راهنمای نرم

افزارآماری spss.16

همراه با

مثالهای کاربردی و تحلیل خروجی ها

- **آمار توصیفی**
- **تحلیل همبستگی**
- **جداول توافقی**
- **شکلهای آماری**
- **آزمون فرض ها**
- **آنالیز واریانس**

- رگرسیون خطی
- آزمون های ناپارامتری
- مثالهای کاربردی
- تحلیل محتوای خروجی ها

علی صادقی

فهرست

عنوان صفحه

پیشگفتار

فصل اول - کلیات علم آمار

فصل دوم -معرفی نرم افزار spss.16 و قابلیتهای آن

2-1- انواع پنجره های موجود درspss ؛ 2-2- منوها در پنجره های spss

فصـــــــل ســـــــوم- ورود داده هـــــــا و ویـــــــرایش آنها

3-1-مقدمه؛ 3-2- وارد کردن داده ها؛ 3-3- امکانات منو File؛

3-4-امکانات منو Edit؛ 3-5- امکانات منو View؛

3-6- امکانات منو Utilites؛ 3-7- امکانات منو Windows

فصـــــل چهـــــارم- آمـــــار توصـــــیفی در نـــــرم افـــــزار spss

4-1-مفدمه؛ 4-2- آمار توصیفی برای متغیرهای گسسته با فرمان Frequencis؛

4-3- آمار توصیفی برای متغیرهای پیوسته با استفاده از زیر منو Descriptives؛

3-5- آمار توصیفی برای متغیرهای پیوسته در سطوح متغیرهای گسسته با استفاده از زیر منو Explor؛

فصـــــــل پـــــــنجم- ویـــــــرایش مقـــــــادیر داده ها

5-1- مقدمه؛ 5-2- منو Data و زیـر منوهـای آن؛ 5-3-منـو Transform و زیـر منوهای آن

فصل ششم- جداول

6-1- مقدمه؛ 6-2- رابطه میان متغیرهای گسسته؛

6-3- تحلیل جداول توافقی با داده های اسمی؛

6-4- تحلیل جداول توافقی با داده های ترتیبی؛ 6-5- نمایش مقادیر خانه های جداول توافقی

فصل هفتم- رسم نمودارها

7-1- مقدمه؛ 7-2- رسم نمودار میله ایی؛ 7-3- رسم نمودار خطی؛

7-4- رسم نمودار سطحی؛ 7-5- رسم نمودار دایره ایی؛ 7-6- ویرایش نمودارها

فصل هشتم- آزمون فرضهای میانگین

8-1- مقدمه؛ 8-2- میانگین ها؛8-3- آزمون T برای مقایسه میانگین یک نمونه؛

8-4- آزمون T برای مقایسه میانگین دو نمونه مستقل؛

8-5- آزمون T برای دو نمونه جفت شده

فصل نهم- آنالیز واریانس

9-1- مقدمه؛ 9-2- آنالیز واریانس یک طرفه ؛ 9-3- آنالیز واریانس دو طرفه

فصل دهم- تحلیل همبستگی

10-1- مفاهیم اولیه؛ 10-2- ضریب همبستگی بین دو متغیر؛

10-3- ضریب همبستگی جزیی متغیرها

فصـــــــــــــل یـــــــــــــازدهم-رگرســـــــــــــــیون خطی

11-1- مبانی اولیه؛ 11-2-رگرسیون خطی؛ 11-3- نحوه ورود متغیرهـای مسـتقل بـه مدل رگرسیونی؛

11-4- محاسبه آماره ها در رگرسیون خطی؛ 11-5- رسـم نمودارهـا در رگرسـیون خطی؛

11-6- بررسی فرض خطی بودن مدل رگرسیونی برازش شده؛

11-7- ذخیره اطلاعات فرعی در رگرسیون خطی

فصل دوازدهم- آزمونهای ناپارامتری

12-1- مفاهیم اولیه؛ 12-2- فرمان **chi square**؛ 12-3- فرمان **Binomial**؛

12-4- فرمان **Runs**؛ 12-5- فرمان **One- sample K-S**؛

12-6- فرمان **2 Independent samples…** ؛

12-7- فرمان**k Independent samples…** ؛

12-8- فر مان**2 Related samples…** ؛

12-9- فرمان**Independent amples…**

پیوست 1- نصـب و راه انـدازی نـرم افـزار SPSS 16 تحـت ویندوز

پیوست 2- داده های پیوست فایل exam

پیوست 3- مبانی انجام پیمایش های میدانی

پیشگفتار

در قرن اخیر استفاده از تکنیکهای آماری با توجه به سرعت روزافـزون تکنولـوژی مـورد توجـه محققـان و پژوهشگران شاخه‌های مختلف علوم قرار گرفته است. در این میان استفاده از نرم افزارهای آماری در استخراج نتایج مورد نظر باعث کاهش زیاد در زمان و هزینهٔ تحقیقات آماری شده است. از جملـه نـرم افزارهـای مـورد توجه و استفاده پژوهشگران و محققان شاخه‌های کاربردی علوم انسانی، پزشـکی و فنـی مهندسـی نـرم افـزار SPSS است.

کلمه SPSS مخفف کلمهٔ Statistical Package For The Social Science است که اولین بار با توجه به کاربرد وسیع روش‌های آماری توسط انجمن علوم اجتماعی آمریکا در دهه هفتاد سفارش داده شد و با توجه به اینکه در دههٔ 90 اولین نرم افزار آماری تحت سیستم عامـل ویندوز نسـخهٔ 3/1 بـود مـورد توجـه تمـامی محققان در سراسر دنیا قرار گرفت و امروزه نیز شهرت جهانی این نرم افزار با توجه به سادگی استخراج نتـایج در آن و تحلیل نتایج و ارائهٔ خروجی‌ها به زبان ساده باعث شده به عنوان پرکاربردترین نرم افزار آمـاری مـورد استفاده قرار گیرد.

نسخه نهایی SPSS که هم اکنون توسط شرکت SPSS به بازار عرضـه گردیـده اسـت نسـخه SPSS.16 می‌باشد که با تغییرات اندکی نسبت به دیگر نسخه‌های قدیمی تر آن ارائه شده است. دلیل اصلی نگارش ایـن

کتاب به دلیل دغدغه ای بود که نویسندگان را پس از تدریس دوره‌های مختلف مقدماتی و پیشرفته این نرم افزار و روبرو شدن با افراد علاقه‌مند به یادگیری با آن مواجه کرد.

تفاوت اصلی نرم افزار SPSS و به طور کلی نرم افزارهای تخصصی در حصول به نتایج و خروجی‌های تخصصی نرم افزار نیست بلکه در نیاز به تحلیل خروجی‌ها است به این معنی که وظیفه اصلی کاربر بعد از خروجی گرفتن از نرم افزار SPSS تازه شروع می‌گردد و باید نتایج را استنباط نماید و این در حالی است که اکثر کتاب‌های تألیف و ترجمه شده موجود در بازار تنها نحوه کار با نرم افزار را آن هم در حد آشنایی مقدماتی با ابزارهای SPSS مطرح می‌نمایند و کاربر را در ارائه و تحلیل نتایج تنها می‌گذارند. به همین دلیل بسیاری از کاربران نتایج مطلوب نظر را از مطالعه کتاب‌های موجود به دست نمی‌آورند.

در این کتاب سعی گردیده است ابزارهای مختلف نرم افزار SPSS همراه با فلسفه وجودی و نمایه ایی از کاربرد عملی ارائه گردد. همچنین خروجی‌ها با بیان تفسیر هر یک به منظور شکل دهی قدرت تحلیل خروجی در کاربران ارائه گردیده است. در فصل اول این کتاب به منظور یکسان سازی مفاهیم آماری یک خلاصه از فلسفه وجودی علم آمار و تکنیک‌های آماری بیان شده که مطالعه آن به تمامی خوانندگان در کلیه رشته‌ها توصیه می‌گردد. سعی شده است ابزارهایی از نرم افزار SPSS با توجه به حجم وسیع آنها مورد توجه قرار گیرد که بیش از سایر ابزارها مورد توجه پژوهشگران و محققان و دانشجویان شاخه‌های مختلف علوم انسانی، علوم پزشکی و فنی و مهندسی است و با توجه به وسعت کاربرد هر یک از ابزارها در شاخه‌های مختلف علوم سعی شده است از مثالهای عملی و راحت و قابل لمس و شبیه سازی در همان شاخه‌ها استفاده گردد. در این کتاب سعی شده است تا با بهره گیری از تجربیات نویسندگان در تدریس این نرم افزار روش انجام تکنیکهای آماری متداول و مورد نیاز پژوهشگران و دانشجویان بخصوص در دوره کارشناسی ارشد پوشش داده شود. همچنین سعی شده در پیوستهای کتاب به موارد مرتبط از جمله راهکارهای انجام یک پیمایش به صورت مرحله ایی به صورت تفضیلی اشاره شود تا خواننده ضمن یادگیری نرم افزار موارد جانبی و با اهمیت یک تحقیق صحیح را بیاموزد.

نوشته حاضر با توجه به تجربیات نویسندگان در نحوه آموزش این نرم افزار تدوین گردیده و بدیهی است

که خالی از اشکال نبوده از ضعف‌هایی برخوردار است. از کلیه اساتید، محققان و دانشجویان عزیـزی کـه ایـن کتاب را مطالعه می‌نمایند خواهشمنداست در صورت برخورد با هرگونه ضعفی نویسندگان را از طریق ایمیـل statsadeghi@Gmail.com از نظرات اصلاحی خود بهره مند سازند.

علی صادقی

تابستان 86

فصل اول

کلیات علم آمار

اهداف آموزشی فصل اول

- فراگیر پس از مطالعه مطالب این فصل کتاب باید بتواند:

- مفهوم جامعه آماری را درک و مثالهایی را در مورد جامعه آماری ذکر نماید.

- انواع متغیرهای آماری را بشناسد و با ذکر مثال هر یک را شرح دهد.

- هدف علم آمار را بیان کند.

- نمونه گیری آماری را شرح دهد.

- وظایف آمار توصیفی و آمار استنباطی را با یکدیگر مقایسه کند.

- پارامترهایی که در آمار توصیفی جهت تلخیص داده ها استفاده می شود را نام برده و هر یک را توضیح دهد.

1-1مقدمه:

نرم افزار spss یک نرم افزار آماری است که به کاربران امکان می دهد در زمان کوتاه و همچنین با دقت زیـادی داده های مورد نیاز خود را مورد تجزیه و تحلیل قرار دهند اما این به آن معنی نیست که می توان فرآیند تجزیه و تحلیـل آماری در این نرم افزار را بدون داشتن اطلاعات تئوریک از علم آمار به طور کامل و درست انجام داد.

انجام یک تحیلی آماری در این نرم افزار مستلزم اطلاعات کافی از فرآیند انجام تحلیل می باشد به همین منظـور بـه همه خوانندگان این کتاب از جمله کسانی که اطلاعات کافی از تئوریهای آماری ندارند پیشنهاد می شود قبل از انجام

تحلیل و یا خوانندن این کتاب پیش نیاز آماری را که مقدمات روشهای مورد نظر را عنوان کرده اند مطالعه نمایند و با اطلاعات کافی نسبت به یاد گیری نرم افزار spss همت گمارند.

در این فصل سعی شده است در حد توان خلاصه ایی از مفاهیم مورد نیاز اولیه کاربران در زمینه کار با نرم ازار عنوان شود اما این به آن معنی نیست که کاربرانی که اطلاعات زیادی در مورد مبانی علم آمار و روشهای آماری ندارند از مطالعه کتب مرجع در این زمینه بی نیاز خواهند شد.

فلسفه وجودی علم شناخت دنیایی است که در آن زندگی می‌کنیم. در مورد علم آمار ساده ترین تعریفی که ارائه گردیده عبارت است از: «شناخت جامعهٔ مورد بررسی» ،اما عبارت جامعه در علم آمار تعریف مشخصی دارد. جامعه در علم آمار عبارت است از مجموعه ای از عناصر که حداقل در یک خصوصیت با یکدیگر مشترک باشند.

برای نمونه در یک کلاس 12 نفره آموزش نرم افزار SPSS تمامی کاربران در یک خصوصیات مشترک هستند و آن نیاز به یادگیری نرم افزار می‌باشد. البته کاربران خصوصیات مشترک دیگری نیز مانند رشته، سطح تحصیلات، جنسیت و ... می‌توانند داشته باشند.

برای نمونه مردم ایران یک جامعه آماری را تشکیل می دهند که هر یک از ایرانیان یک عضو آن می باشند و حداقل در یک خصوصیت ایرانی بودن مشترک هستند

همانطور که عنوان شد یک جامعهٔ آماری تشکیل شده از یکسری عناصر و یکسری خصوصیت که ویژگی‌های عناصر جامعهٔ مورد بررسی هستند. برای نمونه در محل کار شما، کلیه همکاران و خود شما عناصر جامعهٔ آماری و خصوصیاتی مانند: سن، جنسیت، سطح تحصیلات، محل تولد، ردهٔ سازمانی و ... خصوصیت جامعهٔ آماری نیروی انسانی سازمان شما را تشکیل می‌دهند.

حال که مشخص شد یک جامعهٔ آماری تشکیل شده از یکسری عناصر و خصوصیات، می‌توان دریافت که منظور از

شناخت جامعهٔ آماری که فلسفه وجودی علم آمار است در حقیقت شناخت عناصر جامعـهٔ مـورد بررسـی و خصوصـیات آنهاست.

در زبان تخصصی آمار به عناصر جامعـهٔ مـورد بررسـی داده (Data) و بـه خصوصـیات جامعـهٔ مـورد بررسـی متغیـر (Variable) اتلاق می‌شود. پس با توجه به تعریف فوق زمانی می توان ادعا کرد که شناخت نسبت به یک جامعه آماری به وجود آمده است که نسبت به خصوصیات آن جامعه و یا متغیرهای آن جامعه شناخت ایجاد شده باشد. شما چه زمانی می توانید بگویید مردم یک نقطه از ایران را به خوبی می شناسید؟ هنگامی می توانید این ادعا را مطرح نماید که نسبت به خصوصیات مردم آن نقطه از کشور شناخت کافی داشته باشید و یا در اصطلاح آماری متغیرهای رفتاری مردم را خوب بشناسید. پس به صورت کلی می توان اعا کرد که منظور از شناخت جامعه مورد بررسی شناخت متفیرهای مورد بررسی در آن جامعه است. اما ذکر این نکته ضروری است که منظور از شناخت متغیرها در علم آمار شامل شناخت یـک بعـدی متغیرها یعنی رفتار هر متغیر به تنهایی و همچنین شناخت دو یا چند بعدی متغیرها یا شناخت رابطه متغیرها با یکدیگر می باشد.

حال زمان آن فرا رسیده است که با تعریف جامع تری از علم آمار آشنا شوید. آمار، ابزاری برای بدسـت آوردن رابطـهٔ میان متغیرهاست. منظور از بدست آوردن رابطهٔ میان متغیرها یعنی تأثیرگذاری یک متغیر بر روی متغیر یـا متغیرهـای دیگر، برای مثال هنگامی که می‌گویند رابطه میان متغیر جنسیت و فشار خون چگونـه اسـت مـی‌خواهنـد دریابنـد آیـا جنسیت افراد بر روی فشار خون آنها تأثیرگذار بوده است یا خیر، هنگامی که رابطهٔ میان مصرف سیگار و ابتلا به سرطان را استخراج می‌کنیم در حقیقت تأثیر متغیر مصرف سیگار را که دارای دو سطح سیگار کشیدن و سیگار نکشیدن است بر روی متغیر ابتلا به سرطان که دارای دو سطح مبتلا شده و سالم است بدست می‌آوریم.

طبق تعریف ارائه شده علم آمار رابطهٔ میان متغیرهای موجود در جامعه را از طریق شناسایی تأثیرگـذاری گروهـی از متغیرها روی متغیرهای دیگر بدست می‌آورد، در علم آمار متغیرهـای تأثیرگـذار (ماننـد جنسـیت و مصـرف سـیگار در مثالهای بالا) متغیر مستقل (Independent Variable) نامیده می‌شوند و متغیرهایی تأثیرپذیر (مانند فشار خون و ابتلا به سرطان) متغیرهای وابسته(Dependent Variable)نامیده می‌شوند.

پس تعریف جامع تر از علم آمار عبارت است بدست آوردن تأثیر متغیرهای مستقل بر روی متغیرهای وابسته، سوال بسیار مهمی که ممکن است مطرح شود این است که اگر هدف نهایی علم آمار شناخت جامعهٔ مورد بررسی از طریق بدست آوردن تأثیر متغیرها (تأثیر متغیر مستقل بر روی وابسته) است پس چرا تکنیک‌های مختلف آماری نظیر، رگرسیون، آنالیز واریانس، ناپارمتری، آزمون فرض، سریهای زمانی و ... به وجود آمده اند؟ مگر هدف مشخص نیست؟

برای پاسخ گویی به این سوال باید خاطرنشان کرد که متغیرهای مستقل انواع گوناگونی دارند و متغیرهای وابسته هم انواع گوناگونی دارند که این امر باعث شده هرگاه بخواهیم تأثیر متغیر مستقل از نوع خاصی را بر روی متغیر وابسته از نوع خاصی بدست آوریم از تکنیک متفاوتی برای به دست آوردن رابطه استفاده نمائیم. ابتدا لازم است به تعریف و شناسایی انواع متغیرهای آماری بپردازیم تا بعد از آن فلسفه وجودی استفاده از هر یک از تکنیکهای آماری مشخص گردد.

1-2-انواع متغیرها :

در علم آمار فارغ از اینکه متغیرها مستقل یا وابسته باشند به دو گروه بزرگ تقسیم می‌شوند :

*متغیرهای پیوسته (Continuous Variables)

*متغیرهای گسسته (Discreet Variables)

متغیرهای پیوسته متغیرهایی هستند که به بیان ساده می‌توانند بی نهایت عضو داشته باشند، برای نمونه سن افراد یک متغیر پیوسته است، فشار خون، قد، وزن و ... همه متغیرهای پیوسته هستند. در تمامی آنها هر عنصر جامعهٔ مورد بررسی در صورتی که ابزار اندازه گیری دقیق باشد می‌تواند بی نهایت عضو را قبول نماید. برای مثال وقتی که قد شخصی را 165cm در نظر می‌گیریم، چون ابزار دقیق تری نداشته ایم به 165cm بسنده کرده ایم این در حالی است که قد شخص cm ... 165/2436 است که اعداد اعشاری می‌توانند تا بی نهایت ادامه پیدا کنند به همین دلیل امکان ندارد قد دو نفر یکسان باشد و متغیر قد می‌تواند بی نهایت عضو بپذیرد. همانطور که مشاهده می‌شود متغیرهایی مانند زمان، سن، قند خون، فشار خون، سرعت، ضریب هوشی و ... متغیر وابسته هستند.

متغیرهای گسسته برخلاف متغیرهای پیوسته تعداد سطوح محدودی دارند. برای مثال متغیر جنسیت دارای دو سطح زن و مرد است و هر شخص در جامعهٔ مورد بررسی یا زن خواهد بود یا مرد و یا متغیر سطح تحصیلات دارای سطوح محدود، زیردیپلم، دیپلم و بالاتر از دیپلم است و مانند متغیر پیوسته بی نهایت سطح ندارد. متغیرهای نظیر جنسیت، سطح تحصیلات، تعداد فرزند، وضعیت تأهل، استان محل زندگی و ... از این دسته هستند.

متغیرهای گسسته خود به دو دسته متغیرهای اسمی گسسته و ترتیبی گسسته تقسیم می‌شوند. متغیرهایی مانند سطح تحصیلات و تعداد فرزند که بین سطوح آنها یک رابطه منطقی برقرار است متغیر گسسته ترتیبی هستند برای نمونه کسی که سطح تحصیلات دیپلم دارد حتماً دارای سطح تحصیلات بالاتر از کسی است که سطح تحصیلات زیر دیپلم داشته است و یا کسی که 4 فرزند دارد از کسی که 3 فرزند دارد تعداد فرزند بیشتری دارد. به این متغیرهای گسسته که بین سطوح آنها یک ترتیب منطقی برقرار است متغیر گسسته ترتیبی گویند و به متغیرهای گسسته ای که بین سطوح آنها رابطهٔ منطقی (بزرگتر، کوچکتر، مساوی) در صفت مورد بررسی برقرار نیست مانند جنسیت، تأهل و محل زندگی متغیر گسسته اسمی گویند.

با توجه به تعاریف فوق متغیرها به صورت ذیل تقسیم بندی می‌شوند :

- متغیرهای آماری
  - پیوسته (Scale): مانند سن، وزن، قد، درآمد، طول قطر یک لوله، درجه هوا، سرعت و ...
  - گسسته
    - ترتیبی (Ordinal): مانند سطح تحصیلات، تعداد فرزند، ردهٔ سازمانی، تعداد طبقات ساختمان و ...
    - اسمی (Nominal): مانند: جنسیت، رنگ، وضعیت تأهل، استان تولد، محل زندگی (شهر،

حال با توجه به تعریف انواع متغیرهای آماری می‌توان دلیل وجود تکنیک‌های مختلف آماری را توجیه نمود، همانطور که گفته شد هدف بدست آوردن رابطهٔ میان متغیرهای مورد بررسی از طریق شناسایی تأثیر متغیرهای مستقل بـر روی متغیرهای وابسته است. متغیرهای مستقل و وابسته می‌تواند هر یک از انواع پیوسته یا گسسـته باشـند. در جـدول ذیـل موارد استفاده تعدادی از تکنیک‌های آماری بیان شده است.

متغیر وابسته (تأثیر پذیر)

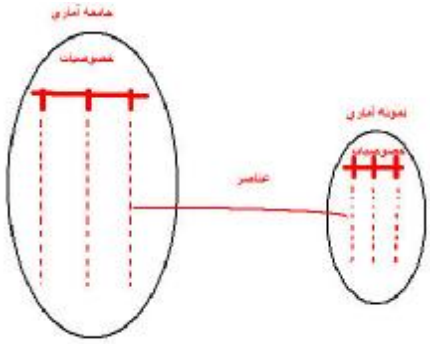
		گسسته	پیوسته
متغیر مستقل (تأثیرگذار)	گسسته	جداول توافقی (X^2) مدلهای همبستگی و ...	آزمون فرضی میانگین‌ها، آنالیز واریانس و ...
	پیوسته	مدلهای گسسته	رگرسیون سری زمانی و ...

برای نمونه هرگاه بخواهید تأثیر متغیر گسسته جنسیت را بر روی متغیـر گسسـته وضـعیت تأهـل بدسـت آوریـد از جداول توافقی استفاده می‌نمائید و یا اگر می‌خواهید تاثیر محل زندگی (شهر یا روستا) را بر روی فشـار خـون افـراد بـه دست آورید چون محل زندگی یک متغیر مستقل گسسته و فشار خون یک متغیر وابسته پیوسته است از آزمـون فـرض میانگین‌ها استفاده می‌نمائیم. یعنی از طریق آزمون فرض، میانگین فشـار خـون دو گـروه شـهری و روسـتایی را بـا هـم مقایسه می‌نمائیم و اگر برای نمونه میانگین فشار خون گروه روستایی به صورت معنی داری پایین تر بود به ایـن نتیجـه می‌رسیم که زندگی در روستا باعث بهبود فشار خون افراد می‌گردد.

همانگونه که گفته شد هدف علم آمار شناسایی جامعهٔ مورد بررسی است که برای شناخت آن نیازمند شناخت

متغیرها و رابطهٔ میان آنها در جامعهٔ آماری هستیم. علم آمار به دو شاخهٔ بزرگ آمار توصیفی و آمار استنباطی تقسیم می‌گردد که وظایف آنها منطبق بر فلسفه وجودی علم آمار است.

در بسیاری از موارد یک محقق به تمامی عناصر یک جامعهٔ مورد بررسی دسترسی ندارد برای مثال برای شناسایی وضعیت جنسیت و نسبت جنسیتی افراد یک کشور دسترسی به همهٔ افراد مقدور نیست و یا دسترسی به همهٔ افراد مستلزم وقت و هزینه‌های هنگفت است که از حدود اختیارات و تواناییهای محقق خارج می‌باشد برای رفع این مشکل شاخه ای از علم آمار به عنوان نمونه گیری (Sampling) پایه گذاری شده که محقق را قادر می‌سازد به جای بررسی همهٔ عناصر جامعهٔ مورد بررسی تعداد محدود و قابل کنترلی از آنها را در نمونه با تکنیکهای خاص طوری انتخاب نماید که نمونه به دست آمده با یک دقت قابل کنترل و مشخص نماینده عناصر جامعهٔ مورد بررسی باشند.



در حقیقت یک نمونهٔ آماری دارای تمام خصوصیات یک جامعهٔ آماری با دقت $(1-\alpha)\%$ است و تنها تعداد عناصر آن کمتر است. برای مثال یک تابلو بزرگ نقاشی و عکس آنرا در نظر بگیرید، عکس تابلو دارای تمام ویژگی‌های تابلو اما در مقیاسی کوچکتر است و مقدار α (خطا) بهایی است که شما به علت استفاده از یک نمونه و استنتاج نتایج در مورد جامعه پرداخت می‌نمایید. در صورتی که به تمامی عناصر جامعهٔ مورد بررسی دسترسی داشتیم مقدار خطا برابر عدد صفر بود.

حال این سوال مطرح می‌گردد، هنگامی که از جامعه یک نمونه با روش‌های نمونه گیری در علم آمار انتخاب نمودیم چه اقداماتی بر روی داده‌ها و متغیرها انجام می‌دهیم؟ هنگامی که یک نمونه گرفته می‌شود باید مراحل آمار توصیفی و

آمار استنباطی بر روی داده‌ها انجام گیرد.

همانطور که گفته شد هدف علم آمار شناخت جامعهٔ مورد بررسی از طریق شناخت متغیرها و رابطهٔ میان آنهاست و چون نمی‌توان تمامی عناصر جامعه را مورد بررسی قرار داد از یک نمونه آماری منتخب عناصر جامعه استفاده می‌شود. تعداد نمونه مورد نیاز یک تحقیق و فرمول مربوط به آن در پیوست نحوه انجام پیمایش های میدانی بیان شده است.

منظور از شناخت متغیرها بررسی یک بعدی متغیرها است که از طریق آمار توصیفی انجام می‌گیرد. آمار توصیفی در حقیقت بیان کننده وضعیت هر متغیر به صورت مجزا در جامعهٔ مورد بررسی در قالب جداول، نمودارها و پارامترهاست که از آن به عنوان تخلیص داده‌ها نیز نام می‌برند. با جداول فراوانی و نمودارهای آماری تا حدودی آشنایی دارید اما در ذیل تعدادی از پارامترهای مهم برای تخلیص اطلاعات متغیرها در قالب آمار توصیفی بیان شده است.

میانگین (Mean): میانگین یک متغیر پیوسته را بیان می‌نماید برای مثال میانگین فشار خون، سن، قد

میانه (Median): عددی را بیان می‌کند که نشان می‌دهد، 50 درصد داده‌های یک متغیر اندازه ای کوچکتر یا مساوی از آن و 50 درصد مقداری بزرگتر از آن دارند. برای مثال میانه نمرهٔ افراد یک کلاس برابر 15 است که نشان می‌دهد 50 درصد افراد نمره ای کمتر یا مساوی از 15 و 50 درصد افراد یک کلاس نمره ای بالاتر از آن کسب نموده اند و یا برای متغیر گسسته سطح تحصیلات، میانه برابر سطح تحصیلات دیپلم است که نشان می‌دهد 50 درصد افراد مورد بررسی تحصیلاتی کمتر یا برابر دیپلم و 50 درصد افراد تحصیلاتی بالای دیپلم دارند.

مد (Mode): نشان دهندهٔ سطح دارای بیشترین فراوانی در یک متغیر (معمولاً گسسته) است. برای مثال مد سطح تحصیلات در جامعه لیسانس است. به این معنی که افراد با سطح تحصیلات لیسانس از لحاظ فراوانی در جامعه بیشتر از دیگر سطوح تحصیلی هستند.

واریانس (Variance): یک معیار برای اندازه گیری پراکندگی (دوری) داده‌ها از میانگین آنهاست برای مثال در دایره علاوه بر مرکز که مرکز ثقل را نشان می‌دهد، شعاع نیز به عنوان یک پارامتر که نشان دهندهٔ بزرگی دایره است بکار می‌رود، وظیفه واریانس نیز درست مانند شعاع در دایره است و هرچه مقدار بزرگتری داشته باشد نشان دهنده دوری بیشتر داده‌ها از مرکز ثقل (میانگین) داده‌ها است.

انحراف معیار مقادیر (Standard deviation): مجذور واریانس و یک معیار برای شناسایی پراکندگی داده‌ها از میانگین آنهاست، برای مثال اگر در 2 کلاس مقدار انحراف معیار نمرات دانش آموزان کلاس اول 3/2 و کلاس دوم 5/5 باشد نشان می‌دهد که در کلاس دوم نمرات دانش آموزان از میانگین نمرات فاصلهٔ بیشتری داشته یا در اصطلاح دیگر در کلاس اول نمرات همگن تر بوده و بیشتر به هم نزدیک بوده است.

چارک اول (Q_1): مقداری از داده‌های یک متغیر که 25 درصد داده‌ها قبل از آن و 75 درصد داده‌ها بعد از آن باشند، برای مثال در یک کلاس چارک اول قد دانشجویان 156cm است به این معنی که 25 درصد افراد زیر 156cm یا برابر آن و 75 درصد دانشجویان بالای 156cm بلندی قد دارند.

چارک سوم (Q_3): مقداری از داده‌های یک متغیر که 75 درصد داده‌ها قبل از آن و 25 درصد داده‌ها بعد از آن هستند برای مثال چارک سوم قد دانشجویان یک کلاس 172cm است به این معنی که 75% درصد دانشجویان قدی کوتاهتر یا برابر 172cm و 25 درصد آنها قدی بلندتر از 172cm دارند.

دامنه (Range): اختلاف بزرگترین (Max) داده‌های یک متغیر از کوچکترین (Min) داده‌های یک متغیر که هرچه قدر کوچکتر باشد نشان دهندهٔ همگن تر بودن داده‌های یک متغیر است.

صدک (Percentile): صدک i ام جایی است که i درصد داده‌ها قبل از آن و 100-i درصد داده‌ها بعد از آن هستند برای مثال صدک 23ام یک متغیر سطح تحصیلات برابر دیپلم است. این به آن معنی است که 23 درصد افراد دارای تحصیلات دیپلم و زیر دیپلم و 77 درصد افراد دارای تحصیلاتی بالاتر از دیپلم هستند.

همانطور که در بالا گفته شد آمار توصیفی با استفاده از جداول آماری، نمودارهایی نظیر نمودار میله ای و پارامترهایی که تعدادی از آنها در بالا شرح داده شد یک نمای کلی از وضعیت هر یک از متغیرها به کاربران ارائه می‌دهد، اما وظیفهٔ آمار استنباطی در بدست آوردن رابطهٔ میان متغیرهاست که از طریق ابزارها و تکنیک‌های مختلف آماری بسته به نوع (پیوسته یا گسسته بودن) متغیرهای مستقل و وابسته مورد استفاده قرار می‌گیرند.

خلاصه مباحث مطرح شده:

- جامعه در علم آمار عبارت است از مجموعه ای از عناصر که حداقل در یک خصوصیت با یکدیگر مشترک باشند.

- شناخت جامعهٔ آماری که فلسفه وجودی علم آمار است در حقیقت شناخت عناصر جامعهٔ مورد بررسی و خصوصیات آنهاست.

- منظور از شناخت متغیرها(خصوصیات) در علم آمار شامل شناخت یک بعدی متغیرها یعنی رفتار هر متغیر به تنهایی و همچنین شناخت دو یا چند بعدی متغیرها یا شناخت رابطه متغیرها با یکدیگر می باشد.

- آمار، ابزاری برای بدست آوردن رابطهٔ میان متغیرهاست.

- علم آمار عبارت است بدست آوردن تأثیر متغیرهای مستقل بر روی متغیرهای وابسته.

- متغیرهای پیوسته متغیرهایی هستند که به بیان ساده می‌توانند بی نهایت عضو داشته باشند،

- متغیرهای گسسته برخلاف متغیرهای پیوسته تعداد سطوح محدودی دارند

- در بسیاری از موارد یک محقق به تمامی عناصر یک جامعهٔ مورد بررسی دسترسی ندارد در این گونه موارد نیازمند استفاده از تکنیکهای نمونه گیری هستیم.

- یک نمونهٔ آماری دارای تمام خصوصیات یک جامعهٔ آماری با دقت $(1-\alpha)\%$ است و تنها تعداد عناصر آن کمتر است.

- آمار توصیفی بیان کننده وضعیت هر متغیر به صورت مجزا در جامعهٔ مورد بررسی یا شناخت یک بعدی متغیرها در قالب جداول، نمودارها و پارامترهاست.

- وظیفه آمار استنباطی بدست آوردن رابطهٔ میان متغیرهاست

تمرین:

1- **مفهوم جامعه آماری را با ذکر حداقل 3 مثال بیان نمائید.**

2- **انواع متغیرهای آماری را نام ببرید و مشخص نمائید که متغیرهای ذیل جزء کدام دسته قرار می گیرند.**

طول عمر انسان، تعداد پرسنل یک شرکت، وزن انسان، وضعیت تاهل، قند خون، انواع دارو، رتبه شغلی، شماره پرسنلی، گروه خونی، تعداد افراد خانواده

3- **سه مورد از مواردی را که برای انجام مطالعه مجبور به استفاده از روش نمونه گیری هستیم نام ببرید.**

4- **تفاوت آمار توصیفی و آمار استنباطی را شرح دهید.**

5- **تفاوت رگرسیون و جدول توافقی را در آمار استنباطی بیان نمائید.**

6- **انواع پارامترهای مرکزی و پارامترهای پراکندگی را نام ببرید.**

فصل دوم

معرفی نرم افزار SPSS.16 و قابلیت های آن

اهداف آموزشی فصل دوم

فراگیر پس از مطالعه این فصل باید بتواند:

- انواع قابلیتهای نرم افزار spss.16 را بازگو نماید.
- انواع پنجره های موجود در نرم افزار spss.16 را بشناسد و هر یک را معرفی کند.

- پنجره های اصلی در نرم افزار را نام برده و وظایف هر یک را شرح دهد.
- پنجره های فرعی در نرم افزار را نام برده و وظایف هر یک را توضیح دهد.
- منوهای پنجره های نرم افزار را نام برده و کار هر یک را بیان کند.
- زیر منوهای منو Help را نام برده و وظایف هر یک را توضیح دهد.

2-1-مقدمه:

نرم افزار SPSS یک بسته نرم‌افزار آماری است که در دهه 1960 جهت کاربرد در دانشگاه استانفورد شیکاگو طراحی شد و از دهه 1970 در پژوهش‌های بسیار زیادی در زمینه‌های موضوعی متعدد مورد استفاده همگانی قرار گرفته است. طراحان این نرم‌افزار آن را بر پایه استفاده در عملیات آماری طراحی نمودند تا داده‌های خام را به اطلاعات پردازش شده و مناسب برای تصمیم‌گیری تبدیل نمایند. سادگی کار با این نرم افزار ، آن را به یکی از کاربردی ترین نرم افزارهای دنیای تحلیل اطلاعات تبدیل کرده است. قابلیت پردازش اطلاعات در محیط این نرم افزار به هیچ عنوان قابل مقایسه با نرم افزارهای متداول در محیط Windows نمی باشد برای مقایسه این نرم افزار قادر است تنها در یک صفحه از تعداد صفحات نامحدود خود 2000000000 (دو میلیارد) سطر و 10000 (ده هزار) ستون اطلاعات را در خود جای دهد که در مقایسه با نرم افزارهایی نظیر Excel و Access قابل قیاس نمی باشد. این نرم افزار علاوه بر قدرت پردازش آماری روی حجم وسیعی از داده ها به عنوان یک بانک اطلاعاتی نیز می تواند مورد استفاده قرار گیرد.

با توجه به قابلیتهای این نرم افزار در این فصل به ارائه توضیحاتی در خصوص پنجره های اصلی و پنجره های فرعی نرم افزار spss و منوهای آن ارائه خواهد شد

2-2- انواع پنجره‌های موجود در SPSS :

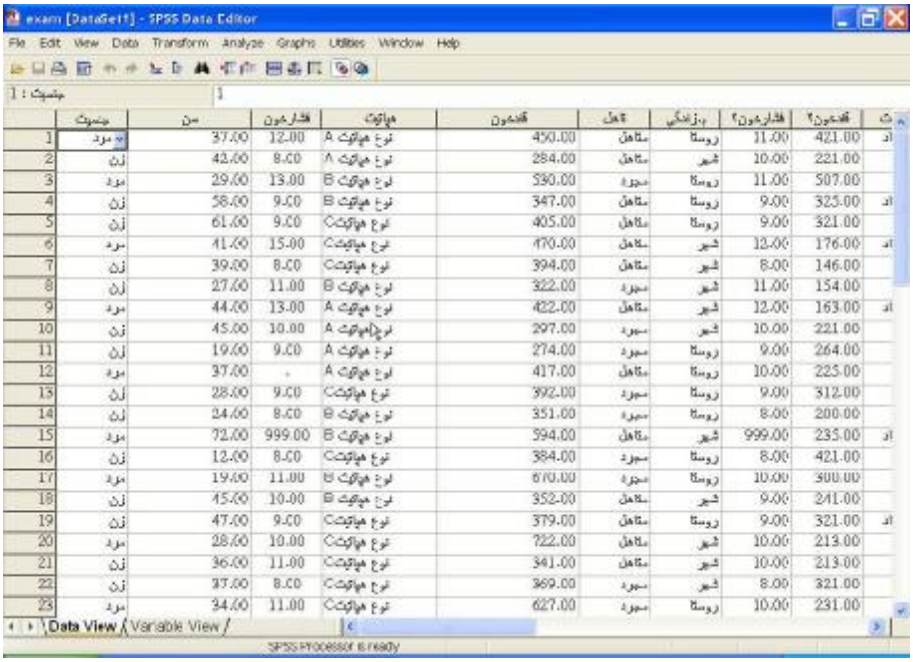
این نرم افزار دارای چهار پنجرهٔ اصلی و چندین پنجرهٔ فرعی است. پنجره‌های فرعی پنجره‌هایی هستند که در مواقع

ضروری برای ویرایش اطلاعات پنجره‌های اصلی از آنها استفاده می‌گردد و به تنهایی ماهیت یک پنجرهٔ مستقل را ندارند.

2-2-1-پنجره های اصلی در spss:

نرم افزار spss دارای چهار پنجره اصلی می باشد که هر یک در ذیل تشریح شده اند.

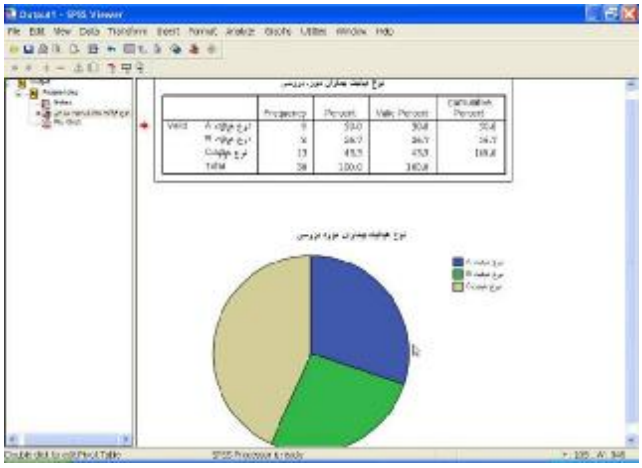
پنجره Dataview : این پنجره محیطی برای ورود داده‌ها به نرم افزار و انجام ویرایش بر روی داده‌ها است. این پنجره مهمترین صفحه ایی است که کاربر با آن در محیط نرم افزار آشنا می شود. این پنجره حاوی تعداد زیادی منو و ابزارهای جانبی است که امکان ورود داده ها، ویرایش داده ها و انجام فرامین آماری مختلف را روی داد ها میسر می سازد. شکل این پنجره در ذیل مشاهده می شود:



کاربر در اولین ورود به نرم افزار این پنجره را مشاهده می‌نماید. بسیاری از اوقات داده های اولیه برای انجام تجزیه و تحلیل ها در محیط نرم افزار مناسب نسیت و کاربر داده های خام را وارد نرم افزار نموده، در این محیط از طریق امکانات منوهای مختلف این پنجره داده ها را ویرایش نموده و آماده ورود به جعبه های گفتگو برای اجرای فرمانهای آماری می نماید. این پنجره دارای یک پنجره فرعی با نام Vaviable View است که دکمه ورود به آن در پائین صفحه با همین نام وجود دارد. این پنجره همانطور که گفته شد بدون وجود پنجره اصلی Dataview ماهیت مستقل ندارد و زمانی می

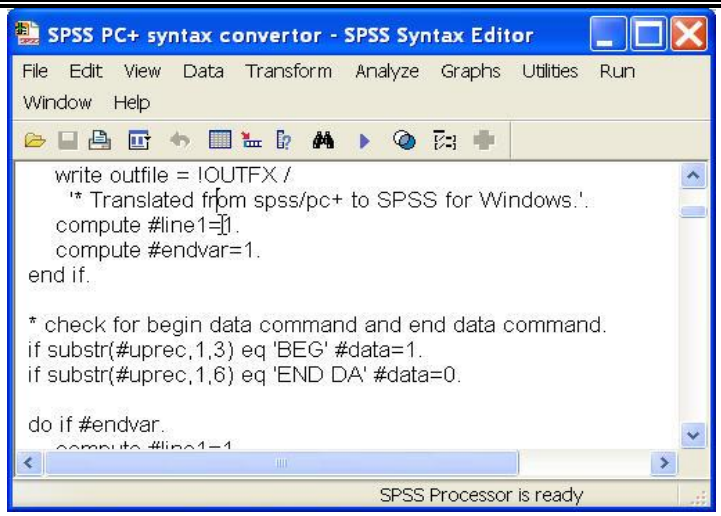
توان از آن بهره گرفت که در پنجره Dataview داده وارد شده باشد. در قسمت پنجره های فرعی به محتویات و وظایف آن اشاره خواهد شد.

پنجرهٔ Viewer : این پنجره که از آن به عنوان پنجرهٔ خروجی‌ها یاد می‌کنند محیطی است که کـاربر در آن نتـایج تجزیه و تحلیل‌های نرم افزار را مشاهده می‌نماید. نمای این پنجره در شکل زیر مشاهده می شود:



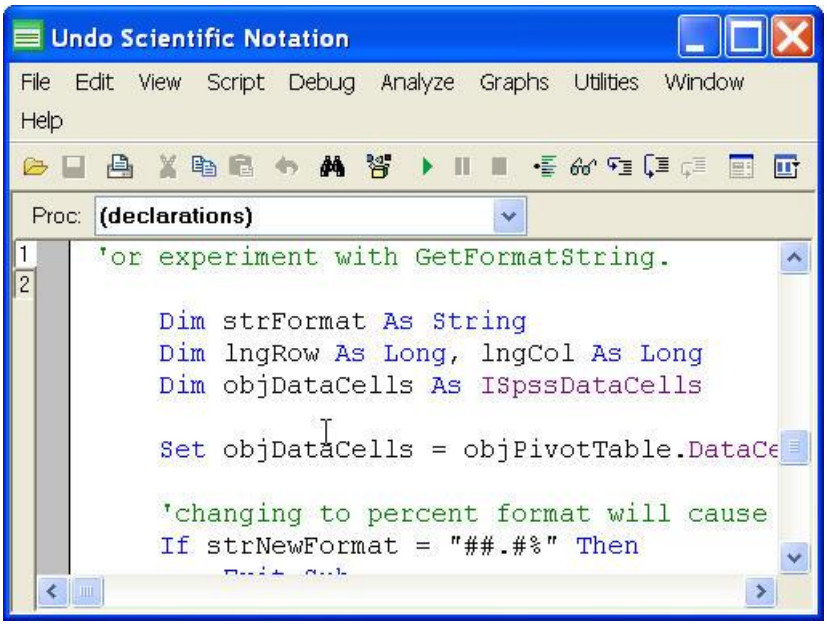
کاربر در این پنجره علاوه بر مشاهده خروجی ها و نتایج تجزیه و تحلیـل هـای آمـاری در قالـب جـداول، نمودارهـا ، خطاها و توضیحات مربوط به اجرای فرامین، می تواند با استفاده از امکانات این پنجره نسبت به ویرایش خروجی ها اقدام نماید. این پنجره دارای یک پنجره فرعی به نام Chart Editor نیز می باشد که محتویات و وظایف آن در قسمت پنجره های فرعی بیان می شود.

پنجرهٔ Syntax : این پنجره یک محیط متنی است که امکان برنامه نویسی را با زبان نرم افزار فراهم می‌آورد، کـاربر می تواند با توجه به نیاز خود آشنا شدن با زبان برنامه نویسی در محیط نرم افزار از امکانات برنامه نویسی در این محـیط و اجرای آنها بهره گیرد. نمای این پنجره در شکل زیر مشاهده می شود:



در حال حاضر با توجه به توانایی‌های زیاد نرم افزار در استفاده از منوهای تحت ویندوز، از کاربرد پنجرهٔ Syntax بـه منظور نوشتن برنامه به صورت متنی و اجرای آن برای گرفتن نتایج کاسته شده است.

پنجرهٔ Script : این امکان را به کاربر می‌دهد که بسیاری از خصوصیات پنجره‌ها و تنظیمات صفحات را مانند رنگ، اندازه، فونت و خصوصیات دیگر را از طریق نوشتن و یا تغییر Scriptها در SPSS به دلخواه انجام دهد. نمای این پنجره در شکل زیر مشاهده می شود:



کاربران با توجه به سلیقه و نیاز خود می توانند با تغییر دادن هر یک از Scriptها که در پوشه محل نسب نرم افزار با

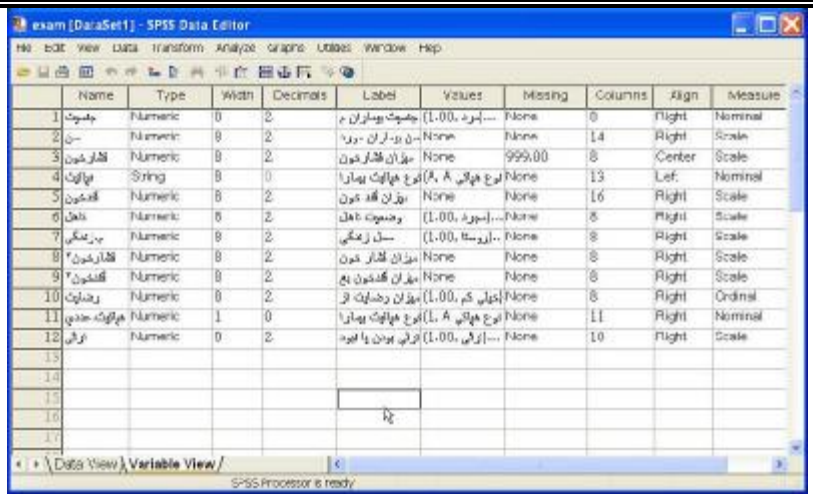
همین نام موجود است نسبت به اعمال تغییرات مورد نظر خود در شکل ظاهری هر یک از پنجره ها و خروجی های نـرم افزار اقدام کنند.

نرم افزار SPSS این توانایی را دارد که در پنجرهٔ Dataview کاربر بتواند ده هزار متغیر را با دو میلیـارد عنصـر وارد نماید و البته این تعداد داده را تنها در یکی از صفحات بدون محدودیت Dataview می توانید داشـته باشـید. بـه طـور مثال می‌توان اطلاعات ده هزار متغیر یک جامعهٔ آماری را برای دو میلیارد عضو آن جامعه را در یک صـفحه Dataview وارد نرم افزار SPSS نمود و مورد تجزیه و تحلیل قرار داد. همانطور که ملاحظه می شود نرم افزار spss از نظرگنجایش داده ها قابل مقایسه با بسیاری از نرم افزارهای دارای بانک اطلاعاتی نمی باشد

از دیگر توانایی‌های نرم افزار SPSS قابل تنظیم بودن کلیه پنجره‌ها و خروجی‌های آن به 13 زبان رایـج و همچنـین استفاده همزمان از چندین پنجرهٔ یکسان است. برای مثال کاربر می‌تواند همزمان چند صفحهٔ مجزا Dataview کـه هـر کدام حاوی اطلاعات متفاوتی هستند را باز و بر روی داده‌های آنها تجزیه و تحلیل انجام دهد.

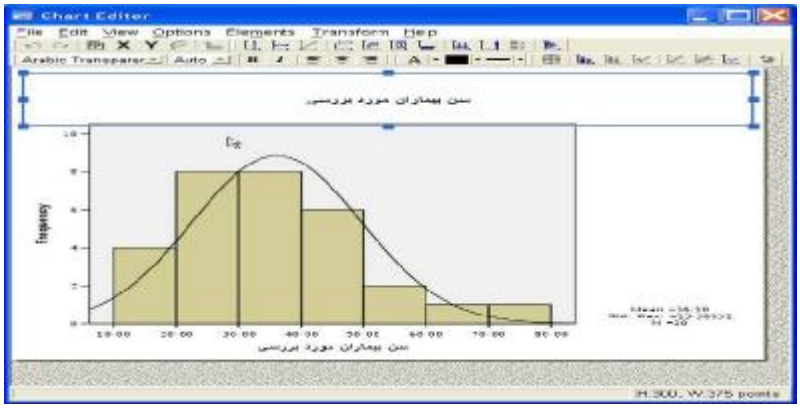
2-2-2-پنجره‌های فرعی در نرم افزار spss :

پنجرهٔ فرعی Vaviable View : این پنجرهٔ فرعی یک پنجره به پیوست پنجرهٔ Data View می‌باشد و وظیفهٔ آن تعریف و کنترل متغیرهایی است که در پنجرهٔ نمایشگر داده‌ها وارد شده اند. نمای پنجره در زیر مشاهده می شود:



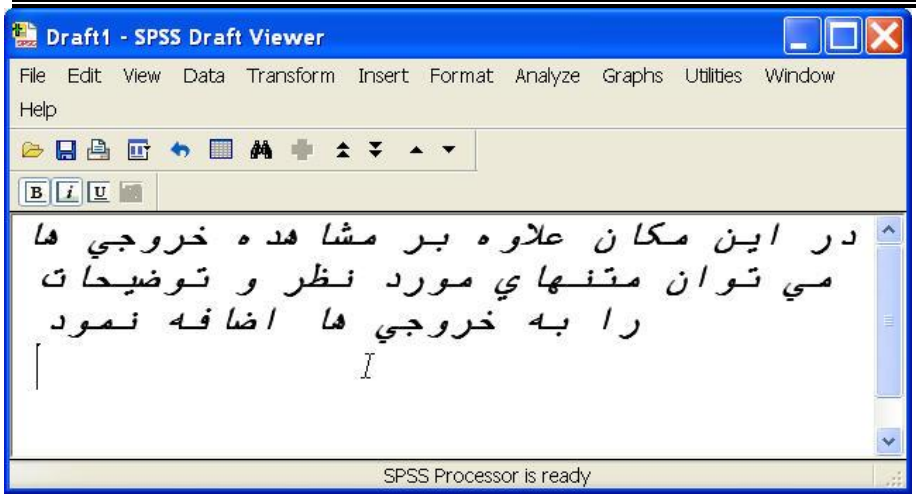
از جمله خصوصیات قابل کنترل و تعریف متغیرها می‌توان به نام متغیر، نوع متغیر، توضیحات در مورد متغیر، تعریف سطح یک متغیر، گسسته یا پیوسته بودن یک متغیر و ... را نام برد که در ادامه به طور کامل قابلیت‌های این پنجره توضیح داده خواهد شد.

پنجرهٔ فرعی Chart Editor : وظیفه این پنجرهٔ فرعی از پنجرهٔ Viewer که خروجی‌ها در آن نمایش داده می‌شوند، ویرایش نمودارهای گرفته شده در خروجی Viewer است. نمای این پنجره در زیر مشاهده می شود:



کاربر می‌تواند با کلیک بر روی هر شکل در خروجی پنجرهٔ Viewer یا Output ، پنجرهٔ Chart Editor را باز نماید و رنگ، فونت، برچسب و دیگر خصوصیات نمودارها را به دلخواه خود تغییر دهد.

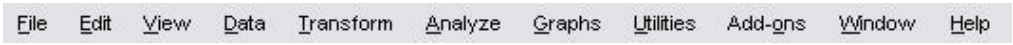
پنجرهٔ فرعی Draft Viwer : این پنجرهٔ فرعی از پنجرهٔ Viewer امکان تایپ در محیط خروجی را علاوه بر مشاهدهٔ نتایج تجزیه و تحلیل‌ها برای کاربر مهیا می‌سازد . نمای این پنجره در شکل زیر مشاهده می شود:



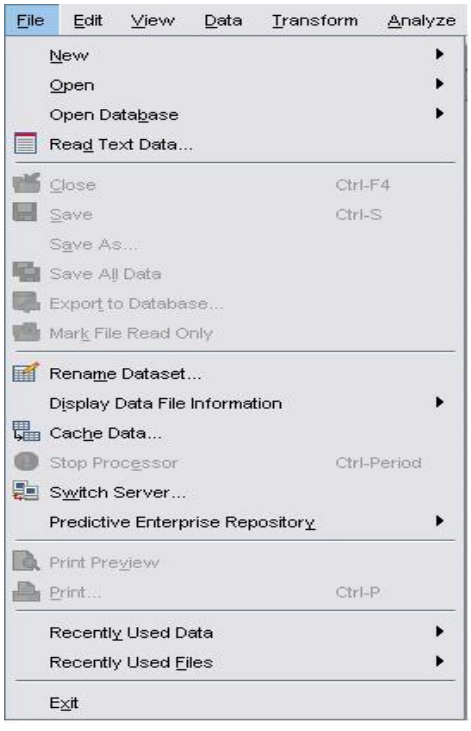
در حقیقت کاربر می‌تواند در این پنجرهٔ فرعی توضیحات متنی خود را با هر فونت و زبان دلخواه به خروجی‌های نرم افزار بیفزاید.

3-2-منوها در پنجره‌های SPSS :

بسیاری از کارهایی که در پنجره‌های مختلف SPSS بر روی محتویات پنجره‌ها انجام می‌گیرد با انتخاب منو آغاز می‌شود. هر پنجره در SPSS نوار منوی مخصوص به خود را دارد و هر منوی آن نیز شامل آیتم‌های مناسب با آن منو هستند که وظیفهٔ خاصی را برعهده دارند.



برای مثال در پنجرهٔ Dataview در نوار منو ده منو مانند شکل بالا وجود دارد که هر یک با وجود زیر آیتم‌های خود وظیفهٔ خاصی را برعهده دارند.



1ـ **منو File (فایل یا پرونده)**: در این منو امکان ایجاد، فراخوان، بستن و ذخیره کردن انواع فایل‌های محتوی داده‌های آماری و پنجره‌های دیگر، چاپ اطلاعات، تنظیم چاپگر، مشاهدهٔ اطلاعات فایلها، نمایش اطلاعات داده‌ها و خروج از نرم افزار موجود است. کاربر با استفاده از امکانات موجود در این منو امکان

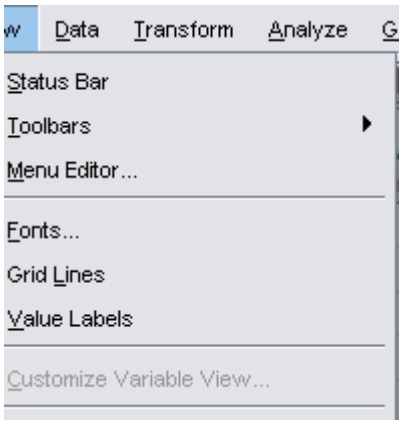
مدیریت فایلهای اطلاعاتی مورد نیاز در نرم افزار را می یابد.

2ـ منو Edit (ویرایش):

در این منو امکان انواع ویرایش بر روی داده‌های موجود در Dataview نظیر برش، کپی، حذف، انتخاب بخشی از فایل، جستجو، جایگذاری و تنظیم کلیات ویرایشگر SPSS موجود است. این منو در هر یک از پنجره های دیگر نرم افزار نیز وظیفه ویرایش اطلاعات موجود در آن نرم افزار را با وجود امکانات موجود و زیر منو ها در اختیار کاربر قرار می دهد. برای نمونه در پنجره خروجی ها وظیفه این منو و زیر منوهای آن ویرایش اطلاعات خروجی ها می باشد.

3ـ منو View (نما):

این منو امکان تغییر شکل ظاهری اطلاعات موجود در پنجره نظیر فونت، سایز، تنظیم



جعبهٔ ابزار و ... را در اختیار کاربر قرار می‌دهد. در سایر پنجره های موجود در نرم افزار هم این منو با توجه به ساختار پنجره دارای ابزارهایی به منظور ویرایش اطلاعات موجود در آن پنجره است. برای نمونه در پنجره خروجی ها تغییر اندازه، فونت خروجی ها، نمای جداول و ... از طریق امکانات این منو انجام می گیرد.

4ـ منو **Data** (داده‌ها):

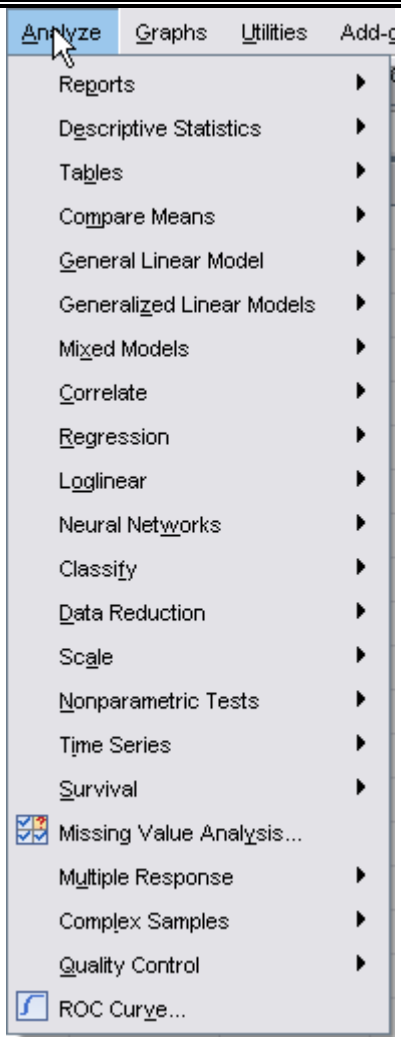
در این منو امکان تغییر متغیرها، و آماده کردن داده های خام وارد شده در محیط ورودی نرم افزار به متغیرهای مورد استفاده در تجزیه و تحلیل های نرم افزار بر عهده دارد. از جمله امکانات موجود در این منو می توان به اضافه کردن متغیرهای جدید به فایل داده‌ها، مرتب کردن داده‌ها، ادغام داده‌ها، انتقال داده‌ها، تعریف خصوصیات متغیرها، گروه بندی متغیرها بر اساس خصوصیات مورد نظر، خوشه بندی و دسته بندی خروجی ها بر اساس سطوح یک متغیر گسسته، انتخاب داده های مورد نظر با شرایط خاص برای شرکت در تجزیه و تحلیل ها و وزن دهی متغیرهای بر اساس یک تغییر مشخص را نام برد که در ادامه به شرح هر یک از زیر منوها و نحوه انجام آنها اشاره خواهد شد.

5ـ منو **Transform** (تبدیل):

در این منو امکاناتی در اختیار کاربر قرار می گیرد که با استفاده از آنها کاربر می تواند مانند منو Data داده ها را برای شرکت در تجزیه و تحلیل ها آماده سازد. از جمله اقداماتی که روی داده ها صورت می گیرد می توان به انجام عملیات ریاضی روی متغیرها، تغییر کد متغیرهای گسسته، تغییر متغیرهای حرفی(String) به متغیرهای عددی(Numeric) ، شمارش داده های خاص درون یک متغیر، تولید داده های سری زمانی، تولید متغیرهای زمان، جایگزین نمودن داده های گمشده و تولید داده های تصادفی را نام برد. که در ادامه فصل به توضیح هر یک از زیر منو ها پرداخته شده است.

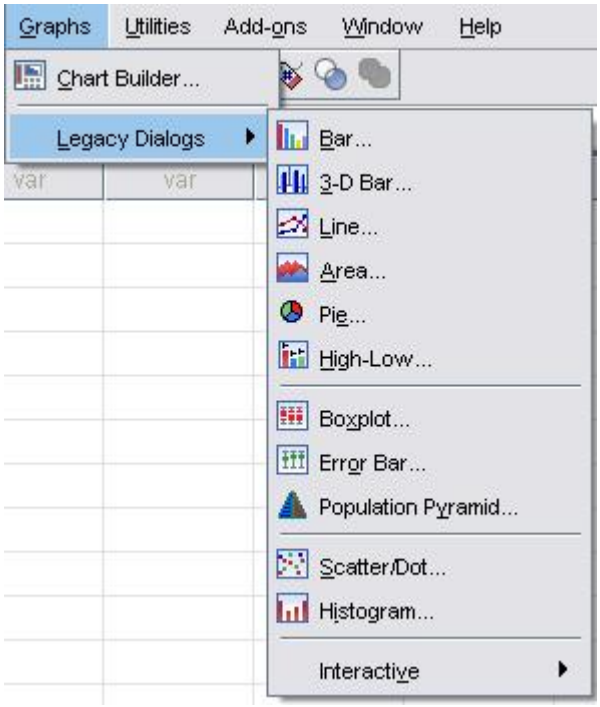
6ـ منو **Analysis** (تحلیل):

کلیه روشهای پردازش آماری از طریق SPSS در این منو متمرکز شده اند، تکنیک های مختلف آمار توصیفی و آمار استنباطی برای تخلیص داده ها و بدست آوردن رابطه میان متغیرها از طریق زیر منوهای مختلف این منو در اختیار کاربر قرار می گیرند. سایر منوهای معرفی شده در واقع ابزارهای مقدماتی برای ورود به منو تحلیل هستند و این منو در حقیقت منو اصلی نرم افزار به حساب می‌آید.



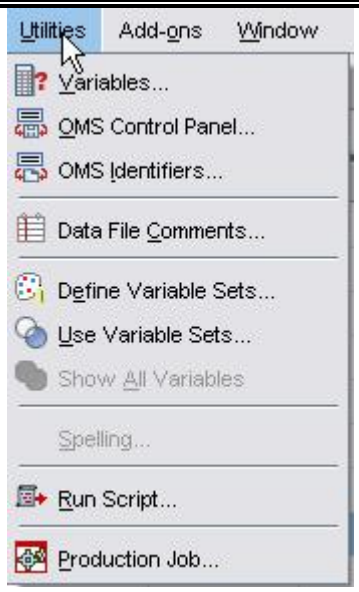
7ـ منو Graphs (نمودارها):

در این منو می‌توان با توجه به هر یک از زیرآیتم‌های آن انواع شکل‌های آماری عمومی و تخصصی را با وضوح بالا رسم نمود. در این منو می توان با استفاده از راهنمای موجود در این منو از نحوه کشیدن هر یک از شکلهای آماری قابل ترسیم از طریق زیر منوهای این منو اطلاعات کسب کرد.



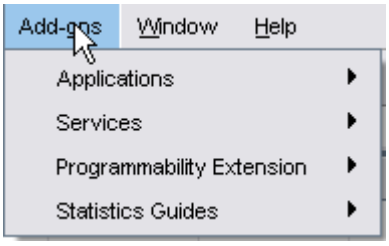
8ـ منو **Utilities** (امکانات):

این منو یکسری توابع فرعی، شامل اطلاعاتی دربارهٔ فایل جاری، ساخت مجموعه ای از متغیرها، اضافه کردن و ترکیب زیر منوها در هر یک از منوهای پنجره های اصلی و فرعی و اجرای فرمانهای Script را در هر یک از پنجره ها در خـود جای داده است.



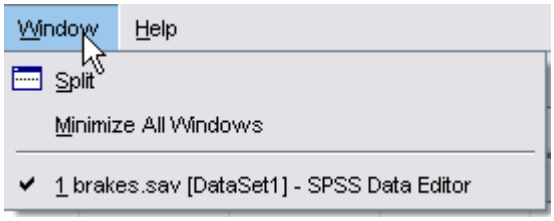
9- منو add-ons(ارتقاء):

این منو راهنمایی لازم در مورد محصولات جانبی شرکت سازنده نرم افزار spss را که می تواند مورد استفاده کاربران قرار گیرد به همراه خلاصه ایی از کارایی هر یک را در اختیار کاربر قرار می دهد.



10ـ منو Windows (پنجره‌ها):

در این منو می‌توان پنجره‌های خاصی را که در نرم افزار تعریف شده اند فعال نمـود و نحـوه آرایـش آنهـا را در کنـار یکدیگر تغییر داد



11ـ منو **Help (کمک رسانی)**:

این منو راهنمایی‌هایی در مورد نحوه استفاده از نرم افزار SPSS و بخش‌های مختلف آن، همچنین امکان جستجوی یک عنوان خاص، نمایش ترکیب فرمان‌های SPSS و خلاصه ای از تعاریف اصطلاحات و فرمان‌ها را فراهم می‌کند.

کاربر در صورتی که به زبان نرم افزار Spss مسلط باشد خواهد توانست با مطالعه راهنماییهای موجود در زیر منوهای موجود این منو بسیاری از مشکلات خود در نحوه یاد گیری این نرم افزار را مرتفع نماید.

در بالا وظیفه ده منو پنجره Data View که در حقیقت پنجرهٔ اصلی نرم افزار است توضیح داده شد که در فصول بعد به طور مفصل زیرمنوهای هر یک از آنها توضیح داده شده است. اما قبل از ورود به نحوه استفاده از نرم افزار SPSS آشنایی با نحوه کمک رسانی نرم افزار از طریق منو Help و زیر منوهای آن می‌تواند سودمند باشد.

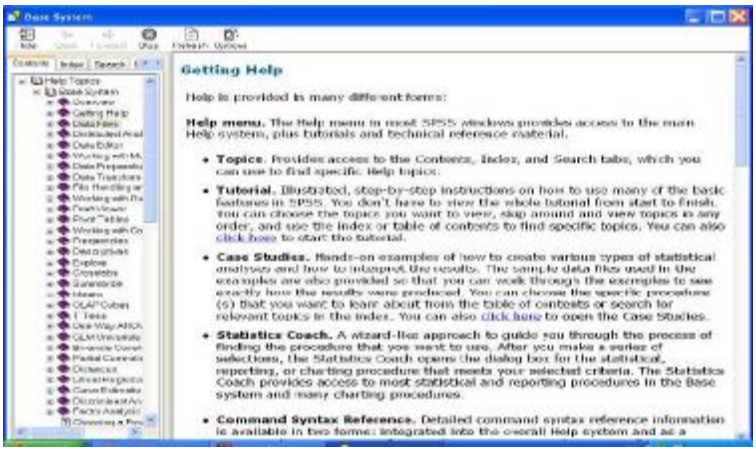
همانطور که در شکل بالا ملاحظه می‌شود منو Help دارای 11 زیر منو می‌باشد که هر یک به شکلی خاص کـاربر را برای رسیدن به مقصود راهنمایی می‌کند. در ذیل تعدادی از زیر منوهای آن توضیح داده شده است.

1- **زیر منو Topics (موضوعات)** :

شرکت SPSS ارائه کننده این نرم افزار، یک کتاب شامل راهنمای محتویات نرم افزار به پیوست نرم افزار، هـر سـاله وارد بازار می‌نماید در حقیقت با زیر منو Topics کاربر می‌تواند اطلاعاتی در مورد محتویات کتـاب مـورد نظـر و شـیوه استفاده از ابزارهای نرم افزار SPSS کسب نماید.

هنگامی که زیر منو Topics را انتخاب می‌نمائید پنجره ای با نام Base System باز می‌شود بـا انتخـاب Contents لیستی از تمامی سرفصل‌های کتاب عنوان شده در اختیار کاربر قرار می‌گیرد که می‌تواند با توجه به هر یک از سرفصـلها توضیح عبارت مورد نظر را در آن مشاهده نماید. برای مثال با انتخاب سرفصل Data File می‌توانـد اطلاعـاتی در مـورد نمایش داده‌ها، وارد کردن خصوصیات متغیرها، وارد کردن داده‌ها، ویرایش داده‌ها و ... را کسب نماید. کاربر با انتخاب هر یک از عناوین در قسمت سمت چپ پنجره و انتخاب زیر گزینه های مورد نظر می تواند به صورت مرحلـه بـه مرحلـه از طریق راهنمایی های متنی که پنجره در قسمت سمت راست مانند شکل زیر در اختیار قرار می دهد از راهنمـایی هـای نرم افزار استفاده نماید.

قابل ذکر است این راهنمایی به نحوه استفاده از نرم افزار و و تعاریف اولیه مربوط به هر یـک از ابزارهـای آمـاری بـه صورت جداگانه می پردازد.



در صورتی که کاربر گزینهٔ Index را انتخاب نماید کلیه اطلاعات موجود در آیتم Topics به صورت مرتب شده در اختیار کاربر قرار می‌گیرد و کاربر با سرعت بیشتر به جستجوی مطلب مورد نظر می‌پردازد. برای مثال تنها با زدن کلمهٔ d تمامی عباراتی که با d شروع ظاهر می گردد و کاربر می‌تواند Data Editor را انتخاب و با زدن Disply در پایین صفحه از تمامی زیر فصل های Data Editor در صفحهٔ ظاهر شده Topics Found استفاده نماید. برای مثال برای راهنمایی گرفتن از نحوه چاپ داده‌ها می‌تواند Data Editor Printing را انتخاب و Disply را کلیک نماید که راهنمای نحوه چاپ داده‌ها در SPSS به صورت متنی ظاهر می‌گردد.

در صورت انتخاب گزینهٔ Search کاربر می‌تواند با تایپ کلمه مورد نظر آن را در Topics نرم افزار جستجو نماید و در صورت وجود از راهنمای متنی آن بهره گیرد.

2-زیر منو Tutorial (راهنمای مرحله به مرحله):

استفاده از این آیتم برای کاربران به منظور آشنایی با نحوه انجام فعالیت در نرم افزار می‌تواند بسیار سودمند باشد. این آیتم کاربر را به صورت مرحله به مرحله برای رسیدن به هدف یاری می‌سازد و در هر مرحله کاربر با انتخاب گزینه های مختلف موجود در صفحات مورد نظر با توجه به هدف خود به مرحله بعد می رود و از راهنمایی های این نرم افزار در مورد خاص با استفاده از متن و مثال و گرافیک بهره می برد.

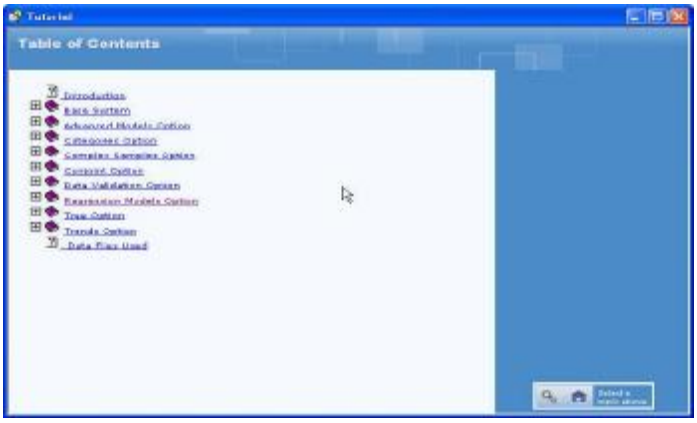
برای مثال برای آشنایی با نحوه خواندن داده‌ها در پنجرهٔ Tutorial آیکون Reading Data را انتخاب و گزینهٔ

Next در پایین سمت راست صفحه را انتخاب می‌کنیم، راهنمای نرم افزار از طریق راهنمایی متنی و تصویری نحوه انجام فرایند خواندن داده‌ها را به کاربر آموزش می‌دهد. برای رفتن به Slide بعد باز هم دکمهٔ Next را می‌زنیم تا اطلاعات تکمیلی بیشتری در اختیار کاربر قرار گیرد.

3-زیر منو Case Studies :

این آیتم امکانی نظیر Tutorial در پنجره ای با همین نام در اختیار کاربر قرار می‌دهد با این تفاوت که در آیتم Tutorial نحوه کار در نرم افزار از قبیل ورود داده‌ها، آشنایی با پنجره‌ها، خروجی گرفتن، باز کردن فایل‌ها و ... آموزش داده می‌شود اما در Case Studies نحوه انجام آمار استنباطی یعنی امکانات نرم افزار در تجزیه و تحلیل متغیرها به صورت مرحله به مرحله به کاربر آموزش داده می‌شود.

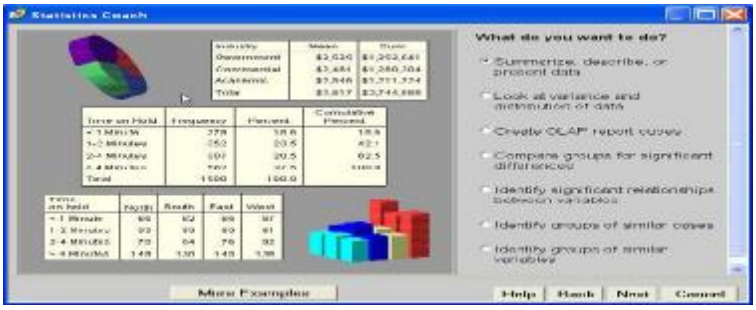
برای مثال با انتخاب Regression Models Option کاربر می‌تواند اطلاعاتی در مورد انواع مدلهای رگرسیونی و نحوه انجام آنها در نرم افزار SPSS با استفاده از راهنمای متنی و تصویری را مشاهده کند و با انتخاب گزینه Next می‌تواند مانند قبل به متن تکمیلی یا Slide تکمیلی بعد برود و از امکانات آن صفحه و راهنمایی های آن استفاده نماید و در نهایت نرم افزار پنجره مورد نظر را که می تواند نیاز کاربر را در انجام آمار استنباطی بر طرف سازد، باز می نماید.



استفاده از این روش‌های کمک رسانی برای کسانی که به یکی از زبان‌های تعریف شده نرم افزار آشنایی دارند بسیار سودمند است.

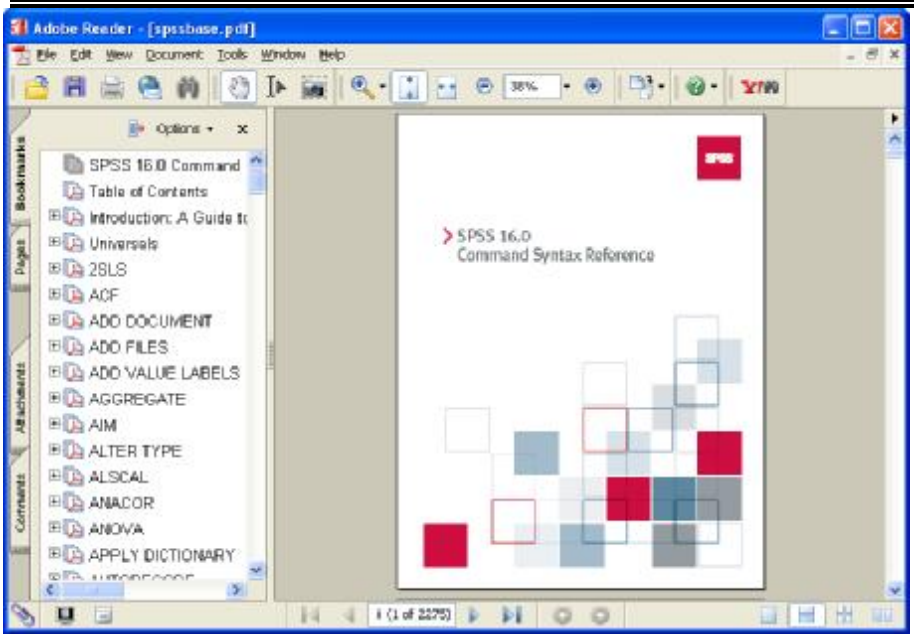
4-زیر منو Statistical Coach :

با انتخاب این آیتم پنجره ای با همین نام باز می‌شود که مانند دو آیتم قبل امکان استفاده از راهنمایی مرحله به مرحله نرم افزار را برای آمار توصیفی یعنی کشیدن شکل‌های آماری، جداول و محاسبه پارامترهای آماری در اختیار کاربر قرار می‌دهد. کاربر با توجه به نیاز خود می‌تواند یکی از گزینه‌های What Do You Want To Do? را انتخاب نماید و با زدن Next و More Example از توضیحات تصویری نرم افزار به صورت مرحله به مرحله بهره گیرد.



5-زیر منو Command Syntax Refrence :

کاربر با انتخاب این آیتم می‌تواند از نحوه برنامه نویسی و لیست دستورات نرم افزار SPSS که در یک فایل PDF با نام SPSSbase.pdf ظاهر می‌شود برای کار در پنجره Syntax استفاده نماید.



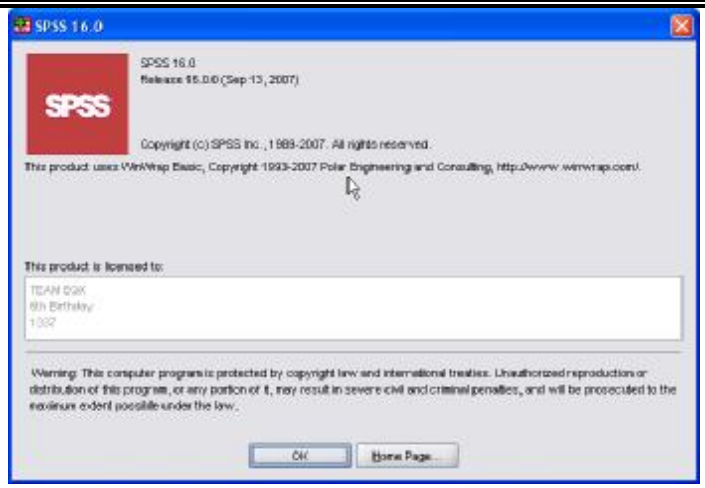
کاربر می تواند با مطالعه محتویات این کتاب با نحوه برنامه نویسی برای انجام دستورات آمار استنباطی و آمـار توصـیفی بهره گیرد.

6-زیر منو **SPSS Developer center** :

در صورتی که کامپیوتر به شبکه اینترنت متصل باشد با انتخاب این زیر منو،صفحه SPSS.Com را که سایت شرکت سازنده این نرم افزار است به منظور آگاهی از اطلاعات جدید شرکت به منظور به روز کردن محصـول خریـداری شـده و ارائه نسخه بهینه شده در اختیار کاربر قرار می‌دهد.

7-زیر منو **About** :

این آیتم با باز کردن پنجره ای به نام SPSS16.0 For Windows اطلاعاتی در مورد نرم افـزار SPSS و همچنـین اطلاعات سخت افزار سیستمی که نرم افزار بر روی آن نصب شده است در اختیار کاربر قرار می‌دهد.



8-زیر منو **Algorithms** :

این زیر منو امکان آگاهی کاربر از فرمولها و الگوریتمهای بکار گرفته شده در نرم افزار برای هر یک از موارد مورد نظر کاربر را در اختیار قرار می دهد.

9- زیر منو **SPSS Home Page** :

در صورتی که کامپیوتر به شبکه اینترنت متصل باشد با انتخاب این زیر منو،صفحه SPSS.Com را که سایت شرکت سازنده این نرم افزار است به منظور آگاهی از اطلاعات جدید شرکت در اختیار کاربر قرار می‌دهد.

10- زیر منو **check for Updars** :

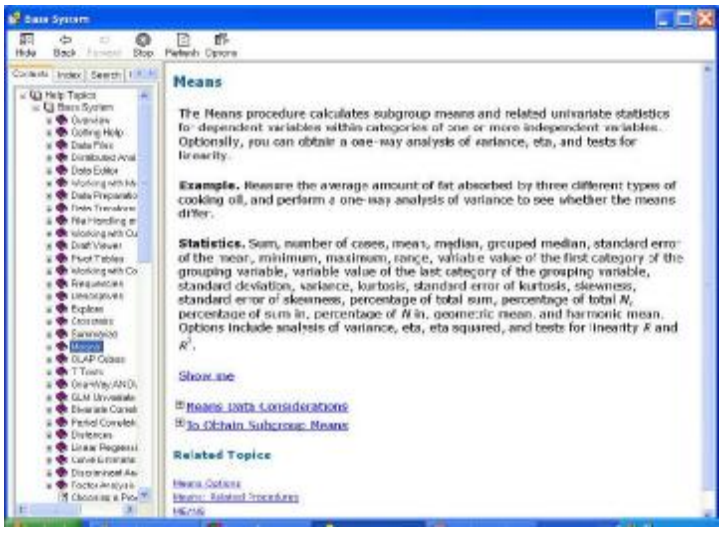
در صورتی که کامپیوتر به شبکه اینترنت متصل باشد با انتخاب این زیر منو،صفحه SPSS.Com را که سایت شرکت سازنده این نرم افزار است به منظور آگاهی کاربر از امکانات جدید به روز رسانی که به صورت رایگان روی سایت شـرکت قرار گرفته است در اختیار کاربر قرار می دهد.

11- زیر منو **Product Registeration** :

با انتخاب این زیر منو اطلاعاتی در مورد نحوه ثبت نام کاربر به عنوان استفاده کننده از امکانات نـرم افـزار SPSS و امکانات جانبی که شرکت در اختیار کاربران ثبت نام شده قرار می‌دهد به کاربران ارائه می‌نماید.

علاوه بر آیتم‌های بالا نرم افزار SPSS از طریق دکمه‌های Help در جعبه‌های محاوره نیز اطلاعـاتی در مـورد همـان جعبهٔ محاوره در اختیار کاربر قرار می‌دهد.

برای مثال در منو Analysis و آیتم Compare Means و زیر آیتم Means، دکمهٔ Help نحوه محاسبهٔ میانگین‌ها از طریق این جعبهٔ محاوره را بیان می‌نماید.



همانطور که ملاحظه می شود با انتخاب این گزینه نحوه راهنمایی در نرم افزار صفحه مربوطه مانند صـفحات موجـود در زیر منو Topics در اختیار کاربر قرار می گیرد که می تواند با استفاده از اطلاعات متنی موجـود در صـفحات آن بـه منظور آشنایی بیشتر با موضوع مورد نظر و نحوه انجام تحلیل مربوط به آن اطلاعات مورد نظر را کسب نماید.

خلاصه فصل دوم:

یکی از کاربردی ترین نرم افزارهای دنیای تحلیل اطلاعات نرم افزار spss می باشد که علاوه بر قدرت پردازش آماری بر روی حجم وسیعی از داده ها، به عنوان یک بانک اطلاعاتی نیز می تواند مورد استفاده قرار گیرد. این نرم افزار دارای تعدادی پنجره اصلی و پنجره فرعی می باشد. چهار پنجره اصلی این نرم افزار شامل :

پنجره Data View : محیطی برای ورود داده ها به نرم افزار و انجام ویرایش بر روی داده ها.

پنجره Viewer: محیطی که کاربر در آن نتایج تجزیه و تحلیل های نرم افزاری را مشاهده می نماید.

پنجره Syntax : محیطی متنی است که امکان برنامه نویسی را با زبان نرم افزار فراهم می آورد.

پنجره Script : محیطی را برای کاربر فراهم می نماید که بسیاری از خصوصیات پنجره ها و تنظیم صفحات مانند رنگ، اندازه، فونت، .... را به دلخواه انجام می دهد.

و اما پنجره های فرعی این نرم افزار، پنجره هایی هستند که در مواقع ضروری برای ویرایش اطلاعات پنجره های اصلی از آنها استفاده می گردد و به تنهایی ماهیت یک پنجره مستقل را دارا نمی باشند. پنجره های فرعی نرم افزار spss شامل:

پنجره فرعی Variable View : این پنجره به پیوست پنجره Data View می باشد و وظیفه آن تعریف و کنترل متغیرهایی است که در پنجره نمایشگر داده ها وارد شده اند.

پنجره فرعی Chart Editor : وظیفه این پنجره فرعی، ویرایش نمودارهای گرفته شده در خروجی Viewer می باشد.

پنجره فرعی Draft Viewer : این پنجره، امکان تایپ در محیط خروجی را علاوه بر مشاهده نتایج تجزیه و تحلیل ها را برای کاربر مهیا می نماید.

بسیاری از فعالیتهای مختلف spss بر روی محتویات پنجره ها انجام می شود با انتخاب منو آغاز می شود. هر پنجره دارای نوار منوی مخصوص به خود و هر منو نیز شامل آیتم هایی متناسب با وظایف آن پنجره در نظر گرفته شده است.

به طور مثـال پنجـره Data View دارای فایـل هـایی چـون: Fill ، Edit، View ، Data ، Transform ، Analysis ، Graphs، Utilities ، add- ones، Window و Help می باشد که هر یک متناسب با وظایف این پنجـره اقـداماتی را انجام می دهند که در این فصل (وفصول بعدی) توضیحاتی در این مورد ارائه شده است . در ایـن فصـل پـیش از ارائـه توضیحاتی در مورد نحوه استفاده از نرم افزار spss ، مطالبی در خصوص نحوه کمک رسانی نرم افزار از طریق منوHelp و زیر منو های آن ارائه شده است.

تمرین:

1. انواع قابلیتهای نرم افزار spss.16 را بازگو نمایید.
2. انواع پنجره های موجود در نرم افزار spss.16 را با ذکر وظایف هر یک نام ببرید.
3. پنجره های اصلی در نرم افزار را نام برده و وظایف هر یک را شرح دهد.
4. پنجره های فرعی در نرم افزار را نام برده و وظایف هر یک را توضیح دهد.
5. منوهای پنجره های نرم افزار را نام برده و کار هر یک را بیان کند.
6. زیر منوهای منو Help را نام برده و وظایف هر یک زیر مجموعه های آن را توضیح دهد.

فصل سوم

ورود داده‌ها و ویرایش آنها

اهداف آموزشی فصل سوم

فراگیر پس از مطالعه این فصل باید بتواند:

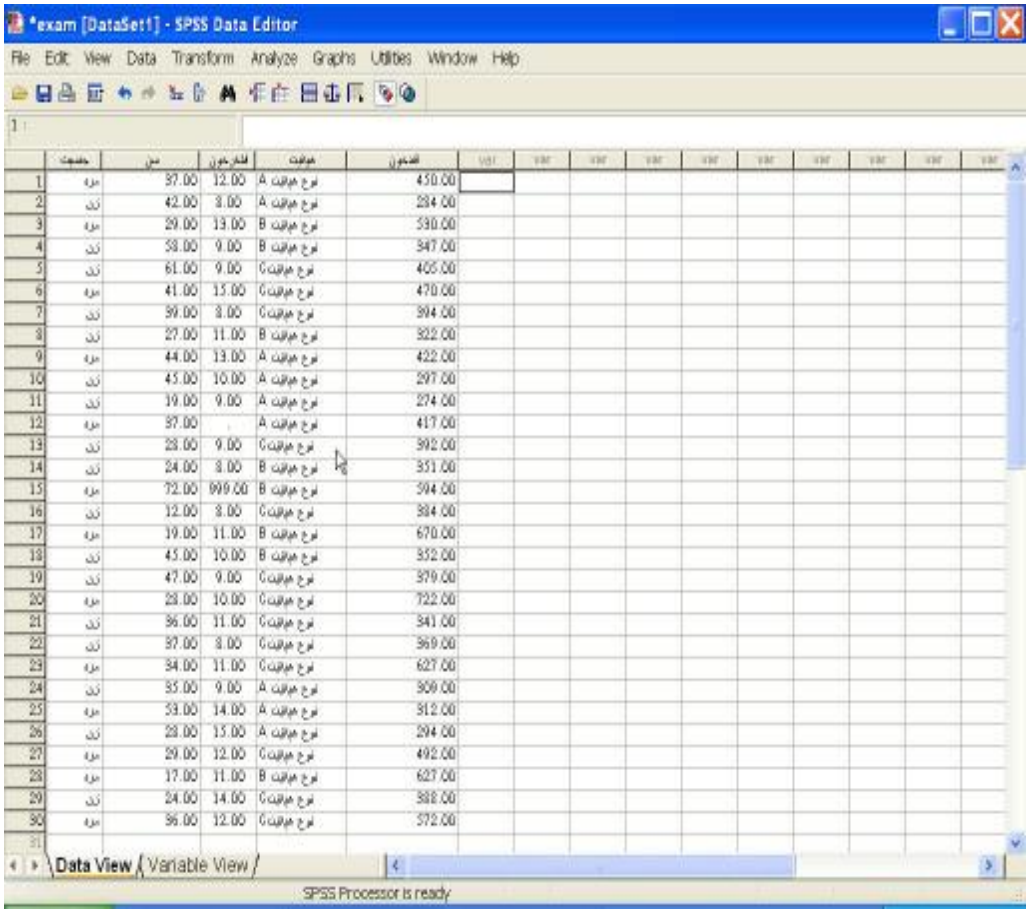
- چگونگی ورود داده ها به پنجره Data View را شرح داده و به طور کامل انجام دهد.
- امکانات منو File را بیان کند و وظایف هر یک از زیر منو های آن را شرح دهد.
- امکانات منو Edit را ذکر کرده و به طور کامل هر یک را توضیح دهد.
- ابزارهای موجود در منو View را نام ببرد و کار هر یک را توضیح دهد.
- امکانات موجود در Utilites را بیان کرده و وظایف هر یک از زیر منو های آن را به طور کار عملی نشان دهد.
- امکانات منو Windows و کاربرد هر یک را شرح دهد.

3-1) مقدمه

اولین قدم در یادگیری نحوه انجام کار با یک نرم افزار آماری نحوه ورود و خواندن اطلاعات مورد نیـاز توسط نرم افزار است. هر نرم افزار با توجه به خصوصیات و ماموریتهای خود به نحوی مشـخص اطلاعـات خام و اولیه را دریافت می نماید. نرم افزار های آماری هم شـرایط خـاص خـود را را بـرای ورود اطلاعـات صحیح و قابل پردازش دارند که عدم آشنایی با این روش ورود اطلاعات، منجر به عدم شناسایی داده هـا توسط نرم افزار و در نتیجه عدم دستیابی کاربر به اهداف خود در استفاده از نرم افزار خواهد شد. در ایـن فصل ابتدا به چگونگی ورود داده‌ها در نرم افزار اشاره می‌گردد و سپس نحوه خواندن فایلهای حاوی داده‌ها بیان می‌شود و سپس به ابزارهای موجود در نرم افزار که برای ویرایش داده‌ها بکار می‌روند ارائه شده است.

3-2) وارد کردن داده‌ها در پنجرهٔ **Data View** :

همانطور که عنوان شد پنجرهٔ نمایشگر داده‌ها، پنجرهٔ اصلی نـرم افـزار اسـت کـه ورود و فراخوانـدن و ویرایش داده‌ها در آن انجام می‌گیرد همانطور که در شکل زیر ملاحظه مـی‌شـود ایـن پنجـره از تعـدادی ستون که در بالای همهٔ آنها کلمه Var که مخفف Variable است و همچنین تعدادی سطر کـه شـمارهٔ هر سطر در قسمت سمت چپ صفحه مشخص شده تشکیل شده است.



ستونها در نرم افزار معرف متغیرها و سطرها معرف عناصر و داده‌ها هستند برای نمونه می‌خواهیم اطلاعات 30 نفر از بیماران مبتلا به هپاتیت را که متغیرهای ذیل از آنها جمع آوری شده است وارد نرم افزار نمائیم.

1- جنسیت بیماران

2- سن بیماران

3- میزان فشار خون بیماران

4- نوع هپاتیت بیماران

5- میزان قند خون بیماران

(داده‌های مربوط به این مثال در انتهای در پیوست شماره 1 کتاب موجود است).

برای این منظور در سطر اول و ستون اول در حقیقت خانهٔ اول، جنسیت نفر اول را وارد می‌نمائیم. نفر اول مرد است. از آنجا که بهتر است خصوصیات را با عدد وارد نمود، عدد 1 را معرف مردها گرفته و وارد می‌نمائیم.

در ادامه ملاحظه می‌شود که می‌توان در نرم‌افزار مرد ≈1 در نظر گرفت، بعد از وارد کردن خانهٔ اول ملاحظه می‌شود که نام ستون اول به Var00001 تغییر کرد که در حقیقت نام متغیر جنسیت است که در آینده می‌توان آن را به Sex یا جنسیت تغییر داد.

به همین طریق خانهٔ در سطر اول و ستون دوم، سن بیمار اول، در خانه سطر اول و ستون سوم میزان فشار خون نفر اول، در خانه سطر اول و ستون چهارم نوع هپاتیت نفر اول (A، B یا C) و در خانهٔ سطر اول و ستون پنجم میزان قند خون نفر اول را وارد می‌نمائیم. به این طریق اطلاعات نفر اول در سطر اول و در 5 ستون با نام‌های Var00001 تا Var00005 وارد شد. حال برای نفر دوم کافی است در سطر دوم این پنج خصوصیات را در سطر دوم در 5 ستون قرار داد. و برای نفر سوم تا نفر 30ام نیز به همین طریق عمل نمود.

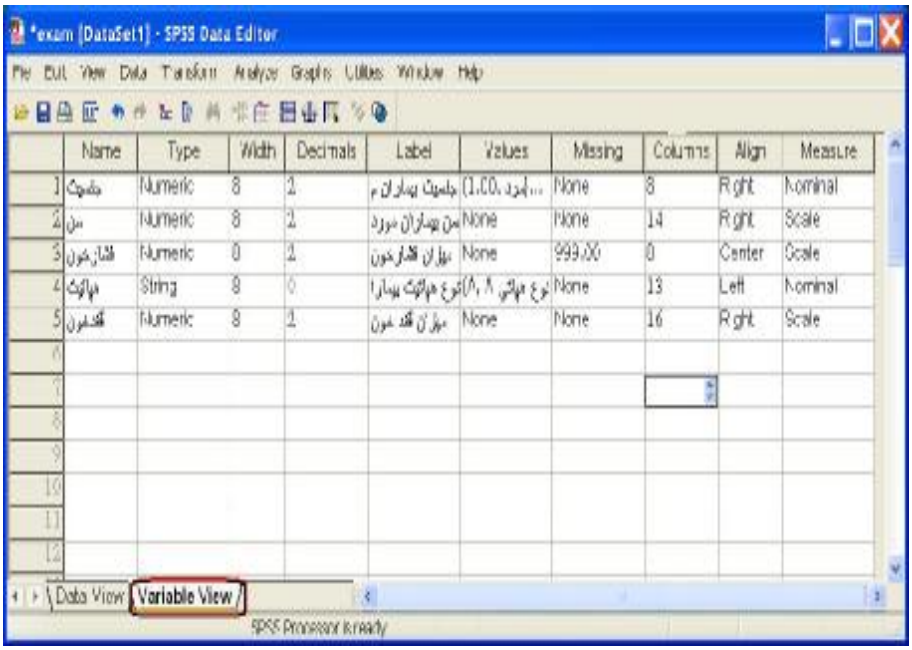
برای مثال عدد 2 در سطر 24ام و ستون Var00001 معرف جنسیت «زن=2» نفر 24ام و عدد 14 در سطر 29ام و ستون Var00003 معرف فشار خون نفر 29ام است. همانطور که گفته شد نرم‌افزار تا حداکثر 10/000 متغیر را در خود جای می‌دهد به همین دلیل نام متغیرها از Var00001 تا Var10000 توسط نرم‌افزار به صورت خودکار قرار می‌گیرد و پس از آن کاربر می‌تواند آن را تغییر دهد.

همانطور که ملاحظه می‌شود Var00001 متغیری حاوی اطلاعات جنسیت 30 نفر مورد بررسی است، Var00002 متغیر حاوی اطلاعات سن بیماران، Var00003 متغیر حاوی اطلاعات فشار خون بیماران، Var00004 متغیر اطلاعات نوع هپاتیت بیماران و Var00005 اطلاعات میزان قند خون بیماران است.

هنوز مرحله ورود اطلاعات تکمیل نشده است زیرا باید هر یک ازکدها در متغیرها، نام متغیرها، نوع متغیرها (پیوسته یا گسسته بودن) و ... در نرم‌افزار تعریف شود که این اقدامات یعنی تعریف خصوصیات

متغیرها در پنجرهٔ فرعی Variable View که یک پنجره به پیوست پنجرهٔ اصلی Data View است صورت می‌پذیرد.

برای رفتن به پنجرهٔ Variable View کاربر می‌تواند دکمهٔ Variable View و یا دکمهٔ Ctrl + t را در پنجرهٔ Data View بزند و همینطور برای بازگشت از Variable View به پنجرهٔ اصلی Data View می‌تواند دکمهٔ Data View و یا Ctrl + t را در پنجرهٔ Variable View بزند.



پنجرهٔ Variable View، ده نوع خصوصیت متغیرها را کنترل می‌نماید. لازم به ذکر است ستونهای 10 گانه Variable View نمایانگر خصوصیات قابل کنترل و سطرها متغیرها می‌باشند. برای مثال سطر سوم خصوصیات متغیر موجود در Data View را که میزان فشار خون بیماران است کنترل می‌نماید.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure

شرح خصوصیات قابل کنترل متغیرها که در ده ستون در بالای پنجرهVariable View برای هر یک از متغیرهاآمده است به قرار زیر است:

Name
جنسیت
سن
فشارخون
هپاتیت
قندخون

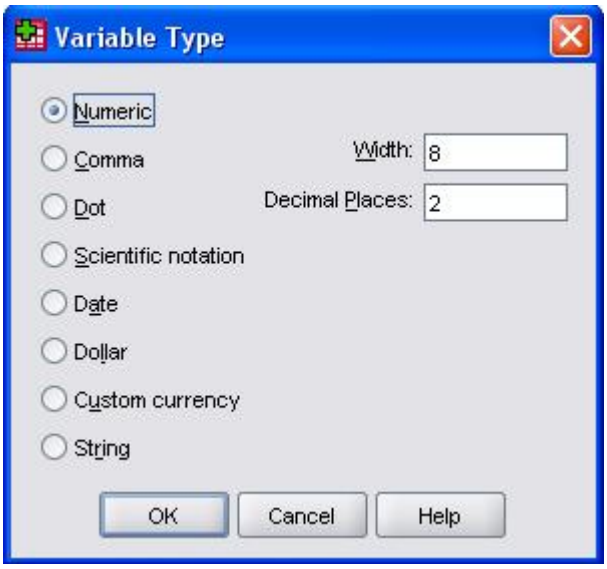
1ـ Name (نام متغیر):

کاربر می‌تواند برای تمامی متغیرهای تعریف شده در Variable View در ایـن سـتون نـام اختیـاری انتخاب نماید برای مثال به جای Var00001 که معرف جنسیت است عبارت Sex یا Gender و حتی در صورت شناسایی فونت‌های فارسی توسط نرم‌افزار عبارت «جنسیت» را به عنوان نام انتخاب نماید.

برای دیگر متغیرها نیز به جای نام ابتدایی نرم‌افزار کاربر می‌تواند نـام مـورد نظـر و متناسـب خـود را انتخاب نماید. ذکر این نکته ضروری است که نام یک متغیر حداکثر باید دارای 8 کاراکتر باشد و با حروف شروع شود و هیچ دو متغیری نمی‌توانند نام یکسان داشته باشند. اگر نام متغیرها را تغییر دهید هنگـامی که به پنجره ورود داده های Data View برگردیم ملاحظه می‌کنیم که نـام متغیرهـا بـه نـام دلخـواه و تعریف شده تغییر کرده است.

2ـ Type (نوع متغیر):

داده‌های یک متغیر می‌توانند انواع گوناگونی داشته باشند در زیر انواع این داده‌ها بیان شده است. برای انتخاب انواع مختلف داده‌ها در هر سطر متغیر بر روی مستطیل سمت راست ستون Type کلیک نمائیـد تا پنجرهٔ Variable Type باز گردد.



ـ نوع Numeric (عددی): اختصاص به ورود مقادیر عددی عادی برای متغیرها دارد برای نمونه متغیر

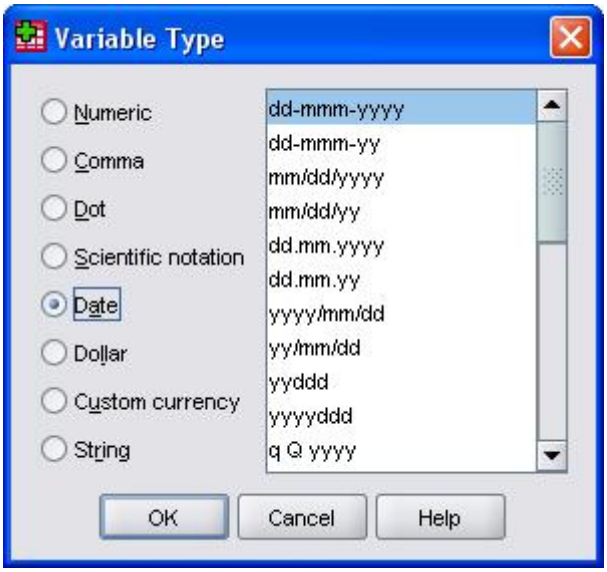
جنسیت که با اعداد 1 و 2 معرف مردها و زنها وارد شده است یک متغیر عددی است در صورتی که متغیر نوع هپاتیت که با حروف A، B یا C معرف هر یک از 3 نوع هپاتیت وارد شده است دیگر یک متغیر عددی نیست، گزینهٔ Width معرف طول کل عدد و Decimal Place تعداد نقاط اعشار قابل نمایش را تعیین می‌کند. پیش فرض این مقادیر به ترتیب 8 و 2 است.

ـ نوع Camma (ویرگول): در این نوع نمایش، هر سه رقم عددی از سمت راست با یک علامت ویرگول از سایر ارقام جدا می‌شود مانند 370 و 246 و 1

ـ نوع Dot (نقطه): در این نوع نمایش، هر سه رقم از سمت راست با یک علامت نقطه از سایر ارقام جدا می‌شود مانند 1.246.370

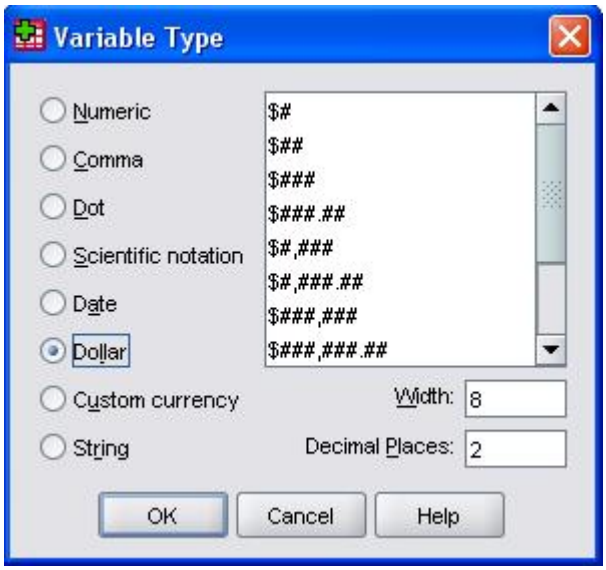
ـ نوع Scientific notation (نماد علمی): هر رقمی بر حسب مضربی از توانهای 10 نوشته می‌شود مانند 1.26E+8 که معادل 126000000 است یا 1.26 که به صورت $1.2\times10^{0} = 1.2E+00$ نوشته می‌شود.

ـ نوع Date (تاریخ): که برای نمایش زمان مشاهدات با دقت‌های متفاوت در نظر گرفته شده است. پس از انتخاب این گزینه جعبه‌ایی که محتوای انواع نمایش زمانی داده‌هاست در جعبهٔ گفتگو ظاهر می‌شود که با انتخاب نوع مناسب، می‌توان زمانها را در متغیر وارد نمود برای مثال با انتخاب dd.mm.yy در جعبهٔ گفتگو متغیر مورد نظر زمان را به صورت دو کلمه معرف روز (dd)، دو کلمه معرف وسط معرف ماه(mm)و دو کلمهٔ yy معرف سال قبول می‌نماید، برای مثال 12.04.99 معرف روز دوازدهم ماه چهارم سال 1999 است.



ـ نوع Dollar (دلار): برای ورود مقادیر قیمت‌ها بر حسب واحد دلار.

پس از انتخاب این گزینه، جعبه‌ای که قالب‌های متفاوت ورود این گزینه را نشان می‌دهد ظاهر می‌شود.



برای مثال در صورتی که کاربر گزینه $###.## را انتخاب نماید مقدار پول او را که در Data View 453/04 معادل (چهارصد و پنجاه و سه دلار و چهار سنت) است به صورت 453.04$ و با انتخاب $### آن را 450$ در متغیر مربوطه در پنجرهٔ Data View نمایش می‌دهد.

ـ نوع Custom Currency (سفارشی):

با انتخاب این گزینه جعبهٔ انتخابها و نمونهٔ قالبهای آن و تعداد کاراکترها و ارقام اعشار هر نوع در جعبهٔ گفتگو ظاهر می‌شود. با انتخاب نوع سفارش مورد نظر می‌توان داده‌ها را با آن قالب وارد نمود.

ـ نوع String (کاراکتری):

این نوع متغیر مخصوص ورودیهایی است که به صورت رشته‌ای از کاراکترها (مانند اسامی خاص) در متغیرها وارد می‌شوند که تعداد حروف آنها در قسمت Characters تا حداکثر 255 کاراکتر قابل تغییر است. برای نمونه متغیر نوع هپاتیت با حروف A و B و C بر حسب نوع هپاتیت مبتلایان وارد شده است پس نوع آن String است. اگر به جای نوع هپاتیت A و B و C از اعداد 1 و 2 و 3 استفاده می‌نمودیم آنگاه باید نوع متغیر را به جای String از نوع Numeric انتخاب می‌کردیم.

ذکر این نکته ضروری است که همواره داده‌ها را در کلیه نرم‌افزارهای آماری به صورت عددی وارد نمائید و سپس مفهوم هر یک از کدهای عددی را برای نرم‌افزار مشخص نمائید، مانند متغیر جنسیت که کد جنسیت افراد به صورت 1 و 2 مشخص شده و در ادامه مفهوم عدد 1 به معنی مرد و 2 به معنی زن را برای نرم‌افزار در پنجرهٔ Variable View مشخص می‌نمائیم.

3ـ Width (پهنا): پهنا یا اندازهٔ واقعی داده‌های وارد شده که می‌خواهیم درData View نمایش داده شود توسط این مشخصه برای هر متغیر تعیین می‌شود. مقدار پهنای پیش فرض برای داده‌ها برابر 8 کاراکتر است که می‌توان آن را کم یا زیاد نمود.

4ـ Decimal (رقم اعشار) : مشخص‌کننده تعداد ارقام اعشار قابل نمایش برای یک متغیر است که پیش فرض آن تنها برای متغیرهای Numeric برابر عدد 2 است برای مثال اگر Decimal را برابر عدد 2 در نظر بگیریم عدد 1272/341 را در آن متغیر در صفحهٔ نمایشگر داده‌ها (Data View) به صورت 1272/34 نمایش می‌دهد.

5ـ Lable (برچسب): یک برچسب متناسب با نام متغیر تعریف می‌کند که در مقایسه با نام متغیر محدود به 8 کاراکتر نمی‌باشد و می‌تواند تا 256 کاراکتر در مورد متغیر توضیح را در خود جای دهد .

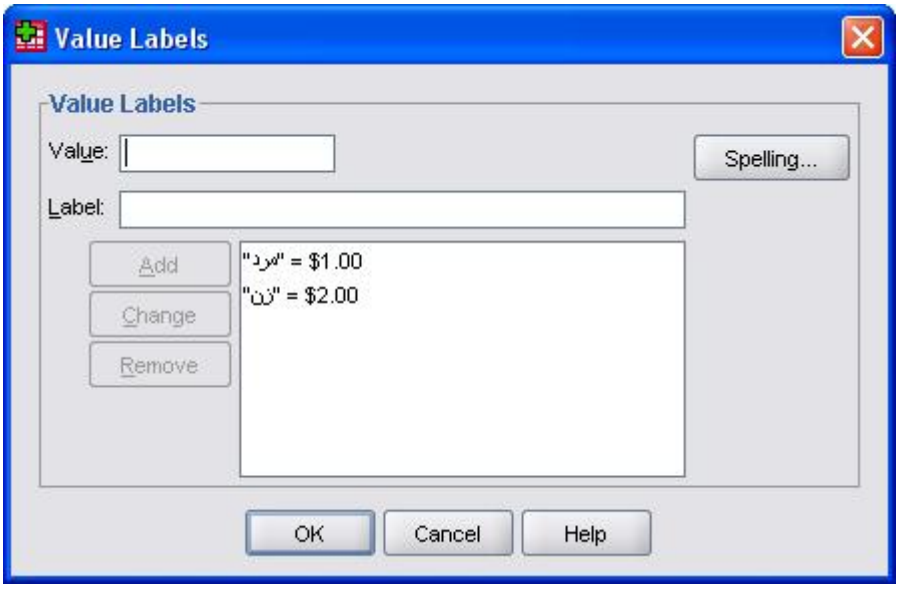
	Name	Type	Width	Decimals	Label
1	جنسیت	Numeric	8	2	جنسیت بیماران مورد بررسي
2	سن	Numeric	8	2	سن بیماران مورد بررسي
3	فشارخون	Numeric	8	2	میزان فشارخون بیماران مورد بررسي
4	هپاتیت	String	8	0	نوع هپاتیت بیماران مورد بررسي
5	قندخون	Numeric	8	2	میزان قند خون بیماران مورد بررسي

در خروجی‌ها نیز می‌تواند به جای اسم متغیرها آورده شود. برای نمونه در شکل ذیل به برچسب هر یک از متغیرهای تعریف شده دقت نمائید.

6ـ Values (مقادیر): برای تعریف هر یک از کدهایی که به سطوح مختلف یک متغیر گسسته نسبت داده شده است به کار می‌رود.

به وسیله کلیک کردن روی Values در سطر هر متغیر پنجرۀ Value Lables بازخواهد شد که توسط آن برچسب‌ها برای کدهای متغیرهای گسسته نسبت داده می‌شود.

شکل ذیل جزئیات انجام این فرآیند را برای متغیر جنسیت نشان می‌دهد. عدد 1 را در قسمت Values و عبارت «مرد» را در قسمت Value Lable وارد می‌نمائیم و دکمۀ Add را می‌زنیم به همین ترتیب مقدار 2 را در قسمت Value و عبارت «زن» را در قسمت Value Lable می‌نویسیم و دکمۀ Add را می‌زنیم نتیجه را در پنجرۀ Value Lables برای متغیر جنسیت ملاحظه می‌نمائید.



ذکر این نکته ضروری است که کاربر تنها زمانی که فونت‌های فارسی بر روی سیستم آنها نصب شده و توسط نرم‌افزار SPSS قابل شناسایی است می‌توانند حروف فارسی را در محیط نرم‌افزار بکار برند. در غیر این صورت بهتر است از حروف انگلیسی برای Name، Lable و Values Lable متغیرها استفاده نمایند.

7ـ Missing (مقادیر گمشده): کاربر می‌تواند برای هر متغیر مقادیر گمشده را مشخص نماید، مقادیر گمشده مقادیری هستند که بنا به دلایل گوناگون امکان دسترسی به آنها مهیا نبوده است. برای مثال نفر 12 و نفر 15 ، میزان فشار خون آنها ثبت نشده است.

در صورتی که در متغیر فشارخون [.] یعنی خانه‌ای بدون اطلاعات مشاهده شود که داخل آن یک نقطه قرار دارد آن خانه اطلاعاتش در دست نبوده و Missing است اما از نوع داده‌های گمشده‌ای که نرم‌افزار برای خود تعریف کرده که به System Missing معروف است.

برای مثال فشار خون نفر 12ام یک System Missing است. اما نوع دیگری از داده‌های گم شده را کاربر خود برای نرم‌افزار مشخص می‌نماید برای مثال کاربر عدد 999 را در یک متغیر دادهٔ گم شده در نظر می‌گیرد و اطلاعات مربوط به هر متغیر را که ندارد با این کد وارد می‌نماید و سپس در پنجرهٔ Variable View و ستون Missing مربوط به آن متغیر مشخص می‌نماید خانه‌ایی که در یک متغیر خاص کد 999 دارد یک دادهٔ گم شده محسوب می‌گردد که به آن User Missing یا دادهٔ گم‌شدهٔ ایجاد شده توسط کاربر می‌گویند.

کاربرد User Missing در این است که ممکن است کاربر بتواند در آینده اطلاعات مربوط به نوع خاصی از داده‌های گم شده را بدست آورده و جایگزین مقادیر گم شده نماید.

برای انتخاب User Missing هر متغیر در ستون Missing برای هر متغیر کلیک نمائید پنجرهٔ محاوره‌ایی مانند شکل ذیل باز می‌شود که دارای 3 گزینه است.

No Missing Value : به این معنی است که شما برای این متغیر دادهٔ گم شده در نظر نگرفته‌اید.

Discrete Missing Value : دارای 3 جای خالی است که کاربر می‌تواند برای هر متغیر حداکثر 3 نوع داده گم شده با کدهای متفاوت ایجاد نماید.

Range Plus One Optional Discrete Missing Value : کاربر می‌تواند یک فاصله از اعداد را به علاوه یک عدد گسسته خارج از آن فاصله به عنوان مقادیر گم شده معرفی نماید. برای مثال اگر Low:12 ،High:14 و Discrete Value:17 باشد همهٔ اعداد بین 12 و 14 همراه با عدد 17 را به عنوان مقادیر گم شده لحاظ می‌کند.

8- Columns (ستونها): پهنای ستون متغیرها را در صفحه گسترده‌ی Data View مشخص می‌کند برای مثال هنگامی که پهنای ستون متغیر جنسیت به 15 افزایش می‌یابد عرض ستون جنسیت در Data View افزایش می‌یابد.

9ـ Align (ترازو): طرز قرار گرفتن داده‌ها در خانه‌ها Data View (Cell) را نمایش می‌دهد که می‌تواند راست چین (Right)، چپ‌چین (Left) و یا وسط‌چین (Center) باشد.

10ـ Measure (معیار اندازه‌گیری متغیرها): این ستون مهمترین ستون پنجرهٔ Variable View است و مشخص می‌کند متغیرها از چه نوعی هستند، همانطور که در فصل اول گفته شد داده‌ها می‌توانند

از نوع پیوسته (Scale)، گسسته اسمی (Nominal) و یا گسسته ترتیبی (Ordinal) باشند و تأکید شد که انواع روشهای آمار استنباطی به دلیل وجود انواع متغیرهاست.

در اطلاعات مورد بررسی، متغیر جنسیت یک متغیر گسسته است اما چون سطوح آن نسبت به هم برتری ندارند از نوع گسسته اسمی یا Nominal، متغیر سن بیماران، میزان فشار خون و قند خون از نوع متغیر پیوسته (Scale) و متغیر نوع هپاتیت نیز از نوع متغیر گسسته اسمی (Nominal) است زیرا سطوح A و B و C هپاتیت نسبت به هم برتری ندارند، اگر اطلاعات سطح سواد افراد را در دست داشتیم آنگاه آن یک متغیر گسسته ترتیبی (Ordinal) بود.

3ـ3ـ امکانات منو File:

همانگونه که پیش از این گفته شد این منو دارای امکاناتی در زیر منوهای خود است که ارتباط میان نرم افزار را با محیط بیرون نرم افزار فراهم می آورد. زیر منوهای موجود در منو هر یک وظایف خاصی را بر عهده دارند که در ذیل بیان شده است:

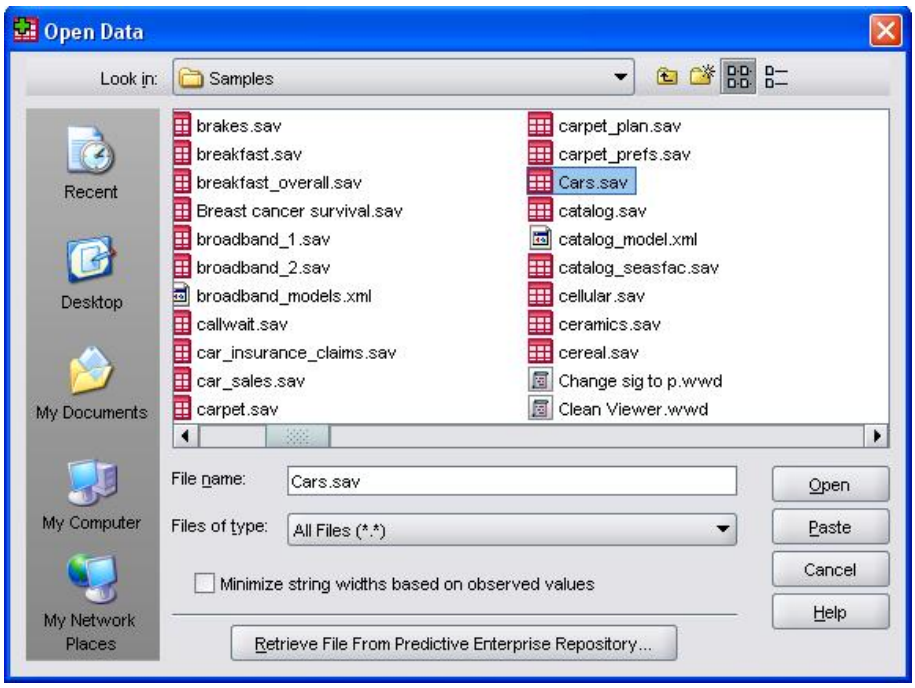
- فراخواندن فایلهای حاوی اطلاعات(از طریق زیر منو File):

برای فراخواندن این فایلها از منو File و آیتم Open استفاده می شود.فایلهای اطلاعاتی در نرم‌افزار SPSS به چهار گروه زیر تقسیم می‌شوند:

1ـ فایلهای حاوی داده‌ها که با پسوند sav.* شناخته می‌شوند و برای وارد کردن در پنجرهٔ Data View فراخوانده می‌شوند.

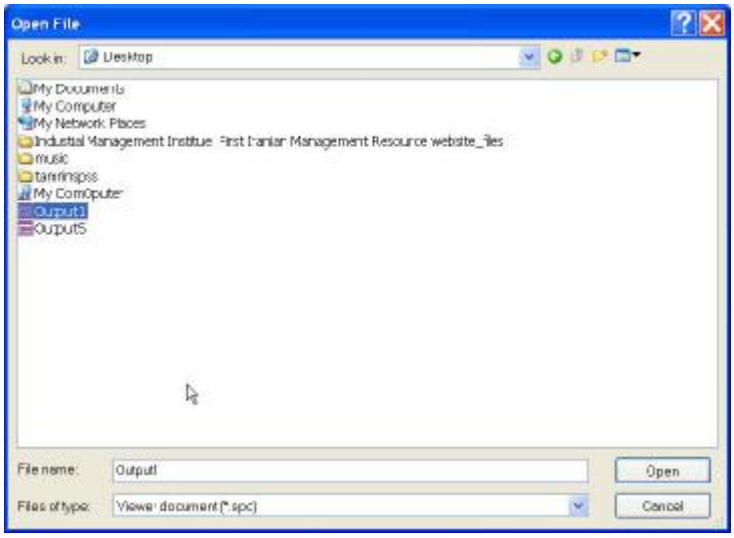
برای فراخواندن این فایلها از منو File و آیتم Open مانند شکل ذیل گزینهٔ Data را انتخاب نموده و بعد از انتخاب فایل مورد نظر آن را از طریق جعبهٔ گفتگوی Open File وارد Data View می‌نمائیم. برای مثال فایل Cars.sav که حاوی اطلاعات خودروهای تولید شده در 3 منطقهٔ آمریکا، اروپا و ژاپن

است را از طریق آیـتم Open و آدرس G:\samples\cars (موجـود در cd کتـاب) در صـفحهٔ Data View باز نمائید.

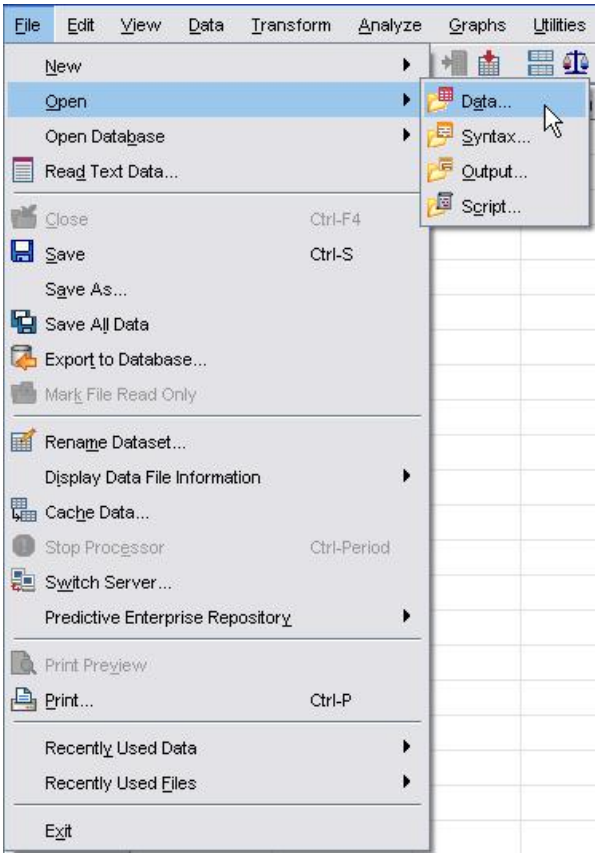


2ـ فایلهای حاوی برنامه‌ها که با پسوند sps.* در نرم‌افزار شناخته می‌شوند و برای وارد کردن آنهـا در پنجرهٔ Syntax باید از طریق منو File و آیتم Open همانند بالا عمل گردد.

3ـ فایلهای حاوی خروجی‌ها که با پسوند spo.* شـناخته مـی‌شـوند. هـر گـاه کـاربر نتـایج تجزیـه و تحلیل‌ها را ذخیره نماید می‌تواند آن فایل را دوباره از طریق منو File و آیـتم Open و انتخـاب Output مجدداً احضار نماید.



4ـ فایلهای حاوی Scriptها که با پسوند sbs.* شناخته می‌شوند و کاربر مـی‌توانـد از طریـق انتخـاب File→Open→Script فایلهای حاوی Script را که قابلیت تغییر ظاهر نرم‌افزار را دارند احضار نماید.



- باز کردن پنجره‌های جدید (از طریق زیر منوNew) :

برای باز کردن پنجره‌های جدید از هر یـک از انـــواع Data ، Syntax، Output و Script کاربر می‌تواند از منو file و آیـتم New ماننـد شکل مقابل استفاده نماید. با انتخاب هر یک از این گزینه‌ها پنجـره‌ایـی فاقـد اطلاعـات بـرای کاربر باز می‌گردد.

ذکر این نکته ضروری است کـه بـا انتخـاب

گزینه Draft Output پنجره‌ایی خروجی برای کاربر باز می‌شود که علاوه بـر گـرفتن خروجـی‌هـا در آن می‌توان متن‌های مورد نظر و توضیحات مکمل را نیز درون آن تایپ نمود.

در بالا نحوه باز کردن یک پنجره و همچنین نحوه فراخواندن انـواع فایلهـا حـاوی داده‌هـا، برنامـه‌هـا، خروجی‌ها و Scriptها در نرم‌افزار که در آیتم‌های New و Open از منو File قـرار دارنـد بیـان شـد در ادامه وظیفه دیگر آیتم‌های منو File را که وظیفه کنترل فایلهای موجود در نرم‌افزار را بر عهده دارند بیان می‌کنیم.

- **فراخواندن یک فایل متنی (ازطریق زیر منو**Read Text Data**)**:

برای خواندن داده‌ها از یک فایل متنی به صورت ذیل عمل می‌نمائیم:

1ـ مسیر زیر را انتخاب می‌کنیم.

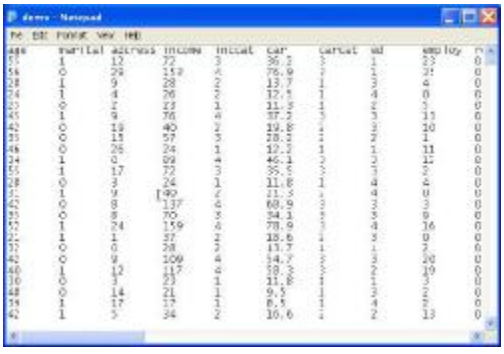
File>Read Text Data

2ـ نوع Text (*.txt) را از قسمت Type of Files انتخاب می‌کنیم.

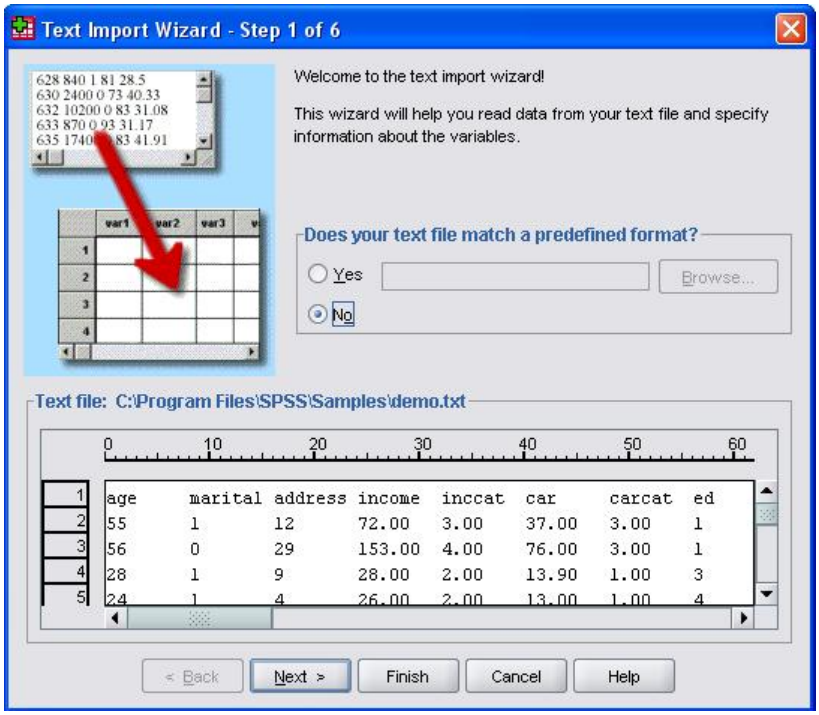
3ـ فایل Demo.txt را از مسیر:

G: \ samples\demo.txt

انتخاب می‌نمائیم و Open می‌کنیم.



جعبهٔ محاوره Text Import Wizard کاربر را در پروسه تعیین چگونگی تفسـیر فایـل متنـی همـراه می‌نماید.



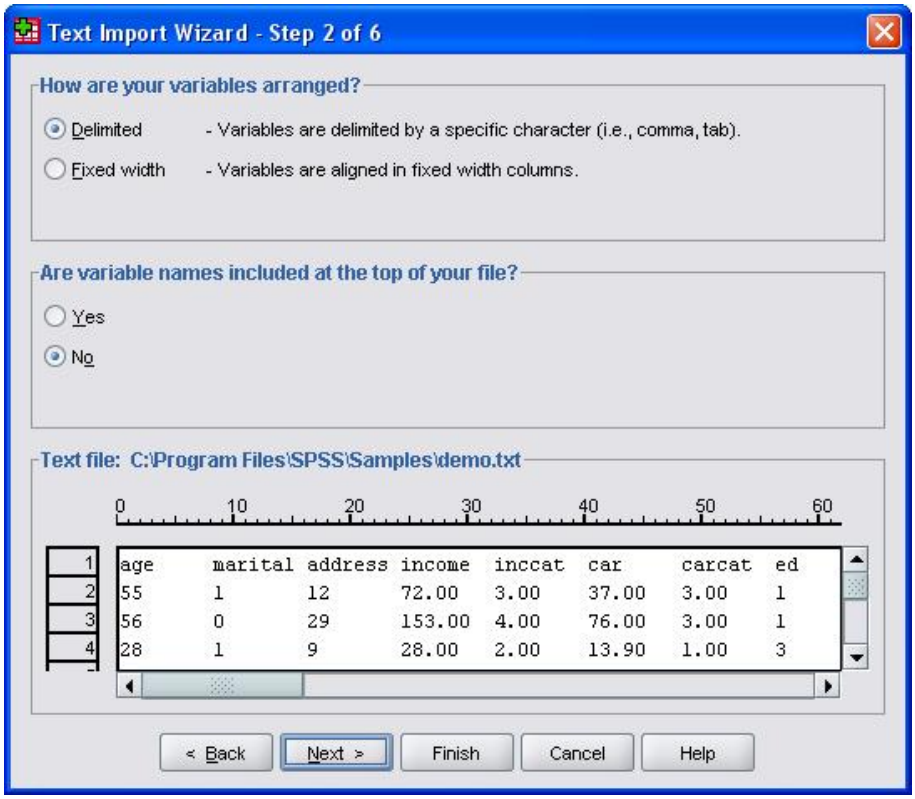
4ـ در ابتدا، می‌توانید یک قالب از پیش تعریف شده را برگزینید یا یک قالب جدید را بسازید. گزینهٔ No را فعال نمائید تا مشخص شود که باید یک قالب جدید ایجاد شود.

5ـ دکمهٔ Next را کلیک کنید.

6ـ گزینهٔ Delimited را برگزینید تا مشخص نمایید که داده‌ها از یک ساختار قالب‌بندی مرزی استفاده نموده‌اند.

7ـ yes را فعال نمائید تا مشخص شود که نام متغیرها باید از بالای فایل خوانده شود.

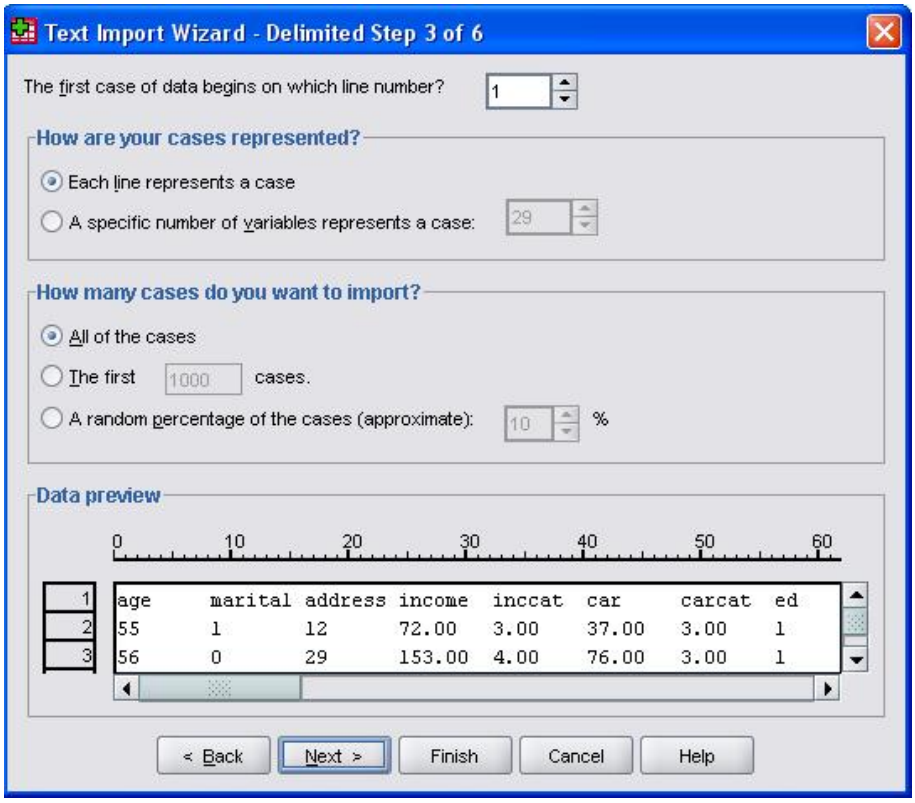
8ـ دکمهٔ Next را انتخاب کنید.



9ـ در کادر The First Case در بالای جعبهٔ محاورهٔ بعدی، عدد 2 را تایپ کنید به این معنی که سطر اول داده‌ها از خط دوم فایل متنی آغاز گردد.

10ـ دیگر گزینه‌ها را تغییر ندهید و روی Next کلیک کنید.

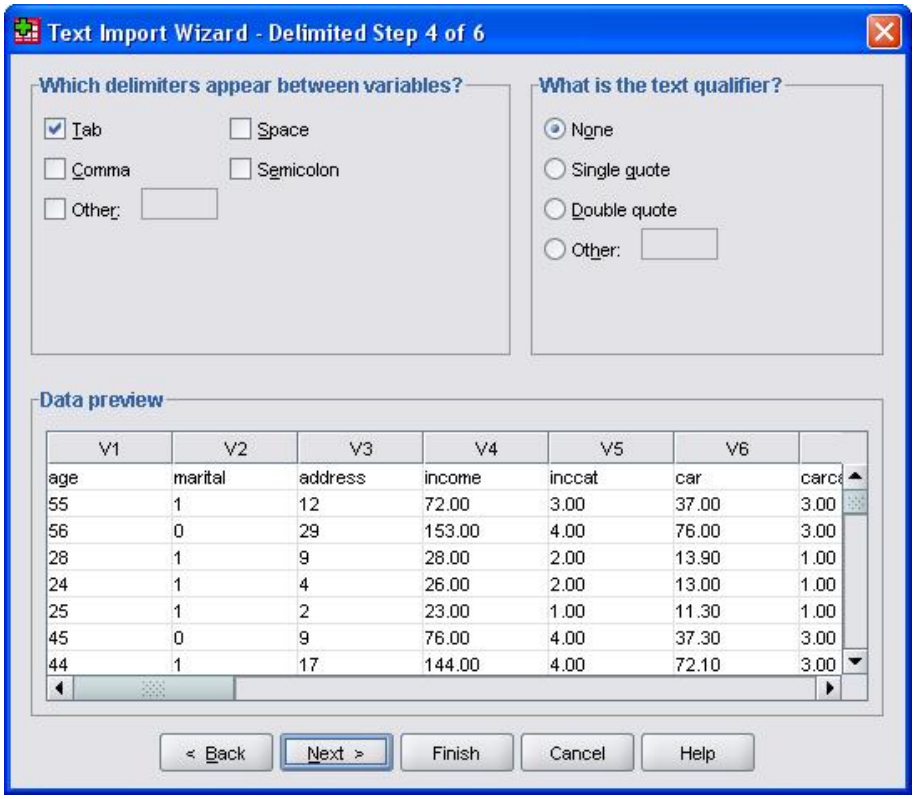
11ـ قسمت Data Preview در جعبهٔ محاورهٔ جدید این امکان را فراهم می‌کند که با سرعت مطمئن شوید داده‌های شما به درستی توسط SPSS خوانده شده‌اند.

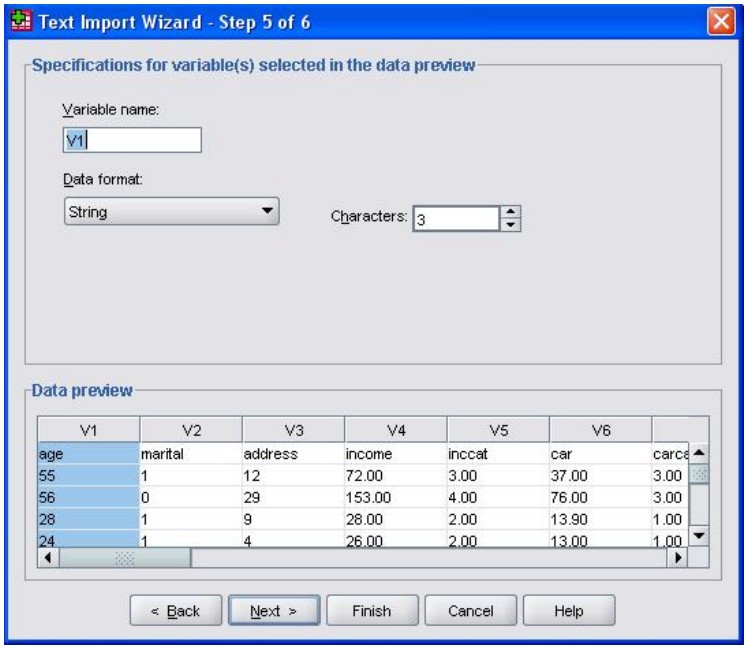


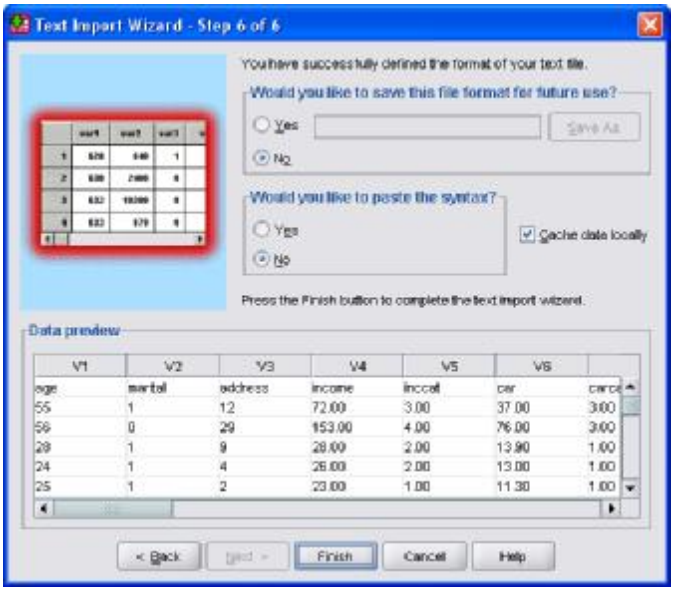
12ـ گزینه Tab را فعال و دیگر گزینه‌ها را غیرفعال نمائید.

13ـ روی دکمه Next کلیک نمائید.

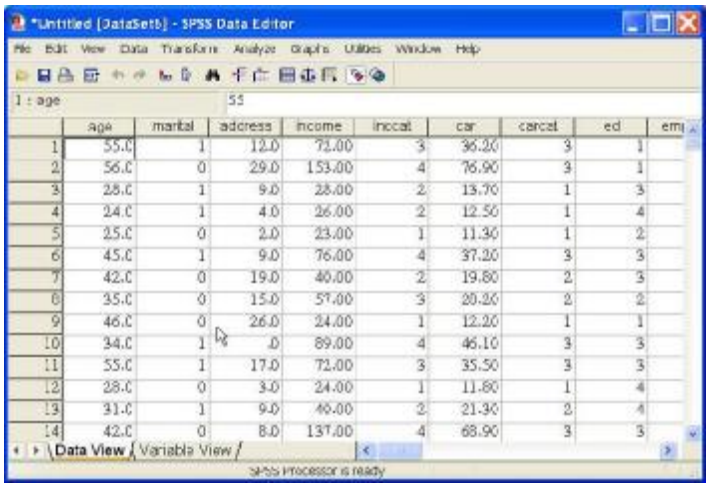
از آنجا که نام متغیرها ممکن است برای انطباق با قالب‌بندی مورد نیاز SPSS خلاصه شده باشد، جعبهٔ محاورهٔ جدید مطابق شکل ذیل امکان ویرایش نام‌های نامطلوب را فراهم می‌آورد.


Text Import Wizard - Delimited Step 4 of 6
Which delimiters appear between variables?
Tab
Space
Comma
Semicolon
Other:
What is the text qualifier?
None
Single quote
Double quote
Other:
Data preview
V1 V2 V3 V4 V5 V6
age marital address income inccat car
55 1 12 72.00 3.00 37.00 3.00
56 0 29 153.00 4.00 76.00 3.00
28 1 9 28.00 2.00 13.90 1.00
24 1 4 26.00 2.00 13.00 1.00
25 1 2 23.00 1.00 11.30 1.00
45 0 9 76.00 4.00 37.30 3.00
44 1 17 144.00 4.00 72.10 3.00
< Back
Next >
Finish
Cancel
Help


Text Import Wizard - Step 5 of 6
Specifications for variable(s) selected in the data preview
Variable name:
V1
Data format:
String
Characters:
3
Data preview
V1 V2 V3 V4 V5 V6
age marital address income inccat car
55 1 12 72.00 3.00 37.00 3.00
56 0 29 153.00 4.00 76.00 3.00
28 1 9 28.00 2.00 13.90 1.00
24 1 4 26.00 2.00 13.00 1.00
< Back
Next >
Finish
Cancel
Help



14ـ در جعبهٔ محاورهٔ بعدی، گزینه‌های پیش‌فـرض را تغییـر ندهیـد و روی Finish کلیـک کنیـد تـا داده‌های متنی در پنجرهٔ Data View ظاهر شوند.



- خواندن داده‌های فایلهای پایگاه داده‌ها (از طریق زیر منو Open Database) :

نرم‌افزارهایی مانند Excel، Access، Foxpro و ... که دارای پایگاه داده‌ها هستند نیز می‌توانند به عنوان منبع ورود داده‌ها به محیط Data View در نظر گرفته شوند، این امکان در منوFileدر زیر منو:

File>Open Database

روشی را برای انتقال داده‌ها از این محیط‌ها به SPSS در اختیار کاربران قرار داده است. اما مـا بـه دلیـل

طولانی بودن مسیر آن از یک روش ساده‌تر استفاده می‌نمائیم.

از آنجا که نرم‌افزار Excel یک نرم‌افزار از مجموعهٔ Office است و فرمت اطلاعات آن توسط تمامی نرم‌افزارها قابل شناسایی است پیشنهاد می‌شود از نرم‌افزار Excel به عنوان یک نرم‌افزار واسطه استفاده گردد یعنی ابتدا داده‌ها از طریق عمل Copy در محیط نرم‌افزار و Paste در محیط Excel به محیط Excel وارد شوند و سپس از طریق Copy کردن از محیط Excel و Paste کردن در محیط SPSS وارد پنجرهٔ Data View شوند. برای مثال می‌توانید به راحتی اطلاعات فایل داده‌ایی demo.xls را از مسیر:

G:\samples\demo.xls

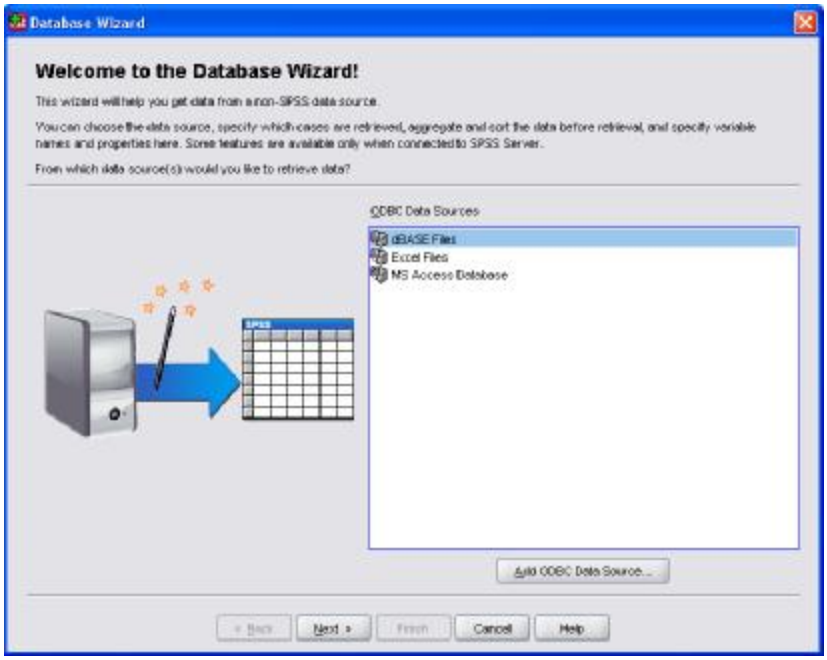
Copy کرده و در محیط Data View، Paste نمائید.

همانگونه که ملاحظه می شود این دستور مانند شکل زیر دارای سه گزینه می باشد.

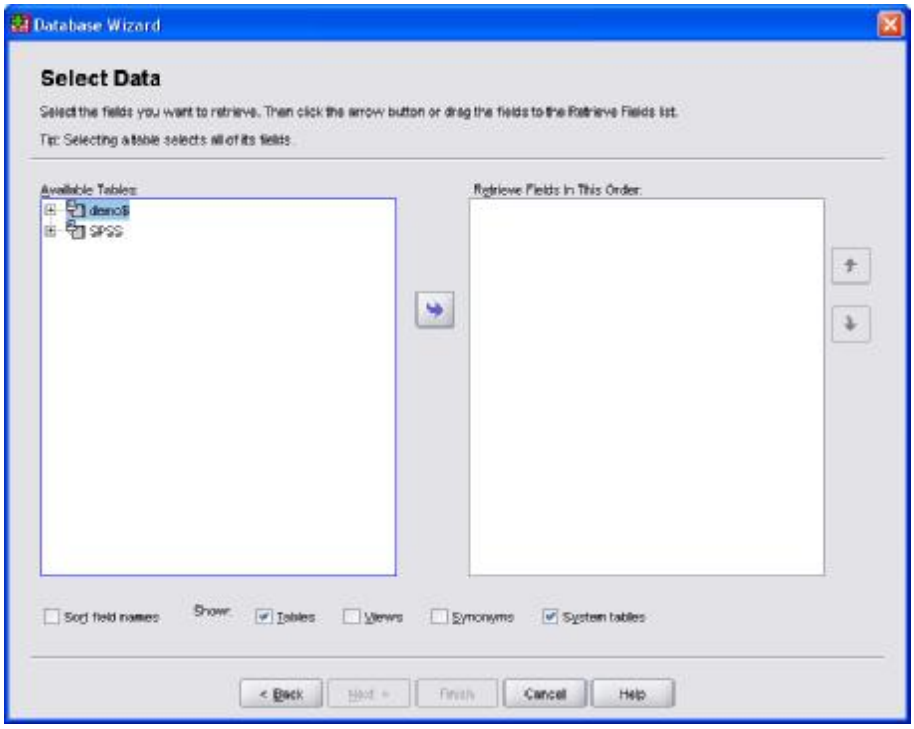
در منو New Query این امکان را به کاربر می دهد که برای اولین بار با استفاده از یک ریز برنامه که به صورت مرحله به مرحله می نویسد امکان فراخواندن یک فایل در یک بانک اطلاعاتی را در محیط spss داشته باشد . برای نمونه می خواهیم فایل demo که یک فایل در محیط اکسل می باشد را به نرم افزار منتقل نمائیم. مسیر فایل مورد نظر به صورت

G:\samples\demo.xls

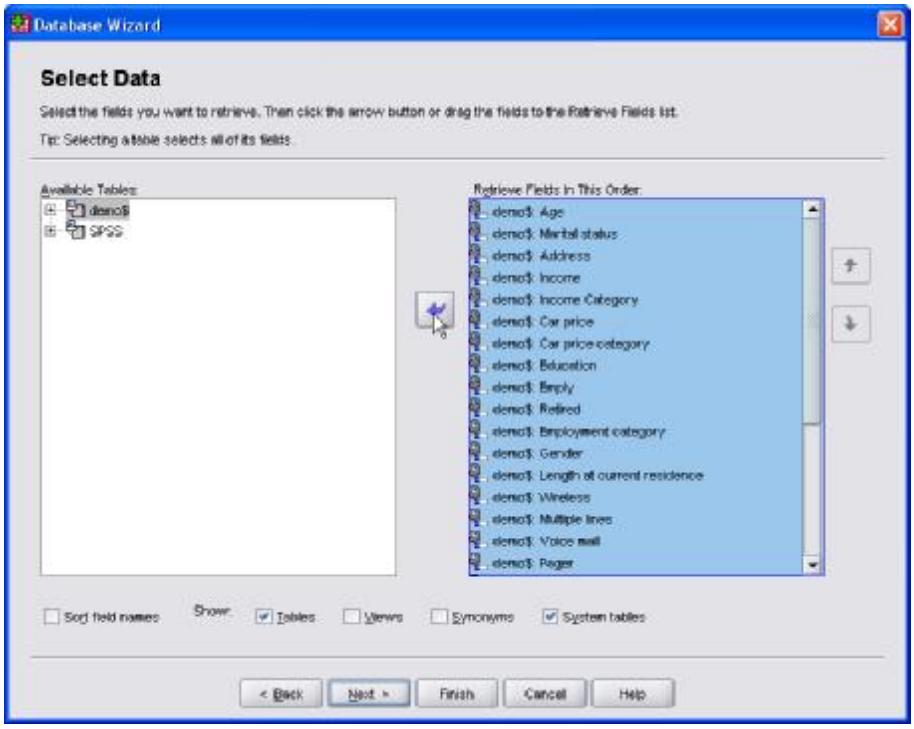
می باشد. ابتدا گزینه NewQuery را از گزینه open Database انتخاب می نماییم تا جعبه محاوره زیر باز گردد.



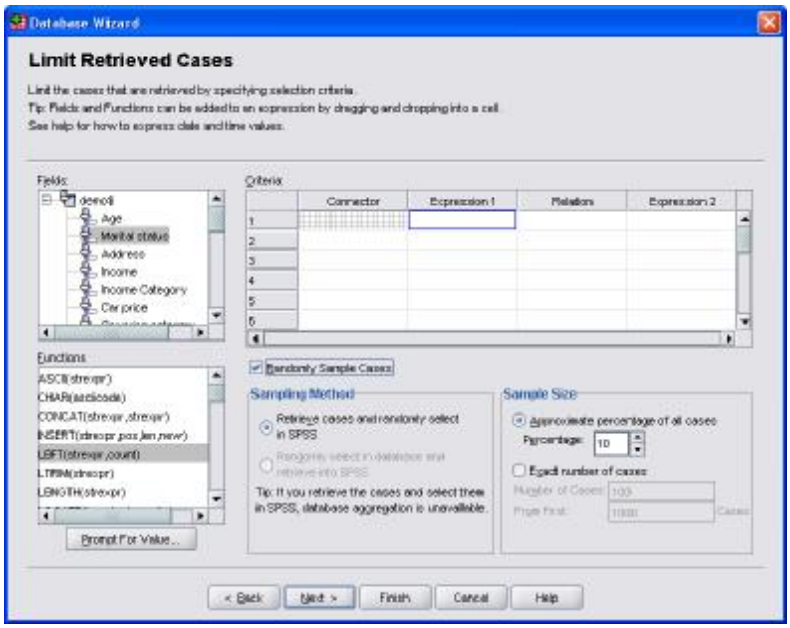
در جعبه محاوره بالا از کاربر پرسیده می شود که بانک اطلاعاتی که فایل مورد نظر تحت آن قـرار دارد را مشخص نماید. کاربر علاوه بر بانک اطلاعاتی اکسل و اکسس که در بالا مشـاهده مـی شـود مـی توانـد از قسمت Add ODBC Data Sourse پایگاه داده مورد نظر خود را معرفی نماید. با توجـه بـه اینکـه در نظر است تا اطلاعات فایل Demo را در قالب اکسل وارد نـرم افـزار نمـائیم پـس فرمـت Exell files را انتخاب نموده و دکمه Next را می زنیم تا به صـفحه زیـر منتقـل شـویم. سـپس بـا اسـتفاده از گزینـه Browse مسیر فایل مورد نظر را وارد نموده و ok را می زنید. تا پنجره زیر باز شود.



از آنجا که نرم افزار اکسل دارای sheet های زیادی می باشد که هر یک دارای اطلاعاتی متفاوتی می توانند باشند در قسمت سمت چپ صفحه همه sheet های موجود در فایل Demo را در قالب Demo$ نمایش می دهد. اگر فایل Demo دارای دو sheet با نامهای Demo1 و Demo2 بود آنگاه در قسمت سمت چپ Demo1$وDemo2$ و Spss$ قابل مشاهده بودند که دو تای اولی معرف هر یک از sheet های موجود اکسل و spss معرف مجموع اطلاعات در هر دو sheet می بود . در حقیقت برای آنکه اطلاعات همه sheet های یک فایل را به نرم افزار وارد نمائیم کافی است تا sheet با نام spss را که در آخر لیست سمت چپ در پنجره بالا قرار دارد را انتخاب نمائیم. در فایل demo با توجه به اینکه تنها یک sheet موجود است فرقی نمی کند Demo$ و یاSpss$ انتخاب گردد. یکی را انتخاب می نمایم و به قسمت سمت راست منتقل می کنیم . مانند شکل ذیل، مشخص می شود که کدام sheet و در داخل آن کدام متغیرها از فایل اکسلی قرار است به نرم افزار وارد شوند.

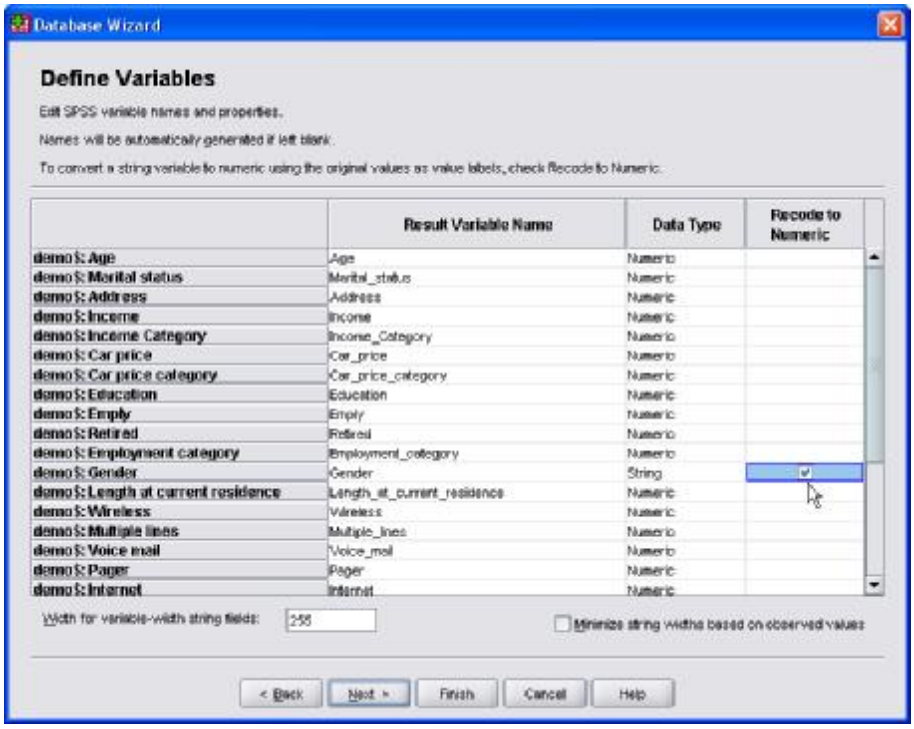


بعد از انتخاب متغیرها دکمه Next را می زینیم تا پنجره زیر باز گردد.

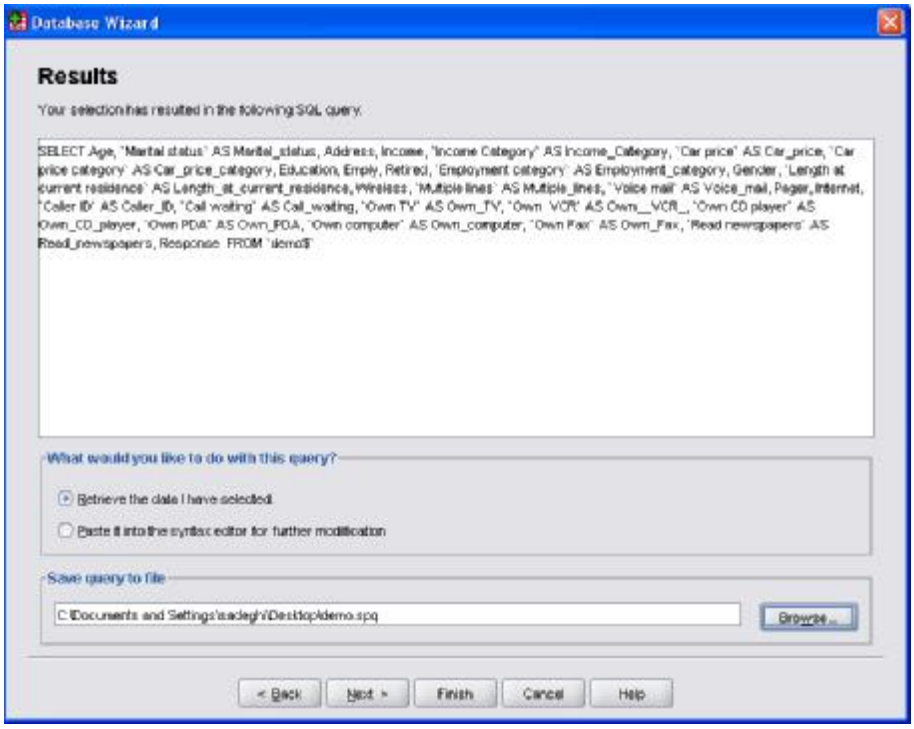


در پنجره بالا مشخص می شود که شما چه حجمی از داده های مورد نظر را قصد دارید به نرم افزار spss وارد نمائید. اگر می خواهید همه داده ها به نرم افزار وارد شوند کافی است گزینه Next را وارد نمائیـد و در غیر این صورت می توانید از قسمت Randomly sample cases و انتخاب حجم نمونه از قسـمت

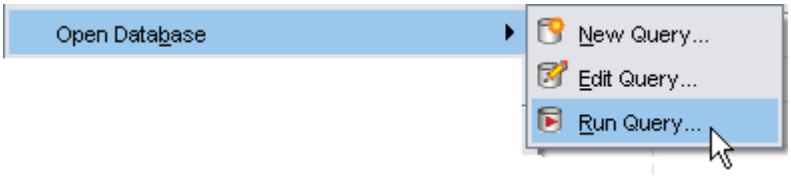
sampl size امکان آوردن تنها نمونه ایی از اطلاعات از نرم افزار اکسل را فراهم نمائید. بعد از انتخاب حجم نمونه گزینه Next را انتخاب می کنیم تا پنجره ذیل باز شود.



این پنجره مشخص می کند هر یک از متغیرها با چه نامی به نرم افزار وارد می شوند . کاربر می تواند در ستون Result Variable Name نام متغیر را به نام دلخواه خود تغییر دهد . همچنین کاربر می تواند در ستون Recode to Numeric نوع متغیر را با توجه به ستون Data Type از متغیر string به متغیر Numeric با انتخاب گزینه تغییر دهد برای نمونه در بالا متغیر Gender با نوع String می باشد که با انتخاب گزینه در نرم افزار spss به صورت فرمت Numeric وارد می شود. بعد از تغییرات مورد نظر دکمه Next را انتخاب می کنیم تا پنجره ذیل ظاهر گردد.



در این قسمت دستور مورد نظر برای انتقال بیان می شود. کاربر با زدن دکمه Finish داده ها را به صفحه نرم افزار منتقل می نماید. در قسمت پائین صفحه عبارت Save query to file این امکان را به کاربر می دهد که با مشخص کردن یک نام و یک مسیر کلیه مراحل را در مکانی ذخیره نماید و در صورتی که دفعات بعد نیازمند انتقال داده های به روز شده فایل Demo با همین فرمت و متغیرها از نرم افزار اکسل بود فایل دیگر لازم نباشد تا مسیر های بالا را تکرار نماید کافی است. در صورتی که کاربر بخواهد در این گزینه مسیر را ذخیره نماید می تواند در دفعات بعدی با استفاده از دستور Run Query مسیر را به صورت ذیل برای نرم افزار مشخص نماید. تا داده ها به سرعت از فایل مورد نظر به نرم افزار Spss منتقل شوند.



با انتخاب گزینه فوق پنجره ایی باز می شود و مسیر فایل اجرایی را سوال می کند. در صورتی که مسیر

فایل اجرایی حاوی مراحل انتقال فایل را مشخص کنیم اطلاعات بـه سـرعت بـه محـیط نـرم افـزار spss منتقل می شود.

هر گاه که لازم باشد تا در فایل اجرایی مشخص شده تغییری شامل تغییر در محل فایل مبـداء و یـا نـوع متغیرهای انتخابی و همچنین نام فایل و ... داده شـود لازم اسـت از طریـق گزینـه Edit Query مسـیر شخص شود تا مراحل مختلف را در اختیار کاربر قرار دهد و امکان تغییر هر مرحله وجود داشته باشد.

- ذخیره کردن فایل جاری (از طریق زیرمنو Save):

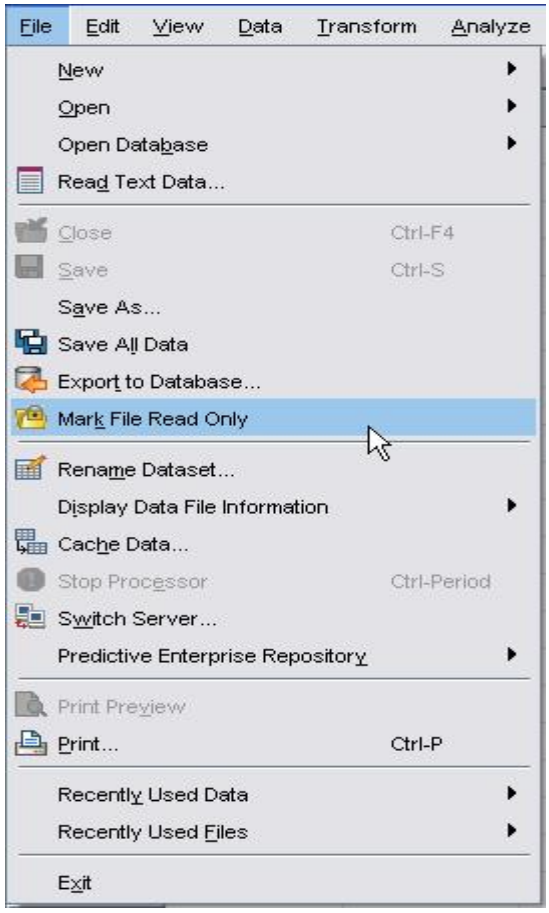
این امکان را به کاربر می‌دهد تا فایل جاری را با تغییراتی که بر روی آن انجام داده است در همان مسیری که فایل را فراخوانده است دوباره ذخیره نماید.

-ذخیره کردن فایل در مسیر دیگر(ازطریق زیر منو Save as):

این امکان را به کاربر می‌دهد که فایل جاری را در مسیری دیگر و بـا نـامی غیـر از نـامی کـه هنگـام فراخواندن فایل داشته است ذخیره نماید.

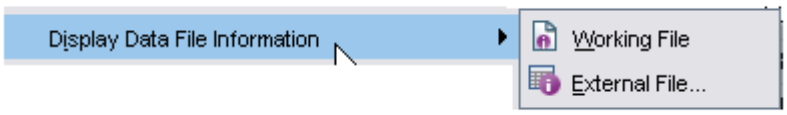
-تنها قابل خواندن کردن فایل(از طریق زیر منوMark File Read Only):

با انتخاب این گزینه کاربر فایل را به صورت یک فایل تنها قابل خواندن می‌نماید که تغییـرات بـر روی آن اعمال نمی‌شود و بعد از پایان تغییرات فایل مورد نظر بدون در نظر گرفتن تغییرات به حالت اولیه خود می‌باشد.



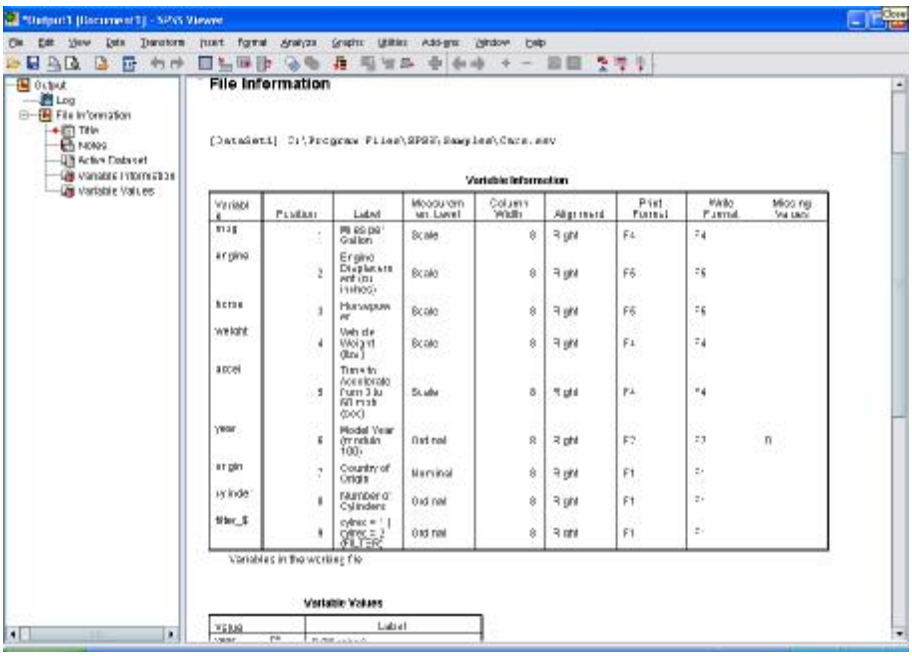
کاربر می‌تواند با انتخاب مجدد File>Mark File Read Write فایل را از حالت فقط قابـل خوانـدن بودن خارج نماید.

-گرفتن اطلاعات فایلهای داده‌ایی(از طریق زیر منو Display Data File Information):



هر گاه کاربر بخواهد در مورد فایلهای داده‌ایی اطلاعاتی شامل تعداد متغیرها، نوع متغیرها، زمـان ورود داده‌ها و ... کسب نماید می‌تواند از این گزینه استفاده نماید اگر می‌خواهد در مورد فایل جاری که در حال حاضر باز است اطلاعات کسب نماید گزینهٔ Working File... و در صورتی که می‌خواهد در مـورد یـک فایل خارجی اطلاعات کسـب نمایـد گزینـهٔ External File... را کلیـک مـی‌نمایـد تـا پنجـرهٔ محـاورهٔ

Display Data Info باز شود، مسیر فایل خارجی را مشخص نموده و Open می‌نماید آنگاه نتایج اطلاعات متغیرها در پنجرهٔ خروجی مانند شکل ذیل برای فایل جاری در خروجی ظاهر می‌شود.



- فشرده کردن داده‌ها(از طریق زیر منو Cache Data):

با انتخاب این گزینه فرمت داده‌ها در قالب فرمت سیستمی در می‌آید و کامپیوتر سریع‌تر تجزیه و تحلیل بر روی داده‌ها را انجام می‌دهد. استفاده از این آیتم در تجزیه و تحلیل داده‌ها با حجم بالا توصیه می‌شود.

-توقف فرآیند تجزیه و تحلیل(از طریق زیر منوStop Processor):

هر گاه کاربر دستوری را در نرم‌افزار اجرا نماید و به علت زمان زیاد اجرای دستور از اجرای آن منصرف گردد می‌تواند با زدن این آیتم فرآیند تجزیه و تحلیل را متوقف نماید.

-رؤیت داده‌ها قبل از چاپ(از طریق زیر منو Print Preview):

کاربر می‌تواند قبل از چاپ داده‌ها و اطلاعات در کلیه پنجره‌های نرم‌افزار با انتخاب این آیتم یک نما از اطلاعات را قبل از چاپ بر روی اندازه‌های کاغذ که مشخص می‌نماید ملاحظه کند.

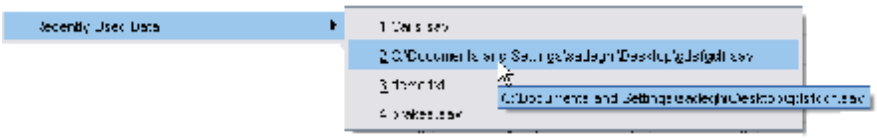
- چاپ اطلاعات پنجره‌ها(از طریق زیر منو Print):

با انتخاب این آیتم پنجرهٔ محاورهٔ Print باز می‌شود و کاربر می‌تواند با انتخاب گزینه‌های مورد نظر آن

امکان چاپ همه یا قسمتی از اطلاعات را در پنجره‌های مختلف SPSS فراهم کند.

-فراخواندن آخرین فایلهای داده‌ایی (از طریق زیر منو Recently Used Data):

کاربر با انتخاب این گزینه می‌تواند آخرین فایلهای داده‌ایی باز شده با پسوند sav.* را که حاوی داده‌های قابل خواندن توسط پنجرهٔ Data View است مجدداً فرابخواند و لزومی به استفاده از آیتم Open برای فراخواندن آنها نیست.

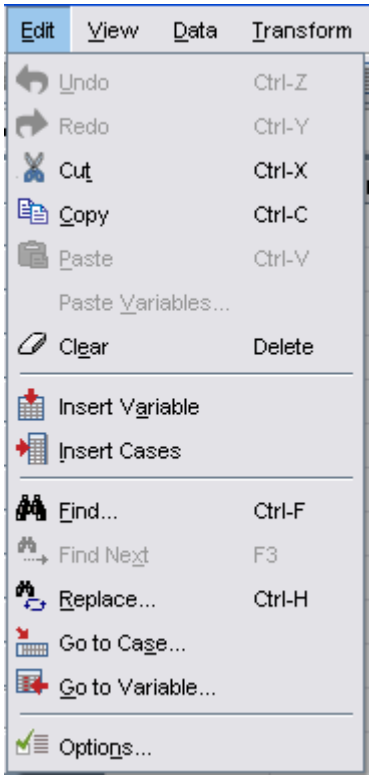


-فراخواندن آخرین فایل‌های غیرداده‌ایی (از طریق زیر منو Recently Used Files):

کاربر با انتخاب این آیتم می‌تواند آخرین فایل‌های غیر داده‌ایی باز شده فایل‌های برنامه ، خروجی و Script) را مجدداً فرابخواند.

- خروج از برنامه spss(از طریق زیر منوExit):

در صورت کلیک بر روی آیتم Exit کلیه پنجره‌های موجود نرم‌افزار SPSS بسته شده و کاربرد از نرم‌افزار SPSS خارج می‌گردد.



3-4- امکانات منو Edit :

همانطور که قبل‌تر گفته شد در این منو امکان ویرایش داده‌ها نظیر برش، کپی، حذف، انتخاب بخشی از فایل، جستجو، جایگذاری و تنظیم کلیات ویرایشگر SPSS موجود می‌باشد در ذیل نحوهٔ کار هر یک از آیتم‌ها بیان شده است.

- بازگشت(از طریق زیر منوUndo):

هنگامی که کاربر فرمانی را روی داده ها اعمال کرده که باعث تغییر داده ها شده است می تواند با انتخاب این زیر منو داده ها را به حالت قبل از اعمال تغییر برگرداند .به طور خلاصه این زیر منو آخرین تغییرات اعمال شده روی داده ها را خنثی می‌نماید.

- خنثی‌سازی Undo(از طریق زیر منوRedo) :

تغییرات انجام گرفته توسط Undo را به حالت قبل برمی‌گرداند. به این معنی که نتیجه فرمانی را که توسط Undo خنثی شده بود دوباره اجرا می نماید.

- بریدن داده ها (از طریق زیر منو Cut):

با استفاده از این دستور می‌توان قسمتی از داده‌ها را از یک قسمت Data View انتخاب کرد و در محل دیگری جایگزین کرد. ذکر این نکته ضروری است که داده‌ها در محیط اولیه از بین می‌روند. البته در هر پنجره از نرم افزار هم با انتخاب این زیر منو می توان قسمتی از محتویات موجود در پنجره را که می توانند خروجی ها یا برنامه ها باشند از طریق این زیر منو به محلی دیگر در همان پنجره منتقل نمود.

- کپی کردن محتویات یک پنجره(از طریق زیر منو Copy):

با استفاده از این دستور می‌توان قسمتی از داده‌ها را از یک قسمت Data View انتخاب کرد و در محل دیگری کپی کرد. ذکر این نکته ضروری است که داده‌ها در محیط اولیه از بین نمی‌روند.

- جایگزینی محتویات انتخاب شده(از طریق زیر منو Paste) :

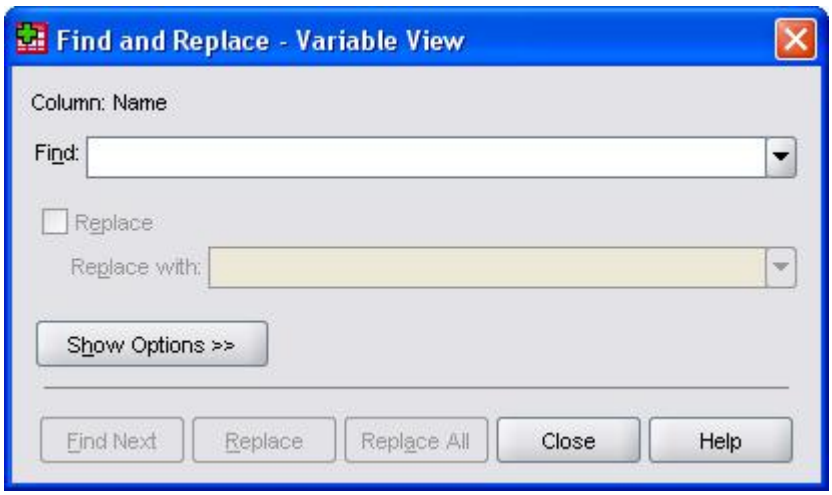
با این دستور می‌توان داده‌های Cut یا Copy شده را در محل دیگر مورد نظر از طریق این دستور جایگزین نمود. اگر دستور Cut را انتخاب کرده باشید داده ها در محلی دیگر که دستور Paste را اجرا می نمایید کپی شده و در محل اول از بین می روند و در اصطلاح با این اقدام تغییر مکان می دهند اما با انتخاب Copy و Paste داده ها علاوه بر حضور در مکان اول در محیط دیگری که انتخاب کرده اید نیز قرار می گیرند.

- حذف محتویات انتخاب شده (از طریق زیر منو Clear):

برای حذف قسمتی از داده‌ها از آن استفاده می‌شود. ابتدا داده‌هایی را که می‌خواهید حذف کنید انتخاب کنید آنگاه دکمه Clear یا دکمه Del را بر روی صفحه کلید را فشار دهید آن داده‌ها حذف شده و به جای آنها . یعنی دادهٔ گم شده سیستمی جایگزین می‌شود. هر گاه بخواهید یک متغیر را حذف نمائید می‌توانید در صفحهٔ Data View روی نام متغیر در بالای صفحه کلیک کنید تا کل ستون متغیر مورد نظر را انتخاب نماید و آنگاه Clear یا دکمه Del را فشار دهید، همچنین برای حذف یک سطر از داده‌ها می‌توانید روی یک شماره سطر مورد نظر در پنجرهٔ Data View کلیک کنید تا انتخاب شود و سپس Clear یا دکمه Del را فشار دهید سپس آن سطر حذف خواهد شد.

- جستجو(از طریق زیر منو Find):

برای جستجوی یک مقدار خاص در یک متغیر بکار می‌رود. ابتدا نشانگر موس را درون یکی از خانه‌های متغیر مورد نظر قرار دهید و سپس مقدار مورد نظر را در قسمت Find What وارد نمائید و دکمهٔ Find Next را بزنید، نشانگر به اولین عدد مورد نظر شما که در صورت وجود منتقل می‌شود.



- اضافه کردن یک متغیر جدید (از طریق زیر منو Insert Variable):

با انتخاب یک متغیر در پنجره Data view و انتخاب این گزینه ، نرم افزار یک متغیر قبل از متغیر انتخابی در لیست متغیرها ایجاد می نماید که خالی از اطلاعات است و کاربر می تواند یک متغیر به

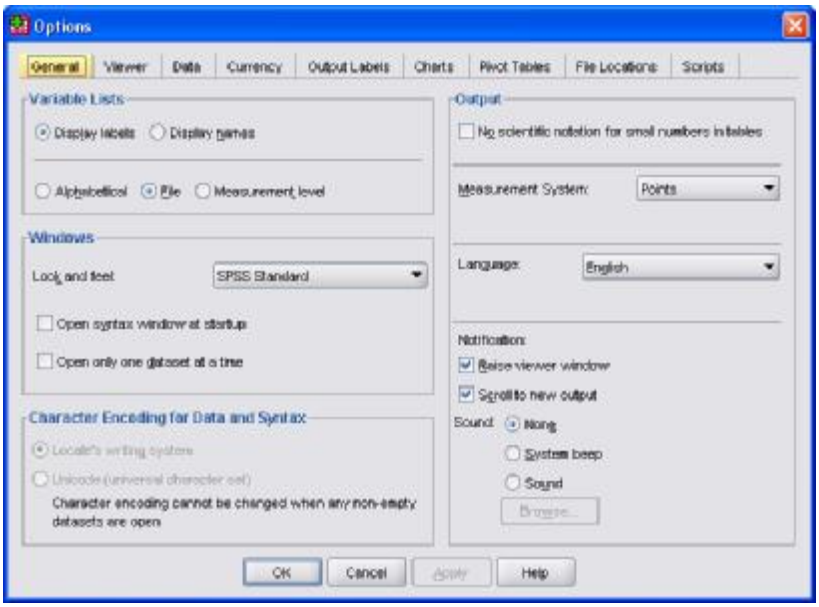
مجموعه متغیرهای موجود اضافه نماید.

- اضافه کردن یک سطرجدید (از طریق زیر منو Insert cases):

با انتخاب یک سطر در پنجره Data view و انتخاب این گزینه ، نرم افزار یک سطر قبل از سطر انتخابی در لیست متغیرها ایجاد می نماید که خالی از اطلاعات است و کاربر می تواند یک سطر به مجموعه داده ه های موجود اضافه می‌نماید. البته ذکر این نکته ضروری است که ترتیب قرار گرفتن سطرها در نرم افزار اهمیتی ندارد.

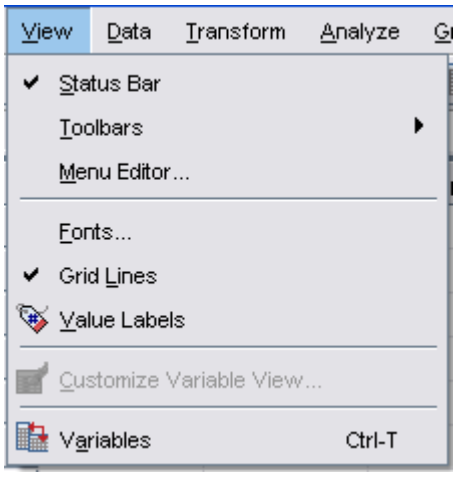
- تنظیمات پنجره‌ها در SPSS(از طریق زیر منو Options) :

با استفاده از این آیتم کاربر می‌تواند تنظیمات کلیه پنجره‌های موجود در SPSS را کنترل نماید.



برای مثال در شکل بالا کاربر می‌تواند زبان نرم‌افزار را از English تغییر دهد یـا در قسـمت Charts می‌تواند فونت نمودارها، رنگ و دیگر خصوصیات آنها را تغییر دهد و یا در قسـمت Pivot Tables کـاربر می‌تواند از قسمت Table look هر نوع جدولی را که قصد دارد جداول در قالـب آن در خروجـی‌هـا بـه

نمایش درآیند انتخاب و با دکمهٔ OK تأئید نماید.

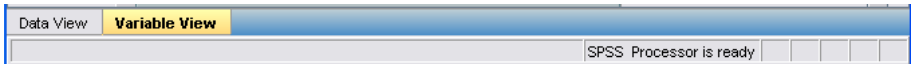


3-5- امکانات منو View :

ابزارهای موجود در این منو برای ویرایش شکل ظاهری پنجره‌های نرم‌افزار مورد استفاده قرار می‌گیرند از جمله ابزارهای آن می‌توان به کنترل‌کننده میله وضعیت، جعبهٔ ابزار، کنترل خطوط و فونت اشاره کرد.

- کنترل میله وضعیت(از طریق زیر منو Statue Bar):

این میله که در پائین تمامی صفحات SPSS وجود دارد و به صورت شکل ذیل است نشان می‌دهد که پردازشگر نرم‌افزار در حال حاضر مشغول به انجام چه کاری است.



در صورتی که کلمهٔ SPSS Processor is ready در آن نوشته شده باشد به این معنی است که پردازشگر در حال هیچ کاری نیست، اگر **ü** کنار این گزینه را برداریم آنگاه این میله از پائین صفحات حذف می‌شود.

-کنترل جعبه ابزارها(از طریق زیر منو Toolbars):

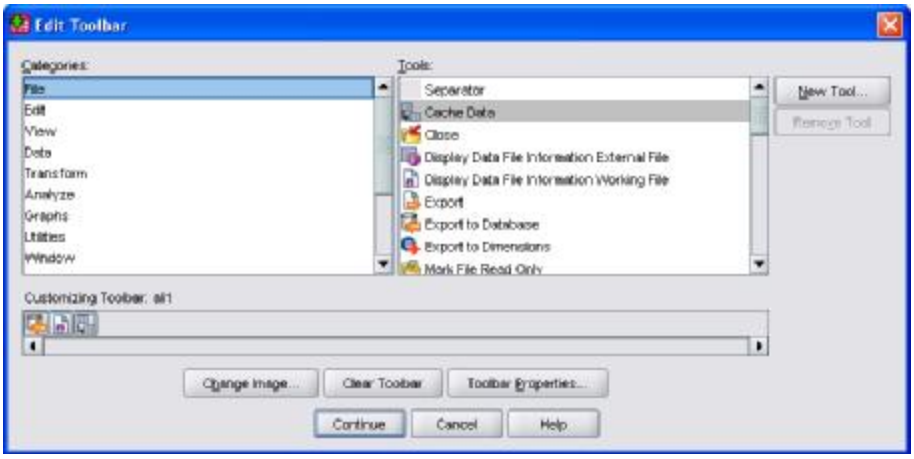
جعبهٔ ابزارها، مانند شکل ذیل تعدادی از زیرآیتم‌ها هستند که به علت کاربرد زیاد آنها در قسمت اصلی صفحه و در جعبهٔ ابزار به نمایش در می‌آیند.

هر پنجره در نرم‌افزار SPSS دارای یک جعبهٔ ابزار با همین نام به صورت مقدماتی است اما کاربر می‌تواند با زدن آیتم Toolbars و باز کردن پنجرهٔ محاورهٔ Show Toolbars هر تعداد جعبهٔ ابزار همراه با آیتم‌های هر کدام را به هر یک از پنجره‌ها اضافه نماید.

برای به وجود آوردن یک جعبه ابزار جدید روی دکمه New Toolbar کلیک کنید تا صفحهٔ Toolbar Properties باز شود، بعد از اینکه یک نام برای جعبهٔ ابزار خود انتخاب نمودید و در قسمت Display on the following windows مشخص می کنید که جعبهٔ ابزار جدید در کدام پنجره‌ها به نمایش در آید.

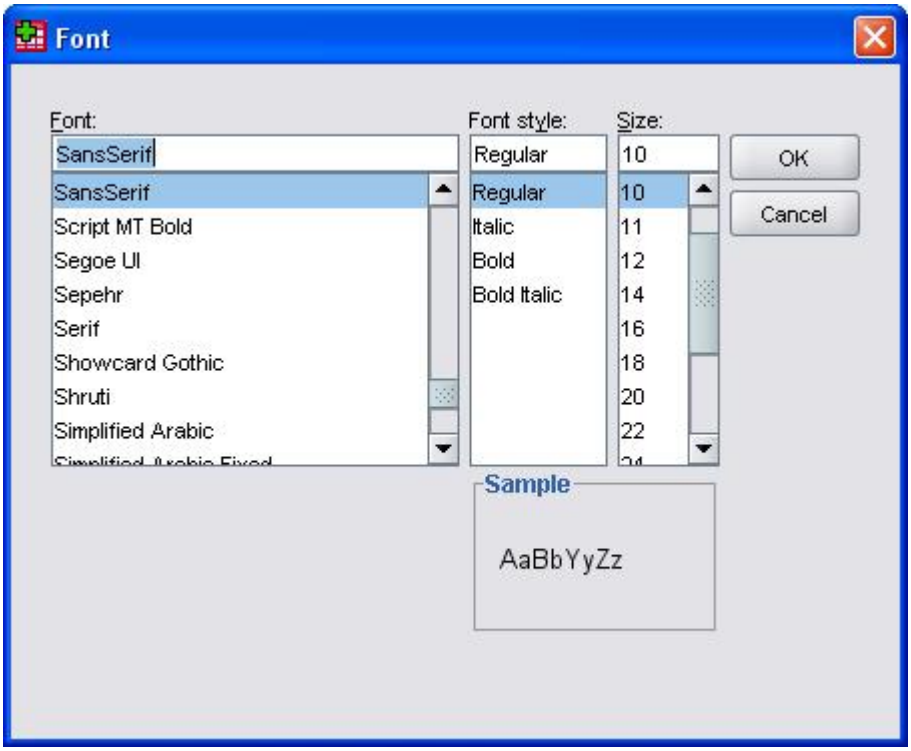
با زدن دکمه Edit به پنجرهٔ EDdit Toolbar بروید و با انتخاب Items مورد نظر خود و قرار دادن هر یک در قسمت Customizing Toolbar جعبهٔ ابزار مورد نظر را بسازید و با انتخاب دکمه Countinue, و دکمهٔ OK تأئید کنید، در نتیجه جعبهٔ ابزار جدید به زیر جعبهٔ ابزار قدیمی در هر یک از پنجره‌ها اضافه می‌شود.

با انتخاب Large Buttons در جعبهٔ محاورهٔ Show Toolbars می‌توانید اندازهٔ هر یک از آیتم‌های جعبهٔ ابزار را به سایز بزرگ در آورید.



- تغییرفونت‌ها (از طریق زیر منو Fonts):

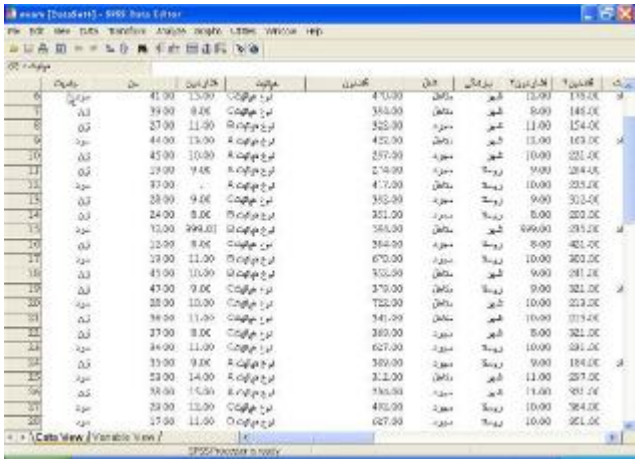
با انتخاب آیتم مورد نظر امکان تغییر فونت و اندازهٔ داده‌ها در پنجرهٔ Data View فراهم می‌آید. به شکل ذیل توجه نمائید:



با انتخاب هر یک از انواع فونت ها و اندازه آنها می توان اندازه داده ها را در هر یک از پنجره ها تغییر داد.

- کنترل خطوط وسط خانه‌ها(از طریق زیر منو Grid Lines):

با برداشتن **ü** مربوطه خطوط وسط خانه‌ها مانند شکل ذیل حذف می‌شود.



همانطور که ملاحظه می شود با اجرای این فرمان خطوط وسط خانه ها از بین می رود و می تـوان در صورتی که نیاز به چاپ داده ها است بدون وجود خطوط داده ها را چاپ نمود.

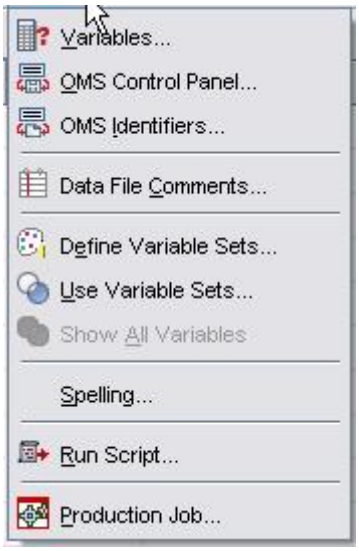
- بر چسب مقادیر(از طریق زیر منو Value Labels):

با قرار دادن **ü** مربوط به آن به جای اعداد، برچسب آنها در پنجرۀ Variable View جایگزین می‌شود.

برای مثال اگرValue Label **ü** باشد در ستون جنسیت کلمات مـرد و زن و اگـر Value Label باشـد اعداد 1 و 2 وجود دارند.

- نمایش پنجره کنترل خصوصیات متغیرها(از طریق زیر منو Variables):

با انتخاب این آیتم به پنجـرۀ کنتـرل کننـده خصوصـیات متغیرهـا یعنـی Variable View منتقـل می‌شویم به جای انتخاب این آیتم می‌توان از دکمه‌های t + Ctrl نیز در صفحۀ کلید استفاده نمود.



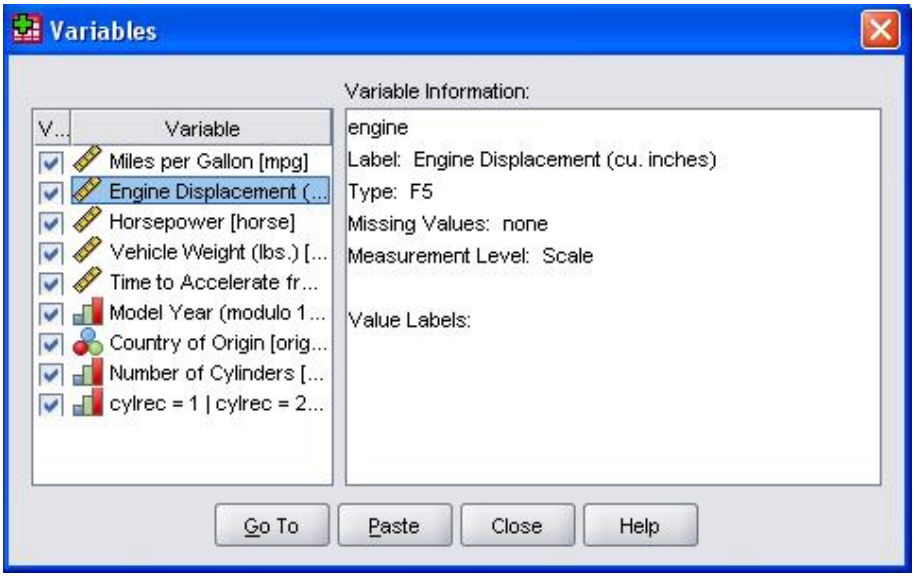
3-6- امکانات منو Utilities :

این منو و زیرمنوهای آن امکانـات جـانبی را در اختیـار کـاربر قـرار می‌دهد. از طریق زیر منوهای این منو می توان نسبت به ایجاد مجموعه ای از متغیرها، اضافه کردن منوهای جدید به پنجره هـا، اجـرای برنامـه

های ویرایشی و ... اقدام نمود. در ذیل به ابزارها و آیتم‌های مهم آن اشاره گردیده است.

- معرفی خصوصیات متغیر(از طریق زیر منو Variables):

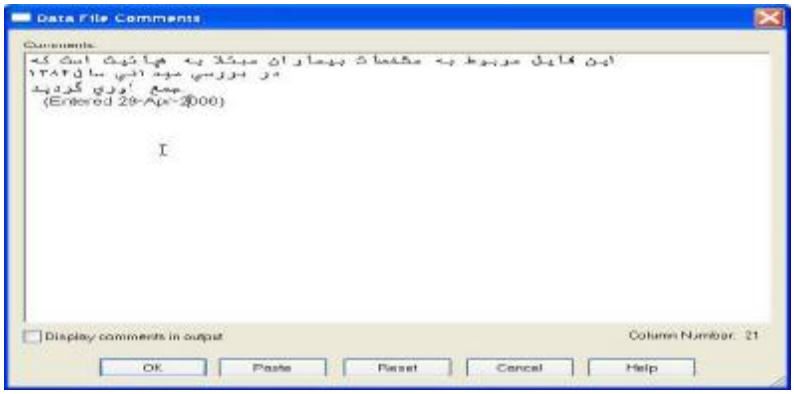
با انتخاب این آیتم پنجره‌ای مانند مقابل باز شده که در قسمت سمت چپ آن لیست تمام متغیرهای موجود در Data View قرار دارد:



با انتخاب هر کدام از آنها، خصوصیات هر کدام از متغیرها شامل نام متغیر، برچسب، نوع، وجود یا عدم وجود داده گم شده، واحد اندازه‌گیری و برچسب مقادیر نمایش داده می‌شود.

- توضیح در مورد فایل داده‌ها(از طریق زیر منو Data File Comments):

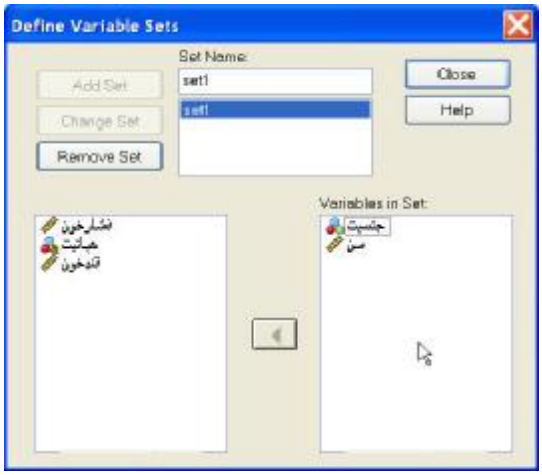
با انتخاب این آیتم یک پنجرۀ متنی با همین نام باز می‌گردد که کاربر می‌تواند کلیه توضیحات متنی خود را در این محیط بنویسید و Ok نماید، این توضیحات همواره به عنوان پیوست در همین محیط محفوظ می‌ماند و کاربر و یا کاربران دیگر در صورتی که به این محیط مراجعه کنند توضیحات را می‌توانند ملاحظه و حتی ویرایش نمایند.



با انتخاب Display Comment in Output توضیحات در خروجی نیز نمایش داده می‌شوند.

- تعریف مجموعه ایی ازمتغیرها(از طریق زیر منو Define Sets):

هر گاه کاربر بخواهد تنها در کلیه پنجره‌های محاوره تعدادی از متغیرها نمایش داده شوند و دیگر متغیرها نمایش داده نشوند باید مجموعه‌ایی از متغیرهایی را که مورد نیاز است در قالب یک مجموعه یا چندین مجموعه انتخاب نماید برای این منظور آیتم Define Sets را انتخاب می کنیم تا پنجرهٔ ذیل باز شود.



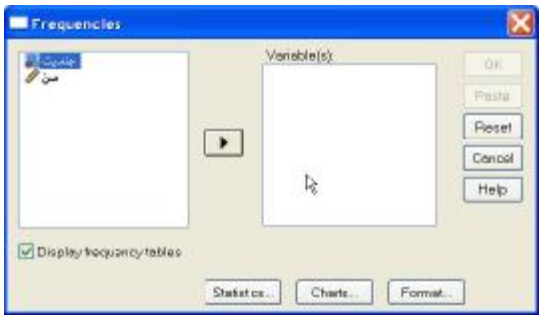
برای مثال در قسمت Set Name یک نام مانند Set1 را می‌نویسیم و از قسمت پائین دو متغیر جنسیت و سن را به قسمت Variable in Set وارد می‌نمائیم و دکمهٔ Add Set را می‌زنیم. با این اقدام یک مجموعه Set1 درست شده است که حاوی دو متغیر یاد شده است. به همین طریق می‌توان هر چند

مجموعه از متغیرها را که مورد استفاده است ایجاد کرد.

- استفاده از مجموعه‌های تعریف شده(از طریق زیر منو Use Sets):

با استفاده از این آیتم پنجره‌ایی به نام Use Sets باز می‌شود، نرم‌افزار SPSS به صورت خودکار دو مجموعه از متغیرها را که با نامهای ALL VARIABLES و NEW VARIABLES به معنی کل متغیرها و متغیرهای تازه ایجاد شده است مورد استفاده قرار می‌دهد. در صورتی که مجموعهٔ Set1 را انتخاب کنیم و جایگزین دو نوع مجموعهٔ بالا نمائیم و ok نمائیم هر جعبهٔ محاوره‌ایی را که در نرم‌افزار SPSS باز نمائیم تنها شامل متغیرهای مجموعهٔ set1 خواهد بود و دیگر متغیرها قابل رؤیت نیستند.

برای مثال به شکل پنجرهٔ محاورهٔ Frequances در ذیل توجه نمائید.



در قسمت سمت چپ لیست متغیرها تنها دو متغیری را که در set1 تعریف نموده بودیم و وارد قسمت

Set in use نمودیم مورد استفاده قرار داده است و دیگر متغیرها را در تجزیه و تحلیل ها دخالت نمی دهد.

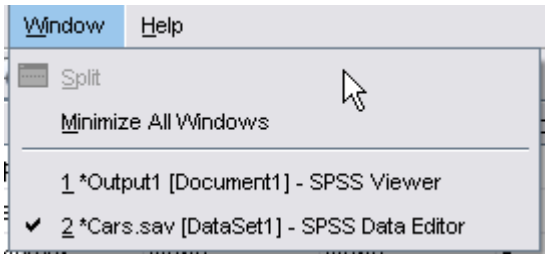
- اجرای Scriptها (از طریق زیر منو Run Script):

با انتخاب این آیتم می‌توان Scriptها را انتخاب و به طور مستقیم از صفحهٔ Data View آنها را اجرا نمود.

- ویرایش منوها (از طریق زیر منو Menu Editor):

با استفاده از ابزارهای این آیتم امکان اضافه کردن منو و زیر آیتم‌های مختلف به منوهای ایجاد شده در کلیه پنجره‌ها که در قسمت Apply to انتخاب می‌شوند از طریق دو دکمهٔ Insert Menu وInsert Item برای کاربر فراهم می‌آید.

3-7- امکانات منو Windows :



امکان کنترل پنجره‌ها و وضعیت هر یک را به کاربر نمایش می‌دهد. با استفاده از گزینهٔ Minimize All windows کلیهٔ پنجره‌های موجود SPSS را کوچک می‌نمائیم، در این منو پنجره‌های باز نرم‌افزار SPSS و پنجره‌های فعال نرم‌افزار (با استفاده از آیکون ü) نمایش داده می‌شوند.

همچنین با استفاده از آیتم Split می‌توان پنجرهٔ Data View را به دو پنجرهٔ مجزا مانند حالت تکثیر سلولی به صورت سطری یا ستونی تقسیم کرد به طوری که هر یک حاوی کلیه اطلاعات باشند.

و برای بازگشت به حالت اول گزینهٔ Remove Split را مجدداً از منو Window انتخاب می‌نمائیم تا پنجرهٔ Data View به حالت اولیه برگردد.

خلاصه مباحث مطرح شده :

در این فصل کاربر نحوه ورود داده ها به نرم افزار spss و همچنین تعریف خصوصیات متغیرها شامل وارد کردن نام تغیر، نوع متغیر، تعداد اعداد و تعداد اعشار متغیر، برچسب متغیر، نحوه مشخص کردن داده های گم شده در متغیر، مشخص کردن نوع متغیر، جانمایی متغیر و طول متغیر را فرا می گیرد.

علاوه بر مشخص نمودن نحوه ورود متغیرها و تعریف خصوصیات آنها انواع منوهای مختلف پنجره اصلی data view نظیر File، Edit ، View، Data، Transform،Windows ، Utilitesو Help مورد بررسی قرار گرفت و همچنین تعدادی از زیر ابزارهای هر یک از این منوها نظیر ابزارهای ورود داده ها به محیط نرم افزار، ابزارهای تعریف مجموعه ها، ابزارهای تغییر کد متغیرها و همچنین ابزارهای تعریف خصوصیات هر یک از پنجره ها و تغییر فونت وشکل ظاهری داده ها مورد بررسی قرار گرفت.

تمرین:

1 .چگونگی ورود داده ها به پنجره Data View را شرح داده و به طور کامل انجام دهد.

2 .مکانات منو File را بیان کند و وظایف هر یک از زیر منو های آن را شرح دهد.

3. امکانات منو Edit را ذکر کرده و به طور کامل هر یک را توضیح دهد.

4. ابزارهای موجود در منو View را نام ببرد و کار هر یک را توضیح دهد.

5. امکانات موجود در Utilites را بیان کرده و وظایف هر یک از زیر منو های آن را به طور کـار عملی نشان دهد.

6. امکانات منو Windows و کاربرد هر یک را شرح دهد.

فصل چهارم

آمار توصیفی در نرم‌افزار SPSS

اهداف آموزشی

فراگیر پس از مطالعه این فصل باید بتواند:

1- ابزارهای مورد نیاز جهت تلخیص داده ها در نرم افزار را بشناسد.

2- مراحل انجام آمار توصیفی برای متغیرهای گسسته رابا ذکر مثال شرح دهد.

3- مراحل انجام آمار توصیفی برای متغرهای پیوسته را با ذکر مثال توضیح دهد.

4ـ1ـ مقدمه:

همانطور که در فصل اول بیان شد منظور از آمار توصیفی یعنی تخلیص اطلاعات متغیرها در قالب نمودارها، جداول و پارامترهای آماری ، به عبارت ساده‌تر شناخت یک بعدی متغیرهای جامعهٔ مورد بررسی را آمار توصیفی گویند.

در این بخش نحوه گرفتن آمار توصیفی برای دو نوع متغیر گسسته و پیوسته را با توجه به ابزارهای مورد نیاز هر کدام را توضیح می‌دهیم. ذکر این نکته در مقدمه این بحث ضروری به نظر می رسد که آمار توصیفی مقدمه آمار استنباطی می باشد یعنی کاربر قبل از اینکه تاثیر متغیرها را بر روی یکدیگر مورد ارزیابی قرار دهد بهتر است نسبت به خصویات هر یک از متغیرها اشراف کافی داشته باشد که برای این

منظور انجام آمار توصیفی ضروری به نظر می رسد.

2-4- آمار توصیفی :

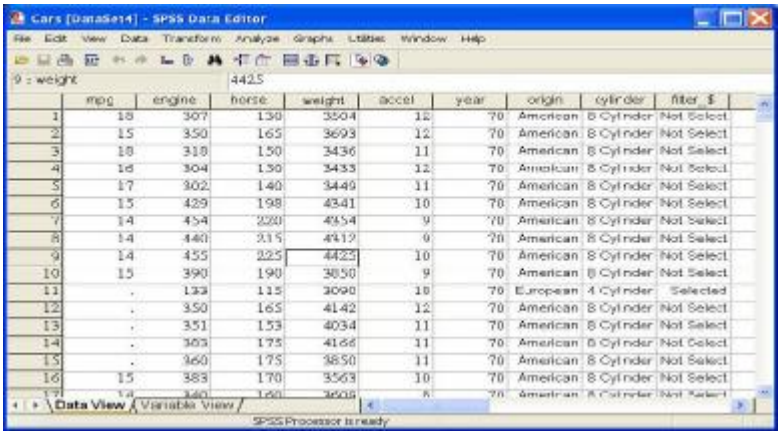
همانطور که در مقدمه بیان شد آمار توصیفی مقدمه آمار تحلیلی و آنالیز داده ها می باشد و عدم انجام آمار توصیفی قبل از آمار تحلیلی در بعضی از موارد به طولانی شدن پروسه تحلیل و عدم دستیابی به نتایج دقیق با توجه به بنود شناخت اولیه منجر می شود.

برای انجام آمار توصیفی از داده‌های نرم‌افزار SPSS که بر روی لوح فشرده نیز موجود است استفاده می‌کنیم. ابتدا فایل Cars.sav را به آدرس ذیل از طریق منو File>Open>Data به پنجرهٔ Data View فرابخوانید:

G:\Samples\cars.sav

البته ممکن است شما لوح فشرده را بر روی درایو دیگری غیر از G داشته باشید که باید به جای درایو G آن را جایگزین نمائید.

بعد از فراخواندن فایل cars اطلاعات مربوط به 406 ماشین که متغیرهای ذیل برای آنها جمع‌آوری شده است ظاهر می‌شود:



mpq: متغیر میزان مسافت طی شده بر حسب مایل در ازای مصرف یک گالن بنزین.

Engine: ظرفیت موتور ماشین.

Horse: قدرت ماشین بر حسب اسب بخار.

Weight: وزن ماشین.

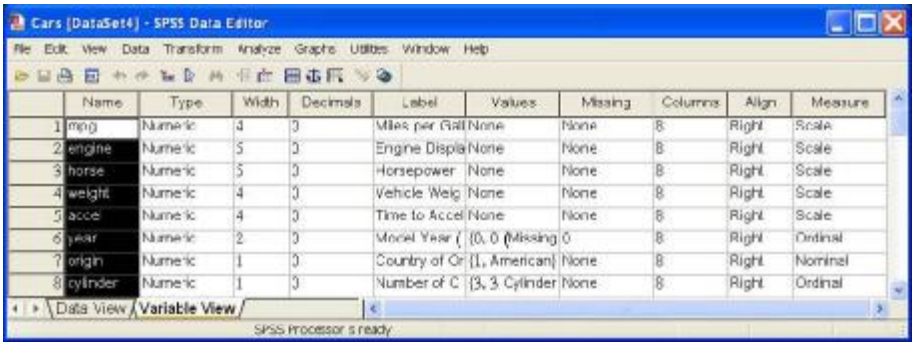
Accel: شتاب ماشین.

Year: سال تولید ماشین.

Origin: منطقهٔ تولید ماشین (1=آمریکا، 2=اروپا، 3= ژاپن).

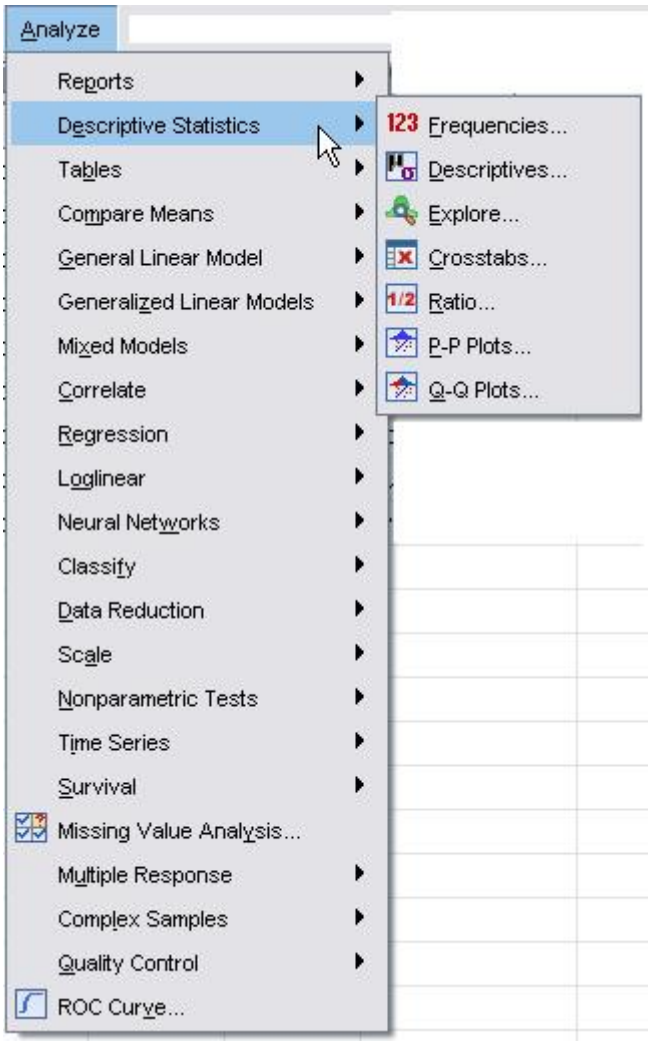
Cylindr: تعداد سیلندر.

در شکل ذیل پنجرهٔ Variable View که خصوصیات این متغیرها را نشان می‌دهد بیان شده است.



تخلـیص داده‌هـا در نـرم‌افـزار SPSS از طریـق منـو Analysis و آیـتمDescriptiveStatistic وزیرآیتم‌های Frequances، Descriptive و Expolre صورت می‌گیرد، بـه طـور کلـی بـا ایـن 3 ابـزار

می‌توان یک شناخت توصیفی نسبت به متغیرها ایجاد کرد، در این بخش تخلیص متغیرها را بـه دو گـروه تخلیص متغیرهای گسسته و تخلیص متغیرهای پیوسته تقسیم می‌کنیم و به شرح هر کدام می‌پردازیم.



4-2ـ آمـار توصـیفی بـرای متغیرهـای گسسته:

یک متغیر گسسته متغیـری اسـت کـه از سطوح مختلف تشکیل شده است. برای مثـال متغیـر جنسـیت دارای 2 سـطح مـرد و زن، متغیر محـل زنـدگی دارای دو سـطح شـهر و روستا، متغیر وضعیت تأهـل دارای دو سـطح مجرد و متأهل، متغیر سال تولیـد ماشـین در فایـل جـاری دارای سـطوح مختلـف، متغیـر منطقهٔ تولید دارای 3 سطح، آمریکـا، اروپـا و ژاپـن و متغیـر تعـداد سـیلندر یـک متغیـر گسسته دارای سطوح 3 تا 8 می‌باشد.

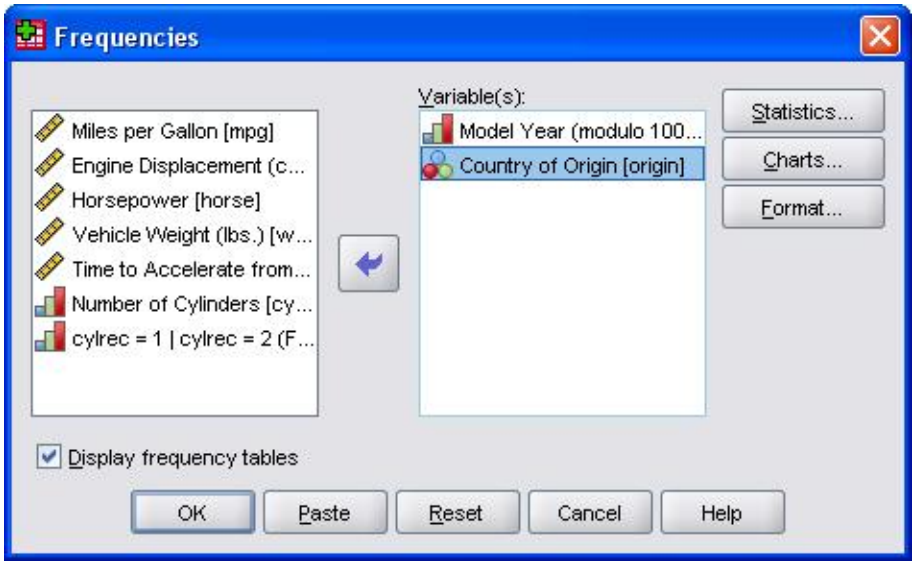
هر گاه صحبت از شناخت نسبت به یک متغیر گسسته می‌شود در حقیقت منظور شناخت هـر یـک از سطوح آن است و در صورتی که کاربر بتواند نسبت به سطوح آن شناخت ایجاد کند می‌تواند وضعیت کلی متغیر گسسته را تخلیص نماید. منظور از تخلیص اطلاعات سطوح یک متغیر در حقیقت بیان درصد و یـا تعداد عناصر هر یک از سطوح در قالب جداول و نمودارهای آماری است.

برای انجام آمـار توصـیفی بررسـی داده‌هـا بایـد از زیـر فرمـان Frequency در آیـتم Descriptive

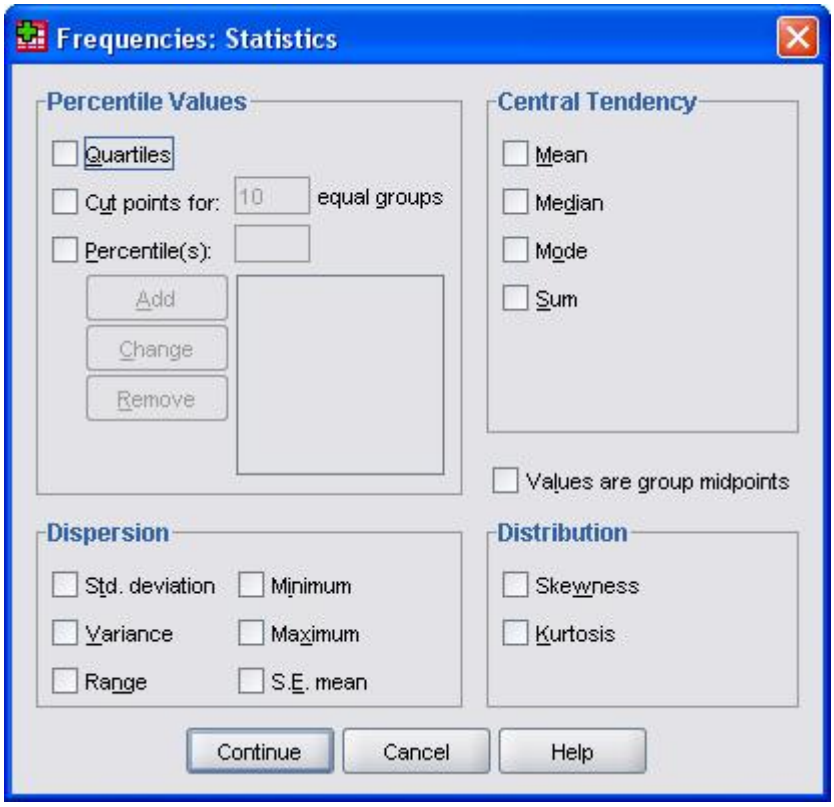
Statistics و منو Analysis مانند شکل بالا استفاده نمائید.

4ـ2ـ1 ـ فرمان Frequencies:

با اجرای این فرمان آمار توصیفی مربوط به متغیرها به صورت گروهی و یا تک به تک برای متغیرهای گسسته در این بخش محاسبه، ترسیم یا جدول‌بندی می‌شوند. جدول فراوانی، نمودارهای میله‌ای و آماره‌های تک‌متغیره به کمک این فرمان به نمایش در می‌آیند. با اجرای این فرمان جعبهٔ گفتگوی زیر باز می‌گردد:



فهرست متغیرهای فایل جاری (Cars.sav) در جعبهٔ سمت چپ مشاهده می‌شود. متغیرهایی که قصد اجرای فرمان بر روی آنها را داریم به جعبهٔ Variable (S) منتقل می‌کنیم. در صورت نیاز به جداول توزیع فراوانی متغیرهای انتخاب شده، گزینهٔ Display Frequencies Tables را فعال می‌کنیم. با فعال کردن کلید Statistics... می‌توان آماره‌های کمی را، که برای متغیرهای گسسته انتخابی محاسبه می‌شود تعیین کرد.



بخش Percentile Values در صورت فعال شدن مقادیر صدکها را محاسبه می‌نماید، با انتخاب Quartiles چارکهای متغیرها محاسبه می‌شود.

گزینه Cut Point For به ازای عدد تعیین شده در ردیف این گزینه داده‌ها را به چند قسمت مساوی تقسیم می‌کند. مثلاً با انتخاب n=5 داده‌ها به پنج بخش مساوی تقسیم می‌شوند. بدیهی است که برای تقسیم داده‌ها به پنج بخش به چهار عدد نیازمندیم که بین هر دو عدد $\frac{100}{5}\%$ یعنی 20% داده‌ها قرار گیرند. پس این چهار عدد باید صدک‌های بیستم، چهلم، شصتم و هشتادم داده‌ها باشند.

گزینهٔ Percentile(s) صدک خاصی از متغیرها را که مورد نیاز کاربر هستند معرفی می‌کند. با انتخاب این بخش بر روی این بخش باید شماره صدک مورد نظر را در جعبه همان ردیف وارد کرد و سپس کلید Add را فعال نمود. در صورت نیاز به تکرار این کار، عدد بعدی را وارد می‌کنیم، اگر قصد جابجایی شماره جدید با شماره قبلی را داشته باشید، کلید Change را فعال می‌کنیم. در غیر این صورت با فعال کردن

کلید Add شماره صدک اخیر به مجموعه شماره‌های قبلی اضافه می‌شود. در صورت نیاز به حذف شماره صدک خاص بر روی آن شمارهرا انتخاب کرده و کلید Remove را فعال می‌کنیم.

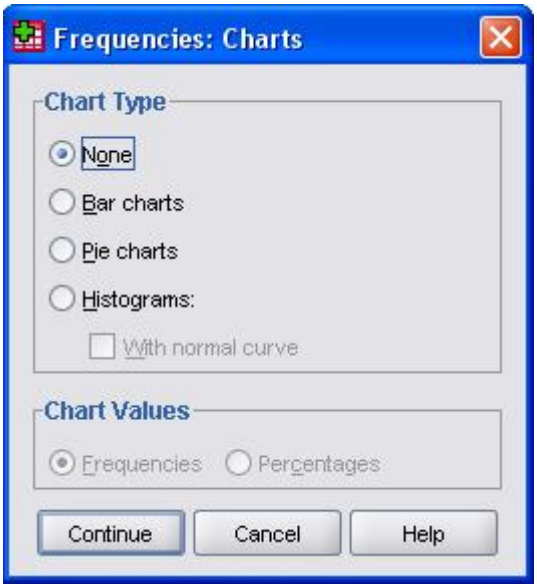
قسمت Central Tendency آماره های مرکزی را که برای متغیرها محاسبه می‌شود، تعیین می‌کند. آماره‌های مرکزی، کمیت‌هایی هستند که تمرکز داده‌ها را از جنبه‌های مختلف (مکانی، عددی، فراوانی) نمایش می‌دهند. کمیت‌های مرکزی در این بخش به ترتیب عبارتند از: میانگین (Mean)، میانه (Median)، مد (Mode) و جمع داده‌ها (SUM) .

در بخش Dispersion، معیارهای پراکندگی داده‌ها حول نقاط تمرکز را تعیین می‌کند. کمیت‌های پراکندگی در این قسمت به ترتیب عبارتند از انحراف معیار (Std.deviation)، واریانس (Variance)، برد (Range)، کمترین مقدار (Minimum)، بیشترین مقدار (Maximum) و انحراف معیار میانگین داده‌ها (S.E.mean).

قسمت Distribution تشخیص شکل کلی توزیع داده‌ها را به کمک محاسبه ضرایب چولگی (Skewness) و ضرایب کشیدگی (Kurtosis) ممکن می‌سازد.

داده‌های آماری می‌توانند داده‌های اولیه جمع‌آوری شده یا خلاصه داده‌ها در قالب یک گروه‌بندی باشند. اگر حالت اخیر در داده‌ها برقرار باشد، گزینهٔ Values are Group Midpoints را فعال می‌کنیم. این عمل سبب می‌شود کلیه محاسبات بخشهای قبل با فرض گروه‌بندی داده‌ها انجام شود.

با فعال نمودن کلید Charts... امکانات نموداری و کشیدن نمودارهای آماری در بخش آمار توصیفی در اختیار کاربر قرار می‌گیرد. جعبهٔ محاورهٔ زیر پس ازانتخاب این کلید باز می‌گردد.



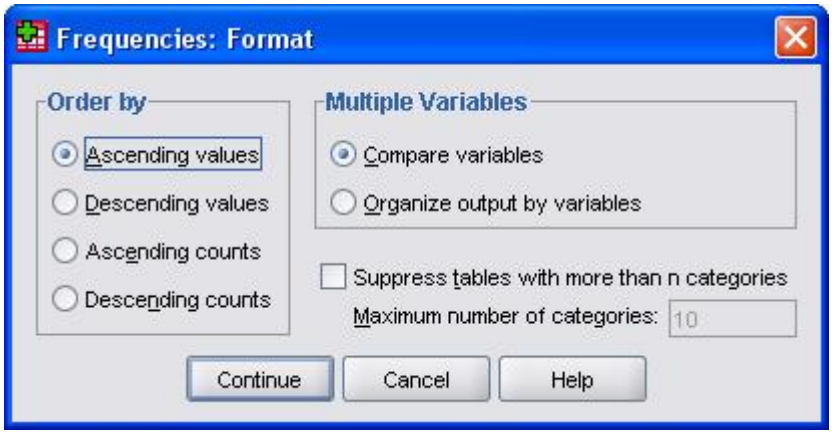
در قسمت Chart Type نوع نمودارها تعیین می‌شود. پیش‌فرض نرم‌افزار، حذف هر گونه نمودار است، که معادل با انتخاب گزینهٔ None است.

با انتخاب گزینهٔ Bar chart(s) نمودار میله‌ای و با انتخاب گزینه Pie Chart (s) نمودار دایره‌ایی و با انتخاب گزینهٔ Histogram(s) نمودار هیستوگرام داده‌های متغیرهای انتخابی در خروجی ترسیم می‌شود.

در صورت انتخاب نمودار هیستوگرام می‌توان منحنی توزیع نرمال را با انتخاب گزینه With Normal Curve هم‌زمان با هیستوگرام و بر روی آن ترسیم کرد. ترسیم هیستوگرام در تشخیص نرمال بودن توزیع داده‌های یک متغیر می‌تواند سودمند باشد.

بخش Axis Label Display تقسیم‌بندی محور عمودی نمودار میله‌ایی را نشان می‌دهد. این تقسیم‌بندی می‌تواند بر حسب فراوانی (Frequancies) یا درصد فراوانی هر یک از سطوح متغیر گسسته از کل داده‌ها (Percentages) باشد.

با فعال کردن کلید Format... می‌توان قالب کلی جداول و صفحه‌بندی خروجی‌ها را کنترل کرد. جعبه گفتگوی زیر پس از Click کردن بر روی این کلید ظاهر می‌شود.



بخش Order By نوع و ترتیب نمایش داده‌ها در جداول فراوانی را نمایش می‌دهد. این ترتیب می‌تواند بر حسب مقادیر هر یک از سطوح متغیر گسسته (Values) و یا فراوانی هر یک از سطوح (Counts) باشد.

ترتیب اعداد نیز می‌تواند افزایشی (Ascending) یا کاهش (Desending) باشد. بنابراین بر حسب نوع داده‌ها و ترتیب قرار گرفتن آنها چهارگزینه (2×2=4) برای جدول‌بندی داده‌ها وجود دارد، که انتخاب هر یک با فعال نمودن ردیف متناظر با آن گزینه ممکن می‌گردد.

در بخش Multiple Variables با انتخاب گزینه Compare Variables جداول به صورت فشرده چاپ شده و می‌توان مقادیر جداول را برای متغیرهای مختلف با هم مقایسه نمود و با انتخاب گزینهٔ Organize Output by Variables جداول فراوانی و شکل‌ها را برای هر متغیر به صورت جدا از هم ارائه می‌دهد و امکان مقایسه تطبیقی میان متغیرها وجود ندارد.

کاربر با انتخاب گزینهٔ Suppres tables with more than n Categories … ، سبب حذف نمایش جداولی می‌شود که تعداد آن را کاربر در قسمت Maximum number of categories: مشخص می‌نماید.

ذکر این نکته بسیار ضروری است که گزینه Format … تنها زمانی فعال است که جداول توزیع فراوانی در خروجی به نمایش در آیند.

4ـ2ـ2ـ نمونه‌ای از اجرای فرمان Frequencies و تحلیل خروجی آن:

در فایل جاری Cars.sav پنجرهٔ محاورهٔ Frequencies را باز می‌نمائیم. همانطور که در شکل ذیل ملاحظه می‌شود. متغیرها در جعبه‌های محاوره با برچسب خود نمایان می‌شوند. و در انتهای برچسب آنها نام متغیر آورده می‌شود برای مثال متغیر Country of Origin [Origin] که به صورت [نام] برچسب متغیر منطقه تولید، در پنجرهٔ محاورهٔ Frequencies آورده شده است.

ابتدا متغیرهای گسسته Origin وارد قسمت Variable(S) می‌نمائیم. هر چند می‌تواند تمامی متغیرهای گسسته را یکجا وارد این قسمت نمود.

سپس با استفاده از دکمهٔ Statistics... گزینه‌های Median، Mode، Quartiles، Std.deviation و Range را به دلخواه انتخاب می‌نمائیم. کاربر می‌تواند با توجه به نیاز خود هر یک از این کمیت‌ها را انتخاب نماید تا نرم‌افزار در خروجی برای او محاسبه کند. سپس دکمه Continue را انتخاب و به صفحهٔ Frequencies بر می‌گردیم.

حال به منظور کشیدن شکل‌های مورد نظر به قسمت Charts... رفته و Bar Charts و قسمت Frequencies را انتخاب می‌نمائیم. و Continue را می‌زنیم. با اجرای این عمل به نرم‌افزار دستور کشیدن یک نمودار میله‌ایی با مقدار فراوانی بر روی محور عددی داده می‌شود حال با بازگشت به پنجرهٔ Frequencies و انتخاب و تائید OK به منظور اجرای فرمان Frequencies نتایج آن در پنجرهٔ خروجی به صورت ذیل نمایش داده می‌شود.

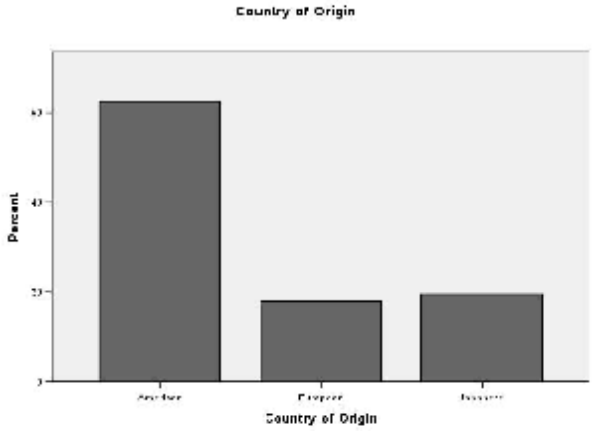
Statistics

Country of Origin

N	Valid	406
	Missing	1
Median		1.00
Mode		1
Std. Deviation		.796
Range		2
Percentiles	25	1.00
	50	1.00
	75	2.00

Country of Origin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	American	252	62.4	62.4	62.4
	European	79	19.6	19.6	82.0
	Japanese	73	18.1	18.1	100.0
	Total	404	100.0	100.0	
Missing	User-defined	1	.2		
	System	1	.2		
	Total	2	.5		
Total		406	100.0		



در جداول اول با نام Statistics تمامی کمیت‌هایی که در قسمت Statistics علامت‌دار کرده‌ایم محاسبه شده‌اند. برای مثال مقدار انحراف معیار متغیر منطقهٔ تولید برابر 0/798 و صدک اول برابر 1 و صدک سوم برابر 2 می‌باشد.

در جدول دوم با برچسب متغیر Origin، جدول فراوانی متغیر گسسته بیان شده است. در ستون اول ابتدا سطوح معتبر (Valid) که سه سطح هستند و سطوح گم‌شده (Missing) از دو نوع سیستمی (System Missing) و تعریف شده توسط کاربر بیان شده و در نهایت کل (Total) داده‌ها بیان شده است.

در ستون دوم فراوانی (Frequency) هر یک از سطوح چه داده‌های معتبر و چه داده‌های گم شده بیان گردیده است. برای مثال تعداد ماشین‌های آمریکایی در نمونه برابر 252 عدد، تعداد کل ماشین‌هایی که معتبر بوده‌اند و محل ساخت آنها مشخص است 404 عدد، تعداد داده‌های گم شده از نوع سیستمی

برابر 1 عدد و تعداد داده‌های از نوع تعریف شده کاربر با نام Miss برابر 1 عدد است. تعداد کل داده‌های گم شده برابر 2 عدد و تعداد کل داده‌ها (مجموع داده‌های معتبر و داده‌های گم شده) برابر 406 می‌باشد.

در ستون سوم درصد (Percent) هر یک از انواع داده‌های معتبر و داده‌های گم شده را از کل داده‌ها بیان کرده است برای مثال از کل ماشین‌ها 62/1 درصد آنها آمریکایی، و 0/5 درصد آنها محل تولید مشخص ندارند.

ستون سوم درصد هر یک از سطوح را از کل داده‌های معتبر نمایش می‌دهد به این معنی که داده‌های گم شده را نادیده در نظر می‌گیرد و پس از کل داده‌های معتبر (404 داده) درصد هر یک از سطوح را مشخص می‌سازد. برای مثال از میان 404 دادهٔ معتبر، 18/1 درصد آنها را ماشین‌های اروپایی تشکیل می‌دهند.

ستون پنجم که درصد تجمعی (Cumulative Percent) هر یک از سطوح را محاسبه می‌کند. این معیار برای داده‌های گسسته ترتیبی (Ordinal) مفید می‌باشد که درصد هر یک از سطوح را با درصد سطح قبل آن جمع می‌نماید.

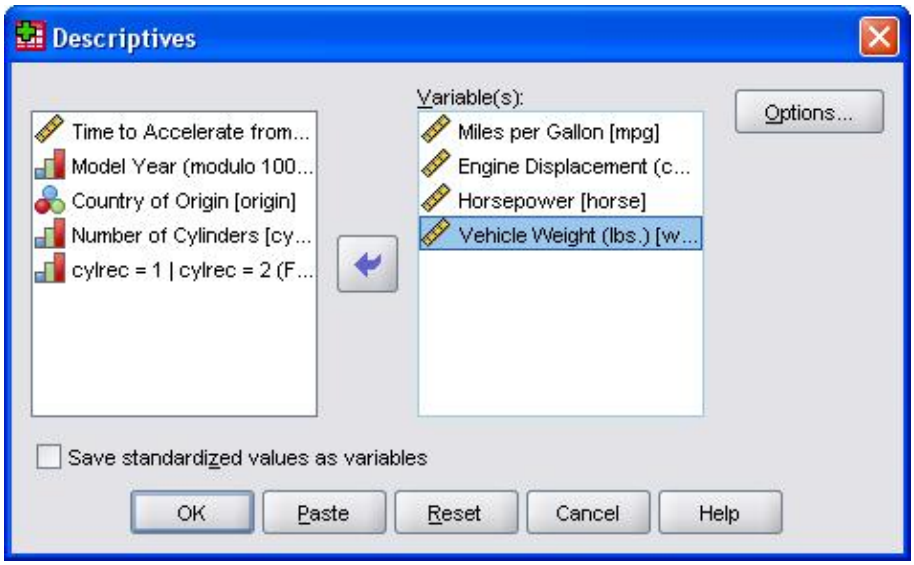
در نهایت نمودار میله‌ای در هر یک از میله‌های خود مقدار فراوانی یکی از سطوح متغیر Origin را نمایش می‌دهد. نکته قابل توجه آن است که همانطور که ملاحظه می‌شود در تمامی خروجی به جای نام متغیر [Origin]، نام برچسب آن [Country of Origin] آورده شده که می‌تواند سودمندتر باشد.

4ـ3ـ فرمان Descriptives:

با اجرای این فرمان آمار توصیفی مربوط به متغیرهای پیوسته در قالب یک جدول با نام Descriptive Statistics در خروجی ظاهر می‌شود. این فرمان بر خلاف فرمان Frequancies قابلیت ترسیم نمودارها و جداول توزیع فراوانی را ندارد. و به منظور ارائه کمیت‌های مرکزی و پراکندگی برای متغیرهای پیوسته مورد نظر مورد استفاده قرار می‌گیرد. با فعال کردن این فرمان از طریق مسیر:

Analyze>Descriptive Statistics>Descriptives…

جعبهٔ محاوره‌ایی به شکل زیر باز می‌شود.



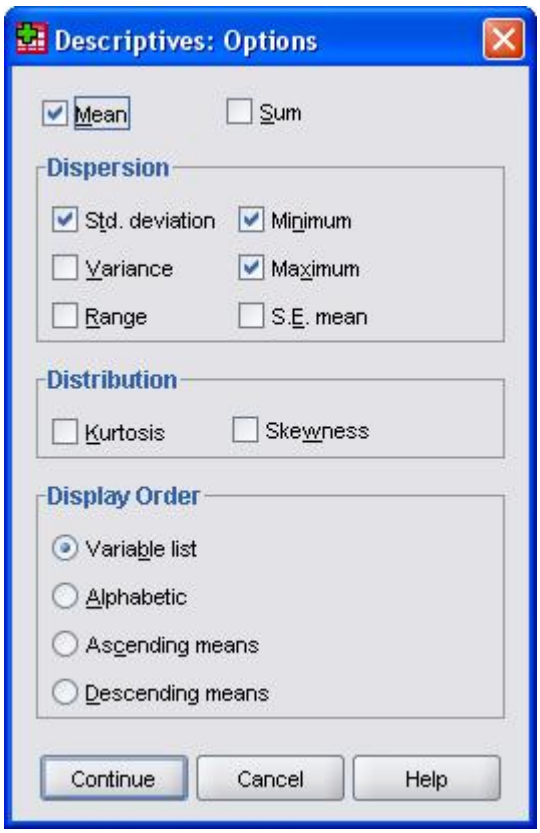
همان طور که ملاحظه می‌شود متغیرهای فایل جاری Cars در سمت چپ جای گرفته‌اند و از میان آنها متغیرهای پیوسته‌ایی را که مورد نظر است انتخاب و به جعبهٔ Variable(s) منتقل می‌کنیم. در این فایل متغیرهای mpg، Engine، Horse و Weight متغیرهای پیوسته هستند و کلیه آنها را به جعبهٔ Variable(s) برای محاسبهٔ کمیت‌های آماری هر یک وارد می‌نمائیم.

گزینهٔ Save Standardize Values as Variables امکان ساخت و ذخیره‌سازی متغیرهای استاندارد شده از متغیرهای مورد نظر را فراهم می‌آورد. متغیر استاندارد شده Z متغیری است که از متغیر اصلی X با میانگین μ و انحراف معیار σ با رابطه $Z = \frac{X - \mu}{\sigma}$ بدست می‌آید و متغیری است بدون واحد با میانگین صفر و واریانس یک. متغیرهای استاندارد شده پس از ساخت با اضافه شدن Z به اول نامشان در کنار متغیرهای فایل جاری قرار می‌گیرند و در حقیقت به مجموعهٔ متغیرهای موجود فایل جاری اضافه می‌شوند.

لازم به ذکر است که کلید Reset در کلیه جعبه‌های محاوره امکان برگرداندن وضعیت جعبه گفتگو به حالت اولیه و کلید Paste امکان کپی کردن فرمان جعبهٔ محاوره باز شده را در پنجرهٔ برنامه‌های Syntax

فراهم می‌آورد.

با انتخاب کلید Options... نوع آماره‌هایی که برای متغیرهای منتخب محاسبه می‌شود و ترتیب قرار گرفتن آنها را در خروجی، تعیین می شود.. با فعال کردن این کلید جعبه گفتگوی زیر باز می‌شود.



از میان پارامترهای مرکزی تنها میانگین (Mean) و مجموع داده‌ها (Sum) قابل انتخاب هستند و در بخش Dispersion معیارهای پراکندگی مانند انحراف معیار، واریانس، برد داده‌ها، کوچکترین مقدار و بزرگترین مقدار قابل انتخاب هستند.

در بخش Distribution نیز مانند فرمان Frequencies معیارهای کمی شناسایی نوع توزیع یعنی ضریب چولگی و ضریب کشیدگی قابل انتخاب و نمایش در خروجی برای داده‌های پیوسته امکان پذیر است.

در بخش Display Order نحوهٔ نمایش خروجی‌های آماری و نحوه قرار گرفتن متغیرها بیان شده

است. گزینه Ascending Means جدول خروجی متغیرها را بـه ترتیـب افـزایش میـانگین آنهـا مرتـب می‌کند، گزینهٔ Descending Means عکس این عمل را انجام می‌دهد.

گزینهٔ Alphabetic کمیت‌های متغیرهای خروجی را به ترتیب الفبای نام متغیرهـا فهرسـت کـرده و گزینه Variable List متغیرها را در جدول خروجی به ترتیب قرار گرفتن آنهـا در جعبـهٔ Variable(s) فهرست می‌نماید.

4ـ3ـ1ـ نمونه‌ایی از اجرای فرمان Descriptives و تحلیل نتایج آن:

با انتخاب متغیرهای mpq، Engine، Horse و Weight که مشخصات ماشین‌های تولید شده هستند و وارد کردن آنهـا بـه جعبـهٔ Variable(s) و انتخـاب گزینـهٔ Save Standardized Values as Variable از پائین پنجرهٔ Descriptives و انتخاب گزینه‌های مورد نظر از کلیـد Options... نظیـر Mean، SUM، Std.Deviation، Variance، Maximum و S.E Means از قسـمت Dispersion و آیـتم Kurtoris که معیار کشیدگی توزیع متغیرهاست و گزینهٔ Ascending Means از قسمت Display Order دکمـهٔ Continue و سپس OK را انتخاب می کنیم..

جدول خروجی‌ها به صورت ذیل می‌باشد:

Descriptive Statistics

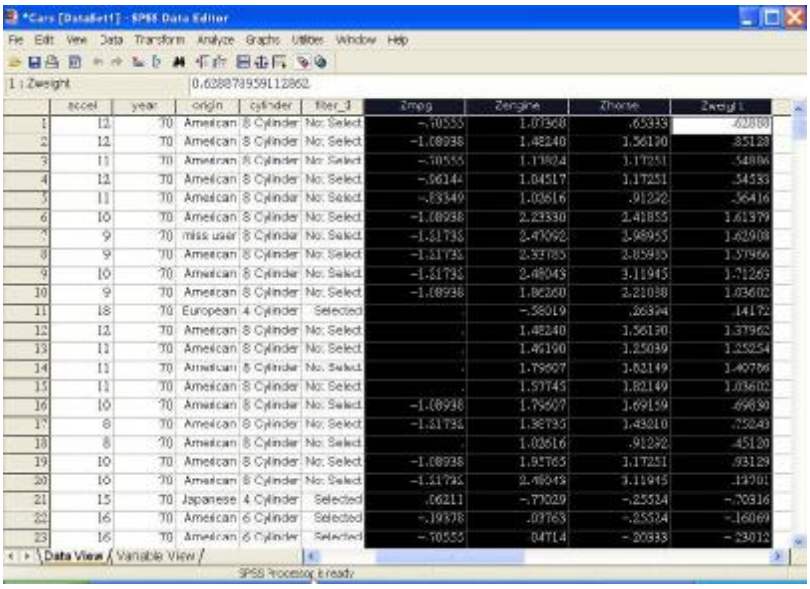
	N	Minimum	Maximum	Mean		Std. Deviation	Variance	Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
Miles per Gallon	398	9	47	23.51	.587	11.716	137.269	-.511	.244
Horsepower	400	46	230	104.89	.979	19.574	383.009	.59	.243
Engine Displacement (cu. inches)	406	4	455	194.04	5.27	105.977	11233.545	-.701	.242
Vehicle Weight (lbs.)	406	756	5140	2969.56	47.774	961.837	925225.4	-.757	.242
Valid N (listwise)	392								

در ستون اول برچسب هر یک از متغیرها بیان شده است. در ستون دوم تعداد داده‌های معتبر هر یـک از متغیرها بیان شده است.

در ستون سوم ماکزیمم مقدار، در ستون چهارم جمع مقادیر، در ستون پنجم مقدار میانگین، در ستون ششم انحراف معیار میانگین [Std.Error Mean] در ستون هفتم انحراف معیار، در ستون هشتم واریانس

و در ستونهای نهم و دهم ضریب کشیدگی هر یک از متغیرها همراه با انحراف معیار ضریب کشیدگی [Std.Error Kurtosis] بیان شده است.

همانطور که ملاحظه می‌شود چون گزینهٔ Ascending Means را از قسمت Display Order انتخاب نموده‌ایم متغیرها به ترتیب صعودی بر حسب میانگین آنها مرتب شده‌اند و متغیر mpg با کمترین میانگین اول و متغیر Weight با بیشترین میانگین آخرین متغیر در جدول آورده شده است. مقادیر استاندارد شده هر یک از متغیرها نیز همانطور که ملاحظه می‌شود به فایل جاری اضافه شده است.



4ـ4ـ فرمان Explore:

با اجرای این فرمان امکان محاسبهٔ آمارهای توصیفی(محاسبه کمیتها، نمودارها و جداول) برای متغیرهای پیوسته در سطوح متغیرهای گسسته فراهم می‌آید. برای نمونه با اجرای این فرمان می‌توانید در فایل اطلاعاتی مربوط به بیماران مبتلا به هپاتیت آمار توصیفی میزان قند خون بیماران را که یک متغیر پیوسته است در سطوح مختلف جنسیت که متغیر گسته است یکبار فقط برای مردها و یکبار هم فقط برای زنها ملاحظه نمایید.

در اجرای این فرمان امکان انتخاب روشهای ترسیمی، محاسباتی یا هر دو وجود دارد. با استفاده از این

فرمان نمودارهای شاخه و برگ (Steam&Leaf) و نمودار جعبه‌ایی، برآوردهای مقاوم میـانگین نتـایج و آزمون نرمال بودن داده‌ها نیز به نمایش در می‌آیند.

جعبهٔ گفتگوی زیر پس از فعال کردن این فرمان از مسیر:

Analyze>Descriptire Statistics>Explore…

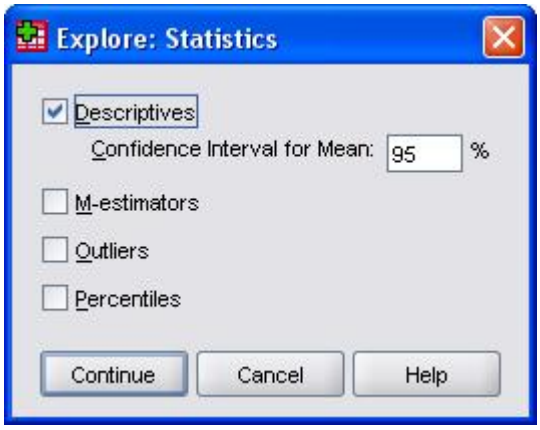
ظاهر می‌شود:

نام متغیرهای فایل جاری در جعبه سمت چپ قرار می‌گیرد. متغیرهایی که قصد ارائـه آمـار توصـیفی آنها را داریم و متغیر پیوسته هستند بـه جعبـهٔ Dependent List و متغیـر گسسـته کـه دارای سـطوح مختلف است به جعبهٔ Factor List منتقل می‌کنیم.

بخش Label Caes by برخی حالتهای غیرعادی در متغیرها را با برچسب ویژه‌ایی معرفی می‌نماید.

بخش Display نوع روشهای آمار توصیفی مورد استفاده بر حسب آماره‌های کمی (گزینه Statistics)، نمودارها (گزینه Plots) یا هر دو (گزینهٔ Both) را مشخص می‌کند، در صورت فعـال شـدن هر گزینه کلید مربوط به آن در جعبه گفتگو برجسته می‌شود.

کلید Statistics… آماره های توصیفی موجود را نمایش می‌دهد.



گزینهٔ Descriptives کلیهٔ آماره‌های معمول، شامل میانگین، میانه، مد، کمینه، بیشینه، برد، دامنه میان چارکی، دامنه حذفی 5% و فاصلهٔ اطمینان با درصد اطمینان تعیین شده در جعبه Confidence Interval for Mean را برای میانگین متغیر پیوسته در سطوح متغیر گسسته نمایش می‌دهد.

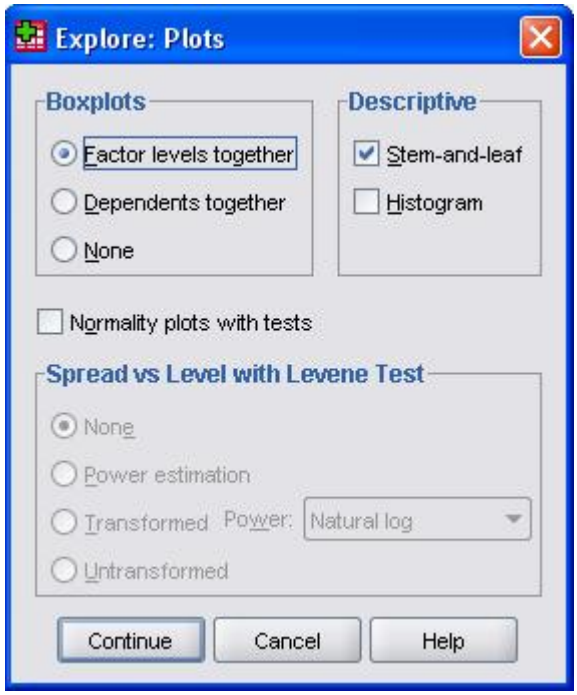
گزینهٔ M-Estimatores برآوردکننده‌های نوع M میانگین را در خروجی آماره‌های توصیفی نمایش می‌دهد. این برآوردکننده‌ها با روش ماکزیمم در تنهایی بدست می‌آیند.

اعداد داخل پرانتز در خروجی فرمان، وزنهای مفروض هر روش را نمایش می‌دهند. گزینهٔ Outliers در صورت انتخاب شدن، داده‌های پرت آماری را در خروجی‌ها نشان می‌دهد.

داده‌های پرت داده‌هایی هستند که در صحت وجود مقدار صحیح آنها شک و تردید وجود دارد. البته با انتخاب این گزینه نرم‌افزار تعداد 5 دادهٔ بالایی و 5 دادهٔ پائینی را به عنوان دادهٔ پرت در نظر می‌گیرد. گزینه Percentiles، صدکهای پنجم، دهم، بیست و پنجم (چارک اول)، پنجاهم (چارک دوم)، نودم و نود و پنجم داده‌ها را در خروجی نمایش می‌دهد.

گزینهٔ Grouped Frequency tables جداول فراوانی گروه‌بندی شده را به تفکیک مقادیر مختلف متغیر گسسته نمایش می‌دهد.

کلید Plots... تنوع نمودارهای ترسیمی را نشان می‌دهد. بعد از انتخاب این کلید پنجره گفتگوی زیر باز می‌شود:



در بخش Box Plots، نمودارهای جعبه‌ای قابل ترسیم را بیان می‌کند. می‌توان با انتخاب Factor Levels Together نمودار جعبه‌ایی را بر اساس سطوح متغیر گسسته به صورت هم‌زمان ترسیم نمود و یا می‌توان با انتخاب Dependents Together هر سطح را به صورت مستقل ترسیم نمود و یا می‌توان با انتخاب None نمودار جعبه‌ای ترسیم نکرد.

در بخش Descriptive در صورتی که کاربر نمودار شاخه و برگ را در خروجی می‌خواهد گزینهٔ Steam-And-Leaf را فعال و در صورتی که نمودار هیستوگرام متغیر پیوسته در سطوح متغیر گسسته را می‌خواهد گزینهٔ Histogram را فعال می‌نماید.

در صورت نیاز به آزمون نرمال بودن داده‌های متغیر پیوسته در هر یک از سطوح متغیر گسسته، گزینهٔ Normality Plots With Tests را فعال می‌کنیم.

ـ آزمونهای Levence جهت آزمونهای همگنی واریانس‌ها و نمودارهای مربوط به آن در بخش Spread VS.Level With Levene Test تعیین می‌شوند.

ـ گزینه None سبب حذف آزمونها و نمودارهای همگنی واریانسها است.

ـ گزینهٔ Power Estimation سبب می‌شود نموداری از لگاریتم گستردگی (بر مبنای طبیعی میان چارکی) در برابر لگاریتم سطوح (لگاریتم طبیعی میانه‌مقادیر) در خروجی به نمایش درآید.

ـ گزینهٔ Transformed در قسمت Power تبدیل خاصی بر روی داده‌ها اعمال می‌نماید. و گزینهٔ Power تبدیل Untransformed داده‌های اولیه را بدون هیچ تبدیلی مورد آزمون همگنی واریانس قرار می‌دهد.

کلید ...Options وضعیت داده‌های گم شده را تعیین می‌کند:

ـ گزینه Exclude Cases Listwise مقادیر گم شده را برای کلیه متغیرها از تحلیل و نمودار داده‌ها خارج می‌کند.

ـ گزینه Exclude Cases Pairwise مشاهدات را تنها برای متغیری که دادهٔ گمشده دارد خارج می‌کند و بر روی سایر متغیرها اثری ندارد.

ـ گزینهٔ Report Values مقادیر گم شده را گزارش و نمودارها و آمارها را مانند گزینهٔ Exclude Cases Listwise برای گروه متغیرهای گم شده در کنار سطوح متغیر گسسته محاسبه می‌نماید.

4ـ4ـ1ـ نمونه‌ای از خروجی‌های فرمان Explore و تحلیل آن

با توجه به اطلاعات فایل جاری که مربوط به مشخصات بیماران هپاتیتی است متغیر قند خون را به عنوان متغیر پیوسته وارد جعبهٔ Dependent list نموده و متغیر جنسیت را به عنوان متغیر گسسته وارد قسمت Factor list می‌کنیم.

از قسمت ...Statistics گزینهٔ M-Estimators ، Outliers و برای نمونه انتخاب کرده و Continue را می‌زنیم تا به صفحهٔ اصلی بازگردیم.

از قسمت ...Plots در قسمت Boxplots گزینهٔ Factor Levels together و از Descriptive قسمت Histogram و از قسمت Normality Plots With Tests گزینهٔ Untrannsformed را انتخاب نموده و Continue را می‌زنیم و OK می‌نمائیم.

نتایج خروجی‌ها به صورت جداول ذیل است:

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	جنسیت	N	Percent	N	Percent	N	Percent
قندخون	مرد	13	100.0%	0	.0%	13	100.0%
	زن	17	100.0%	0	.0%	17	100.0%

جدول Case Processing Summary تعداد داده‌های معتبر و داده‌های گم شده متغیر قند خون (متغیر پیوسته) را برای دو سطح متغیر گسسته بیان می‌کند.

Descriptives

	جنسیت			Statistic	Std. Error
قندخون	مرد	Mean		531.1538	32.47540
		95% Confidence Interval for Mean	Lower Bound	460.3960	
			Upper Bound	601.9117	
		5% Trimmed Mean		532.7265	
		Median		530.0000	
		Variance		13710.474	
		Std. Deviation		117.09173	
		Minimum		312.00	
		Maximum		722.00	
		Range		410.00	
		Interquartile Range		191.00	
		Skewness		-.149	.616
		Kurtosis		-.580	1.191
	زن	Mean		346.0000	10.30634
		95% Confidence Interval for Mean	Lower Bound	324.1515	
			Upper Bound	367.8485	
		5% Trimmed Mean		346.7222	
		Median		351.0000	
		Variance		1805.750	
		Std. Deviation		42.49412	
		Minimum		274.00	
		Maximum		405.00	
		Range		131.00	
		Interquartile Range		83.00	
		Skewness		-.317	.550
		Kurtosis		-1.269	1.063

جدول Descriptive آمارهای توصیفی از قبیل میانگین، سطوح بالا و پائین فاصلهٔ اطمینان 95 درصد برای میانگین، میانگین 5% کوتاه شده (که 2/5 درصد از بالا و 2/5 درصد از داده‌های پائین آن حذف شده و سپس میانگین محاسبه شده است) ، میانه، واریانس، انحراف معیار، کمترین مقدار، بیشترین مقدار، برد درون چارکی، ضریب چولگی و ضریب کشیدگی برای قند خون هر یک از سطوح مرد و زن ارائه شده است.

M-Estimators

	جنسیت	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
قندخون	مرد	532.5314	532.4690	533.6660	532.4466
	زن	348.4588	347.9187	346.5981	347.9209

a. The weighting constant is 1.339.

b. The weighting constant is 4.685.

c. The weighting constants are 1.700, 3.400, and 8.500

d. The weighting constant is 1.340*pi.

در جدول بالا مقادیر میانگین بیشینهٔ درستنمایی قند خون برای دو گروه زن و مردها با چهار روش ارائه شده است.

Extreme Values

	جنسیت			Case Number	Value
قندخون	مرد	Highest	1	20	722.00
			2	17	670.00
			3	23	627.00
			4	28	627.00
			5	15	594.00
		Lowest	1	25	312.00
			2	12	417.00
			3	9	422.00
			4	1	450.00
			5	6	470.00
	زن	Highest	1	5	405.00
			2	7	394.00
			3	13	392.00
			4	29	388.00
			5	16	384.00
		Lowest	1	11	274.00
			2	8	284.00
			3	26	294.00
			4	10	297.00
			5	24	309.00

این جدول نتایج اجرای گزینهٔ داده‌های پرت را نشان می‌دهد که برای هر یک از دو سطح تعداد 5 داده ابتدایی و انتهایی را با شماره سطر آنها بیان کرده است.

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
قندخون	Based on Mean	14.676	1	28	.001
	Based on Median	14.508	1	28	.001
	Based on Median and with adjusted df	14.508	1	16.560	.001
	Based on trimmed mean	14.827	1	28	.001

در جدول Test of Homogenity of Variance آزمون هم‌گرایی واریانس با استفاده از چهار روش، بر اساس میانگین، بر اساس میانه، بر اساس درجه آزادی تعدیل شده و بر اساس میانگین کوتاه شده ارائه شده است برای قضاوت در مورد اینکه آیا واریانس‌ها همگرا هستند یا نه ،مقدار Sig یا سطح معنی‌داری را برای هر یک از روشهای فوق در نظر می‌گیریم و آن را با مقدار α (خطای نوع اول) تحقیق مقایسه می‌کنیم. اگر $Sig>\alpha$ آنگاه می‌توان گفت که واریانس‌ها در دو گروه برابر هستند و اگر $Sig<\alpha$ آنگاه می‌توان گفت واریانس‌ها در دو گروه برابر نیستند، همان طور که ملاحظه می‌شود در روشهای فوق

Sig>α (با در نظر گرفتن α=0/05) است که می‌توان گفت واریانس‌ها با هم برابرند.

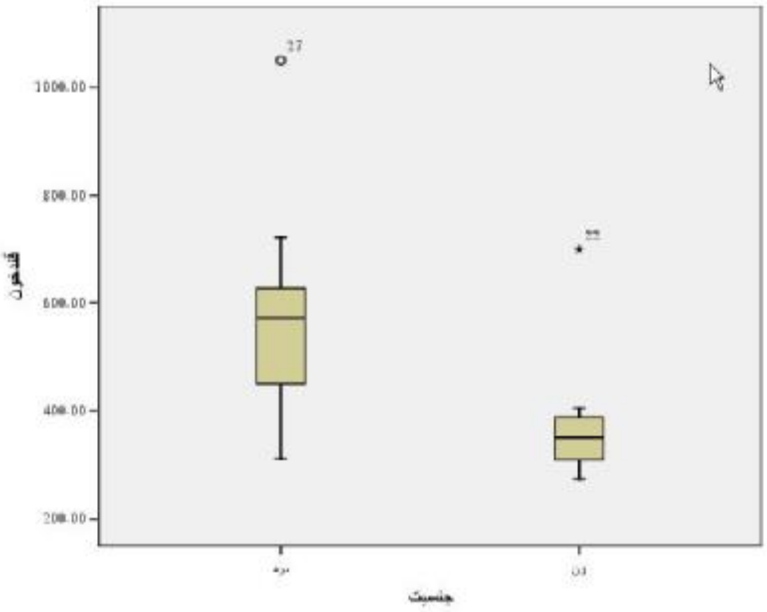
جنسیت= مرد

Frequency	Stem & Leaf
1.00	3 . 1
5.00	4 . 12579
3.00	5 . 379
3.00	6 . 227
1.00	7 . 2

جنسیت= زن

Frequency	Stem & Leaf
4.00	2 . 7899
4.00	3 . 0244
8.00	3 . 55678899
1.00	4 . 0

نمودار بالا هیستوگرام افقی را برای قند خون خانم‌ها و آقایان مبتلا بیان می‌کند. این نمودار به نمودار شاخه و برگ مشهور است. در گذشته با توجه به اینکه امکان کامپیوترهای شخصی در دسترس همگان نبود با استفاده از کشیدن شکل شاخه و برگ مشخص می نمودند که آیا داده ها دارای توزیع نرمال هستند یا خیر؟ اما در حال حاضر با توجه به اینکه می توان به طور دقیق با استفاده از نمودارهای هیستوگرام و آزمونهای آماری که در فصول بعد مورد بررسی قرار می گیرد کاربرد این شکل بسیار محدود شده است و حتی در کتب آماری نیز به آن اشاره ایی نمی شود.



نمودار Box Plot معیاری برای شناسایی شکل توزیع و شناسایی داده‌های پرت است. در شکل بالا نمودار جعبه‌ای برای هر دو سطح زنها و مردها بیان شده است. خط سیاه وسط جعبه بیان کننده میانه داده‌ها و ابتدا و انتهای جعبه نشان دهندهٔ چارک اول و چارک سوم یا صدک 25ام و صدک 75 ام هستند. هر چه خط مشکی بیشتر در وسط مستطیل باشد نشان‌دهنده شکل نرمال داده‌هاست. در دنبالهٔ بالا و پائین مستطیل که دو خط افقی در انتهای دو خط عمودی بالا و پائین مستطیل هستند مرز بین داده‌های پرت و مرزی می‌باشند. برای مثال در شکل بالا نفر 27ام که یک مرد است مقدار قند خون پرت (داده‌های پرت با علامت O مشخص می‌شوند) و نفر 22ام که یک زن است مقدار فشار خون مشکوک مرزی (داده‌های مرزی با علامت * مشخص می‌شوند) دارند، معیار پرت بودن داده‌ها در این روش بیرون قرار گرفتن از فاصلهٔ $(\mu-3\sigma$ و $\mu+3\sigma)$ می‌باشد.

خلاصه مباحث مطرح شده :

در این فصل سه ابزار ارائه آمارهای توصیفی در نرم‌افزار SPSS بیان گردید. در این فصل ابزارهای آمار

توصیفی که در زیر منو Analyze وجود دارند اشاره شد . این ابزارهای در قالب دستور Frequencies برای آمار توصیفی متیرهای گسسته، دستور Descriptive برای آمار توصیفی متغیرهای پیوسته و دستور Explore برای ارائه آمار توصیفی متغیرهای پیوسته به تفکیک سطوح یک متغیر گسسته دیگر ارائه شد. در این فصل علاوه بر ارائه نحوه کار با هر یک از دستور های یاد شده به انجام دو مثال آمار توصیفی بر روی متغیرهای گسسته و پیوسته اشاره گردید. دیگر زیر آیتم‌های منو Analyze مربوط به آمار استنباطی و بدست آوردن رابطهٔ میان متغیرهاست که در فصول بعد به تعدادی از آنها مانند جداول توافقی، آزمون فرض‌ها، رگرسیون، ناپارامتری، ضرایب همبستگی و کنترل کیفیت به تفضیل پرداخته خواهد شد.

تمرین:

برای فایل employe data که درون خود نرم افزار spss به آدرس

G:\sampels:\employee data.sav قرار دارد مربوط به اطلاعات پرسنل یک شرکت است که شامل متغیرهای ذیل می باشد.

Id :شماره پرسنلی

Gender : جنسیت

Bdate : تاریخ تولد

Educ: سابقه سالهای تحصیل به سال

Jobcat: رده شغلی

Salary: درآمد

Salbegin: دآمد در هنگام شروع به کار

Jobtime: سابقه در شرکت

Prevexp: سابقه کار در قبل از ورود به شرکت

Minority: اقلیت بودن یا نبودن

حال با توجه به اطلاعات موجود در فایل مورد نظر به سوالات ذیل پاسخ دهید.

1. آمار توصیفی متغرهای گسسته فایل مورد نظر را بدست آورید.
2. آمار توصیفی متغیرهای پیوسته فایل مورد نظر را بدست آوید.
3. آمار توصیفی متغیرهای پیوسته فایل مورد نظر به تفکیک جنسیت را بدست آورید.
4. آمار توصیفی متغیرهای پیوسته فایل مورد نظر را به تفکیک اقلیت بودن بدست آورید.
5. آمار توصیفی متغیرهای پیوسته را به تفکیک رده شغلی بدست آورید.

فصل پنجم

ویرایش مقادیر داده‌ها

اهداف آموزشی فصل پنجم

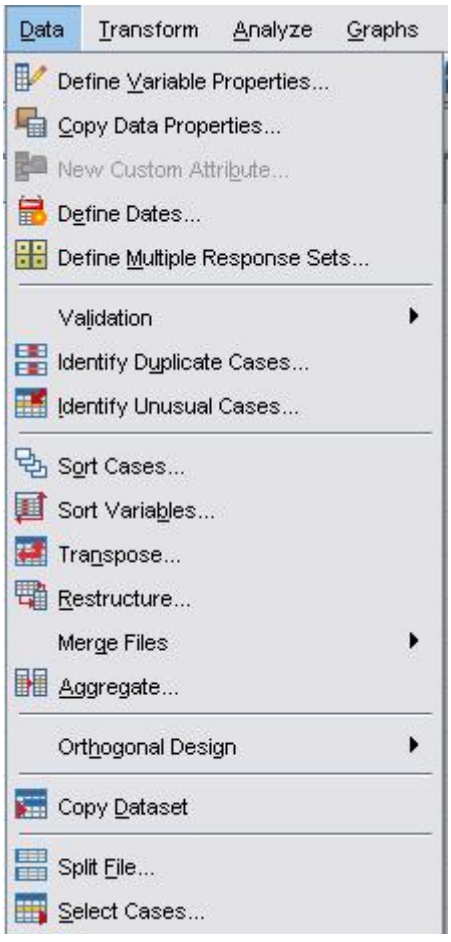
فراگیر پس از مطالعه این فصل باید بتواند:

4- امکانات منو Data به منظور ویرایش مقادیر داده ها را بیان کند.

5- کاربر هر یک از زیر آیتم های منو Data را به همراه ذکر کثال شرح دهد.

6- قابلیت های منو Transform را بازگو نماید.

7- کاربر هر یک از ابزارهای منو Transform را با ذکر مثال شرح دهد.



5ـ1ـ مقدمه:

در این بخش ابزارهایی معرفی می‌گردند که به کمک آنها می‌توان بر روی داده‌های فایل جاری تغییراتی نظیر اعمال محاسبات ریاضی بر روی متغیرها، تغییر کد متغیرها، رتبه‌بندی سطوح متغیرها، مرتب کردن داده‌ها، دسته‌بندی داده‌ها، وزن‌دهی داده‌ها، چسباندن دو فایل داده به هم و .. را از طریق ابزارهای دو منو Data و Transform انجام داد. در ادامه هر یک از زیر منوهای عنوان شده به همراه کاربرد هر یک با استفاده ازمثالهای متنوع بیان شده

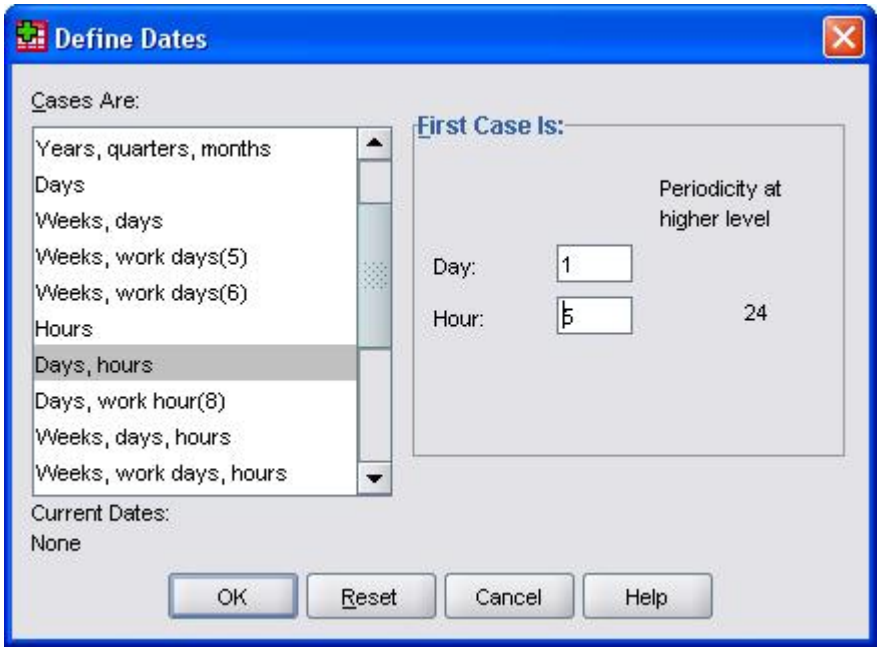
است.

5ـ2ـ منو Data (داده‌ها):

در این منو امکان تغییر متغیرها، اضافه کردن متغیرهای جدید به فایل داده‌ها، مرتب کردن داده‌ها، ادغام داده‌ها، انتقال داده‌ها و ... را در اختیار کاربر قرار می‌دهد. درذیل تعدادی از ابزارهای پرکاربرد این منو بیان شده است. در ذیل تعدادی از زیر آیتم های کاربردی این منو توضیح داده شده است.

- تعریف متغیر زمان (از طریق زیر منو Define Dates):

با انتخاب این زیر آیتم ، پنجرهٔ زیر باز می‌شود و کاربر می‌تواند با انتخاب هر یک از قالبهای زمان یک یا چند متغیر را به فایل جاری اضافه نماید.

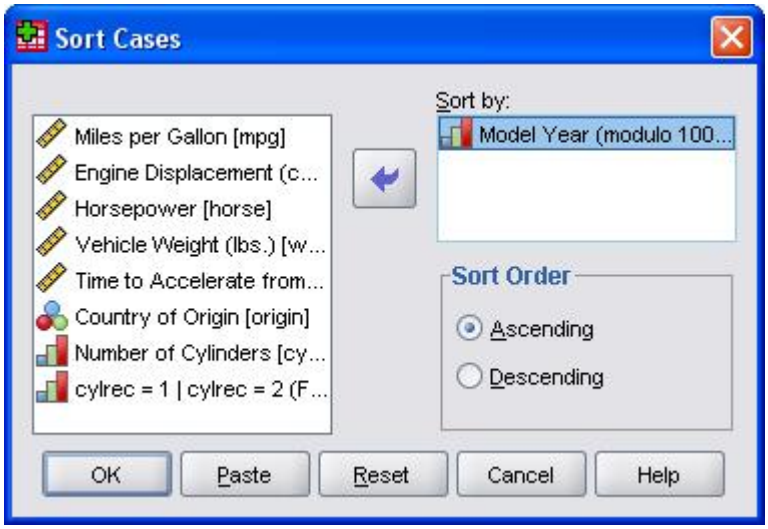


برای نمونه در صورتی که کاربر از قسمت Case Ave نوع Hours و Days را انتخاب نماید و از قسمت First Case is، در قسمت Day عدد 1 و در قسمت Hour عدد 0 را بنویسید آنگاه در پایان متغیرها در فایل جاری سه ستون با نام‌های DAY- ، HOUR- و DATE- اضافه می‌شود که روز ساعت

و زمان (روز و ساعت) را به ترتیب برای هر سطر نوشته و به پیش رفته‌اند برای مثال تاریخ -DAY بعد از 24 سطر به عدد 2 تغییر می‌کند و -HOUR نیز بین عدد 0 و 23 در سطرها تکرار می‌گردد.

- مرتب کردن سطرها(از طریق زیر منو Sortcases):

اگر بخواهیم کل داده‌ها و سطرهای پنجرهٔ Data View را بر حسب مقادیر یک متغیر مرتب نمائیم از این آیتم استفاده می‌نمائیم. در صورت انتخاب این فرمان جعبهٔ گفتگوی زیر ظاهر می‌شود.



در قسمت سمت چپ فهرست متغیرهای فایل جاری دیده می‌شود. در صورت نیاز می‌توان هر متغیری را برای مرتب کردن انتخاب و با Click کردن بر روی آن و سپس Click کردن بر روی نشانه 4 آن را در جعبه Sort by قرار داد. در این حالت کلیه متغیرهای موجود در فایل بر اساس مقادیر متغیر یا متغیرهای منتخب در جعبهٔ Sort by مرتب می‌شوند. بخش Sort Order چگونگی مرتب کردن داده‌ها را نشان می‌دهد. برای مرتب کردن داده‌ها به صورت افزایشی گزینهٔ Ascending و کاهشی گزینهٔ Descending را انتخاب کنید.

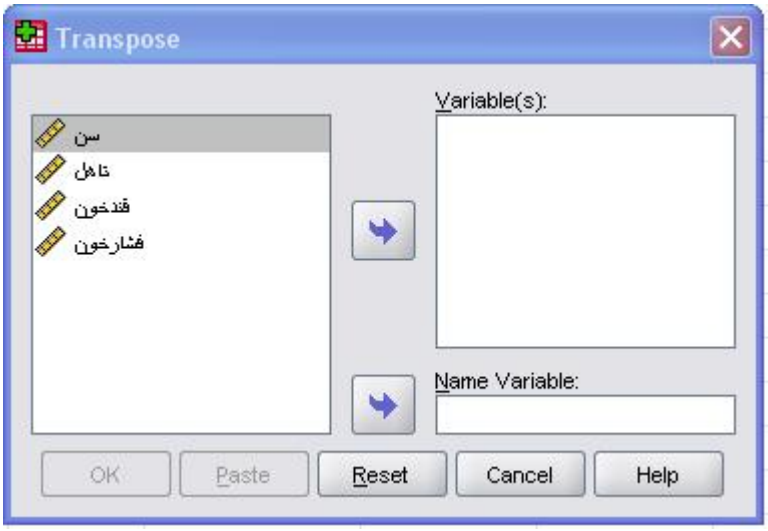
برای مثال اگر در فایل Cars متغیر Weight را مبنای مرتب کردن داده‌ها قرار داده و آن را به جعبهٔ Sort by منتقل کنیم و گزینهٔ Ascending را انتخاب می کنیم در نتیجه در فایل جاری ابتدا اطلاعات متغیرهای ماشینی را در سطر اول می‌دهد که وزن آن از همه کمتر است و در نتیجه آخرین سطر

اطلاعات ماشینی است که دارای بیشترین وزن بوده است.

- برگرداندن داده‌ها(از طرق زیر منو Transpose):

با انتخاب این دستور در منو Data می‌توان جای سطرها و ستون‌ها را در مجموعه داده‌ها جابجا کرد.

پس از انتخاب این فرمان، جعبهٔ گفتگوی زیر ظاهر می‌شود:



متغیرهای موجود در جعبهٔ سمت چپ، ستونهای فایل جاری هستند، با انتخاب حداقل یکی از این متغیرها و انتقال آن به جعبه Variable(s) ستونهایی که قصد برگرداندن آنها به سطر را داریم، انتخاب می‌کنیم (ذکر این نکته ضروری است که سایر ستونهای انتخاب نشده پس از اجرای فرمان حذف می‌شوند) جعبهٔ Name Variable به ذخیره نام متغیرهایی اختصاص دارد که برگردان می‌شوند، در صورتی که چنین متغیری معرفی نشود نرم‌افزار به صورت خودکار آن را در متغیر Case-LbL جای می‌دهد. برای مثال 3 متغیر با نام‌های X_1 و X_2 و X_3 به صورت ذیل موجود هستند.

Name	X_1	X_2	X_3	X_4
1	1	10	17/5	100
2	2	30	12/4	200
3	1	40	13/7	200

نتیجهٔ برگردان این متغیرها به صورت ذیل است:

	CASE_LBL	var001	var002	var003
1	x1	1.00	2.00	1.00
2	x2	10.00	30.00	40.00
3	x3	17.50	12.40	13.70
4	x4	100.00	200.00	200.00

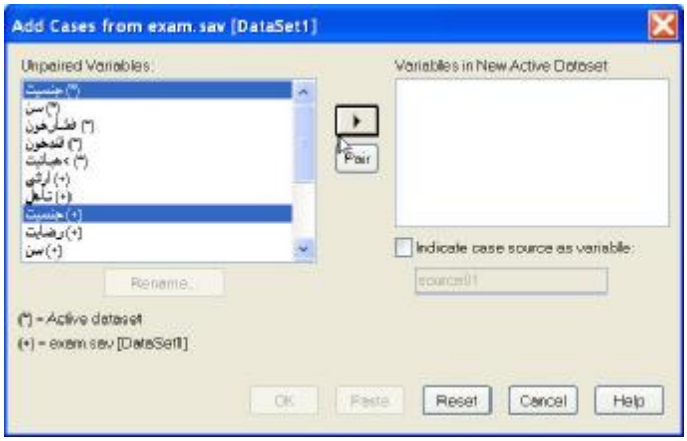
همانطور که ملاحظه می‌شود در ستون Case-LbL نام متغیرها قبل از برگرداندن حفظ شده است.

-اضافه کردن داده‌ها از فایل‌های دیگر(از طریق زیر منو Marge Files):

هر گاه کاربر بخواهد متغیر یا سطری از فایل دیگری را به فایل جاری اضافه نماید از زیر منو Marge File در منو Data استفاده می‌نماید. این زیر منو مانند شکل ذیل دارای دو فرمان Add Cases... برای اضافه نمودن سطرهای جدید از یک فایل بیرونی به فایل جاری و فرمان Add Variables... جهت اضافه کردن متغیرهای جدید از یک فایل بیرونی به فایل جاری است. در ذیل نحوه کار با هر دو بیان شده است:

- اضافه کردن سطرهای جدید(از طریق زیر منو Add Cases):

ابتدا دستور Add Cases را فعال می‌کنیم. پنجره‌ایی برای انتخاب فایل داده‌ایی که سطر یا سطرهایی از آن به فایل جاری اضافه خواهد شد فعال می‌گردد. شکل پنجره دقیقاً شبیه به پنجرهٔ Save as می‌باشد. با انتخاب فایلی که قصد اضافه کردن سطرهای آن را به فایل جاری داریم پنجره‌ایی به صورت زیر باز می‌شود:



عنوان جعبه، نشان‌دهنده دستور فرعی و مسیر فایل انتخابی برای ادغام با فایل جاری است برای مثال متغیرهای فایل Exam2.sav برای اضافه شدن به سطرهای متغیرهای Exam انتخاب شده‌اند. ذکر این نکته ضروری است که Add Cases زمانی به کار می‌رود که هر دو فایل دارای متغیرهای مشابه باشند و تنها بخواهیم داده‌های آنها را در یک فایل با یکدیگر جمع نمائیم. برای مثال اطلاعات یک بررسی میدانی که در استانهای اصفهان و تهران انجام گرفته و پرسش‌نامه‌های آن کاملاً یکسان بوده است را بخواهیم از این طریق با یکدیگر ادغام کنیم. در این صورت اگر فایل اطلاعاتی تهران فایل جاری و فایل اطلاعاتی اصفهان فایل بیرونی باشد با این دستور می‌توان اطلاعات افراد مورد بررسی در اصفهان در هر یک از متغیرها را به اطلاعات افراد مورد بررسی فایل تهران در همان متغیرها اضافه نمود و فایلی تشکیل داد که حاوی اطلاعات بررسی میدانی از شهروندان تهرانی و اصفهانی است.

در قسمت سمت چپ جعبه، فهرستی با عنوان Un Paired Variables (متغیرهای جفت نشده) موجود است که می‌توان از میان آنها متغیرهای مناسب را برای ترکیب با یکدیگر انتخاب کرد. این متغیرها به دو دسته تقسیم می‌شوند، متغیرهایی با نماد [*] مربوط به فایل جاری و متغیرهایی با نماد [+] مربوط به فایل خارجی (فایل منتخب جهت ادغام) است. در سمت راست جعبهٔ متغیرهای جفت شده قرار می‌گیرند و فرد می‌تواند متغیرهای جفت نشده را از هر یک از فایلها انتخاب و با انتخاب گزینهٔ Pair به صورت جفت شده به قسمت سمت راست منتقل کند.

کلید Rename... امکان تغییر نام متغیرهای هر دو فایل جاری و خارجی را فراهم می‌کند. این گزینه برای زوج کردن متغیرهایی با نام‌های متفاوت یا تمایز بیان دو متغیر از دو فایل مختلف که هم نام هستند به کار می‌رود. با تأئید فرمان توسط دکمهٔ OK می‌توانید نتیجه ادغام دو فایل را در پنجرهٔ Data View ملاحظه کنید. ذکر این نکته ضروری است که نتیجهٔ ادغام در یک فایل جدید با نام جدید نمایش داده می‌شود.

- اضافه کردن متغیرها(از طریق زیر منو Add Variables) :

در صورتی که بخواهیم متغیرهایی را از یک فایل بیرونی به متغیرهای فایل جاری اضافه نمائیم از دستور Add Variables استفاده می‌نمائیم. در ابتدا پس از انتخاب فرمان Add Variable از زیر منو Marge Files و منو Analyze، پنجره‌ایی باز می‌شود که در آن باید مسیر فایل خارجی را مشخص و دکمهٔ Continue را انتخاب کنیم. سپس پنجرهٔ گفتگوی ذیل باز می‌شود:

در جعبهٔ سمت چپ متغیرهای هر دو فایل جاری و خارجی با عنوان Excluded Variables (متغیرهای حذف شده) جای می‌گیرد. در صورت تمایل به ادغام متغیرهای فایل خاص، آنها را فعال کرده و به جعبهٔ New Active Datasetدر سمت راست که متغیرهای فایل جدید حاصل از ادغام دو فایل می‌باشد، منتقل می‌کنیم. اگر متغیرهای ادغامی قبلاً مرتب شده باشند، نام متغیر یا متغیرهایی که مبنای این ترتیب بوده‌اند در جعبه Key Variables جای می‌گیرد. این متغیرها می‌توانند در هر دو فایل جاری و خارجی یا یکی از آنها قرار داشته باشند.

گزینهٔ Match Cases محل این متغیرها را نشان می‌دهد. اگر متغیرهای کلیدی در هر دو فایل جاری و خارجی قرار دارند بر روی گزینهٔ ...Both File ، Click کرده، اگـر در فایـل خـارجی هسـتند گزینـهٔ ...External File و اگر در فایل جاری‌اند گزینهٔ ...Working Data را فعال می‌کنیم.

در صورت تمایل به تغییر نام متغیر، بر روی نام متغیر Click کرده و کلید Rename را برای تغییر نام آن بزنید. با تأئید فرمان توسط کلید OK در پنجرهٔ Data View یک فایل حاوی تمامی متغیرهایی که از فایل خارجی و فایل جاری دارد پنجرهٔ New Active Dataset شده‌اند به وجود می‌آید.

- تجزیه فایل جاری(از طریق زیر منو ...Split File):

هر گاه بخواهید تمامی تجزیه و تحلیل‌ها بر روی متغیرها در خروجی در سـطوح مختلـف یـک متغیـر گسسته انجام گیرد از این فرمان استفاده می‌شود. برای مثال اگر کاربر بخواهد فرمان Descriptive را بر روی متغیرهای پیوسته فایل جاری که اطلاعات بیماران مبتلا به هپاتیت است در سطوح مختلـف متغیـر گسسته نوع هپاتیت بدست آورد می‌تواند متغیر گسسته نوع هپاتیت را مبنی Split داده‌ها نمایـد. بعـد از اجرای زیرمنو Split پنجره زیر باز می‌شود:

در قسمت سمت چپ کلیه متغیرهای فایل وجود دارند. در صورتی که هیچ یک از متغیرهای گسسته مبنای شکستن خروجی‌ها نیستد گزینهٔ Analyze all Cases… فعال می‌باشد در صورتی که گزینهٔ Compare Groups را انتخاب و یک متغیر گسسته را وارد Groups Based On: نمائید تمام خروجی‌هایی که توسط فرامین آمار توصیفی و آمار استنباطی انجام می‌دهید در سطوح مختلف نوع هپاتیت و در کنار یکدیگر برای امکان مقایسه نتایج بیان می‌شود. و در صورتی که گزینهٔ Organize Output By Groups: را انتخاب و متغیر نوع هپاتیت را وارد Groups Based On: نمائید. کلیه خروجی‌هایی که بعد از آن به وسیلهٔ آمار توصیفی و آمار استنباطی گرفته می‌شود برای هر یک از سطوح نوع هپاتیت و بدون امکان مقایسه نتایج ارائه می‌شود.

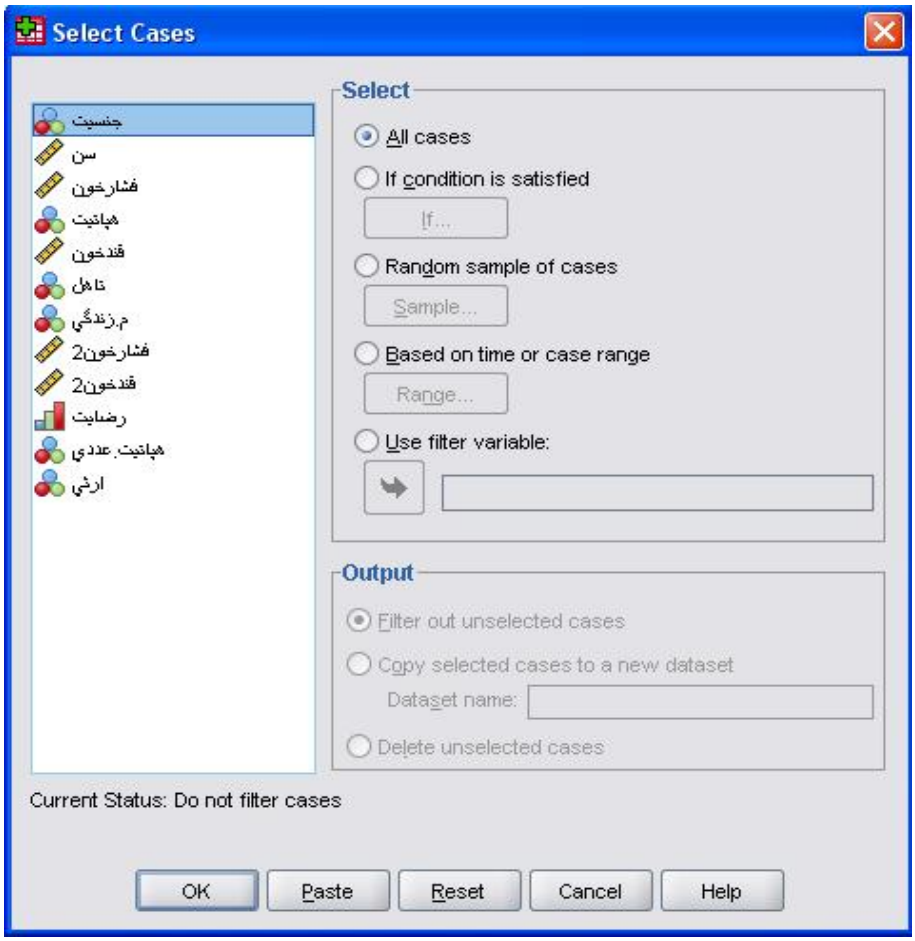
در صورتی که فایل داده‌ها قبلاً مرتب شده باشد گزینهٔ Sort the File… و در صورتی که فایل قبلاً مرتب نشده است گزینهٔ File is Already Sorted را می‌زنیم. ذکر این نکته ضرور ی است که بعد از تأئید فرمان Split File از طریق دکمهٔ OK هیچ تغییر خاصی ملاحظه نمی‌گردد بلکه این فرمان تنها پیش‌زمینه اجرای فرمانهای بعدی است و هر فرمانی که بعد از آن انجام گیرد نتایج را در سطوح مختلف

متغیری که مبنای Split است در خروجی نشان می‌دهد.

در صورتی که نمی‌خواهید عمل Split بر روی داده‌ها و خروجی‌ها انجام پذیرد باید آن را به حالت اولیه یعنی گزینه Analyze all Cases… برگردانده و با دکمهٔ OK تأئید نمائید.

-انتخاب کردن سطرها(از طریق زیر منو Select Cases):

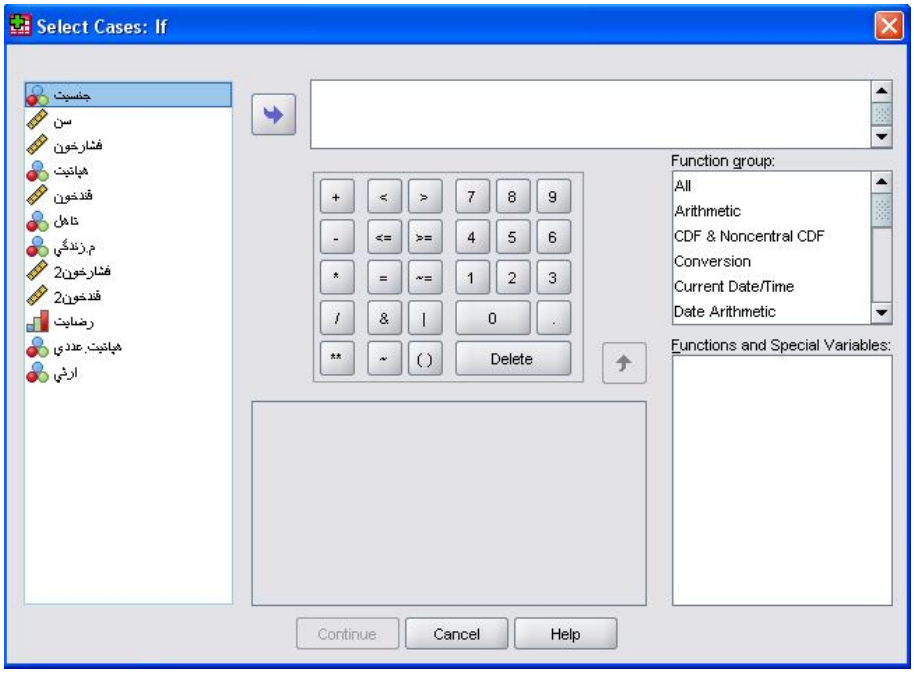
هنگامی که می‌خواهیم کلیه تجزیه و تحلیل‌های آماری نرم‌افزار SPSS تنها بر روی سطرهایی از متغیرها که دارای شرایط خاصی هستند انجام گیرد از این فرمان در منو Data برای انتخاب سطرهای با شرایط مورد نظر استفاده می‌کنیم. با اجرای فرمان Select Cases جعبهٔ گفتگوی زیر باز می‌شود:



برای مثال این فرمان را بر روی فایل Cars.sav اجرا می‌نمائیم. در قسمت سمت چپ این پنجره کلیه متغیرهای موجود در صفحه ویرایشگر داده‌ها نمایش داده می‌شوند و در قسمت سمت راست انواع روشهای انتخاب سطرها بیان شده است.

ـ گزینه All Cases که گزینهٔ اولیه این پنجره است بیان می‌کند که کلیه سطرها و داده‌ها در تجزیه و تحلیل‌های آماری شرکت نمایند.

ـ گزینهٔ If Condifion is Satisfied بیان می‌کند که می‌توان شرط خاصی را از طریق دکمهٔ If... بر روی داده‌ها اعمال نمود. در صورت فعال کردن این گزینه دکمهٔ If... فعال می‌شود و با انتخاب آن پنجرهٔ زیر ظاهر می‌شود:



برای مثال با وارد کردن متغیر Engine در قسمت پنجره سمت راست و اعمال شرط Engine>300 تنها سطرهایی را انتخاب می‌نمائیم که ظرفیت موتور در آنها بالاتر از مقدار 300 است و ماشینهای با

ظرفیت موتور کمتریا برابر مقدار 300 هر چند در پنجرهٔ Data View هستند اما در تجزیه و تحلیل‌های آماری از طریق زیرمنوهای Analyze شرکت نمی‌کنند. با اعمال شرط و زدن دکمه Continue و تأئید فرمان توسط دکمهٔ OK این شرط به نرم‌افزار اعمال می‌شود.

ـ گزینهٔ Random Sample of Cases یک نمونهٔ تصادفی از سطرها و داده‌ها را برای شرکت در تجزیه و تحلیل‌ها انتخاب می‌نماید. با فعال کردن آن گزینه Sample… فعال می‌گردد و با Click بر روی آن پنجرهٔ زیر باز می‌شود:

در پنجره بالا به دو صورت می‌توان یک نمونهٔ تصادفی از سطرها انتخاب کرد. در روش نخست یک درصد از کل داده‌ها انتخاب می‌شود و می‌توان عددی بین 1 تا 99 را در قسمت Approximatly وارد نمود.

در روش دوم که دو عدد در قسمت:

وارد می‌شود برای مثال با نوشتن عدد 100 در خانهٔ اول و عدد 300 در خانهٔ دوم تعداد 100 سطر را به صورت تصادفی از 300 سطر اول انتخاب می‌نماید.

ـ گزینهٔ Based on Time or Case range که با انتخاب آن دکمهٔ Range… فعال می‌شود و با Click بر روی آن پنجرهٔ زیر باز می‌شود:

در این پنجره برای نمونه با نوشتن عدد 10 در قسمت First Cases و عدد 50 در قسمت Last Cases تنها سطرهای 10 تا 50 را برای تجزیه و تحلیل انتخاب می‌نمائیم و نرم‌افزار فقط این 41 سطر را در تجزیه و تحلیل‌های آماری انتخاب می‌نماید.

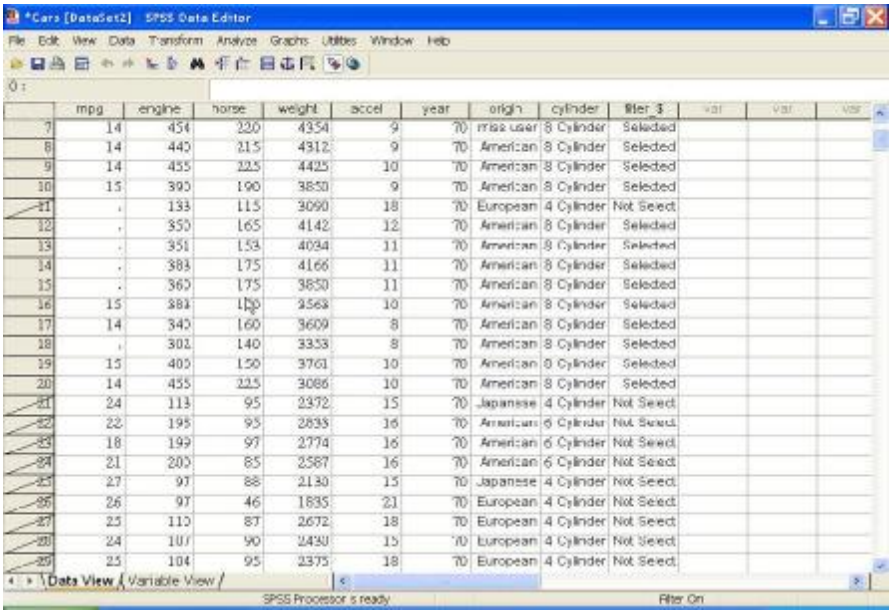
ـ گزینه Use Filter Variable این امکان را به کاربر می‌دهد که یک متغیر را در قسمت پائین آن وارد نماید. این متغیر انتخابی باید شامل داده‌های 0 و اعداد غیرصفر باشد. در این روش هر سطری که متغیر مورد نظر مقدار 0 دارد را انتخاب نمی‌کند و هر سطری را که متغیر مورد نظر عددی غیر از صفر را در خود جای داده انتخاب می‌نماید.

در قسمت Unselected Case Are نرم‌افزار نحوه برخورد با سطرهایی را که با یکی از روشهای بالا انتخاب نشده‌اند معین می‌نماید. با انتخاب Filtered این سطرها در پنجرهٔ Data View مشاهده می‌شوند و با خطی که بر روی شماره سطرهای آنها کشیده می‌شود، انتخاب نشدن آنها در تجزیه و تحلیل‌ها مشخص می‌گردد.

با انتخاب گزینهٔ Deleted تمامی سطرهایی که انتخاب نشده‌اند از پنجرهٔ نمایشگر داده‌ها حذف می‌شوند.

برای مثال نتیجهٔ انتخاب ماشین‌هایی که با روش If Condition is Satisfied و یا شرط Engine>300 در گزینهٔ If و انتخاب Filtered به وجود آمده در پنجرهٔ Data View به صورت ذیل

است:

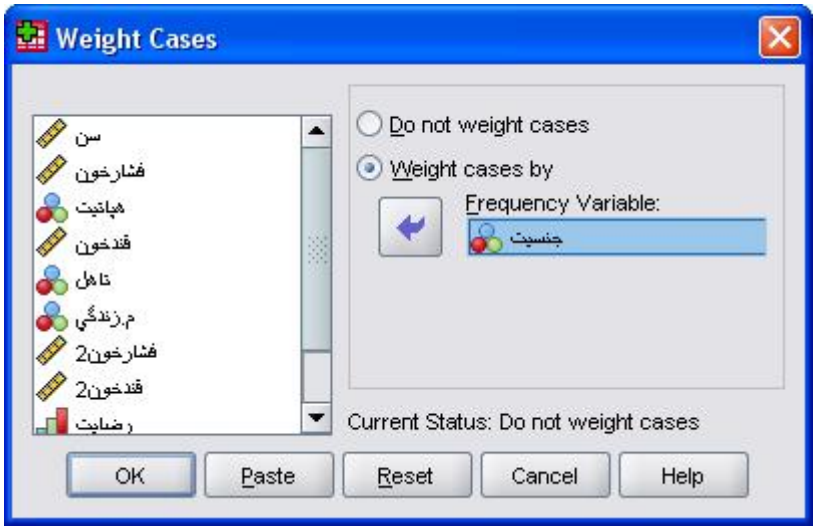


همانطور که ملاحظه می‌شود سطرهایی که مقدار Engine در آنها کمتر یا برابر عدد 300 بوده با علامت / بر روی شماره سطر آنها مشخص شده‌اند و همچنین متغیری با نام Filter_$ به مجموعهٔ متغیرها اضافه شده که هر سطر انتخابی با عدد 1 و هر سطر انتخاب نشده در آن با عدد 0 مشخص شده است.

حال اگر بر روی داده‌ها فرمانی نظیر Frequency یا هر فرمان دیگری از منو Analyze صورت پذیرد، تجزیه و تحلیل فقط بر روی این سطرهای انتخاب شده اعمال می‌شود.

-وزن‌دهی سطرها(از طریق زیر منو Weight Cases):

هر گاه بخواهیم اعداد یک متغیر را فراوانی خانه‌ها در نظر بگیریم می‌توانیم از این دستور استفاده نمائیم. با اجرای این فرمان در منو Data جعبهٔ گفتگوی زیر باز می‌شود:



اگر هیچ متغیری را به عنوان متغیر وزنی در نظر نگرفته باشیم گزینـهٔ Do Not Weight Cases و اگر متغیری را به عنوان فراوانی متغیر وزنی در نظر بگیـریم گزینـهٔ Weight Cases by را انتخـاب و آن متغیر را از قسمت سمت چپ وارد قسمت سمت راست می‌نمائیم.

برای مثال در داده‌های ذیل اگر X_3 را به عنوان متغیر وزنی وارد نموده باشیم به این معنـی اسـت کـه 40 خانم با تحصیلات لیسانس (سطر اول) و 30 نفر مرد با تحصیلات دیپلم (سطر دوم) و ...، اطلاعاتشـان وارد پنجرهٔ Data View گردیده است و اعداد 40 و 30 در حقیقت به معنی فراوانی سطرها می‌باشد کـه به دلیل جلوگیری از تکرار با استفاده از این فرمان به صورت زیر برای نرم‌افزار قابل درک است:

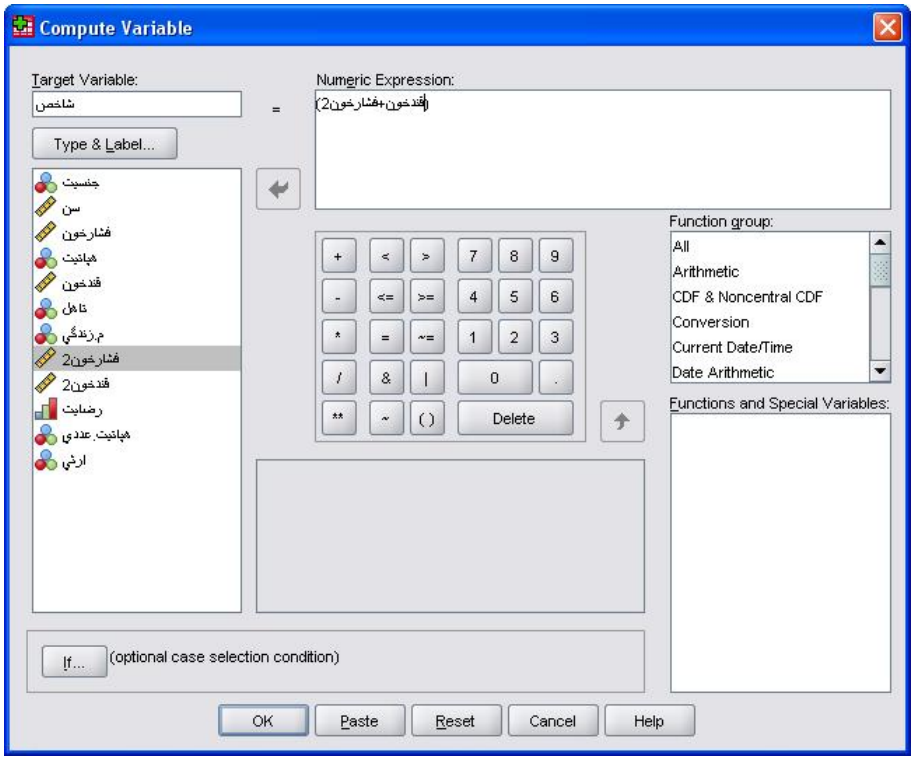
	x1	x2	x3
1	زن	لیسانس	40.00
2	مرد	دیپلم	30.00
3	مرد	فوق اسیانس	20.00
4	زن	لیسانس	45.00

5ـ3ـ منو Transform (تبدیل):

همانطور که قبلاً عنوان شد از زیر منوهای این منو برای انجام عملیات تبـدیل بـر روی داده‌هـا، انجـام عملیات ریاضی بر روی داده‌ها، تغییر کد دادن داده‌های عددی، جایگزاری مقادیر گـم شـده و ... اسـتفاده می‌شود. در ادامه تعدادی از مهم‌ترین ابزارهای این منو بیان گردیده است.

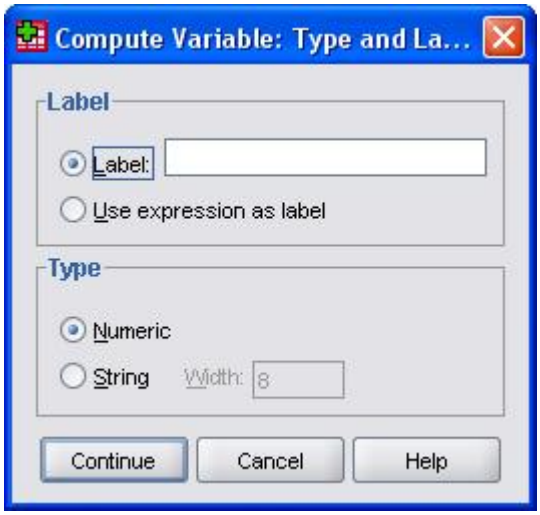
-انجام محاسبات ریاضی بر روی متغیرها(از طریق زیر منو Compute):

به منظور ایجاد متغیرهای جدید با اعمال محاسبات و گزاره‌های ریاضی بر روی متغیرهای فایل جـاری می‌توان از فرمان compute از منو Transform استفاده نمود. با اجرای این فرمان جعبهٔ گفتگوی زیر باز می‌شود:



در سمت چپ و بالای جعبه، Target Variable (نام متغیر مقصد) را وارد می‌نمائیم. به محـض ورود اولین کاراکتر، کلید Type & Label... (نوع و برچسب) برجسته شده و می‌توان نوع و بر چسب متغیـر جدید را تعریف کرد.

لازم به ذکر است که نام متغیر جدید می‌تواند یکی از متغیرهای جدید هـم باشـد امـا در ایـن حالـت متغیر قدیمی از بین خواهد رفت. با فعال کردن کلید Type & Label... جعبهٔ گفتگوی زیر باز می‌شود:



در بخش Label با انتخاب گزینه Label، بر چسب متغیر جدید را وارد می‌نمائیم. می‌توان برای به خاطر سپردن نوع محاسباتی که منجر به ایجاد متغیر جدید شده است عبارت ریاضی آن را نیز به عنوان برچسب معرفی کرد. در این حالت گزینه Use Expresion as Label یعنی «انتخاب عبارت به عنوان برچسب» را انتخاب می‌نمائیم و در قسمت Type یکی از انواع عددی (Numeric) و یا رشته‌ای (String) را همراه با تعداد کاراکترهای آن بسته به نوع عملیات انتخاب می‌نمائیم و برای تأئید فرایند دکمهٔ Continue را انتخاب می‌کنیم.

جعبهٔ Numeric Expression جعبه‌ای است که آماده دریافت متغیرها، علائم، توابع و شرایط منطقی و ترکیب آنهاست. علائم ریاضی مورد نیاز در وسط صفحهٔ قرار دارند و توابع آماری و ریاضی در قسمت Function Group قرار دارند.

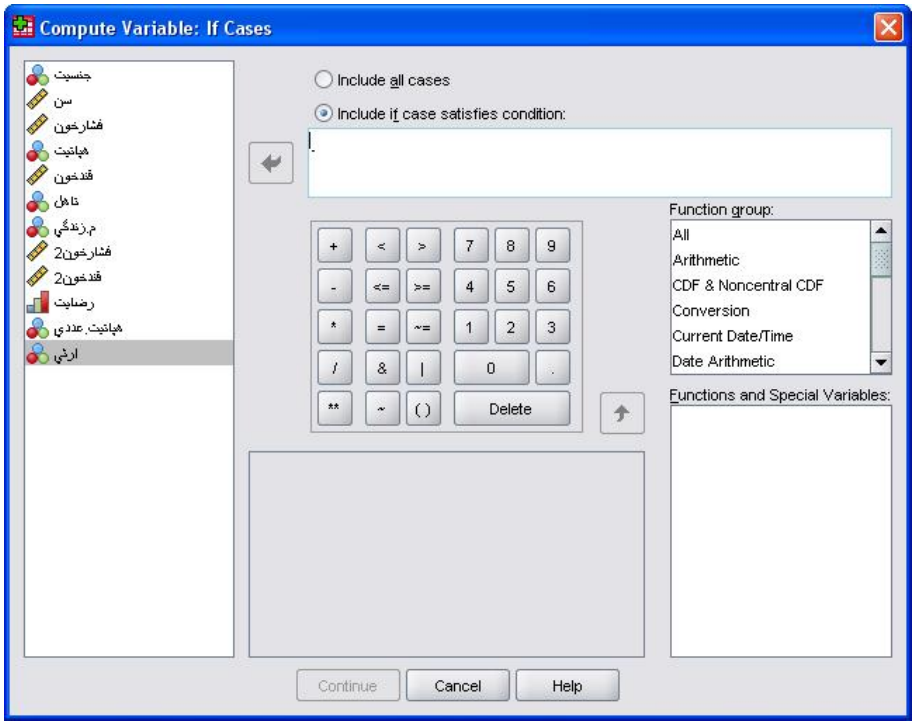
برای ساخت یک عبارت ریاضی از اجزای مختلف این صفحه استفاده می‌نمائیم. برای مثال اگر می‌خواهیم متغیر جدیدی با نام $S=[A_1+A_2^{12}\times\ln(A_3)]$ ایجاد کنیم باید هر یک از گزینه‌های زیر را به ترتیب وارد قسمت Numeric Expression نمائیم:

ابتدا] وسپس با Click کردن بر روی A_1 و 4 را وارد می‌کنیم، سپس +، A_2، ** (به معنی توان)،

12، 0 (?)lN ، A_3 و] ، حـــال در محـــیط Target Variable متغیـــر S و در قســـمت Numeric Expression عبارت:

$$[A_1 + A_2 **12 * LN(A_3)]$$

نوشته شده است. اگر می‌خواهیم این عملیات فقط بر روی مقادیری از یک متغیر که شرایط خاصـی را دارند اعمال شود می‌توان آن شرط را با دکمه If... بر روی نتایج اعمال کرد. با Click کردن بـر روی If پنجرهٔ زیر باز می‌شود:



فرض کنید می‌خواهیم عملیات ریاضی بر روی سطرهایی انجام پذیرد که مقدار متغیر A4 در هر یـک از سطرهای آن بزرگتر از عدد 2 باشد و این عمل را برای مقادیر A4 کمتر یا مساوی عدد 2 اعمال نکنـد در این صورت مانند بالا با انتخاب Include if Case Satisfies Condition، شرط A4>2 را می‌نویسم و دکمهٔ Continue را انتخاب می‌نمائیم تا به پنجرهٔ گفتگوی اصلی Compute بر گردیم. حال با استفاده

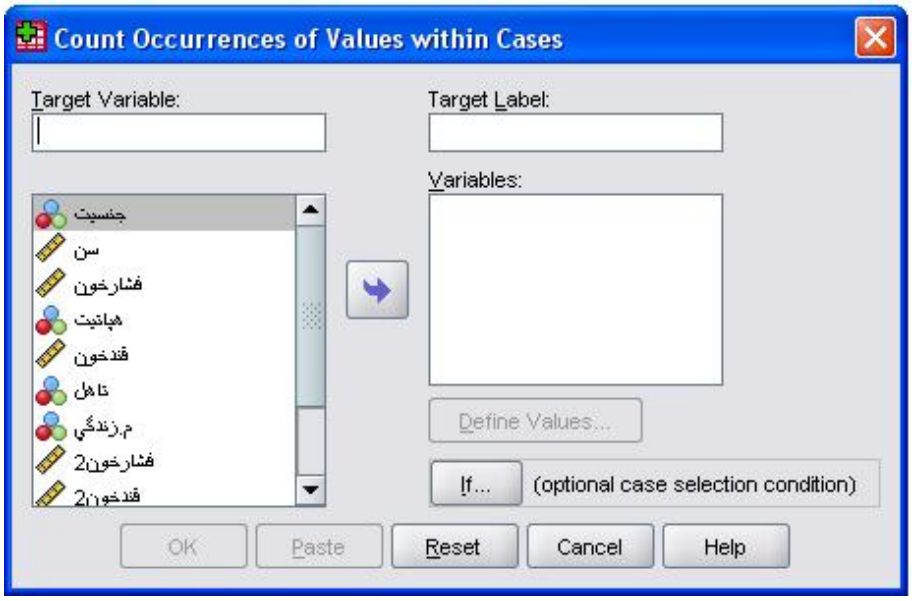
از دکمهٔ OK فرمان Compute را تأئید نمائیم.

در پنجرهٔ Data View متغیر S ظاهر می‌گردد که نتایج آن را ملاحظه می‌نمائید. همانطور که ملاحظه می‌شود این عملیات ریاضی تنها بر روی سطرهایی انجام پذیرفته است که شرط A4>2 را که از طریق دکمهٔ If... اعمال شد دارا بوده‌اند.

در صورتی که در جریان اجرای فرمان نام متغیر مقصد با یکی از نام‌های متغیرهای موجود یکسان باشد پیامی مبنی بر تأئید جایگذاری داده‌های جدید به جای داده‌های قدیم ظاهر می‌شود، که تنها در صورت تأئید جایگزاری، متغیر قدیم حذف می‌شود.

- شمارش تعداد داده‌های یک متغیر با شرایط خاص(از طریق زیر منو Count):

اگر بخواهیم مقادیر متغیرهایی را که دارای شرایط خاصی هستند شمارش کنیم، از فرمان Count در منو Transform استفاده می‌کنیم. با انتخاب این فرمان جعبه گفتگوی زیر باز می‌شود:



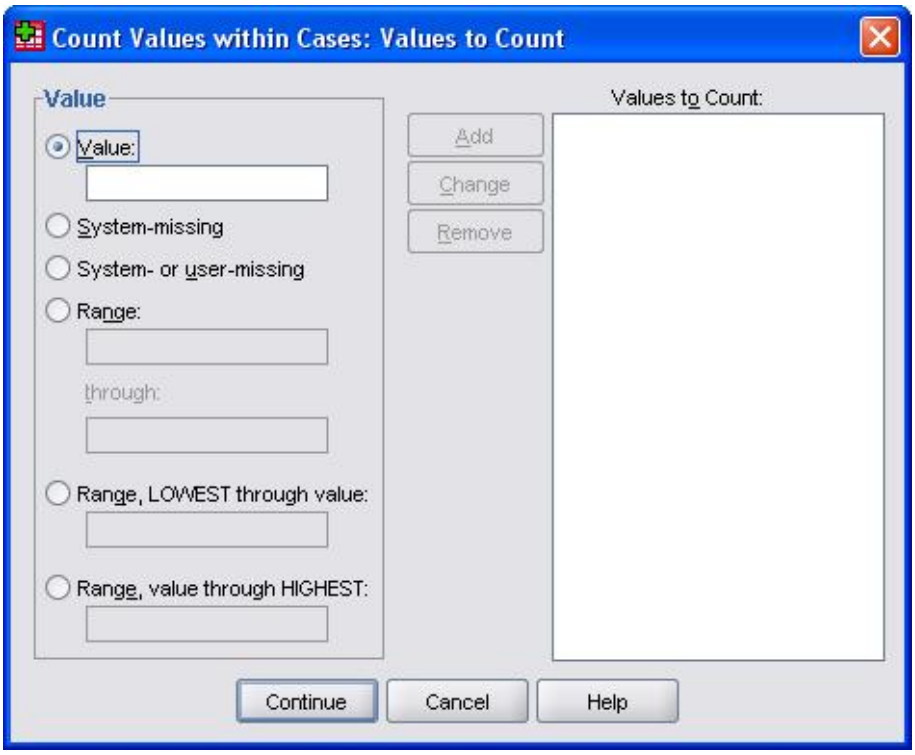
در جعبهٔ Target Variable نام متغیری که شمارشها در آن جای می‌گیرد، نوشته می‌شود و در جعبهٔ پائین آن فهرست تمامی متغیرهای موجود در فایل جاری نمایش داده می‌شوند. اگر بخواهیم به متغیر جدید برچسب بزنیم نام آن برچسب را در قسمت Target Labelوارد می‌نمائیم.

در جعبهٔNumeric Variable متغیرهایی را که قصد شمارش حالتهای مختلف آن را داریم وارد می‌نمائیم. این شمارش به دو صورت انجام می‌شود:

الف ـ ورود شرایط شمارش از طریق تعیین محدوده‌ای برای مقادیر متغیر.

ب ـ صدق کردن شرایط شمارش در عبارات منطقی (کلیدهای Define Values... و If...

با انتخاب کلید Define Values جعبهٔ گفتگوی دیگری به شکل زیر ظاهر می‌شود:



ـ اگر می‌خواهید عددی خاص را شمارش کنید بخش Value را انتخاب کرده عدد را داخل پنجرهٔ آن نوشته و دکمهٔ Add را می‌زنیم تا عدد در پنجرهٔ Values to Count قرار گیرد.

- گزینه System-Missing تعداد داده‌های گم شده سیستمی را که با جای . در متغیر موجود است شمارش می‌نماید همانند بالا با دکمهٔ Add می‌توان آن را وارد Values to Count نمود که شمارش شود.

ـ گزینهٔ System or User-Missing همهٔ انواع داده‌های گم شده، چه داده‌هایی را که گم شده سیستمی هستند و چه داده‌های گم شده تعریف شده توسط کاربر را همانند بالا با قرار دادن در Values to Count شمارش می‌کند.

ـ گزینهٔ Range یک فاصله که پائین و بالای آن را وارد کرده‌ایم و تعداد اعداد یک متغیر را که درون این فاصله است شمارش می‌نماید.

ـ انتخاب گزینهٔ Rang و باز شدن پنجرهٔ Lowest Though باعث شمارش هر عدد کوچکتر از عدد وارد شده در این قسمت می‌شود.

ـ انتخاب گزینهٔ Range و باز شدن پنجرهٔ Though Highest باعث شمارش هر عدد بزرگتر از عدد وارد شده در این قسمت می‌شود.

لازم به ذکر است بعد از انتخاب هر یک از روشهای بالا باید دکمهٔ Add انتخاب می کنیم تا نتایج در قالب پنجرهٔ Values to Count وارد گردد.

برای مثال در فایل زیر مقادیر کمتر از عدد 3 و بین 5 و 7 و مقادیر گم شده در متغیر A_2 و A_1 شمارش شده و نتایج آن در قالب متغیر جدید S بیان شده است:

	A1	A2	S
1	3.00	12.00	1.00
2	1.00	15.00	.00
3	2.00	9.00	.00
4	5.00	8.00	1.00
5	7.00	4.00	1.00
6	8.00	4.00	.00
7	9.00	12.00	.00
8	14.00	10.00	.00
9	12.00	11.00	.00
10	4.00	6.00	1.00
11	4.00	3.00	1.00
12	5.00	1.00	1.00

هر گاه بخواهیم مانند فرمان Count ، کلیه اقدامات فقط بر روی سطرهایی انجام گیرد که شرایط خاصی داشته باشند، می‌توان آن شرایط را از طریق دکمهٔ If... بیان نمود. نحوه انجام کار در پنجره ظاهر شده درست مانند قبل می‌باشد.

-تغییر کد داده‌ها(از طریق زیر منو Recode):

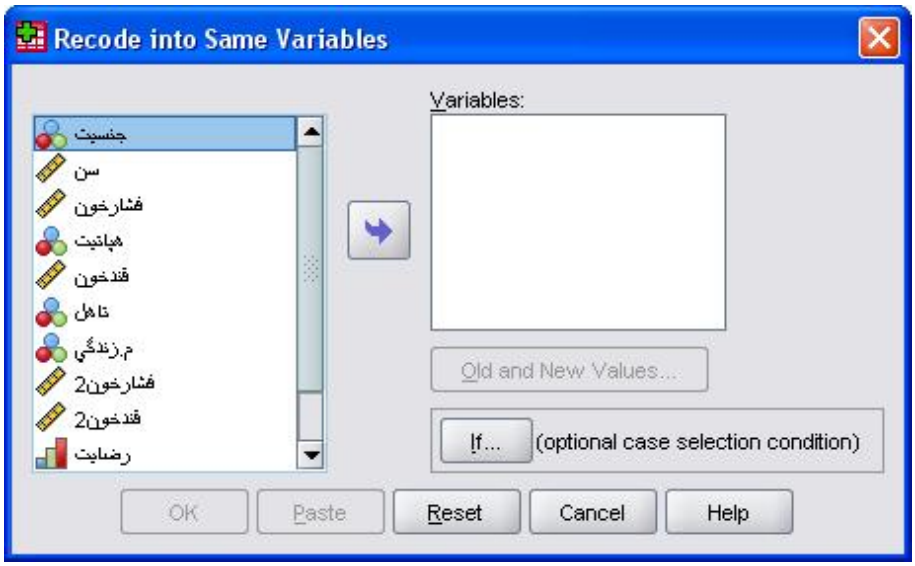
هر گاه کاربر بخواهد کدهای عددی یک یا چند متغیر را تغییر دهد می‌تواند از این فرمان در منو Transform استفاده نماید. برای مثال اگر سطح تحصیلات از زیر دیپلم تا لیسانس به ترتیب با کدهای 1 تا 4 مشخص شده است و کاربر می‌خواهد افراد را به دو گروه دارای سطح تحصیلات دیپلم و زیر دیپلم و افراد بالای سطح تحصیلات دیپلم تقسیم نماید لازم است که کدهای 1 و 2 را که معرف سطح تحصیلات زیر دیپلم و دیپلم است به کد 1 با برچسب تحصیلات دیپلم و پائین‌تر و کدهای 3 و 4 را که معرف سطح تحصیلات فوق دیپلم و لیسانس و بالاتر است را به کد 2 با برچسب سطح تحصیلات بالای دیپلم تغییر کد دهد.

به مثال دیگر در این مورد توجه کنید، معلم می‌خواهد دانش‌آموزان کلاس را بر اساس متغیر نمرهٔ آنها که عددی بین 0 تا 20 است به دو گروه زیر ده و بالا یا مساوی عدد ده تقسیم نماید. برای این منظور لازم است که یک تغییر کد انجام شود و به همهٔ نمرات پائین عدد ده کد عدد 1 با برچسب «نمره کمتر از

ده» و به همهٔ نمرات مساوی یا بالاتر از عدد ده کد عدد 2 با برچسب «نمره بیشتر یا برابر ده» داده شود در این صورت متغیر جدیدی ایجاد می‌شود که به جای نمرات پیوسته افراد دارای اعداد 1 و 2 می‌باشد. متغیر جدید ایجاد شده می‌تواند جایگزین متغیر قبلی گردد و یا یک متغیر جدید در پنجره داده‌ها ایجاد شود به همین دلیل عمل تغییر کد دارای دو حالت با نام‌های "Into Same Variable" به معنی جایگزینی در داخل همان متغیر و "Into different variable" یعنی جایگذاری در داخل متغیر جدید می‌باشد.

ـ تغییر کد و جایگذاری در همان متغیر(از طریق زیر منو Into Same Variable):

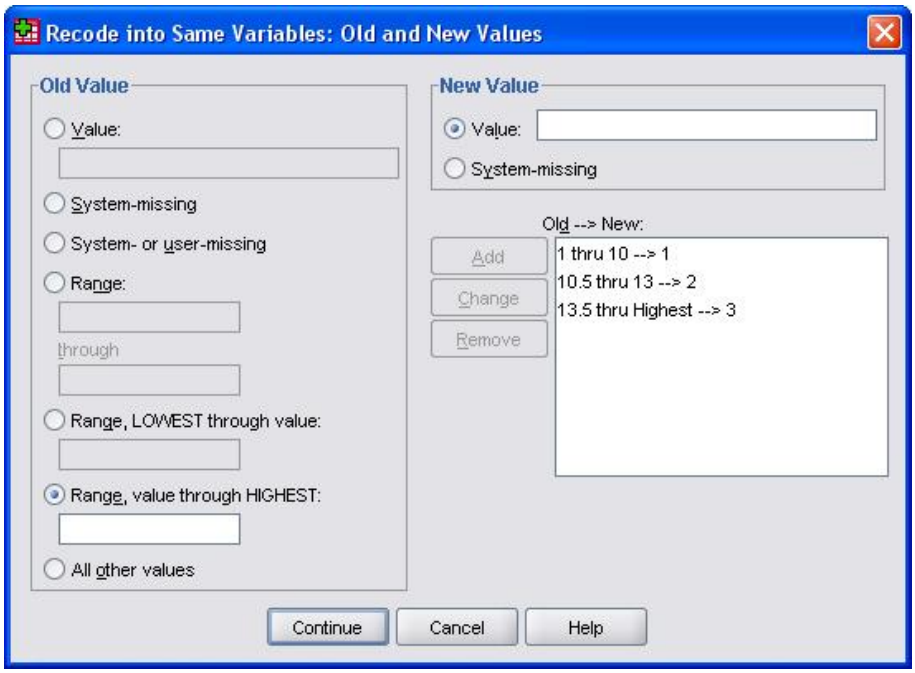
جهت اجرای این نوع کدگذاری دستور Into Same Variable را از زیر منو Recode در منو Transform انتخاب می‌کنیم، پنجره‌ایی به صورت ذیل باز می‌شود:



در قسمت سمت چپ فهرست متغیرهای موجود فایل جاری وجود دارد. در این حالت متغیرها یا متغیرهایی را که قصد تغییر کدهای آنها را داریم به جعبهٔ Numeric Variables وارد می‌نمائیم برای مثال در فایل Exam متغیر فشار خون را وارد این قسمت می‌نمائیم. کلید ... If امکان اعمال شرط بر

روی سطرها را فراهم می‌آورد و نحوه عمل این کلید همانند فرمانهای Compute و Count می‌باشد.

کلید Old and New Values… (مقادیر قدیم و جدید) امکان معرفی و جایگزینی مقادیر کدهای منتخب (NewValues) را با مقادیر قبلی(Old Values) فراهم می‌کند. در صورت فعال شدن این کلید، پنجرهٔ زیر باز می‌گردد. بخش Old Value شامل گزینه‌های مختلف معرفی مقادیر قبلی متغیرهاست که مقادیر کدهای جدید به آن منتسب می‌شود.



این گزینه‌ها می‌تواند Value (مقدار یک متغیر)، System Missing (مقادیر گمشده سیستمی)، System-or user-Missing (مقادیر گمشده سیستمی یا کاربری) و مقادیری در فاصله عددی معین باشند (مشابه دستور Count) تنها تفاوت در گزینهٔ All Other Values (سایر مقادیر) است که شامل تمامی مقادیری است که در گزینه‌های قبلی قرار نمی‌گیرند ولی قصد تغییر کد آنها را هم داریم.

برای اعمال یک تغییر کد، ابتدا نوع مقدار قبلی را از بخش Old Value انتخاب می‌کنیم، سپس مقدار

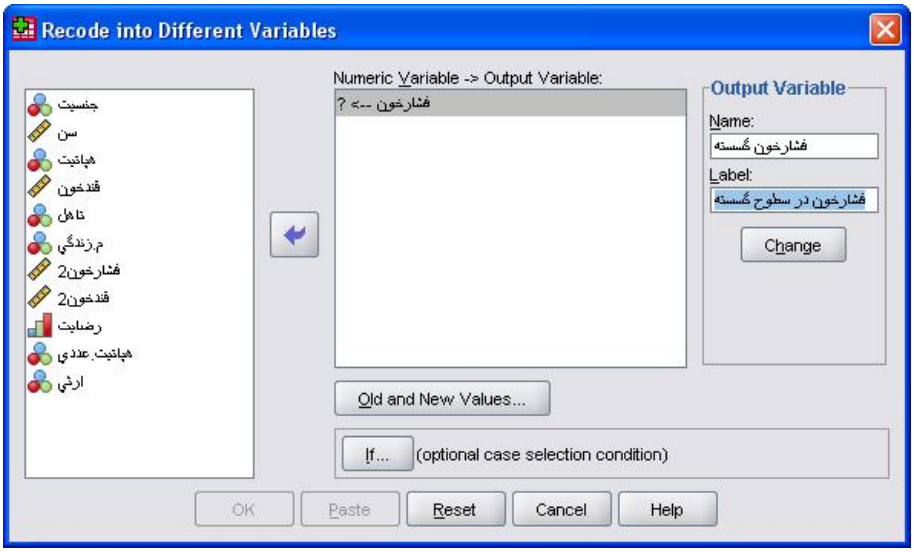
جایگزین آن را در جعبهٔ New Value انتخاب می‌نمائیم که می‌تواند یک مقدار معین Value یا Missing-System (مقدار گمشده) باشد. در صورت جایگزینی یک کد ثابت، مقدار آن را می‌توان در جعبهٔ مقابل گزینه Value وارد کرد و پس از تعیین مقدار کد جدید دکمهٔ Add را که برجسته شده است فعال نمود. با این عمل تغییرات مورد نظر در جعبهٔ Old→New جای می‌گیرد.

برای مثال از فایل Exam که حاوی اطلاعات افراد مبتلا به هپاتیت بود، متغیر فشار خون را انتخاب و وارد پنجرهٔ Numeric Variable می‌نمائیم و دکمهٔ…Old and New Values را فعال می‌نمائیم. با استفاده از گزینه Range در قسمت Old Value ،به همهٔ فشار خون‌های بین عدد صفر و 10 عدد 1 را در قسمت Range عدد بین 0 و 10 را مشخص می نماییم و در قسمت Value برای New Value عدد 1 را زده و دکمهٔ Add را می‌زنیم. به همین ترتیب فشار خون بین 10/5 تا 13 را به کد 2 و فشار خون بالای 13 را به کد 3 تبدیل نموده و دکمهٔ Continue و سپس دکمهٔ OK را برای اجرای فرمان Click می‌کنیم.

اگر به پنجرهٔ نمایش داده‌ها بروید متغیر فشار خون به جای اعداد پیوسته حاوی اعداد 1، 2 و 3 می‌باشد. در حقیقت با انجام این فرآیند متغیر پیوسته فشار خون تبدیل به یک متغیر گسسته ترتیبی با 3 سطح شده است که افراد با کد عدد 1 دارای فشار خون پائین، افراد با کد عدد 2 دارای فشار خون طبیعی و افراد با فشار خون با کد 3 دارای فشار خون بالا می‌باشند.

ـ تغییر کد در داخل یک متغیر جدید(از طریق زیر منو Into Different Variables:

هر گاه بخواهیم بر روی کدهای یک متغیر تغییر کد ایجاد نمائیم و نتایج آن را بر خلاف حالت قبل در داخل یک متغیر جدید قرار دهیم از این دستور استفاده می‌نمائیم. با فرا خواندن این دستور پنجره‌ایی مانند شکل زیر باز می‌شود:



همانطور که ملاحظه می‌شود دکمهٔ Old and New Values… و … If درست مانند حالت قبل عمل می‌نمایند تنها تفاوت این دستور در جایگزاری نتایج در داخل یک متغیر جدید است. ابتدا متغیری را که می‌خواهیم تغییر کد بر روی آن اعمال گردد به پنجرهٔ

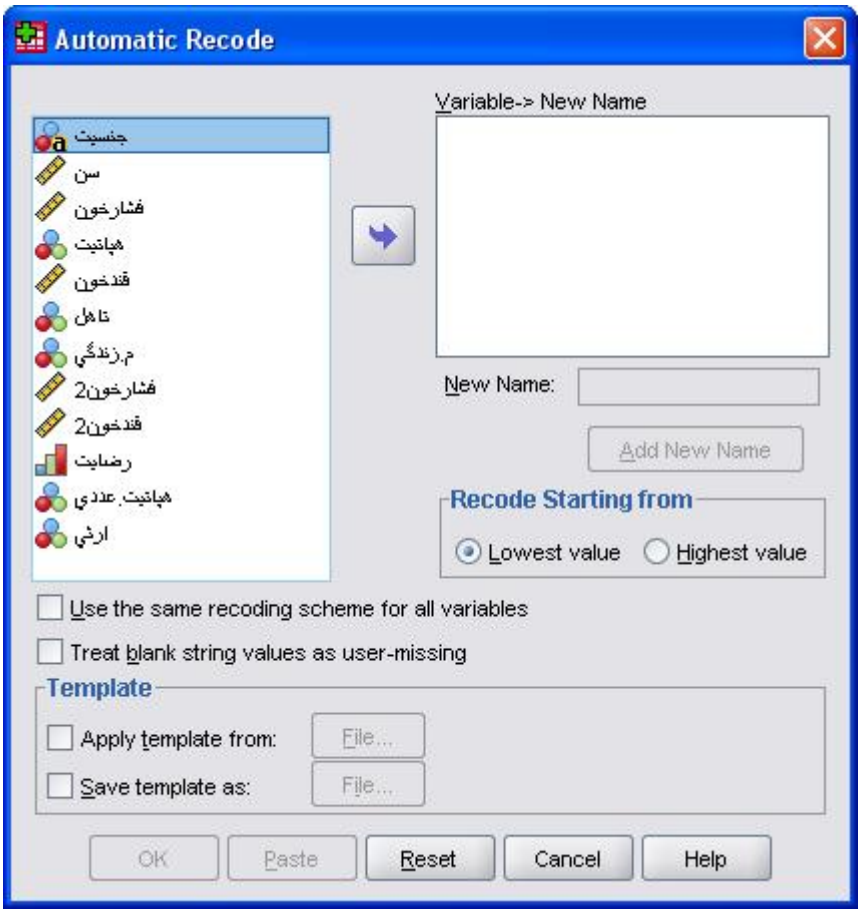
Input Variable ® Output Variable

وارد می‌نمائیم و در آن صورت پنجرهٔ Output با دو گزینهٔ Name و Lable در قسمت بالای سمت راست پنجره باز می‌شود که باید نام و برچسب متغیر جدید را در آن وارد کرده و دکمهٔ Change را انتخاب می کنیم تا نام متغیر در جلو نام متغیر انتخابی در پنجرهٔ Input Variable→Output Variable قرار گیرد. حال می‌توان مانند حالت قبل مقادیر قدیم و جدید را از پنجرهٔ Old and New Values… تعیین نمود. پس از اجرای این فرمان کدهای جدید درون متغیری جدید با نامی که درون پنجرهٔ Name وارد شده است به عنوان یک متغیر جدید در پنجرهٔData View ظاهر می‌شود.

-تغییر کد خودکار(از طریق زیر منو Auto Recode):

بسیاری از پنجره‌ها در نرم‌افزار SPSS تنها برای متغیرهایی که از نوع عددی هستند فعال هستند و برای متغیرهای رشته‌ای (String) عمل نمی‌کند بهتر است همواره متغیرهای با کدهای عددی وارد نرم‌افزار شوند.

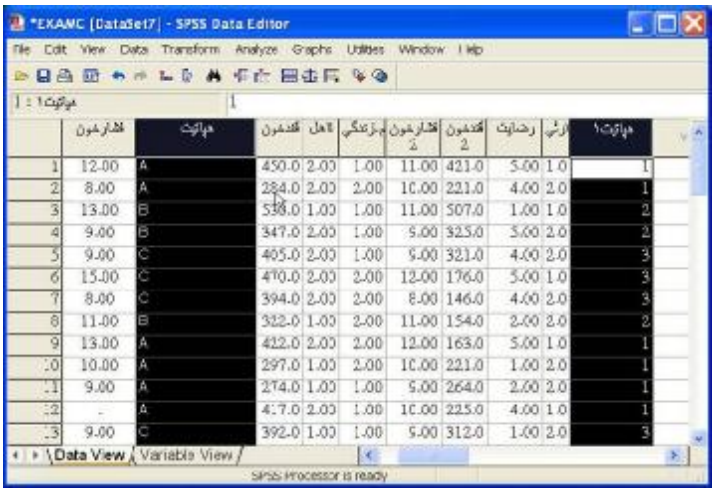
این فرمان امکان تغییر کد متغیرهای String (رشته‌ای) را به متغیرهای Numeric (عددی) به صورت خودکار فراهم می‌آورد. با فراخواندن این زیرمنو از منو Transform پنجرهٔ زیر باز می‌شود:



ابتدا متغیر از نوع String را وارد پنجرهٔ Variable → New Name می‌نمائیم و سپس نام متغیر جدیدی را که می‌خواهیم تغییر کد درون آن قرار گیرد در قسمت New Name وارد کرده و دکمهٔ Add New Name را انتخاب می کنیم. حال اگر می‌خواهیم کدگذاری از کوچکترین مقدار صورت پذیرد

گزینهٔ Lowest Value را در Recode Starting From و اگر می‌خواهیم از بزرگترین مقدار صورت پذیرد گزینهٔ Highest Value را فعال می‌نمائیم.

برای مثال متغیر نوع هپاتیت در فایل Exam که داده‌های آن در پایان کتاب موجود است یک متغیر String با کدهای A، B و C می‌باشد، در صورتی که این متغیر را وارد Variable→New Name نموده و یک متغیر جدید مثلاً با نام هپاتیت انتخاب نمائیم و دکمهٔ Add New Name را انتخاب می کنیم و گزینهٔ Lowest Value را فعال کرده و دکمهٔ OK را انتخاب می کنیم. در پنجرهٔ ویرایشگر داده‌ها متغیر هپاتیت 1 به صورت زیر ظاهر می‌شود:



همانطور که ملاحظه می‌شود در هر سطر متغیر هپاتیت عدد A نوشته شده در متغیر هپاتیت1به کد 1، هر سطر نوع هپاتیت B بوده در متغیر هپاتیت1 کد 2 و هر چه سطر نوع هپاتیت C باشد در متغیر هپاتیت1 کد 3 منظور شده است.

نکته جالب توجه آن است که در تغییر کدها، برچسب مقادیر نیز منتقل می‌شود. برای مثال کد 1 در متغیر هپاتیت1 دارای همان برچسب مقادیر A در متغیر نوع هپاتیت خواهد بود.

خلاصه مباحث مطرح شده:

در این فصل ابزارهایی مورد بررسی قرار گرفت که از نظر کاربردی دارای اهمیت ویژا ایی می باشند. اهمیت استفاده از این ابزارها در آماده کردن داده ها برای استفاده از نرم افزار می باشد. در این فصل زیر منوهای منو Data و Transform مورد بررسی قرار گرفت از مهمترین ابزارهای این منوها می توان به ابزارهایی برای ویرایش داده ها نظیر جابجایی سطرها و ستونها، ابزارهای برای فیلتر کردن داده ها و عدم ظاهر شدن آنها در محاسبات، چند تکه کردن خروجی ها بر حسب سطوح یک متغیر گسسته ، تغییر کد متغیرهاو .. اشاره کرد.

تمرین:

1- با استفاده از امکانات recode در داخل یک متغیر دیگر در داخل فایل مبتلایان به هپاتیت، سن بیماران را که یک متغیر پیوسته است به متغیر گسسته با طول رده های 10 ساله تبدیل نمائید.

2- متغیرهای مربوط به فایل cars را با استفاده از دستور marge files به متغیرهای مربوط به employe data اضافه نمائید.

3- با استفاده از دستور split files و descriptive اطلاعات متغیرهای پیوسته فایل cars را به تفکیک متغیر گسسته منطقه تولید ماشین بدست آورید.

4- اطلاعات متغیرهای پیوسته فایل مبتلایان به هپاتیت را با استفاده از دستور select case فقط برای افراد روستایی بدست آورید.

5- با استفاده از دستور Autorecode متغیر هایی را که غیر عددی هستند در فایل employe dataتبدیل به متغیر عددی نمائید.

6- اطلاعات مربوط به فشار خون بعد را از فشار خون قبل کم نموده و با تقسیم نمودن بر فشار

خون قبل میزان (درصد) تغییر فشار خون را برای هر نفر در اثر استفاده از داروی جدید بدست آورید

7- متوسط تغییر میزان قند خون را در اثـر مصـرف داروی جدیـد بـر روی بیمـاران مبتلایـان بـه هپاتیت بدست آورید.

8- ماکزیمم مقدار تغییر قند خون در اثر مصرف داروی جدید برای چه فردی با چه سنی به وجود آمده است؟

فصل ششم

اهداف آموزشی فصل ششم

فراگیر پس از مطالعه این فصل باید بتواند:

- مراحل اجرای فرمان جدول توافقی در نرم افزار Spss.16 را بشناسد.
- مراحل انجام فرمان جدول توافقی در نرم افزار Spss.16 را با ذکر مثال شرح دهد.
- ابزار تحلیل جداول توافقی با داده های اسمی و ترتیبی را بشناسد.
- مراحل انجام تحلیل جداول توافقی با داده های اسمی و ترتیبی را با ذکر مثال نشان دهد.
- مراحل انجام نمایش مقادیر خانه های جدول توافقی را با ذکر مثال توضیح دهید.

جداول توافقی

6ـ1ـ مقدمه:

جداول توافقی ابزارهایی از آمار استنباطی هستند که به وسیله آنها تأثیر یک متغیر گسسته (اسمی یا ترتیبی) را بر روی یک متغیر گسسته (اسمی یا ترتیبی) در حالت دو بعدی بدست می‌آوریم.

متغیری که متغیر مستقل یا تأثیرگذار است را متغیر سطری (Row Variable) و متغیری را که متغیر وابسته یا تأثیرپذیر است متغیر ستونی (Column Variable) می‌نامند. ذکر این نکته ضروری است که هر گاه میان دو متغیر گسسته رابطه وجود نداشته باشند در حقیقت آنها از یکدیگر مستقل هستند یا در اصطلاح آماری دارای مقدار ضریب همبستگی (R) برابر عدد صفر هستند و هر گاه میان آنها رابطه وجود داشته باشد یا در اصطلاح متغیر گسسته مستقل بر روی متغیر گسسته وابسته تأثیرگذار باشد مقدار همبستگی برابر $R \neq 0$ خواهد بود.

در حقیقت نتیجه یک تجزیه و تحلیل جدول توافقی باید با نتیجه آزمون فرض $H_0 : R = 0$ در مقابل

$H_1 : R \neq 0$ برابر باشد. زیرا اگر فرض H_0 رد شود نشان‌دهندهٔ تأثیرگذاری متغیر گسسته مستقل بر روی متغیر وابسته و در صورت رد نشدن فرض H_0 ،نشان‌دهندهٔ عدم تأثیرگذاری یا عدم وجود رابطه میان متغیر گسسته مستقل و متغیر گسسته وابسته خواهد بود.

6ـ2ـ رابطهٔ میان متغیرهای گسسته(از طرق زیر منو...Crasstabs):

برای اجرای فرمان جدول توافقی ابتدا وارد منو Analyze و سپس وارد زیر منو Descriptive Statistics می‌شویم، بعد از انتخاب گزینهٔ ...Crosstabs پنجرهٔ گفتگوی زیر باز می‌شود:

در قسمت سمت چپ این پنجره لیستی از متغیرهای فایل جاری مشاهده می‌شود. متغیر مستقل را وارد قسمت Row (s) و متغیر وابسته را وارد قسمت Column (s) می‌نمائیم. برای مثال می‌خواهیم تأثیر جنسیت را بر روی نوع هپاتیت بدست آوریم به صورت بالا عمل می‌نمائیم و دکمهٔ OK را فعال

می‌نمائیم. خروجی به صورت ذیل است.

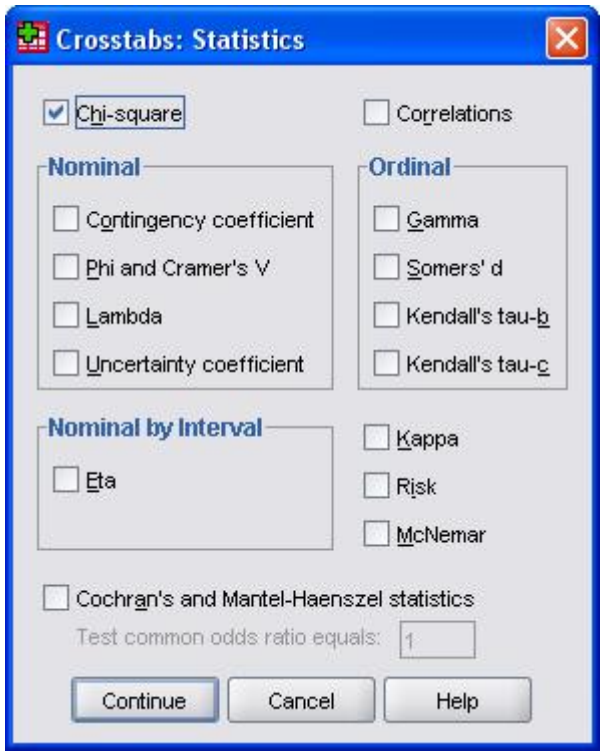
جنسیت * هپاتیت Crosstabulation

Count

		هپاتیت			Total
		نوع هپاتیت A	نوع هپاتیت B	نوع هپاتیت C	
جنسیت	مرد	۱	۱	۵	۱۶
	زن	۵	۴	۸	۱۷
Total		۶	۵	۱۳	۳۳

همانطور که ملاحظه می‌شود در این مرحله تنها جدول توافقی میان دو متغیر در خروجی به صورت بالا ظاهر می‌شود و معیاری برای قضاوت در مورد تأثیر متغیر جنسیت بر روی متغیر نوع هپاتیت ظاهر نمی‌شود.

معیاری عمودی که توسط آن در مورد رابطهٔ میان دو متغیر گسسته در جداول توافقی قضاوت می‌شود به آماره χ^2 یا آمارهٔ خی دو و یا پیرسون کای اسکور معروف است که با انتخاب دکمهٔ ... Statistics و انتخاب Chi-Square آن را فعال می‌نمائیم



سپس دکمهٔ Continue و سپس OK را انتخاب می کنید تا خروجی ذیل به خروجی بالا اضافه گردد.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.279[a]	2	.870
Likelihood Ratio	.279	2	.870
N of Valid Cases	30		

a. 5 cells (83.3%) have expected count less than 5. The minimum expected count is 5.47.

همانطور که ملاحظه می‌شود در قالب جدول بالا با نام Chi-Square Tests مقدار آمارهٔ Pearson Chi-Square در ستون Value و با درجه آزادی در ستون df بیان شده است اما عدد مورد نظر برای رد کردن یا پذیرفتن تأثیر متغیر گسسته که در آن Row (s) وارد شده بر روی متغیری که در پنجرهٔ Column (s) وارد شده است ستون سوم یعنی Asymp.sig(2-sided) است.

این مقدار که به سطح معنی‌داری دو طرفه یا به طور خلاصه به Sig معروف است عددی است بین 0 و 1 و معیاری برای رد کردن یا پذیرفتن فرض $H_0 : R = 0$ می‌باشد. همانطور که گفته شد در صورتی که فرض $H_0 : R = 0$ رد نشود یعنی میان دو متغیر استقلال وجود دارد و اگر $H_0 : R = 0$ رد شود به این معنی است که میان دو متغیر وابستگی وجود دارد و یا متغیری که در Row (s) وارد نموده‌ایم بر روی متغیری که در Column (s) وارد نموده‌ایم تأثیرگذار است.

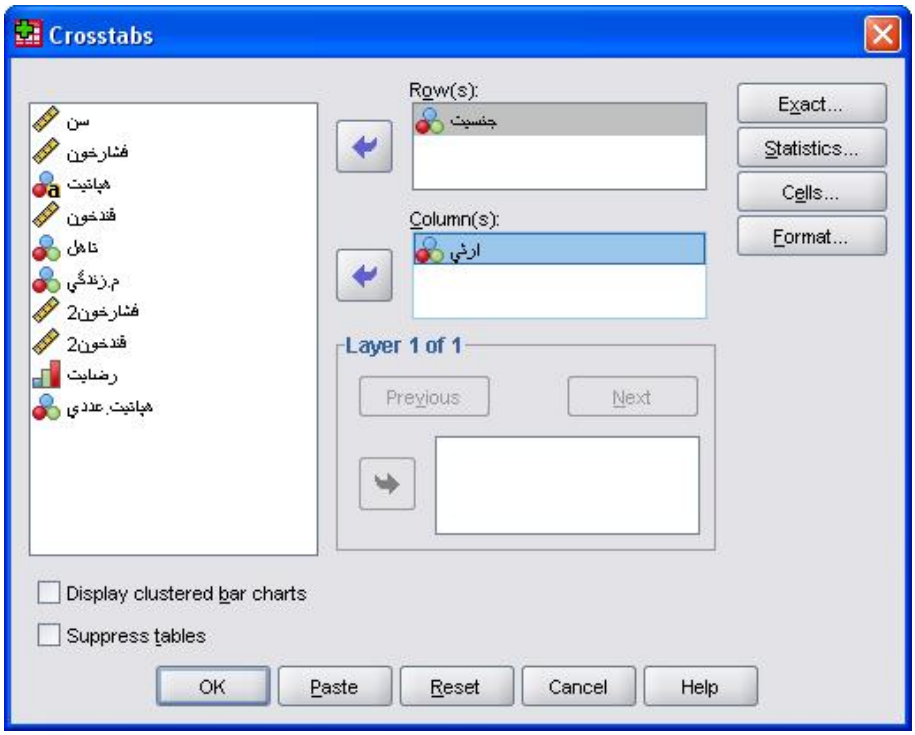
اگر مقدار sig که عددی بین 0 و 1 است از مقدار α (خطای نوع اول تحقیق آماری) بیشتر باشد ($sig>\alpha$) آنگاه فرض H_0 رد نمی‌شود یا در اصطلاح میان دو متغیر استقلال وجود دارد.

اگر مقدار sig که عددی بین 0 و 1 است از مقدار α (خطای نوع اول تحقیق آماری) کوچکتر یا مساوی باشد ($sig\leq\alpha$) آنگاه فرض H_0 رد می‌شود یا در اصطلاح میان دو متغیر همبستگی وجود دارد و یا متغیر مستقل بر روی متغیر وابسته تأثیرگذار است.

در مثال مورد نظر ما یعنی تاثیر جنسیت افراد بر روی نوع هپاتیت مبتلایان مؤثر است و یا افراد با جنسیت خاص بیشتر به نوع خاصی از هپاتیت مبتلا می‌شوند. برای کسب اطلاعات بیشتر فایل Exam را

با متغیرهای جدید وضعیت تأهل، محل زندگی (روستا یا شهر)، میـزان فشـار خـون بعـد از مصـرف دارو، میزان قند خون بعد از مصرف دارو، میزان رضایت از زندگی، متغیر عددی نوع هپاتیـت و ارثـی بـودن یـا نبودن هپاتیت تجهیز نمودهایم. [دادهها در انتهای کتاب موجود میباشند].

برای مشاهدهٔ رابطهٔ میان دو متغیر گسسته، رابطهٔ میان متغیر جنسیت و ارثی بودن یا نبودن بیماری را به طریق زیر بدست میآوریم.



از جعبهٔ Statistics عبارات Chi-Square و Correlation را انتخاب وسـپس Continue را فعـال و بعد از آن OK مینمائیم. خروجی به صورت زیر ظاهر میشود:

جنسیت * ارثی Crosstabulation

Count

		ارثی		Total
جنسیت	مرد	17	3	19
	زن	2	15	17
Total		19	18	36

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	14.043[a]	1	.000		
Continuity Correction[b]	10.454	1	.001		
Likelihood Ratio	14.053	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	13.937	1	.000		
N of Valid Cases	80				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.20.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Interval by Interval	Pearson's R	.419	.139	4.637	.000
Ordinal by Ordinal	Spearman Correlation	.416	.149	4.647	.000
N of Valid Cases		80			

همانطور که در جدول Chi-Square ملاحظه می‌شود آماره‌های:

Pearson Chi-Square

Continuity Correlation

Fisher's Exact test ،Likeli hood Ratio

Liner-by-liner Association

هر یک معیارهایی برای قضاوت در مورد استقلال یا وابستگی جنسیت و ارثی بودن است اما روش Pearson Chi-Square از کلیه روشهای بالا قوی‌تر و با دقت بیشتری است مگر در مواقعی که درجه آزادی (df) برابر عدد 1 باشد. در این گونه موارد روش آزمون دقیق فیشر (Fisher Exact Test) از دقت بیشتری برخوردار است و معیار قضاوت در مورد استقلال متغیرها ($H_0 : R = 0$) مقدار دقیق سطح معنی‌داری[Exact sig(2-sided)] است که در این تحقیق برابر مقدار 0/001 است که با در نظر گرفتن α=0/05 و در حالی که sig<α است فرض $H_0 : R = 0$ رد می‌گردد یعنی میان جنسیت و ارثی بودن بیماری رابطه وجود دارد.

جدول Symetric Measures مقدار R را هنگامی که فرض برابری آن با عدد صفر رد می‌شود بیان

می‌کند. ذکر این نکته ضروری است که $-1 \leq R \leq 1$ است و هنگامی که $R > 0$ است یعنی بین هر دو متغیر رابطه مستقیم وجود دارد و اگر $R < 0$ باشد میان دو متغیر رابطه معکوس وجود دارد.

در جدول آخر در صورتی که تعداد داده‌ها بیش از 30 عدد باشد معمولاً از روش Pearson' R و در صورتی که تعداد داده‌ها کمتر از مقدار فوق باشد از روش Spearman Correlation استفاده می‌نمایند. ذکر این نکته ضروری است که در داده‌های با حجم زیاد این دو مقدار بسیار به هم نزدیک هستند.

در مثال رابطهٔ میان جنسیت و ارثی بودن بیماری هپاتیت از روش Pearson'R استفاده می‌نمائیم و در نتیجه مقدار R=0/659 است و چون $R > 0$ است به این معنی است که با افزایش کد متغیر جنسیت [مرد=1 و زن=2] کدهای متغیر ارثی بودن [1=ارثی، 2=ارثی نبودن] افزایش می‌یابد و به عبارت دیگر آقایان با کد 1 بیشتر از طریق ارثی [1] و خانم‌ها با کد 2 بیشتر از طریق غیرارثی [2] مبتلا به هپاتیت می‌گردند.

هر گاه مانند شکل ذیل در هر یک از خانه‌های Row (s) و Column (s) بیش از یک متغیر وارد شده باشد:

نرم‌افزار رابطه کلیه متغیرهای موجود در Row (s) را با کلیه متغیرهای موجود Column (s) بدست می‌آورد در حقیقت در اینجا 4 آزمون همبستگی میان متغیرهای گسسته انجام می‌پذیرد.

با انتخاب گزینهٔ Display Clustered bar Charts نمودار خوشه‌ای سطوح مختلف متغیر Row (s) در سطوح خوشه‌های متغیر Column (s) ارائه می‌شود و با انتخاب گزینهٔ Suppress tables در خروجی جدول توافقی متغیرها نمایش داده نمی‌شود.

ذکر این نکته ضروری است که برای گرفتن جداول Chi-Square Test و Symetric Measure باید گزینه‌های Chi-Square و Correlations از پنجرهٔ Statistics فعال باشند.

اگر همانند شکل ذیل یک متغیر گسسته وارد قسمت Layer نمائیم آنگاه رابطهٔ میان متغیرهای گسسته‌ای که در پنجرهٔ Row (s) و column (s) وارد نموده‌ایم در هر یک از سطوح این متغیر محاسبه می‌شود.

Chi-Square Tests

		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
روستا	Pearson Chi-Square	4.254	1	.039		
	Continuity Correction	3.405	1	.065		
	Likelihood Ratio	5.231	1	.022		
	Fisher's Exact Test				.050	.029
	Linear-by-Linear Association	4.053	1	.032		
	N of Valid Cases	32				
شهر	Pearson Chi-Square	9.479	1	.002		
	Continuity Correction	6.285	1	.012		
	Likelihood Ratio	10.910	1	.001		
	Fisher's Exact Test				.005	.005
	Linear-by-Linear Association	8.752	1	.003		
	N of Valid Cases	38				

Symmetric Measures

			Value	Asymp. Std. Error	Approx. T	Approx. Sig.
روستا	Interval by Interval	Pearson's R	.357	.156	2.157	.039
	Ordinal by Ordinal	Spearman Correlation	.357	.156	2.157	.039
	N of Valid Cases		32			
شهر	Interval by Interval	Pearson's R	.551	.138	3.913	.000
	Ordinal by Ordinal	Spearman Correlation	.551	.138	3.913	.000
	N of Valid Cases		38			

همانطور که ملاحظه می‌شود رابطهٔ میان جنسیت و ارثی بودن بیماری یک بار برای افراد ساکن روستا و مقدار sig=0/05 فرض وجود همبستگی را تأئید می‌نماید و برای افراد ساکن شهر نیز sig=0/005 فرض

همبستگی تأئید می‌شود. (چون df=1 است از آمارهٔ Fisher Exat Test استفاده می‌نمائیم).

حال با توجه به R=0/537 برای افراد ساکن روستا و R=0/854 برای افراد ساکن شهر می‌توان قضاوت نمود که چه در شهر و چه در روستا آقایان بیش از خانم‌ها از راه ارثی مبتلا به هپاتیت می‌شوند.

برای روشن شدن موضوع از طریق این زیر منو رابطهٔ میان متغیر گسسته وضعیت تأهل و میزان رضایت از زندگی را در بین بیماران هپاتیتی بدست می‌آوریم،

وضعیت تأهل دارای دو سطح و میزان رضایت یک طیف لیکرت 5 سطحی است که از گزینهٔ بسیار کم، کم، متوسط، زیاد و خیلی زیاد با کدهای 1 تا 5 تشکیل شده است. فرض اولیه $H_0 : R = 0$ نشان‌دهنده استقلال و فرض جایگزین $H_1 : R \neq 0$ نشان‌دهنده وابستگی است.

از دکمهٔ ... Statistics گزینهٔ Chi-Square و Correlation را برای گرفتن جداول Chi-Square Test و Symetric Measure در خروجی انتخاب کرده، Continue را فعال و سپس OK می‌نمائیم. خروجی به صورت زیر ارائه می شود:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
تاهل * رضایت	30	100.0%	0	.0%	30	100.0%

جدول Case Processing تعداد و درصد داده‌های معتبر و گم شده را نشان می‌دهد.

Crosstabulation تاهل * رضایت

Count

		رضایت					Total
		خیلی کم	کم	متوسط	زیاد	خیلی زیاد	
تاهل	مجرد	6	5	2	1	1	15
	متاهل	3	0	1	5	6	15
Total		9	5	3	6	7	30

جدول Crosstabulation جدول توافقی میان وضعیت تأهل * میزان رضایت از زندگی را نشان می‌دهد.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.571	4	.014
Likelihood Ratio	15.164	4	.004
Linear-by-Linear Association	8.088	1	.004
N of Valid Cases	30		

جدول Chi-Square Test جدولی است که با اطلاعات موجود در مورد $H_0 : R = 0$ قضاوت می‌شود. با توجه به اینکه sig=0.014 برای Pearson Chi-Square است و با در نظر گرفتن $\alpha = 0/05$ می‌توان مشاهده کرد که $sig<\alpha$ پس‌فرض $H_0 : R = 0$ یعنی استقلال میان متغیرها رد می‌شود. و متغیرها به هم وابسته هستند.

Symmetric Measures

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Interval by Interval	Pearson's R	.528	.152	3.291	.003
Ordinal by Ordinal	Spearman Correlation	.498	.157	3.042	.005
N of Valid Cases		30			

در جدول Symmetric Measure در صورتی که $H_0 : R = 0$ رد شده باشد می‌توان جهت R یعنی $R > 0$ یا $R < 0$ را مشخص کرد با استفاده از 'R Pearson مقدار R=0/528 را ملاحظه می کنیم که عددی مثبت است و به این معنی است که میان وضعیت تأهل و رضایت از زندگی رابطه مستقیم وجود دارد و چون تأهل (مجرد=1 و متأهل =2) دارای 2 کد است می‌توان نتیجه گرفت که افراد متأهل دارای رضایت بیشتری از زندگی هستند.

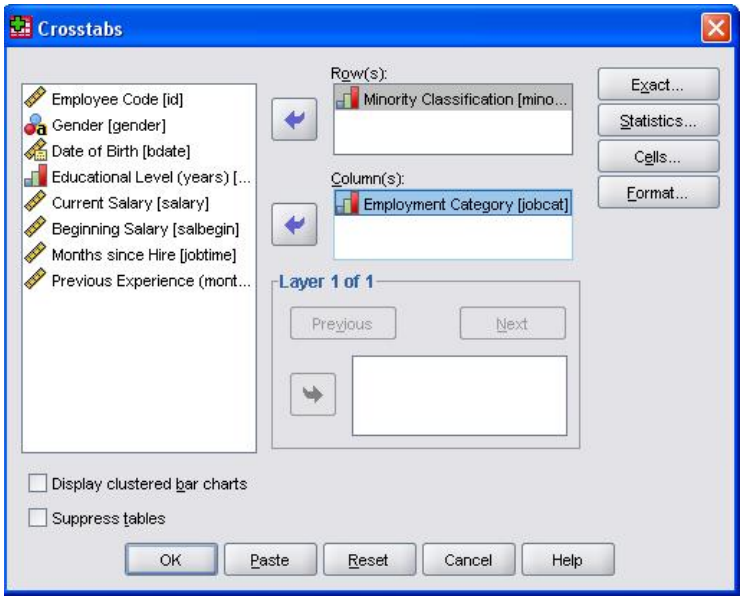
برای بدست آوردن رابطهٔ میان متغیرهای گسسته می‌توان در سطح مثالهای بیان شده در بالا بسنده نمود. در صورت علاقه بیشتر می‌توان تحلیل‌های زیر را هم همراه خروجی‌ها دریافت نمود.

یک مثال دیگر با استفاده از مثال employeedata که در خود نرم افزار spss موجود است می زنیم.

این فایل حاوی اطلاعات مربوط به تعدادی از پرسنل یک سازمان دولتی می باشد.

این فایل حاوی اطلاعات شماره پرسنلی(id)، جنسیت(gender)، تاریخ تولد به میلادی(bdate)، تحصیلات به سال(educ)، رده شغلی(jobcat)، درآمد جاری(salary)، درآمد در بدو شروع به کار(salbegin)، مدت زمان اشتغال به کار در سازمان(jobtime)، سابقه قبلی(prevexp)، اقلیت دینی (minority) می باشد. می خواهیم بدانیم آیا میان اقلیت که دارای دو سطح اقلیت(yes)، و غیر اقلیت(no)، و متغیر رده شغلی با سه رده مدیران(management)، کارمند میانی(custodial) و کارمند رده پائین(clerical) رابطه وجود داردیا خیر و یا به عبارت دیگر می خواهیم دریابیم آیا می توان ادعا کرد ، افرادی که در اقلیت مذهبی هستند کمتر از سایر افراد به رده های مدیریتی مشغول هستند و اقلیت بودن آنها یک عامل بازدارنده در رسیدن به جایگاه مدیریتی هست یا خیر.

فرض اولیه $H_0 : R = 0$ نشان‌دهنده استقلال دو متغیر و به نوعی عدم تاثیر اقلیت بودن بر انتخاب افراد به شغلهای مدیریتی و فرض جایگزین $H_1 : R \neq 0$ نشان‌دهنده وابستگی این دو متغیر و تاثیر اقلیت بودن بر رده شغلی افراد می باشد. برای این منظور مانند شکل ذیل با استفاده از جدول توافقی عمل می نمائیم.



از دکمهٔ ... Statistics گزینهٔ Chi-Square و Correlation را برای گرفتن جـداول Chi-Square Test و Symetric Measure در خروجـی انتخـاب کرده،دکمـه Countinue را فعـال و سـپس OK می‌نمائیم. خروجی ذیل ظاهر می شود.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Minority Classification * Employment Category	474	100.0%	0	.0%	474	100.0%

جدول Case Processing تعداد و درصد داده‌های معتبر و گم شده را نشان می‌دهد.

Minority Classification * Employment Category Crosstabulation

Count

		Employment Category			Total
		Clerical	Custodial	Manager	
Minority Classification	No	276	14	80	370
	Yes	87	13	4	104
Total		363	27	84	474

جدول Crosstabulation جدول توافقی میان وضعیت اقلیت * رده شغلی را نشان می‌دهد.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.172[a]	2	.000
Likelihood Ratio	29.436	2	.000
Linear-by-Linear Association	9.778	1	.002
N of Valid Cases	474		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.92.

جدول Chi-Square Test جدولی است که با اطلاعات موجود در مورد $H_0 : R = 0$ قضاوت می‌شود. با توجه به اینکه sig=0.014 برای Pearson Chi-Square است و با در نظر گرفتن $\alpha = 0/05$ می‌توان مشاهده کرد که $sig<\alpha$ پس‌فرض $H_0 : R = 0$ یعنی استقلال میان متغیرها رد می‌شود. و متغیرها به هم وابسته هستند.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Interval by Interval	Pearson's R	-.144	.033	-3.157	.002[c]
Ordinal by Ordinal	Spearman Correlation	-.113	.038	-2.466	.014[c]
N of Valid Cases		474			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

در جدول Symmetric Measure در صورتی که $H_0 : R = 0$ رد شده باشد می‌توان جهت R یعنی $R > 0$ یا $R < 0$ را مشخص کرد با استفاده از Pearson 'R مقدار R= -0/144 را ملاحظه می کنیم که عددی منفی است و به این معنی است که میان وضعیت اقلیت و رده شغلی رابطه معکوس وجود دارد و چون اقلیت (اقلیت=1 و اقلیت نبودن =0) دارای 2 کد است می‌توان نتیجه گرفت که افراد اقلیت کمتر به رده های بالای مدیریتی که کدهای بالا (3)را می گیرد می رسند.

6ـ3ـ تحلیل جدول توافقی با داده‌های اسمی:

بخش Nominal در پنجرهٔ Statistics... به اندازه‌گیری آماره‌های توافق، بر مبنای داده‌های اسمی می‌پردازد، این آماره‌ها عبارتند از:

ـ Contingency Coefficient (ضریب توافق).

ـ Phi and Cramers V (آماره فی و آماره کرامر).

ـ Lambda (با نماد λ).

گودمن و کردسکال این معیار را به منظور شناسایی تأثیر متغیرهای اسمی در پیش‌بینی مقادیر آنها ارائه نمودند.

ـ Uncertainty Coefficient (ضریب عدم اطمینان).

با انتخاب این گزینه‌ها جداول ذیل برای دو متغیر وضعیت تأهل و جنسیت در خروجی‌ها ارائه می‌شود:

Directional Measures

			Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.154	.074	1.902	.044
		marital Status Dependent	.077	.040	1.902	.044
		gender Dependent	.200	.082	1.902	.044
	Goodman and Kruskal tau	marital Status Dependent	.026	.014		.000
		gender Dependent	.041	.016		.000
	Uncertainty Coefficient	Symmetric	.020	.010	1.961	.047
		marital Status Dependent	.020	.010	1.961	.047
		gender Dependent	.020	.010	1.961	.047

Symmetric Measures

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Nominal by Nominal	Phi	.200			.044
	Cramer's V	.200			.044
	Contingency Coefficient	.196			.044
Interval by Interval	Pearson's R	.200	.073	1.960	.050
Ordinal by Ordinal	Spearman Correlation	.200	.073	1.960	.050
N of Valid Cases		80			

6ـ4ـ تحلیل جداول توافقی با داده‌های ترتیبی:

بخش Ordinal در پنجرهٔ ... Statistics به اندازه‌گیری آماره‌های مفید برای سنجش میان متغیرهای مستقل و وابسته از نوع ترتیبی می‌پردازد این آماره‌ها عبارتند از:

ـ Gamma (گامای کروسکال ـ والیس).

ـ Samers'd.

ـ Kendall's tau-b (ضریب تاوb کندال).

ـ Kendall's Tau-C (ضریب تاوC کندال).

ذکر این نکته ضروری است که این روشها در حقیقت روشهایی هستند که می‌توان در مواقعی کـه هـر دو متغیر گسسته از نوع ترتیبی باشند مورد استفاده قرارگیرند.

برای مثال اگر نوع هپاتیت 1، 2 و 3 را در متغیر هپاتیت ـ عددی به نوعی نشـان دهنـده یـک متغیـر ترتیبی در نظر بگیریم و رابطهٔ آن را با متغیر ترتیبی میزان رضایت از زندگی بدست آورید، این آماره‌ها به صورت ذیل در خروجی ظاهر می‌شوند:

Directional Measures

			Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	-.022	.[illegible]	-.[illegible]	.[illegible]
		[illegible] Dependent	.[illegible]	.[illegible]	.[illegible]	.[illegible]
		[illegible] Dependent	.[illegible]	.[illegible]	.[illegible]	.[illegible]

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	-.044	.[illegible]88	-.[illegible]77	.[illegible]5%
	Kendall's tau-c	-.047	.[illegible]60	-.[illegible]77	.[illegible]5%
	Gamma	-.[illegible]61	.[illegible]50	-.[illegible]77	.[illegible]5%
	Spearman Correlation	-.[illegible]55	.[illegible]87	-.[illegible]81	.[illegible]78[illegible]
Interval by Interval	Pearson's R	.[illegible]52	.[illegible]07	.[illegible]341	.[illegible]7[illegible]
N of Valid Cases		60			

همانطور که ملاحظه می‌شود هر کدام از آماره‌های بالا، یک مقدار جایگزین برای بدست آوردن ضریب همبستگی ارائه می‌کنند و در داده‌های با حجم بالا مقادیر آنها به یکدیگر بسیار نزدیک خواهد بود.

بخش Nominal by Interral به تحلیل داده‌هایی که در جدول توافقی با متغیرهای فاصله‌ای در نظر گرفته می‌شود اختصاص دارد در این حالت باید متغیر وابسته بر حسـب مقیـاس فاصـله‌ای (ماننـد فشـار خون) و متغیر مستقل بر حسب داده‌های اسمی (مانند جنسیت) باشد.

با انتخاب گزینه‌های Kappa، Risk یا Mc Nemar با روشهای متفاوت میزان توافق و قدرت بـین دو رتبه دهنده نشان داده می‌شود.

بـا انتخـاب Cochran's and Mantel-Heanszel Statistics و انتخـاب عـدد 1 در قسـمت Test Common Odds Ratio Equal در حقیقت معیاری برای بررسی استقلال بین یـک متغیـر مسـتقل دو دویی(مانند جنسیت) و یک متغیر وابسته دو دویی (مانند تاهل)ارائه می‌گردد که تابیر sig در هر یک از آنها مانند حالات قبل است و در انتخاب عدد 1 در حقیقت آزمون فرض برابری نسبت‌ها آزمون می‌گردد که در نتیجه یک فاصله اطمینان 95% برای این نسبت‌ها با فرض توزیع نرمال برای متغیرها توسط آمارهٔ t به صورت زیر ارائه می‌گردد:

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	1.222	1	.269
Mantel-Haenszel	.525	1	.469

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			.438
ln(Estimate)			-.827
Std. Error of ln(Estimate)			.754
Asymp. Sig. (2-sided)			.273
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.100
		Upper Bound	1.916
	ln(Common Odds Ratio)	Lower Bound	-2.304
		Upper Bound	.650

6ـ5ـ نمایش مقادیر خانه‌های جدول توافقی

برای مشاهده مقادیر خانه ها در هر جداول توافقی از کلید Cells استفاده می شود. با انتخـاب کلیـد Cells ... جعبهٔ گفتگوی زیر باز می‌شود:

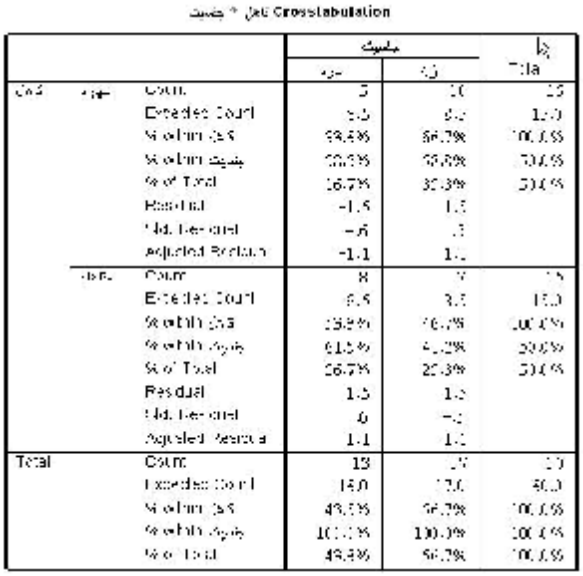
بخش Counts نوع مقادیری را که در هر خانه جدول مشاهده می‌شود نشان می‌دهد. گزینهٔ Observed تعداد مشاهدات در هر خانه و گزینه Expected تعداد مورد انتظار در هر خانه با وجود شرط استقلال میان متغیرها را نشان می‌دهد.

بخش Percentages نوع درصدهای جدول را نشان می‌دهد، درصدها می‌توانند از تقسیم تعداد مشاهدات هر خانه بر تعداد مشاهدات هر سطر (گزینه Row) یا هر ستون (گزینه Column) و یا محل جدول (گزینه Total) به دست آیند.

بخش Residuals نوع مانده‌ها را محاسبه می‌نماید. گزینه Unstandardaized مقدار باقی‌مانده‌های R=Observed – Expected را در هر خانه محاسبه می‌نماید.

گزینه Standardized مقدار $\frac{\text{Observed} - \text{Expected}}{\sqrt{\text{Expected}}}$ را برای هر خانه محاسبه می‌نماید. و مقدار Adj.Standardized تعداد خطای استاندارد مانده‌ها (R) برای هر خانه را محاسبه می‌نماید.

برای متغیر مستقل جنسیت و متغیر وابسته وضعیت تأهل با انتخاب این گزینه‌ها جدول توافقی به صورت زیر تغییر می‌کند و هر یک از این مقادیر را برای هر خانه از جدول محاسبه می‌نماید.



با انتخاب کلید ... Format، قالب کلی نمایش جداول تعیین می‌شود.

بخش Row Order ترتیب قرار گرفتن سطرها را نشان می‌دهد.

گزینهٔ Ascending مقادیر متغیرها را به صورت افزایش و گزینهٔ Descending آنها را به ترتیب کاهش در جدول نمایش می‌دهند.

تمرین:

1-رابطه میان متغیرهای گسسته فایل employe data را با متغیر گسسته رده شغلی بدست آورید و هر یک را شرح دهید.

2-رابطه میان متغیرهای گسسته فایل cars با متغیر منطقه تولید را بدست آورید و هر یک را شرح دهید.

3- رابطه میان متغیرهای گسسته فایل مبتلایان به هپاتیت را با متغیر جنسیت بدست آورید.

4- رابطه میان متغیر گسسته جنسیت را در فایل employ data با متغیر گسسته رده شغلی بدست آورید مشروط به آنکه افراد با رده شغلی مدیر()manager در بررسی حضور نداشته باشند.

5- متغیر سن افراد مورد بررسی را به فواصل مساوی 10 ساله در فایل بیماران مبتلا به هپاتیت (پیوست 1) تقسیم نموده و سپس رابطه متغیر سن از نوع گسسته را با متغیر نوع ابتلا از راه ارثی بدست آورید.

6- رابطه میان متغیر جنسیت و ارثی بودن و نبودن بیماری هپاتیت را بدست اورید.

7- رابطه میان شهری و روستایی بودن و نوع ابتلا به هپاتیت را بدست اورید.

8- تاثیر ارثی بودن و نبودن بیماری هپاتیت را بر روی کلیه متغیرهای گسسته داده های بیماران هپاتیتی بدست آورید.

فصل هفتم

رسم نمودارها

اهداف آموزشی فصل هفتم

فراگیر پس از مطالعه این فصل باید بتواند:

1-چگونگی رسم نمودار در نرم افزار spss.16 را بیان کند.

2- مراحل رسم نمودارها ی میله ایی در نرم افزار Spss.16 را با ذکر مثال توضیح دهد.

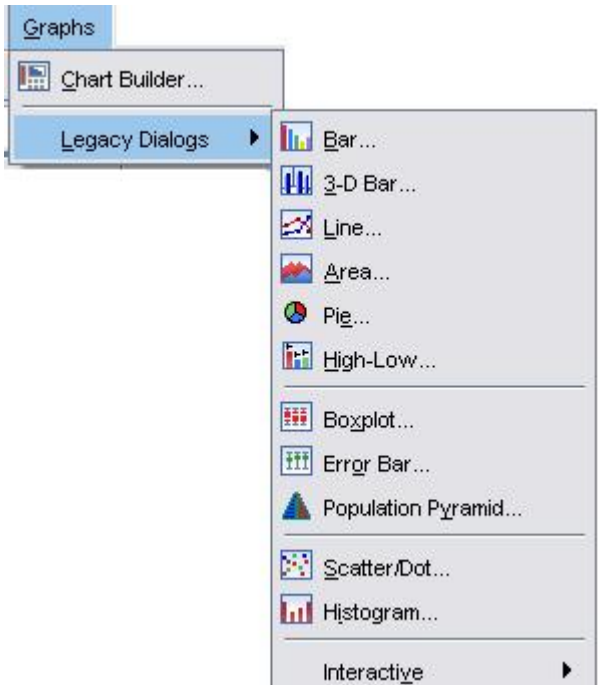
3- مراحل رسم نمودارهای خطی در نرم افزار Spss.16 را با ذکر مثال توضیح دهد.

4- مراحل رسم نمودارهای سطحی در نرم افزار Spss.16 را با ذکر مثال توضیح دهد.

5-مراحل رسم نمودارهای دایره ایی در نرم افزار Spss.16 را با ذکر مثال توضیح دهد.

6-چگونگی ویرایش نمودارها در نرم افزار Spss.16 را با ذکر مثال شرح دهد.

7ـ1ـ مقدمه



نرم‌افزار SPSS در دو حالت نمودارهای آماری را توسط دستورات زیر رسم می‌نماید.

الف ـ مجموعه‌ای از گزینه‌ها در زیر منوهای Analyze که نمودار مربوط به همان فرمان را

ترسیم می‌کنند، مانند گزینهٔ Charts در فرمان Frequency که با انتخاب آن می‌توان نمودارهای جانبی فرمان فوق را اجرا نمود.

ب ـ زیر منوهای منو Graph که می‌توان مانند شکل زیر یکی از انواع نمودارهای آماری قابل ترسیم را انتخاب و بعد از طی مراحل مورد نظر در پنجرهٔ خروجی نمودارها را ملاحظه کرد.

انواع نمودارهایی که در این منو قابل ترسیم است عبارتند از:

ـ Bar Charts (نمودار میله‌ای، خوشه‌ای یا انباشته).

ـ Line Charts (نمودار خطی ساده یا چند گانه).

ـ Area Charts (نمودار سطحی ساده یا پشته‌ای).

ـ Pie Charts (نمودار ـ دایره‌ای ساده یا مرکب).

ـ High-Low Charts (نمودار کراندار زوج مقادیر یا مقادیر سه گانه).

ـ Control Charts (نمودارهای کنترل کیفیت آماری).

ـ Pareto Charts (نمودار پارتو).

ـ Box Plot (نمودار جعبه‌ای).

ـ Error Bar (نمودار خطا).

ـ Scatter Plot (نمودار پراکنش داده‌ها).

ـ Histogram (نمودار هسیتوگرام).

ـ Normal P-P Plots (نمودار خط نرمال با استفاده از نسبت‌های تجمعی) توزیع داده‌ها در برابر نسبت‌های تجمعی توزیع نرمال).

ـ Normal Q-Q Plots (نمودار خط نرمال با استفاده از چندک‌های توزیع داده‌ها در برابر چندکهای توزیع نرمال).

ـ Sequance Charts (نمودار رشته‌ای بر حسب زمان تولید محصولات).

ـ Time Series (نمودارهای سریهای زمانی).

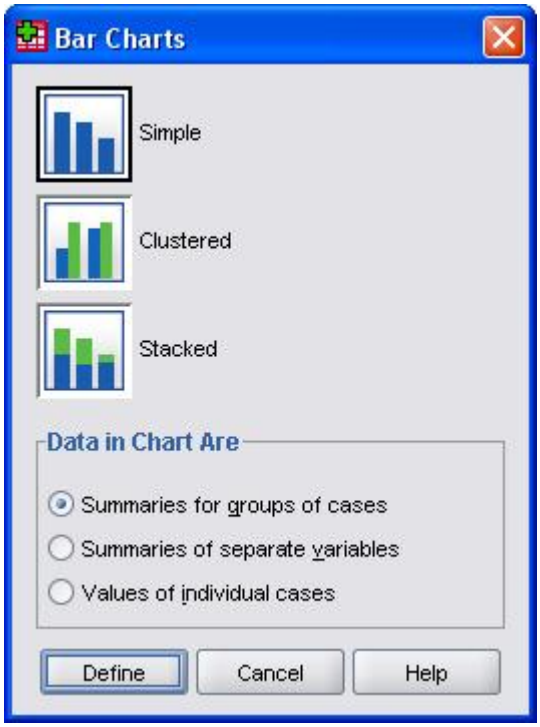
در این فصل سعی می‌شود تعدادی از شکل‌ها که کاربرد وسیع‌تری دارند توضیح داده شوند.

رسم نمودارها در نرم افـزار توسـط دو روش صـورت مـی گیـرد . در روش اول کـه بـا گزینـه Chart Builder که این امکان را به کاربر می دهد که با توجه به جاگذاری متغیرهـا در محـل هـای مـورد نظـر بتواند شکل را تولید کند و در حقیقت یک روش خودکار برای کشیدن شکلها در نرم افزار بـه شـمار مـی رود و روش دیگر که استفاده از گزینه Legacy Dialogs این امکان فراهم می شود که کاربر بـه صـورت دستی شکلها را تولید کند . برای روشن شدن مسئله ابتدا به روش legacy dialog که یک روش دستی برای تولید شکلها می باشد اشاره می شود و در ادامه روش Chart Builder نیز مطـرح خواهـد شـد. در قسمت نمودارها به روش دستی هر یک از نمودارها را جدا بررسی می نمائیم و در ادامه بـه روش Chart Builder که روش خودکار برای ترسیم نمودارها می باشد خواهیم پرداخت.

7ـ2ـ رسم نمودار میله‌ای

برای ترسیم نمودارهای میله‌ای زیر منو Bar ... از زیر منو Legacy Dialogs را انتخاب می کنیم

و Bar را انتخاب می‌نمائیم تا جعبهٔ زیر ظاهر شود:



در قسمت بالا سه نوع نمودار میله‌ای ساده (Simple)، خوشه‌ای (Clusterd) و انباشته (Stacked) را می‌توان انتخاب نمود و از قسمت Data in Charts Are نحوه تقسیم‌بندی محور X را بر روی یک نمودار میله‌ای انتخاب می‌نمائیم.

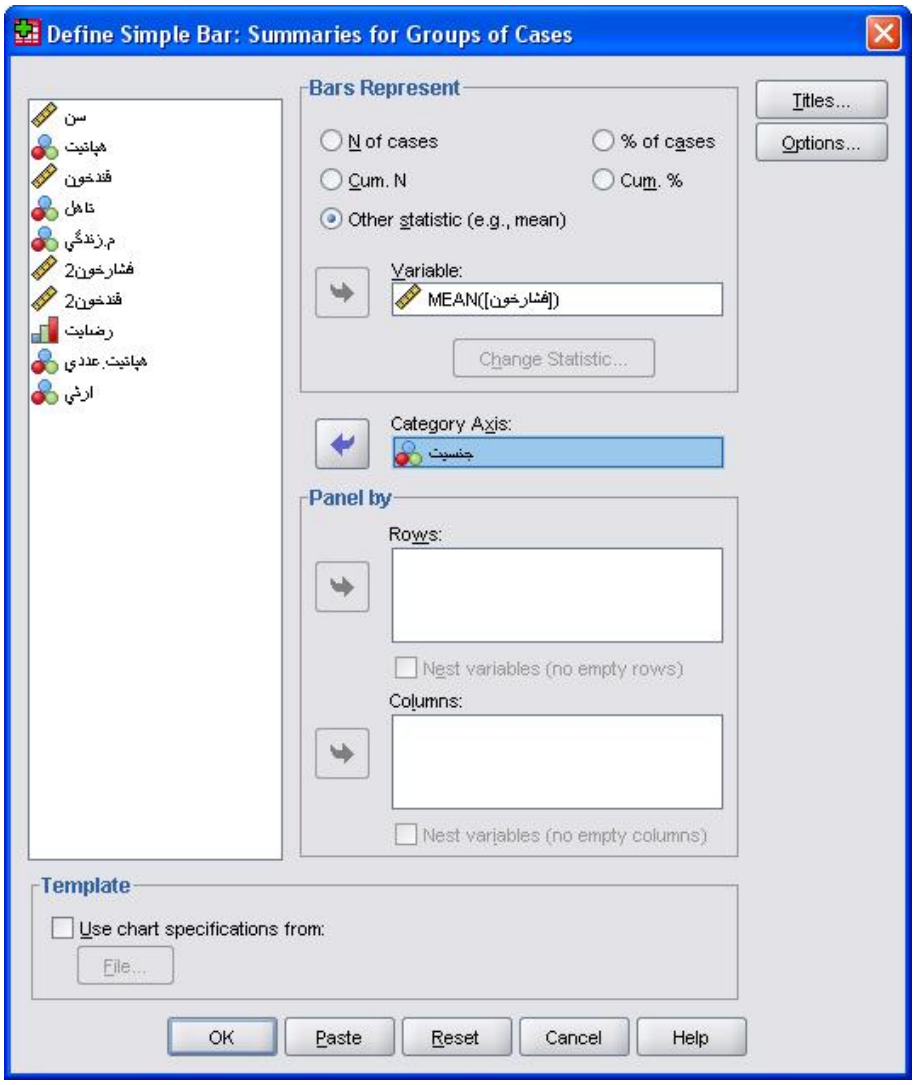
ـ گزینه Summaries for groups of cases مقادیر سطوح مختلف یک متغیر گسسته را بر روی محور X نشان می‌دهد. مثلاً سطوح مختلف جنسیت (زن و مرد) وامکان مقایسه یک معیار را که می‌تواند، تعداد داده‌ها، میانگین یک خصوصیت، درصد و ... باشد برای این دو سطح فراهم می‌آورد.

ـ گزینهٔ Summaries of separate Variables امکان قرار گرفتن در متغیر را بر روی محور X بیان می‌کند. مثلاً اگر بخواهید میانگین متغیرهای فشار خون قبل و بعد را با هم مقایسه کنید در محور Xها باید دو متغیر فشار خون قبل و بعد قرار داشته باشند و معیار میانگین نیز باید معیار محور Y باشد.

ـ گزینهٔ Values of Individual Cases که بر روی محور Xها شماره سطرهای یک متغیر را نشان می‌دهدو مقدار هر سطر آن را بر روی محور Yها نمایش می‌دهد.

7ـ2ـ1ـ رسم نمودار میله‌ای ساده

برای انتخاب نمودار میله‌ای ساده از پنجرهٔ Bar Charts حالت Simple را انتخاب و از قسمت Data in Chart Are حالت اول یعنی Summaries For Group of Cases را انتخاب می‌نمائیم و بعد از انتخاب دکمهٔ Define پنجرهٔ زیر باز می‌شود.



در قسمت سمت چپ لیستی از تمامی متغیرها نشان داده شده است.

قسمت Bar Represent معیار تقسیم‌بندی محور y را نشان می‌دهد که شامل حالات زیر است:

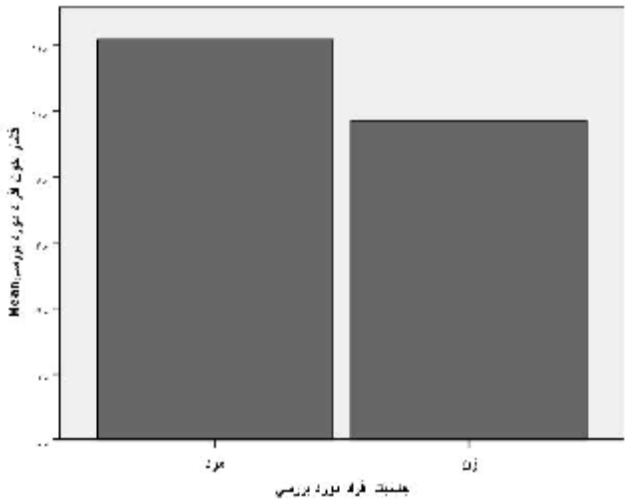
ـ Nof Cases (تعداد مشاهدات هر سطح).

ـ % Of Cases (درصد مشاهدات هر سطح).

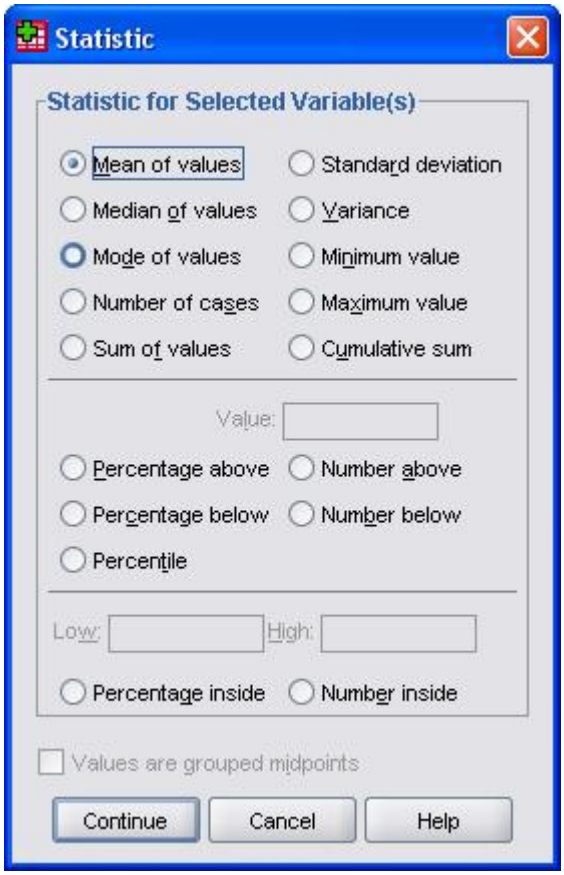
ـ Cum.N (تعداد تجمعی مشاهدات هر سطح).

ـ Cum.% (درصد تجمعی هر سطح).

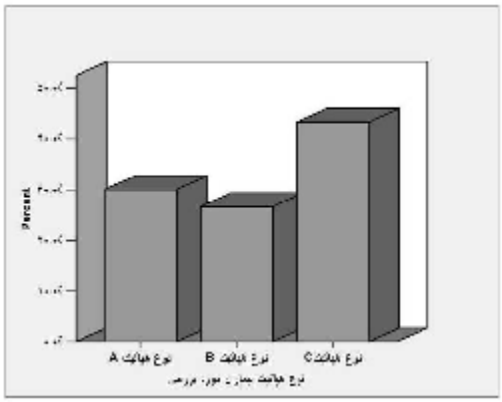
ـ Other Statistic با انتخاب این گزینه جعبهٔ Variable باز می‌شود و می‌توان محور y را بر اساس یک پارامتر از یک متغیر دیگر محاسبه نمود. برای مثال اگر بخواهید برای سطوح مختلف جنسیت میانگین فشار خون زنها و مردها را با هم مقایسه نمائید می‌توانید با انتخاب این گزینه و قرار دادن فشار خون در قسمت Variable و انتخاب گزینه Mean (میانگین) از گزینهٔ ... Change Statistics و قرار دادن متغیر جنسیت در داخل پنجرهٔ Category Axis و تأئید از طریق OK شکل زیر در خروجی ظاهر می‌شود:



همانطور که در شکل بالا ملاحظه می‌شود میانگین فشار خون در سطوح مختلف متغیر جنسیت محاسبه شده است. کاربر می‌تواند با انتخاب ... Change Statistic پنجرهٔ ذیل را باز نماید و هر یک از پارامترهای پائین را معیار تقسیم‌بندی محور yها قرار دهد.



اگر کاربر از قسمت Bar Represent گزینهٔ % Of Cases را انتخاب نماید و در قسمت Category Axis متغیر نوع هپاتیت را وارد نماید آنگاه شکل زیر در خروجی ظاهر می‌شود:



در این شکل درصد هر یک از مبتلایان به انواع هپاتیت بیان شده است. اگر کاربر در قسمت Panel

By در پنجره‌های Rows یا Columns یک متغیر گسسته را وارد نماید. شکل‌های نمودار میله‌ای برای هر یک از سطوح این متغیر در خروجی ظاهر می‌شود.

دکمهٔ Tiles پنجره‌ای مانند زیر را باز می‌نماید:

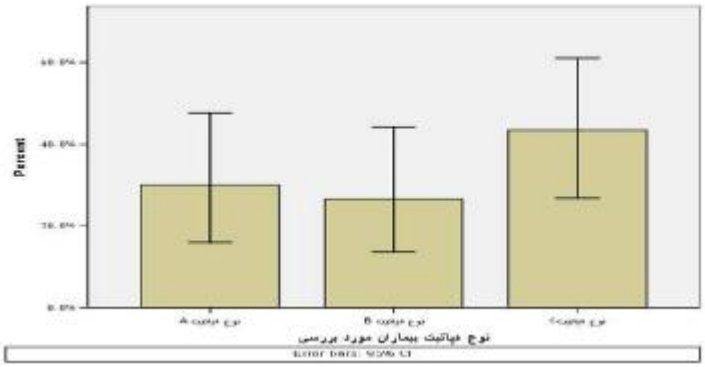
این پنجره این امکان را به کاربر می‌دهد که تا دو خط را به عنوان Title، یک خط را به عنوان زیر عنوان Subtitle و دو خط را در پائین شکل در قسمتFoot Note بنویسید.

با فعال نمودن دکمهٔ ... Options پنجره‌ای باز می شود که امکان گرفتن شکلها با خصوصیات خاص را در خروجی فراهم می آورد.

با انتخاب گزینهٔ Display Group Defined by Missing values مقادیر گم شده هم به عنون یک سطح در نظر گرفته شده و شکل آنها رسم می‌شود.

با انتخاب گزینهٔ Display Error Bars امکان انتخاب یک فاصله اطمینان با انتخاب آن در Level در

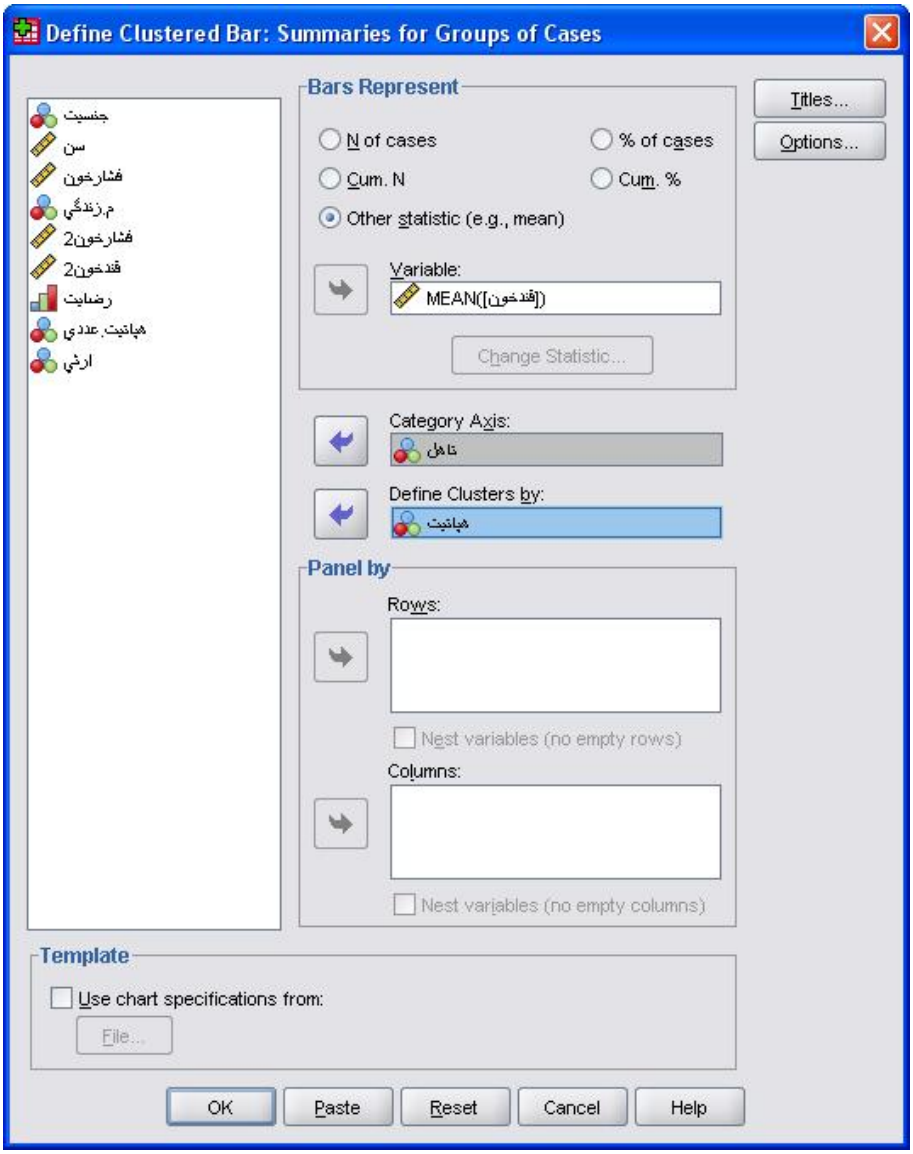
شکل‌ها مانند زیر فراهم می‌شود:



در پنجرهٔ اصلی با انتخاب گزینهٔ Use Chart Specification From و انتخاب ... File این امکان برای کاربر فراهم می‌شود که بتواند مسیر یک شکل را که بر روی خصوصیات آن ویرایش‌های لازم را انجام داده است به نرم‌افزار بدهد تا تمامی شکل‌هایی که بعد از آن کشیده می‌شوند با خصوصیات آن شکل در خروجی‌ها ترسیم شوند. از جمله این ویژگیها می‌تواند، رنگ، سایز، فونت و ... باشد.

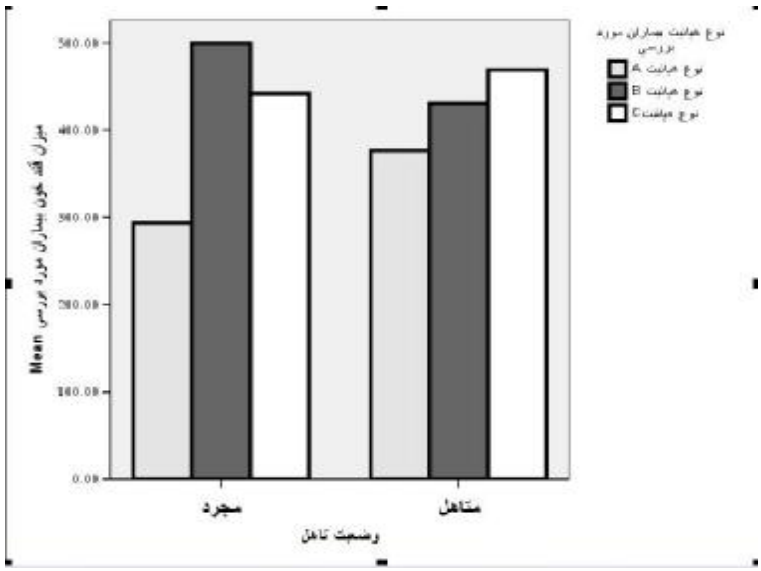
7ـ2ـ2ـ نمودار میله‌ای خوشه‌بندی شده

برای ترسیم یک نمودار میله‌ای خوشه‌بندی از پنجره Bar Charts نمودار Cluster و از قسمت Data In Chart Are حالت اول یعنی Summarise For Groups of Cases را انتخاب و Define را می‌زنیم. پنجره‌ای به صورت زیر باز می‌شود:



همانند قبل محور طبقه‌بندی محور y را از قسمت Bar Represent انتخاب می‌نمائیم.

در قسمت Category Axis متغیر گسسته‌ای را که سطوح آن در محور X قرار می‌گیرند و در قسمت Define Cluster By متغیر گسسته‌ای را که سطوح آن مبنای خوشه‌بندی هستند وارد می‌نمائیم. در اینجا مثالی از نمودار میله‌ای سطوح مختلف کامل که میانگین قند خون آنها در خوشه‌های مختلف نوع هپاتیت با هم مقایسه شده است (نحوه ورود متغیرها در شکل بالا موجود است). بعد از تأئید فرمان، نموداری مانند زیر در خروجی ظاهر می‌شود:

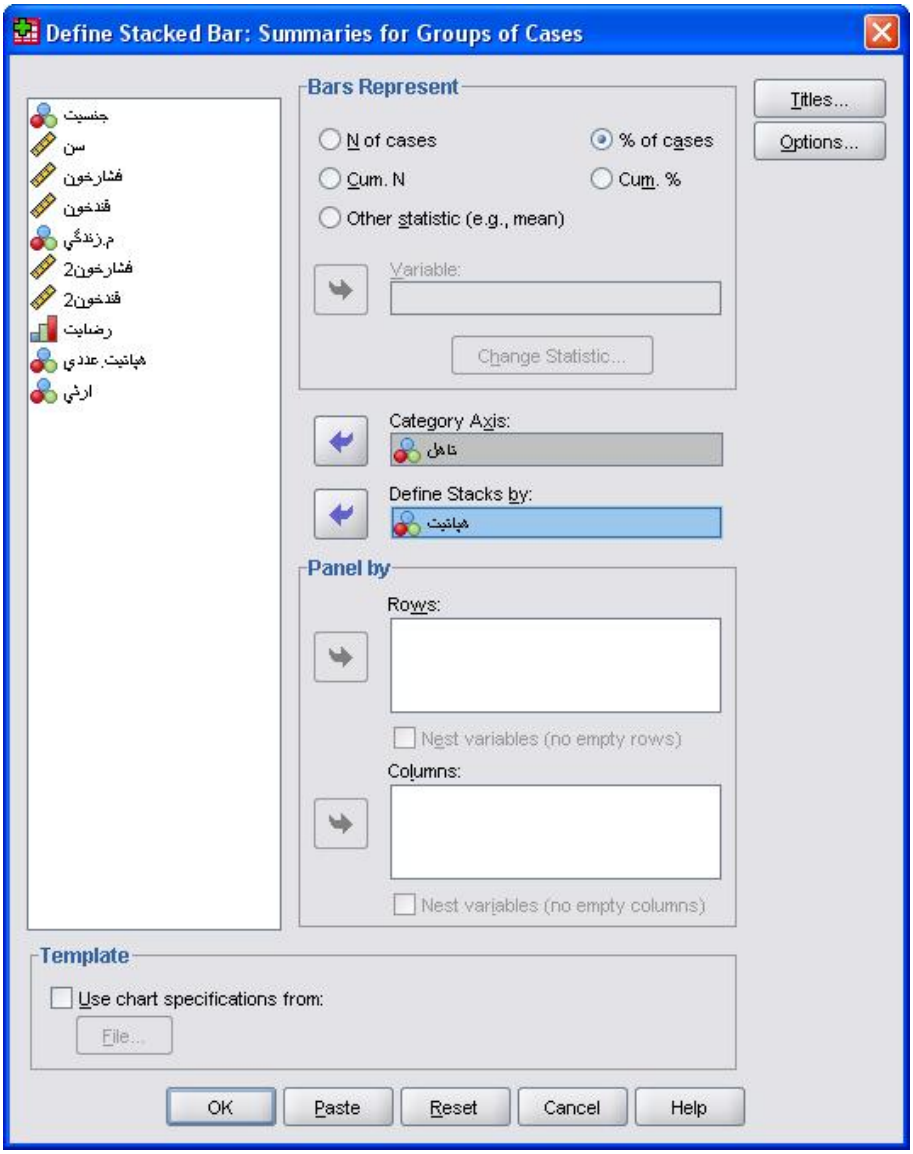


شکل بالا امکان مقایسه تطبیقی میان میانگین قند خون در هر یک از انواع هپاتیت میان افراد مجرد و متأهل را فراهم می‌آورد.

دیگر گزینه‌های موجود در صفحه مانند گزینه‌های نمودار میله‌ای ساده عمل می‌کند.

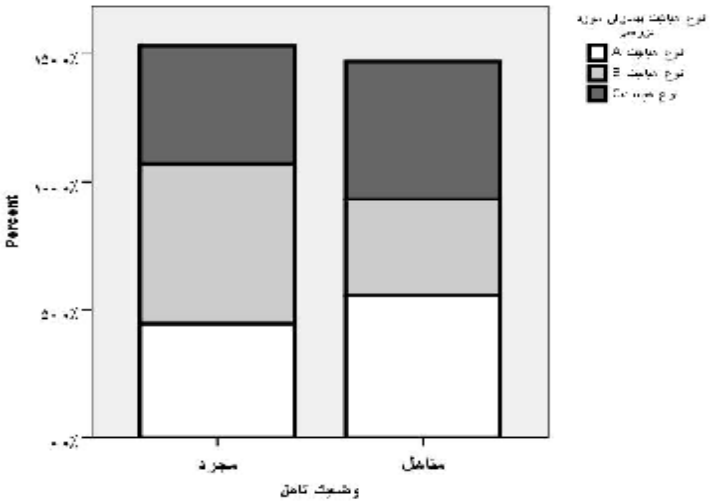
7ـ2ـ3ـ نمودار میله‌ای انباشته شده

برای ترسیم یک نمودار میله‌ای انباشته شده از قسمت پنجرهٔ Bar Charts نمودار Stacked و از قسمت Data in Chart Are حالت اول Summaries For Groups of Cases را انتخاب و Define را می‌زنیم. پنجره‌ای به صورت زیر باز می‌شود:



همانند شکل‌های قبل معیار طبقه‌بندی محور y را از قسمت Bar Represent انتخاب می‌نمائیم.

در قسمت Category Axis متغیر گسسته محور xها را وارد می‌نمائیم. در قسمت Define Stacks By متغیر گسسته‌ای را که سطوح آن مبنای انباشت هستند وارد می‌نمائیم. در اینجا مثالی از نمودار میله‌ای انباشته که درصد نوع هپاتیت را به عنوان متغیر انباشته در سطوح مختلف متغیر گسسته وضعیت تأهل به صورت شکل بالا وارد نموده‌ایم ارائه شده است. نتیجه اجرای فرمان از طریق OK به صورت زیر است:

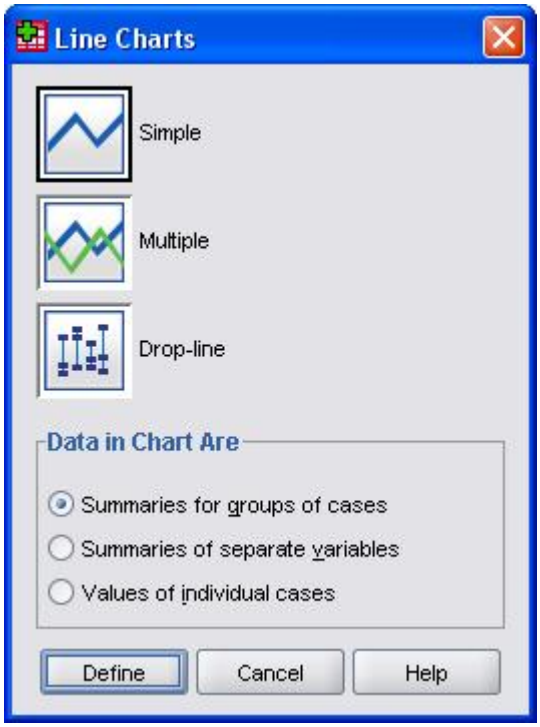


این شکل امکان مقایسه تطبیقی درصد ابتلا به هر یک از انواع هپاتیت را در گروه افراد مجرد و متأهل فراهم می‌آورد.

دیگر گزینه‌ها نظیر ... Titles، ... Options و Template مانند توضیحات نمودار میله‌ای ساده عمل می‌نمایند.

7ـ3ـ رسم نمودار خطی

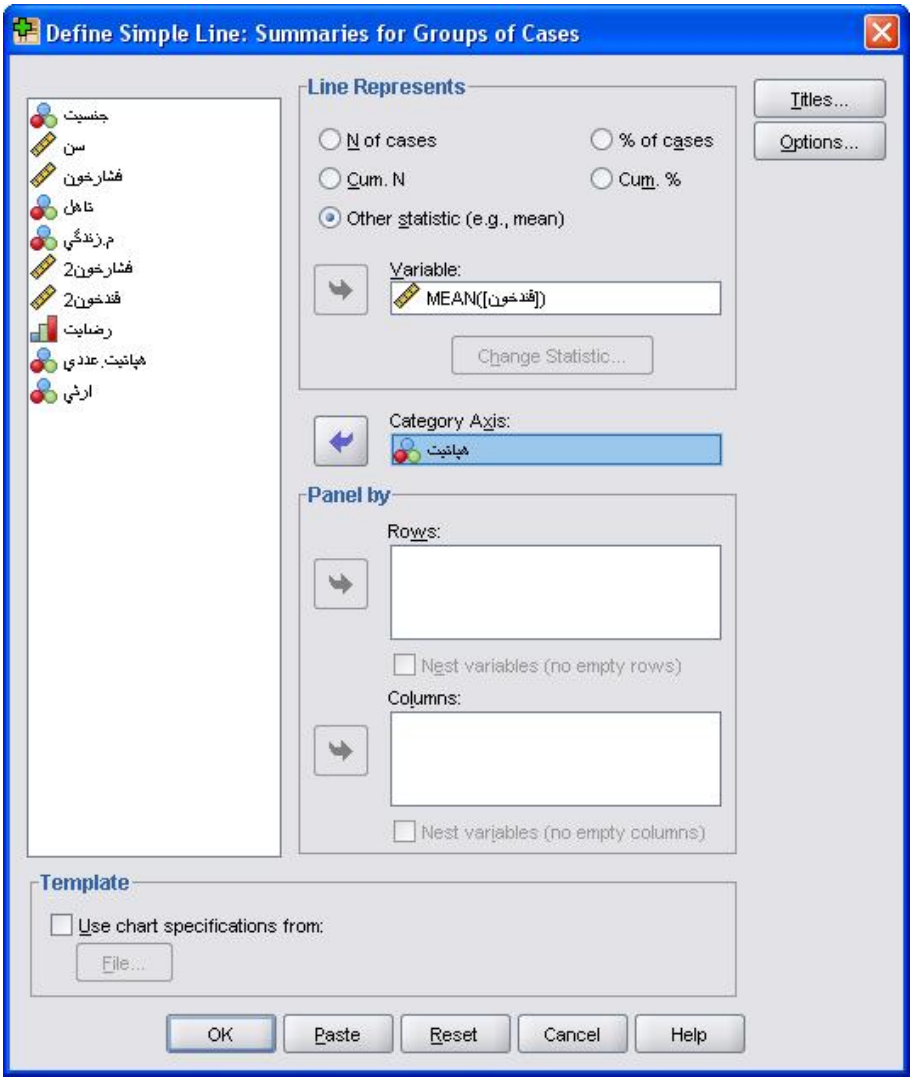
برای ترسیم نمودارهای خطی زیر منو ... Line از منو Graphs را انتخاب می‌نمائیم جعبه‌ای به صورت زیر ظاهر می‌شود:



در قسمت بالای پنجرهٔ Line Charts سه نوع نمودار خطی ساده (Simple)، چندبعدی (Multiple) و خط و نقطه (Drop-Line) را می‌توان انتخاب نمود و از قسمت Data In Charts Are نحوه تقسیم‌بندی محور X را بر روی نمودار خطی انتخاب می‌نمائیم. هر یک از این گزینه‌ها و معیارهای تقسیم‌بندی محور Xها در قسمت ترسیم نمودار میله‌ای بیان شده است.

7ـ3ـ1ـ رسم نمودار خطی ساده:

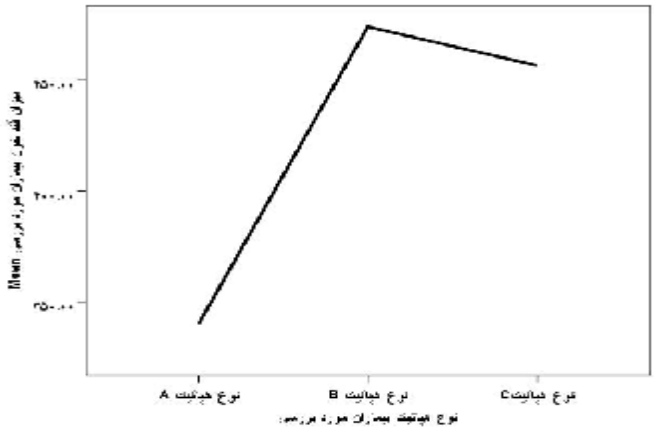
برای انتخاب نمودار خطی ساده از پنجرهٔ Line Charts حالت Simple و از قسمت Data in Chart Are گزینهٔ اول یعنی Summaries For Groups For Cases را انتخاب کرده و بعد از انتخاب دکمهٔ Define پنجرهٔ زیر باز می‌شود:



در قسمت سمت چپ لیستی از تمامی متغیرهای موجود فایل نشان داده شده است. قسمت Line Represents معیار تقسیم‌بندی محور y را نشان می‌دهد که شامل حالات توضیح داده شده مانند قسمت ترسیم نمودار میله‌ای است.

در قسمت Category Axis متغیر تقسیم‌بندی محور xها را واردمی‌نمائیم. برای مثال اگر قصد ترسیم نمودار خطی میانگین قند خون افراد مبتلا به انواع هپاتیت را داشته باشیم مانند بالا متغیر نوع هپاتیت را وارد Category Axis کرده و گزینهٔ Other Statistic را از قسمت Line Represent انتخاب نموده و متغیر قند خون را وارد قسمت Variable نموده و از پنجرهٔ Change

... Statistic گزینهٔ Mean of Values را انتخـاب نمـوده یادکمـه Continue را فعـال وسـپس OK می‌نمائیم.شکل نمودار خطی زیر در خروجی ظاهر می‌شود:

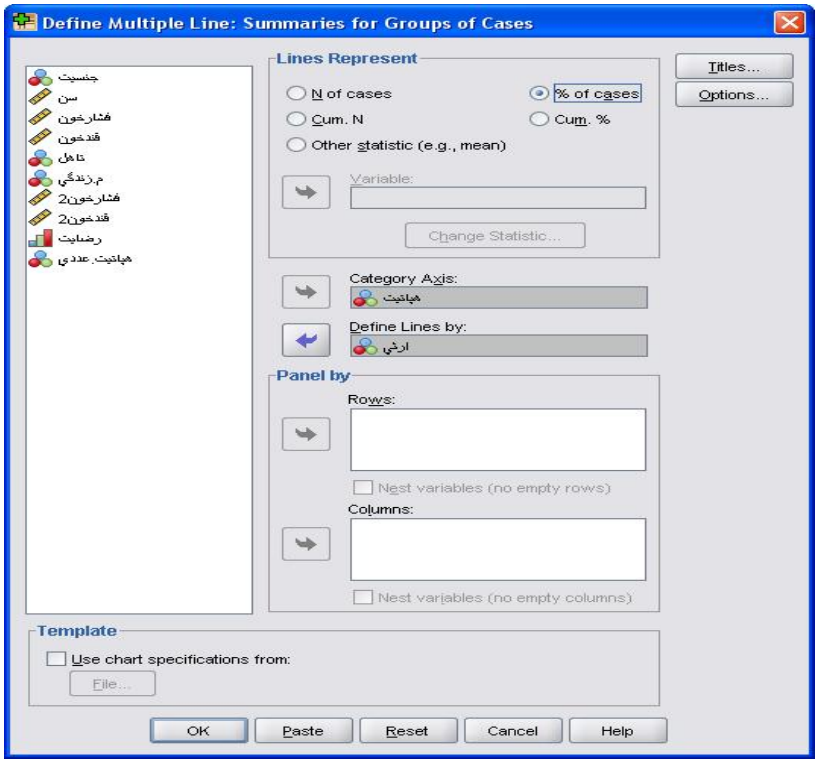


گزینه‌های ...Titles ، Options و Template مانند انواع شکل‌های توضیح داده شده در قبـل عمـل می‌نمایند.

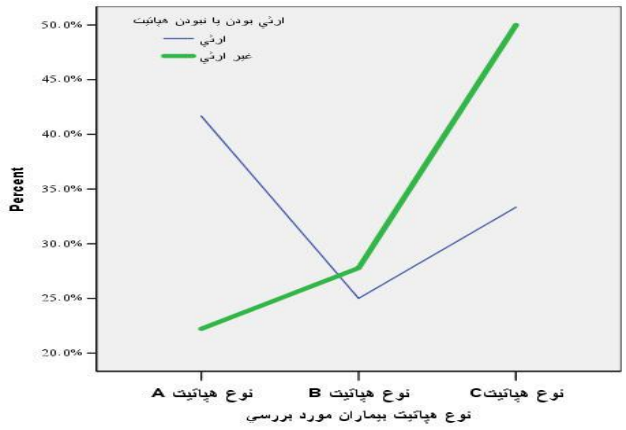
7ـ3ـ2ـ رسم نمودار خطی چندگانه

توسط این نمودار چند خط به طور همزمان، تغییرات متغیرها را نشان می‌دهنـد، لـذا مـی‌تـوان عمـل مقایسه بین متغیرها را انجام داد.

برای انتخاب نمودار خطی چندگانه از قسمت Line Charts حالـت Multiple را انتخـاب نمـوده و از قسمت Data in Chart Are حالت اول را انتخاب نموده و Define را می‌زنیم، پنجره‌ای به صـورت زیـر باز می‌شود:



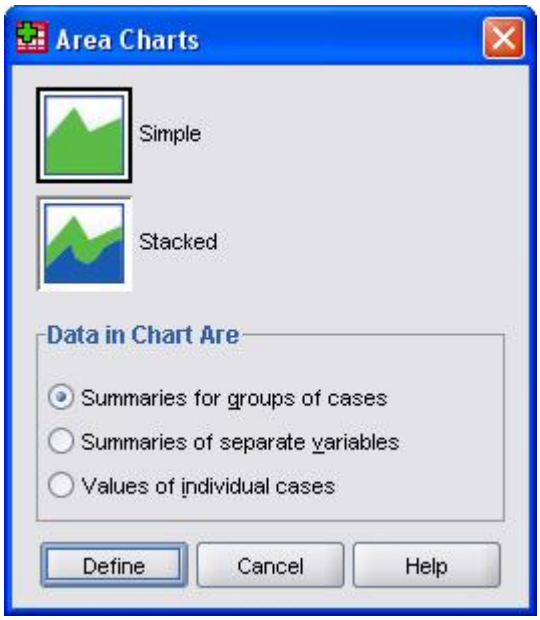
از قسمت Line Reprecent معیار اندازه‌گیری محور y را مشخص می‌نمائیم. در قسمت Category Axis متغیری را که سطوح آن مبنی تقسیم‌بندی محور xها می‌شوند را وارد می‌نمائیم و در قسمت Define Lines by متغیری را که سطوح آن مبنی ایجاد خطوط چندگانه است وارد نموده وسپس OK را انتخاب می کنیم. برای مثال نمودار خطی چندگانه برای درصد ارثی بودن افراد مبتلا به انواع هپاتیت به صورت زیر ارائه می شود.



همانطور که ملاحظه می‌شود 50 درصد افرادی که از راه ارثی مبتلا به هپاتیت می‌شوند دچار هپاتیت نوع C می‌شوند و حدود 42 درصد افرادی که از راههای غیرارثی مبتلا به هپاتیت می‌شوند به هپاتیت نوع A مبتلا می‌شوند.

7ـ4ـ رسم نمودار سطحی

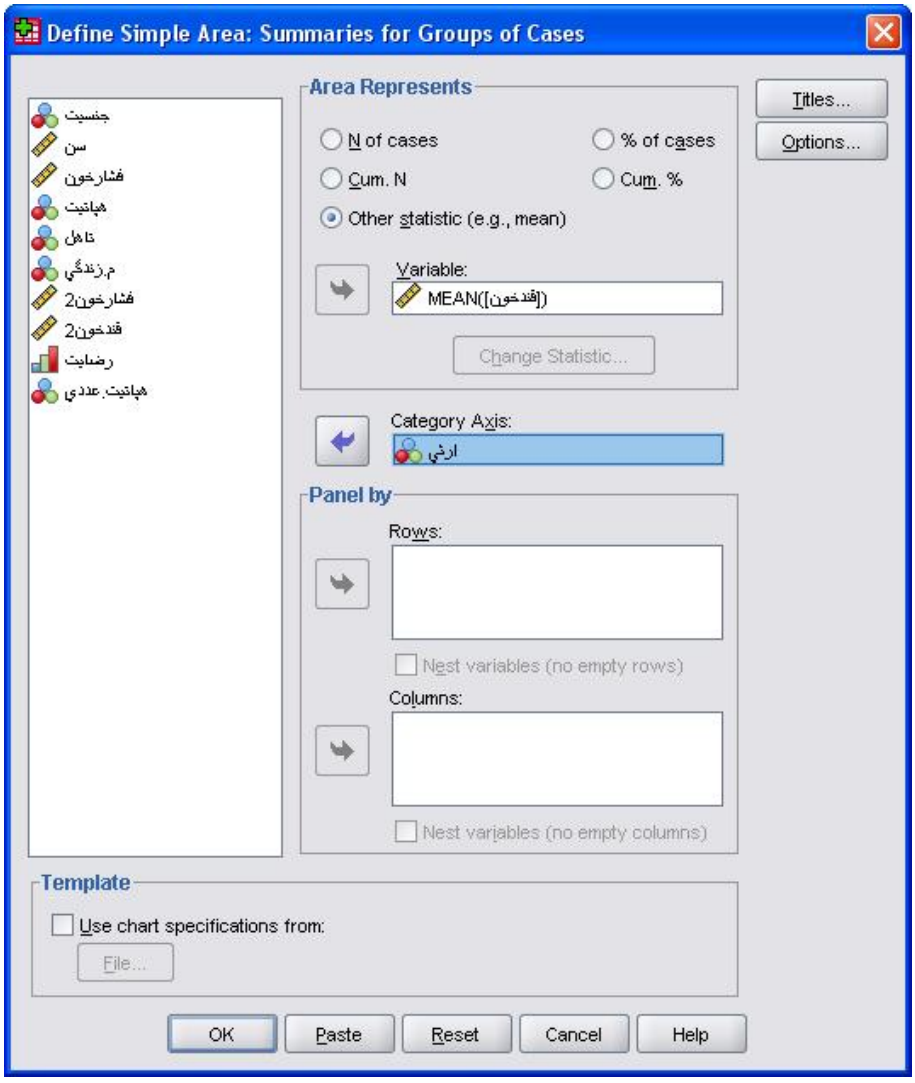
برای ترسیم نمودارهای سطحی زیر منو ... Area از منو Graphs را انتخاب می‌نمائیم تا پنجرهٔ زیر باز شود:



همانطور که ملاحظه می‌شود نمودارهای سطحی دارای دو نوع ساده (Simple) و انباشته (Stacked) هستند و داده‌های مورد نیاز این نمودارها بر روی محور Xها نیز به صورت همان سه صورت توضیح داده شده در قبل می‌باشد.

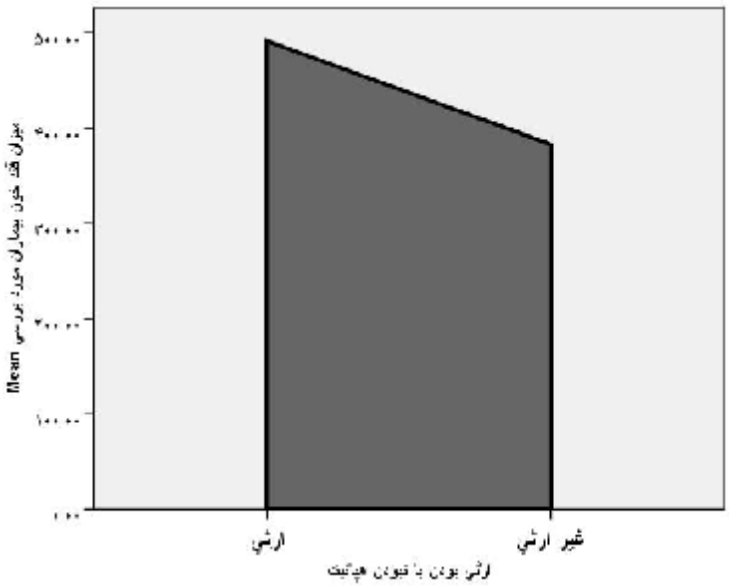
7ـ4ـ1ـ نمودار سطحی ساده

برای ترسیم نمودار سطحی ساده در پنجرهٔ Area Charts گزینهٔ Simple را انتخاب نموده و از قسمت Data in Chart Are حالت Summaries For Groups Of Cases را انتخاب نموده و Define را می‌زنیم تا پنجره‌ای به صورت زیر باز شود:



تمامی مراحل کشیدن نمودار Area Simple مانند Line Simple است تنها در نتیجه زیر خطوط Line Chart به عنوان سطح زیرین نمایش داده می‌شود. برای نمونه میانگین قند خون را به عنوان معیار Area Represent وارد نموده و متغیر ارثی بودن را به عنوان معیار Category Axis نموده وسپس

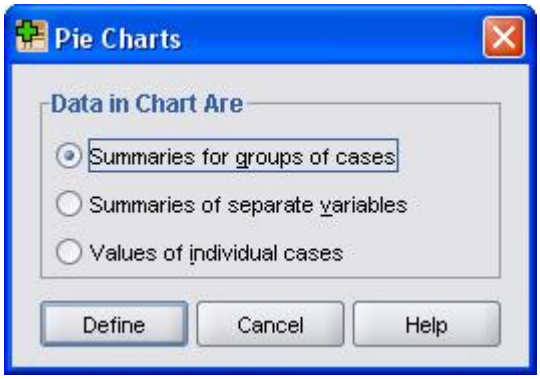
OK را انتخاب می‌نمائیم. خروجی یک نمودار سطحی ساده به صورت زیر است:



همانطور که ملاحظه می‌شود کسانی که از راه ارثی مبتلا به هپاتیت می‌شوند به طور متوسط قند خون بالاتری دارند.

7ـ5ـ ترسیم نمودار دایره‌ای

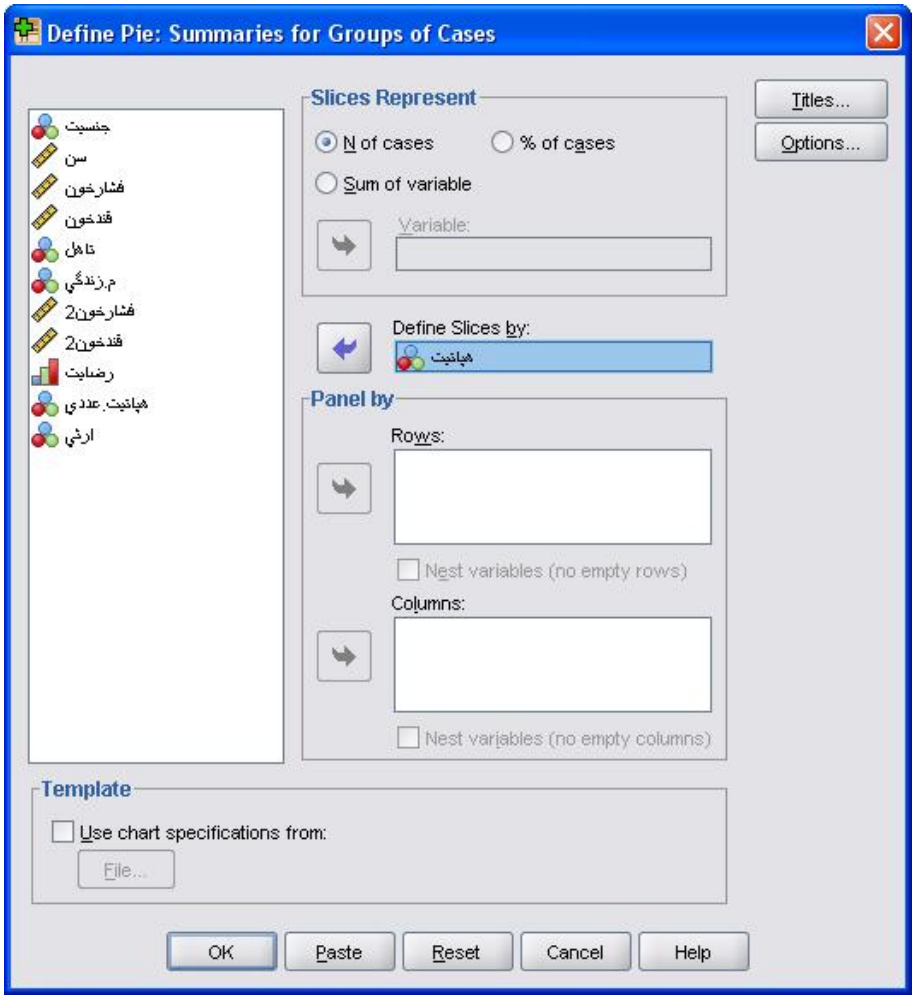
با انتخاب فرمان ... Pie در منوی Graphs جعبهٔ گفتگوی زیر ظاهر می‌شود:



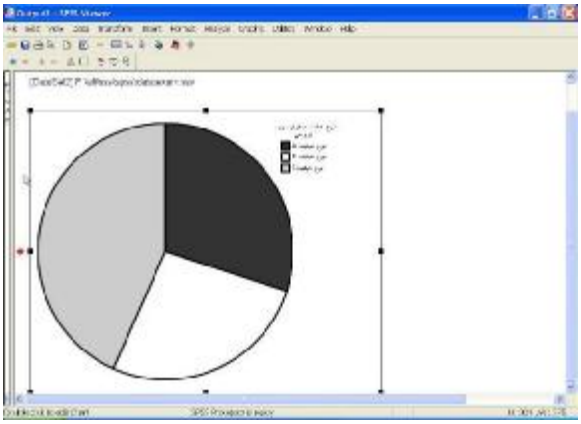
تنوع داده‌ها مانند نمودارهای قبلی است. اما در هر حالت تنها نمودار دایره‌ای ساده ترسیم مـی‌شـود و انواع پشته‌ای یا خوشه‌ای در نمودار دایره‌ای معنی ندارد.

7ـ5ـ1ـ نمودار دایره‌ای ساده

با انتخاب حالت نخست داده‌ها جعبهٔ گفتگوی زیر ظاهر می‌شود:

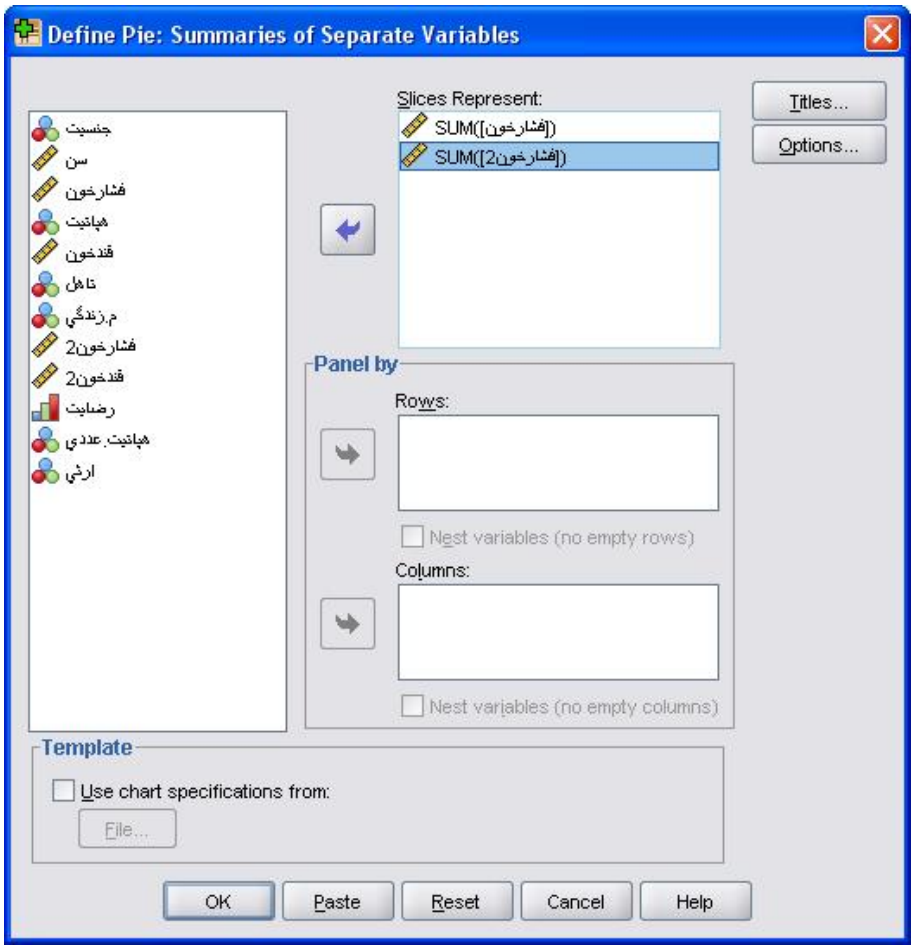


متغیری که سطوح آن مبنای ایجاد برشهایی از دایره خواهد بود در جعبهٔ Define Slices by وارد می‌کنیم و مقداری که مبنای محاسبه درصد مساحت ناحیه برش خورده از کل است، در جعبهٔ Slices Represent تعیین می‌شود. برای مثال متغیر نوع هپاتیت را وارد Define Slices by نموده و از Slies Represent حالت % Of Cases را انتخاب نموده و OK را می‌زنیم، نتیجه به صورت شکل زیر است:



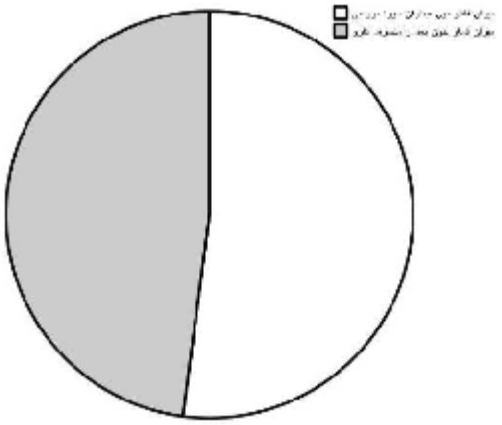
7ـ5ـ2ـ نمودار دایره‌ای متغیرهای مجزا

برای ترسیم این نمودار باید از پنجرهٔ Pie Charts گزینهٔ Summaries of Separate Variables را انتخاب نموده و Define را انتخاب می کنیم تا پنجره زیر باز شود:



دو متغیری را که می‌خواهیم مجموع آنها را با هم مقایسه نمائیم وارد قسمت Slices Represent می‌نمائیم. میزان سطح اشغال شده توسط هر برش، مستقیماً به مقدار کمیت بستگی دارد. مثلاً اگر مقدار مجموع یکی از متغیرها 9 و دیگری 1 باشد دایره را به دو قسمت که هر یک به ترتیب 90% و 10% درصد آن را تشکیل می‌دهند تقسیم می‌کند.

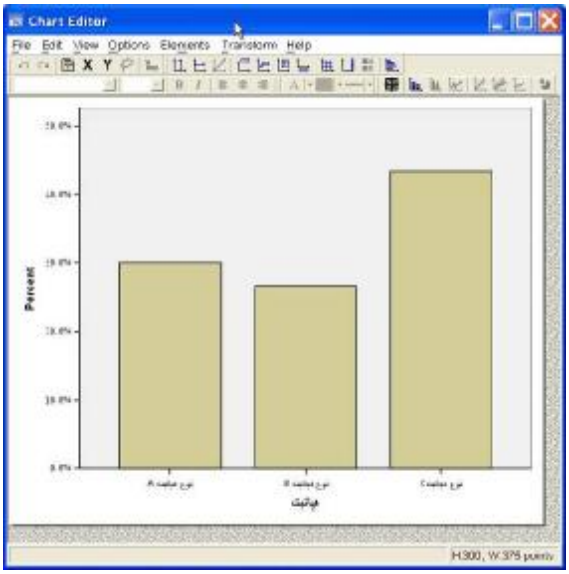
در مثال زیر نمودار دایره‌ای برای مقایسه فشار خون قبل و بعد را به صورت زیر ملاحظه می‌کنید:



همانطور که ملاحظه می‌شود فشار خون بعد از مصرف دارو کاهش یافته است.

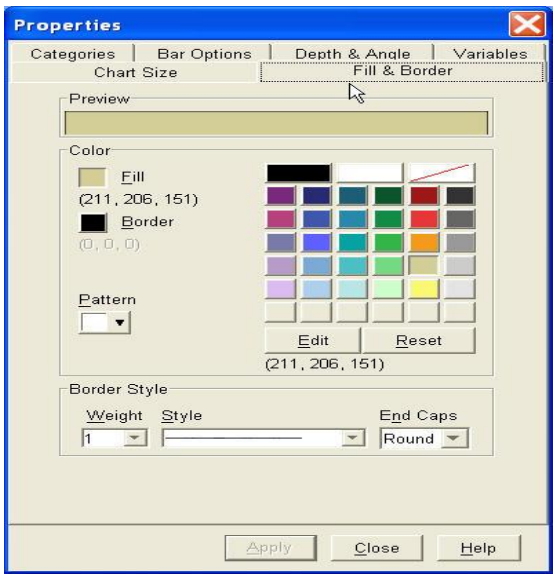
7ـ6ـ ویرایش نمودارها:

همانگونه که در قبل ملاحظه گردید، نمودارها در پنجرهٔ خروجی به نمایش در می‌آیند اما برای ویرایش آنها از پنجرهٔ دیگری به نام Chart Editor استفاده می‌شود. این پنجره دارای ابزارها و منوهایی است که از طریق آنها می‌توان هر گونه تغییر بر روی شکل‌ها و نمودارها نظیر تغییر رنگ، تغییر فونت، اضافه کردن توضیح به نمودارها، ایجاد برچسب و ... را بر روی نمودارها اعمال نمود. برای باز شدن پنجرهٔ Chart Editor کافی است و بر روی شکل مورد نظر که می‌خواهیم آن را ویرایش نمائیم دوبار Click نمائیم تا مانند شکل زیر پنجرهٔ Chart Editor باز شود:



کاربر می‌تواند در پنجرهٔ Chart Editor بر روی هر قسمت از شکل مورد نظر Click نماید تا ابزارهای مربوط به تغییر در آن قسمت باز شوند. این عمل مانند ویرایش شکل در نرم‌افزار Excel می‌باشد.

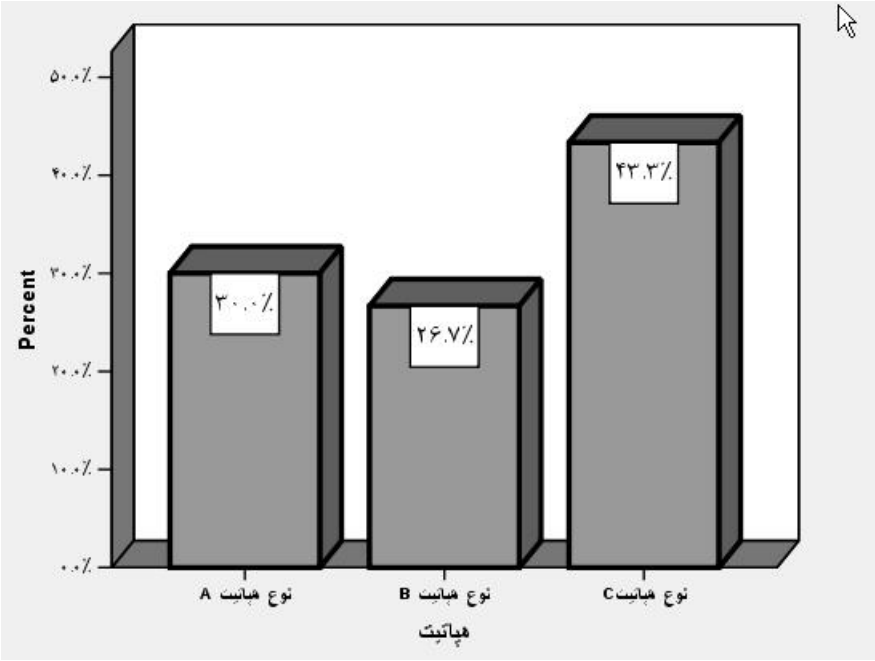
برای مثال کاربر برای تغییر رنگ نمودارهای میله‌ای می‌تواند بر روی رنگ دو بار Click نماید تا پنجرهٔ Properties به صورت زیر باز گردد:



در پنجرهٔ Properties می‌توان انواع خصوصیات شکل‌ها نظیر رنگ (در قسمت Fill&Border)، خصوصیات میله‌ها (در قسمت Bar Options)، اندازهٔ نمودار (در قسمت Chart Size)، ابعاد میله‌ها (در قسمت Depth&Angle) و گروه‌بندی متغیرها (در قسمت Categories) را کنترل نمود.

همچنین کاربر می‌تواند فونت و اندازهٔ هر قسمتی از برچسب‌ها در شکل را توسط امکانات تغییر فونت و اندازه در جعبه ابزار پنجرهٔ Chart Editor پس از انتخاب قسمت مورد نظر ویرایش نماید.:

برای برچسب دار نمودن سطوح نمودار میله‌ای یا انواع دیگر نمودارها نظیر نمودار دایره‌ای و ... می‌توان از زیرمنو Show Data Labels در منو Elements و انتخاب یکی از حالات گزینه Custom در قسمت Label Position استفاده نمود. بعد از انتخاب دکمهٔ Apply نمودار در پنجرهٔ Chart Editor دارای برچسب خواهد شد.

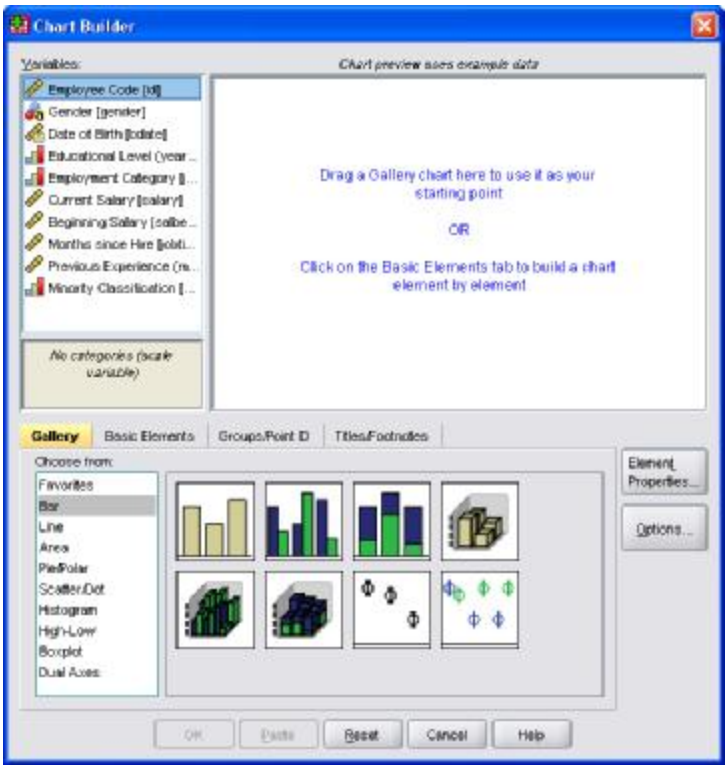


ذکر این نکته ضروری است که برای اعمال تغییراتی که در پنجرهٔ Chart Editor انجام گرفته است بر روی نمودار اصلی که در پنجرهٔ خروجی (Viewer) قرار دارد باید پنجرهٔ Chart Editor در پایان انجام

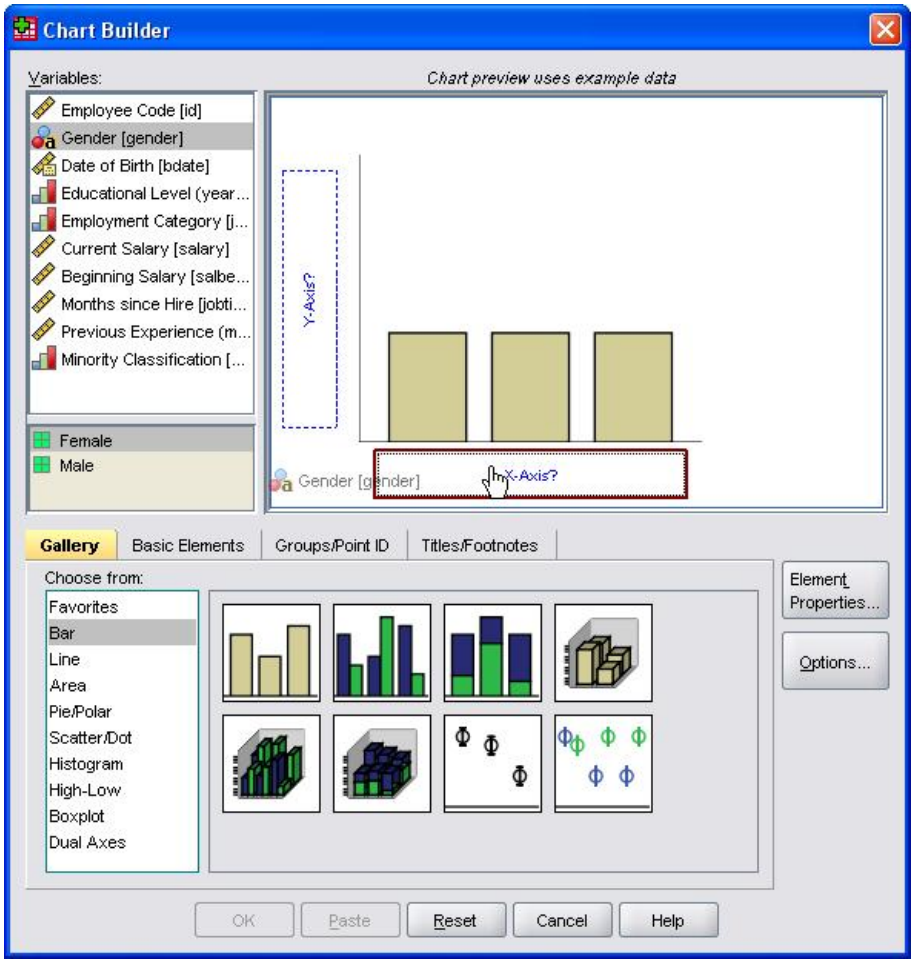
عمل ویرایش بر روی متغیرها بسته شود، آنگاه کلیه تغییرات داده شده بر روی نمودارها در نمودار پنجـرۀ خروجی قابل رؤیت خواهد بود.

7ـ7ـ ویرایش نمودارهابه صورت خود کار:

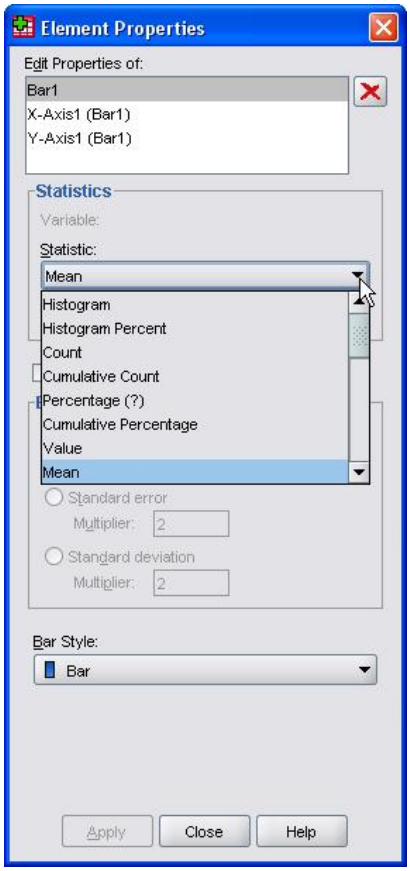
با استفاده از زیر منو Chart Builder امکان کشیدن شکل با کمک نرم افزار وجـود دارد بـا زدن ایـن گزینه پنجره ایی به شکل ذیل باز می شود.



از قسمت choose from شکل مورد نظر را انتخاب می نمـائیم . بـرای نمونـه گزینـه Bar را بـرای کشیدن یک نمودار میله ایی انتخاب می کنیم . آنگاه در قسمت روبرو انواع نمودار میله ایی را نمایش می دهـد. بـرای مثـال نمـودار میلـه ایـی سـاده را ماننـد شـکل ذیـل انتخـاب مـی کنـیم.

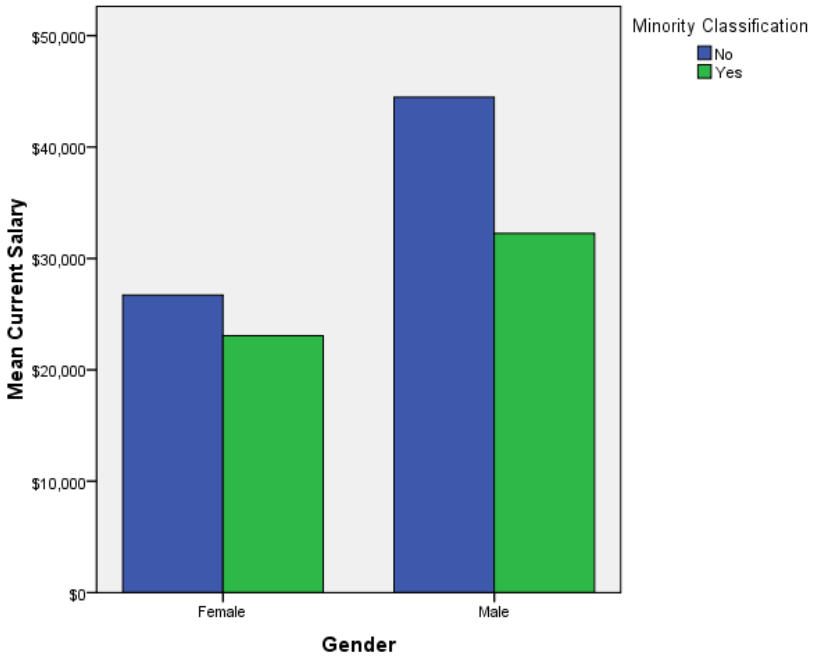


آنگاه در قسمت نمایش بالای صفحه نمودار محور x و محور y ها نمایان می شود . باید مد نظر داشت که همانند نمودار میله ایی روی محور x ها همواره متغیر گسسته و روی محور y ها علاوه بر پارامترهـایی مانند تعداد و درصد می تواند خصوصیات یک متغیر پیوسته نظیر میانگین و ... هم قرار داشته باشد. برای نمونه متغیر جنسیت را از فایل employee data انتخاب کرده و در قسمت متغیـر x قـرار دهـیم. روی متغیر y کلیک راست نموده و گزینه properties window یا Ctrl+T را انتخاب می کنیم تـا پنجـره زیر باز شود.



حال در قسمت Edit properties of: می توان خصوصیات هر کدام از بخش های نمودار را مشخص نمود. با انتخاب Bar می توان مشخص نمود که محور y ها کدامیک از پارامترهای زیر باشد. برای مثال با انتخاب count این امکان فراهم می آید که تعداد هر سطوح متغیر گسسته در نمودار میله ایی ارائه شود. برای مثال در نظر داریم تا یک نمودار خوشه ایی از میانگین درآمد هر یک از سطوح مختلف جنسیت(Gender) ترسیم نمائیم برای این منظور متغیر جنسیت(Gender) را در محور x ها قرار داده متغیر ارثی بودن(Minority) را وارد قسمت Cluster X on: می نمائیم تا مبنی خوشه بندی قرار گیرد و متغیر درآمدرا (Salary) وارد محور y می نمائیم. حال با انتخاب Bar از قسمت Edit properties of: و انتخاب mean از قسمت Statistics: این امکان را فراهم می آوریم که نمودار

خوشه ایی متوسط میزان درآمد هر یک از سطوح جنسیتی به تفکیک اقلیت بودن و نبودن آنها به صورت نمودار ذیل بعد از تائید به صورت ذیل ارائه شود.



با انتخاب هر یک از شکلهای مورد نظر امکان کشیدن آنها با انتخاب متغیرهای مورد نظر و قـرار دادن آنها در محل های مشخص شده برای هر متغیر و انتخاب پارامترهای مـورد نظـر از قسـمت properties وجود دارد.

تمرین:

1- نمودار میله ایی متغیرهای گسسته فایل cars را بدست آورید.

2- نمودار میله ایی متغیرهای گسسته فایل employe data را به تفکیک رده شغلی بدست آورید.

3- نمودار خطی چند جمله ایی متغیرهای گسسته فایل بیماران هپاتیتی را به تفکیک متغیر جنسیت بدست آورید.

4- نمودار میله ایی مقایسه میانگین فشار خون بیماران را به تفکیک جنسیت در فایل بیماران هپاتیتی بدست آورید.

5- نمودار میله ایی انباشته میانگین فشار خون بیماران را به تفکیک جنسیت در فایل بیماران هپاتیتی در دو گروه شهری و روستایی بدست آورید.

6- نمودار میله ایی خوشه ایی میانگین فشار خون بیماران را به تفکیک ارثی و غیر ارثی در فایل بیماران هپاتیتی در دو گروه شهری و روستایی بدست آورید.

7- تفاوت میان نمودارهای خوشه ایی و نمودارهای انباشته را با ذکر مثال بیان نمائید.

8- با استفاده از نمودار خودکار نمودار میانگین فشار خون قبلا و بعد را مقابل هم ترسیم نمائید.

9- با استفاده از روش خودکار ترسیم نمودارها، ماکزیمم مقدار قند خون مبتلایان مرد را در انواع مختلف ابتلا به هپاتیت بیان نمائید.

فصل هشتم

اهداف آموزشی فصل هشتم

فراگیر پس از مطالعه این فصل باید بتواند:

- مراحل به دست آوردن میانگین متغیرهای پیوسته در سطوح مختلف متغیرهای گسسته در نرم افزار Spss.16 را با ذکر مثال انجام دهد.

- مراحل انجام آزمون T برای مقایسه یک نمونه را در نرم افزار Spss.16 با ذکر مثال شرح دهد.
- مراحل انجام آزمون T برای مقایسه دو نمونه مستقل را در نرم افزار Spss.16 با ذکر مثال شرح دهد.
- مراحل انجام آزمون T برای مقایسه دو نمونه جفت شده را در نرم افزار Spss.16 با ذکر مثال شرح دهد.

آزمون فرضهای میانگین

8ـ1ـ مقدمه:

آزمون فرض‌های میانگین یکی از پرکاربردترین روشهای آمار استنباطی در شاخه‌های مختلف علوم است. از آنجا که میانگین یکی از پارامترهای مهم یک متغیر پیوسته در جامعه آماری است، قضاوت در مورد وضعیت و نحوه رفتار میانگین‌ها می‌تواند راهگشای بسیاری از مسائل باشد. در این فصل به بررسی انجام آزمون فرض‌های میانگین به یکی از سه صورت زیر می‌پردازیم:

الف ـ آزمون فرض میانگین یک متغیر با یک مقدار ثابت (One Sample T.test).

ب ـ آزمون میانگین‌های دو نمونه مستقل (Independent Samples T.test).

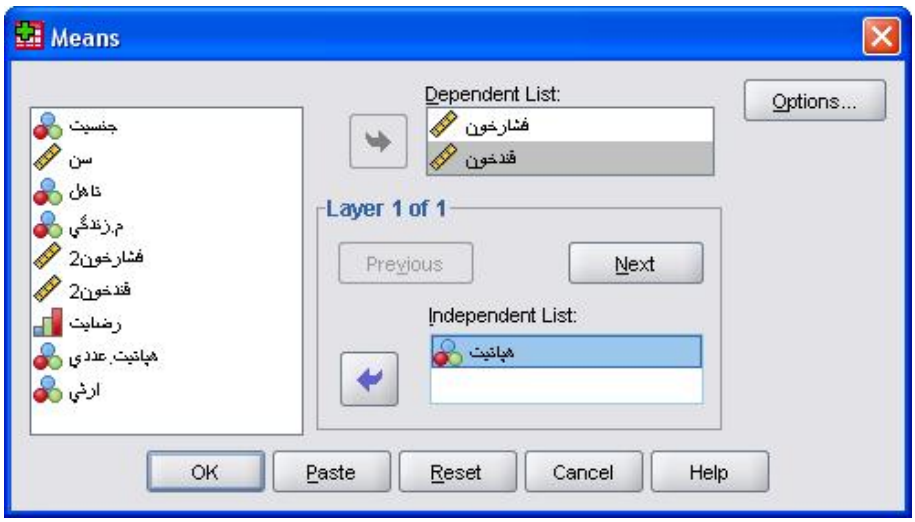
ج ـ آزمون فرض میانگین‌های دو نمونه جفت شده (Paired-Samples T.test)

در ادامه هر یک از این روشها مورد بررسی قرار گرفته‌اند.

8ـ2ـ میانگین‌ها (Means):

قبل از بیان آزمون فرضها در نرم‌افزار SPSS با استفاده از آیتم Means در زیرمنو Compare Means و منو Analyze نحوه بدست آوردن میانگین متغیرهای پیوسته در سطوح مختلف متغیرهای

گسسته را بیان می‌نمائیم. با انتخاب Means پنجرهٔ زیر باز می‌شود.



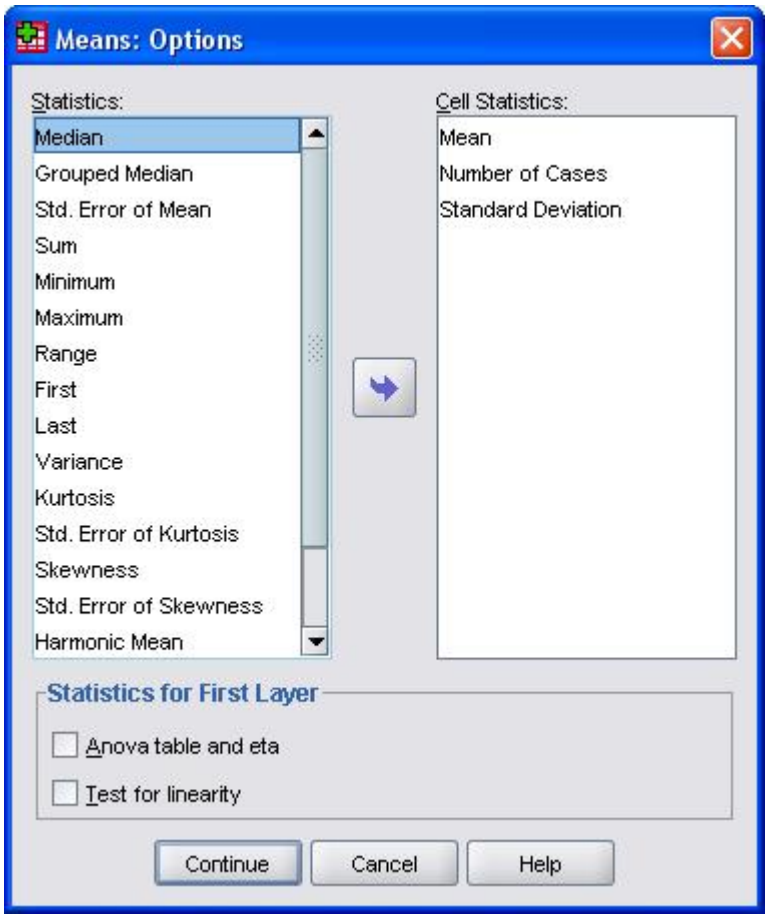
با استفاده از این ابزار می‌توان اطلاعات یک متغیر پیوسته را در سطوح متغیرهای گسسته بدست آورد. برای مثال از فایل جاری Exam، متغیر پیوسته فشار خون و قند خون را وارد پنجـرهٔ Dependent List می‌نمائیم و متغیر گسسته نوع هپاتیت را وارد پنجرهٔ Independent List می‌نمائیم ودکمه OK را فعال می‌کنیم. خروجی به صورت زیر ظاهر می‌شود:

Report

هپاتیت		فشارخون	قندخون
نوع هپاتیت A	Mean	11.2500	339.8889
	N	8	9
	Std. Deviation	2.60494	68.87932
نوع هپاتیت B	Mean	10.4286	474.1250
	N	7	8
	Std. Deviation	1.61835	145.7036
نوع هپاتیتC	Mean	10.4615	456.5385
	N	13	13
	Std. Deviation	2.29548	116.1942
Total	Mean	10.6786	426.2333
	N	28	30
	Std. Deviation	2.19517	124.0080

همانطور که ملاحظه می‌شود اطلاعاتی در مورد کمیت‌های تعـداد داده‌هـا (N)، میـانگین (Mean) و انحراف معیار (Std.Deviation) برای دو متغیر پیوسته فشار خون و قند خون در سطوح مختلف متغیـر گسسته نوع هپاتیت ارائه شده است.

با انتخاب دکمهٔ ... Options پنجرهٔ زیر باز می‌شود:



با انتخاب هر یک از کمیت‌ها از قسمت Statistics و انتقـال آنهـا بـه قسـمت Cell Statistics می‌توان این مقادیر را نیز برای هر یک از متغیرهای پیوسته در هر یک از سـطوح متغیرهـای گسسـته در خروجی دریافت نمود.

با انتخاب Anovatable And eta در قسـمت Statics For First Layer جـدول آنـالیز واریـانس و

همچنین ضریب Eta در خروجی ظاهر می‌شود که در فصول بعد بیشتر در مورد آنها صحبت خواهد شد. همچنین با انتخاب Test Of Linearity آزمون خطی بودن نیز بر روی داده‌ها انجام می‌گیرد. البته برای گرفتن نتایج آزمون خطی بودن حتماً باید متغیر گسسته از نوع عددی و دارای بیش از دو سطح باشد.

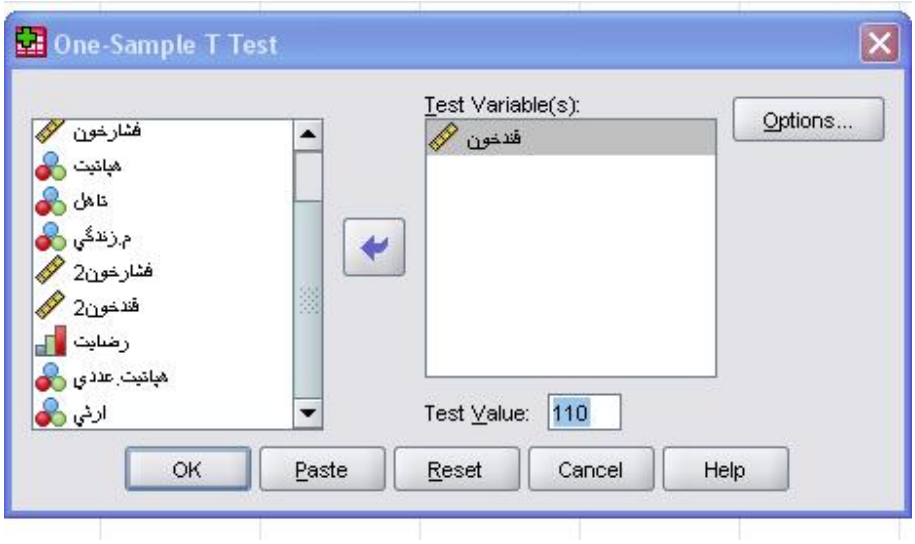
3-8- آزمون T مقایسه میانگین یک نمونه (One Sample T.test)

هر گاه یک نمونه از داده‌های یک متغیر پیوسته موجود باشد و بخواهیم قضاوت کنیم که آیا میانگین واقعی این متغیر در جامعه آماری برابر عددی خاص هست یا نه از این روش استفاده می شود..

برای مثال طول عمر متوسط لامپ‌های تولید با یک روش خاص 1000 ساعت است، یک شرکت ادعا کرده است که لامپ‌های تولیدی آن شرکت دارای طول عمری بیشتر از لامپ‌ها با روش قبلی است. برای آزمون ادعای این شرکت لازم است یک نمونه از لامپ‌های تولیدی آن استخراج گردد و میانگین لامپ‌های نمونه با مقدار ثابت 1000 که طول عمر لامپ‌های تولید شده با روش قدیمی است آزمون شود برای این منظور از روش آزمون T برای مقایسه میانگین یک نمونه با یک عدد استفاده می‌شود.

مثالی دیگر: یک شرکت داروسازی ادعا می‌کند دارویی را تولید کرده است که باعث کاهش فشار خون به مقدار قابل توجهی می‌شود. برای اثبات ادعای این شرکت یک نمونه از افراد با وضعیت جسمانی طبیعی را در نظر می‌گیریم و میانگین فشار خون آنها بعد از استفاده از داروی شرکت را با عدد 12 که میانگین فشار خون افراد طبیعی است آزمون می‌نمائیم. در صورتی که میانگین این نمونه از افراد در آزمون فرض $H_0 : \mu = 12$ رد شود و $\mu < 12$ باشد ادعای شرکت ثابت می‌شود.

برای آزمون فرض میانگین گزینهٔ One Sample T.test را از زیر منوی Compare Means و منو Analyze انتخاب می‌کنیم تا پنجرهٔ گفتگوی زیر باز شود:



متغیری را که می‌خواهیم میانگین آن را با عدد خاصی مقایسه نمائیم وارد قسمت Test Variable می‌نمائیم و عدد خاص را در قسمت Test Value وارد می‌نمائیم. برای مثال قند خون یک فرد عادی برابر عدد 110 است، می‌خواهیم بدانیم که در نمونهٔ آماری گرفته شده از بیماران مبتلا به هپاتیت آیا قند خون آنها با افراد عادی متفاوت است. به همین منظور متغیر قند خون را وارد Test Variable و عدد 110 را وارد Test Value نموده و OK را می‌زنیم.

خروجی به صورت جدول زیر ظاهر می‌شود:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
قندخون	30	426.2333	124.00798	22.64066

One-Sample Test

	Test Value = 110				
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower / Upper
قندخون	15.459	29	.000	316.23333	269.9080 / 362.5587

ابتدا در جدول One Sample Statistics اطلاعاتی نظیر میانگین، انحراف معیار و خطای استاندارد میانگین ارائه می‌گردد.

جدول One-Sample T.test جدولی است که بر اساس محتویات آن می‌توان در مورد رد یا عـدم رد فرض $H_0 : m = 110$ در مقابل $H_1 : m \neq 110$ قضاوت نمود.

برای قضاوت در مورد فرض $H_0 : m = 110$ سه راه زیر به وسیله محتویات جدول امکان‌پذیر است.

- راه اول: مقدار آمارهٔ t را با درجهٔ آزادی آن با مقدار توزیع t در جداول آماری که معمـولاً در انتهـای کتابهای آماری موجود است مقایسه نمائید و اگر این مقدار از مقادیر جدول بزرگتـر اسـت فـرض H_0 رد می‌شود.

- راه حل دوم: مقدار آماره sig یا سطح معنی‌داری را با مقدار خطـای نـوع اول تحقیـق (α) مقایسـه نمائیم اگر $sig > \alpha$ فرض H_0 رد نمی‌شود و اگر $sig \leq \alpha$ فـرض H_0 رد مـی‌شـود. در ایـن مثـال چـون sig = 0/000 است فرض برابری میانگین قند خون بیماران مبتلا به هپاتیت با افراد عادی رد می‌شود.

- راه حل سوم: استفاده از دو ستون آخر جدول One Sample T.test یعنی حد بالا و پـائین فاصـله اطمینان 95% برای اختلاف میانگین (μ ـ 110) است. اگر عدد صفر شـامل ایـن بـازه باشـد یعنـی عـدد Lower منفی و عدد Upper مثبت باشد به این معنی است که با احتمال 95% اختلاف میانگین μـ110 برابر عدد صفر شده است و μ=110 است و فرض H_0 رد نمی‌شود.

در این مثال Lower و Upper هر دو مثبت هستند پس چون صفر در داخل این بازه قرار نمـی‌گیـرد فرض H_0 از این طریق هم رد می‌شود.

البته ذکر این نکته ضروری است که می‌توان سطح فاصلهٔ اطمینان را از 95% به هر عدد دلخـواهی در گزینه ... Options تغییر داد.

حال که فرض $H_0 : m = 110$ رد شده است این سؤال مطرح می‌شود کـه آیـا $m > 110$ یـا $m < 110$ است به عبارت دیگر میانگین قند خون افراد مبتلا به هپاتیت آیا کمتر از افراد عادی است یا بیشـتر از آن است، برای پاسخ به این سؤال دو راه حل وجود دارد.

- راه حل اول: به مقدار آمارهٔ Mean Different در جدول One-Sample T.test توجـه کنیـد ایـن

مقدار در حقیقت اختلاف میانگین نمونه و عدد ثابت $(m-110)$ است، اگر مقدار آن عددی مثبت باشد به معنی آن است که $m-110>0$ و $m>110$ است و یا میانگین قند خون افراد مبتلا به هپاتیت بیش از افراد عادی است و اگر مقدار آن منفی باشد یعنی $m-110<0$ و یا $m<110$ که به معنی کمتر بودن میانگین قند خون افراد مبتلا به هپاتیت از افراد عادی است. در این مثال همانطور که ملاحظه می‌شود مقدار این آماره مثبت است پس $\mu>110$ خواهد بود.

- راه حل دوم: استفاده از مقدار فاصلهٔ اطمینان برای اختلاف میانگین‌ها نیز یک راه حل دیگر است. اگر حد بالا و پائین فاصله اطمینان هر دو اعدادی مثبت باشند به این معنی است که با احتمال مشخص (در اینجا 95%)، مقدار اختلاف میانگین $m-110$ مقداری بزرگتر از عدد صفر و مثبت است پس در نتیجه می‌توان نتیجه گرفت $m>110$ خواهد بود و اگر حدود بالا و پائین هر دو عددی منفی باشند به این معنی است که با احتمال مشخص $m-110$ عددی منفی خواهد بود و می‌توان نتیجه گرفت که $m<110$ است.

در مثال بالا با احتمال 95% مقدار حد بالا و پائین فاصله اطمینان مقادیر مثبت است پس $m-110>0$ و در نتیجه $m>110$ خواهد بود.

برای واضح‌تر شدن موضوع میانگین فشار خون افراد مبتلا را با عدد 12 که مقدار فشار خون افراد عادی است مقایسه نمائید. نتایج در قالب جدولOne-Sample T.test به صورت زیر است:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
فشار خون	28	10.6786	2.19517	.41485

One-Sample Test

	Test Value = 12					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
فشار خون	-3.185	27	.004	-1.32143	-2.1726	-.4702

فرض اولیه $H_0 : m = 12$ و فرض مقابل $H_1 : m \neq 12$ خواهد بود. با توجه به مقدار sig که برابر عدد 0/004 است و از مقدار خطای نوع اول تحقیق(α = 0/05 برای نمونه) کمتر است پس فرض H_0 رد می‌شود یعنی $m \neq 12$ خواهد بود. البته این نتیجه را می‌توان از روش فاصلهٔ اطمینان نیز بدست آورد چون عدد صفر در فاصله Lower و Upper صدق نمی‌کند پس $m \neq 12$ خواهد بود.

حال این سؤال مطرح می‌شود که $m > 12$ یا $m < 12$ است، برای پاسخگویی به این سؤال مقدار Mean Difference را نگاه می‌کنیم که برابر 1/32ـ است و چون عددی منفی است $\mu - 12 < 0$ و یا $m < 12$ خواهد بود، همین نتیجه را می‌توان از روش فاصله اطمینان نیز بدست آورد. با توجه به اینکه Lower و Upper هر دو عددی منفی هستند پس با احتمال 95% می‌توان گفت که $m - 12 < 0$ است و یا $M < 12$ خواهد بود.

ملاحظه می‌شود که با استفاده از هر یک از روشها به این نتیجه می‌رسیم که میانگین فشار خون افراد مبتلا به هپاتیت از میانگین فشار خون افراد عادی کمتر است.

8ـ4ـ آزمون T برای مقایسه میانگین دو نمونه مستقل (Independen-Samples T-Test)

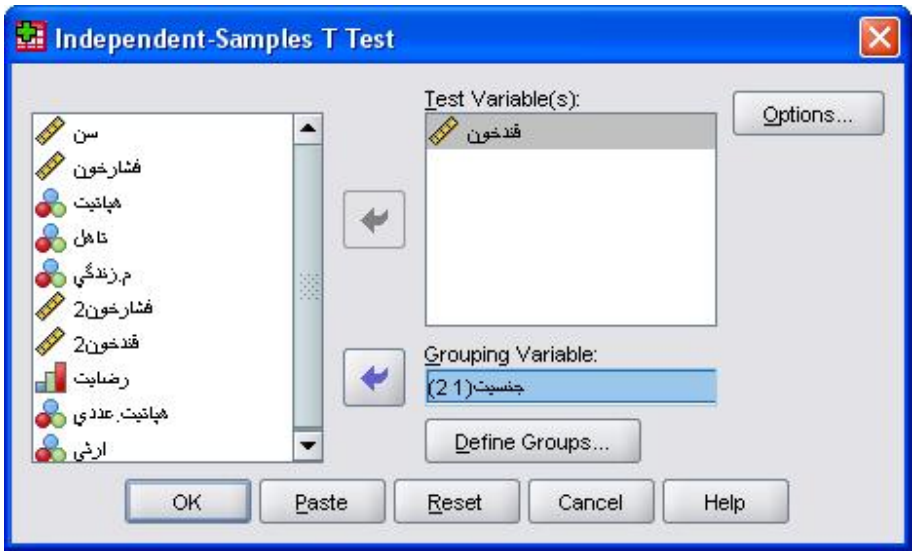
آزمون T برای مقایسه میانگین دو نمونهٔ مستقل زمانی به کار می‌رود که هدف مقایسه کردن دو نمونهٔ آماری باشد که از یکدیگر مستقل هستند، منظور از استقلال دو نمونه آن است که هیچ عضو مشترکی نداشته باشند، برای مثال نمونهٔ مردها و نمونهٔ زنها دو نمونهٔ مستقل از یکدیگر هستند یا نمونهٔ افراد متأهل و مجرد دو نمونهٔ مستقل از یکدیگر هستند چون هیچ فردی نمی‌تواند هم عضو جامعهٔ افراد متأهل باشد و هم عضو جامعهٔ افراد مجرد باشد.

پس هدف از آزمون T برای مقایسه میانگین دو نمونه مستقل، آزمون $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : m_1 \neq m_2$ است که در آن μ_1 میانگین متغیر پیوسته از نمونه اول و μ_2 میانگین متغیر پیوسته از نمونهٔ دوم است.

برای مثال اگر بخواهید میانگین فشار خون خانم‌ها و آقایان را با یکدیگر مقایسه کنید با توجه به اینکه جامعهٔ خانم‌ها و آقایان از یکدیگر مستقل هستند باید از روش مقایسه میانگین‌ها برای دو نمونهٔ مستقل این عمل صورت گیرد.

در پزشکی هر گاه می‌خواهند تأثیر دارویی را بر روی پارامترهای پیوسته جسمی مانند فشار خون، قند خون، تعداد گلبولهای قرمز و ... بدست آورند. دو نمونهٔ از افراد را انتخاب می‌نمایند، گروه اول افرادی هستند که به آنها دارو می‌دهند و به گروه تیمار معروف هستند و گروهی که به آنها دارو نمی‌دهندو به آنها گروه شاهد گفته می‌شود. اگر میانگین گروه شاهد و گروه تیمار را با یکدیگر مقایسه نمایند ، این عمل را باید از طریق آزمون T برای دو نمونه مستقل انجام دهند زیرا هر یک از عناصر تحقیق تنها می‌تواند در یکی از گروه‌های شاهد یا تیمار حضور داشته باشد.

برای انجام آزمون T برای دو نمونه مستقل باید گزینهٔIndependent Samples T.Test را از زیرمنوی Compare Means و منو Analyze انتخاب نمود تا پنجرهٔ زیر باز شود:



متغیر با متغیرهای پیوسته را وارد قسمت Test Test Variable (s) می‌نمائیم و متغیر گسسته دو بعدی را که نمونه را به دو قسمت مستقل تقسیم می‌کند وارد قسمت Grouping Variable می‌نمائیم.

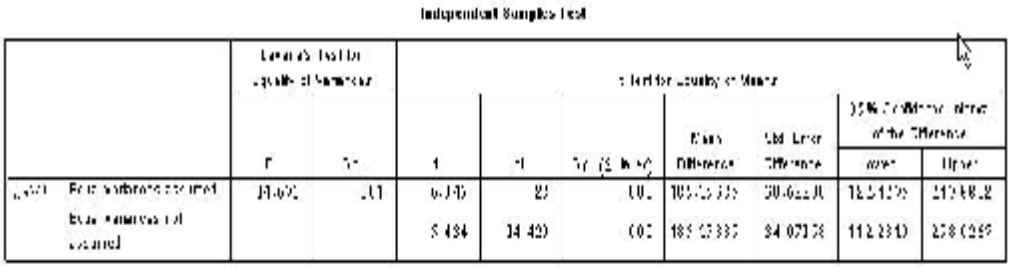
دکمهٔ Define Groups فعال می‌شود، با Click بر روی آن پنجرهٔ زیر باز می‌شود:

کد سطحی از متغیر گسسته را که می‌خواهیم میانگین آن بـه عنـوان میـانگین نمونـه اول وارد شـود داخل پنجرهٔ Group 1 و کد سطحی از متغیر گسسته که میانگین آن به عنوان میانگین نمونه دوم درنظر گرفته می‌شود را وارد پنجره Group 2 نموده وگزینـه Continue را فعـال وسپسـکلید OK را انتخـاب می‌نمائیم تا نتایج آزمون میانگین‌های متغیر پیوسته که در پنجرهٔ Test Variable (s) وارد نموده‌ایـم در دو سطح متغیر گسسته که در پنجرهٔ Grouping Variable و کد سطح‌هـای آن را در پنجـرهٔ Define Groups وارد کرده‌ایم در خروجی ظاهر شود.

برای مثال می‌خواهیم میزان قند خون مردها و زن‌ها را با یکدیگر مقایسـه نمـائیم. بـرای ایـن منظـور متغیـر قنـد خـون را وارد Test Variable (s) نمـوده و متغیـر جنسـیت را وارد پنجـره Grouping Variable نموده و با توجه به اینکه کد 1 در متغیر جنسیت متعلق به مردان اسـت در پنجـرهٔ Define Groups مقابل Group 1 کد 1 و مقابل Group 2 کد 2 را وارد می‌نمائیم. با این کار μ_1 میـانگین قنـد خون آقایان و μ_2 میانگین قند خون خانم‌ها خواهد بود، برای اجـرای فرمـان دکمـهٔ Continue وسـپس OK را انتخاب می کنیم . خروجی به صورت زیر است:

Group Statistics

	جنسیت	N	Mean	Std. Deviation	Std. Error Mean
قندخون	مرد	13	531.1538	117.09173	32.47540
	زن	17	346.0000	42.49412	10.30634



همانطور که ملاحظه می‌شود در جدول Group Statistics آمار توصیفی مربوط به میزان قند خون برای دو سطح خانم‌ها و آقایان بیان شده است. در جدول Independent Sample Test نتیجه آزمون $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : m_1 \neq m_2$ که در آن μ_1 میانگین قند خون آقایان و μ_2 میانگین قند خون خانم‌ها است بیان شده است.

ذکر این نکته ضروری است که آماره آزمون T برای مقایسه میانگین دو نمونهٔ مستقل هنگامی که واریانس دو نمونه برابر است با زمانی که دو نمونه واریانس‌های متفاوتی دارند، یکسان نخواهد بود. به همین دلیل قبل از آزمون $H_0 : \mu_1 = \mu_2$ باید بدانیم که آیا $H_0 : \sigma_1^2 = \sigma_2^2$ برای این منظور از آزمون لونت برای آزمون $H_0 : s_1^2 = s_2^2$ در مقابل $H_1 : s_1^2 \neq s_2^2$ استفاده می‌کنیم.

نتایج آزمون لونت در قالب آماره F و سطح معنی‌داری (sig) در ستون‌های دوم و سوم بیان شده است اگر $sig > \alpha$ باشد فرض H_0 رد نمی‌شود و اگر $sig \leq \alpha$ باشد فرض H_0 یعنی برابری واریانس‌ها در دو نمونه رد می‌شود و باید از روش نابرابری واریانس‌ها در آزمون $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : m_1 \neq m_2$ استفاده کرد.

در مثال فوق همانطور که می‌بینیم در صورتی که $\alpha = 0/05$ در نظر گرفته شود $sig = 0/001$ برای آزمون لونت خواهد بود که نشان می‌دهد $\sigma_1^2 \neq \sigma_2^2$ است حال برای قضاوت در مورد $H_0 : \mu_1 = \mu_2$ باید از سطر دوم یعنی سطر با شرط نابرابری واریانس (Equal Variance Not Assumed) استفاده نمود.

همانند حالت آزمون T برای یک نمونه یکی از ابزارهای قضاوت در مورد فرض $H_0 : \mu_1 = \mu_2$ مقدار

sig است که برابر 0/000 در سطر دوم است پس $\mu_1 \neq \mu_2$ یا فرض H_0 رد می‌شود. و درست مانند حالت قبل برای قضاوت در مورد $\mu_2 > \mu_1$ یا $\mu_1 > \mu_2$ به مقدار Mean Difference که همان آماره $\mu_1 - \mu_2$ نگاه می‌کنیم چون مقدار آن مثبت است پس $\mu_1 - \mu_2 > 0$ یا $\mu_1 > \mu_2$ است و در نتیجه میزان قند خون آقایان از خانم‌ها در بیماران مبتلا به هپاتیت بالاتر است.

-یک مثال دیگر: آیا میانگین فشار خون افراد مجرد با افراد متأهل در بیماران مبتلا به هپاتیت یکسان است؟

برای پاسخ گویی به این سؤال با توجه به اینکه نمونهٔ افراد مجرد و متأهل از هم مستقل هستند از روش آزمون T برای دو نمونه استفاده می‌نمائیم. متغیر فشار خون را وارد Test Variable (s) نموده و متغر دو سطحی تأهل را وارد Grouping Variable نموده و در پنجرهٔ Define Groups در پنجرهٔ Group 1 عدد 1 (کد افراد مجرد) و در Group 2 عدد 2 (کد افراد متأهل) را وارد نموده و فرمان را اجرا می‌نمائیم. یک خروجی به صورت زیر ظاهر می‌شود:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
فشار خون	Equal variances assumed	1.05	.748	.200	26	.842	.36923	.64702	-1.96330	1.70176
	Equal variances not assumed			.200	23.140	.844	.36923	.64950	-1.97791	1.71638

ابتدا باید مشخص شود که آیا $H_0 : \sigma_1^2 = \sigma_2^2$ رد می‌شود یا نه؟ با توجه به مقدار Sig = 0/748 که از مقدار α (با در نظر گرفتن α=0/05) بزرگتر است پس فرض H_0 رد نمی‌شود یعنی باید از سطر Equal Variabances Assumed برای آزمون $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : \mu_1 \neq \mu_2$ استفاده نمود.

با توجه به مقدار sig=0/843 در سطر اول با فرض برابری واریانس‌ها می‌توان گفت که sig>α پس فرض $H_0 : \mu_1 = \mu_2$ رد نمی‌شود یا میانگین فشار خون افراد مجرد با میانگین فشار خون افراد متأهل

برابر است.

ذکر این نکته ضروری است که در مثال اول که قند خون آقایان با قند خون خانم‌ها متفاوت بود می‌توان استنباط کرد متغیر مستقل جنسیت که دارای دو سطح است بر روی متغیر وابسته قند خون که یک متغیر پیوسته است تأثیرگزار بوده است و در مثال دوم می‌توان استنباط کرد که چون فشار خون افراد متأهل و افراد مجرد متفاوت نبوده است پس متغیر گسسته دو سطحی وضعیت تأهل بر روی متغیر پیوسته فشار خون مؤثر نبوده است.

به طور کلی آزمون فرض در نمونهٔ مستقل ابزاری برای شناسایی تأثیر یک متغیر گسسته دو سطحی با سطوح مستقل بر روی یک متغیر پیوسته است.

می‌توان یک متغیر پیوسته را نیز در قسمت Grouping Variable وارد نمود و در جعبهٔ Define Groups گزینه Cut Point را انتخاب کرده و یک عدد را در آن وارد نمود. با این روش متغیر پیوسته وارد شده در Grouping Variable به دو قسمت کوچکتر یا مساوی آن عدد و بزرگتر از آن عدد کدگذاری شده و به عنوان یک متغیر گسسته دو بعدی مبنی گروه‌بندی قرار می‌گیرد.

8ـ5ـ آزمون T برای دو نمونه جفت شده (Paired-Sample T Test)

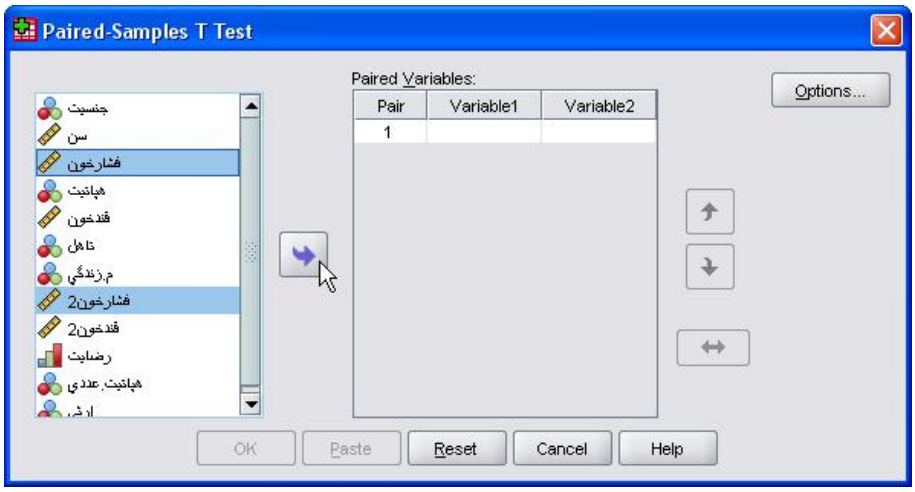
هر گاه قصد مقایسه میانگین‌های یک نمونه را در دو وضعیت متفاوت با یکدیگر داریم از این ابزار استفاده می‌نمائیم. برای مثال اگر طبق مقایسه میانگین‌های دو نمونه مربوط به فشار خون یک نمونه از بیماران را قبل از مصرف داروی خاص با میانگین فشار خون همین نمونه از بیماران بعد از مصرف دارو داریم باید از این روش استفاده نمائیم.

تفاوت این روش با روش میانگین‌های دو نمونه مستقل در آن است که این دو نمونه دارای عناصر یکسان هستند و تنها در شرایط متفاوتی اطلاعات از عناصر جمع‌آوری شده است ، در حقیقت اگر میانگین این دو نمونه با یکدیگر متفاوت باشند نشان دهنده مؤثر بودن تغییر شرایط و در غیر این صورت

نشان دهندهٔ بی‌اثر بودن تغییر شرایط در دو نمونه است.

برای مثال اگر میانگین فشار خون بیماران قبل از مصرف دارو با میانگین فشار خون آنها بعد از مصرف دارو متفاوت باشد نشان‌دهندهٔ اثر بخش بودن دارو می‌باشد.

برای انجام این آزمون بر روی گزینهٔ Paird-Sample T Test در زیر منو Compare Means و منوی Analyze ، Click کرده تا پنجره زیر باز شود:



برای مثال می‌خواهیم فشار خون بیماران را با فشار خون بعد از مصرف دارو مقایسه نمائیم برای این منظور دو متغیر فشار خون و فشار خون 2 را با هم وارد قسمت Paired Variables نموده و OK می‌نمائیم. جداول زیر در خروجی ظاهر می‌شوند:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	میزان فشارخون بیماران مورد بررسی	10.6786	28	2.19517	.41485
	میزان فشار خون بعد از مصرف دارو	9.8214	28	1.15642	.21854

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	میزان فشارخون بیماران مورد بررسي & میزان فشار خون بعد از مصرف دارو	28	.852	.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			
Pair 1 میزان فشارخون بیماران مورد بررسي - میزان فشار خون بعد از مصرف دارو	[illegible]	1.35303	[illegible]	[illegible]	1.38179	[illegible]	27	.002

همانطور که ملاحظه می‌شود جدول Paired Sample T.Test ملاک قضاوت فرض $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : \mu_1 \neq \mu_2$ است. همانند حالات توضیح داده شده در قبل مقدار sig=0/002 است که از مقدار α (α=%5) در نظر گرفته شده است کوچکتر است پس فرض H_0 رد می‌شود و با توجه به مقدار Mean که عددی مثبت است می‌توان نتیجه گرفت: $\mu_1 - \mu_2 > 0$ یا $\mu_1 > \mu_2$ به این معنی که فشار خون بعد از مصرف دارو (μ_2)کوچک تر از فشار خون قبل از مصرف دارو (μ_1) می‌باشد و می‌توان نتیجه گرفته که دارو در کاهش فشار خون افرادمبتلا به بیماری مؤثر بوده است.

تمرین:

1- انواع مقایسه میانگین ها در نرم افزار spss را بیان نمائید.

2- میانگین سن بیماران مبتلا به هپاتیت برابر با 35 سال را آزمون نمائید.

3- میانگین قند خون مبتلایان به هپاتیت نوع اول را با میانگین قند خون مبتلایان با هپاتیت نوع سوم آزمون نمائید.

4- میانگین وزن خودروهای آمریکایی با خودروهای اروپایی را در بین ماشین هایی که از سال 1973 به بعد تولید شده اند آزمون نمائید.

5- آیا می توان اعدا نمود که وضعیت تاهل افراد مبتلا به هپاتیت در مزان قند خون آنها موثر است/

6- آیا می توان ادعا نمود که میزان فشار خون خانمهای بالای 25 سال از آقایان بالاتر از 25 سال کمتر است، در صورت مثبت بودن ادعا میزان تفاوت چند درصد خواهد بود.

7- آیا می توان ادعا نمود که ماشین های آمریکایی وزن بیشتری نسبت به ماشین های ژاپنی دارند؟ میزان تفاوت وزن چقدر می باشد؟

فصل نهم

اهداف آموزشی فصل نهم

فراگیر پس از مطالعه این فصل می تواند:

- مراحل انجام آنالیز واریانس یک طرفه در نرم افزار Spss.16 را با ذکر مثال شرح دهید.
- مراحل انجام آنالیز واریانس یک متغیره در نرم افزار Spss.16 را با ذکر مثال شرح دهید.

آنالیز واریانس

9ـ1ـ مقدمه:

از مدلهای آنالیز واریانس یک بعدی زمانی استفاده می‌شود که هدف شناسایی تأثیر یک متغیر گسسته دارای بیش از دو سطح بر روی یک متغیر پیوسته باشد، حال اگر تعداد متغیرهای گسسته که در مدلهای آنالیز واریانس زیاد شود آنگاه ابعاد مدلهای آنالیز واریانس نیز افزایش می‌یابد برای نمونه یک مدل آنالیز واریانس دو بعدی مدلی است که در آن دو متغیر گسسته که حداقل یکی از آنها دارای بیش از دو سطح است بر روی یک متغیر پیوسته تأثیرگزار است و یا در مدل آنالیز واریانس چند بعدی هدف شناسایی اثر چندین متغیر گسسته که حداقل یکی از آنها دارای تعداد سطوح بیشتر از دو می‌باشد بر روی یک متغیر پیوسته است. در ادامه این فصل انجام مدل آنالیز واریانس یک بعدی که به آنالیز واریانس یک طرفه مشهور است و مدل آنالیز واریانس یک متغیره بیان شده است.

9ـ2ـ آنالیز واریانس یک طرفه (.. One-Way ANOVA):

همانطور که در بالا عنوان شد آنالیز واریانس یک طرفه یا یک بعدی، تأثیر یک متغیر گسسته دارای بیش از دو سطح را بر روی یک متغیر پیوسته مورد ارزیابی قرار می‌دهد. برای مثال اگر در داده‌های مربوط به بیماران مبتلا به هپاتیت کاربر بخواهد اثر نوع هپاتیت را که دارای 3 سطح می‌باشد بر روی میزان قند خون بیماران بدست آورد باید از آنالیز واریانس یک طرفه استفاده نماید و یا اگر در یک کارخانه مدیریت می‌خواهد میزان تولید محصول را در 4 خط تولید با یکدیگر مقایسه نماید باید از آنالیز واریانس یک طرفه استفاده نماید.

در آنالیز واریانس یک متغیر گسسته K سطحی بر روی یک متغیر پیوسته تاثیر می گذارد و در حقیقت هدف انجام آزمون

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

H_1 : [حداقل یکی از میانگین‌ها مخالف دیگر میانگین‌ها باشد]

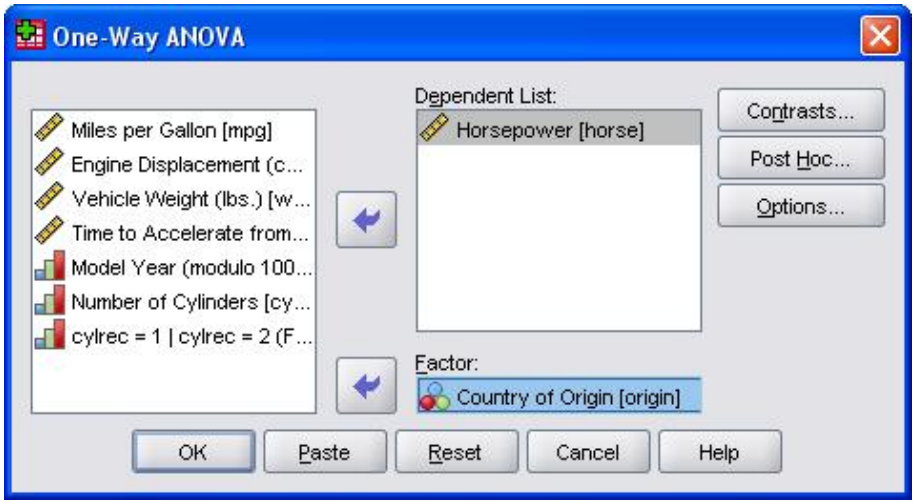
می‌باشد.

اگر فرض H_0 رد نشود به این معنی است که سطوح متغیر گسسته تأثیر یکسانی را بر روی متغیر پیوسته می‌گذارند یا در اصطلاح آماری متغیر گسسته تأثیری بر روی متغیر پیوسته ندارد و اگر فرض H_0 رد شود به این معنی است که حداقل تأثیر یکی از سطوح با دیگر سطوح متفاوت است پس متغیر گسسته بر روی متغیر پیوسته موثر بوده است.

برای مثال در شناسایی تأثیر متغیر گسسته نوع هپاتیت که دارای سطوح A، B و C می‌باشد فرض اولیه $H_0 : \mu_A = \mu_B = ... = \mu_C$ می‌باشد. اگر حداقل یکی از میانگین‌ها با دیگر میانگین‌ها متفاوت باشد یعنی نوع هپاتیت موثر بوده است و اگر فرض اولیه رد نشود یعنی انواع هپاتیت تأثیر یکسانی روی میانگین قند خون بیماران دارند پس نوع هپاتیت بر روی میزان قند خون مؤثر نخواهد بود.

برای انجام تحلیل واریانس یک طرفه گزینه One-Way ANOVA را از زیر منو Compare Means

در منوی Analyze انتخاب می‌نمائیم تا پنجرهٔ زیر باز شود:



در قسمت سمت چپ لیست کلیه متغیرهای فایل جاری ملاحظه می‌شود، در اینجا از فایل cars.sav استفاده شده است. در پنجرهٔ Dependent List متغیر یا متغیرهای پیوسته را وارد می‌نمائیم و در پنجرهٔ Factor متغیر گسسته بیش از دو سطح را وارد می‌نمائیم.

برای مثال می‌خواهیم تأثیر منطقه تولید ماشین (Origin) را بر روی قدرت اسب بخار ماشین‌ها (horse) بدست آوریم. در این صورت Horse را وارد Dependent List و Origin را وارد Factor می‌نمائیم و OK را انتخاب می‌کنیم. خروجی زیر ظاهر می‌شود:

ANOVA

Horsepower

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140868.6	2	70434.303	63.549	.000
Within Groups	437796.8	395	1108.346		
Total	578665.4	397			

همانطور که ملاحظه می‌شود خروجی جدولی به نام ANOVA را در خود جای داده است، محتویات

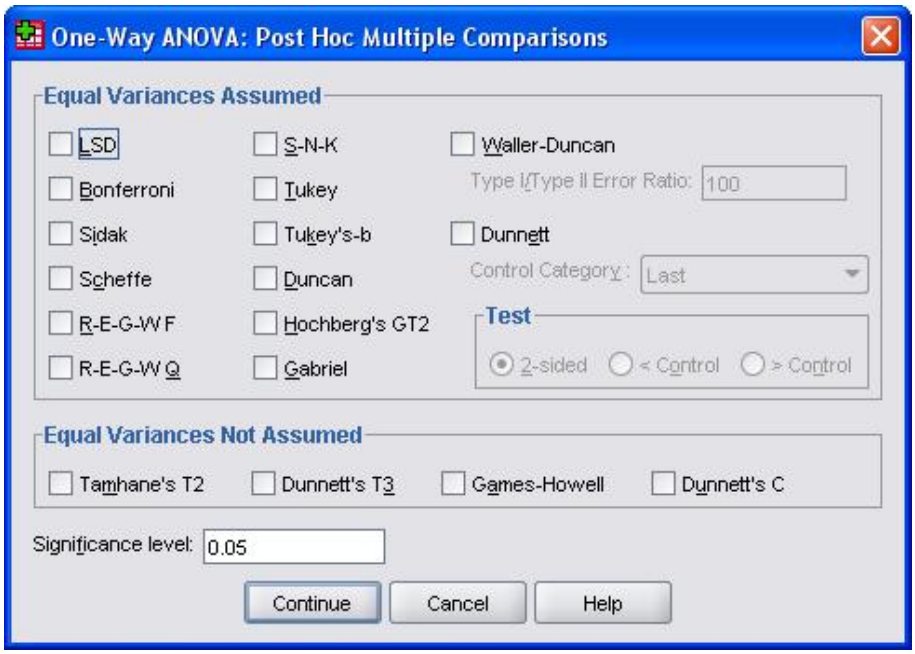
این جدول ملاک تصمیم‌گیری در مقابل فرض $H_0 : m_A = m_U = m_j$ می‌باشد که در آن μ_A میانگین قدرت اسب بخار ماشین‌های تولید شده در آمریکا، μ_U میانگین قدرت اسب بخار ماشین‌های تولید شده در اروپا و μ_j میانگین قدرت اسب بخار ماشین‌های تولید شده در کشور ژاپن است.

اگر فرض H_0 رد نشود می‌توان گفت که منطقهٔ تولید تأثیری بر روی میزان قدرت ماشین‌ها ندارد و اگر فرض H_0 رد شود به این معنی است که منطقهٔ تولید مؤثر بوده است.

ملاک رد شدن فرض H_0 در جدول آنالیز واریانس بالا، مقدار سطح معنی‌داری sig است. به طور کلی هر جایی که در نرم‌افزار SPSS آزمونی صورت می‌گیرد ملاک رد یا قبول فرض H_0 آزمون، مقدار سطح معنی‌داری (sig) می‌باشد، از آنجا که هر تحقیق آماری با خطا همراه است و همواره α از ابتدای تحقیق مشخص است، برای قضاوت در مورد فرض H_0، مقدار sig را با α مقایسه می‌نمائیم. اگر $sig \leq \alpha$ آنگاه فرض H_0 رد می‌شود و اگر $sig > \alpha$ آنگاه نمی‌توان فرض H_0 را رد نمود. در این خروجی اگر $\alpha = 0/05$ در نظر گرفته شود با توجه به مقدار $sig = 0/000$ می‌توان فرض H_0 را رد نمود و استنباط نمود که منطقه تولید یک ماشین بر روی قدرت آن موثر است.

حال این سؤال مطرح می‌شود که ترتیب میانگین‌ها به چه صورت است و یا کدام منطقه دارای ماشین‌های با قدرت بیشتر است؟ و کدام منطقه دارای ماشین‌های با قدرت اسب بخار کمتر است؟

برای پاسخ‌گویی به این سؤال نیاز به روشهایی است که به وسیله آنها میانگین‌ها را از کوچکتر به بزرگتر مرتب نمایند، برای استفاده از این روشها گزینهٔ ... Post Hoc را انتخاب می‌نمائیم تا پنجرهٔ زیر باز شود:

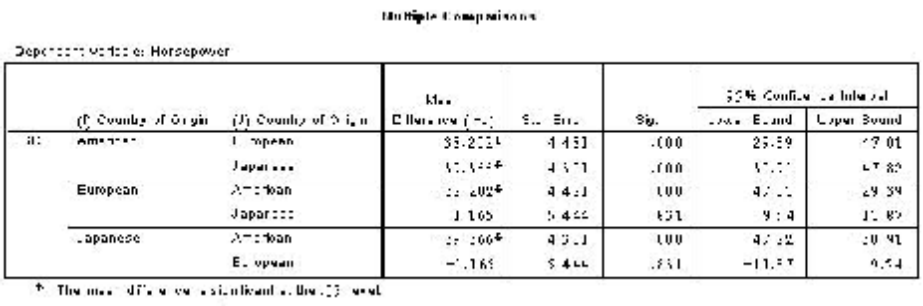


در این پنجره انواع روشهایی را که برای رتبه‌بندی میانگین‌ها بکار می‌روند در دو حالت با فرض برابری واریـانس‌هـا (قسـمت Equal Variances Assumed) و فـرض نـابرابری واریـانس‌هـا (Equal Variances Not Assumed) بیان شده است. ذکر این نکته ضروری است که روشهای تحلیل واریـانس در اکثر موارد با فرض برابری واریانس استفاده می‌شود مگر آنکه فرض نابرابری واریانس‌هـا در گـروه‌هـای مختلف برای متغیر پیوسته از قبل معلوم باشد.

در هر دو حالت بالا روشهای متفاوتی وجود دارد که هر کدام در شاخه‌های مختلف آمار کاربردی، مورد استفاده قرار می‌گیرند برای مثال در رشته‌های مهندسی از قبیل مهندسی صنایع محققان ترجیح می‌دهند از روش TUKEY یا Tuke's-b استفاده نمایند، در رشته‌های علوم اجتماعی Bonferroni، Duncan و Dunnett بیشتر مورد توجه است، در رشته کشاورزی Sidak، Scheffe و Gabriel مورد توجه است، اما محققان آماری LSD و Duncan را ترجیح می‌دهند.

برای مثال در آزمون قـدرت اسـب بخـار در سـه منطقـه روش LSD و Duncan را انتخـاب وسـپس Continue را فعال کرده ودکمه OK را انتخاب می‌نمائیم. خروجی‌های زیر به جـدول ANOVA اضـافه

می‌شود:



Multiple Comparisons

Dependent Variable: Horsepower

	(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	American	European	38.223*	4.451	.000	29.39	47.01
		Japanese	41.388*	4.571	.000	32.31	50.46
	European	American	-38.223*	4.451	.000	-47.01	-29.39
		Japanese	3.165	5.444	.561	-7.64	13.97
	Japanese	American	-41.388*	4.571	.000	-50.46	-32.31
		European	-3.165	5.444	.561	-13.97	7.64

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

در جدول Multiple Comparisons، به صورت دو به دو با روش LSD میانگین‌ها را با یکدیگر مقایسه نموده است. در سطر اول جدول میانگین اسب بخار ماشین‌ها در منطقهٔ America با European و Japanes و به همین طریق در ستون دوم میانگین European با دو تای دیگر و در ستون سوم میانگین Japanes با دو گروه دیگر مقایسه شده است. البته تعدادی از آنها تکراری هستند. تعبیر نتایج این آزمونها مانند مقایسه میانگین‌های دو نمونه است با توجه به اینکه مقدار sig با μ_A و μ_E و μ_j برابر صفر است می‌توان نتیجه گرفت $m_A \neq m_U$ و $\mu_A \neq \mu_j$ اما مقدار Mean Diffrence برای $m_A - m_U > 0$ و $\mu_A - \mu_J > 0$ پس $m_A > m_U$ و $\mu_A > \mu_j$ حال کافی است m_U را با μ_j مقایسه نمائیم.

همانطور که ملاحظه می‌شود sig=0/831 که sig>α در سطح α=%5 است پس $m_U = m_j$ خواهد بود و در نتیجه $m_A > m_U = m_j$ یا به عبارت دیگر ماشین‌های تولید شده در آمریکا دارای قدرت بیشتر از ماشین‌های تولیدی در ژاپن و اروپا هستند و ماشین‌های تولیدی در اروپا و ژاپن دارای میانگین قدرت یکسان می‌باشند.

یک راه دیگر رتبه‌بندی میانگین‌ها استفاده از خروجی جدول Homogeneous Subsets است.

Homogeneous Subsets

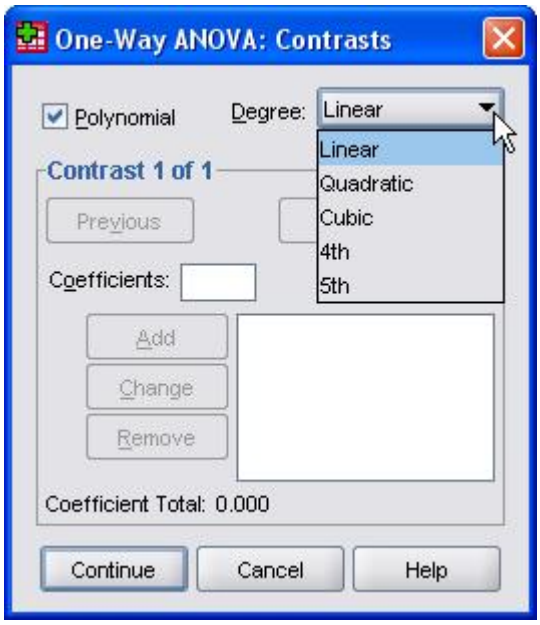
Horsepower

	Country of Origin	N	Subset for alpha = .05	
			1	2
Duncan	Japanese	79	79.84	
	European	71	81.00	
	American	248		119.20
	Sig.		.807	1.000

در این جدول با استفاده از روش Duncan و با مقدار مشخص α کـه در سـتون Subset for Alpha نوشته شده است، میانگین‌ها را به زیرگروههایی تقسیم می‌نماید. همانطور که در بـالا ملاحظـه مـی‌شـود میانگین دو گروه ژاپن و اروپا در گروه 1 و میانگین گروه آمریکا در گروه 2 قرار گرفتـه اسـت و میـانگین گروه دو از گروه یک بالاتر است و به عبارت دیگر نتیجهٔ $\mu_A > \mu_E = \mu_j$ با این ابزار در سطح α=%5 هم تأئید می‌شود.

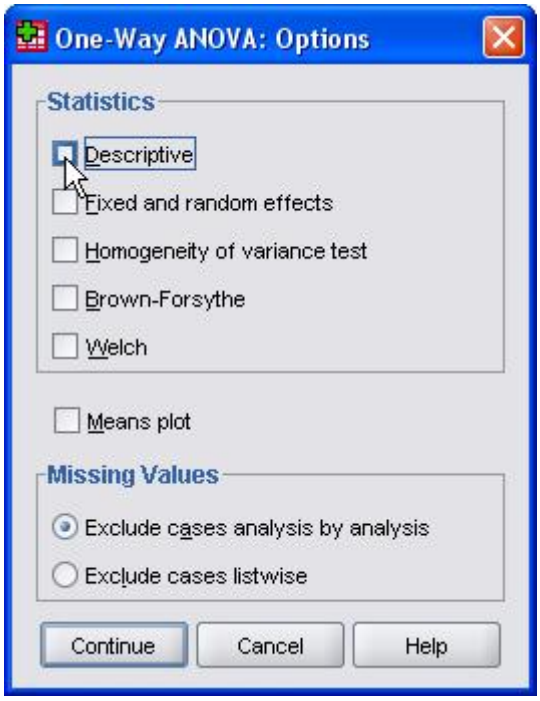
می‌توان مقدار α را در گزینهٔ Significance Level با توجه به مقدار مشخص آن در هـر تحقیـق در گزینهٔ Post Hoc... تغییر داد.

با انتخاب گزینهٔ Contrasts ... پنجره‌ای مانند زیر باز می‌شود:



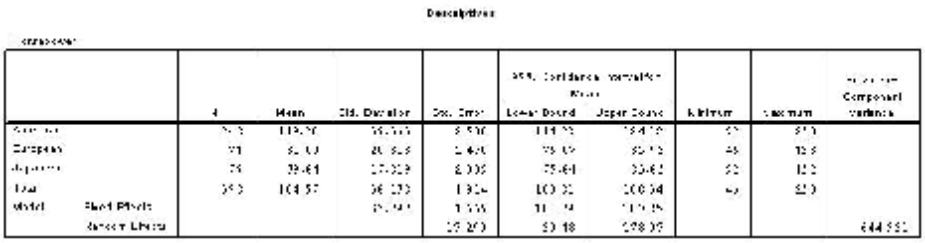
کاربر می‌تواند با انتخاب گزینهٔ Polynomial، درجه مدل را از خطی به مربع، مکعب، چهار یا پنج بعدی افزایش دهد که در آن صورت نتایج جدول آنالیز واریانس متناسب با آن تغییر می‌کند. علاوه بر آن کاربر می تواند ضریب هر یک از سطوح را نیز به صورت دستی از طریق گزینهٔ Cofficients وارد نماید. در حالت عادی در روش آنالیز واریانس مجموع ضرایب عدد صفر در نظر گرفته می‌شود.

با انتخاب گزینهٔ ... Options پنجرهٔ زیر باز می‌شود:



با انتخاب گزینهٔ Descriptive اطلاعات توصیفی متغیرهای پیوسته را در هر یک از سطوح متغیر گسسته در خروجی نمایش می‌دهد، با انتخاب گزینهٔ Fixed and Random Effects در پائین جدول Descriptive یک فاصله اطمینان 95% برای میانگین با فرض ثابت (Fixed) یا تصادفی (Random) بودن مدل آنالیز واریانس ارائه می‌شود. با انتخاب Homogeneity of Variance test ، آزمون فرض برابری واریانسهای متغیر پیوسته در هر یک از سطوح متغیر گسسته در خروجی در قالب همین نام ارائه

می‌شود که برای مثال قدرت اسب بخار به صورت زیر است:



		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
Model	Fixed Effects									
	Random Effects									

Test of Homogeneity of Variances

Horsepower

Levene Statistic	df1	df2	Sig.
49.765	2	395	.000

همانطور که ملاحظه می‌شود با توجه به sig = 0/000 فرض $H_0 : s_A^2 = s_U^2 = s_j^2$ رد می‌شود و در قسمت Post Hoc … باید از روشهای زیر مجموعهٔ Equal Variance Not Assumed استفاده نمود.

با انتخاب Brown-Forsthe یا Welch دو روش برای آزمون اعتبار آزمون برابری میانگین‌ها در اختیار کاربر قرار می‌گیرد.

بعد از انتخاب این گزینه‌ها خروجی زیر را می‌توان مشاهده کرد:

Robust Tests of Equality of Means

Horsepower

	Statistic[a]	df1	df2	Sig.
Welch	87.050	2	197.779	.000
Brown-Forsythe	117.944	2	354.799	.000

a. Asymptotically F distributed.

با هر یک از روشهای بالا اگر مقدار sig≤α باشد آنگاه می‌توان اعتبار آزمون برابری میانگین‌ها را تأئید نمود و اگر sig>α باشد به این معنی است که روش آزمون میانگین‌ها از اعتبار لازم به دلیل وجود داده‌های کم در بعضی از سطوح و .... برخوردار نیست. همانطور که ملاحظه می‌شود در مثال قدرت اسب بخار sig=0/000 است که نشان‌دهندهٔ صحت آزمون میانگین‌ها می‌باشد.

8ـ3ـ آنالیز واریانس یک متغیره (Univariate ANOVA):

تحلیل واریانس یک متغیره هنگامی مورد استفاده قرار می‌گیرد که بخـواهیم تـأثیر حـداقل دو متغیـر گسسته را بر روی یک متغیر پیوسته بدست آوریم.

برای ایـن منظـور بـا انتخـاب گزینـهٔ Univariate از زیـر منـو General Linear Model از منـوی Analyze پنجرهٔ زیر باز می‌شود:

همانطور که ملاحظه می‌شود در قسمت سمت چپ این پنجره لیستی از تمام متغیرهای فایـل جـاری مشاهده می‌شود.

می‌خواهیم تأثیر دو متغیر گسسته را بر روی متغیر پیوسـته مشـاهده نمـائیم، متغیـر پیوسـته را وارد پنجرهٔ Dependent Variable می‌نمائیم و متغیرهای گسسته را با توجه به ثابت یا تصـادفی بـودن آنهـا وارد پنجرهٔ Fixed Factor (s) یا Random Factor (s) می‌نامیم و در صورتی که متغیر پیوسته کمکی در اختیار است آن را وارد پنجرهٔ Covariate می‌نمائیم.

برای مثال در فایل مربوط به بیماران مبتلا به هپاتیت می‌خواهیم اثر متغیر نوع هپاتیت (در سه سطح)

و اثر جنسیت را بر روی متغیر میزان قند خون بیماران بدست آوریم، متغیر سـن را هـم بـه عنـوان یـک متغیر پیوسته کمکی در نظر می‌گیریم. برای این منظور مانند شکل بالا متغیرها را در پنجره‌های مربوطـه قرار داده و فرمان را از طریق Click بر روی دکمهٔ OK اجرا می‌نمائیم. خروجی زیـر در پنجـرهٔ خروجـی ظاهر می‌شود:

Between-Subjects Factors

		Value Label	N
جنسیت	1.00	مرد	13
	2.00	زن	17
هپاتیت .عددی	1	نوع هپاتیت A	9
	2	نوع هپاتیت B	8
	3	نوع هپاتیتC	13

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33318.648[a]	5	6663.730	28.100	.000
Intercept	69107.208	1	69107.208	291.458	.000
سن	1753.128	1	1753.128	7.393	.012
هپاتیت .عددی	17644.022	2	8822.011	37.202	.000
جنسیت	2524.806	1	2524.806	10.646	.004
جنسیت * هپاتیت .عددی	2176.004	2	1088.002	4.589	.021
Error	5728.352	24	237.015		
Total	108154.000	30			
Corrected Total	39046.963	29			

a. R Squared = .853 (Adjusted R Squared = .823)

همانطور که در بالا ملاحظه می‌شـود ابتـدا در جـدولی بـا نـامBetween-Subjects Factors تعـداد مشاهدات در هر یک از سطوح متغیرهای گسسته جنسیت و نوع هپاتیت بیماران بیان شده است.

در جـدول Tests of Between –Subjects Effects ماننـد جـدول ANOVA معیارهـایی بـرای سنجش اثر متغیرهای گسسته بر روی متغیر پیوسته ارائه شـده اسـت. در سـطر Corrected Model در حقیقت قدرت مدل در شناسایی عوامل مؤثر بر تغییرات قند خون بیان شده است. فرض،

H_0 : مدل مناسب نیست

H_i مدل مناسب است.

توسط آماره‌های این سطر آزمون می‌شود و معیار پذیرش یا رد آزمون در سطح خطـای نـوع اول α ،

مقدار sig است، چون sig=0/000 و sig≤α می‌توان نتیجه گرفت که فرض H_0 رد می‌شود و مدل مناسب است، همانطور که در پائین جدول مشاهده می‌شود مقدار R^2 که یک معیار برای قضاوت در مورد مناسب بودن مدل می‌باشد برابر 0/852 است که عددی مناسب است.

در سطر مربوط به Intercept فرض $H_0 : A = 0$ در مقابل $H_1 : A \neq 0$ از طریق مقدار sig آزمون می‌شود. مقدار A یا عوض از مبدأ یا مقدار ثابت، مقداری از قند خون است که تغییرات آن تحت تأثیر عواملی مانند جنسیت یا سن و نوع هپاتیت افراد تغییر نمی‌کند.

در سطر مربوط به سن، فرض

H_0 : عدم تاثیر سن بر روی قند خون

H_1 تاثیر سن بر روی قند خون

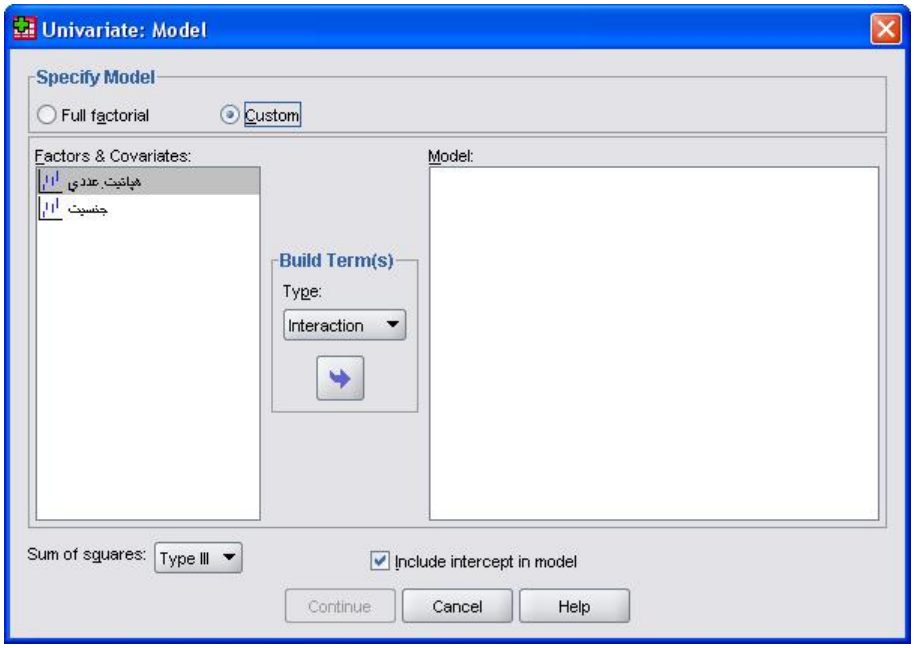
مورد ارزیابی قرار گرفته است و با توجه به sig=0/442 و α=0/05 می توان بیان نمود که متغیر سن بر روی میزان قند خون مبتلایان مؤثر نیست.

همان طور که در سطرهای مربوط به جنسیت و نوع هپاتیت ملاحظه می‌شود با توجه به مقدار sig=0/000 در این دو سطر، می‌توان گفت جنسیت و نوع هپاتیت بر روی میزان قند خون مبتلایان موثر بوده است. مؤثر بودن جنسیت بر روی میزان قند خون به معنی تفاوت میان میانگین قند خون خانم‌ها و آقایان و مؤثر بودن نوع هپاتیت بر روی قند خون به معنی تفاوت میانگین قند خون مبتلایان به هپاتیت در سه نوع هپاتیت است.

در سطر جنسیت * هپاتیت، اثر متقابل متغیر جنسیت و نوع هپاتیت بر روی قند خون بیماران بررسی شده است که با توجه بهsig=0,038 کمتر از α=%5 نشان دهنده اثربخش بودن تأثیر متقابل آنها بر روی میزان قند خون است.

کاربر با گزینه‌های موجود در پنجرهٔ ... Model امکان انتخاب مدل تحلیل واریانس را با فعال کردن

این کلید انجام می‌دهد. با انتخاب این گزینه پنجرهٔ زیر باز می‌شود:



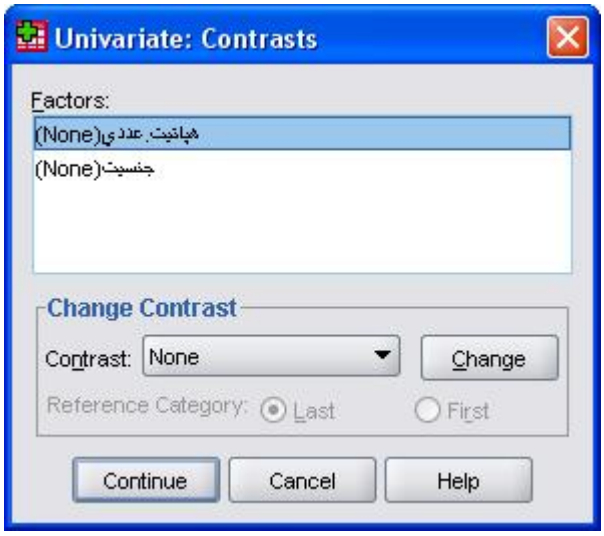
در جعبهٔ Specify Model امکان تعیین نوع مدل بر حسب گزینه‌های Full Factorial (مدل اشباع‌شده) یا Custom (مدل تعریف شده توسط کاربر یا سفارشی) وجود دارد، که مدلی ناقص است. در صورتی که کاربر به سلیقه خود مدلی را تعریف کند، جعبه‌ها فعال شده و امکان انتخاب اجزاء مدل برحسب نوع متغیرهای عامل و اثرات متقابل موجود در آن، فراهم می‌شود، جعبهٔ Factors & Covariates ورودی‌های مدل را به تفکیک عامل با نماد (F) یا متغیر کمکی با نماد (R) نمایش می‌دهد. پس از انتخاب یک یا چند اثر، می‌توان نوع ارتباط بین آنها را برحسب درجه اثرات، از جعبه Build Term(s) انتخاب کرد. با Click بر روی عامل مورد نظر و بر روی ↓ اثر اصلی یا متقابل با درجه مورد نظر را انتخاب و بر روی Click 4 می‌نمائیم تا اثر مورد نظر در جعبهٔ Model قرار گیرد. این مراحل بر روی تمامی اثرات مورد نیاز تکرار می‌شود.

پس از ساخت مدل، گزینه‌های انتخابی مدل شامل روشهای محاسبه مجموع مربعات و جملات خطا در بخش Sum Of Squares تعیین می‌شود که می‌تواند از انواع 1 تا 5 باشد.

با انتخاب Include intercept in Model مدل با وجود مقدار ثابت A برازش می‌شود و در صورتی

که آن را انتخاب ننمائیم مقدار ثابت از جدول خروجی Test of Between-Subjects Effects حذف خواهد شد.

کلید Contrasts پنجره‌ای مانند زیر را ظاهر می‌سازد:



می‌توان با فعال کردن این کلید مقایسات مختلفی را بر سطوح مختلف متغیر عامل گسسته اعمال کرد. در پنجرهٔ Factor متغیری گسسته نمایش داده می‌شوند و در بخش Change Contrast به تعویض نوع مقایسه بین سطوح اختصاص دارد.

انواع این مقایسه‌ها در کرکره Contrast موجود است. با انتخاب هر یک از آنها و Click بر روی دکمهٔ Change روش مقایسه بین سطوح برای هر متغیر گسسته تغییر می‌کند. برای مثال در شکل بالا برای متغیر جنسیت از حالت Deriation و برای نوع هپاتیت از حالت Simple استفاده شده است. با اجرای این فرمان خروجی‌های زیر برای مقایسه بین سطوح σ در پنجرهٔ خروجی اضافه می‌شوند:

Test Results

Dependent Variable: میزان قند خون بیماران مورد بررسی

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	251259.6	1	251259.608	89.110	.000
Error	67671.90	24	2819.663		

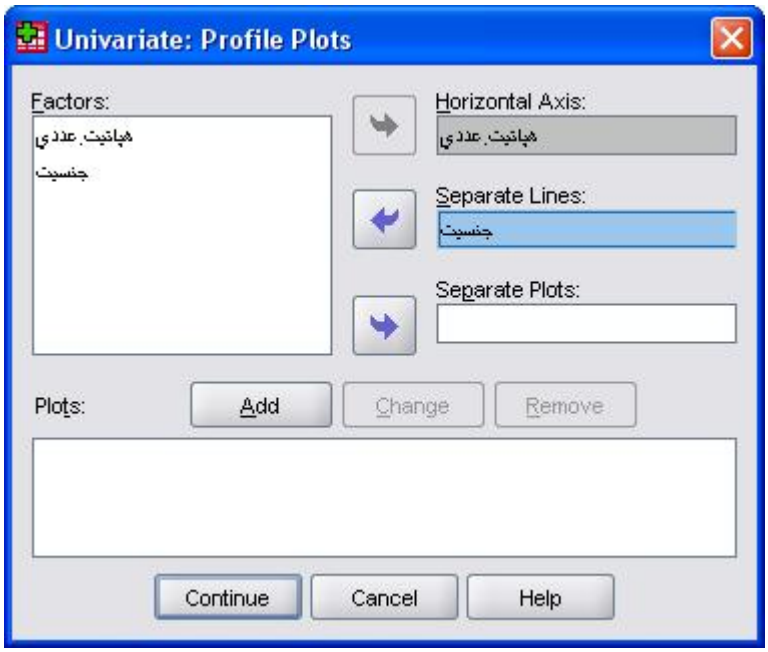
Test Results

Dependent Variable: میزان قند خون بیماران مورد بررسی

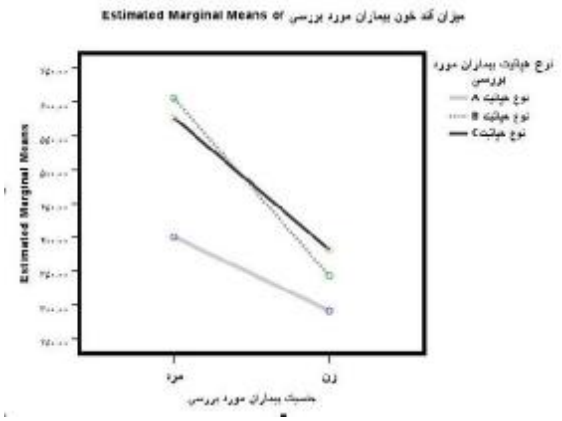
Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	106518.0	2	53259.020	18.888	.000
Error	67671.90	24	2819.663		

در جدول اول فرض برابری میانگین قند خون مختلف جنسیت و در جدول دوم فرض برابری میانگین قند خون سطوح مختلف نوع هپاتیت مورد آزمون قرار گرفته است که با توجه به مقدار sig=0/000 فرض اولیه رد می‌شود.

کلید ... Plots پنجره‌ای مانند زیر را ظاهر می‌سازد:

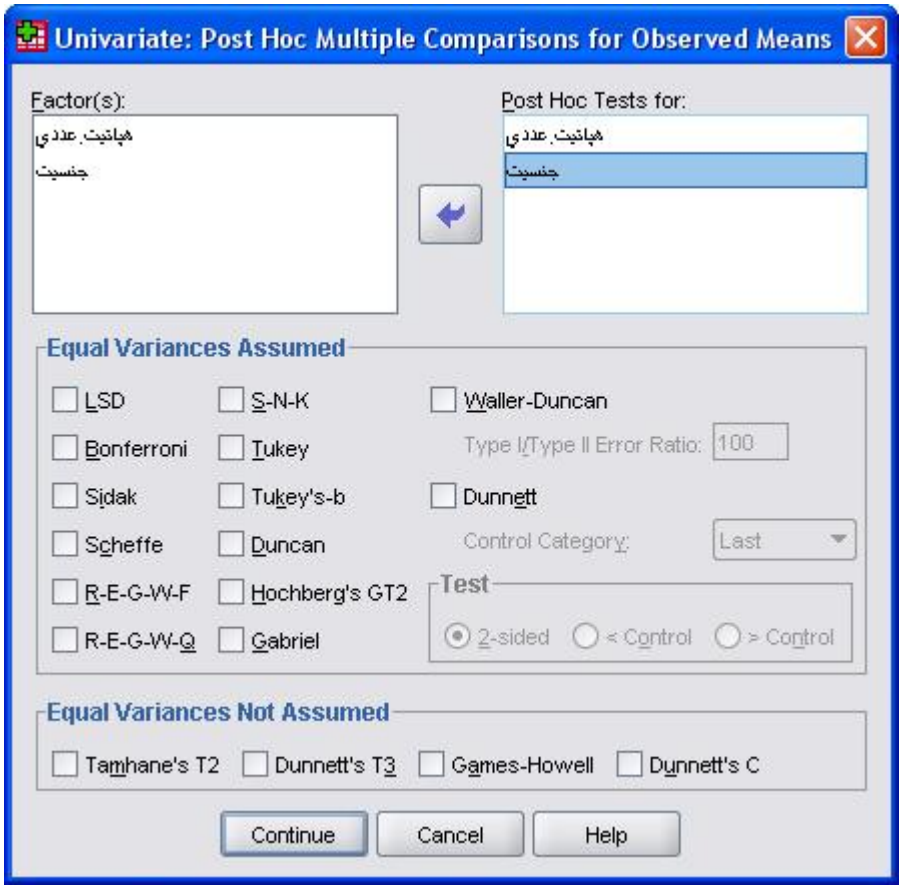


این پنجره امکان رسم نمودارهای اثرات اصلی و متقابل را مهیا می‌سازد. برای مثال مانند شکل بالا با انتخاب متغیر جنسیت به عنوان متغیر مبنای محور xها در Horizontal Axis و متغیر نوع هپاتیت به عنوان عامل دوم در separate Lines و فشردن دکمهٔ Add و اجرای فرمان خروجی زیر ظاهر می‌شود:



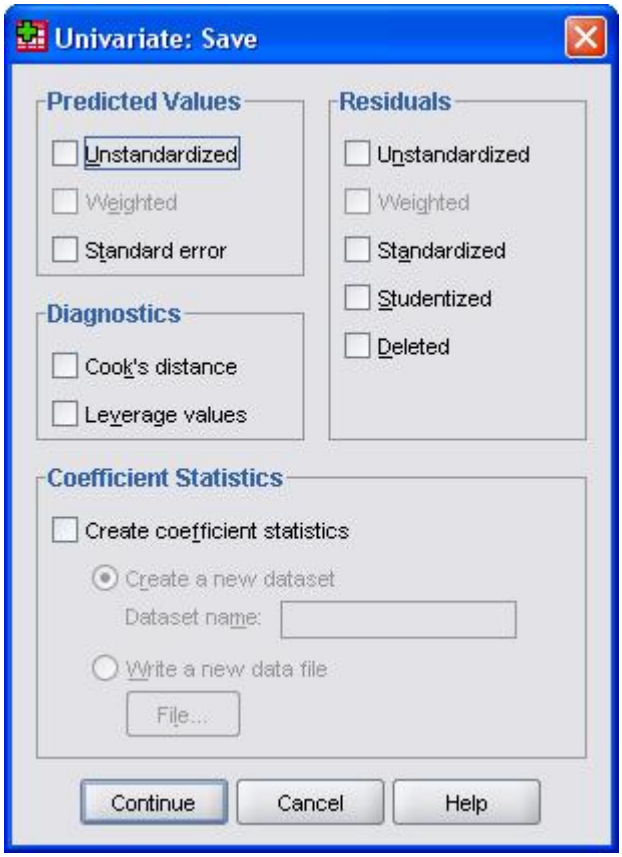
همانطور که در شکل بالا مشاهده می‌شود میانگین قند خون در هر دو سطح جنسیت نوع هپاتیت A از دیگر انواع هپاتیت پائین‌تر است اما در سطح B و C نوع هپاتیت اثر متقابل میان جنسیت و نوع هپاتیت دیده می‌شود.

کلید Post Hoc ... پنجره‌ای مانند زیر را ظاهر می‌سازد:



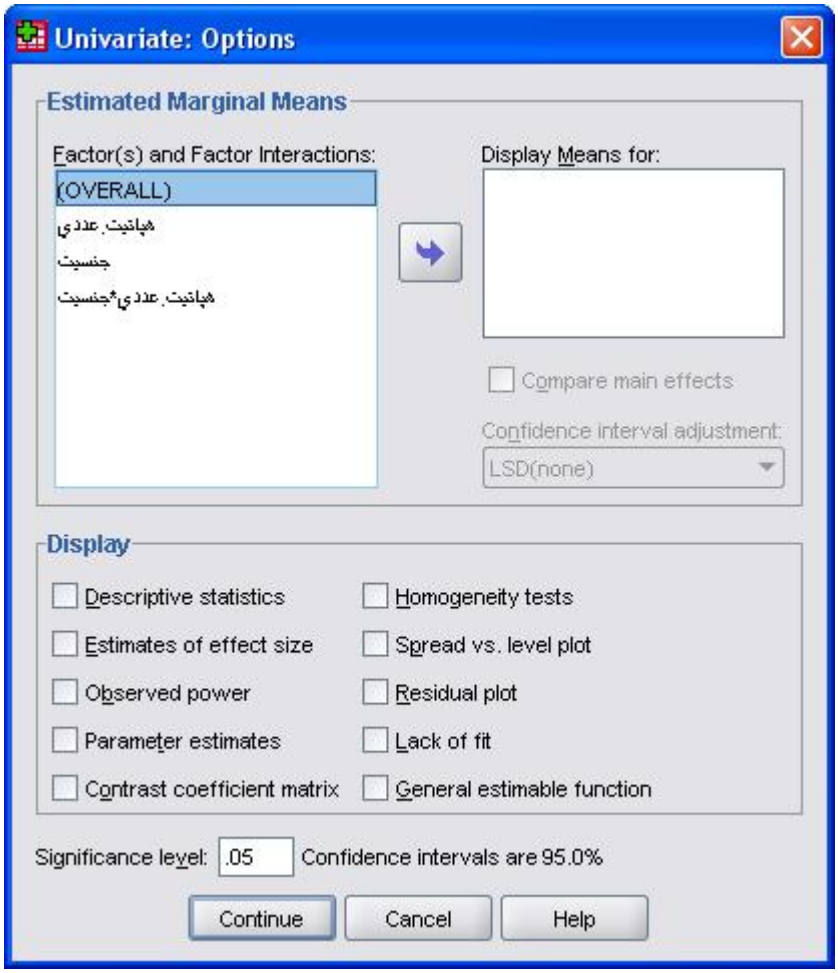
همانند حالت تحلیل واریانس یک طرفه ابزارهای این پنجره می‌توان سطوح مختلف متغیرهای گسسته را از نظر تأئیدگذاری بر روی قند خون مبتلایان رتبه‌بندی نمود.

کلید ... Save پنجره‌ای مانند زیر را باز می‌نماید:



کاربر با انتخاب هر یک از این گزینه‌ها می‌تواند مقادیر آنها را برای هر یک از داده‌ها در قالب ستونهای جدید مشاهده نماید. برای مثال با انتخاب گزینهٔ Standard Error مقادیر استاندارد شده خطاها برای هر یک از داده‌ها در مدل در پنجرهٔ داده‌ها با نام متغیر SEP-1 نمایش داده می‌شود.

کلید ... Options پنجره‌ای مانند شکل زیر را بازی می نماید:



در قسمت Estimated Marginal Means کاربر با انتخاب هر یک از عامل‌ها و اثرات متقابل بین آنها و انتقال آنها به قسمت Display Means For امکان محاسبه میانگین‌ها در هر یک از سطوح این عامل‌ها و اثرات متقابل آنها را در خروجی فراهم می‌آورد.

در قسمت Display انواع آماره‌های مفید مانند آماره‌های توصیفی، برآورد اثرات، برآورد پارامترها، نیکویی برازش، آزمون همگرایی میانگین‌های سطوح، ماتریس مقایسه ضرایب و شکل‌های باقیمانده و سطوح با انتخاب هر یک از گزینه‌ها در خروجی نمایش داده می‌شود.

در پنجره Significence Level کاربر می‌تواند میزان خطای نوع اول را از عدد اولیه 0/05 به هر مقدار مورد نظر در تحقیق تغییر دهد.

تمرین:

1- اطلاعات متغیر درآمد، درآمد در هنگام ورود به شرکت و همچنین سابقه کار در شرکت را به تفکیک جنسیت و اقلیت بودن بدست آورید.

2- سابقه افرادی را که در شرکت کار می نمایند در فایل employe data به یک متغیر گسسته با سه گروه زیر 20 ماه، بین 20 ماه و 60 ماه و بالاتر از 60 ماه تقیسم نمایئد و تاثی این متغیر گسسته تازه ایجاد شده را بر روی متغیر درآمد افراد دست آورید.

3- در فایل cars اثر منطقه تولید ماشین را بر روی وزن آن بدسیت آورید و بگوئید ماشین های تولیدی کدام منطقه به طور متوسط سنگین تر هستند.

4- اثر متغیر منطقه تولید ماشین و نعداد سیلندر را با هم بر روی قدرت ماشین های تولیدی بدست اورید.

5- آیا می توان ادعا کرد که ماشین های 8 سیلندری که در منطقه آمریکا نولید می شوند از دیگر ماشین های تولیدی در سراسر دنیا به طور متوسط سنگین تر هستند.

6- آیا این ادعا که افرادی که مدیر هستند و جزء گروه اقلیت نیستند در فایل employe data بیشترین حقوق را دریافت می کنند صحت دارد.

7- اثر تعداد سیلندر را بر روی میزان قدرت ماشین ها بدست آورید.

8- اثر رده های مختلف استخدام درشرکت را بر روی میزان درآمد بدست آورید.

9- رابطه میان جنسیت و رده های مختلف استخدامی را به صورت توام بر روی میزان درآمد به دست اورید. کدام یک از رده های شغلی آقایان و خانمها بیشترین و کدامیک کمترین حقوق را دریافت می کنند.

فصل دهم

اهداف آموزشی فصل دهم

فراگیر پس از مطالعه این فصل باید بتواند:

1-مفاهیم اولیه در تحلیل همبستگی را بازگو نماید.

2-مراحل انحام محاسبه میزان ضریب همبستگی میان دو متغیر رادر نرم افزار Spss.16با ذکر مثال شرح دهد.

3-مراحل انجام ضریب همبستگی جزئی متغیرها در نرم افزار Spss.16 را با ذکر مثال شرح دهد.

تحلیل همبستگی

10ـ1ـ مفاهیم اولیه

همبستگی یک معیار برای سنجش نوع و شدت رابطه میان دو یا گروهی از متغیرهاست. هنگامی که بیان می‌شود میان دو متغیر همبستگی وجود دارد به این معناست که تغییرات مقادیر متغیرها بر روی یکدیگر اثرگذار می‌باشد و هنگامی که از عدم همبستگی میان دو متغیر که به آن استقلال دو متغیر هم گفته می‌شود، صحبت می‌کنیم منظور این است که تغییرات مقادیر متغیرها بر روی یکدیگر اثر نمی‌گذارد.

ضریب همبستگی یک معیار عددی برای شناسایی شدت رابطه میان دو متغیر است. ضریب همبستگی که آن را با r نمایش می‌دهند عددی بین 1- و 1 می‌باشد و اگر $r = 0$ باشد نشان‌دهندهٔ استقلال میان دو متغیر X و Y است.

اگر $r > 0$ باشد نشان‌دهندهٔ وجود همبستگی مستقیم میان دو متغیر X و Y است به این معنی که با

زیاد شدن مقادیر یا کدهای متغیر X، مقادیر یا کدهای متغیر Y هم افزایش می‌یابند و هر چه ضریب همبستگی r به مقدار 1+ نزدیکتر باشد نشان‌دهندهٔ شدت همبستگی مستقیم بین دو متغیر X و Y می‌باشد.

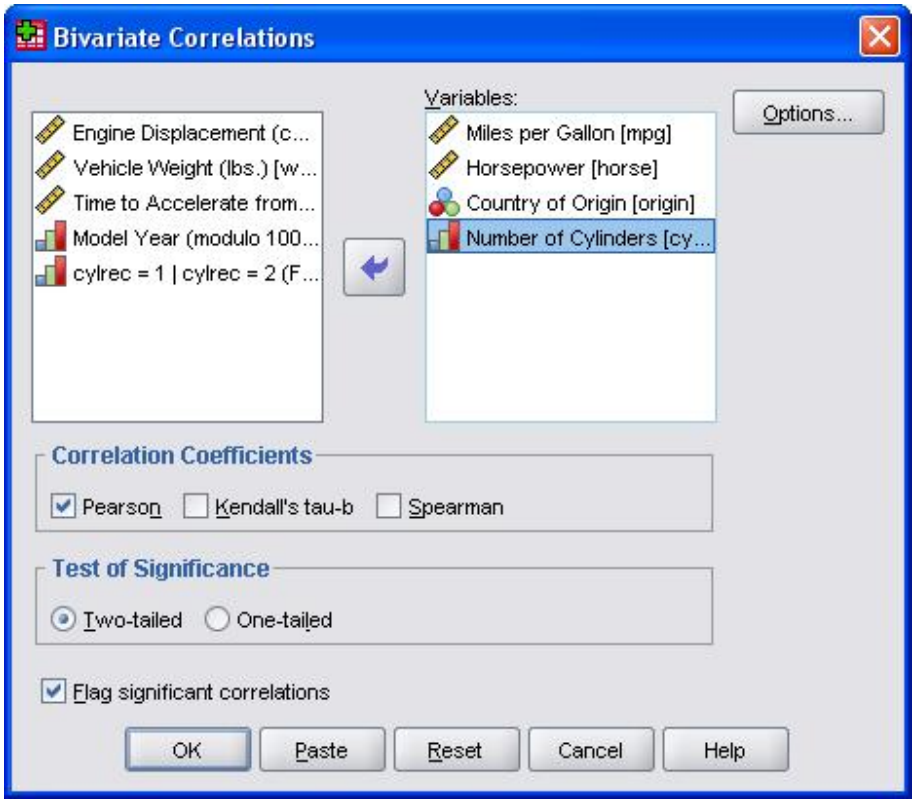
اگر $r < 0$ باشد نشان‌دهندهٔ وجود همبستگی معکوس میان دو متغیر X و Y است، به این معنی که با زیاد شدن مقادیر یا کدهای متغیر X، مقادیر یا کدهای متغیر Y کاهش می‌یابد و هر چه ضریب همبستگی r به مقدار 1- نزدیکتر باشد نشان‌دهندهٔ شدت همبستگی معکوس میان دو متغیرX و Y می‌باشد.

در این فصل دو نوع متغیر همبستگی دودویی و ضریب همبستگی جزئی در بخش‌های بعد مورد بررسی قرار گرفته‌اند.

10ـ2ـ ضریب همبستگی بین دو متغیر

همانطور که در بخش قبل عنوان گردید محاسبه میزان ضریب همبستگی میان دو متغیر می‌تواند از اهمیت ویژه‌ای برخوردار باشد. برای مثال قبل از اینکه در مدلهای مختلف رابطهٔ میان متغیر مستقل X بر روی متغیر وابسته Y سنجیده شود با استفاده از نتایج ضریب همبستگی میان این دو متغیر می‌توان به نتایج اولیه در مورد نوع رابطهٔ میان متغیرها دست یافت.

برای محاسبهٔ ضریب همبستگی میان دو متغیر از فرمان ... Bivariate در زیرمنو Correlate از منوی Analyze را اجرا می‌نمائیم تا پنجرهٔ زیر ظاهر شود:



در قسمت سمت چپ این پنجره لیستی از تمامی متغیرهای فایل جاری Cars موجود است. متغیرهایی را که می‌خواهیم ضریب همبستگی بین آنها را به صورت دو به دو محاسبه نمائیم به پنجرهٔ سمت راست Variables وارد می‌نمائیم.

در بخش Correlation Coefficients نوع ضریب همبستگی که محاسبه می‌شود، نشان داده می شود.

انتخاب پیش‌فرض، ضریب همبستگی نمونه‌ای (پیرسون) است. این انتخاب زمانی معتبر خواهد بود که داده‌ها کمی و توزیع آنها به طور تقریبی نرمال باشد تا نتایج ضریب همبستگی پیرسون بیان‌کننده نوع و شدت رابطهٔ میان دو متغیر باشد.

گزینهٔ Kendall's Tau-b برای محاسبه ضریب همبستگی میان متغیرهای ترتیبی کاربرد وسیع‌تر دارد، مقادیر ممکن این نوع ضریب همبستگی نیز اعدادی بین 1+ و 1- خواهد بود.

ضریب همبستگی Spearman به محاسبهٔ ضریب همبستگی ناپارامتری میان دو متغیر می‌پردازد که این متغیرها می‌توانند از هر نوعی از داده‌ها شامل داده‌های پیوسته، ترتیبی یا اسمی باشد. در حقیقت ضریب همبستگی Spearman هنگامی که متغیرها دارای تعداد کمی داده هستند مورد استفاده قرار می‌گیرد و هنگامی که تعداد داده‌های درون متغیرها به اندازه کافی زیاد باشد می‌توان با فرض قانون حد مرکزی مبنی بر وجود توزیع نرمال برای میانگین داده‌های یک متغیر از ضریب همبستگی پیرسون برای هر نوعی از متغیرها بهره برد.

با توجه به اینکه در مثال Cars تعداد داده‌ها به اندازه کافی بزرگ است، برای محاسبهٔ ضریب همبستگی هر نوعی از متغیرها به صورت دو به دو، آنها را وارد پنجرهٔ Variables می‌نمائیم.

در بخش Test Of Significance نوع آزمون ضریب همبستگی تعیین می‌شود.

گزینهٔ Two-Tailed آزمون فرض مساوی بودن ضریب همبستگی (از هر نوع) را انجام می‌دهد یعنی آزمون

$H_0 : r = 0$

$H_1 : r \neq 0$

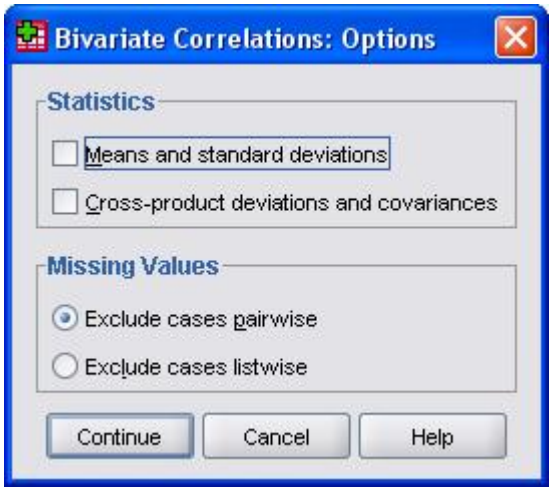
اجرا می‌شود. گزینه One-Tailed آزمون فرض

$H_0 : r > 0$

$H_1 : r \leq 0$

را آزمون می‌نماید.

انتخاب Flag Significance Correlations باعث می‌شود نرم‌افزار فرض معنی‌داری همبستگی میان دو متغیر را در سطح α=0/05 و α=0/01 آزمون نماید. با فعال کردن کلید ... Options نوع آماره‌های مربوط به زوج متغیرها و نحوه عمل مقادیر گمشده تعیین می‌شود. جعبهٔ گفتگوی این کلید در شکل زیر مشاهده می‌شود:



در بخش Statistics، گزینهٔ Means and standard Deviations برای نمایش میانگین‌ها و انحراف معیارهای متغیرهای انتخابی در پنجرهٔ Variables منظور شده است. با انتخاب گزینهٔ Cross-Product Derivations and Covariances حاصلضربهای متقاطع و کواریانس زوج متغیرها به شکل ماتریس در پنجره خروجی نشان داده می‌شود. البته ذکر این نکته ضروری است که این خروجی‌ها تنها با انتخاب ضریب همبستگی از نوع Pearson فعال می‌شوند.

بخش Missing Values نحوه برخورد با مشاهدات گمشده را نشان می‌دهد. گزینهٔ Exclude Case Pair wise برای حذف جفت مشاهدات، در صورت برخورد با مقدار گمشده و گزینهٔ دیگر این بخش برای حذف فهرست گونه مشاهدات در نظر گرفته شده است. در حذف زوج گونه مشاهده‌ای که مقدار گمشده‌ای دارد، از دو متغیر مانند X و Y حذف می‌شود اما در حذف فهرست گونه مشاهده‌ای که در یک متغیر مقدار گمشده دارد از کلیه متغیرهای موجود در تحلیل همبستگی حذف می‌شود.

10ـ2ـ1ـ نمونه‌ای از خروجی فرمان Bivariate

برای نمونه متغیرهای Engine، Hourse، Origin و Sylinder را از فایل Cars انتخاب نموده و وارد پنجرهٔ Variables می‌نمائیم. از قسمت Correlation Coefficients نوع Pearson را با

توجه به تعداد زیاد داده‌ها انتخاب نمـوده و از قسـمت Test of Significance حالـت Two-Tailed را انتخاب می‌کنیم.

قسمت Flag Significant Correlations را هم فعال نموده و فرمان را اجرا می‌نمائیم. خروجی زیـر ظاهر می‌شود:

Correlations

		Engine Displacement (cu. inches)	Horsepower	Country of Origin	Number of Cylinders
Engine Displacement (cu. inches)	Pearson Correlation	1	.895**	.607**	.951**
	Sig. (2-tailed)	.	.000	.000	.000
	N	406	400	406	406
Horsepower	Pearson Correlation	.895**	1	.456**	.842**
	Sig. (2-tailed)	.000	.	.000	.000
	N	400	400	398	398
Country of Origin	Pearson Correlation	.607**	.456**	1	.626**
	Sig. (2-tailed)	.000	.000	.	.000
	N	406	398	406	406
Number of Cylinders	Pearson Correlation	.951**	.842**	.626**	1
	Sig. (2-tailed)	.000	.000	.000	.
	N	406	398	406	406

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

همانطور که ملاحظه می‌شود در یک ماتریس با نام Correlations مقادیر ضریب همبستگی پیرسون، مقدار سطح معنی‌داری دو طرفه بـرای آزمـون وجـود همبسـتگی یـا اسـتقلال میـان دو متغیـر و مقـدار مشاهدات (N) بیان شده است.

همانطور که ملاحظه می‌شود مقدار ضریب همبستگی برای وجود رابطه هر متغیر بـا خـود آن عـدد 1 می‌باشد که حداکثر همبستگی میان دو متغیر است و تنها زمانی اتفاق می‌افتد که دو متغیر دقیقـاً ماننـد هم باشند.

در سطر دوم در هر خانه از ماتریس مقدار sig یـا سـطح معنـی‌داری آزمـون $H_0 : rxy = 0$ در مقابـل $H_1 : rxy \neq 0$ بیان شده است.

همانطور که ملاحظه می‌شود بین همهٔ متغیرها فرض استقلال به دلیل وجود sig=0/00 در هر سطحی از α رد می‌شود. فرض H_0 زمانی رد می‌شود که $sig \leq \alpha$ باشد و چون sig=0/000 است از هر سطحی از

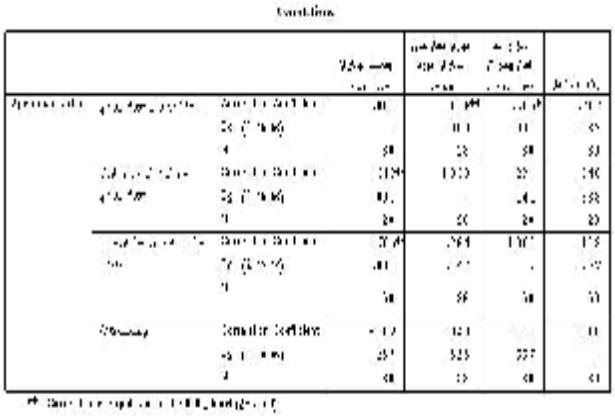
α کوچکتر است.

خانه‌هایی از ماتریس که با ** مشخص شده‌اند با توجه به توضیح ** در پائین جدول رابطهٔ میان متغیرهای X و Y آنها در سطح $\alpha = 0/01$ بیان شده است همانطور که ملاحظه می‌شود ضریب همبستگی میان Engine و Hourse برابر 0/897 است و در سطح $\alpha=0/01$ معنی‌دار است، این به معنی وجود رابطه میان دو متغیر است و چون مقدار ضریب همبستگی عددی مثبت است با افزایش هر یک از این متغیرها (مثلاً Engine یا ظرفیت موتور) مقدار متغیر دیگر (مثلاً Hourse یا قدرت ماشین) نیز افزایش می‌یابد.

در خانهٔ رابطهٔ میان Origine و Hourse همانطور که ملاحظه می‌شود مقدار ضریب همبستگی پیرسون برابر 0/459- است و سطح معنی‌داری نیز برابر عدد صفر می‌باشد که نشان از وجود همبستگی میان دو متغیر است. با وجود ضریب همبستگی پیرسون با مقدار منفی و کدهای Origin [1= آمریکا، 2=اروپا، 3=ژاپن] می‌توان استنباط نمود که با افزایش کدهای متغیر Origin، مقدار Hourse (قدرت اسب بخار) کاهش می‌یابد یا در اصطلاح ماشین‌های آمریکایی از قدرت اسب بخار بالاتر از دیگر ماشین‌ها و ماشین‌های ساخت کشور ژاپن دارای کمترین قدرت اسب بخار هستند.

حال اگر تعداد داده‌های تحقیق به اندازهٔ کافی زیاد نباشد، استفاده از ضریب همبستگی Pearson مناسب نیست و توصیه می‌شود از ضریب همبستگی ناپارامتری Spearman استفاده گردد.

برای مثال در داده‌های مربوط به بیماران مبتلا به هپاتیت با توجه به تعداد کم داده‌ها (30 عدد) توصیه می‌شود که برای محاسبه ضرایب همبستگی میان متغیرها از روش Spearman استفاده گردد. خروجی زیر آزمون و برد همبستگی میان متغیرهای جنسیت، میزان فشار خون، میزان قند خون و تأهل در فایل Exam نمایش می‌دهد:



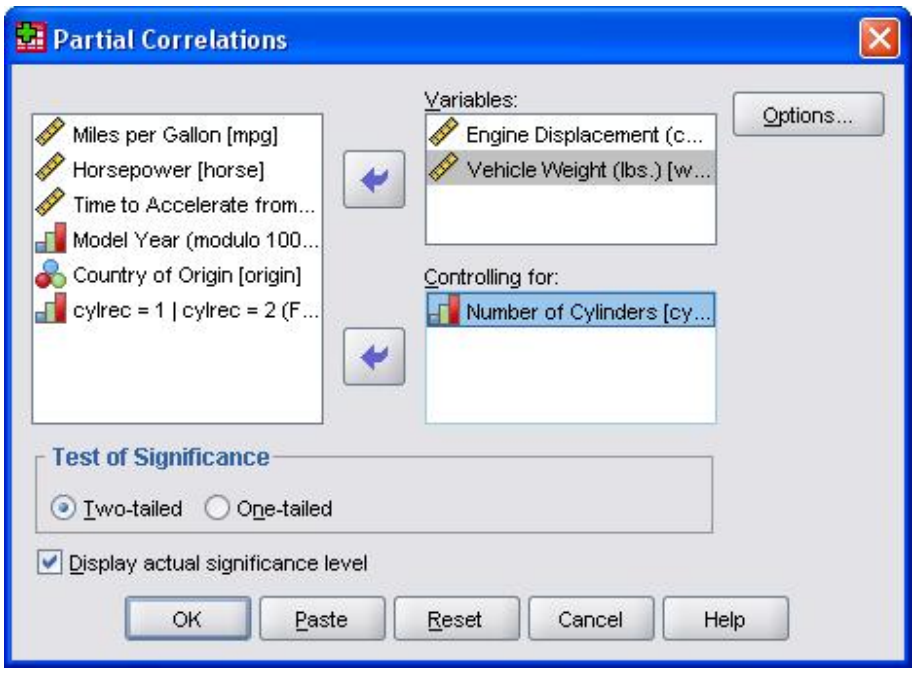
همانطور که ملاحظه می‌شود به دلیل تعداد وجود کم داده‌ها از روش ضریب همبستگی Spearman استفاده شده است و تنها می‌توان گفت که میان جنسیت و میزان فشار خون و قند خون بیماران رابطه مستقیم وجود دارد و با توجه به اینکه در متغیر جنسیت آقایان با کد 1 و خانم‌ها با کد 2 مشخص شده‌اند، با افزایش کد جنسیت میزان قند خون و فشار خون مبتلایان نیز افزایش می‌یابد یا در اصطلاح قند خون و فشار خون خانم‌ها بیش از قند خون و فشار خون آقایان است. نکته جالب این آزمون بدست آوردن نتایج تأثیر یک متغیر گسسته دو سطحی مانند جنسیت بر روی متغیر پیوسته است که مانند نتایج آزمون مقایسه میانگین دو نمونه مستقل (Independent Sample T.test) می‌باشد.

10ـ3ـ همبستگی جزئی متغیرها (partial Correlations)

هر گاه بخواهیم ضریب همبستگی میان دو متغیر را با شرط ثابت بودن سایر متغیرها محاسبه نمائیم، ضریب همبستگی حاصل را ضریب همبستگی جزئی می‌نامیم. در حقیقت در این روش از محاسبه ضریب همبستگی در صورتی که متغیر خاصی بر روی هر دو متغیر تأثیرگذار است، اثر آن بر روی دو متغیر ثابت فرض شده و میزان همبستگی واقعی میان دو متغیر محاسبه می‌شود.

برای مثال فرض کنید می‌خواهیم ضریب همبستگی میان متغیرهای وزن ماشین (Weight) و ظرفیت موتور (Engine) را بدست آوریم، با توجه به اینکه هر دو متغیر بالا تحت تأثیر تعداد سیلندر

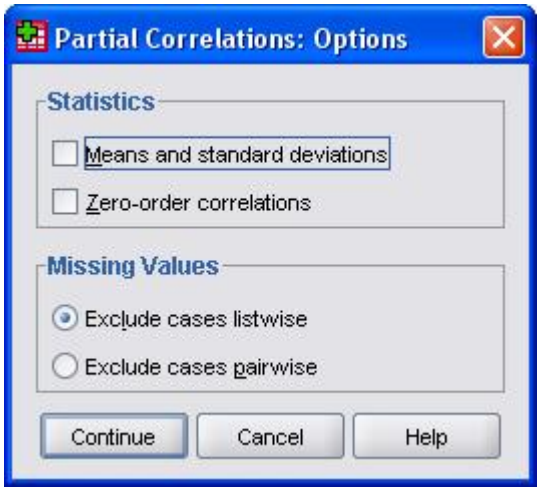
(Cylinder) تغییر می‌نمایند، بهتر است اثر متغیر تعداد سیلندر با ثابت در نظر گرفتن آن حذف شود تا ضریب همبستگی جزئی بین دو متغیر وزن ماشین و ظرفیت موتور آن مقدار واقعی‌تر از ضریب همبستگی را نشان دهد. برای اجرای این فرمان گزینهٔ Partial... را از زیر منو Correlate و منوی Analyze انتخاب می‌نمائیم تا پنجره زیر باز شود:



همانند قبل در قسمت سمت چپ پنجره لیستی از متغیرهای فایل جاری نمایش داده شده است. متغیرهایی که محاسبه ضریب همبستگی جزئی آنها مورد نظر است به جعبهٔ Variables و متغیر یا متغیرهایی که در این محاسبه ثابت فرض می‌شوند به جعبهٔ Controlling for منتقل می‌نمائیم.

بخش Test of Significance مانند بخش قبل عمل می‌نماید و با عدم انتخاب Display Actual Significance Level مقایر سطح معنی‌داری sig و تعداد مشاهدات در خروجی ظاهر نخواهند شد.

با فعال کردن کلید ... Options جعبهٔ گفتگوی زیر را ظاهر می‌شود:



جزئیات جعبهٔ گفتگو مشابه فرمان Bivariate است با این تفاوت که گزینهٔ Zero-Order Correlations در بخش Statistics جایگزین گزینهٔ مشابه شده است. با انتخاب این گزینه، ضرایب همبستگی کلیه متغیرها بدون اعمال متغیرهای کنترلی (ثابت) محاسبه و آزمون می‌شود.

10ـ3ـ1ـ نمونه‌ای از خروجی فرمان Partial

برای نمونه، متغیرهای Engine، Hourse و Weight را وارد پنجرهٔ Variables نموده و مقدار همبستگی جزئی آنها با ثابت در نظر گرفتن اثرات متغیرهای Origin و Sylinder و وارد کردن آنها در قسمت Controlling for و تأئید فرمان در خروجی به صورت زیر ظاهر می‌شود:

Correlations

Control Variables			weight	horse	engine
cylinder & origin	weight	Correlation	1.000	.480	.571
		Significance (2-tailed)	.	.000	.000
		df	0	394	394
	horse	Correlation	.480	1.000	.618
		Significance (2-tailed)	.000	.	.000
		df	394	0	394
	engine	Correlation	.571	.618	1.000
		Significance (2-tailed)	.000	.000	.
		df	394	394	0

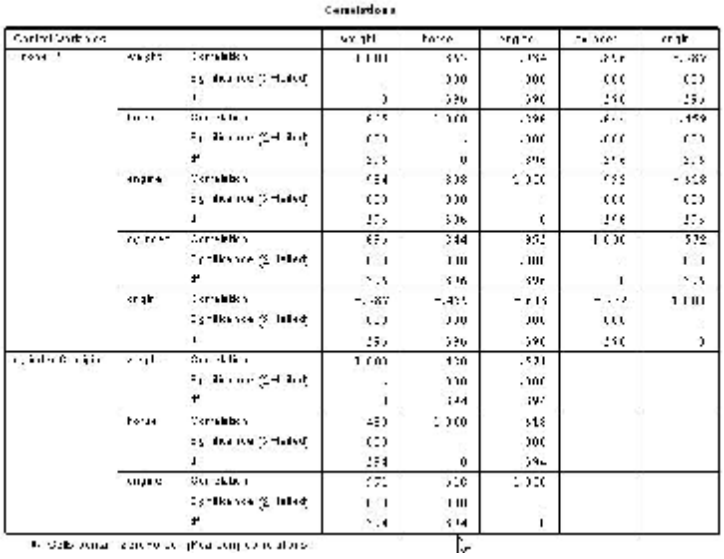
همانطور که ملاحظه می‌شود در ماتریس Correlations، آزمون همبستگی میان سه متغیر Engine

(ظرفیت موتور)، Hourse (قدرت موتور) و Weight (وزن ماشین) به صورت دو بـه دو و بـا فـرض ثابـت بودن اثرات متغیرهای Origin (محل ساخت) و Sylinder (تعداد سیلندر) به صورت بالا است. با توجه به مقدار ضریب همبستگی 0/63 و سطح معنی‌داری sig=0/000 بین دو متغیر Engine و Houre و ضریب هبستگی r=0/568 و sig=0/000 برای Engine و Weight به این نتیجـه مـی‌رسـیم کـه میـان متغیـر Engine و هر یک از متغیرهای Hourse و Weight رابطه معنی‌دار و در جهت مستقیم برقرار است و با افزایش ظرفیت موتور به صورت معنی‌داری قدرت و وزن ماشین افزایش خواهد یافت.

همینطور میان متغیر Hourse و Weight مقدار ضریب همبستگی برابر 0/48 و sig=0/000 است که باز هم نشان‌دهندهٔ رابطهٔ معنی‌دار و مستقیم است.

البته این نتایج با در نظر گرفتن متغیرهای Origin و Sylinder به متغیرهای با اثـرات ثابـت بدسـت آمده است و در صورتی که اثرات این دو متغیر ثابت در نظر گرفته نشود مقدار ضریب همبستگی با مقدار ضریب همبستگی جزئی بدست آمده در این بخش به طور حتم متفاوت خواهد بود.

بـا انتخـاب گزینـهٔ Zero-Order Correlation از قسـمت Statistics در کلیـد ... Options در خروجی یکی با در نظر گرفتن متغیرهای ثابت و دیگری بدون در نظر گرفتن متغیرهای ثابت ارائه می‌شود که در حقیقت کاربر با مقایسه مقادیر ضریب همبستگی در این دو حالت مـی‌توانـد تفـاوت مقـادیر روش Bivariate (بدون در نظر گرفتن متغیر ثابت) را با مقادیر روش Partial مشاهده کنـد. خروجـی اجـرای این فرمان به صورت زیر است:



همانطور که ملاحظه می‌شود متغیر کنترلی در قسمت اول خروجی در نظر گرفته نشده و در قسمت دوم نتایج با در نظر گرفتن متغیرهای کنترلی تعداد سیلندر و کشور سازنده خودورها بیان شده است.

تمرین:

1- تفاوت همبستگی دودویی و همبستگی جزئی را بیان نمائید.

2- به نظر مشان کاربرد کدام نوع از همبستگس بیشتر است و چرا؟

3- همبستگی همه متغیرهای فایل cars را با با استفاده از همبستگی دو دویی با متغیر وزن خودرو بدست آورید و به ترتیب میزان همبستگی عوامل آنها را بیان نمائید.

4- همبستگی همه متغیرهای فایل بیماران مبتلا به هپاتیت را قند خون بیماران بدست آورید و بر اساس شدت همبستگی انها را رتبه بندی نمائید.

5- به نظر شما رتبه بندی عوامل موثر بر روی حقوق جاری افراد در فایل employe data به چه صورت می باشد.

6- با توجه به روش همبستگی جزئی عوامل موثر بر حقوق افراد را در شرکت بر اساس میـزان خـالص شدت آنها رتبه بندی نمائید.

7- با توجه به روش همبستگی جزئی عوامل موثر بر فشار خـون افـراد را در فایـل بیمـاران مبـتلا بـه هپاتیت بر اساس میزان خالص شدت آنها رتبه بندی نمائید.

8- با توجه به روش همبستگی جزئی عوامل موثر بر وزن خودرو را در فایـل cars بـر اسـاس میـزان خالص شدت آنها رتبه بندی نمائید.

فصل یازدهم

رگرسیون خطی

اهداف آموزشی فصل دوازدهم

فراگیر پس از مطالعه این فصل باید بتواند:

- مفاهیم اولیه در رگرسیون خطی را بیان کند.
- مراحل انجام رگرسیون خطی در نرم افزار Spss.16 را ب ذکر مقال شرح دهید.
- نحوه ورود متغیرهای مستقل به مدل رگرسیونی در نرم افزار Spss.16 را انجام دهد.
- مراحل انجام محاسبه آمارها در رگرسیونی را با ذکر مثال توضیح دهد.
- دلایل رسم نمودارها در رگرسیون خطی را بیان کند.
- مراحل رسم نمودار رگرسیون خطی در نرم افزار Spss را با ذکر مثال شرح دهد.
- بررسی فرض خطی بودن مدل رگرسیونی برازش شده را شرح دهد.
- مراحل ذخیره اطلاعات فرعی رگرسیون خطی در نرم افزار Spss را با ذکر مثال انجام دهد.

11ـ1ـ مبانی اولیه:

همانطور که در فصل اول بیان شد، یکی از اهداف استفاده از تکنیک‌های آماری بدست آوردن رابطهٔ میان دو متغیر یا بین گروهی از متغیرها می‌باشد، رگرسیون هم مانند تمامی ابزارهای آمار استنباطی وظیفهٔ بدست آوردن رابطهٔ میان متغیرها را بر عهده دارد. در رگرسیون ساده تنها 2 متغیر پیوسته وجود دارد و هدف

بدست آوردن تأثیر یک متغیر پیوسته بر روی متغیر پیوسته دیگر است. برای مثال در مثال ماشین‌های انتخاب شده که اطلاعات آن در فایل Cars موجود می‌باشد، برای بدست آوردن تأثیر متغیر Weight که وزن خودروها می‌باشد بر روی mpq که میزان طی مسافت‌ها در ازای یک گالن سوخت است باید از رگرسیون استفاده نمود زیرا هر دو متغیر مستقل تأثیرگذار و هم متغیر وابسته تأثیرپذیر از نوع متغیرهای پیوسته می‌باشند.

فرض کنید k متغیر مستقل $X_1, X_2, ..., X_k$ وجود دارند و می‌خواهیم تأثیر آن را بر روی یک متغیر پیوسته وابسته که در رگرسیون آن را با نماد Y نشان می‌دهند بدست آوریم، رابطه میان متغیرهای مستقل $X_1, ..., X_k$ و متغیر پیوسته Y می‌تواند به صورت زیر باشد:

$$Y = f(X_1, ...X_k) + e$$

که در آن f تابعی احتمالی بر حسب متغیرهای $X_1, ..., X_k$ می‌باشد و e مقدار تفاوت میان $f(X_1, ...X_k)$ و مقدار Y می‌باشد، e متغیری تصادفی است که دارای توزیع آماری نرمال با میانگین صفر می‌باشد. اگر تابع احتمالی f یک تابع خطی از متغیرهای $X_1, X_2, ..., X_k$ باشد مدل $Y = f(X_1, ...X_k) + e$ را رگرسیون خطی و اگر f یک تابع خطی از متغیرهای $X_1, X_2, ..., X_k$ نباشد مدل $Y = f(X_1, ...X_k)$ را غیرخطی می‌نامند.

برای مثال رابطهٔ میان میزان کوددهی (بر حسب کیلوگرم) و میزان آب‌دهی (بر حسب لیتر) در یک کرت زمین کشاورزی با میزان برداشت محصول به صورت:

$$Y = 427 + 0/827\ X_1 + 0/324\ X_2$$

که در آن X_1 میزان کود، X_2 میزان آب دهی و Y میزان برداشت محصول است. چون رابطه میان Y با تمامی متغیرهای مستقل خطی است به این مدل رگرسیون خطی می‌گویند، حال اگر X_1 طول میزان گالری دریافتی کودکان در سال، X_2 افزایش قد کودکان در سال باشد و Y میزان افزایش وزن کودکان در سال بر حسب گرم باشد آنگاه

$$Y = 243 + 4LnX_1 + 12e^{-X_2}$$

آنگاه چون میان X_1 و X_2 و Y رابطه غیرخطی برقرار است، برای بدست آوردن تأثیر X_1 و X_2 از رگرسیون غیرخطی استفاده می‌نمائیم. ذکر این نکته ضروری است که در بسیاری از موارد می‌توان با استفاده از تبدیلات آماری رگرسیون غیرخطی را به رگرسیون خطی تبدیل نمود.

11ـ2ـ رگرسیون خطی:

در رگرسیون خطی همانطور که عنوان شد رابطهٔ میان مجموعهٔ متغیرهای مستقل و متغیر وابسته y به صورت خطی است برای مثال:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + e_i \quad i = 1,..., N$$

که در آن $\beta_0, \beta_1, ..., \beta_k$ پارامترهای مجهول، e متغیر تصادفی و $X_1, ..., X_k$ متغیرهای مستقل هستند.

برای انجام یک رگرسیون خطی رعایت چند پیش‌فرض اولیه لازم و ضروری است که تعدادی از آنها عبارتند از:

$E(e_i) = 0$

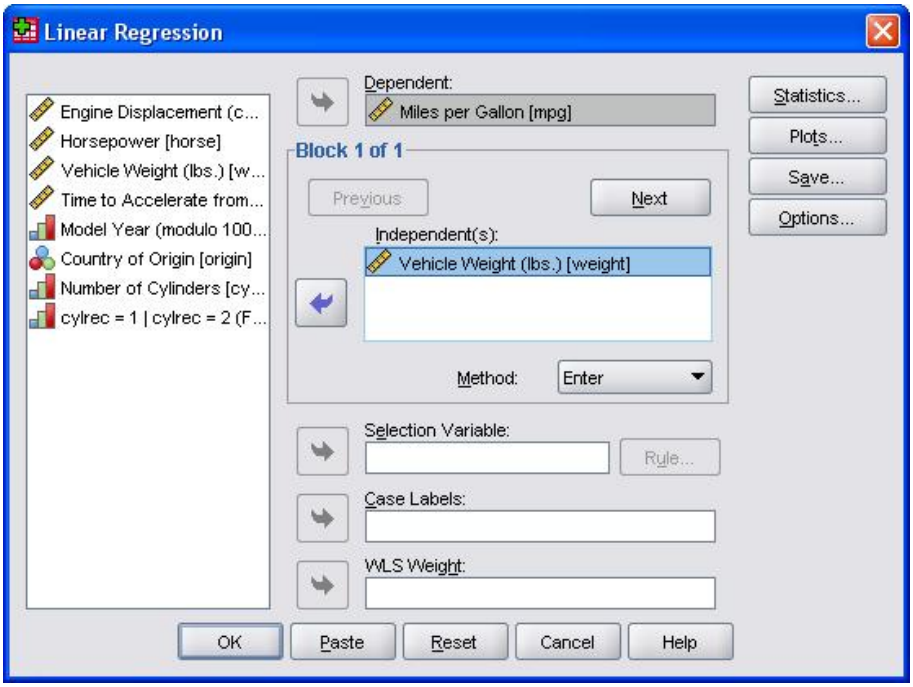
$Var(e_i) = \sigma^2$ و σ^2 مقداری ثابت.

$Cov(e_i, e_j) = 0$

$Cov(X_i, X_j) = 0$

منظور از $Cov(X_i, X_j) = 0$ عدم وابستگی میان متغیرهای مستقل است، اگر دو متغیر مستقل به هم وابسته هستند، لزومی ندارد که هر دو متغیر وارد مدل شوند، مدلی را وارد مدل می‌نمائیم که دارای همبستگی بیشتر با متغیر وابسته Y می‌باشد.

برای انجام یک رگرسیون خطی در نرم‌افزار SPSS، فرمان Linear را از زیر منو Regression و منو Analyze انتخاب نموده و اجرای نمائیم تا پنجرهٔ زیر باز شود:



همانطور که مشاهده می‌شود در قسمت سمت چپ پنجره لیستی از تمامی متغیرهای موجود در فایل جاری مانند پنجره‌های قبل ظاهر شده است.

در پنجرهٔ Dependent کاربر متغیر وابسته را وارد و در پنجرهٔ Independent (s) مجموعهٔ متغیرهای مستقل را وارد می‌نمایید، برای مثال برای انجام رگرسیون خطی و بدست آوردن تأثیر وزن ماشین (Weight) بر روی طی مسافت در ازای یک گالن سوخت (mpq)، متغیر وزن ماشین را در پنجرهٔ Independent(s) و متغیر طی مسافت در ازای یک گالن سوخت را در پنجرهٔ Dependent وارد می‌نمائیم. در مثال بالا اگر قبل از بیان اطلاعاتی در مورد دیگر محتویات پنجره بخواهیم یک خروجی ساده از تأثیر متغیر Weight بر روی mpq مشاهده کنیم، دکمهٔ OK را می‌زنیم تا خروجی زیر ظاهر شود:

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Vehicle Weight (lbs.)	.	Enter

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.807[a]	.651	.650	4.622

a. Predictors: (Constant), Vehicle Weight (lbs.)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15794.63	1	15794.632	739.503	.000[a]
	Residual	8457.943	396	21.358		
	Total	24252.58	397			

a. Predictors: (Constant), Vehicle Weight (lbs.)

b. Dependent Variable: Miles per Gallon

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	46.492	.794		58.537	.000
	Vehicle Weight (lbs.)	-.007	.000	-.807	-27.196	.000

a. Dependent Variable: Miles per Gallon

در پنجرهٔ اول با نام Variable Entered/Remore لیستی از متغیرهای مستقل که وارد شده‌انـد را بیـان می‌کند، چون تنها یک متغیر Weight را وارد نموده‌ایم، نام آن را بیان کرده است.

در جدول دوم با نام Model Summary ، معیاری را بیان کرده است که از طریـق آن مـی‌تـوان قضـاوت نمود که آیا تعداد متغیرهای مستقل موجود در مدل مناسب است یا خیر؟ این معیار به ضریب تعیـین (R^2) یا Rsquare معروف است و عددی است که بین 0 و عدد 1 و هر چه به عدد یک نزدیکتر باشد نشان دهنـدهٔ مناسب بودن تعداد و نوع متغیرهای انتخابی مستقل برای توجیه تغییرات متغیـر وابسـته در مـدل رگرسـیون است.

در بسیاری از شاخه‌های کاربردی علم آمار مقدار بالاتر از 0/7 را برای ضریب تعیین به عنوان مناسب بودن نوع انتخاب و تعداد متغیرهای مستقل تأثیرگذار در مدل در نظر می‌گیرند. در این مثال مقدار ضـریب تعیـین برابر حدود 0/65 است که نشان‌دهندهٔ ناکافی بودن تعداد متغیرهای مستقل وارد شده در مدل است.

جدول سوم با عنوان ANOVA، بیان‌کننده مناسب بودن یا نبودن رگرسـیون خطـی بـرای بـرازش مـدل است. فروض این آزمون به صورت زیر است:

H_0: مدل خطی مدل مناسبی در برازش مدل رگرسیون نیست.

H_1: مدل رگرسیون خطی در برازش مدل مناسب است.

نتایج آزمون فوق از طریق آمارهٔ F و سطح معنی‌داری (sig) قابل تحلیل است اگر $sig>\alpha$ باشد به معنی رد فرض اولیه و به معنی نامناسب بودن روش رگرسیون خطی برای برازش مدل می‌باشد و چنانکه $sig \leq \alpha$ باشد، نشان‌دهندهٔ مناسب بودن رگرسیون خطی در برازش مدل رگرسیونی و بدست آوردن تأثیر متغیرهای مستقل بر روی متغیر وابسته است.

در جدول فوق با در نظر گرفتن $\alpha=0/05$ مشاهده می‌شود که sig=0/000 خواهد بود که $sig \leq \alpha$ خواهد بود و نشان‌دهندهٔ مناسب بودن برازش رگرسیون خطی است.

جدول Coefficients حاوی اطلاعات ضرایب مجهول مدل می‌باشد اگر مدل خطی به صورت :

$$Y = \beta_0 + \beta_1 X_1$$

باشد که در آن X_1 متغیر Weight و Y متغیر mpq است، مقدار β_0 که از آن به عنوان عرض از مبدأ یا مقدار ثابت نام می‌برند، مقدار متوسطی از mpq در هر خودرو است که هر ماشین به طور متوسط دارای آن است و ربطی به تغییرات وزن ماشین‌ها ندارد و ضریب β_1 مقدار تغییر در واحد mpq در ازای یک واحد تغییر Weight است. همانطور که در جدول بالا مشاهده می‌شود مقدار ضریب β_0 برابر حدود 45/5 و ضریب β_1 در حدود 0/007- می‌باشد که بیان می‌کند با یک واحد تغییر وزن ماشین (Weight)، مقدار مسافت طی شده در ازای یک گالن سوخت 0/007 کم خواهد شد. در ستون‌های آمارهٔ t و sig برای هر یک از ضرایب حذفی اولیه برابر صفر بودن آنها مورد آزمون قرار گرفته است که با توجه به sig=0/000 برای هر دو پارامتر β_0 و β_1 فرض برابر صفر بودن آنها رد می‌شود.

با توجه به اطلاعات جداول بالا می‌توان مدل رگرسیون خطی میان متغیر Weight و mpq را به صورت $Y = 45/5 - 0/007X$ کرد که در آن Y متغیر mpq و X متغیر، Weight می‌باشد.

اگر مقدار ضریب تعیین به اندازه کافی بزرگ بود می‌توانسیتم مدل فوق را مدلی مناسب ارزیابی نموده و

در جهت بهبود آن قدمهای بعدی را برداریم اما با توجه به مناسب نبودن ضریب تعیین نیاز به ورود متغیرهای مستقل جدیدی که مقدار بیشتری از تغییرات متغیر وابسته را توجیه نماید.

اگر متغیر Hours را نیز به عنوان متغیر مستقل وارد مدل نمائیم آنگاه مدل نهایی به صورت:

$$mpq = 44/7 - 0/005(weight) - 0/061(Hourse)$$

در خواهد آمد، با توجه به مقدار sig=0/000 می‌توان استنباط نمود که مدل خطی، مدلی مناسب است اما با توجه به مقدار ضریب تعیین که برابر 0/675 است باز هم نیاز به ورود متغیرهای مستقل تأثیرگذار بیشتری به مدل احساس می‌شود.

همانگونه که در فروض مدل عنوان شد، متغیرهای مستقلی که در مراحل بعد برای بهبود ضریب تعیین وارد می‌شود نباید با متغیرهای مستقل موجود در مدل همبستگی داشته باشند.

ورود متغیرهای مستقل می‌تواند به صورت بلوکی صورت گیرد، این عمل سبب می‌شود مدل نهایی رگرسیون، به صورت مرحله‌ای ساخته شود و مدل در هر مرحله (بلوک) با مدل بلوک قبلی در کلیه پارامترها و جداول بالا قابل مقایسه باشد.

برای مثال اگر ابتدا متغیر Weight وارد شود و دکمهٔ Next را در قسمت Block بزنید و دو متغیر Weight و Hourse را وارد نمائید آنگاه در تمام جداول خروجی امکان مقایسه خروجی‌ها برای مدلی با متغیر مستقل Weight و مدلی با متغیرهای مستقل Weight و Hourse را فراهم می‌آورد، با مقایسه نتایج می‌توانید مؤثر بودن یا نبودن متغیر جدید Hourse را که وارد مدل نموده‌اید، ملاحظه کنید.

11ـ3ـ نحوه ورود متغیرهای مستقل به مدل رگرسیونی

جعبهٔ Method روش ورود متغیرهای مستقل را به مدل رگرسیونی نشان می‌دهد، این روشها عبارتند از:

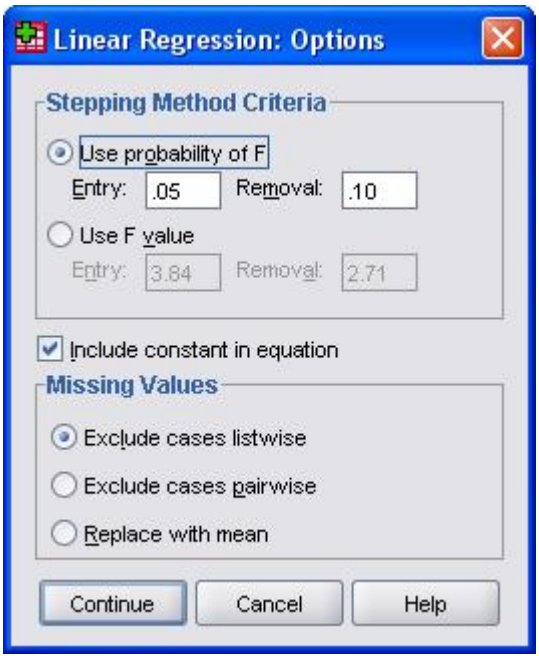
ـ روش Enter

در این روش، کلیه متغیرها در یک مرحله و به ترتیبی که در پنجرهٔ Independent (s) وارد شده‌اند، به

مدل رگرسیون وارد می‌شوند. شدت یا ضعف وابستگی متغیر وابسته به این متغیرها اثری در شکل نهایی مدل نخواهد داشت و در نهایت تمامی متغیرهای مستقل به مدل وارد می‌شوند. روش ورود متغیرهای مستقل از بلوک متغیرهاست و هر بلوک از متغیرها می‌تواند روش خاص خود را در ورود به مدل داشته باشد. انتخاب یک بلوک جدید با فعال کردن کلید Next و برگشت به بلوکهای قبلی با فعال کردن Previous انجام می‌شود.

ـ روش ورود مرحله‌ای متغیرها

اگر در جعبهٔ Method گزینهٔ Stepwise را انتخاب کنیم، ورود متغیرهای جعبهٔ Independent (s) در مدل از متغیری که بیشترین همبستگی را با متغیر وابسته دارد تا متغیری که کمترین همبستگی را با متغیر وابسته دارد صورت می‌پذیرد، البته معیار میزان اهمیت متغیر در مدل مقدار سطح معنی‌داری یا آمارهٔ F متناظر با آن در جدول آزمون معنی‌داری متغیرهاست. این مقادیر به همراه برخی ویژگی‌های دیگر مدل را می‌توان با فعال نمودن کلید ... Options تعیین کرد:



در بخش Stepping Method Criteria معیار توقف روش مرحله‌ای را تعیین می‌کنند. گزینهٔ Use

Probability of F برای انتخاب معیارها بر اساس مقدار سطح معنی‌داری آمارهٔ F، به کار می‌رود.

مقدار جعبه Entry حداکثر مقدار سطح معنی‌داری را نشان می‌دهد که ضرایب رگرسیونی متغیرهای ورودی به مدل تحمیل می‌کنند.

در مقابل جعبهٔ Removal حداقل سطح معنی‌داری را برای خروج متغیرها در هر مرحله از مدل رگرسیونی نشان می‌دهد. برای نمونه اگر مقدار سطح معنی‌داری در آزمون ضریب رگرسیونی متغیری از عدد 0/05 کمتر باشد، متغیر متناظر آن به مدل وارد می‌شود و اگر مقدار از 0/1 بیشتر باشد از مدل خارج می‌شود. به همین ترتیب گزینهٔ Use F Value می‌تواند همین حدود را برای آماره F متناظر با هر سطح معنی‌داری تکرار نماید.

گزینهٔ Include Constanst in Equation سبب حضور عدد ثابت (عرض از مبدأ) در مدل رگرسیونی خواهد شد.

نحوهٔ رفتار با مقادیر گمشده داده‌ها در بخش Missing Values تعیین می‌شود گزینهٔ Exclude Cases Listwise سبب حذف فهرست گونه مشاهدات دارای مقادیر گمشده است (حذف مشاهده از کلیه متغیرها)، گزینهٔ Exclude Cases Pairwise سبب حذف جفت گونه مشاهدات گمشده از متغیر مستقل و وابسته خواهد شد و با انتخاب Replace with Mean مقادیر گمشده با میانگین مشاهدات جایگزین می‌شوند. همانطور که در جدول Variable Entered / Removed ملاحظه می‌شود از آنجا که متغیر Weight بیشترین همبستگی را با mpq داشته است، آن را وارد مدل نموده و در قالب مدل شماره یک آن را معرفی نموده است، در این مدل مقدار ضریب تعیین برابر 0/652 بوده است و در جدول ANOVA، مدل خطی مناسب تشخیص داده شده است، همینطور مدل به صورت زیر با توجه به اطلاعات جدول Coefficients برازش شده است.

$$mpq = 45/38 - 0/007(weight)$$

در مرحلهٔ دوم، متغیر Hourse به مدل اضافه شده است و ضریب تعیین بـه مقـدار 0/675 افـزایش یافتـه است، همینطور با وجود متغیر جدید Hourse همچنان نتایج جدول ANOVA (sig≤α=0/05) نشان‌دهندهٔ مناسب بودن رگرسیون خطی در برازش مدل است. نتایج برآورد پارامترهای مدل بـه صـورت زیـر در جـدول Coefficients موجود است.

$$mpq = 44/7 - 0/005(weight) - 0/061(Hourse)$$

در آخرین جدول خروجی با نام Exclude Variables لیست متغیرهای خارج شده از مدل در هر مرحلـه را نشان می‌دهد، همانطور که ملاحظه می‌شود متغیر Engine، به دلیل عدم وجود همبستگی جزئی شدید بـا متغیر mpq از مدل خارج شده است و دلیلی برای وارد کردن آن در مدل وجود ندارد.

ـ روش حذف متغیرهای ورودی

با انتخاب گزینهٔ Remove در جعبهٔ Method می‌توان متغیرهای با اثربخشی کم را از هر بلوک در مـدل رگرسیونی حذف نمود. البته ذکر این نکته ضروری است که این روش را نمی‌توان به عنوان روش اولین بلـوک انتخاب کرد زیرا می‌بایست متغیرها در یکی از بلوک‌های قبلی بـه مـدل وارد شـوند و در بلوکهـای بعـدی بـا انتخاب این روش آنها را حذف نمود.

ـ روش حذف تدریجی متغیرها:

با انتخاب گزینهٔ Backward می‌توان روش حذف تدریجی را بر روی متغیرهای مسـتقل اعمـال نمـود. بـا اجرای این روش همان معیارهای پذیرش و رد متغیرها از طریق پنجرهٔ Removal در گزینهٔ ... Options که در روش Enter توضیح داده شد برای حذف تدریجی متغیرها به ترتیب کم اهمیت بودن آنها صورت می‌گیرد. برای مثال خروجی اجرای این روش هنگامی که سه متغیر مستقل Weight، Engine و Hourse را در برابـر متغیر وابسته mpq در مدل رگرسیون خطی برازش می‌نمائیم به صورت زیر است:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Horsepower, Vehicle Weight (lbs.), Engine Displacement (cu. inches)[a]	.	Enter
2	.	Engine Displacement (cu. inches)	Backward (criterion: Probability of F-to-remove >= .100).

a. All requested variables entered.

b. Dependent Variable: Miles per Gallon

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.822[a]	.676	.673	4.461
2	.822[b]	.675	.674	4.459

a. Predictors: (Constant), Horsepower, Vehicle Weight (lbs.), Engine Displacement (cu. inches)

b. Predictors: (Constant), Horsepower, Vehicle Weight (lbs.)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	46.374	1.295		35.806	.000
	Vehicle Weight (lbs.)	-.005	.001	-.501	-4.130	.000
	Engine Displacement (cu. inches)	-.006	.007	-.097	-.783	.435
	Horsepower	-.056	.012	-.672	-4.552	.000
2	(Constant)	46.341	.859		53.930	.000
	Vehicle Weight (lbs.)	-.005	.001	-.504	-4.548	.000
	Horsepower	-.047	.008	-.564	-5.945	.000

a. Dependent Variable: Miles per Gallon

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
2	Engine Displacement (cu. inches)	-.097[a]	-.783	.435	-.082	.054

a. Predictors in the Model: (Constant), Horsepower, Vehicle Weight (lbs.)

b. Dependent Variable: Miles per Gallon

همانطور که در جدول اول خروجی بالا مشاهده می‌شود ابتدا سه متغیر مستقل در مدل 1 وارد مـی‌شـوند

اما در مدل دوم متغیر Engine از لیست متغیرهای ورودی حذف می‌شود و مدل بهینه دوم که مقدار ضریب تعیین تبدیل شده آن (Adjust R Square) بالاتر از مدل اول با وجود متغیر Engine در مدل است، برازش می‌شود.

همانطور که در جدول Coefficient ملاحظه می‌شود در مدل اول فرض برابر صفر بودن ضریب متغیر Engine با توجه به مقدار sig=0/432 رد نمی‌شود و این به معنی عدم تأثیر متغیر Engine بر روی متغیر mpq در حضور متغیرهای Weight و Hourse است به همین دلیل متغیر Engine از مدل حذف شده و مدل برازش شده خطی رگرسیونی به صورت :

$$mpq = 44/7 - 0/061(\text{Hourse}) - 0/005(\text{weight})$$

خواهد بود که ضریب تعیین برابر با 0/67 دارد و به این معنی است که 67% تغییرات متغیر mpq از طریق تغییرات متغیرهای Weight و Hourse در یک رابطهٔ خطی قابل توجیه است.

در جدول آخر با نام Exclude Variables نام متغیرهای حذف شده از مدل بیان شده است، همانطور که ملاحظه می‌شود متغیر Engine از مدل شماره 2 حذف گردیده است.

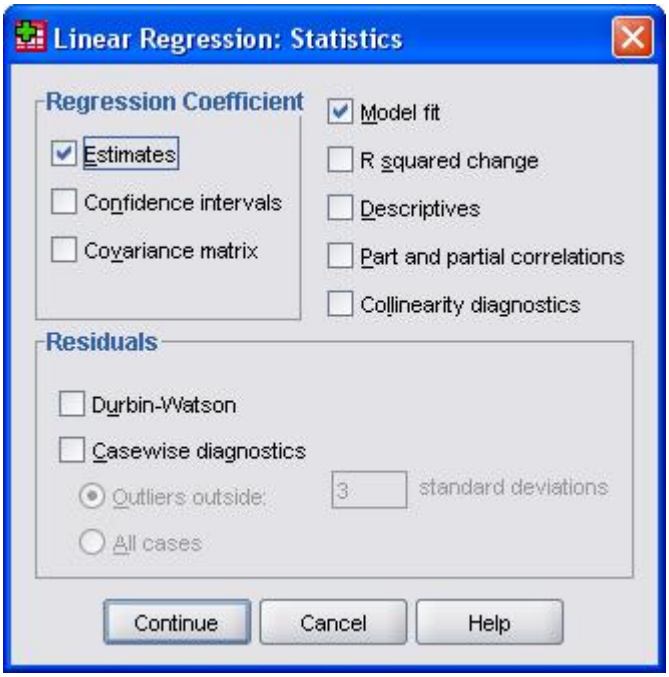
ـ روش ورود تدریجی متغیرها

با انتخاب گزینهٔ Forward می‌توان متغیرها را یکی پس از دیگری (در صورت معنی‌دار بودن همبستگی با متغیر وابسته) به مدل وارد کرد. معیار ورود متغیرها، اهمیت بیشتر آنها است. میزان اهمیت هر متغیر از مقایسه ضریب همبستگی با مقدار Y بدست می‌آید. مراحل انجام فرآیند انتخاب توسط این روش درست بر عکس روش Backward است یعنی ابتدا متغیر با همبستگی شدید را وارد مدل اول می‌نماید و پس در مراحل بعد در صورت معنی‌داری ورود متغیرهای دیگر آنها را به تدریج واردمدل می‌کند.

11ـ4ـ محاسبه آماره‌ها در رگرسیون خطی

برای دریافت اطلاعات جانبی و گرفتن آماره‌ها در مدل رگرسیون، کلید

... Statistics را فعال می‌کنیم تا پنجره‌ایی به شکل زیر ظاهر شود:



در بخش Regression Coefficient نحوه نمایش ضرایب رگرسیونی نمایش داده می‌شود. گزینهٔ Estimates برآوردهای ضرایب رگرسیونی را در قالب اعداد ستون B در جدول Coefficient نمایش می‌دهد.

با انتخاب گزینهٔ Confidence Intervals فاصلهٔ اطمینان برآورد ضرایب نشان داده می‌شود. با انتخاب این گزینه برای مدل با متغیرهای مستقل Weight و Hourse و متغیر وابسته mpq و انتخاب روشی Enter ستونهای با عنوان 95% Confidence Interval for B به جدول Coefficients به صورت زیر اضافه می‌شود:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	44.777	.825		54.307	.000	43.156	46.398
	weight	-.005	.001	-.551	-9.818	.000	-.006	-.004
	horse	-.061	.011	-.299	-5.335	.000	-.084	-.039

a. Dependent Variable: mpg

با انتخاب گزینهٔ Covariance matrix، ماتریس همبستگی میان متغیرهای مستقل به صورت زیـر ظـاهر می‌شود:

Coefficient Correlations[a]

Model			horse	weight
1	Correlations	horse	1.000	-.857
		weight	-.857	1.000
	Covariances	horse	.000	-5E-006
		weight	-5E-006	2.6E-007

a. Dependent Variable: mpg

با انتخاب گزینهٔ Model Fit (برازش مـدل)، جـدول تحلیـل واریـانس و معیارهـای ضـریب همبسـتگی و انحراف معیارهای متغیرهای مستقل در خروجی بیان می‌شود که نتیجه انتخاب این گزینه که در حالت عادی فعال است را در خروجی‌های قبلی ملاحظه نموده‌ایم.

انتخاب گزینهٔ R Square Change، مقدار ضریب تعیین تبدیل شـده را در جـدول Model Summary ارائه می‌کند.

انتخاب گزینهٔ Descriptive باعث ارائه خلاصه‌های توصیفی از متغیرهای مستقل و وابسته در پنجره‌ای به نام Descriptive Statistics به صورت زیر می‌شود:

Descriptive Statistics

	Mean	Std. Deviation	N
Miles per Gallon	23.45	7.805	392
Vehicle Weight (lbs.)	2967.38	852.294	392
Horsepower	104.21	38.233	392

با انتخاب گزینهٔ Part and Partial Correlations ، ضریب همبستگی هر یک از متغیرهـای مسـتقل بـا

متغیر وابسته در صورت ثابت ماندن سایر متغیرهای مستقل ارائه شده است. نتیجه انتخاب گزینهٔ فـوق بـرای متغیرهای مستقل Weight و Hourse و متغیر وابسته mpq به صورت زیر است:

Correlations

		Miles per Gallon	Vehicle Weight (lbs.)	Horsepower
Pearson Correlation	Miles per Gallon	1.000	-.807	-.771
	Vehicle Weight (lbs.)	-.807	1.000	.857
	Horsepower	-.771	.857	1.000
Sig. (1-tailed)	Miles per Gallon	.	.000	.000
	Vehicle Weight (lbs.)	.000	.	.000
	Horsepower	.000	.000	.
N	Miles per Gallon	392	392	392
	Vehicle Weight (lbs.)	392	392	392
	Horsepower	392	392	392

گزینهٔ Collinearity Diagnostics به منظور تشخیص همخطی میان متغیرهای مستقل به کار مـی‌رود، هم خطی میان متغیرهای زمانی رخ می‌دهد که یک یا چند متغیر مستقل در مدل رگرسیونی، ترکیب خطـی از سایر متغیرهای مستقل موجود در مدل باشد. خروجی انتخاب این گزینه به صورت زیر برای است:

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	Vehicle Weight (lbs.)	Horsepower	Engine Displacement (cu. inches)
1	1	3.853	1.000	.00	.00	.00	.00
	2	.124	5.572	.15	.00	.00	.07
	3	.017	15.100	.03	.08	.97	.17
	4	.006	24.806	.83	.92	.02	.76

a. Dependent Variable: Miles per Gallon

ستون Dimension بعد ریشه راکد را نشان می‌دهد. تعداد مقادیر ویژه برابر تعداد پارامترهای مدل اسـت. ستون Eigen Value مقدار ریشه راکـد را نشـان مـی‌دهـد. سـتون Conditional Index شـاخص شـرطی متناظر با هر ریشه را نشان می‌دهد که در آن K تعداد ریشهٔ راکد اسـت. مـاتریس Variance Proportions سهم هر یک از متغیرهای مستقل موجود در مدل را در ریشه‌های راکد محاسبه شده (بر حسب نسبت) نشان می‌دهد. برای مثال در ستون اول با عنوان Constant 0/83، سهم واریانس این متغیر به ریشـهٔ راکـد چهـارم

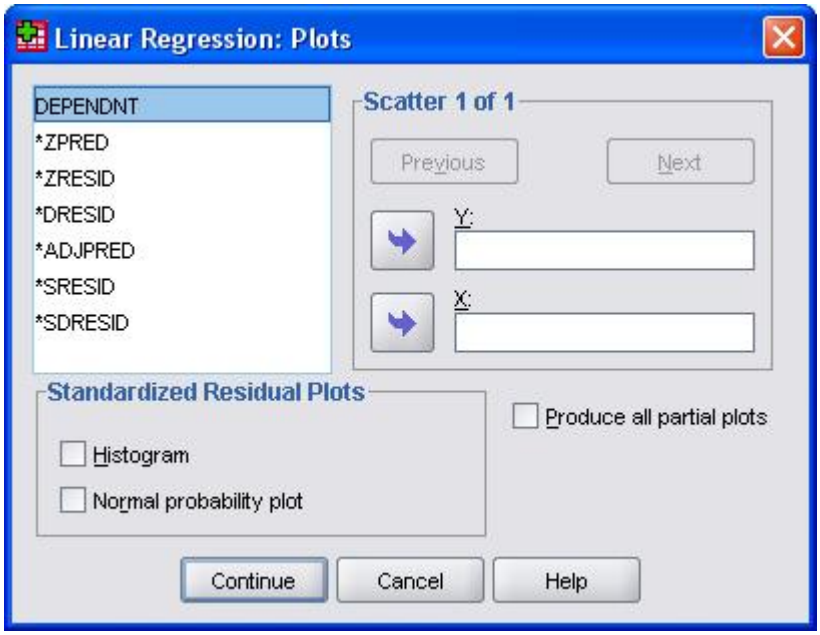
اختصاص دارد. به همین ترتیب برای متغیرهای مستقل این جفت‌ها بیان شده است. ضرایبی که درصد بالایی را در یک ریشه راکد نشان دهند به یکدیگر وابسته‌اند برای نمونه در خروجی بالا Weight و Engine با توجه به مقدار بالای آنها در ریشهٔ راکد چهارم بـه یکـدیگر نزدیکنـد یـا در حقیقـت میـان متغیرهـای Engine و Weight یک رابطهٔ خطی برقرار است به همین دلیل در یک مدل مناسب لزومی ندارد هـر دو آنهـا در مـدل خطی حضور داشته باشند.

11ـ5ـ رسم نمودارها در رگرسیون خطی:

رسم نمودارها در رگرسیون از اهمیت ویژه‌ای برخوردار است، رسـم نمودارهـا در رگرسـیون خطـی بـه دو منظور صورت می‌گیرد:

1ـ آزمون فرض‌های اساسی مدل مانند $E(e_i)=0$ یا $Var(e_i)=\sigma^2$ و مقدار σ^2 یک مقدار ثابت.

2ـ یافتن نقاط مؤثر یا داده‌های پرت مؤثر در برازش مدل رگرسیونی برای رسم نمودارها، کلیـد ... Plots را فعال می‌نمائیم تا جعبهٔ گفتگوی زیر ظاهر شود:

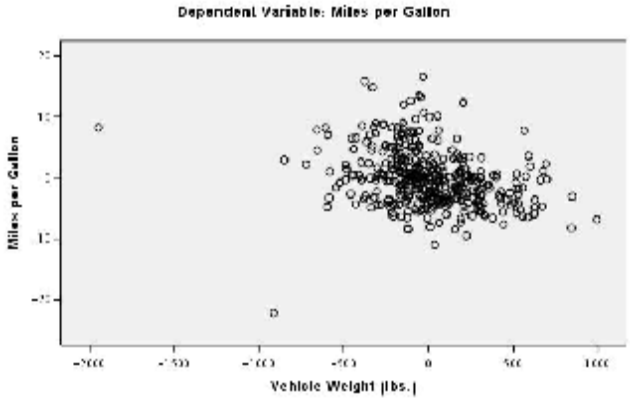


جعبهٔ سمت چپ فهرست کلیهٔ متغیرهایی را که می‌توان نمودار آنها را در مقابل یکدیگر رسم نمـود نشـان می‌دهد. این متغیرها عبارت هستند از:

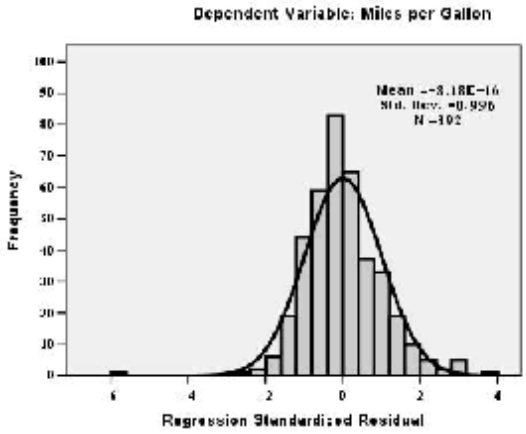
- DEPENDENT، متغیر وابسته در مدل رگرسیونی.
- ZPRED*، متغیر مقادیر پیش‌بینی‌های استاندارد شده متغیرهای وابسته.
- ZRESID*، متغیر مقادیر خطاهای استاندارد شده.
- DRESID*، متغیر مقادیر باقیمانده‌های حذفی.
- ADJPRED*، متغیر مقادیر تعدیل شده پیش‌بینی‌ها.
- SRESID*، متغیر مقادیر استودنت شده.
- SDRESID*، متغیر مقادیر استودنت شده حذفی.

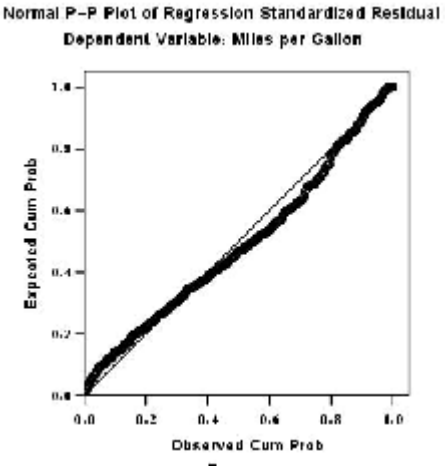
جعبهٔ سمت راست، به ورود متغیرهای متناظر با محورهای افقی و عمـودی اختصـاص دارد. متغیـر محـور عمودی در جعبهٔ Y و متغیر محور افقی در جعبه X جای می‌گیرد. با فعال کردن Next می‌تـوان محورهـای سایر نمودارها را فعال کرد. برای تغییر مشخصه‌های محورهای قبلی Previous را فعال می‌نمائیم.

نمودارهای قابل ترسیم در ایـن پنجـره از نـوع نمودارهـای پـراکنش (Scatter) هسـتند. انتخـاب گزینـهٔ Produce all Partial Plots، سبب تولید کلیه نمودارهای جزئی مرتبط با نمودار ترسیمی خواهد شـد. نـوع نمودارهای جزئی به نوع مدل انتخابی و متغیرهای موجود در آن بستگی دارد. برای مثال در شکل زیر نمـودار پراکنش mpq در مقابل متغییر Hourse رسم شده است:



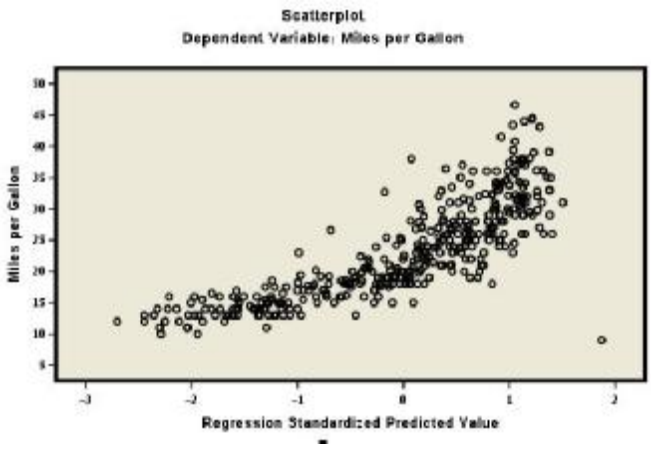
در بخش Standardized Residual Plots نوع نمودارهای ممکن از خطاها را نشان می‌دهد. با انتخاب گزینهٔ Histogram می‌توان نمودار هیستوگرام خطاها را ترسیم نمود و گزینهٔ Normal Probability Plots نمودار خط نرمال خطاها را رسم می‌نماید. نتایج انتخاب این دو گزینه، شکل‌های زیر می‌باشد:





در نمودار P-P Plot هر چه داده‌ها به خط وسط نزدیکتر باشند نشان دهندهٔ نرمال بودن توزیع خطاها است.

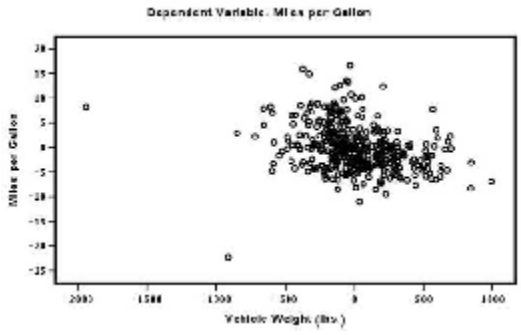
نمودار پراکنش خطاها در برابر مقادیر برازش شده متغیر وابسته (DEPENDENT در مقابل ZRESID*) به نوعی نشان‌دهنده تصادفی بودن نقاط خطا در برابر مقادیر برازش شده متغیر پاسخ است. شکل زیر نمودار پراکنش متغیر پاسخ در مقابل مقادیر خطا را نشان می‌دهد:

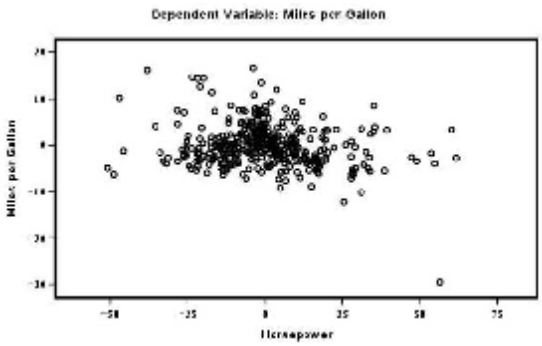


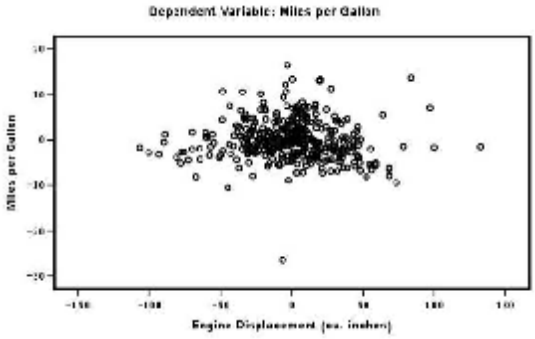
ذکر این نکته ضروری است که میان برازش شده Y و مقادیر e باید استقلال وجود داشته باشد و نتوان بین آنها یک رابطهٔ خطی ایجاد نمود.

11ـ6ـ بررسی فرض خطی بودن مدل رگرسیون برازش شده

هنگامی که متغیر وابسته یک تابع غیرخطی از متغیرهای مستقل باشد، با استفاده از ترسیم مانده‌ها در مقابل مقادیر برازش شده متغیر وابسته می‌توان به مناسب بودن یا نبودن مدل رگرسیون خطی در برازش مدل پی برد، در شکل بالا مقادیر خطا در مقابل مقادیر برازش شده متغیر وابسته ملاحظه نمودید. همانطور که ملاحظه شد می‌توان وجود یک روند به صورت حرف V را در شکل فوق ملاحظه نمود که نشان‌دهندهٔ عدم خطی بودن رابطهٔ میان متغیرهای مستقل با متغیر وابسته است. اگر رابطه هر یک از متغیرهای مستقل را با متغیر وابسته در شکل‌های زیر ملاحظه نمائید خواهید دید در همهٔ آنها یک رابطهٔ غیرخطی بین متغیر مستقل و متغیر وابسته برقرار است:

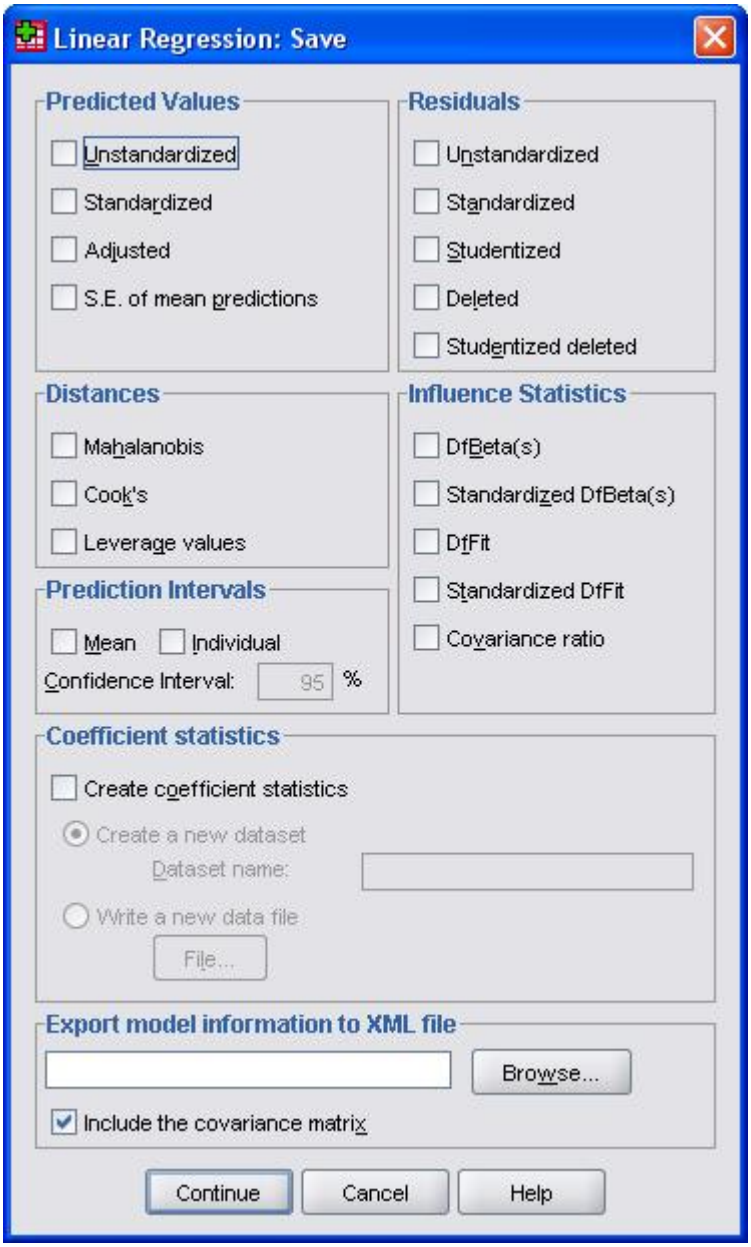






11ـ7ـ ذخیره اطلاعات فرعی در رگرسیون خطی

با فعال کردن کلید ... Save در پنجرهٔ اصلی رگرسیون خطی جعبهٔ زیر ظاهر می‌شود:



با استفاده از گزینه‌های مختلف این جعبه، می‌توان اطلاعات مختلف حاصل از برازش مدل رگرسیون خطی را در پنجرهٔ DataView در کنار متغیرهای موجود ذخیره نمود.

بخش Predicted Values انواع متغیرهای پیش‌بینی متغیر وابسته را که قابل ذخیره‌سازی هستند نشـان می‌دهد.

بخش Residual's ، انواع خطاها را در پنجرهٔ Data View نشان می‌دهد.

بخش Distances انواع فواصل را برای هر نمونه به منظور دست‌یابی به مدل بهینه در Data View نمایش می‌دهد.

بخش Influence Statistics آماره‌هایی را محاسبه می‌نماید که میزان مؤثر بودن نمونه‌ها را نشان می‌دهد و بخش Prediction Intervals فاصلهٔ اطمینان پیش‌بینی‌ها را مشخص می‌کند.

با انتخاب گزینهٔ Coefficient Statistics در قسمت Save to New File و دادن آدرس یک مکان برای ذخیره، می‌توان تمامی اطلاعاتی را که خواهان نمایش آنها در Data View سیستم به فایل دیگری منتقل نمائیم.

با مشخص کردن مسیر یک فایل از طریق Browse برای گزینه Export Model Information To XML File اطلاعات برازش مدل رگرسیونی را در قالب یک فایل HTML در قالب مسیری که در قسمت Browse بیان شده است نمایش می‌دهد.

با انتخاب گزینهٔ Include The Covariance Matrix ، ماتریس کواریانس متغیرها نیز در خروجی‌ها ظاهر می‌گردد.

تمرین:

1- مدل رگرسیونی mpg را در فایل cars بدست آورید. آیا مدل مورد نظر خطی است؟ مهمترین عامل تاثیر گذار بر روی mpg کدام متغیر است؟

2- مدل مناسب رگرسیونی را بر روی مزان درآمد افراد در فایل emlpoye data بدست آورید. آیا مدل مورد نظر خطی است؟ در صورت خطب نبودن آیا می توانید با استفاده از روشهای متداول تبدیل بر روی داده ها مدل ررا خطی نمائید.

3- مدل مناسب رگرسیونی بر روی وزن خودرو در فایل cars را بدست آورید.

فصل دوازدهم

آزمونهای ناپارامتری

اهداف آموزشی فصل دوازدهم

فراگیر پس از مطالعه این فصل باید بتواند:

- روشهای ناپارامتری در علم آمار را بشناسد.
- انواع فرمانهای آزمون ناپارامتری را نام برده و کاربر هر یک را شرح دهید.
- مراحل انجام فرمان chi-square ناپارامتری در نرم افزار Spss.16 را با ذکر مثال شرح دهد.
- مراحل انجام فرمان Binomial ناپارامتری در نرم افزار Spss.16 را با ذکر مثال شرح دهد.
- مراحل انجام فرمان Runs ناپارامتری در نرم افزار Spss.16 را با ذکر مثال شرح دهد.
- مراحل انجام فرمان One Sample K-S ناپارامتری در نرم افزار Spss.16 را با ذکر مثال شرح دهد.
- مراحل انجام فرمان 2 Independent Sample ناپارامتری در نرم افزار را با ذکر مثال شرح دهد.
- مراحل انجام فرمان K Independent Samples ناپارامتری در نرم افزار را با ذکر مثال شرح دهد.
- مراحل انجام فرمان 2 Independent Samples ناپارامتری در نرم افزار را با ذکر مثال شرح دهد.
- مراحل انجام فرمان K Independent Samples ناپارامتری در نرم افزار را با ذکر مثال شرح دهد.

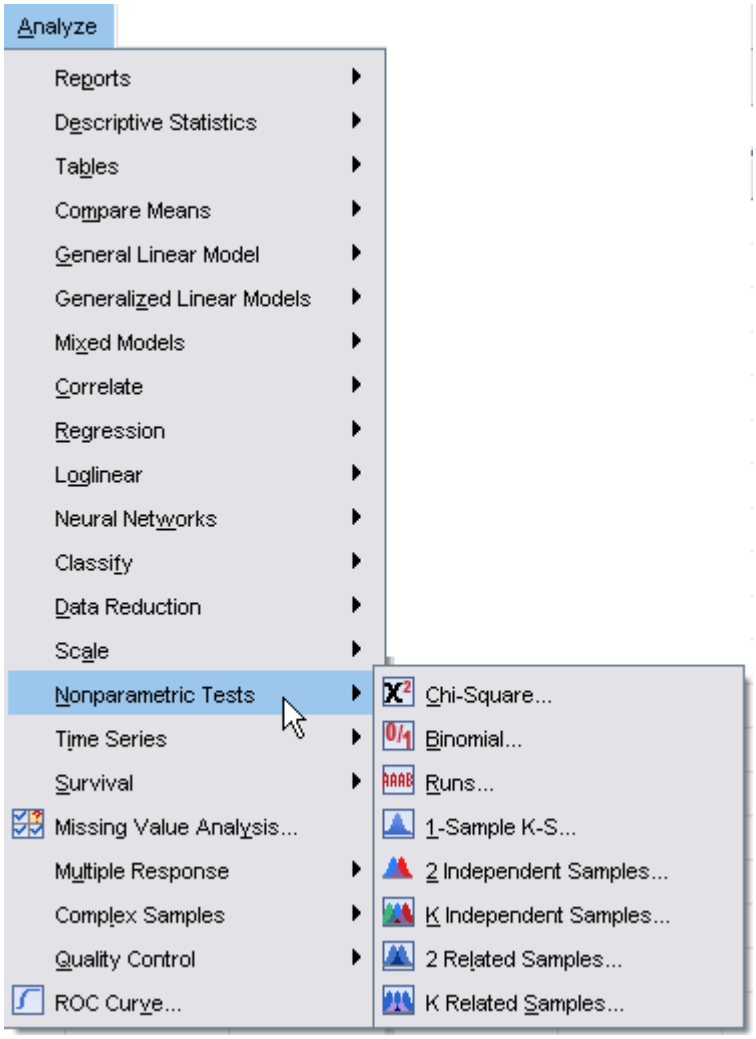
12ـ1ـ مفاهیم اولیه:

بسیاری از تکنیکهای آماری در آزمون فرض‌های آماری خود دارای چندین پیش‌فرض اولیه هستند که در صورت برقرار نبودن این پیش‌فرضها نمی‌توان به نتایج بدست آمده با دقت مورد نظر اطمینان نمود. یکی از مهم‌ترین این پیش‌فرضها داشتن توزیع نرمال متغیرهای پیوسته است که در صورت عدم وجود، استفاده از تکنیک‌های آماری را دچار مخاطره می‌نماید، از دیگر پیش‌شرط‌های مهم وجود تعداد به اندازه کافی داده‌ها برای استنباط آماری است، اگر تعداد داده‌های آماری کم باشد، آماره‌ها در بسیاری از آزمونهای آماری کارایی خود را از دست خواهند داد.

برای رفع این مشکل در زمان نبود توزیع مورد نظر در داده‌ها و یا تعداد کم داده‌ها به جای استفاده از روشهای عمومی و کلاسیک آماری از روشهای استفاده می‌نماید که به نام روشهای ناپارامتری یا توزیع آزاد شناخته شده‌اند.

روشهای ناپارامتری در علم آمار همانند روشهای محاسبات عددی در ریاضیات عمل می‌نمایند. مسائلی را که در ریاضیات با استفاده از روشهای کلاسیک قابل حل نیست در محاسبات عددی با استفاده از تکنیک‌هایی نظیر روشهای تکرار حل می‌نمایند.

زیرمنوی Non Parametric Tests در منو Analyze برای اجرای انواع مختلف آزمونهای پارامتری در نظر گرفته شده است.



هر یک از فرمانهای آزمون ناپارامتری خاصی را با هدف معینی اجرا می‌نمایند. وظایف ایـن فرمانهـا بـه طـور خلاصه به شرح زیر است:

ـ Chi-Square: آزمون مربع کای دو، فرض برابری نسبت هر یک از سطوح یـک متغیـر گسسـته را آزمـون می‌نماید. این آزمون از آمارهٔ کای دو استفاده نموده و اساس آن تفاوت‌های بین فراوانیهای مشاهده شده و مـورد انتظار در هر سطح از متغیر گسسته است.

ـ Binomial، آزمون دو جمله‌ای، این آزمون برای مقایسه P (نسبت موفقیت) بـا یـک عـدد خـاص در یـک متغیر گسسته دو سطحی به کار می‌رود.

ـ Runs، آزمون گردش، نتیجه این آزمون تصادفی بودن یا نبودن ترتیب رخداد مقادیر متغیرهـا را مشـخص

می‌نماید.

ـ 1-Sample K-S، آزمون یک نمونه‌ای کولموگروف ـ اسمیرنوف ، مقایسه تابع توزیع مشاهدات یک متغیر با توزیع مفروض نرمال، یکنواخت، پواسون یا نمایی را انجام می‌دهد.

ـ 2 Independent Samples، آزمون دو نمونه مستقل، آزمون برابری میانگین‌ها در دو نمونه مستقل را آزمون می‌نماید.

ـ K Independent Samples، آزمون چند نمونه مستقل، میانگین چند نمونه مستقل را با یکدیگر آزمون می‌نماید.

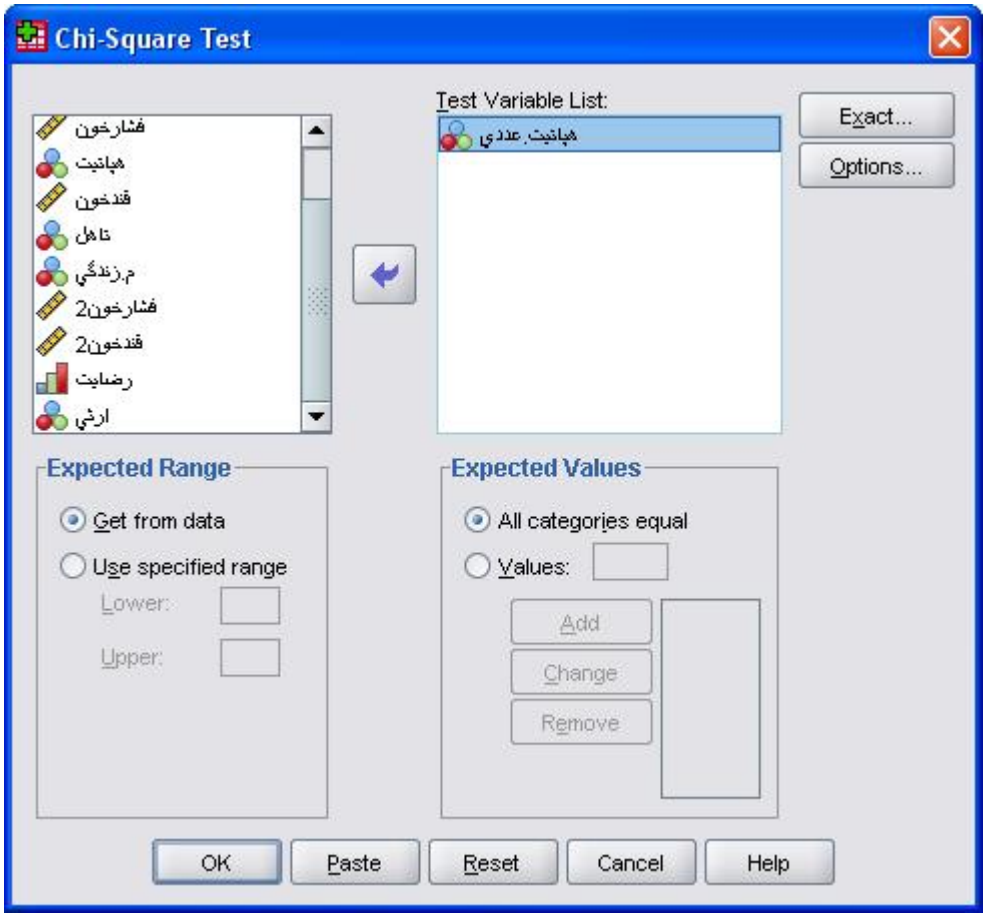
ـ 2 Related Samples، آزمون در نمونه وابسته، توزیع در متغیر وابسته به هم را با یکدیگر مقایسه می‌نماید.

ـ K Related Samples ، آزمون چند نمونه وابسته، توزیع چند متغیر وابسته به هم را با یکدیگر مقایسه می‌نماید.

12ـ2ـ فرمان Chi-Square

هر گاه بخواهیم نسبت سطوح یک متغیر گسسته را با یکدیگر مقایسه نمائیم یا در اصطلاح، توزیع یک متغیر گسسته چندسطحی را با توزیع یک متغیر مفروض که بین نسبت سطوح آن رابطهٔ خاصی برقرار است مقایسه نمائیم از آزمون Chi-Square استفاده می‌کنیم.

برای مثال می‌خواهیم آزمون نمائیم که آیا نسبت ابتلا به هر یک از انواع هپاتیت با هم برابر است یا نه؟ در حقیقت آزمون فرض $H_0 : P_A = P_B = P_C$ را در مقابل فرض جایگزین که حداقل یکی از نسبت‌ها با دیگر نسبت‌ها برابر نباشد آزمون می‌نمائیم. در فرمان Chi-Square، نرم‌افزار با شبیه‌سازی توزیع داده‌ها در فرض H_0 و مقایسه آن با توزیع داده‌های متغیر گسسته عمل می‌نمایند. با اجرای فرمان Chi-Square، جعبهٔ گفتگوی زیر باز می‌شود:



فهرست کلیه متغیرهای عددی در قسمت سمت چپ جعبهٔ بالا مشاهده می‌شود. از بین متغیرهای موجود، متغیرهایی را که آزمون بر روی مقادیر آنها اجرا می‌شود به جعبهٔ Test Variable List منتقل می‌نمائیم. در بخش Expected Range برد مورد انتظار متغیرها، تعیین می‌شود، اگر برد مقادیر از داده‌ها گرفته شود گزینهٔ Get From Data انتخاب می‌شود و اگر برد جدیدی برای اجرای آزمون در نظر گرفته شده گزینهٔ Use Specified Range را انتخاب می‌کنیم، سپس حد پائین مقادیر را در جعبه Lower و حد بالا را در جعبه Upper وارد می‌کنیم. در بخشExpected Values، مقادیر مورد انتظار سطوح مختلف را معرفی می‌نمائیم. اگر نسبت تعداد مشاهدات مورد انتظار در فرض H_0 در هر رده یکسان فرض شود، گزینه All Categories Equal انتخاب می‌شود، در غیر این صورت گزینهٔ Values انتخاب می‌شود.

با وارد کردن مقادیر به تعداد سطوح متغیر گسسته و فعال کردن کلید Add مقادیر نسبت‌های مورد انتظار به ترتیب هر یک از کدهای سطوح در جعبه پائین، زیر هم قرار خواهند گرفت، در صورت نیاز به تغییر مقادیر، بر روی نسبت مورد نظر Click کرده و عدد جدید را در جعبه Value وارد می‌نمائیم تا کلید Change برجسته شود، با فعال شدن این کلید، عدد جدید جایگزین عدد قبلی می‌شود.

اگر بخواهیم مقدار خاصی را حذف نمائیم بر روی عدد مورد نظر Click نموده و سپس کلید Remove را فعال می‌کنیم.

برای مثال اگر بخواهیم فرض برابر بودن نسبت افراد مبتلا به هر یک از انواع هپاتیت را آزمون نمائیم، متغیر نوع هپاتیت را که از طریق دستور Auto Recode به صورت کدگذاری عددی درآمده وارد جعبهٔ Test Variable List می‌نمائیم و چون فرض $H_0 : P_A = P_B = P_C$ است در قسمت Expected Values گزینهٔ All Categories Equal را انتخاب می‌نمائیم. خروجی فرمان به صورت زیر است:

نوع هپاتیت بیماران مورد بررسی

	Observed N	Expected N	Residual
نوع هپاتیت A	9	10.0	-1.0
نوع هپاتیت B	8	10.0	-2.0
نوع هپاتیتC	13	10.0	3.0
Total	30		

Test Statistics

	نوع هپاتیت بیماران مورد بررسی
Chi-Square	1.400
df	2
Asymp. Sig.	.497

در جدول اول، تعداد مشاهدات و تعداد مقادیر مورد انتظار و تفاضل آنها در هر یک از سطوح نوع هپاتیت بیان شده است.

در جدول Test Statistics با استفاده از آمارهٔ کای دو، معیاری برای قضاوت در مورد فرض H_0 بیان شده است.

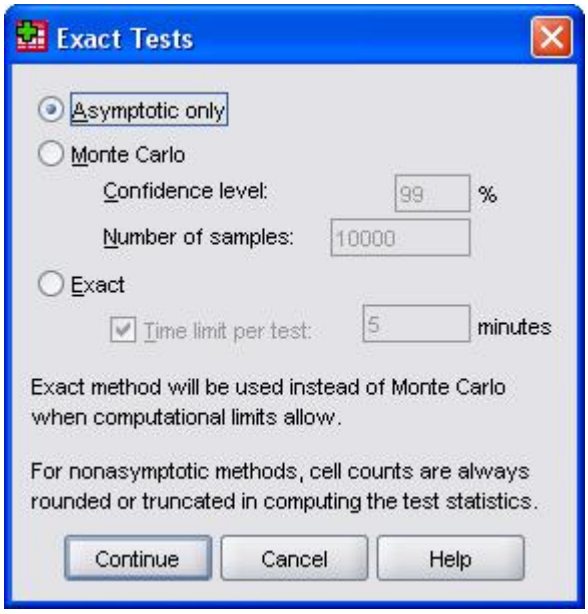
نحوهٔ قضاوت در مورد فرض H_0 در آزمونهای ناپارامتری هم مانند قبل با استفاده از سطح معنی‌داری (Asymp.sig) یا به اختصار sig صورت می‌گیرد، اگر $sig>\alpha$ باشد نمی‌توان در سطح اطمینان α-1% فرض H_0 را رد نمود و اگر $sig\leq\alpha$ آنگاه فرض H_0 رد خواهد شد، در آزمون جدول بالا در نظر گرفتن α=0/05 مشاهده می‌شود که $sig>\alpha$ و در نتیجه نمی‌توان فرض H_0 را رد نمود و این معنی آن است که نسبت افراد مبتلا به هپاتیت در هر یک از انواع آن یکسان است.

با فعال کردن کلید ... Options نوع آماره‌هایی را که محاسبه می‌شوند و چگونگی رفتار با مقادیر گمشده در محاسبات تعیین می‌شوند. با انتخاب این کلید جعبه زیر ظاهر می‌شود:

در بخش Descriptive Statistics با انتخاب آماره‌های توصیفی مختلف نظیر میانگین، انحراف معیار، کمترین مقدار و بیشترین مقدار برای متغیر گسسته و با انتخاب Quartiles چارکهای اول، دوم (میانه) و سوم برای متغیر گسسته در جدول Descriptive Statistics خروجی ظاهر می‌شود.

در بخش Missing Values روش حذف مقادیر گممشده انتخاب می‌شود، اگر گزینه Exclude Cases test-by-test انتخاب شود، در صورت اجرای آزمون بر روی یک متغیر، مقادیر گمشده آن را حذف می‌کند، اما اگر گزینهٔExclude Cases Listwise انتخاب شود، مشاهده‌ای که در یک متغیر مقدار گمشده دارد از تمامی متغیرها حذف می‌شود.

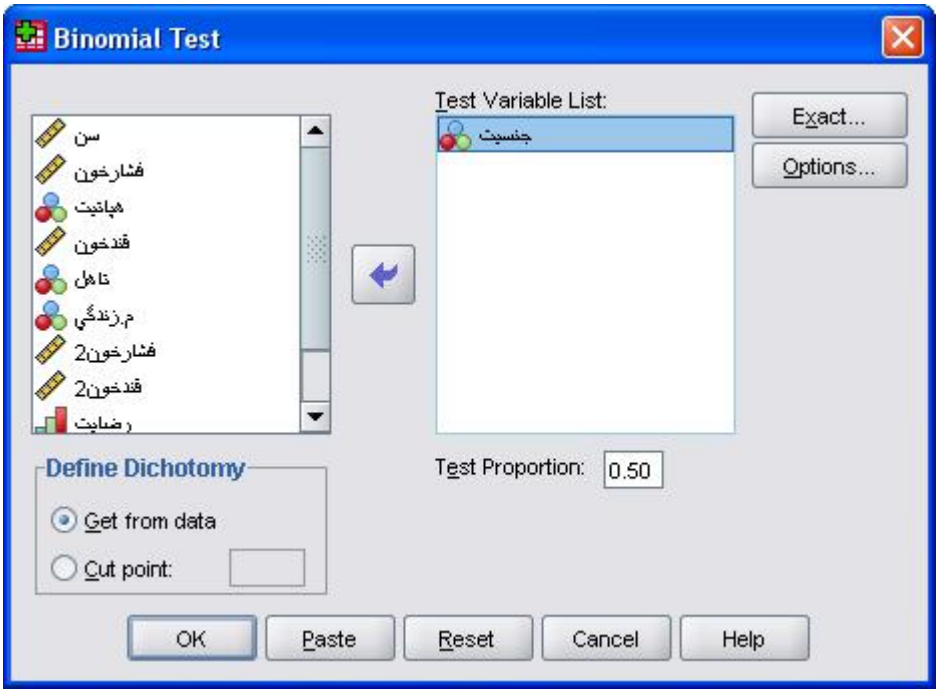
با فعال کردن کلید ... Excact جعبهٔ گفتگوی زیر ظاهر می‌شود:



محتویات این جعبه تعداد تکرارها و شرایط تکرار در آزمون را مشخص می‌نماید و در کلیه روشهای آزمونهای پارامتری نیز وجود دارد با استفاده از گزینهٔ Mont Carlo و Exact دو روش دیگر محاسبهٔ سطح معنی‌داری (sig) برای هر آزمون در اختیار کاربر قرار می‌گیرد، پیشنهاد می‌شود به منظور اجتناب از دست‌یابی به نتایج متفاوت در آزمونهای مختلف ناپارامتری گزینه‌های این جعبه را تغییر ندهید.

12ـ3ـ فرمان Binomial

با انتخاب این فرمان آزمون دو جمله‌ای اجرا می‌گردد، آزمون دو جمله‌ای بر روی داده‌های دودویی (دوسطحی) اجرا می‌شود، هر گاه بخواهیم توزیع داده‌ها را در این دو سطح با یکدیگر مقایسه نمائیم از این آزمون استفاده می‌کنیم. با اجرای این فرمان جعبه محاوره زیر ظاهر می‌شود.



در قسمت سمت چپ، لیست تمامی متغیرهای موجود، در فایل جاری Exam مشاهده می‌شود، متغیر گسسته دودویی را به قسمت Test Variable List منتقل می‌نمائیم و نسبت موفقیت یا نسبت سطح اول یا عدد P را که بین 0 و 1 خواهد بود وارد قسمت Test Proportion می‌نمائیم.

بخش Define Dichotomy، نحوه گروه‌بندی را نشان می‌دهد، اگر داده‌ها دو سطحی باشند، گزینه Get From Data انتخاب می‌شود، اما اگر داده‌ها بیش از دو سطح داشته باشندع یا پیوسته باشند، گزینهٔ Cut Point را انتخاب می‌نمائیم. در این حالت نرم‌افزار داده‌ها را به دو قسمت کوچکتر یا مساوی از عددی که در قسمت Cut Point وارد نموده‌ایم و قسمت بزرگتر از آن عدد تقسیم می‌کند و به این ترتیب متغیر پیوسته یا بیش از دو سطح تبدیل به یک متغیر دوسطحی می‌شود.

برای مثال می‌خواهیم بدانیم آیا نسبت ابتلا به هپاتیت در بین خانم‌ها با آقایان یکسان است یا خیر؟ برای این منظور از فرمان Binomial استفاده می‌نمائیم و متغیر جنسیت را وارد Test Variable List نموده و مقدار Test Proportion را عدد 0/5 در نظر می‌گیریم. در این حالت در حقیقت آزمون فرض

$H_0 : P = 0/5$

$H_1 : P \neq 0/5$

را انجام می‌دهیم. با تأئید فرمان خروجی زیر ظاهر می‌شود:

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
جنسیت	Group 1	مرد	13	.43	.50	.585[a]
	Group 2	زن	17	.57		
	Total		30	1.00		

a. Based on Z Approximation.

همانطور که در جدول Binomial Test ملاحظه می‌شود برای قضاوت در مورد رد یا قبول فرض H_0 باید سطح معنی‌داری (Asymp.sig) معیار قرار گیرد و با در نظر گرفتن α=0/05 ملاحظه می‌شود که sig>α پس در نتیجه فرض H_0 رد نخواهد شد و می‌توان نتیجه گرفت که خانم‌ها و آقایان به یک نسبت به هپاتیت مبتلا می‌شوند یا به عبارت دیگر جنسیت عامل اثرگذار در ابتلا به بیماری هپاتیت نیست.

کلیدهای ... Options و Exact همانند فرمان Chi-Square عمل می‌نمایند.

12ـ4ـ فرمان Runs

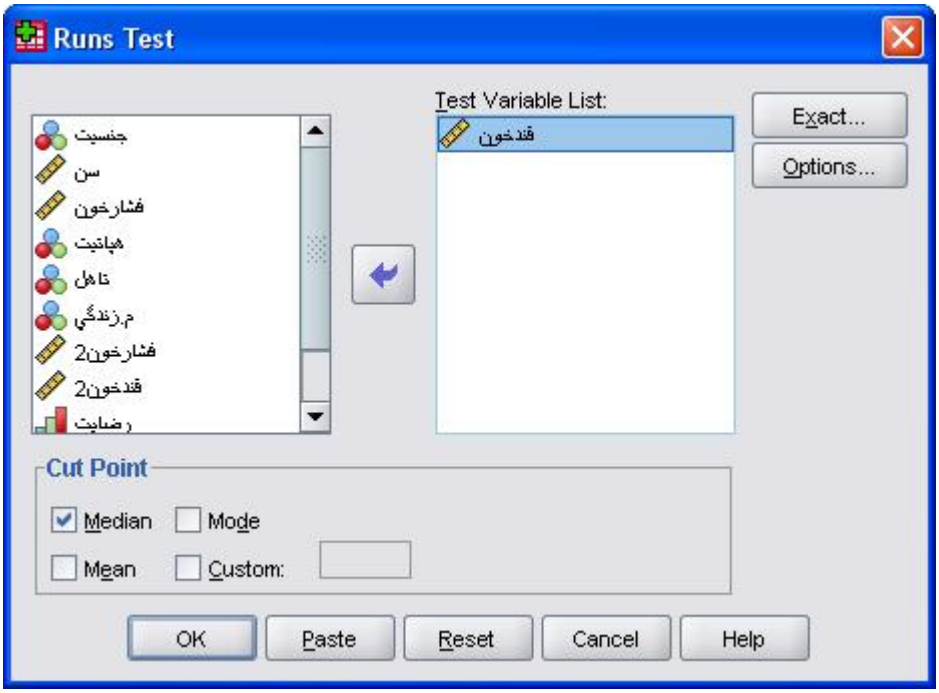
این فرمان به منظور بررسی تصادفی بودن داده‌های عددی به کار می‌رود، یکی از فرضهای اصلی در پردازش‌های آماری تصادفی بودن داده‌ها می‌باشد. اگر فرض تصادفی بودن داده‌ها برقرار نباشد هیچ یک از روشهای آمار استنباطی نمی‌توانند مورد استفاده قرار گیرند.

آزمون Runs بر این پایه استوار است که اگر مشاهدات مستقل از یکدیگر باشند از به هم پیوستن آنها دنباله‌ای تصادفی ایجاد می‌شود و اگر از هم مستقل نباشند به طور هم این دنباله تصادفی نخواهد بود، برای آزمون فرض

H_0 : داده‌ها تصادفی انتخاب شده‌اند

H_1: داده‌ها تصادفی انتخاب نشده‌اند

چهار معیار در قسمت Cut Point مشخص شده است، اگر داده‌ها به صورت تصادفی انتخاب شده باشند انتظار می‌رود پراکندگی آنها حول پارامترهای مرکزی نظیر میانه (Median)، مد (Mode) یا میانگین (Mean) و یا هر عدد سفارشی (Custom) با هم تفاوت معنی‌داری نداشته باشد. با اجرای این فرمان جعبه زیر ظاهر می‌شود.



متغیرهای مورد آزمون در جعبهٔ Test Variable List جای می‌گیرند، بخش Cut Point انتخابهای ممکن خط مبنا را نشان می‌دهد.

برای نمونه متغیر پیوسته میزان قند خون را دارد Test Variable List می‌کنیم و می‌خواهیم ببینیم داده‌های آن به صورت تصادفی حول خط مبنی میانگین پراکنده شده‌اند یا خیر؟ از قسمت Cut Point گزینه Mean را انتخاب می‌نمائیم با اجرای فرمان، نتایج زیر در خروجی ظاهر می‌شود:

Runs Test

	فندخون
Test Value[a]	390.00
Cases < Test Value	15
Cases >= Test Value	15
Total Cases	30
Number of Runs	21
Z	1.672
Asymp. Sig. (2-tailed)	.094

a. Median

همانطور که در جدول خروجی بالا با عنوان Run Test مشاهده می‌شود، در قسمت Test Value خط مبنی که میانگین داده‌هاست بیان شده است و در سطرهای دوم و سوم تعداد داده‌های بالاتر یا مساوی میانگین و تعداد داده‌های کوچکتر از میانگین بیان شده است.

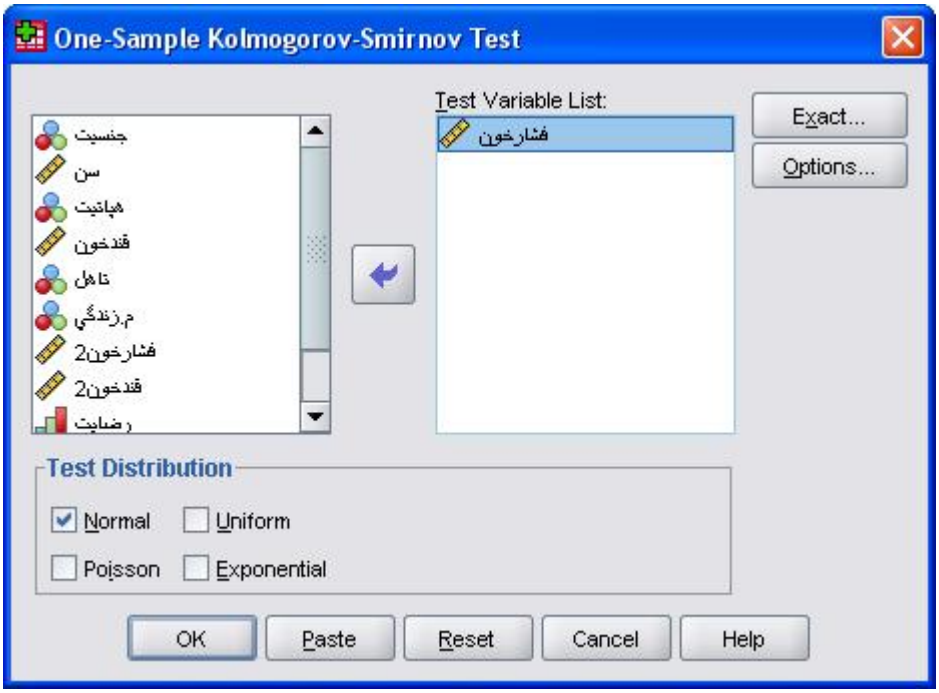
تعداد کل داده‌ها، تعداد گردش‌ها، مقدار آماره Z و سطح معنی‌داری دیگر خروجی‌های جدول هستند، همانطور که در قبل بیان شد معیار پذیرش یا رد فرض اولیه، مقدار سطح معنی‌داری دوطرفه است و اگر $sig>\alpha$ باشد فرض H_0 رد نمی‌شود و اگر $sig\leq\alpha$ باشد آنگاه فرض H_0 در مقابل فرض H_1 رد می‌شود.

با در نظر گرفتن $\alpha=0/05$ و مقدار $sig=0/363$ مشاهده می‌شود که فرض H_0 رد نخواهد شد و فرض تصادفی انتخاب شدن داده‌ها در سطح $\alpha=0/05$ رد نمی‌شود.

امکانات کلیدهای ... Options و ... Exact مانند فرمان Chi-Square است.

12ـ5ـ فرمان One-Sample K-S

یکی از روشهایی که آزمون می‌کند آیا داده‌ها از توزیع خاص پیروی می‌کنند یا نه؟ آزمون تک‌نمونه‌ای کولموگروف ـ اسمیرنوف (K-S) است. این آزمون با مقایسه تابع تجمعی احتمال مشاهدات و تابع تجمعی احتمال توزیع مفروض، فرض پیروی مشاهدات از توزیع احتمالی خاص را آزمون می‌نماید. با فعال کردن فرمان One Sample K-S جعبه گفتگوی زیر ظاهر می‌شود:



در قسمت Test Variable List، متغیرهایی را که می‌خواهیم توزیع آنها را با توزیع خاصی مقایسه نمـائیم وارد می کنیم، در قسمت Test Distribution امکان مقایسه توزیع متغیرهای وارد شده در جعبه Test Variable List با توزیع‌های نرمال (Normal)، یکنواخت (Uniform)، پواسون (Poisson) و نمایی (Exponenetial) وجود دارد. برای مثال می‌خواهیم فرض زیر را انجام دهیم:

H_0: فشار خون بیماران مورد بررسی از توزیع نرمال پیروی می‌کند

H_1: فشار خون بیماران مورد بررسی از توزیع نرمال پیروی نمی‌کند

بـرای ایـن منظـور متغیـر میـزان فشـار خـون را وارد پنجـرۀ Test Variable List نمـوده و از بخـش Test Distribution گزینه Normal را انتخاب نموده و اجرای فرمان را با کلیـد OK تأئیـد مـی‌نمـائیم. جـدول زیـر در خروجی ظاهر می‌شود:

One-Sample Kolmogorov-Smirnov Test

		فشار خون
N		28
Normal Parameters[a,b]	Mean	10.6786
	Std. Deviation	2.19517
Most Extreme Differences	Absolute	.171
	Positive	.171
	Negative	-.111
Kolmogorov-Smirnov Z		.903
Asymp. Sig. (2-tailed)		.389

a. Test distribution is Normal.

b. Calculated from data.

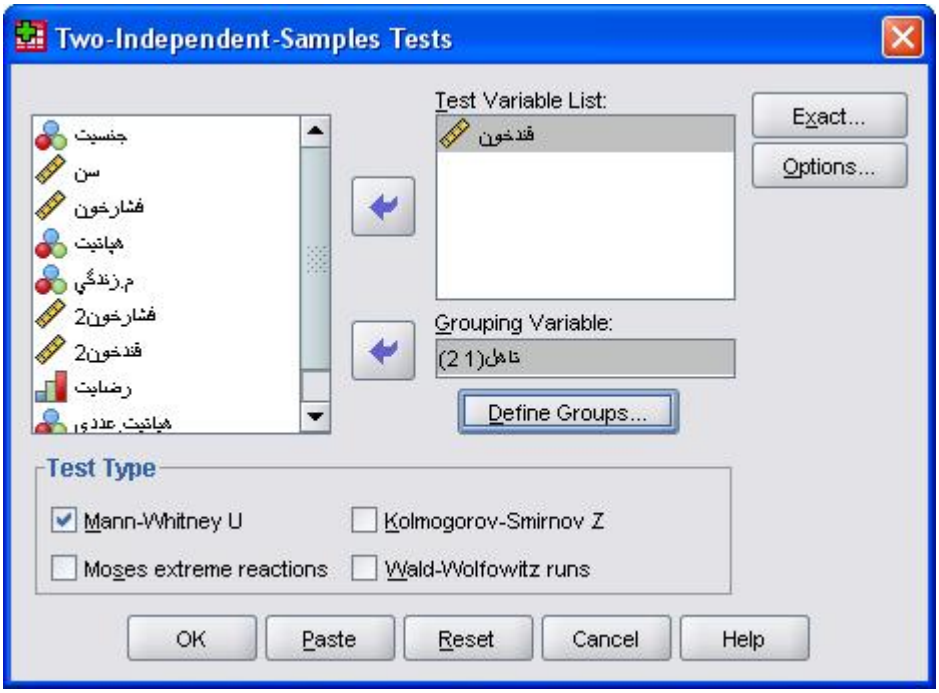
جدول بالا حاوی اطلاعات میانگین، انحراف معیار، آماره z و سطح معنی‌داری آزمون است. همانند قبل معیار رد یا قبول آزمون مقدار سطح معنی‌داری (Asymp.sig) و مقایسه آن با مقدار α (در این تحقیق 0/05 در نظر گرفته شده) می‌باشد.

همانطور که ملاحظه می‌شود sig>α خواهد بود و فرض پیروی فشار خون از توزیع نرمال رد نمی‌شود.

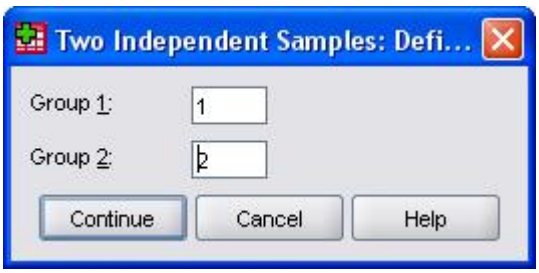
کلیدهای ... Options و ... Exact مانند فرمانهای قبل در فرمان Chi-Square عمل می‌کنند.

12ـ6ـ فرمان ... 2 Independent Samples

با اجرای این فرمان در آزمونهای ناپارامتری امکان مقایسه توزیع‌های دو نمونه مستقل از هم فراهم می‌آید. اگر هر دو نمونه مستقل باشند این فرمان امکان مقایسه میانگین‌های آنها را با یکدیگر فراهم می‌آورد، در حقیقت اجرای این فرمان در آزمونهای ناپارامتری معادل اجرای آزمون T برای مقایسه میانگین‌های دو نمونه مستقل است، هنگامی که توزیع نمونه‌ها نرمال نباشد و یا تعداد داده‌ها کم است باید از این فرمان به جای آزمون T برای مقایسه میانگین‌های دو نمونه مستقل بهره برد، با اجرای این فرمان جعبه زیر ظاهر می‌شود:



همانند آزمون T برای دو نمونه مستقل، متغیر پیوسته را وارد جعبهٔ Test Variable List نموده و متغیر مبنای گروه‌بندی را که وارد قسمتGrouping Variable می‌نمائیم و با فعال شدن کلید Define Groups، کد سطوحی را که می‌خواهیم میانگین متغیر پیوسته آنها را با هم مقایسه نمائیم مانند شکل زیر وارد می‌کنیم:



برای مثال می‌خواهیم میزان قند خون افراد متأهل و مجرد را با استفاده از روشهای ناپارامتری آزمون نمائیم، متغیر پیوسته قند خون را وارد Test Variable List نموده و متغیر دو سطحی وضعیت تأهل را وارد جعبه Grouping Variable می‌نمائیم و با فعال کردن Define Groups ، کد در سطح وضعیت تأهل را در Group 1 و Group 2 وارد نموده و از قسمت Test Type یکی از انواع زیر را انتخاب می‌نمائیم:

ـ آزمون من و تینی (Mann-Whitaey U)

ـ آزمون کولموگروف ـ اسمیرنوف (Kolmogrov-Smirnov Z)

ـ آزمون به روش Moses Exterme Reactions

ـ آزمون به روش wald-Wolfowitz Runs

هر یک از روشهای فوق با استفاده از راههای متفاوتی آزمون فرض برابری توزیع دو نمونه مستقل را آزمون می‌نمایند، با توجه به استفاده بیشتر آزمون کولموگروف ـ اسمیرنوف در ذیل نتایج آزمون زیر برای متغیر پیوسته قند خون در دو سطح مستقل متغیر وضعیت تأهل بیان شده است.

Test Statistics[b]

	قندخون
Mann-Whitney U	98.000
Wilcoxon W	218.000
Z	-.601
Asymp. Sig. (2-tailed)	.548
Exact Sig. [2*(1-tailed Sig.)]	.567[a]

a. Not corrected for ties.

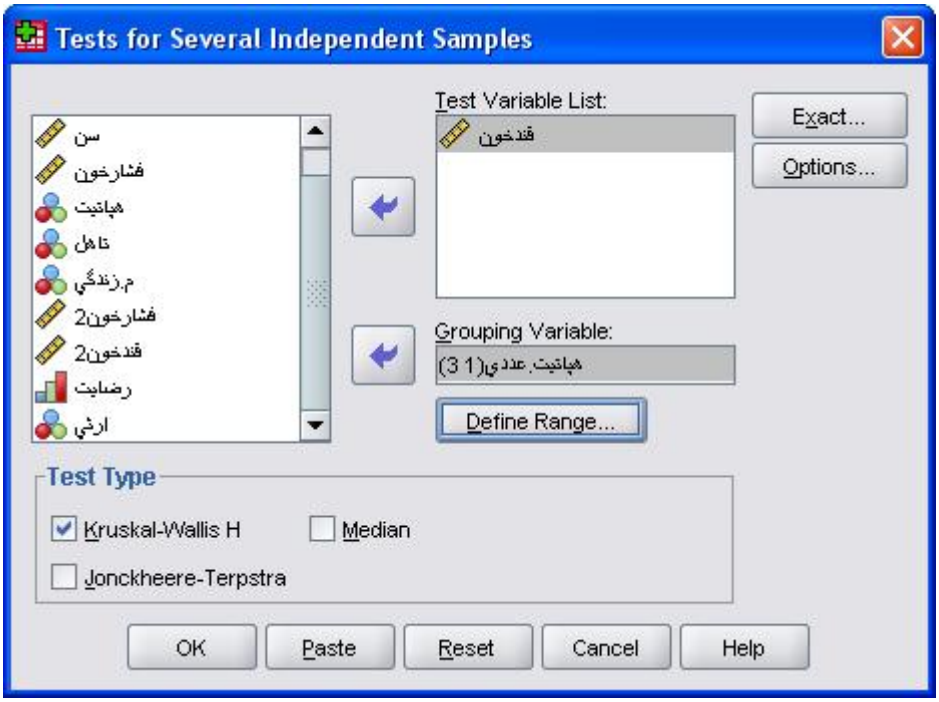
b. Grouping Variable: تاهل

در جدول Test Statistics ابتدا به آمارهٔ کولموگروف ـ اسمیرنوف و سطح معنی‌داری دو طرفه دقت می‌نمائیم با توجه به اینکه sig=0/000 است با در نظر گرفتن هر سطحی از α می‌توان عنوان کرد که میانگین قند خون در دو گروه خانم‌ها و آقایان با یکدیگر متفاوت است یا در اصطلاح، عامل جنسیت بر روی قند خون بیماران مؤثر است، اما این سؤال مطرح می‌شود که قند خون خانمها بالاتر است یا قند خون آقایان؟ برای پاسخ به این سؤال به مقدار اختلاف‌ها (Differences) که دارای دو مقدار مثبت (Positive) و منفی (Negative) است دقت می‌نمائیم اگر |Positive|<|Negative| به این معنی است که $\mu_1 > \mu_2$ است. در مثال بالا با توجه به اینکه |Positive|>|Negative| است پس $\mu_1 < \mu_2$ که میانگین قند خون مردها است از قند خون خانم‌ها بالاتر است.

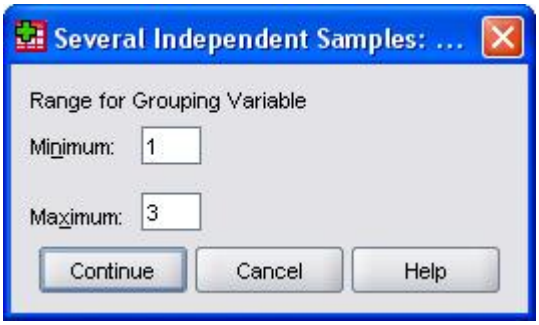
کلیدهای ... Options و ... Exact مانند فرمانهای قبل عمل می‌نمایند.

12ـ7ـ فرمان K Independent Samples

با اجرای این فرمان امکان مقایسه میانگین‌های یک متغیر پیوسته در سطوح مختلف یـک متغیـر گسسـته دارای بیش از دو سطح امکان‌پذیر می‌شود، با اجرای این فرمان نتایجی مانند آنالیز واریانس یک طرفـه در خروجـی ظـاهر می‌شود، در حقیقت این فرمان حالت ناپارامتری آنالیز واریانس یک طرفه است هنگامی که تعداد داده‌ها کم باشـد و یا متغیر پیوسته دارای توزیع نرمال نباشد. با اجرای این فرمان، جعبه گفتگوی زیر ظاهر می‌شود:



متغیر پیوسته مورد آزمون را به جعبهٔ Test Variable List منتقل می‌کنیم و متغیر گسسته چندسطحی را بـه جعبهٔ Grouping Variable وارد می‌نمائیم. با فعال کردن کلید Define Range جعبه گفتگوی زیر ظاهر می‌شود:



همانند آنالیز واریانس یک طرفه باید کد سطح کمترین و کد سطح بیشترین را در قسمت Minimum و Maximum وارد نمائید.

ذکر این نکته ضروری است که در کلیه فرمانهای آزمون‌های ناپارامتری تنها متغیرهای عددی ظاهر می‌شوند برای وارد کردن متغیرهای رشته‌ای (String) باید ابتدا از طریق دستور Auto Recode در منو Transform متغیرها را تبدیل به متغیرهای با کدهای عددی نمود تا بتوان آنها را جعبه‌های مورد نظر وارد نمود.

با اجرای آزمون فوق در حقیقت فرض H_0 زیر در مقابل H_1 آزمون می‌شود:

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

H_1: حداقل یکی از میانگین‌ها با سایر آنها مخالف است

برای اجرای این آزمون در بخش Test Type سه نوع روش زیر بیان شده است:

- Kruskal-Wallis H
- Median
- Jon Ckheere-Terpstra

هر یک از روشهای فوق با روشهای متفاوتی آزمون بالا را انجام می‌دهند. با توجه به کاربرد بیشتر روش کروسکال ـ والیس این روش در مثال زیر توضیح داده شده است.

می‌خواهیم تأثیر سه نوع مختلف هپاتیت را در متغیر گسسته نوع هپاتیت بر روی میزان قند خون مبتلایان آزمون نمائیم در حقیقت آزمون فرض زیر مد نظر است.

$H_0 : \mu_A = \mu_B = \mu_C$

H_1 : حداقل یکی از میانگین‌ها مخالف باشد

برای انجام این آزمون متغیر پیوسته میزان قند خون را وارد جعبهٔ Test Variable List نموده و متغیر گسسته عددی نوع هپاتیت را وارد جعبهٔGrouping Variable نمود و با فعال کردن Define Rane مقادیر 1 و 3 را در Minimum و Maximum که حداقل کدها و حداکثر کدهای نوع هپاتیت هستند وارد می‌نمائیم، از بخش Test

Type نوع آزمون را به روش Kruskal-Wallis H انتخاب نموده و فرمان را اجرا می‌نمائیم تـا خروجـی زیـر ظـاهر شود:

Ranks

	هپاتیت.عددي	N	Mean Rank
قندخون	نوع هپاتیت A	9	9.00
	نوع هپاتیت B	8	17.94
	نوع هپاتیتC	13	18.50
	Total	30	

Test Statistics[a,b]

	قندخون
Chi-Square	7.031
df	2
Asymp. Sig.	.030

a. Kruskal Wallis Test

b. Grouping Variable: هپاتیت.عددي

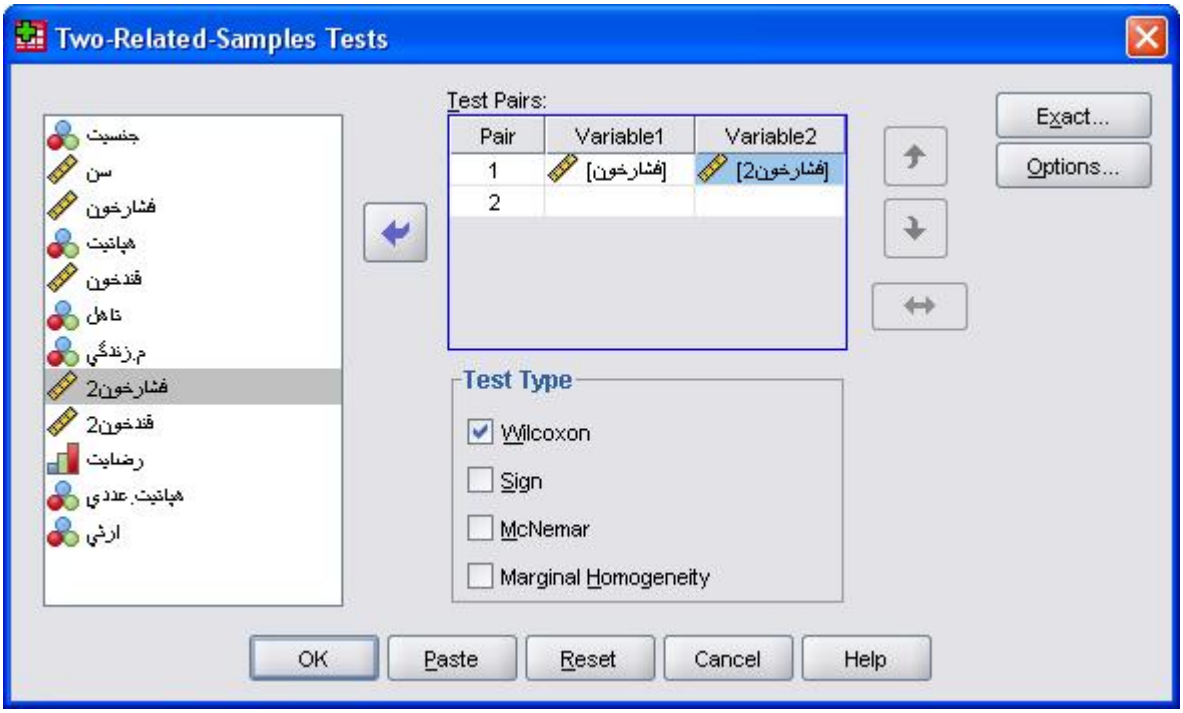
در جدول Ranks، میانگین رتبه‌ای هر یک از سطوح مختلف نوع هپاتیت بدست آمـده اسـت کـه در صـورت رد شدن فرض اولیه می‌تواند مورد استفاده قرار گیرد.

در جدول Test Statistics با استفاده از آماره توزیع کای دو معیاری بـرای پـذیرفتن یـا رد کـردن فـرض اولیـه مطرح شده است. با توجه به این که Asymp sig=0/03 است. با در نظر گرفتن α = 0/05 می‌توان نتیجه گرفت که $sig \leq \alpha$ و در نتیجه فرض H_0 رد می‌شود یعنی میانگین قند خون افراد مبتلا به انواع هپاتیـت بـا یکـدیگر متفـاوت است. برای مقایسه دو به دو میانگین قند خون در انواع مختلف هپاتیت بهتر و دقیـق‌تـر از روش 2 Independent Samples استفاده می‌نمائیم اما می‌توان با اطلاعات میانگین رتبه‌ها در جدول Rank میانگین‌ها را رتبه‌بندی نمود و با استفاده از نتایج این جدول $\mu_C > \mu_B > \mu_A$خواهد بود.

کلیدهای ... Options و ... Exact مانند فرمان Chi-Square عمل می‌کنند.

12ـ8ـ فرمان 2 Related Samples

حالت ناپارامتری آزمون t برای دو نمونه وابسته جفت شده در حالتی که توزیع نمونه‌ها نرمال نیست یا داده‌ها تعداد کمی هستند از طریق فرمان 2 Related Samples انجام می‌پذیرد. با انتخاب این فرمان می‌توان توزیع و پارامترهای دو متغیر وابسته را با هم مقایسه نمود. پس از انتخاب این فرمان جعبه گفتگوی زیر ظاهر می‌شود:



همانند آزمون t برای دو نمونه وابسته، دو متغیر پیوسته زوج شده را از قسمت سمت چپ انتخاب نموده و به جعبهٔ Test Pair(s) List وارد می‌نمائیم. این آزمون نیز در بخش Test Type با استفاده از روشهای زیر قابل انجام است:

- روش Wilcoxon
- روش Sign
- روش Mc Nemar
- روش Marginal Homogenity

با توجه به کاربرد بیشتر روش ویلکاکسون در مثال زیر با این روش آزمون صورت گرفته است.

در مثالی بیماران هپاتیتی ابتدا قند خون بیماران ثبت شده است و بعد داروی خاصی برای آنها تجویز شده است. پس از مصرف دارو دوباره قند خون بیماران اندازه‌گیری و در متغیر قند خون 2 ثبت شده است. می‌خواهیم تفاوت میان قند خون بیماران قبل از مصرف دارد و میزان قند خون آنها بعد از مصرف دارو را آزمون نمائیم در حقیقت با انجام این آزمون تأثیر دارو را بر روی قند خون بیماران مبتلا به هپاتیت آزمون می‌نمائیم که می‌تواند بی‌تأثیر باشد یا باعث افزایش یا کاهش قند خون مبتلایان گردد. برای انجام این آزمون متغیر قند خون قبل و بعد را با هم وارد پنجرهٔ Test Pair(s) Test می‌نمائیم و از بخش Test Type نوع Wilcoxon را انتخاب نموده و فرمان را اجرا می‌نمائیم، خروجی زیر ظاهر می‌شود:

Ranks

		N	Mean Rank	Sum of Ranks
میزان قندخون بعد از مصرف دارو - میزان قند خون بیماران مورد بررسي	Negative Ranks	28[a]	16.11	451.00
	Positive Ranks	2[b]	7.00	14.00
	Ties	0[c]		
	Total	30		

a. میزان قندخون بعد از مصرف دارو < میزان قند خون بیماران مورد بررسي

b. میزان قندخون بعد از مصرف دارو > میزان قند خون بیماران مورد بررسي

c. میزان قندخون بعد از مصرف دارو = میزان قند خون بیماران مورد بررسي

Test Statistics

	میزان قندخون بعد از مصرف دارو - میزان قند خون بیماران مورد بررسي
Z	-4.494
Asymp. Sig. (2-tailed)	.000

ابتدا در جدول Test Statistics و با استفاده از سطح معنی‌داری آماره z که برابر sig=0/000 است می‌توان نتیجه گرفت که فرض H_0 یعنی برابری میانگین قند خون با قبل و بعد از مصرف دارو رد می‌شود حال این سؤال مطرح است که قند خون بعد از مصرف دارو کاهش یافته یا افزایش یافته است؟ برای یافتن پاسخ این سؤال به نتایج جدول Ranks مراجعه می‌نمائیم. با توجه به اینکه میانگین رتبه‌های منفی (16/11) بیشتر از میانگین رتبه‌های مثبت (7) است، می‌توان با توضیحات انتهای جدول در مورد تعبیر مثبت یا منفی بودن رتبه‌ها دریافت که میزان قند

خون بیماران بعد از مصرف دارو کاهش می‌یابد. زیرا negative Ranks بالاتر است و به تـأقیر آن در پـائین جـدول توجه می‌نمائیم که بیان می‌نماید.

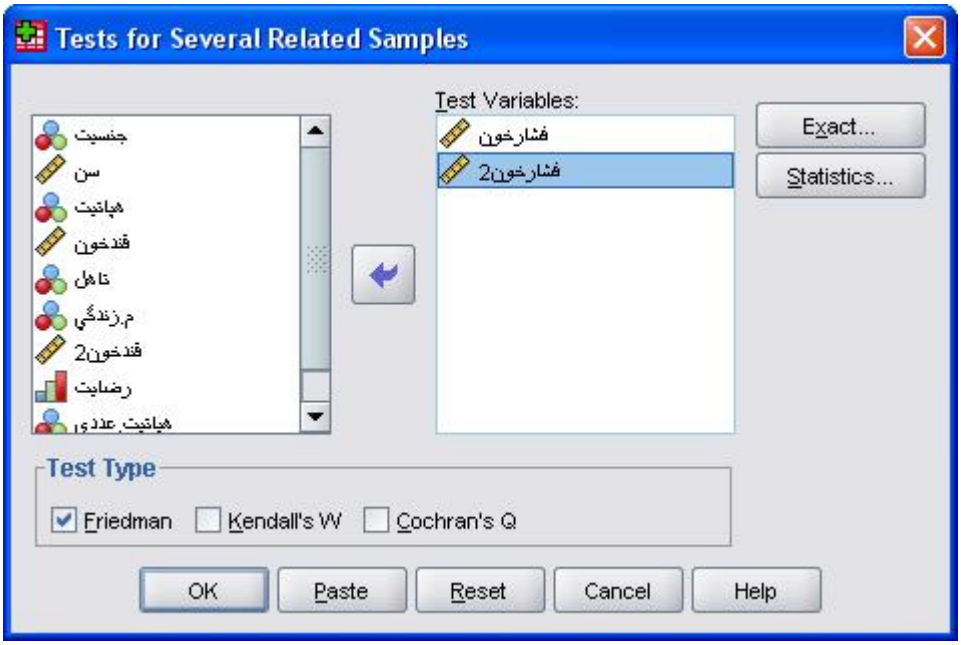
میزان قند خون بعد از مصرف دارو < میزان قند خون بیماران مورد بررسی

که نشان دهنده تأثیر دارو در کاهش قند خون افراد مبتلا خواهد بود.

دکمه‌های ... Options و ...Exact همانند فرمان Chi-Square عمل می‌نمایند.

12ـ9ـ فرمان K Related Samples

با انتخاب فرمان ... K Related Samples می‌توان توزیع (میانگین) دو یا چند متغیر وابسته را با هـم مقایسـه کرد. با انتخاب این فرمان جعبه گفتگوی زیر ظاهر می‌شود:



متغیرهای وابسته مورد آزمون را به جعبهٔ Test Variables منتقل می‌کنیم. روش‌های اجرای آزمـون در بخـش Test Type به صورت زیر است:

ـ روش Fridman

- روش Kendall's W

- روش Cochran's Q

با توجه به کاربرد بیشتر روش فریدمن، در مثال زیر از این روش استفاده شده است.

می‌خواهیم میزان فشار خون بیماران را قبل از مصرف دارو با میزان فشار خون آنها را بعد از مصرف دارو با یکدیگر مقایسه نمائیم. (البته این آزمون برای بیش از دو متغیر وابسته نیز کاربرد دارد) این دو متغیر را وارد جعبه‌ی Test Variables نموده و از بخش Test Type گزینه روش Fridman را انتخاب نموده و فرمان را اجرا می‌نمائیم. خروجی زیر ظاهر می‌شود:

Ranks

	Mean Rank
میزان فشار خون بیماران مورد بررسی	1.73
میزان فشار خون بعد از مصرف دارو	1.27

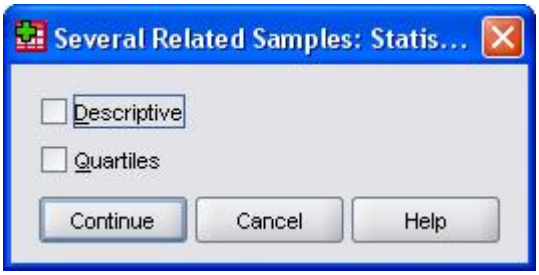
Test Statistics[a]

N	28
Chi-Square	11.267
df	1
Asymp. Sig.	.001

a. Friedman Test

ابتدا با توجه به آماره کای دو و درجه آزادی آن و مقدار سطح معنی‌داری (sig=0/001) و با در نظر گرفتن α=0/05 فرض برابری میانگین دو متغیر رد می‌شود، همانند حالت 2 Related Sample در صورت رد شدن فرض اولیه، می‌توان با استفاده از میانگین رتبه هر متغیر نتیجه را بدست آورد. با توجه به اینکه میانگین رتبه فشار خون قبل از مصرف دارو بالاتر است به این نتیجه می‌رسیم که مصرف دارو باعث کاهش رتبه میانگین و در نتیجه میانگین فشار خون افراد می‌شود.

با فعال کردن کلید ... Statistics پنجرهٔ زیر باز می‌شود.



با انتخاب گزینهٔ Descriptive و خلاصه‌های آمار توصیفی نظیر میانگین، انحراف معیار، کمترین مقدار و بیشترین مقدار برای هر یک از متغیرها در خروجی ظاهر می‌شود و با انتخاب Quartiles اطلاعات چارک‌های اول، چارک دوم (میانه) و چارک سوم برای هر یک از متغیرها در خروجی ظاهر می‌شود.

کلید ... Exact نیز مانند فرمان Chi-Square عمل می‌کند.

تمرین:

1- آیا می توان ادعا نمود که نرخ ابتلا به انواع هپاتیت یکسان است؟

2- آیا می توان ادعا نمود که سهم هر یک از مناطق مختلف خودرو ساز در تامین خودرو یکسان است؟

3- آیا می توان ادعا نمود که جنسیت افراد در ابتلا به بیماری هپاتی موثر است؟

4- آیا می توان ادعا نمود که ارثی بودن و نبودن در ابتلا به بیماری هپاتیت موثر است؟

5- آیا می توان ادعا نمود که متغیر داده های مربوط به بیماران هپاتیتی تصادفی هستند؟

6- آیا می توان ادعا نمود که داد های مربوط به افراد مشغول به کار در شرکت(employe data) تصادفی انتخاب شده اند؟

7- آیا می توان ادعا نمود که میزان درآمد افراد شاغل در شرکت دارای توزیع نرمال است؟

8- آیا می توان ادعا نمود که قند خون و فشار خون بیماران مبتلا به هپاتیت دارای توزیع نرمال است؟

9- هر گاه بخواهیم تاثیر یک متغیر گسسته دو سطحی را بر روی یک متغیر گسسته ترتیبی بدست

آوریم نیز می توان از روش ناپارامتری آزمون میانگین دو نمونه مستقل استفاد نمود. روش من ویتنی و روش کلموگروف اسمیرنوف روش های مناسبی برای آزمون فوق هستند با توجه به این مسئله تاثیر جنسیت را در میزان رضایت مبتلایان به هپاتیت که یک متغیر گسسته ترتیبی(طیف لیکرت) در نظر گرفته شده است. بدست آورید.

10- تاثر ارثی بودن و نبودن بیماری را با روش کلموگروف اسمیرنوف بر روی میزان رضایت از زندگی بدست آورید.

11- هر گاه بخواهیم تاثیر یک متغیر گسسته بیش از دو سطح را بر روی یک متغیر گسسته ترتیبی بدست آوریم نیز می توان از روش ناپارامتری آزمون میانگین چند نمونه مستقل استفاد نمود. روش کروسکال والیس و روش میانه روش های مناسبی برای آزمون فوق هستند با توجه به این مسئله تاثیر نوع هپاتیت را در میزان رضایت مبتلایان به هپاتیت که یک متغیر گسسته ترتیبی(طیف لیکرت) در نظر گرفته شده است. بدست آورید.

12- از روش کلموگروف اسیرنوف و روش کروسکال والیس در روشهای تحلیل پرسشنامه ها که سوالات عموما به صورت طیف لیکرت می باشند بسیار استفاده می شود. یک پرسشنامه تهیه و با استفاده از روشهای فوق نتایج بدست آمده را تحلیل نمائید

پیوست 1

نصب و راه اندازی نرم افزار SPSS 16 تحت ویندوز

1-مراحل نصب نرم افزار SPSS 16

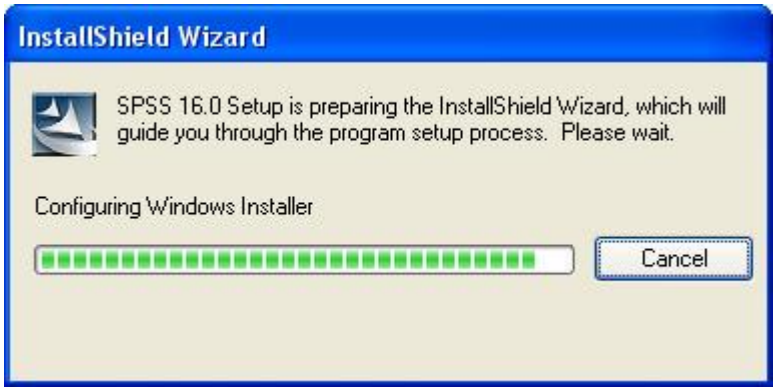
مراحل نصب و راه اندازی نرم افزار spssدر همه نسخه های ویندوز، یکسان و به شرح زیر است.

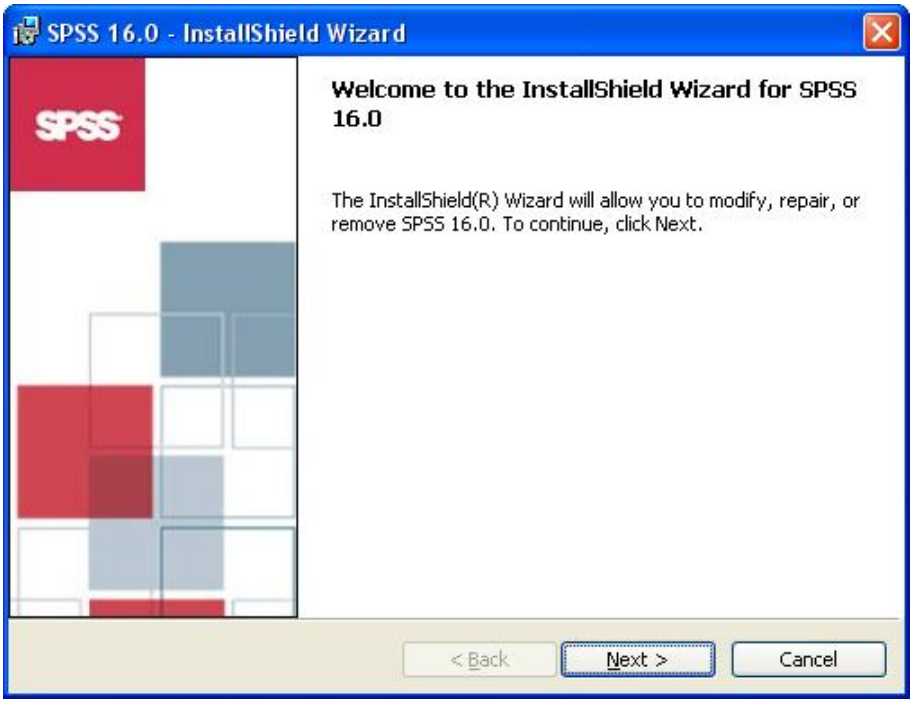
1) CD مربوط به برنامه SPSS 16 را در CD-ROM قرار می‌دهیم.

2) با انجام مرحله 1 در صورت وجود فایل Autorun ،کادر محاوره ای Installshield Wizard خودبخود ظاهر می‌شود. در غیر این صورت باید از مسیر مناسب فایلSetup برنامه SPSS را به صورت زیر بیابید و اجرا کنید.

فایل را اجراء کرد ه تا کادر محاوره ای Installshield Wizard ظاهر شود. در کادر محاوره

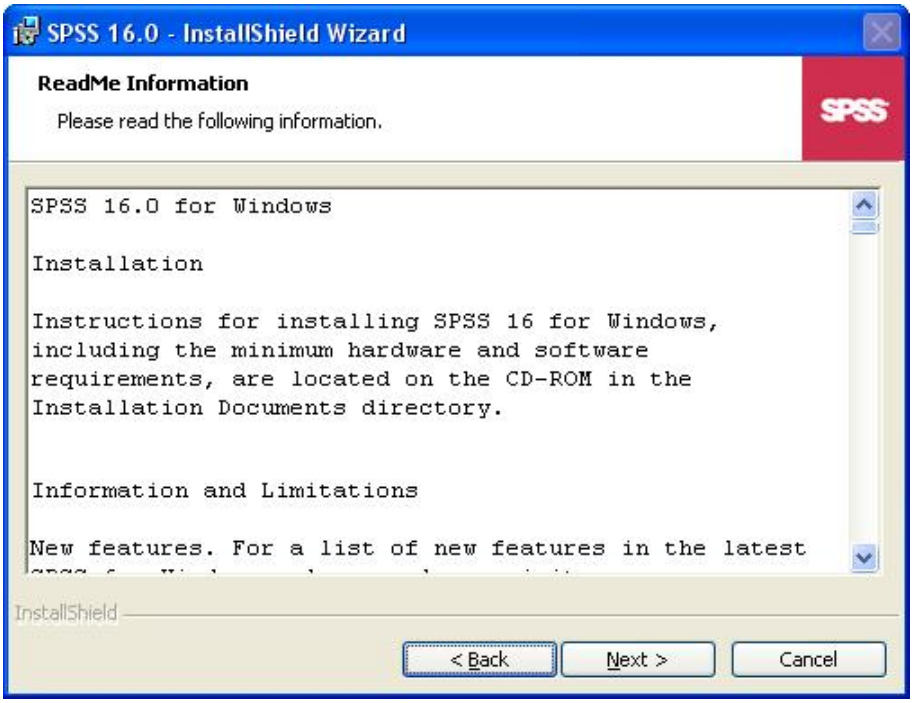
ای Installshield Wizard دکمه Next را کلیک می‌کنیم.



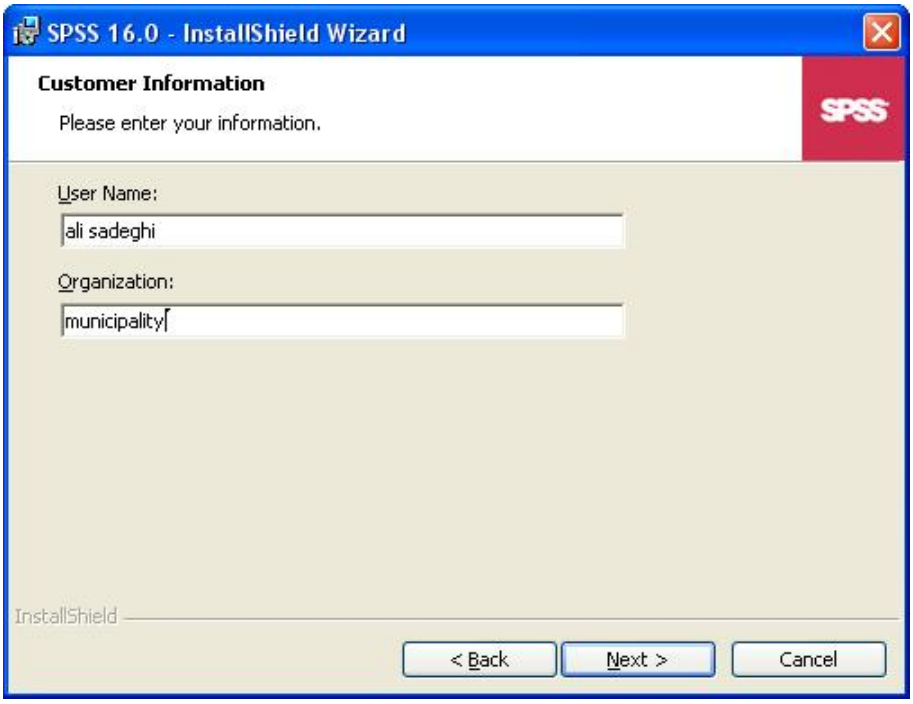


3) با انجام مرحله 2، پنجره دیگری باز شده و توضیحاتی راجع به اجازه نشر برنامه می‌دهد. گزینه I accept the terms in the license agreement را انتخاب و روی دکمه Next کلیک می‌کنیم.

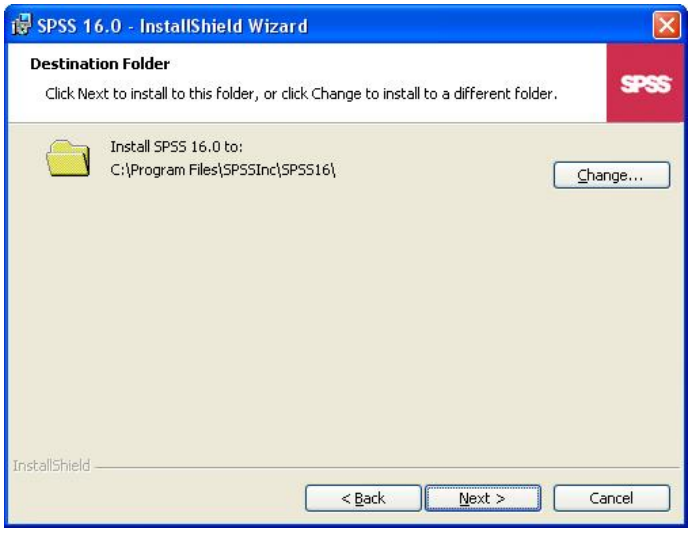
4) با انجام مرحله 3، کادر دیگری باز شده و اطلاعاتی راجع به نرم افزار SPSS را ارائه می‌دهد، در این کادر روی دکمه Next کلیک می‌کنیم.



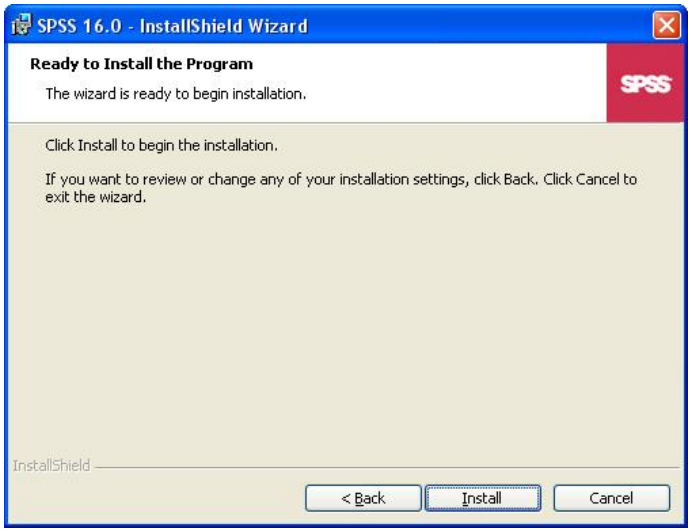
5) با انجام مرحله4، کادر دیگری باز می‌شود. در این کادر، نامی دلخواه در قسمت نام کاربر(User Name) و سازمان(Organization) وارد کرده و روی دکمه Next کلیک می‌کنیم.

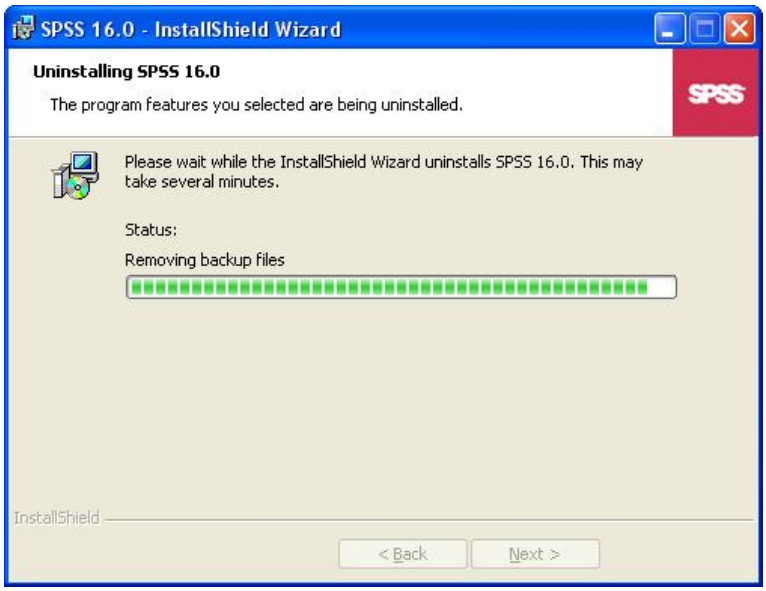


6) با انجام مرحله 5، کادر دیگری باز می‌شود. در این کادر، برای تغییر مسیر مقصد نصب برنامه در محلی غیر از محل پیش گزیده، گزینه Change ... را کلیک نموده و مسیری را که می‌خواهیم برنامه نصب شود انتخاب می‌کنیم. سپس برروی دکمه Next کلیک می‌کنیم. بعد از این مرحله اگر شماره سریال برنامه خواسته شد، باید شماره سریال برنامه را که همراه نرم افزار ارائه می شود، در قسمت License Code وارد کرده و روی دکمه Update کلیک کنیم.

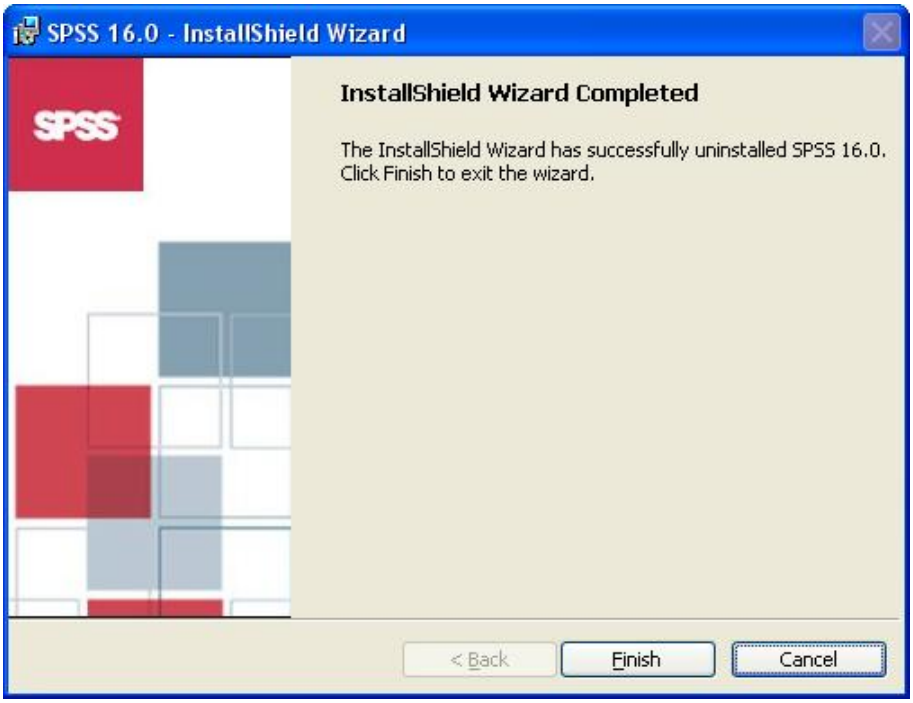


7) با انجام مرحله 6، کادر دیگری باز می‌شود. روی گزینه Install کلیک می کنیم تا نصب نرم افـزار SPSS 14 آغاز گردد.





8) بعد از اجرای مرحله نصب نرم افزار، در کادر محاوره ای زیرگزینه Finish را برای اتمام مراحل نصـب برنامـه کلیک می کنیم. نصب نرم افزار خاتمه یافته و آماده اجرا می گردد.

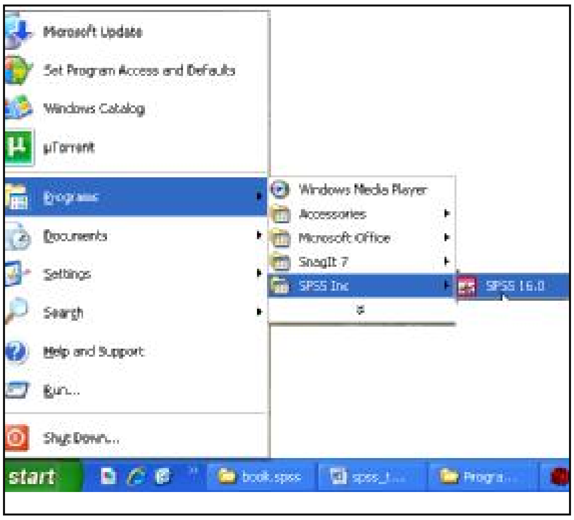


2-شروع کار با SPSS و اعمال مقدماتی

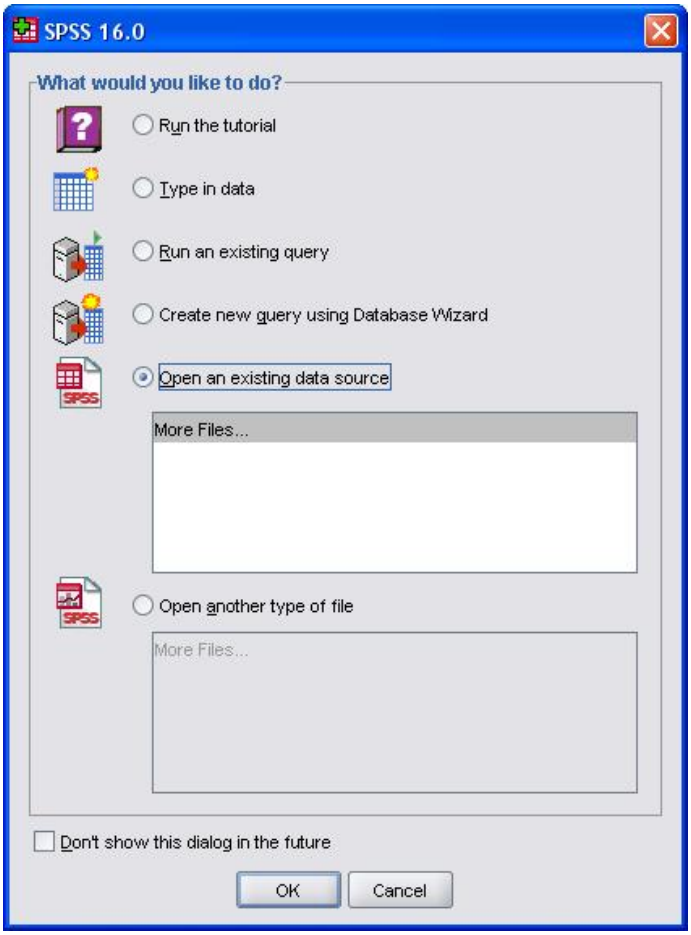
برای ورود به محیط نرم افزار SPSS در صورت وجود Icon نـرم افـزار SPSS در Desktop روی آن Double-

click خـــواهيم كـــرد.

یا از منوی Start برنامه SPSS و آیکون اجرائی آن را فراخوانی می‌کنیم.



با اجرای این دستور محیط SPSS فعال شده وکادر محاوره ای زیر ظاهر می‌شود. در این پنجره، نوع کـاری کـه کاربر از SPSS انتظار دارد سئوال می‌شود



اگر می خواهیم با نحوه کارکردن با نرم افـزار SPSS آشـنا شـویم، گزینـه Run The Tutorial را انتخـاب مـی کنیم. اگر بخواهیم داده‌های جدیـدی را وارد SPSS کنـیم، گزینـه Type In Data را انتخـاب مـی نمـائیم. اگـر می‌خواهیم از فایلهای بانک اطلاعاتی SPSS استفاده کنیم،گزینه

an exiting query Run را انتخاب می نمائیم. اگر بخواهیم از فایلهای بانک اطلاعاتی تهیه شـده در سـایر نـرم

افزارها مانند SAS, dBASE, EXEL, FOXPRO, EpiINFO, Minitab و... استفاده کنیم، ازگزینه Creat query using new database wizard جهت استفاده از آن برای تطابق جهت استفاده و انتقال بانک اطلاعاتی به بخش نمایش داده‌ها (dara veiw) در SPSS استفاده می کنیم. اگر بخواهیم از مجموعه داده های ذخیره شده یا فایلهای داده ای آماده SPSS استفاده کنیم، گزینه Open an exiting data source را انتخاب می کنیم. اگر بخواهیم سایر انواع فایلهای SPSS برون داد[1]را مشاهده و استفاده کنیم، گزینه Open another type file را انتخاب می کنیم.اگر گزینهDon’t show this dialog in the future را در کادر محاوره ای انتخاب کنیم، در اجرای مجدد SPSS و به صورت پیش فرض، یک صفحه داده خالی باز خواهد شد(این کار برای مبتدیان توصیه نمی‌شود).

پردازشگرSPSS همواره وضعیت جاری خود را در کادری در پائین صفحهSPSS Data Editor به اطلاع ما می رساند، توجه کنید که زمانی برنامه به خوبی کار خواهد کرد که عبارت زیر ، SPSS Processor is ready در این کادر نوشته شده باشد.

پیوست 2

داده های پیوست فایل exam

	ارثی	هیاتیت.عددی	رضایت	قندخون۲	فشارخون۲	م.زندگی	تاهل	قندخون	هپاتیت	فشارخون	سن	جنسیت
1	ارثی	نوع هپاتیت A	خیلی زیاد	421.00	11.00	روستا	متاهل	450.00	نوع هپاتیت A	12.00	37.00	مرد
2	غیر ارثی	نوع هپاتیت A	زیاد	221.00	10.00	شهر	متاهل	284.00	نوع هپاتیت A	8.00	42.00	زن
3	ارثی	نوع هپاتیت B	خیلی کم	507.00	11.00	روستا	مجرد	530.00	نوع هپاتیت B	13.00	29.00	مرد
4	غیر ارثی	نوع هپاتیت B	خیلی زیاد	325.00	9.00	روستا	متاهل	347.00	نوع هپاتیت B	9.00	58.00	زن
5	غیر ارثی	نوع هپاتیتC	زیاد	321.00	9.00	روستا	متاهل	405.00	نوع هپاتیتC	9.00	61.00	زن
6	ارثی	نوع هپاتیتC	خیلی زیاد	176.00	12.00	شهر	متاهل	470.00	نوع هپاتیتC	15.00	41.00	مرد
7	غیر ارثی	نوع هپاتیتC	زیاد	146.00	8.00	شهر	متاهل	394.00	نوع هپاتیتC	8.00	39.00	زن
8	غیر ارثی	نوع هپاتیت B	کم	154.00	11.00	شهر	مجرد	322.00	نوع هپاتیت B	11.00	27.00	زن
9	ارثی	نوع هپاتیت A	خیلی زیاد	163.00	12.00	شهر	متاهل	422.00	نوع هپاتیت A	13.00	44.00	مرد
10	غیر ارثی	نوع هپاتیت A	خیلی کم	221.00	10.00	شهر	مجرد	297.00	نوع هپاتیت A	10.00	45.00	زن
11	غیر ارثی	نوع هپاتیت A	کم	264.00	9.00	روستا	مجرد	274.00	نوع هپاتیت A	9.00	19.00	زن
12	ارثی	نوع هپاتیت A	زیاد	225.00	10.00	روستا	متاهل	417.00	نوع هپاتیت A	.	37.00	مرد
13	غیر ارثی	نوع هپاتیتC	خیلی کم	312.00	9.00	روستا	مجرد	392.00	نوع هپاتیتC	9.00	28.00	زن
14	غیر ارثی	نوع هپاتیت B	خیلی کم	200.00	8.00	روستا	مجرد	351.00	نوع هپاتیت B	8.00	24.00	زن
15	ارثی	نوع هپاتیت B	خیلی زیاد	235.00	999.00	شهر	متاهل	594.00	نوع هپاتیت B	999.00	72.00	مرد
16	غیر ارثی	نوع هپاتیتC	کم	421.00	8.00	روستا	مجرد	384.00	نوع هپاتیتC	8.00	12.00	زن
17	ارثی	نوع هپاتیت B	خیلی کم	300.00	10.00	روستا	مجرد	670.00	نوع هپاتیت B	11.00	19.00	مرد
18	غیر ارثی	نوع هپاتیت B	زیاد	241.00	9.00	شهر	متاهل	352.00	نوع هپاتیت B	10.00	45.00	زن
19	غیر ارثی	نوع هپاتیتC	خیلی زیاد	321.00	9.00	روستا	متاهل	379.00	نوع هپاتیتC	9.00	47.00	زن
20	ارثی	نوع هپاتیتC	متوسط	213.00	10.00	شهر	متاهل	722.00	نوع هپاتیتC	10.00	28.00	مرد
21	غیر ارثی	نوع هپاتیتC	خیلی کم	213.00	10.00	شهر	متاهل	341.00	نوع هپاتیتC	11.00	36.00	زن
22	غیر ارثی	نوع هپاتیتC	متوسط	321.00	8.00	شهر	مجرد	369.00	نوع هپاتیتC	8.00	37.00	زن
23	ارثی	نوع هپاتیتC	زیاد	231.00	10.00	روستا	مجرد	627.00	نوع هپاتیتC	11.00	34.00	مرد
24	غیر ارثی	نوع هپاتیت A	خیلی زیاد	184.00	9.00	روستا	مجرد	309.00	نوع هپاتیت A	9.00	35.00	زن
25	ارثی	نوع هپاتیت A	خیلی کم	297.00	11.00	شهر	متاهل	312.00	نوع هپاتیت A	14.00	53.00	مرد
26	ارثی	نوع هپاتیت A	خیلی کم	321.00	11.00	شهر	مجرد	294.00	نوع هپاتیت A	15.00	28.00	زن
27	غیر ارثی	نوع هپاتیتC	کم	364.00	10.00	روستا	مجرد	492.00	نوع هپاتیتC	12.00	29.00	مرد
28	غیر ارثی	نوع هپاتیت B	متوسط	351.00	10.00	روستا	مجرد	627.00	نوع هپاتیت B	11.00	17.00	مرد
29	ارثی	نوع هپاتیتC	کم	372.00	10.00	روستا	مجرد	388.00	نوع هپاتیتC	14.00	24.00	زن
30	غیر ارثی	نوع هپاتیتC	خیلی کم	489.00	11.00	روستا	متاهل	572.00	نوع هپاتیتC	12.00	36.00	مرد

پیوست 3

مبانی انجام پیمایش های میدانی

مقدمه :

در این پیوست سعی گردیده است علاوه بر روشهای تهیه پرسشنامه با زبانی ساده و به صورت مختصر یک نمای ساده اما علمی و آموزنده و اثر بخش در اختیار افراد علاقه مند به روشهای تحقیقات میدانی آموزش داده شود تا با مطالعه آن علاوه بر آشنایی با انواع روشها ی پیمایش ها ی میدانی نسبت به مراحل مختلف هر کدام از این روشها و همچنین مزایا و محدودیتهای احتمالی استفاده از هر کدام به همراه انواع خطاهای آماری و غیر آماری که ممکن است در مراحل مختلف تحقیق انجام پذیرد آشنا شوند و قادر باشند که خود پرسشنامه هایی با دقت آماری قابل قبول و حداقل خطاهای غیر آماری تولید نمایند تا نتایج بدست آمده توسط این پرسشنامه ها قابل دفاع و قابل بسط به جامعه مورد بررسی باشد.

1- دلایل انجام پیمایش ها:

همواره بنا به دلایل مختلفی تحقیقات که در این جا از آن به پیمایش یاد می نماییم شکل می گیرند. به ساده ترین عبارت دلیل اصلی وجود همه پیمایش ها پاسخ گویی به سوالاتی است که در ذهن افراد ی که پیمایش را انجام می دهند رسیده است و برای پاسخ گویی به این مجموعه از سوالات دست بـه انجـام پیمایش ها می زنند.

به طور کلی یک پیمایش با هدف پاسخ گویی بـه سـوالات اصـلی پیمـایش شـکل مـی گیـرد ودارای مراحلی است که تعدادی از مهمترین آنها در زیر بیان شده است:

1 – تعریف اهداف (مفروضات) پیمایش

2 – تشخیص جمعیت و نمونه آماری

3 – تعیین چگونگی جمع آوری پاسخ ها

4 – طراحی پرسشنامه تحقیق

5 – انجام پیمایش مقدماتی

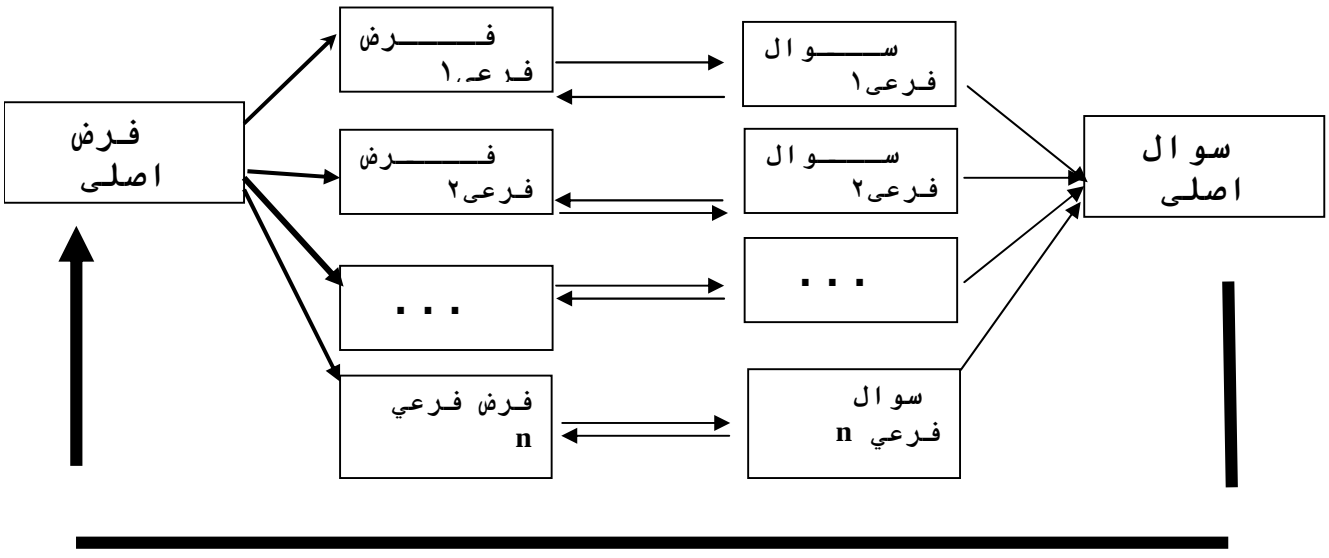
6 – انجام پیمایش اصلی

7 – تحلیل داده ها

رعایت صحت انجام هر یک از موراد بالا از اهمیت ویژه ایی برخوردار است. اگر محقق به عمد یـا بـه طور سهوی هر یک از موارد با لا را با دقت مورد نظر انجام ندهد نخواهد توانست به نتایج به دسـت آمـده اطمینان داشته باشد و یا به عبارت دیگر نخواهد توانست جواب سوالات پیمایش را آنطور کـه انتظـار دارد بدست آورد.

2-تعریف اهداف (مفروضات) پیمایش:

تعریف اهداف و مفروضات پیمایش یکی از مهمترین مراحل انجام یک پیمایش است. محقق باید به صورت کامل و روشن بداند که منظور از انجام پیمایش چیست و این هدف را در صورت کلی بودن به اهداف جزئی تقیسم نماید تا بتوان در راستای هر یک از اهداف جزئی و مفروضات، سوالاتی را مطرح کرد که هدف از انجام تحقیق در حقیقت یافتن پاسخ آن سوالات است و در نتیجه محقق با جمع بندی پاسخ سوالات فرعی به یافتن پاسخ سوال اصلی و قضات در مورد فرض یا فرضیات اصلی دست می یابد.



مثال1:در یک پیمایش انجام گرفته در شهرداری تهران هدف شناسایی میزان رضایت شهروندان از خدمات شهری قابل ارائه توسط شهرداری تهران بوده است. این هدف یک هدف کلی است . در صورتی که

نخواهیم آن را به مفروضات یا اهداف کوچکتر تقسیم نماییم چگونه خواهیم توانست میزان رضایت شهروندان را در مورد خدمات شهری بدست آوریم؟

یکی از دانشجویان توصیه می کند که از شهروندان پرسیده شود که چه میزان از خدمات شهری ارائه شده توسط شهرداری رضایت دارید؟ نظر شما چیست؟

آیا مفهوم خدمات شهری برای تمامی شهروندان قابل درک است؟

آیا با پرسش تنها یک سوال می توان میزان رضایت شهروندان را از خدمات شهری استخراج کرد؟

آیا میزان اهمیت هر یک از خدمات قابل ارائه توسط شهرداری تهران برای همه شهروندان یکسان است؟

پس اگر نمی توان با یک سوال بالا به هدف مورد نظر رسید شما چه پیشنهادی دارید؟

برای حل این مشکل ابتدا پیشنهاد می شود که هدف اصلی در اهداف کوچکتری خرد شود اما چگونه؟

در مثال بالا با استفاده از بررسیهای گذشته مشخص شد که منظور از خدمات شهری که شهرداری به شهروندان ارائه می نماید شامل 25 فعالیت عمده است که شهروندان می‌تواند در مورد آنها اظهار نظر نمایند. (ذکر این نکته ضروری است که شهرداری فعالیتهای بسیاری انجام می دهد که شهروندان به دلیل عدم اطلاع از آنها و یا عدم رویت اثر آنها بر زندگی اجتماعی خود قادر به اظهار نظر در مورد آنها نیستنداین گزینه ها نباید از شهروندان مورد سوال قرار گیرد.) در ذیل 25 فعالیت خرد شده، به قرار زیر هستند:

1 – جمع آوری منظم زباله و پسماند های جامد

2 – رسیدگی به کانالها، جویها و کنترل آبهای روان

3 – توسعه و نگهداری فضای سبز شامل بوستانها، جنگل کاری، درختان، فضای سبز بلوارها و ...

4 – کنترل و بهبود وضعیت ترافیک منطقه

5 – تامین ایمنی شهروندان از طریق نصب نرده،پل هوایی، خط کشی عابرپیاده، نصب سرعت گیر و...

6 – برخورد با صنایع آلاینده هوا، صوت و محیطی

7 – رسیدگی به چهره ظاهری منطقه از نظر زیبایی، مبلمان شهری و ...

8 – مقابله با حیوانات مزاحم و موذی (سگ و موش و ...)

9 – وضعیت سطح آسفالت خیابانها و معابر در منطقه

10 – امکانات فرهنگی، هنری، ورزشی و تفریحی اماکن وابسته به شهرداری منطقه (فرهنگسرا، شهرکتاب، خانه فرهنگ، خانه مطبوعات و ...)

11 – نصب تابلوهای راهنما، علائم راهنمایی و رانندگی در خیابانهای مورد نیاز

12- سرعت و جدیت مسئولین شهرداری در رسیدگی به مشکلات شهروندان

13- قاطعیت برخورد با ساخت و ساز یا عدم تخریب غیرمجاز

14 – نظرخواهی و دخالت دادن نظرات شهروندان در تصمیم گیریها

15- ارتباط مداوم ومستمر و آسان مسئولین شهرداری منطقه با مردم

16- برخورد شایسته با مراجعه کنندگان به ناحیه یا شهرداری

17 – وضعیت بازارچه ها و میادین میوه و تره بار عرضه مستقیم محصولات

7 – رسیدگی به وضعیت کف پوش پیاده روها

19 – وضعیت روشنایی خیابانها در هنگام شب

20- هماهنگی با ارگانهای ذیربط در امور عمرانی منطقه

21- رسیدگی به سرویس های بهداشتی در سطح منطقه و نقاط پرجمعیت

22- وضعیت دسترسی به حمل و نقل همگانی (اتوبوس و مترو)

23- حل مشکل متکدیان ودست فروشان در سطح منطقه

24 – وضعیت پارکینگهای عمومی شهرداری در سطح منطقه

25 – وضعیت کیوسک های مطبوعاتی و گل فروشی ها

همانطور که ملاحظه می شود حال باید 25 سوال بالا به جای تنها یک سوال مورد نظر قرار گیرند. ذکر این نکته مبرهن است که هر یک از 25 آیتم بالا وزن مشخص در میزان رضایت شهروندان دارند که باید در تحقیقات مشخص گردد هر چند در بسیاری از موارد برای راحتی کار وزن هر یک از آیتمها و سوالات فرعی را در سوال اصلی مساوی در نظر می گیرند که دارای مقداری خطای غیر آماری خواهد بود.

مثال2: در یک مطالعه بازار محصولات یک شرکت ، هدف شناسایی میزان رضایت مصرف کنندگان از محصول غذایی ارائه شده شرکت برای مصرف کنندگان است، اگر تنها این سوال مطرح شود که میزان رضایت شما به عنوان مصرف کننده محصولات شرکت به چه میزان است؟ آیا می توان به نتایج بدست آمده پیمایش اعتماد نمود؟ چرا؟

بسیاری از افراد معتقد هستند که با توجه به اینکه افراد نظر خود را در مجموع برای همه جنبه های محصول بیان می کنند نتایج قابل استناد خواهد بود. نظر شما چیست؟

با توجه به اینکه برای یک مصرف کننده هنگام تهیه محصول شما موارد زیر مهم است باید میزان رضایت از هر یک از آنها مورد سوال واقع شود:

1- قیمت محصول
2- در دسترس بودن محصول و توزیع مناسب آن
3- شکل ظاهری و بسته بندی محصول
4- رنگ و مواد بکار رفته در تولید محصول
5- طعم و مزه محصول
6- نحوه رفتار متصدیان فروش(در صورت دارا بودن نمایندگی فروش)

7- و...

با توجه به اینکه میزان اهمیت هر یک از موارد بالا نیز با یکدیگر متفاوت است باید میزان اهمیت هر یک از آیتمها در کل میزان رضایت مصرف کنندگان از محصول مشخص گردد تا وزن مخصوص به هر یک از آیتمها به درستی لحاظ شود.

شاید این سوال در ذهن شما ایجاد شده باشد که چگونه از سوال اصلی ، سوالات فرعی را نتیجه می گیریم به طور حتم این امر نیاز به نظر افراد متخصص در موضوع مورد بررسی(برای مثال اول مدیران شهری و برای مثال دوم صاحب نظران در آن شاخه خاص از صنایع غذایی) دارد برای این منظور روشی با نام روش دلفی وجود دارد که هدف آن شکستن فرضیات اصلی تحقیق در فرضیات کوچکتر قابل اندازه گیری می باشد. به طور خلاصه روش کار دلفی به این صورت است که ابتدا جنبه های مختلف موثر بر روی فرضیات اصلی تحقیق را از روشهای مختلفی نظیر روش جمع آوری اطلاعات به صورت کتابخانه ایی یا میدانی(پرسش از شهروندان در مثال 1 و پرسش از مصرف کنندگان محصولات در مثال2) جمع آوری می کند و سپس توسط افراد صاحب نظر و مطلع جمع بندی و به صورت گزینه های اصلی فرضیات (سوالات) در می آیند.

3-تشخیص جمعیت و نمونه آماری:

در بسیاری ازموارد دیده می شود که افرادی که پیمایش را انجام می دهند بدون در نظر گرفتن خصوصیات جامعه مورد بررسی هم از نظر کمی و هم از نظر کیفی اقدام به انجام پیمایش می نمایند در اینگونه موارد نتایج بسیار دور از حقیقت و دور از انتظار ظاهر می شوند زیرا بدون در نظر گرفتن خصوصیات جامعه مورد بررسی نتایج بدست آمده را نمی توان به آن نسبت داد و نمی توان از صحت

و سقم نتایج دفاع کرد.

منظور از شناخت کمی جامعه مورد بررسی شناخت تعداد عناصر جامعه مورد بررسی و انتخاب تعداد نمونه لازم با دقت آماری مناسب است. منظور از دقت نمونه آماری پارامتری است که از آن به عنوان میزان اعتماد به نتایج بدست آمده از نمونه یاد می‌کنند. شاید این سوال در ذهن شما ایجاد شده باشد که چرا اصولا نمونه گیری انجام می گیرد؟ پاسخ به این سوال واضح و روشن است. از آنجا که هزینه بررسی یک جامعه آماری از عناصر به دلیل داشتن تعداد عناصر زیاد یا در دسترس نبودن همه عناصر آن کاری دشوار، وقت گیر و در بعضی موارد غیر قابل انجام در محدوده زمانی مشخصی است پس از یک نمونه که دارای همه خصوصیات جامعه و تنها در اندازه عناصر کوچکتری است استفاده می شود.

اثبات شده است که اگر روشها و اصول نمونه گیری رعایت شوند نتایجی هم دقت روش انجام بررسی جامعه می دهند . به بیان ساده تر یک نمونه مانند یک عکس از یک تابلوی نقاشی است که دارای اندازه کوچکتری است اما دارای تمامی ویژگیهای همات تابلوی نقاشی است. برای اینکه با دقت یک روش نمونه گیری در صورت رعایت همه اصول آن آشنا شویم مثالی می زنیم.

مثال: شرکت گالوپ یک شرکت امریکایی است که در زمینه انجام پیمایش های اجتماعی، اقتصادی و سیاسی در سراسر دنیا فعالیت می نماید. در سال 1996 در انتخابات ریاست جمهوری کشور آمریکا شرکت گالوپ 8 روز قبل از انتخابات با گرفتن یک نمونه 29 هزار نفری از جمعیت 300 میلیونی کشور امریکا اعلام کرد که آقای کلینتون ریاست جمهوری وقت آمریکا با 49,5 درصد آرا پیروز انتخابات خواهد بود و 8 روز بعد کلینتون در انتخابات ریاست جمهوری کشور آمریکا با 49,2 درصد بار دیگر رئیس جمهور آمریکا شد. نظر شما چیست؟ چگونه هر یک نفر در نمونه تقریبا نماینده 10 هزار نفر در جامعه است؟ آیا همه چیز تصادفی بوده؟

در بسیاری از موارد به طور فرضی همواره 10 درصد تعداد عناصر نمونه را به صورت تصادفی

انتخاب می کنند و آن را مبنی تجزیه و تحلیل ها و استنتاج در مورد جامعه قرار مـی دهنـد و یـا در بسیاری از موارد بخصوص در شاخه های مهندسی تعداد 25 داده در هر گروه انتخـاب مـی نماینـد و فارق از تعداد عناصر جامعه مورد بررسی با همان تعداد نمونه در مـورد جامعـه آمـاری عناصـر مـورد بررسی قضاوت می کنند. نظر شما چیست؟ آیا بـرای هـر تعـداد جامعـه بـا هـر مقـدار پراکنـدگی و خصوصیت آماری تعداد 25 داده یا 10 درصد داده ها مناسب به نظر می‌رسد؟

انتخاب تعداد عناصر مورد نظر نمونه تابع شرایط خاص آماری و خصوصیات جامعه مـورد بررسـی است برای مثال اگر بخواهیم یک نمونه را طوری از یک جامعه آماری انتخاب نماییم که این نمونه بـا دقت 95 درصد خصوصیات جامعه مورد بررسی را در خود داشته باشد با زمانی که بخواهیم نمونـه بـا دقت 99 درصد خصوصیات را در خود داشته باشد متفاوت خواهد بود.

فرمول انتخاب نمونه تصادفی از یک جامعه آماری بصورت زیر است:

$$n = \frac{z^{2}_{a/2}\, pqN}{(N-1)\,d^{2} + z^{2}_{a/2}\, pq}$$

که در آن d حداکثر خطای مطلق است کـه در پژوهشـهای اجتمـاعی و صـنعتی معمـولاً 0,07 درنظر گرفته می شود، N تعداد عناصر جامعهٔ مورد بررسـی، $Z_{a/2}$ مقـدار توزیـع نرمـال در سـطح خطای آماری a است که اگر a برابر 0,05 در نظر گرفته شـود مقـدار $Z_{a/2}$ برابـر 1,96 و اگـر aبرابر 0,1 در نظر گرفته شود مقدار $Z_{a/2}$ برابر 1,64 و اگر a برابـر 0,01 در نظـر گرفتـه شـود مقدار $Z_{a/2}$ برابر 2,3 منظور می شود.

در تحقیقات اجتماعی و بازاریابی معمولا مقدار خطای نوع اول پیمایش برابـر 0,05، در تحقیقـات

صنعتی برابر 0,1 و در تحقیقات پزشکی یا تحقیقات خیلی حساس برابر 0,01 در نظر گرفته می شودو ذکر این نکته ضروری است که با کاهش مقدار خطای آماری حجم نمونه افزایش می یابد.

مقدار p در احتمال حضور هر عضو جامعه آماری در نمونه است و q احتمال عدم حضور یک عضو جامعه در نمونه است با توجه به اینکه pq از قبل مشخص نیست مقدار ماکزیمم این مقدار را که برابر 0,25(p برابر 0,5 و q برابر 0,5 با استفاده از بسط تیلور) است در نظر می گیریم . N تعداد عناصر جامعه مورد بررسی را نشان می دهد. و n تعداد نمونه آماری مورد نیاز را برای داشتن دقت a 100- درصد بیان می کند.

برای مثال تعداد نمونه آماری لازم برای داشتن دقت 95 درصد آماری را از یک جامعه با 270 هزار عنصر مشخص نمایید؟

$$n = \frac{(1.96)^2_* 05 \ _* 0.5 \ _* 2700000}{(269999) \ _* (0.8)^2 + (1.96)^2_* 05 \ _* 0.5}$$

حال با توجه به مقادیر بالا مقدار تعداد نمونه برابر 196 نمونه خواهد بود. ذکر این نکته ضروری است که این تعداد نمونه از روش بالا با استفاده از روشی به نام روش نمونه گیری تصادفی ساده بدست آمده است حال اگر عوامل دیگری در جامعه وجود داشته باشند که بر روی تعداد نمونه تاثیر بگذارند باید از روشهای نمونه گیری که در زیر نمونه ایی از آنها عنوان می شود استفاده نمود.

مثال: در یک جامعه مورد بررسی هدف میزان رضایت مصرف کنندگان پوشاک یک شرکت است با توجه به اینکه جنسیت افراد بر روی نظراتشان موثر است و با توجه به اینکه 53 درصد جامعه مورد بررسی را آقایان تشکیل می دهند پس 53 درصد نمونه را هم باید آقایان تشکیل دهند با توجه به اینکه 60 درصد آقایان زیر 35 سال هستند پس باید 60 درصد نمونه از آقایان هم زیر 35 سال باشد و اگر عامل دیگری نظیر منطقه سکونت و ... نیز بر روی نظرات افراد موثر است باید وزن مطابق جامعه آن در نمونه لحاظ گردد.

شما چند مثال فرضی بزنیدو تعداد نمونه مورد نظر را بدست آورید.

4-تعیین چگونگی جمع آوری پاسخ ها:

تعیین جمع آوری پاسخ ها در انجام یک پیمایش از اهمیت ویـژه ایـی برخـوردار اسـت. اتخـاذ روش نامناسب جمع آوری پاسخ ها در بعضی موراد منجر به نتایج غیر قابل اعتماد و در نتیجـه بـه هـدر رفـتن وقت و انرژی و هزینه های صرف شده برای انجام پیمایش خواهد شد. در زیر تعدادی از روشهای معروف جمع آوری اطلاعات را بیان نموده و به اختصار در مورد هر یک و زمان استفاده از هر کدام توضیحی بیان می شود.

1- روش پستی
2- روش خود اظهاری
3- روش تلفنی
4- روش حضوری
5- روش کتابخانه ایی
6- روش پست الکترونیکی
7- روش تحت وب

4-1- روش پستی:

این روش زمانی استفاده می شود که حجم نمونه بسیار زیاد بوده و افراد مـورد بررسـی کـاملا در مقابل پاسخ به سوالات شما ذینفع باشند و خود را مقید به پاسخ گویی به سوالات بدانند. در بسـیاری از کشورها که فرهنگ اینگونه از جمع آوری اطلاعات برای مردم ایجاد شده است بسیاری از شـرکتها برای اطلاع از ویژگیهای محصول مورد نظر مصرف کنندگان و یا اطلاع از میزان رضـایت آنهـا از ایـن روش استفاده می نمایند و در عوض از ابزارهای مشوق برای ذینفع کردن پاسخ دهندگان به سـوالات

استفاده می‌نمایند. ذکر این نکته ضروری است که اهرمهای تشویقی نباید چنان باشند که در نظـرات پاسخ دهندگان تاثیر گذار باشند و یک نکته بسیار مهم دیگر واضح و شفاف بودن سوالات است بگونه ایی که هر سوال در ذهن همه پاسخ دهندگان یک جنبه را تداعی کند و در مواردی که شما از واضح و شفاف بودن یک سوال اعتماد ندارید می‌توانید توضیحی را در ابتدای سوال ضمیمه کنید تـا پاسـخ دهنده بعد از مطالعه آن به سوالات پاسخ دهد. نکته دیگـری کـه بایـد مـد نظـر قـرار دهیـد نحـوه برگرداندن پاسخ نامه است در اینگونه موارد یک پاکت که دارای بسته بندی شکیل و در صـورت نیـاز دارای پارامترهای تشویقی باشد و همچنین دارای پاکت برای بازگرداندن پاسخ نامه ها از طریق پست باشد می تواند مور د استفاده قرار گیرد، هر چند بررسی بر روی 27 پیمایش کـه بـا اسـتفاده از ایـن روش در طول سالهای 84-81 در شهر تهران انجام گرفته نشان دهنده نرخ بازگشت به طور متوسط 37 درصد و بین 7 تا 54 درصد بوده است.

باید دقت شود چنانکه با سوالاتی سرو کار دارید که جزء مسائل زندگی روزمره زندگی شـهروندان نمی باشد از جمع آوری اطلاعات از این طریق خود داری نمائید. زیرا این روش برای سوالاتی که برای همه شهروندان واضح و عینی است و شهروندان در مورد آنها نظـر دارنـد مناسـب اسـت. از پرسـیدن سوالاتی که احتیاج به تحلیل و سپس پاسخ گویی دارند در این روش باید خود داری شود زیرا شرایط پاسخ گویان برای کنترل خطاهای غیر آماری تحت کنترل نخواهد بود و نمی توان به جوابهای تحلیلی این افراد اعتماد نمود.

2-3- روش خود اظهاری:

در این روش پرسش نامه به شهروندان یا عناصر مورد بررسی داده شده و از آنها خواسته می شـود ضمن مطالعه سوالات به آنها پاسخ دهند و در صورتی که سوالی دارند و یا ابهامی برای آنها در پاسـخ گویی پیش آمده همان موقع از پرسشگر بپرسند.

این روش زمانی که عناصر شما در گروههای قابل دسترسی حضور دارند می تواند باعث کاهش وقت و زمان گردد.

برای مثال به یک گروه که در محل جمع شده اند مراجعه می کنید و یا اگر عناصر مورد بررسی خانمها هستند در محلی که همه دسته‌های خانمها ی مورد نظر در جامعه شما در آن حضور دارند می توانید مکانی برای توزیع پرسشنامه های خود اظهاری باشد.

دقت کنید که پرسشنامه خود اظهاری باید در شرایط مکانی و زمانی مناسب در اختیار افراد قرار گیرد، در خیابانها، در مواقع صبح که همه افراد برای رفتن به محل کار و ... عجله دارند، در هنگام تاریکی هوا، در زمان نهار، در اتوبوس، در صف و ... مکانهای مناسبی برای توزیع و جمع آوری پرسشنامه های خود اظهاری نیست زیرا باید همه عوامل موثر بر نظرات افراد تحت کنترل و ثابت باشند . توصیه می شود دراینگونه موارد اقلام مورد نیاز مانند خودکار و در صورت نیاز زیر دستی در اختیار افراد قرار گیرد تا افراد برای پاسخ گویی به سوالات شما دچار مشکل نگردند. نکته مهم دیگر این است که نباید پرسشنامه ها به صورت مشورتی پاسخ داده شوند و از افراد بخواهید به صورت منحصر نظرات خود را در پرسشنامه قید فرمایند. نرخ بازگشت پرسشنامه در این روش معمولا بالای 90 درصد است و تنها مشکلی که استفاده از این روش دارد اطمینان از نمونه تصادفی متناسب با جامعه مورد بررسی است که نتایج را دچار خطاهای آماری کرده و تحلیل را دچار انحراف می نماید.

بسیاری از پرسشگران تمایل زیادی به استفاده از این روش دارند زیرا در زمان کمتری تعداد پرسشنامه زیادی را جمع آوری می نمایند اما در بسیاری از موارد این امر منجر به هم خوردن نسبت خصوصیات موثر در نمونه مطابق جامعه خواهد شد.

3-4- روش تلفنی:

در این روش که معمولا برای پیمایش های با سوالات محدود مورد استفاده قرار می گیرد با تهیه یک

لسیت کامل از تلفن خانوارهای جامعه مورد بررسی و انتخاب یک نمونه مورد بررسی از آنها سوالات پرسیده می شود. ذکر این نکته مهم است که در اینگونه بررسی دیگر عناصر جامعه مورد بررسی افراد نیستند بلکه برای نمونه خانوارها هستند زیرا شما نمی دانید که چه شخصی تلفن را بر می دارد و اگر هم برای مثال پدر خانواده را برای صحبت فرا می خوانید باز از طرف خانواده و با در نظر گرفتن تجمیع نظرا ت خانواده به سوالات شما پاسخ می دهد. در صورتی که از قبل این مسئله در نظر گرفته نشده باشد دچار انحراف در تحلیل ها خواهیم شد.

یکی دیگر از مسائلی که باید مد نظر قرار گیرد این است که حداکثر زمان مناسب این بررسی ها حدود 3 دقیقه است و گرفتن زمان بیشتر از افراد باعث کاهش دقت پاسخ گویی به سوالات خواهد شد. این روش جمع آوری اطلاعات در بسیاری از کشورها بخصوص در پیمایش های سیاسی و در تحقیقات بازاریابی برای کالاهای شناخته شده بسیار مورد استفاده قرار می گیرد اما در کشور ما بنا به دلایل مختلف فرهنگی به دلیل وجود خطاهای غیر آماری توصیه نمی شود. تماس با تلفن ثابت و همراه اشخاص در صورتی که خود شماره تلفن را در اختیار قرار داده باشند و در پاسخگویی به سوالات شما ذینفع بدانند مورد استفاده قرار ی گیرد. خدمات پس از فروش کالاها از جمله موارد فوق می باشند.

4-4-روش حضوری:

این روش یکی از موثر ترین روشهای جمع آوری اطلاعات در پیمایشها است در این روش مشکلاتی نظیر نامفهوم بودن سوالات و عدم تفهیم پاسخ دهندگان، عدم بازگشت و ابطال پاسخ نامه ها وخطاهای غیر آماری موثر در انتخاب گزینه ها و ... به حداقل می رسد و پرسشگر با توان و قدرت خود که ناشی از آموزش نحوه برخورد با چنین مشکلاتی است سعی کنند آنها را کنترل نماید.

در این روش که نیاز به انتخاب و آموزش پرسشگران با توجه به موضوع مورد بحث دارد، پرسشگران در حقیقت کلید انجام صحیح پیمایش محسوب می شوند و لازم است تا ضمن آموزش نحوه مواجهه با

کنترل خطاهای آماری و غیر آماری تحقیق ، کار آنها با ابزارهای کنترلی نمونه گیری مجدد با حجم محدود، سوالات کنترلی و یا بازرسی نامحسوس و ... کنترل شود.

انتخاب و آموزش پرسشگران زبده رمز موفقیت اینگونه پیمایش ها می باشد به همین دلیل توصیه می شود در انتخاب پرسشگران علاوه بر در نظر گرفتن خصوصیات عمومی مانند جنسیت، سن، قابلیتهای فردی و ... مورد نیاز به عواملی نظیر لهجه، شکل ظاهری، مسئولیت پذیری، قدرت تصمیم گیری آنی، مدیریت، قدرت کار گروهی و... نیز توجه شود. توصیه می شود پرسشگران را از نظر حق الزحمه کاملا راضی نمایید زیرا در صورتی که این مشکل وجود داشته باشد پرسشگران با علاقه مندی کم و با سرعت زیاد و در نظر نگرفتن توصیه های انجام شده به جمع آوری اطلاعت مورد نظر می‌پردازند.

4-5-روش کتابخانه ایی:

در بسیاری از پیمایش ها عناصر مورد بررسی را افراد تشکیل نمی دهند و یا اطلاعات از عناصر مورد بررسی یکجا جمع است که باید از روی آنها اطلاعات را در فرمهای ثبت اطلاعاتی مانند پرسشنامه، ثبت نمود.

برای مثال دولت می خواهد اطلاعاتی در مورد کارمندان خود نظیر سابقه، تحصیلات ، سمت و ... کسب نماید با توجه به اینکه این اطلاعات تماما" در اختیار امور اداری هر وزارت خانه وجود دارد تنها نیاز به مراجعه و ثبت این اطلاعات در فرمهای مربوطه است و نیازی به پرسش از کارمندان وجود ندارد.

4-6-روش پست الکترونیکی:

این روش که جزء یکی از جدید ترین روشهای جمع آوری اطلاعات از عناصر مورد بررسی است در کشورهای پیشرفته به دلیل سرعت بالای جمع آوری اطلاعات و کاهش هزینه های پیمایش و از همه مهمتر کاهش خطا های غیر آماری بسیار مورد توجه قرار گرفته است.

در این روش پرسشنامه به همراه اطلاعات جانبی برای پاسخ گویی صحیح به آدرس الکترونیکی شخص ارسال می شود و شخص بعد از پاسخ گویی به سوالات ، پاسخ نامه را بـه آدرس اعـلام شـده ارسـال مـی نماید. در این روش ذکر چند نکته ضروری است:

1- این روش برای بررسی جامعه آماری مفید و موثر است که ضریب استفاده از اینترنت در آنها بالا باشد.

2- قبل از ارسال پرسش نامه به پست الکترونیکی افراد حتما باید آنها در جریان قرار داده شده باشند و رضایت آنها جلب شده باشد . توجه نمایید که جمع آوری اطلاعات از طریق پست الکترونیکی با تبلیغات از طریق پست الکترونیکی متفاوت است.

3- باید کاملا مشخص باشد که برای چه شخصی پرسشنامه از طریق پست الکترونیکی ارسال شده است . معمولا شرکت ها دارای یک بانک اطلاعاتی از مصرف کنندگان محصولات و خدمات خود هستند که از قبل این افراد رضایت خود را در همکاری با شرکت اعلام نموده اند و اطلاعات این افراد در یک بانک اطلاعاتی وجود دارد و تعداد مورد نیاز این افراد از بانک اطلاعاتی به صورت تصادفی انتخاب شده و پرسشنامه به پست الکترونیک آنها ارسال می شود.

4- سیستم پرداخت پاداش باید برای تمایل به همکاری های بعدی دیده شده باشد. معمولا از طریق پرداخت الکترونیکی به تمامی افرادی که با شما همکاری می نمایند هدیه ایی ارسال می نمایید و یا تسهیلاتی برای این افراد پیش بینی می شود تا آنها در پیمایش های آتی نیز تمایل به همکاری با شما داشته باشند

5- توجه نمائید که سیستم هدیه و پاداش و یا پرداخت اینترنتی به هیچ عنوان نباید باعث تاثیر پذیری اشخاص و تاثیر در پاسخ های قابل ارائه آنها گردد.

6- شکل ظاهری سوالات در پست الکترونیک و همچنین نحوه ایجاد سوالات، تعداد سوالات و چیدمان سوالات باید به گونه ایی باشد که باعث تاثیر منفی در رقبت شخص به ادامه پاسخگویی نگردد.

7- از پرسش سوالات باز که نیاز به تایپ مطلب برای پاسخ گویی دارد اجتناب کنید.

8- از پرسش سوالات خصوصی نظیر سن، تحصیلات، منطقه زندگی و ... خود داری نمایید همانطور که گفته شد از قبل باید بدانید که پرسشنامه را برای چه فردی ارسال می کنید.

9- در صورتی که ناگزیر به پرسش سوالات خصوصی از افراد هستید بهتر است در انتهای پرسشنامه و به صورت اختیاری مورد پرسش واقع شوند.

10- نرخ بازگشت پاسخ نامه در این روش بسیار متغیر است و بستگی به عوامل متعددی نظیر زمان ارسال پرسشنامه ها، تقویم رسمی کشور، نوع سوالات، تعداد سوالات، ابزارهای تشویقی، سطح فرهنگی ، سرعت اینترنت و ... دارد.

7-4- روش تحت وب :

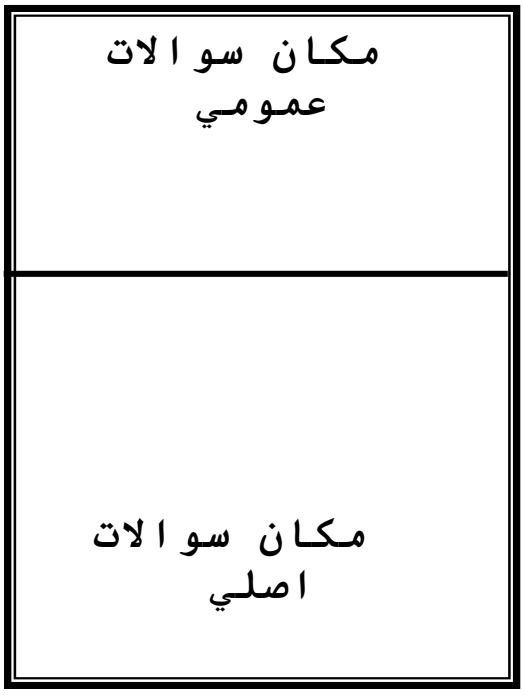
در این روش پرسشنامه پیمایش بر روی صفحه وب قرار گرفته و افراد برای پاسخ گویی به این سایت مراجعه نموده و نسبت به پاسخ گویی به سوالات اقدام می نمایند. این روش معمولا برای مواردی که با تعداد مشخصی از عناصر مورد بررسی و ذینفع در پاسخ گویی مواجح هستیم می تواند سودمند باشد. این روش نیز مانند روش قبل دارای مزایای زیادی از جمله کاهش زمان، هزینه تحقیق و کاهش خطاهای غیر آماری تحقیق است . معمولا در نظر سنجی از گروهی مشخص مثل اعضای یگ گروه یا کلوپ و یا اعضای یک شرکت در مورد موضوعی که آنها در آن ذینفع هستند می‌توان از این روش استفاده نمود. البته ذکر این نکته ضروری است که باید تدابیری اندیشیده شود که هر شخص نتواند بیش از یکبار به پرسش نامه پاسخ دهد. برای مثال در نظر سنجی برای انتخاب

قویترین مرد قرن بیستم که در سال 2005 توسط فدراسیون بین المللی وزنه برداری انجام شد به دلیل اینکه تعدادی از مردم ایران و یونان به قهرمان مورد نظر خود بارها رای داده بودند نتایج نظر سنجی از طریق وب مردود اعلام شد.

5-طراحی پرسشنامه:

طراحی یک پرسشنامه مناسب که بتواند تمامی اهداف تحقیق را پوشش دهد نیازمند اطلاع کامل به دستور زبان ، داشتن اطلاعات کامل از موضوع مورد نظر، داشتن آشنایی با مبانی نظری طراحی پرسشنامه و همچنین آشنایی با انواع خطاهای آماری و غیر آماری قابل ایجاد در مراحل مختلف طراحی پرسشنامه و رشهای کنترل آنهاست.

همانگونه که قبلا عنوان شد هر یک از سوالات پرسش نامه باید منطیق بر یکی از اهداف یا مفروضات پیمایش باشند و نباید هیچ سوال اضافی و بدون هدفی در پرسشنامه موجود باشد. به طور کلی نمای عمومی یک پرسشنامه به صورت زیر است:



این سه قسمت ، سه بخش اصلی یک پرسشنامه را تشکیل می دهند. سوالات عمومی معمولا" سوالاتی هستند که هدف از پرسش آنها بدست آوردن رابطه آنها با سوالات اصلی است. برای مثال اگر هدف از انجام تحقیق شناسایی میزان رضایت شهروندان از محصول شما است و فکر می کنید که جنسیت مصرف کنندگان و یا منطقه زندگی و یا سن آنها بر روی میزان رضایتشان از محصول شما موثر است می توانید این گزینه ها را به عنوان سوالات عمومی مطرح نمائید تا در بخش تحلیل نتایج داده های جمع آوری شده رابطه آنها با سوالات اصلی قابل ارزیابی و اندازه گیری باشد.

بهتر است در صورتی که سوالات عمومی حساسیت در پاسخ دهندگان ایجاد نمی کند در ابتدای

پرسشنامه سوال شوند تا شخص در پاسخ به سوالات اصلی تحقیق احساس راحتی و مسئولیت بیشتری نماید. در این قسمت از پرسشنامه از پرسیدن سوالاتی نظیر درآمد، قومیت و ... که باعث حساسیت افراد و جبهه گرفتن در مقابل سوالات پرسش نامه می‌شود جدا" خود داری نمائید و چنانچه چاره ایی جز پرسیدن سوالات فوق ندارید بهتر است در آخر پرسش نامه واقع شوند.

منظور از سوالات فرعی سوالاتی هستند که مانند سوالات اصلی نقشی در پیمایش ندارند و مانند سوالات عمومی نیز به دنبال ارزیابی رابطه آنها با سوالات اصلی نیستید و تنها به این دلیل که یک پیمایش میدانی انجام داده اید که می توانستید برای حداکثر کردن بهره‌وری اطلاعاتی جانبی را که ممکن است در آینده به کار آیند نیز جمع آوری نموده‌اید.

از نقاط قوت یک پیمایش از طریق پرسشنامه این است که پرسشنامه طراحی شده تمام نیاز تحقیق را پوشش دهد و این تنها زمانی اتفاق می افتد که هر یک از سوالات پرسشنامه طراحی شده با دقت و با خوبی تهیه شده باشند. برای دستیابی به نتایج و اهداف یک تحقیق باید نقش هر یک از سوالات پرسشنامه کاملاً مشخص باشد. در این مقوله سعی شده به این مورد مهم پرداخته شود.

طراحی یک پرسشنامه دارای مراحل زیر است:

A – مشخص کردن سوالاتی که پرسیده می شوند.

B – مشخص کردن نوع سوال(باز یا بسته بودن، چند گزینه ای بودن و ...) برای هر سوال و جمله بندی آن.

c- طراحی توالی سوالات و چیدمان کلی پرسشنامه

حال به بیان مختصر هر یک از مراحل می پردازیم.

مشخص کردن سوالاتی که پرسیده می شوند:

پرسیدن سوالاتی که در جهت اهداف تحقیق تهیه نشده است علاوه بر طولانی کردن پرسشنامه و افزایش خطاهای غیر آماری و هزینه های تحقیق، منجر به کاهش کارایی پرسشنامه نیز می گردند. همانطور که در بخش مشخص کردن اهداف به آن اشاره شد سعی نمایید هر یک از سوالات به گونه ایی جنبه های مختلف مفروضات تحقیق را پوشش دهند و از طرفی دیگر تمامی جنبه های مورد نظر یک فرضیه نیز پوشش داده شود

یک ارتباط کلیدی باید میان اهداف تحقیق و هریک از سوالات وجود داشته باشد، سوالات به واسطه ارتباط با هر یک از اهداف تحقیق مشخص می گردند. یک شرح مختصر در قالب یک مثال مطرح شده است:

هدف تحقیق کشف عواملی است که باعث می شود کاندیداهای مجلس برنامه های خود برای توسعه را به فراموشی بسپارند.

سوالات مرتبط	دلایل مطرح شده نتیجه بحث
تغییر سابقه استفاده از مسند برای پیشبرد اهداف پنهان دستمزد بیشتر وجهه اجتماعی بیشتر خدمت به شهروندان	چه عاملی باعث کاندیداشدن برای مجلس می شود
کاندیدا چند سال سابقه کار دارد	آیا سابقه گذشته تأثیر گزار بوده ؟
کاندیدا مرد است یا زن	آیا تفاوت جنسیتی عامل تاثیر گزاری است؟
بالاترین سطح تحصیلات چیست؟	آیا سطح تحصیلات و پیشرفت تأثیرگزار

پیشرفت کاری در کدام زمینه بوده است؟ در کلام رشته تحصیلات دارد؟	است؟

همانطور که در بالا ملاحظه می شود ابتدا از طریق افراد صاحب نظر و حتی در صورت عمومی بودن موضوع از نظر مردم عادی دلایل کلی مطرح شده را که به عنوان مفروضات تحقیق مطرح هستند بدست آورده و سپس سوالات مرتبط با هر یک از مفروضات را که در حقیقت هر کدام از جنبه ای خاص فرض را مورد ارزیابی قرار می دهند مطرح می‌نماییم.

ذکر این نکته ضروری است که هر پرسشنامه باید دارای یک شماره یا کد منحصر به فرد باشد. بهتر است که این شماره قبل از توزیع پرسشنامه مشخص باشد. این مشخصه می‌تواند حروف یا اعداد یا تلفیقی از هر دو باشد ، این کد باید برای همه پرسشنامه ها قید شود، علاوه بر آن این کدهای منحصر به فرد شما را قادر می سازد تا اطلاعات موجود پرسشنامه و اطلاعات وارد شده در نرم افزار را هر زمان که خواستید با یکدیگر مرتبط نمایید.

مشخص شدن نوع سوال و جمله بندی آن:

به طور کلی سوالات می تواند به دو صورت مورد پرسش واقع گردند. در حالت اول که که معروف به سوال "باز" است . برای مثال شما می توانید نظر افراد را در مورد مورد کیفیت محصول خود سوال نمایید و از افراد بخواهید توضیحی در مورد کیفیت محصول بیان کرده یا در پرسشنامه وارد نمایند و یا سن افراد را در قالب سوال باز مطرح می نمایید تا هر فرد سن مورد نظر را بر حسب سال وارد نماید.

سوالات باز در موراد زیر بیشتر مورد استفاده قرار می گیرد:

1- زمانی که امکان پرسش سوال به صورت بسته وجود ندارد . برای مثال رنگ مورد علاقه شما چه رنگی است؟ و یا استان محل تولد شما کجاست؟ در اینگونه موارد اگر بخواهیم تمامی گزینه

های پاسخ را در پرسشنامه قرار دهیم تا افراد یکی را متناسب با نظر خود انتخاب نمایند ممکن است در نهایت با یک پرسشنامه چند صفحه ایی مواجه باشیم که کارایی پیمایش را به حداقل می رساند.

2- زمانی که قرار دادن گزینه های مشخص بسته باعث جهت گیری شخص و انتخاب یکی از گزینه ها بدون انجام تجزیه و تحلیل و دقت در پاسخگویی می شود.

برای مثال اگر از مصرف کنندگان یک محصول غذایی بپرسید کدام مارک محصول را برای مصرف ترجیح می دهید؟ در صورتی که پیمایش شما مقارن با تبلیغات وسیع یک شرکت تولید کننده محصول مورد نظر باشد آنگاه مصرف کننده به محض دیدن مارک مورد نظر نسبت به انتخاب آن اقدام می نمایند هر چند که تا کنون آن را تجربه نکرده باشد . در صورتی که از سوال باز استفاده نمایید امکان تاثیر گذاری عوامل خارجی مانند مثال بالا در آن کمتر خواهد بود.

3- زمانی که شما به عنوان انجام دهنده پیمایش احساس می کنید که با توجه به اهمیت سوال نیاز به تجزیه و تحلیل ذهنی بیشتری توسط پاسخ دهندگان برای پاسخ وجود دارد . زیرا افراد در پاسخ گویی به سوالات باز بیشتر تامل و تجزیه و تحلیل می نمایند.

4- هنگامی که قصد سنجش نظرات افراد خبره و متخصص را دارید.

5- هنگامی که یک پیمایش مقدماتی قبل از انجام پیمایش اصلی تحقیق انجام می‌دهید.

ذکر این نکته ضروری است که در بسیاری از روشهای جمع آوری اطلاعات که قبل تر گفته شد وجود سوالات باز به جز در مواردی که ناگزیر به استفاده از آنها هستیم(مانند گزینه ا و 2) توصیه نمی شود.

برای مثال در روشهای پیمایش از طریق پست الکترونیکی و پیمایش تحت وب استفاده از سوالات باز با توجه به محدودیتهای که افراد در انجام تایپ در کامپیوتر با آن مواجه هستند توصیه نمی شود اما در روشهایی نظیر روشهای رو در رو و کتابخانه ایی می تواند مفید باشد.

در بسیاری از پرسشنامه ها سوالات به صورت چند گزینه ایی و با انتخابهای محدود در اختیار پاسخ

دهنده قرار می گیرند . برای طراحی اینگونه سوالات ابتدا توسط یک پیمایش مقدماتی تمامی گزینه های ممکن که افراد مورد بررسی در نمونه انتخاب نمایند را شناسایی کرده و با دسته بندی آنها به تعدادی گزینه محدود و مشخص می رسیم که آنها مبنی پاسخ های افراد را تشکیل می دهند.

برای سوالات با گزینه های بسته مشخص مانند جنسیت و سطح تحصیلات تعداد گزینه ها مشخص است اما در مواردی که تعداد گزینه ها کاملا مشخص نیست بهتر است که یک قسمت با عنوان غیره..... نیز قرار داده شود تا شخص اگر پاسخ مورد نظر خود را در میان گزینه ها نیافت از این گزینه برای نوشتن پاسخ خود استفاده نماید.

علاوه بر نوع سوال، جمله بندی سوال نیز از همیت وپژه ایی برخوردار است . در صورتی که سوال جمله بندی مناسبی نداشته باشد هر فرد برداشت متفاوتی را از سوال مطرح شده خواهد داشت و در نتیجه نمی توان به نتایج پیمایش اطمینان داشت در اینگونه موارد در اصطلاح می گویند پرسشنامه طراحی شده در روایی دچار مشکل است. طراحی یک پرسشنامه مناسب با توجه به روش جمع آوری اطلاعات نیاز به آشنایی کامل با ادبیات و همچنین سابقه لازم در امر طراحی پرسشنامه دارد. پیشنهاد می شود چنانکه از روایی پرسشنامه طراحی شده خود اطمینان ندارید از روشهای جمع آوری اطلاعات رو در رو و یا تلفنی که به طور مستقیم پرسش شونده را با سوال شما مواجه نمی کند (در این روش پرسشگران سوالات را مطرح می کنند) استفاده نمائید. به طور کلی در جمله بندی سوالات باید موارد زیر را مد نظر قرار دهید:

- مختصر و مفید بودن (عدم دو پهلو بودن سوالات)
- اجتناب از سوالات دو منظوره
- دوری از سوالات معکوس ،
- دقت در مراعات ملاحظات عامه در سوالات
- پرهیز از سوالات جهت دار

طراحی توالی سوالات و چیدمان سوالات:

طراحی چیدمان سوالات و نحوه قرار گرفتن هر کدام از آنها نیز از اهمیت ویژه ایی برخوردار است. برای مثال در صورتی که در یک سوال از کارمندان میزان رضایت از مدیریت مستقیم آنها سوال شود و بلافاصله میزان رضایت از مدیریت ارشد مورد سوال واقع شود به نظر شما چه مشکلی به وجود خواهد آمد؟

بسیاری از افرادی که از مدیریت مستقیم خود رضایت ندارند برای سوال بعد هم بدون تجزیه و تحلیل گزینه های میزان رضایت کم را انتخاب می نمایند.

در مثالی دیگر از شهروندان پرسیده اید در میان شرکت های تولید کننده محصولات لبنی کدام مارک را ترجیح می دهید؟ در صورتی که یک مارک بخصوص برای یک شهروند با توجه به مطلوبیتی که در استفاده آن ایجاد شده ارجحیت دارد آنگاه باید سوال بعد را سوالی قرار دهید که مصرف کننده باز همین مارک را به دلیل مطلوبیت مصرف در محصول اول برای محصول دوم هم انتخاب نکند زیرا ممکن است هیچگاه تجربه محصول دوم با مارک انتخابی را نداشته باشد.

برای رفع این مشکل معمولا ابتدا یک تحقیق مقدماتی به منظور آزمون اثر بخش بودن پرسشنامه انجام می گیرد که در صورت مشاهده این مشکلات نسبت به رفع آنها در پرسشنامه پیمایش اصلی اقدام شود.

6-انجام پیمایش مقدماتی:

معمولا قبل از پیمایش اصلی یک پیمایش با تعداد نمونه محدود با اهداف زیر صورت می پذیرد:

1- چک کردن روش جمع آوری اطلا عات و نرخ برگشت پاسخ نامه ها و در صورت نیاز تغییر روش جمع آوری اطلاعات

2- آزمون روایی هر یک از سوالات و در صورت نیاز تغییر جمله بندی سوالات

3- مشخص شدن مناسب بودن گزینه های سوالات بسته و در صورت نیاز تعدیل گزینه ها

4- استفاده از نظرات افراد کارشناس و خبره در مورد پرسشنامه

5- مطرح کردن سوالات به صورت باز و انتخاب گزینه های بسته برای پرسشنامه اصلی

6- شناخت مقدماتی جامعه مورد بررسی و تعدیل فرضیات تحقیق و ...

ذکر این نکته ضروری است که در صورت عدم استفاده از پیمایش مقدماتی ، خطاهایی را کـه ممکـن بوده با ابتکار ساده ایی تحت کنترل درآیند به تحقیق وارد نموده اید و در نتیجه تحلیـل نتـایج دیگـر بـا دقت آماری پیمایش قابل بیان نخواهند بود. در یک تحقیق انجام گرفتـه بـر روی حـدود 300 پیمـایش اجتماعی در کشور آمریکا که توسط دانشگاه نبراسکا در سال 1992صورت گرفته است، نشان می دهـد استفاده از پیمایش‌های مقدماتی و استفاده از نتایج آن در پیمایش های اصلی باعـث کـاهش حـدود 65 درصدی خطاهای غیر آماری در این پیمایش ها در مقایسه با پیمایش های مشابه بـدون وجـود پیمـایش مقدماتی شده است . نظر شا چیست؟ آیا کنترل این میزان خطا ضرورت انجام پیمایش های مقدماتی را روشن می سازد؟

7- روشهای کنترل خطاهای آماری و غیر آماری:

قبل از اینکه روشهای کنترل خطاهای آماری و غیر آماری در پیمایش ها مورد بررسی قرار گیرد لازم است تا تعریفی از این دو نوع خطا ارائه شود.

خطای آماری به خطاهایی گفته می شود که محقق برای اسـتفاده از نمونـه بـه جـای جامعـه و آنگـاه

استنباط در مورد جامعه باید بپردازد. به عبارت دیگر با توجه به اینکه هدف محقق بررسی جامعه آماری است اما نمونه ایی از جامعه آماری انتخاب می نماید و سپس بر اساس این نمونه انتخابی در مورد تمام اعضای جامعه آماری قضاوت می کند مرتکب نوعی خطا شده است . این خطا که در علم آمار آنرا با a نمایش می دهند و هر چه مقدار

a کوچکتر باشد محقق نیاز مند تعداد نمونه بیشتری از جامعه مورد بررسی است.

مقدار خطای آماری در ابتدای تحقیق و بر اساس میزان بودجه تعیین شده تحقیق مشخص می شود. برای مثال یک مدیر می خواهد نتایج تحقیقی که به وی ارائه می شود حداکثر دارای 5 درصد خطای آماری باشد و 95 درصدبتوان به آن اعتماد نمود. در این حالت مقدار خطای آماری 5 درصد در نظر گرفته شده و با توجه به فرمول ارائه شده در حجم نمونه در بخش دوم تعداد نمونه ایی که باید با روشهای آماری اتخاذ گردد تا حداکثر 5 درصد خطای آماری در پیمایش وجود داشته باشد مشخص می شود.

خطای غیر آماری به خطایی گفته می شود که توسط محقق یا تیم تحقیق در مراحل مختلف انجام پیمایش صورت می گیرد. چند مورد از عوامل ایجاد خطاهای غیر آماری در پیمایش ها به صورت زیر هستند:

1- جمله بندی نامناسب سوالات
2- رعایت نکردن چیدمان سوالات در پرسشنامه
3- عدم انجام تحقیق مقدماتی
4- طرح سوالات زائد و غیر ضروری
5- وجود سوالات با جنبه خصوصی برای افراد
6- انتخاب روش نامناسب جمع آوری اطلاعات
7- عدم آموزش پرسشگران
8- عدم تفهیم هدف تحقیق برای پرسش شوندگان

9- ندادن اعتماد به پرسش شوندگان برای محرمانه ماندن نظرات آنها

10- تهیه نکردن نقشه نمونه گیری

11- ورود اشتباه کدهای اطلاعاتی پاسخ ها در رایانه

12- عدم وجود شرایط مطلوب مکانی و زمانی جمع آوری اطلاعات

13- و غیره...

مشکل وجود خطاهای غیر آماری بسیار جدی است زیرا هنگامی پیمایش که پایان می‌یابد اینگونه خطاها خود را نشان می دهند و در بسیاری از مواقع کل نتایج پیمایش به دلیل وجود خطاهای غیر آماری زیاد غیر قابل استفاده می ماند.

در یک بررسی توسط یک تیم پژوهشی بر روی 720 پایان نامه -انجام گرفته با استفاده از ابزارهای آماری- در دانشگاههای کشور، مشخص شد که حدود 92 درصد از تحقیقات آماری صورت گرفته دارای خطاهای آماری بوده اند به طوری که از 15 تا 100 درصد نتایج تحقیقات آنها را فاقد اعتبار می نموده است.

در زیر تعدادی از نکات را که می تواند به عنوان روشهایی برای کاهش خطاهای غیر آماری در کنار موارد مطرح شده در بالا مورد استفاده قرار گیرد بیان می نمائیم:

1-اکثر محققان به اشتباه برای جواب به مفروضات سوالات زیادی را مطرح می نمایند. که این امر منجر به نتیجه گیری و تحلیل ناقص مفروضات پیمایش می شود. اما بزرگترین خلاء در پیمایش ممکن است سرعت پاسخ باشد و یک پرسشنامه واضح و مختصر می تواند بهترین پاسخ را بدهد

2- تصمیم در مورد اینکه پیمایش به کدام روش انجام گیرد بستگی به طرح و چارچوب پرسشنامه دارد .

3- ساختار پرسشنامه‌هایی که به صورت رودررو جمع‌آوری می‌شوند می‌تواند خیلی متفاوت باشد. برای

مثال سوالاتی که شفاهی پرسیده می‌شوند دارای اصطلاحات متفاوتی از پرسشنامه‌هایی هستند که بصورت نوشته جمع می‌شوند دقت داشته باشید که پرسشگران شما در مورد وارد کردن پاسخ های متنی پاسخ دهندگان به پرسشنامه باید آموزش لازم را دیده باشند.

4- اگر پرسشنامه از طریق پست یا پست الکترونیک یا خود اظهاری و یا تحت وب جمع آوری می شود. یک نامه ضمیمه کنید که به تشریح بیشتر موضوع و تضمین کار کیفیت کار بینجامد.

5- لزومی ندارد تنها به یک روش جمع آوری اطلاعات بسنده کنید شما می توانید از روشهای ترکیبی جمع آوری اطلاعات هم سود ببرید.

6- این را در نظر داشته باشد که در سوالات بسته ممکن است یک پرسش شونده بیش از یک گزینه را مد نظر دارد امکان انتخاب بیش از یک گزینه را هم به او بدهید. برای مثال:

از کدامیک از وسایل زیر برای رفتن به دانشگاه استفاده می نمائید؟

اتوبوس ☐

ماشین شخصی ☐

دوچرخه ☐

غیره.....

7- از افقی نمون و آرایش متن ها دوری کنید ، استفاده از یک فونت خوانا باعث افزایش سرعت ودقت خواندن می شود

8- بهتر است از کلمات عامیانه و مصطح و از فرمت Bold استفاده نماییم

9- اگر شما مطمئن هستید که پاسخ دهندگان به همه سوالات شما بصورت کامل پاسخ می دهند، با سوالاتی شروع کنید که برای آنها جذابیت بیشتری دارد.

10- بعضی معتقد هستند که سوالات آسانتر برای پاسخ در اول پرسشنامه مناسب تر هستند زیرا باعث می شوند که پاسخ دهنده از لحاظ ذهنی برای به دیگر سوالات آمادگی بیشتری کسب نماید. اما

بعضی از افراد معتقد هستند که عدم پوشش سوالات اساسی واصلی و پیچیده در ابتدای پرسشنامه های بلند باعث خستگی پاسخ دهنده و در نیتجه عدم پاسخ گویی با دقت لازم به آنها خواهد شد. دقت داشته باشید این شما هستید که باید جریان روانی سرتاسر یک پرسشنامه را حفظ کنید.

11- در صورتی که سوالات رتبه بندی شده می پرسید مانند سوال زیر:

((در صورتی که بخواهید تعطیلات آخر هفته را در مکانی بگذرانید اهمیت هر یک از موارد زیر رتبه بندی می نمائید. { شامل نمره گزاری از 1 تا 4 در حالی که 1 با اهمیت ترین است }))

- ☐ محیط زیبا
- ☐ آزادی محل بیشتر
- ☐ قیمت مناسب
- ☐ وجود مردم
- ☐ هوای آزاد و سالم

گزینه ها را با توجه به پیمایش مقدماتی و میزان اهمیت آنها قرار دهید.

12- از پرسیدن سوالات جهت دار دوری کنید برای مثال سوال:

« آیا شما با نظر اکثریت مردم که سیستم بهداشت را سیستمی شکست خورده می دانند موافقید؟ »

سوال مناسبی نیست ، باید از هرگونه ایجاد ذهنیت که روی نظر شخص اثر گذار است دوری کرد.

13- برای رسیدن به اهداف از طرح سوالات مربوط به جنبه های خصوصی افراد باید اجتناب کرد، زیرا در غیر این صورت نرخ پاسخ کاهش خواهد یافت.

14- اگر شما مجبور هستیدکه بعضی سوالات شخصی افراد را بپرسید باید آنها را در 2 چیز حتماً متقاعد کنید اول اینکه پاسخ های آنها برای چه کاری از جانب شما مورد استفاده قرار می گیرد و بعد اینکه چه تمهیدی برای محرمانه ماندن اطلاعات آنها اندیشیده‌اید؟

15- اگر محقق قولی به افرادی که پرسشنامه ها را تکمیل می کنند می دهد حتماً باید به قول خود

عمل نماید و بهتر است که قسمتی از عمل به وعده در قبل یا حین و یا دقیقاً پس از تکمیل پرسشنامه باشد.

16- سوالاتی که افراد در مورد پاسخ دادن به آنها از خود واکنش نشان می دهند باید در انتهای پرسشنامه گنجانده شوند.

17- از پرسیدن سوالاتی که دوبار فعل منفی در آن بکار رفته دوری کنید. برای مثال: ((آیا با این جمله موافقید که افرادی که سیگار نمی کشند دارای طول عمر بیشتری نیستند؟))

علاوه بر موارد مطرح شده نکات دیگری نیز وجود دارد که به نظر می رسد در صورت رعایت نکات بالا خود محقق در مراحل مختلف پیماش با آنها روبر خواهد شد و با توجه به شرایط روش مناسبی را برای مرتفع کردن آن خواهد اندیشید.