱۵۵	B	good	no	۶۲
۹	۲۹	female	۷۰	۱۶۵	B	good	no	۶۵
۱۰	۵۷	male	۷۵	۱۷۱	B	good	no	۷۱
۱۱	۵۱	female	۷۰	۱۵۰	B	moderate	no	۶۶
۱۲	۲۹	female	۸۰	۱۶۱	B	moderate	no	۷۹
۱۳	۵۷	male	۶۲	۱۷۵	B	poor	no	۶۹
۱۴	۲۵	male	۷۳	۱۵۹	B	moderate	no	۷۰
۱۵	۴۹	male	۵۹	۱۵۰	A	good	yes	۵۵
۱۶	–	male	۶۰	۱۶۵	A	good	no	۵۷
۱۷	۳۸	male	۶۵	۱۵۶	A	moderate	no	۶۰

آزمون برابری میانگین دو گروه مستقل

هدف از انجام این آزمون مقایسه میانگین های دو گروه مستقل در حامعه مـی باشـد. (مـثلا ، مقایسـه میانگین سن در دو گروه مذکر و مونث) فرضیه انجام این آزمون نرمال بودن توزیع متغیـر مـورد مطالعـه در دو گروه می باشد. در اینجا دو متغیر داریم، یک متغیر همان است که می خواهیم از آن میانگین بگیریم مثل سـن که حتما باید کمی باشد و متغیر دیگر کیفی دو حالته است (مثل جنس).برای اجرای آزمـون مسـیر زیـر را در SPSS انتخاب می نماییم.

نخست نرمال بودن متغیر سن را در دو گروه کنترل می کنیم . چون بایستی سن در هر یک از دو گروه نرمال باشد می توان با استفاده از Split کردن روی جنس ، آزمون نرمال بودن را بطور جداگانه روی هر جنس انجام داد .



sex =female

One-Sample Kolmogorov-Smirnov Test[c]

		age
N		۱۹
Normal Parameters[a]	Mean	۳۸.۵۷۸۹
	Std. Deviation	۹.۹۹۰٦٤
Most Extreme Differences	Absolute	.۲۰۰
	Positive	.۲۰۰
	Negative	-.۱۰۸
Kolmogorov-Smirnov Z		.۸۷۰
Asymp. Sig. (۲-tailed)		.٤۳٥

a. Test distribution is Normal.

One-Sample Kolmogorov-Smirnov Test[c]

		age
N		۱۹
Normal Parameters[a]	Mean	۳۸.۵۷۸۹
	Std. Deviation	۹.۹۹۰٦٤
Most Extreme Differences	Absolute	.۲۰۰
	Positive	.۲۰۰
	Negative	-.۱۰۸
Kolmogorov-Smirnov Z		.۸۷۰
Asymp. Sig. (۲-tailed)		.٤۳٥

c. sex = female

چون p value =۰.۴۳۵ فرضیه نرمال بودن در گروه افراد مونث رد نمی شود

sex = male

One-Sample Kolmogorov-Smirnov Test[c]

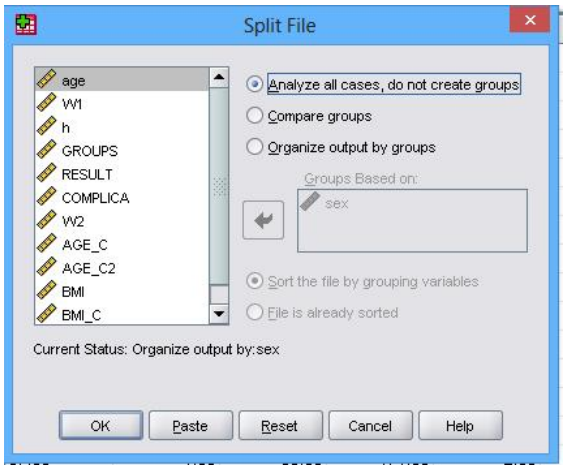
		age
N		۱۹
Normal Parameters[a]	Mean	٤۱.۹٤۷٤
	Std. Deviation	۱.۰۹۳٦٦E۱
Most Extreme Differences	Absolute	.۱۰۹
	Positive	.۱۰٥
	Negative	-.۱۰۹
Kolmogorov-Smirnov Z		.٤۷٥
Asymp. Sig. (۲-tailed)		.۹۷۸

a. Test distribution is Normal.

c. sex = male

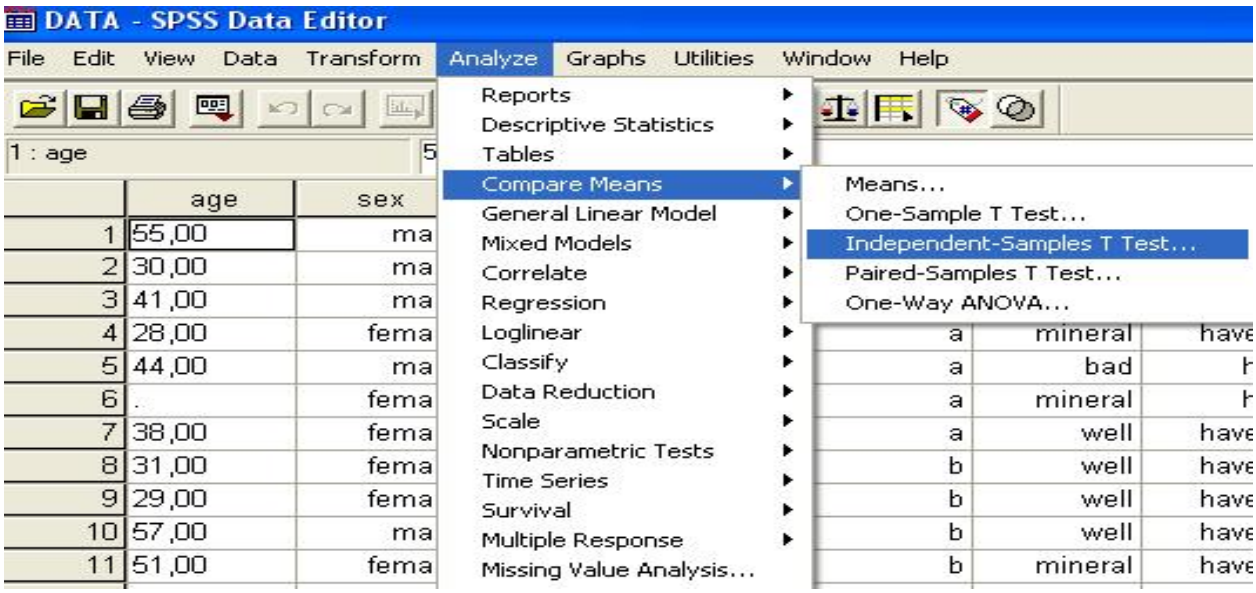
چون p value =۰.۹۷۸ فرضیه نرمال بودن در گروه افراد مذکر رد نمی شود

پس از آزمون نرمال بودن ابتدا از حالت Split خارج می شویم .

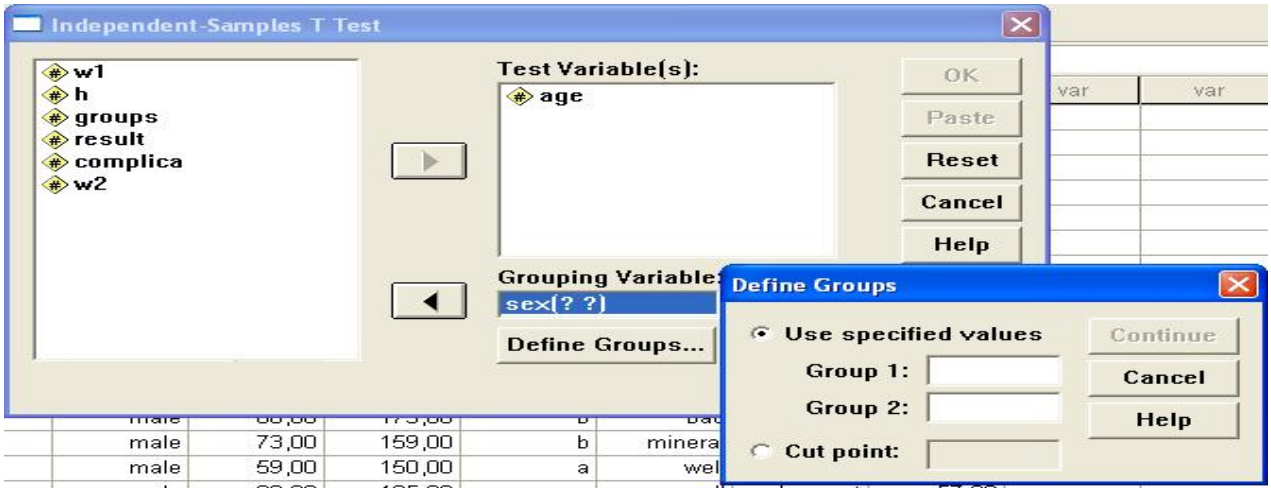


با توجه به نرمال بودن سن در دو گروه به مقایسه میانگین ها و انجام آزمون t مستقل می پردازیم .

از منوی Analyze به compare Mean و Independent Sample t test را برمی گزینیم .



با انتخاب مسیر فوق منوی Independent-Samples T-Test ظاهر می شود. حال متغیرکمی مورد آزمون را از لیست متغیرها به قسمت Test Variable و متغیر گروه بندی را به قسمت Group Variable فراخوانی نموده و کدهای متغیر گروه بندی را تعریف می نماییم. مثلا برای جنس کد ۱ و ۲ را در نظر گفته بودیم .



با توجه به این که مقدار آماره آزمون و درجه آزادی آن بستگی به برابری یا نابرابری واریانس های متغیر مورد مطالعه در دو گروه دارد، لازم است آزمون برابری واریانس ها نیز به موازات این آزمون صورت گیرد. نتایج آزمون برابری واریانس های تحت عنوان آزمون لون (Leven's Test) در ابتدای جدول خروجی آزمون برابری میانگین ها ارائه می گردد. با مشاهده میزان معنی داری آن می توان نسبت به برابری واریانس ها تصمیم گیری نمود. در نهایت چنانچه فرض برابری واریانس ها پذیرفته شود از نتایج موجود در سطر اول و چنانچه فرض برابری واریانس ها رد گردد از نتایج موجود در سطر دوم آزمون میانگین استفاده و بر اساس مقدار Sig. آزمون میانگین، نسبت به برابری میانگین ها تصمیم گیری می شود.خروجیSPSS به صورت ذیل است.

Group Statistics

	SEX	N	Mean	Std. Deviation	Std. Error Mean
AGE	male	19	41,9474	10,93655	2,50902
	female	19	38,5789	9,99064	2,29201

معنی داری از خط اول p-value>۰.۰۵

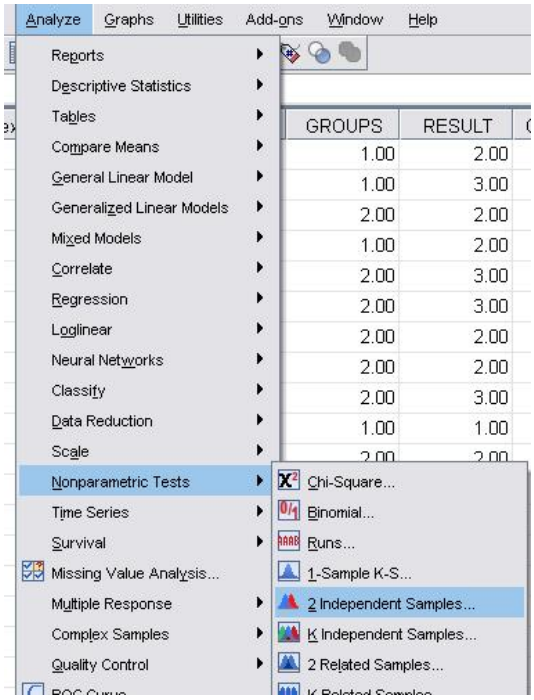
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Differ
AGE	Equal variances assumed	,093	,762	,991	36	,328	3,3684	3,3
	Equal variances not assumed			,991	35,709	,328	3,3684	3,3

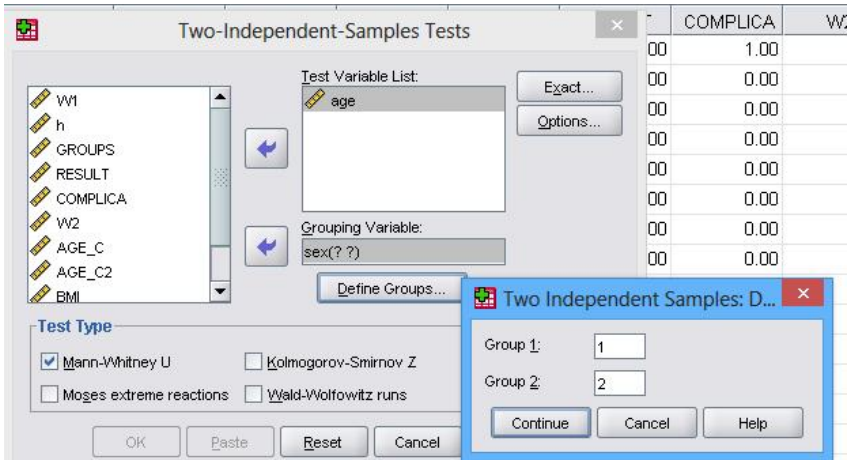
آزمون من – ویتنی :

اگر متغیر کمی مورد نظر نرمال نباشد دیگر نمی توان از آزمون فوق استفاده نمود بلکه از آزمون های ناپارامتری معادل استفاده می شود که برای مقایسه دو گروه مستقل نام آزمون مربوطه من – ویتنی می باشد .

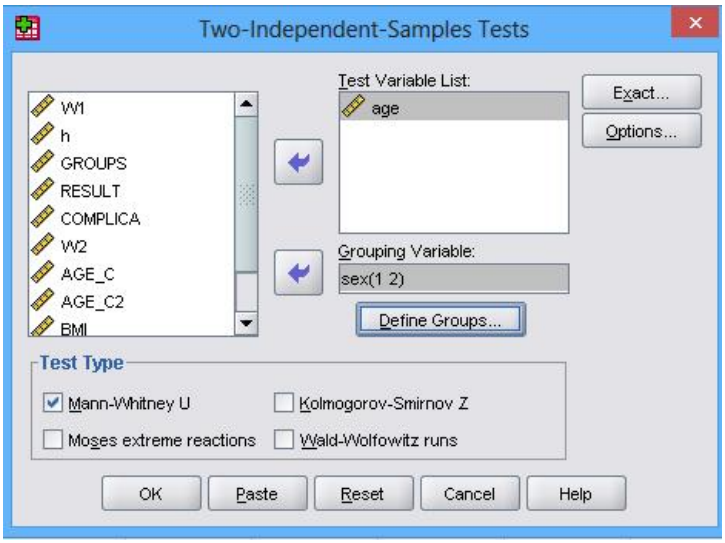
مسیر آزمون من- ویتنی به صورت ذیل می باشد .



باشد .



با انتخاب مسیر فوق منوی Two independent-Samples Test ظاهر می شود. حال متغیرکمی غیر نرمال مورد نظر را از لیست متغیرها به قسمت Test Variable و متغیر گروه بندی را به قسمت Group Variable فراخوانی نموده و کدهای متغیر گروه بندی را تعریف می نماییم. مثلا برای جنس کد ۱ و ۲ را در نظر گفته بودیم .بعد کلید ادامه را میزنیم .



در نهایت OK و به این ترتیب پاسخ در خروجی به صورت ذیل خواهد بود .

Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
age	male	۱۹	۲۱.۱۳	٤۰۱.۵۰
	female	۱۹	۱۷.۸۷	۳۳۹.۵۰
	Total	۳۸		

Test Statistics[b]

	age
Mann-Whitney U	۱٤۹.۵۰۰
Wilcoxon W	۳۳۹.۵۰۰
Z	-.۹۰٦

Asymp. Sig. (۲-tailed)	.۳٦٥
Exact Sig. [۲*(۱-tailed Sig.)]	.۳۷۰[a]

p-value>۰.۰۵ معنی دار نیست

a. Not corrected for ties.

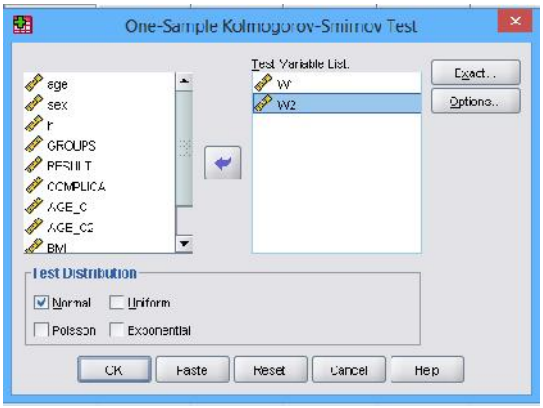
b. Grouping Variable: sex

نتیجه نشان میدهد که توزیع سن در دو گروه تفاوت معنی داری ندارد

آزمون تی زوجی : paired sample t-test

مقایسه ی میانگین های دو گروه وابسته (مشاهدات زوجی)

هدف از انجام این آزمون که به آزمون قبل و بعد نیز معروف می باشد برابری میانگین های دو گروه وابسته می باشد. پیش فرض انجام آزمون، نرمال بودن متغیر مورد مطالعه در دو گروه می باشد. در این آزمون برای هر آزمودنی دو مشاهده وجود دارد. برای انجام آزمون لازم است گروه اول و گروه دوم (مشاهدات قبل و بعد) را به صورت موازی وارد نموده به گونه ای که مشاهدات هر آزمودنی در مقابل یکدیگر باشند. (مشاهدات قبل به عنوان یک متغیر مثل وزن اولیه و مشاهدات بعد متغیری دیگر مثل وزن ثانویه) برای انجام آزمون لازم است ابتدا متغیر وزن قبل و وزن بعد نرمال باشند .



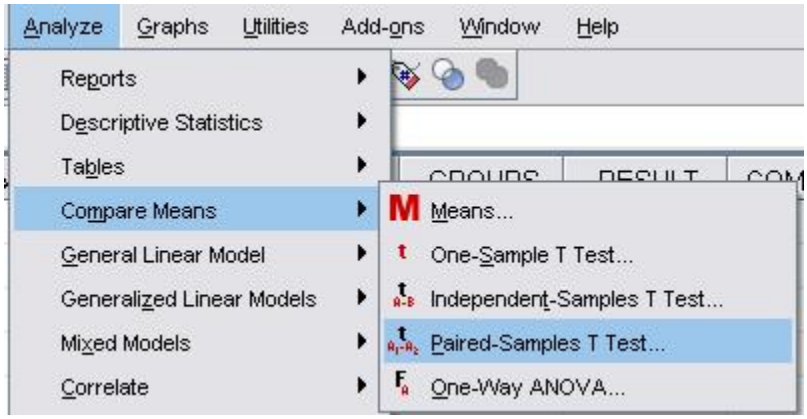
One-Sample Kolmogorov-Smirnov Test

		W۱	W۲
N		٤۰	٤۰
Normal Parameters[a]	Mean	٦٥.٦٧٥۰	٦۱.۳۷٥۰
	Std. Deviation	۹.۱۷٤٦۲	۹.٤۳۹۹۲
Most Extreme Differences	Absolute	.۱۳۲	.۱۸۳
	Positive	.۱۳۲	.۱۸۳
	Negative	-.۰۹۱	-.۰۷٦

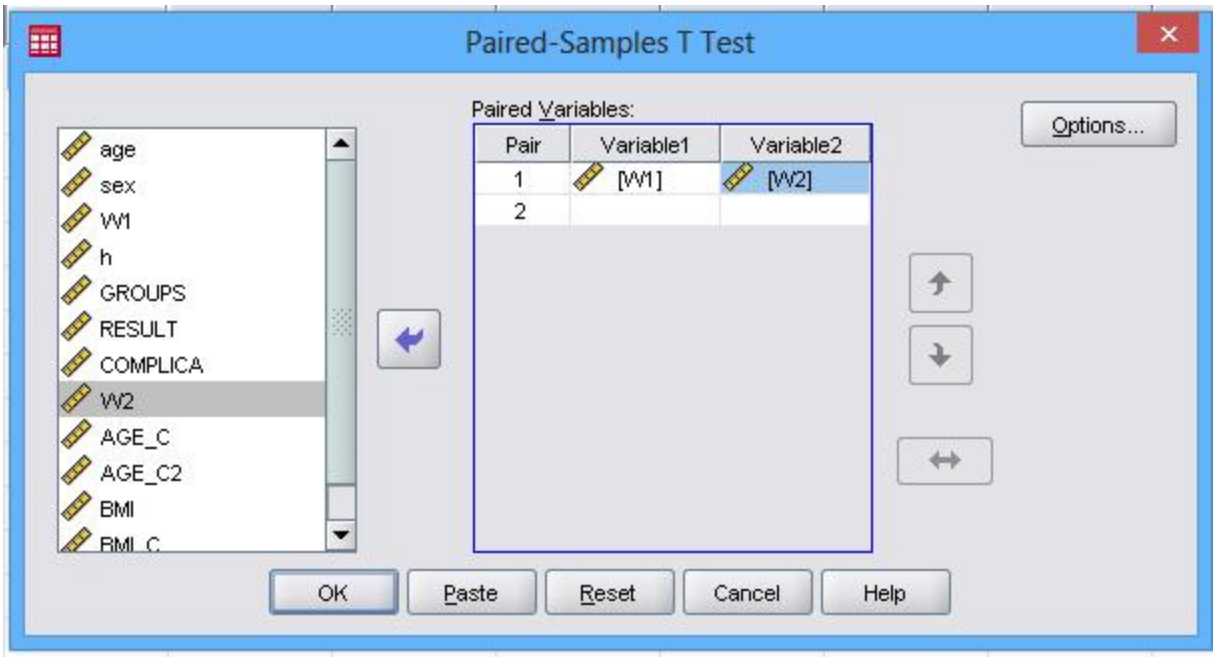
Kolmogorov-Smirnov Z	.٨٣٤	١.١٥٧
Asymp. Sig. (٢-tailed)	.٤٩٠	.١٣٨

a. Test distribution is Normal.

با توجه به p value های فوق وزن اولیه و ثانویه نرمال هستند . پس آزمون tزوجی را از مسیر ذیل اجرا می کنیم .



با انتخاب مسیر فوق منوی Paired-Samples T-Test ظاهر می گردد در این منو لازم است دو متغیر قبل و بعد را همزمان انتخاب نموده و آنها را به قسمت Paired Variable فراخوانی نمود.



در خروجی این آزمون، ابتدا مقادیر میانگین و انحراف معیار دو متغیر جهت توصیف داده ها ظاهر مـی شود و سپس نتیجه آزمون همبستگی بین مقادیر قبل و بعد (این قسمت در تفسیر مقایسه میانگین قبل و بعـد تاثیری ندارد)، سپس نتیجه آزمون برابری میانگین دو گروه که بر اساس برابری میانگین اختلافات قبل و بعد بـا مقدار صفر می باشد، ارائه می گردد. فرض برابری میانگین قبل و بعد در سطح خطای α رد می گردد اگر مقدار Sig. آزمون کمتر از α باشد.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	W1	65.6750	40	9.17462	1.45064
	W2	61.8750	40	9.51433	1.50435

Paired Samples Correlations

		N	Correlation	Sig.
Pair1	W1 & W2	40	.837	.000

Paired Samples Test

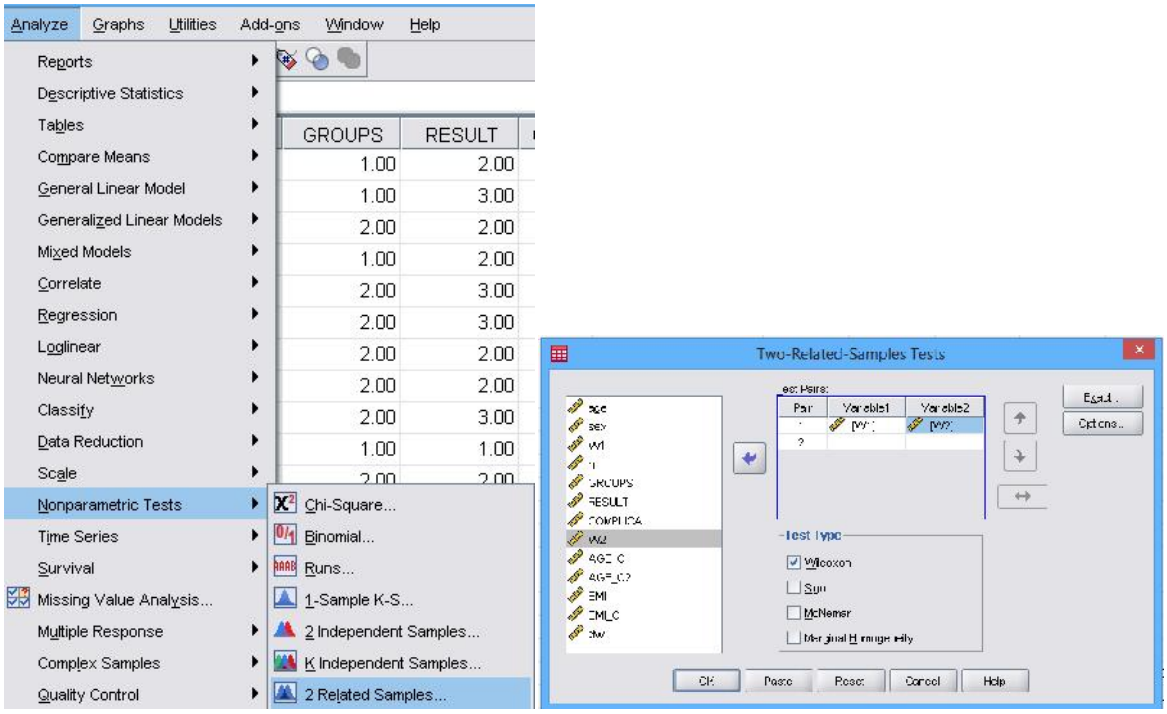
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair1	W1 - W2	3.8000	5.34070	.84444	2.0920	5.5080	4.500	39	.000

p-value<۰.۰۵ بین میانگین قبل و بعد اختلاف معنی دار وجود دارد

آزمون ویلکاکسون :

مقایسه دو گروه وقتی متغیر کمی نرمال نباشد:

چنانچه حداقل یکی از متغیر های کمی نرمال نباشد از آزمون ناپارامتری ویلکاکسون استفاده می شود .



Ranks

		N	Mean Rank	Sum of Ranks
W۲ - W۱	Negative Ranks	۳۷[a]	۲۱.۱۵	۷۸۲.۵۰
	Positive Ranks	۳[b]	۱۲.۵۰	۳۷.۵۰
	Ties	.[c]		
	Total	٤۰		

a. W۲ < W۱

b. W۲ > W۱

c. W۲ = W۱

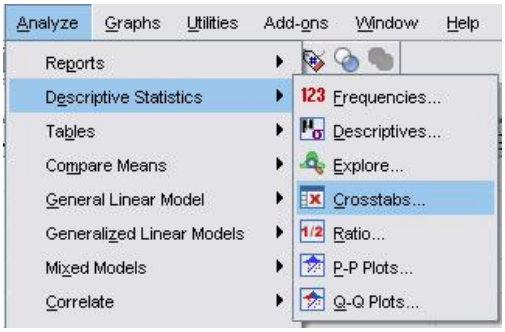
Test Statistics[b]

	W۲ - W۱
Z	-۵.۰۱۷[a]
Asymp. Sig. (۲-tailed)	.۰۰۰

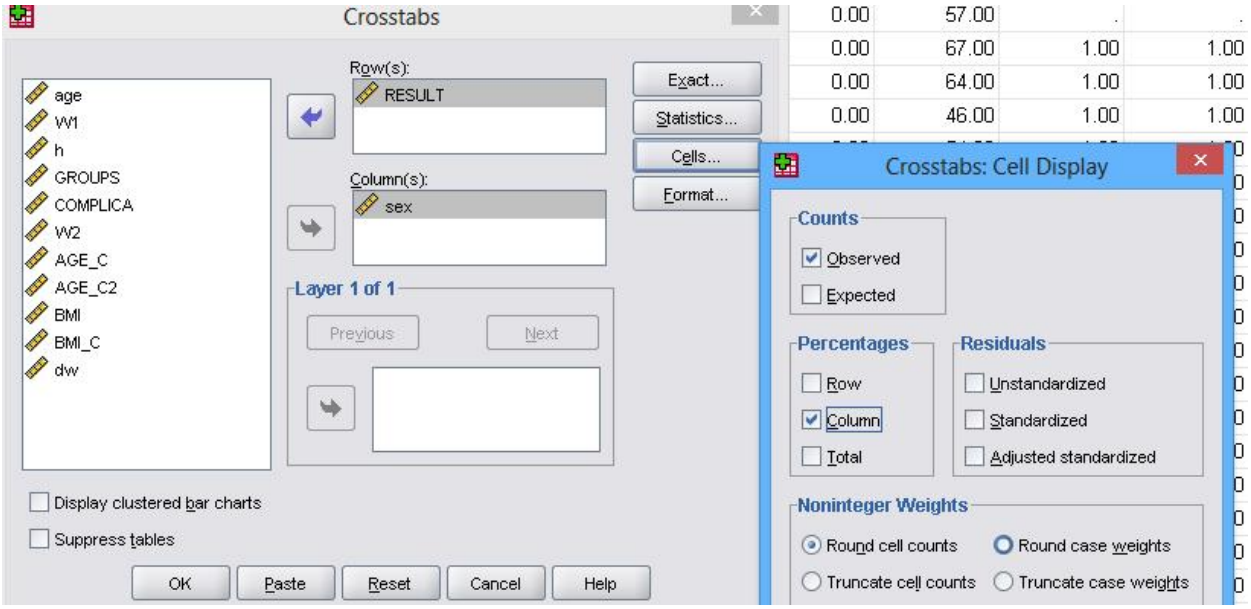
آزمون کای اسکوئیر (chi-square)

مقایسه یک متغیر کیفی در دو گروه :

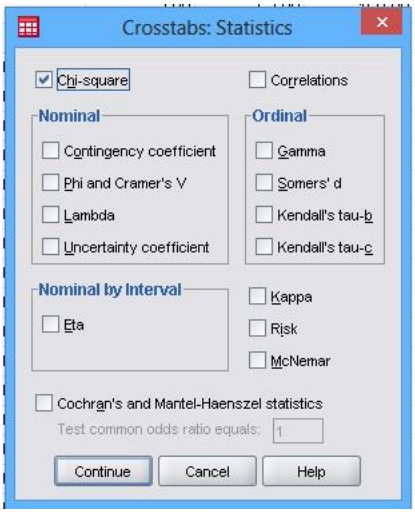
اگر بخواهیم نتیجه درمان را در دو گروه جنسی مقایسه کنیم . از آزمون chi- square استفاده می کنیم.



با انتخاب مسیر فوق پنجره ذیل ظاهر می شود که دو متغیر مورد نظر که نتیجه درمان و جنس را انتخاب می کنیم . و گزینه cell را برای تعیین در صد ها انتخاب نموده و نسبت به متغیر گروه درصد می گیریم . در اینجا چون می خواستیم دو جنس را از نظر نتیجه درمان مقایسه کنیم و جنس را در ستون قرار دادیم لذا column را علامت می زنیم .



بعد از ادامه گزینه Statisticsرا کلیک نموده در پنجره ذیل chi- Squareرا فعال می کنیم .



به این ترتیب خروجی به صورت ذیل خواهد شد.

RESULT * sex Crosstabulation

			sex		Total
			male	female	
RESULT	poor	Count	۸	٤	۱۲
		% within sex	٤۰.۰%	۲۰.۰%	۳۰.۰%
	moderate	Count	۷	۸	۱۵
		% within sex	۳۵.۰%	٤۰.۰%	۳۷.۵%
	good	Count	۵	۸	۱۳
		% within sex	۲۵.۰%	٤۰.۰%	۳۲.۵%
Total		Count	۲۰	۲۰	٤۰
		% within sex	۱۰۰.۰%	۱۰۰.۰%	۱۰۰.۰%

Chi-Square Tests

	Value	df	Asymp. Sig. (۲-sided)
Pearson Chi-Square	۲.۰۹۲[a]	۲	.۳۵۱
Likelihood Ratio	۲.۱۲٤	۲	.۳٤٦
Linear-by-Linear Association	۱.۹۱۳	۱	.۱٦۷
N of Valid Cases	٤۰		

a. ۰ cells (.۰%) have expected count less than ۵. The minimum expected count is ٦.۰۰.

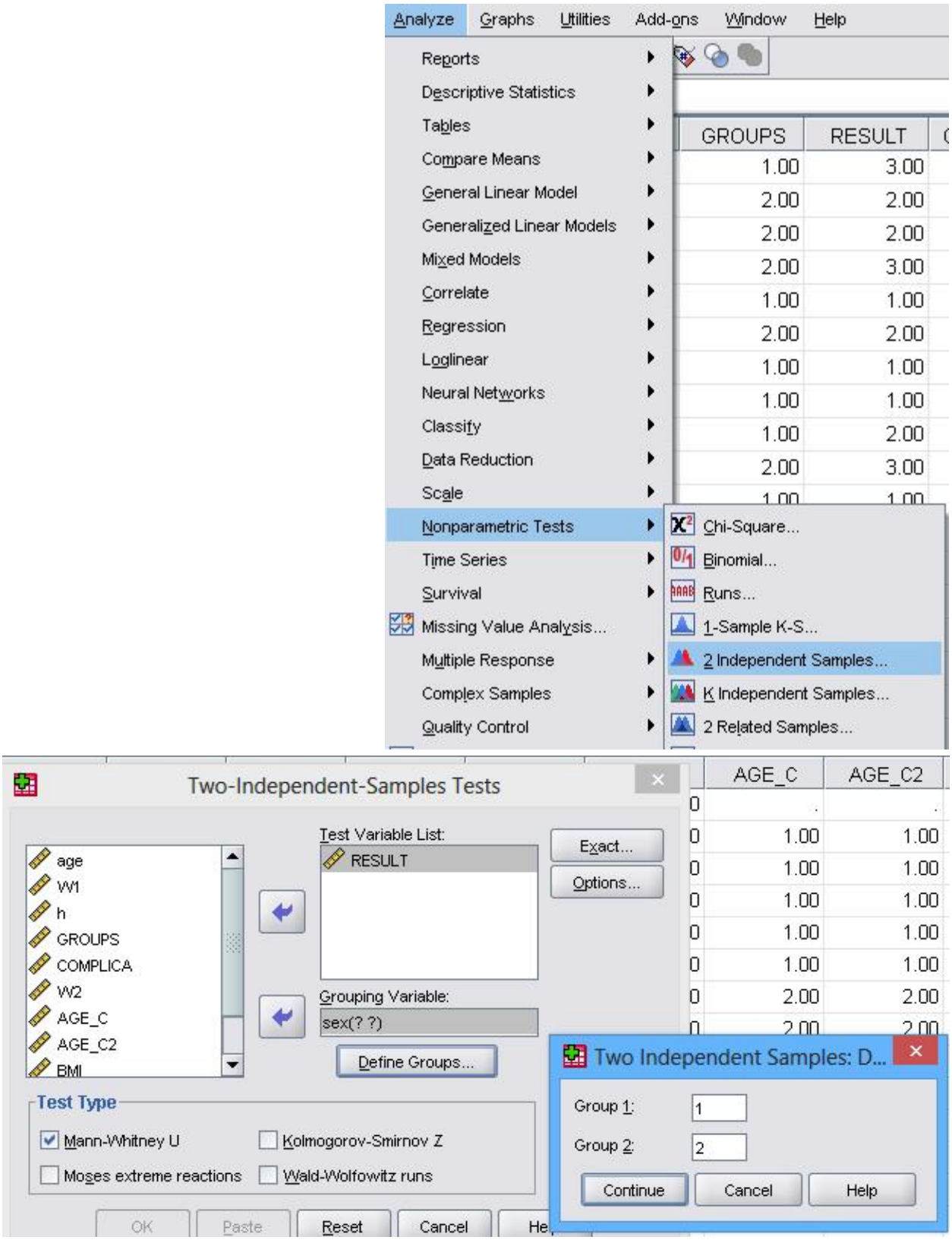
در جدول اولی مشخص می شود که ۴۰/۰ درصد افراد مذکر نتیجه درمان ضعیف داشته اند در صورت که در افراد مونث تنها ۲۰/۰ درصد نتیجه درمان ان ها ضعیف بوده است . در جدول سطر اول در ذیل ستون Asymp. Sig. مقدار p- را ۰.۳۵۱ به دست آورده که حکایت از عدم معنی داری است . یعنی نتیجه درمان در دو گروه جنسی تفاوت معنی داری ندارد .

شرایط آزمون :

در جدول فوق اگر داخل پرانتز بیشتر از ۲۰٪ باشد این آزمون اعتبار ندارد . همچنین مقدار عددی که در سطر دوم جدول نوشته شده نباید از ۱ کمتر باشد. در مثال فوق ۶ است .(The minimum expected count is ۶.۰۰.)

تذکر مهم :

برای متغیر های کیفی با مقیاس رتبه ای می توان همانند متغیر های کمی غیر نرمال رفتار کرد و از آزمون من – ویتنی استفاده نمود . این آزمون دارای توان بیشتری می باشد.


Analyze
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Missing Value Analysis...
Multiple Response
Complex Samples
Quality Control
Chi-Square...
Binomial...
Runs...
1-Sample K-S...
2 Independent Samples...
K Independent Samples...
2 Related Samples...
GROUPS
RESULT
AGE_C
AGE_C2
Two-Independent-Samples Tests
Test Variable List:
RESULT
Grouping Variable:
sex(? ?)
Define Groups...
Exact...
Options...
age
W1
h
GROUPS
COMPLICA
W2
AGE_C
AGE_C2
BMI
Test Type
Mann-Whitney U
Kolmogorov-Smirnov Z
Moses extreme reactions
Wald-Wolfowitz runs
OK
Paste
Reset
Cancel
Two Independent Samples: D...
Group 1:
1
Group 2:
2
Continue
Cancel
Help

Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
RESULT	male	٢٠	١٨.١٠	٣٦٢.٠٠
	female	٢٠	٢٢.٩٠	٤٥٨.٠٠
	Total	٤٠		

Test Statistics[b]

	RESULT
Mann-Whitney U	١٥٢.٠٠٠
Wilcoxon W	٣٦٢.٠٠٠
Z	-١.٣٧٩
Asymp. Sig. (٢-tailed)	.١٦٨
Exact Sig. [٢*(١-tailed Sig.)]	.٢٠١[a]

a. Not corrected for ties.

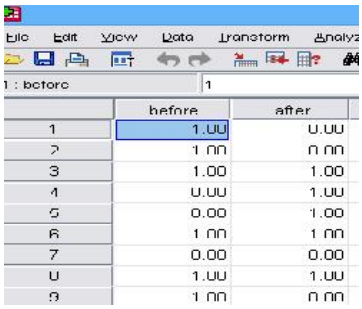
b. Grouping Variable: sex

که در این آزمون p-value=۰.۱۶۸ می شود . که البته در این آزمون هم معنی دار نیست . ولی p-value آن کمتر از **آزمون** chi- squareمی باشد .

آزمون مک نیمار

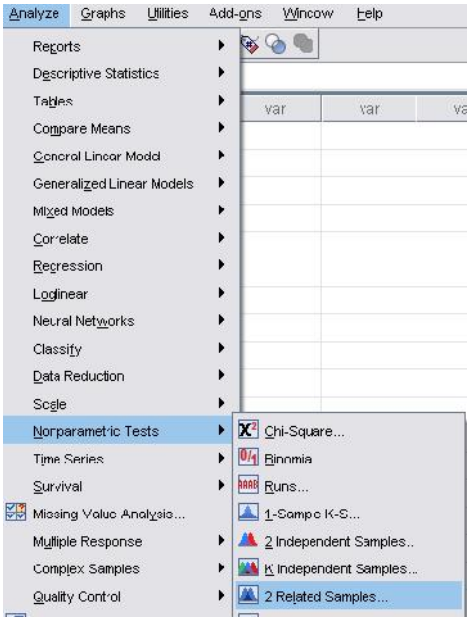
مقایسه یک متغیر کیفی دو حالتی دردو گروه وابسته (مشاهدات زوجی)

فرض کنید تعدادی از افراد را در مورد نگرش آن ها مورد بررسی قرار داده ایم .برخی دارای نگرش مثبت و برخی منفی می باشد . پس از یک دوره آموزش مجددا از همان افراد نگرش سنجی می شود و پاسخ های بعد نیز به صورت مثبت یا منفی است . می خواهیم آزمون کنیم آیا نوع نگرش تغییر یافته است یا خیر ؟



	before	after
1	1.00	0.00
2	1.00	0.00
3	1.00	1.00
4	0.00	1.00
5	0.00	1.00
6	1.00	1.00
7	0.00	0.00
8	1.00	1.00
9	1.00	0.00

برای مقایسه این گونه داده ها از آزمون MC Nemar (مک نیمار) استفاده می شود .



before & after

before	after	
	no	yes
no	5	7
yes	1	7

Test Statistics[b]

	before & after
N	20
Exact Sig. (2-tailed)	.070[a]

a. Binomial distribution used.

b. McNemar Test

کار عملی جلسه اول:

تعریف متغیر ها و ورود داده ها

۱- نرم افزار SPSS را اجرا نموده و با توجه به پرسشنامه ای که در اختیار دارید، متغیرها را تعریف کنید.
۲- برای متغیرهای کیفی کدها را تعریف کنید.
۳- داده های پرسشنامه را وارد نرم افزار SPSS نمایید.
۴- فایل ایجاد شده را در محل مناسبی به نام Tamrin۱ ذخیره نمائید.
۵- از SPSS خارج شده و مجددا SPSS را اجرا کنید.
۶- فایل داده ها را بخوانید.
۷- داده های پرسشنامه را وارد نرم افزار EXCEL نماییدو آن را save کنید.(۹۷-۲۰۰۳)
۸- فایل ایجاد شده را با spss بخوانید.

کار عملی جلسه دوم

توزیع فراوانی و شاخص های آماری

۱- توزیع فراوانی را برای متغیرهای جنس، گروه درمانی و عارضه بدست آورید.
۲- میانگین ، واریانس ، انحراف معیار، حداقل ، حداکثر ، میانه، چارک اول، چارک سوم، صدک ۹۷ و صدک سوم را برای متغیر وزن ثانویه بدست آورید.
۳- میانگین، انحراف معیار، حداقل و حداکثر را برای متغیرهای کمی (سن, وزن اولیه و وزن ثانویه) بدست آورید.
۴- میانگین، انحراف معیار، حداقل و حداکثر را برای متغیرهای کمی (سن, وزن اولیه و وزن ثانویه) بر حسب جنس بدست آورید.
۵- برای متغیرهای جنس و نتیجه درمان نمودارستونی Bar chart رسم کنید.(درصدی)
۶- نتیجه درمان را در دو گروه در یک نمودارستونی نشان دهید.
۷- هیستوگرام داده های سن را رسم کنید.
۸- نمودار جعبه ای (Box-plot) را برای وزن ثانویه رسم کنید.

جلسه سوم :

مدیریت داده ها

۱- افراد را در گروه های سنی زیر دسته بندی کنید و آن را age_c بنامید. (طبقه بندی کردن داده های کمی) و سپس کد های جدید را تعریف کنید.

(۲۵-۳۴)، (۳۵ – ۴۴)، (۴۵ – ۵۴)، (۵۵ - ۶۴)

۲- جدول توزیع فراوانی age_c را بدست آورید

۳- گروه های سنی (۴۵ – ۵۴) را با گروه سنی (۵۵ – ۶۴) ادغام کنید و نام آن را age_c۲ بنامید. مجددا کدها را برای گروه بندی جدید تعریف کنید.

۴- توزیع فراوانی age_c۲ را بدست آورید.

۵- شاخص BMI را با توجه به فرمول $BMI = \frac{weight(kg)}{(height(m))^2}$ یا (قد به متـر)/ وزن بـه کیلـوگرم= BMI بـرای وزن اولیه محاسبه کنید و آن را BMI۱ بنامید.

۶- میانگین، انحراف معیار، حداقل و حداکثر را برای BMI۱ بدست آورید.

۷- با توجه به BMI۱ افراد را بصورت لاغر، طبیعی و چاق طبقه بندی کنید و آن را BMI۱_C بنامید.(فاصله ها بصـورت ذیل باشد)

- کوچکتر یا مساوی ۱۸/۵ را لاغر
- بین ۱۸/۵ تا ۲۵ را طبیعی
- بیشتر از ۲۵ را چاق

۱۰- میانگین و انحراف معیار سن را برای کسانیکه نتیجه درمان آنها خوب بوده است، بدست آورید.

۱۱ – میانگین ، میانه و انحراف معیار سن را برای کسانیکه چاق و نتیجه درمان آنها خوب بوده است بدست آورید.

سوالات پیشرفته:

۱۲- اگر متغیر date۲ تاریخ ترخیص و متغیر date۱ تاریخ پذیرش باشد طول دوره بستری را بدست آورید .

۱۳- مشخص کنید چند نفر اطلاعاتشان کامل نیست ؟

۱۴- فایل data۱ مربوط به گروه۱ و data۲ مربوط به گروه ۲ را به هم متصل کنید و متغیری به نام گروه را ایجاد کنید .

۱۵- فایل w را به فایل datawithout w الحاق کنید.

۱۶ –فایل datanew دادهای مربوط به تعدادی از افراد می باشد که بسیاری از ویژگی های آن هـا در فایـل data ذخیـره شده است . داده های datanew را توسط الحاق سایر متغیرها که در فایل dataذخیره شده است ،کامل کنید .

جلسه چهارم:مقایسه های دو گروهی

الف) متغیر کمی

فایل section۴ را باز کنید و تمرین های ذیل را انجام دهید .

۱- آیا وزن اولیه ، وزن ثانویه، سن و BMI دارای توزیع نرمال می باشند.

۲- میانگین سن و همچنین BMI را در دو گروه درمانی مقایسه کنید. (آزمون کنید)

۳- میانگین وزن اولیه را در دو گروه مقایسه کنید. (آزمون کنید)

۴- میانگین وزن ثانویه را در دو جنس مقایسه کنید. (آزمون کنید)

۵- آیا بین وزن اولیه و ثانویه تفاوت وجود دارد؟

۶- آیا در هر یک از دو گروه درمانی A و B ، تفاوتی بین وزن اولیه و ثانویه وجود دارد؟

۷- کاهش وزن را بدست آورده و بعنوان یک متغیر محاسبه و به نام dw ثبت کنید.

۸- آیا کاهش وزن در دو گروه درمانی تفاوت دارد؟

۹- با فرض این که وزن اولیه دارای توزیع نرمال نباشد آیا تفاوتی بین وزن اولیه و ثانویه وجود دارد؟

ب) متغیر کیفی:

۱- آیاجنس در دو گروه درمان یکسان توزیع شده است؟

۲-آیا تفاوتی بین وجود عارضه در دو گروه وجود دارد؟

۳-آیا تفاوتی بین نتیجه درمان در دو گروه وجود دارد؟

۴-سوال ۳ را در هر یک از دو جنس پاسخ دهید.

۵-آیا تفاوتی بین سطوح BMI (BMI۱_C) در دو گروه وجود دارد ؟

۶-آیا بین گروه سنی(age_c۲) و عوارض رابطه ای وجود دارد؟

۷- آیا بین سطوحBMI۱_c اولیه و گروه سنی(age_c۲) ارتباطی وجود دارد؟

۸ – در افراد مذکر چه ارتباطی بین سطوح BMI۱_c و نتیجه درمان وجود دارد؟

۹- سوال های ۵ تا ۸ را با آزمون های من- ویتنی یا کروسکال – والیس انجام دهید .