

بناام خدا

دانشگاه فردوسی مشهد  
دانشکده ادبیات و علوم انسانی  
گروه علوم اجتماعی

# راهنمای تحلیل داده ها با SPSS ( در پژوهشهای اجتماعی )

احمد رضا اصغرپور ماسوله

۱۳۸۶

## بخش اول

### آغاز تحلیل

تحلیل داده ها تنها بخشی از یک پژوهش اجتماعی است. قبل از آن شما باید طرح تحقیق مناسبی تهیه کرده باشید و بر اساس آن داده های لازم را جمع آوری کنید. بعد از جمع آوری داده ها باید آنها برای ورود به SPSS آماده کرد. برای این کار باید تمام متغیرها و مقادیری که هر متغیر می تواند داشته باشد، مشخص شوند. برای این منظور از کدنامه استفاده می شود. برای اینکه بتوانید داده ها را وارد SPSS کنید باید بدانید که این نرم افزار چگونه کار می کند و چگونه می توانید با آن ارتباط برقرار کنید. هر کدام از این مراحل در بخش اول مورد بحث قرار گرفته اند. در فصل اول برخی نکات و پیشنهادات درباره طراحی پژوهش ارائه شده اند و هدف آن است که با رعایت این نکات شما بتوانید داده هایی با کیفیت بالا جمع آوری کنید. در فصل دوم درباره چگونگی تهیه کدنامه و تبدیل اطلاعاتی که جمع آوری کرده اید به چیزی است که به بهترین شکل برای SPSS قابل فهم باشد. در فصل سوم مروری بر SPSS خواهیم داشت و برخی مهارت های لازم بررسی خواهند شد. اگر تاکنون با SPSS کار نکرده اید لازم است که قبل از انجام هر تحلیلی آنچه در فصل سوم ارائه شده است را مطالعه کنید.

## فصل اول

### طراحی یک پژوهش

شاید عجیب باشد که در کتابی که درباره تحلیل داده ها با SPSS است درباره چگونگی طراحی پژوهش فصلی ارائه شود اما باید گفت که شیوه طراحی پژوهش تأثیری مستقیم بر کیفیت داده ها و تحلیل آنها دارد. اطلاعاتی که شما وارد SPSS می کنید باید از جایی آمده باشند- مثلاً پرسشنامه ها، اطلاعات جمع آوری شده از مصاحبه ها، مشاهده رفتارها و کدگذاری آنها یا اندازه گیری عینی نتایج و عملکرد یک پدیده. کیفیت داده ها رابطه مستقیمی دارد با کیفیت ابزار اندازه گیری. ابزار اندازه گیری هم بر اساس مباحث نظری و روشی پژوهش فراهم می شود. بنابراین بدون توجه به بخشهای نظری و روشی پژوهش نمی توان داده های خوبی جمع آوری کرد.

### برنامه ریزی پژوهش

انجام یک پژوهش خوب در گرو برنامه ریزی و اجرای دقیق آن است. کتابهای عالی بسیاری در زمینه طراحی پژوهش نوشته شده اند- مراحل کلی چون بررسی منابع پژوهش، تنظیم فرضیه ها، انتخاب طرح تحقیق، انتخاب و تخصیص سوژه ها، ثبت مشاهدات و جمع آوری داده ها- که می تواند راهنمای شما باشند. تصمیم هایی که در هر کدام از این مراحل گرفته می شوند می توانند بر کیفیت داده های جمع آوری شده و شیوه پاسخگویی به سوالاتی که در پی یافتن پاسخ آنها هستید تأثیر بگذارند. پیشنهاد می شود که برای طراحی تحقیق به بهترین شیوه وقت کافی اختصاص دهید. مطالعه متون گوناگون و متفاوت می کند به این امر کمک کند.

برای شروع طراحی پژوهش خود به این نکات توجه کنید:

- بررسی کنید که کدام طرح تحقیق ( به عنوان مثال آزمایش، پیمایش، مشاهده) بهترین شیوه پاسخگویی به سوال پژوهش شماست. هر کدام از رهیافتهای پژوهشی مزایا و معایبی دارند. برای سوال پژوهشی خود مناسب ترین رهیافت را انتخاب کنید. پژوهش های قبلی که در حوزه پژوهشی شما انجام شده اند به خوبی بررسی کنید و سعی کنید آنها را خوب بفهمید.

- اگر آزمایش را انتخاب کردید، تصمیم بگیرید که طرح مقایسه بین گروهی (سوژه های متفاوت در هر شرایط آزمایشی) یا طرح اندازه گیری تکراری (سوژه های مشابه تحت تمام شرایط آزمایشی) برای سوال پژوهش شما مناسب است. هر کدام از این دو محاسن و معایبی دارند پس هر دو رهیافت را به دقت بررسی کنید.
- در پژوهشهای آزمایشی مطمئن شوید که سطوح کافی را برای متغیر وابسته خود به کار برده-اید. استفاده از تنها دو سطح (یا دو گروه) به آن معنی است که سوژه های کمتری مورد نیاز هستند اما این امر نتیجه گیری های شما را محدود می کند. آیا وجود گروه کنترل لازم و مطلوب است؟ آیا نبود گروه کنترل نتایج شما را محدود می کند؟
- همواره بیش از آنچه نیاز دارید سوژه انتخاب کنید. به خصوص هنگامی که نمونه شما از انسانها تشکیل شده است. مردم چندان قابل اتکاء نیستند – یعنی همواره آنطور که انتظار می-رود رفتار نمی کنند، بیمار می شوند، پرسشنامه ها را به خوبی پر نمی کنند. پس متناسب با چنین اتفاقاتی برنامه ریزی کنید. در اینجا بدین بودن بهتر از خوش بین بودن است.
- در پژوهشهای آزمایشی بررسی کنید که در هر گروه به اندازه کافی سوژه داشته باشید (هنگامی که امکان پذیر است سعی کنید که تعداد آنها برابر باشند). در گروههای کوچک استخراج نتایج از نظر آماری معنی دار بین گروهها مشکل است (بحث توان در مقدمه بخش پنجم مورد بحث واقع شده است). محاسباتی وجود دارد که با استفاده از آنها می توان حجم نمونه مورد نیاز را محاسبه نمود. برای این بحث می توانید به کتابهای آماری مراجعه کنید.
- هر جا امکانپذیر است سوژه ها را به صورت تصادفی به گروههای آزمایشی انتساب دهید. این کار مسایل مربوط به نابرابری گروهها در طرحهای بین گروهی کاهش می دهد. همچنین خوب است که متغیرهای بیشتری بین گروهها مقایسه شوند تا مطمئن شویم که بین گروهها تفاوت ماهوی وجود ندارد. می توان تفاوت های مشخص شده را با آزمون های آماری مورد بررسی قرار داد (به عنوان مثال با آنالیز واریانس)
- متغیرهای مستقل مناسبی را انتخاب کنید که دارای اعتبار و قابلیت اعتماد باشند (به بحثی در همین زمینه در ادامه همین فصل توجه کنید). خوب است که سنجه های گوناگونی وارد شوند – برخی سنجه ها حساس تر از دیگران هستند. تمام تخم مرغهای خود را در یک سبد قرار ندهید.
- انتظار متغیرهای ویژه و غافلگیرکننده را داشته باشید. متغیرهایی وجود دارند که می توانند تبیین دیگری برای نتایج شما ارائه دهند. گاهی اوقات هنگامی که شما در طرح پژوهش خود بسیار بر خود تکیه کرده اید، قرار دادن چنین متغیرهایی مشکل می شود. همواره شخص

دیگری را داشته باشید (استاد راهنما، پژوهشگر همکار) که طرح تحقیق شما را قبل از انجام پژوهش بررسی کند. هر کاری که می توانید برای کنترل این متغیرهای غافلگیرکننده به کار برید. داشتن شناخت نسبت به حوزه پژوهش هم به شما کمک می کند که متغیرهای غافلگیرکننده را مشخص کنید. اگر متغیرهای دیگری وجود دارد که شما نمی توانید آنها را کنترل کنید، بررسی کنید که آیا می توانید آنها را اندازه گیری کنید؟ و آیا بوسیله اندازه گیری می توان آنها را به طور آماری کنترل کرد (به عنوان مثال به آنالیز واریانس).

- اگر در حال انجام یک پیمایش هستید، اجرای پایلوت به شما کمک می کند که مطمئن شوید که راهنمایی ها، سوالات و طیف روشن و واضح است. هر جا که امکان پذیر است انجام پایلوت با همان کسانی که پژوهش اصلی با آنها انجام خواهد شد، توصیه می شود. شما باید مطمئن شوید که پاسخگویان شما گویه های پیمایش و پرسشنامه را به خوبی می فهمند و به طور مناسبی به آنها پاسخ خواهند داد. آزمون پایلوت باید بتواند سوالات آزاردهنده برای پاسخگویان را مشخص کند تا حذف شوند.

- اگر می خواهید یک آزمایش را انجام دهید کل آزمایش را تمرین کنید و پایلوت را هم انجام دهید. این کار را برای آزمایش و آنچه می خواهید بسنجید هر دو انجام دهید. اگر از وسیله ای استفاده می کنید، مطمئن شوید که به درستی کار می کند. اگر از آزمایش کنندگان یا مصاحبه کنندگان متفاوتی استفاده می کنید، مطمئن شوید که به خوبی آموزش دیده اند و می دانند چه باید بکنند. اگر مشاهده کنندگان متفاوتی برای رتبه بندی رفتارها لازم هستند مطمئن شوید که می توانید به خوبی آنچه آنها دیده اند را رتبه بندی کنید. به صورت تمرینی یک بار این کار را انجام دهید و قابلیت اطمینان بین رتبه دهندگان را بررسی کنید (ببینید نمرات رتبه دهندگان متفاوت تا چه حد با هم سازگار هستند). آزمون پایلوت رویه ها و سنجه ها به شما کمک می کند که خطا باشند و عوامل مزاحم را در پژوهش اصلی بیاید. برخی از اینها را ممکن است نتوانید پیش بینی کنید (به عنوان مثال پشت پنجره آزمایشگاه کارگران سروصدا کنند)، اما تلاش کنید آنهایی را که می توانید، کنترل کنید.

### **انتخاب مقیاسها و سنجه های مناسب**

شیوه های گوناگونی برای جمع آوری داده ها وجود دارد که انتخاب آنها به ماهیت پژوهش شما بستگی دارد. این کار ممکن است سنجش نتیجه و عملکرد بر اساس برخی معیارهای عینی یا مجموعه ای از معیارهای عینی مشخص برای رتبه بندی رفتار باشد. همچنین ممکن است در بردارنده

طیفهایی باشد که برای عملیاتی کردن سازه ها یا ویژگی هایی طراحی شده اند که به طور مستقیم قابل سنجش نیستند (مانند اعتماد به نفس).

هزاران طیف اعتباریابی شده وجود دارند که می توانند در پژوهش مورد استفاده قرار گیرند. یافتن طیفی که برای پژوهش شما مناسب است ممکن است بسیار مشکل باشد. اولین قدم مطالعه آثار موجود در حوزه مورد پژوهش شماست. دیگر پژوهشگران از چه سنجه هایی استفاده کرده اند؟ گاهی اوقات آیتم های مورد نظر شما در ضمیمه مقالات پژوهشی وجود دارد. ممکن است مجبور شوید به مقاله اولیه مراجعه کنید که در آن گویه های مورد نظر، اعتبار بایی شده اند. برخی طیف ها حق تالیف دارند و این به آن معنی است که شما باید نسخه ای از آن را از ناشر خریداری کنید. برای دیگر طیفها که در مقالات پژوهشی منتشر شده اند فرض می شود که در حوزه عمومی قرار دارند و می توان از آنها بدون پرداخت پول استفاده کرد. با این حال مهم است به طور مناسبی از پژوهشگرانی که از کار آنها استفاده شده است تشکر شود و با جزئیات کامل به آنها ارجاع داده شود.

در انتخاب طیفهای مناسب دو ویژگی وجود دارد که باید به آنها توجه کرد: اعتبار و قابلیت اعتماد. هر دو این عوامل می توانند بر کیفیت داده هایی که شما جمع آوری کرده اید تاثیر بگذارند. هنگامی که طیف های در دسترس را به قصد استفاده بررسی می کنید باید برای طیف اطلاعات مربوط به اعتبار و قابلیت اعتماد آن را هم جمع آوری کنید. شما این اطلاعات را برای بخش روش پژوهش خود لازم دارید. بدون توجه به اینکه پژوهشگران قبلی تا چه حد به خوبی به بحث اعتبار و قابلیت اعتماد طیفهای مورد استفاده شما پرداخته اند، باید شما هم در آزمون پایلوت خود با نمونه خود این عامل را مورد بررسی قرار دهید. گاهی اوقات برخی طیفها برای برخی گروه ها قابل اعتماد هستند ( به عنوان مثال بزرگسالانی که می توانند انگلیسی صحبت کنند) اما هنگامی که در گروههای دیگر استفاده شود به کلی غیر قابل اعتماد باشد ( به عنوان مثال کودکانی که هیچ زمینه ای در زبان انگلیسی ندارند).

### قابلیت اعتماد

قابلیت اعتماد یک طیف نشان دهنده آن است که تا چه حد آن طیف از خطاهای تصادفی مبرا است. دو شاخص متداول برای بررسی قابلیت اعتماد طیفها عبارتند از قابلیت اعتماد آزمون-بازآزمون ( که به آن پایداری موقتی<sup>1</sup> هم می گویند) و سازگاری درونی. قابلیت اعتماد آزمون-بازآزمون بوسیله بررسی در یک جمعیت، در دو موقعیت متفاوت و به دست آوردن همبستگی بین دو دسته نمره بررسی می شود. همبستگی های بالا در آزمون-بازآزمون بالا نشان دهنده آن است که طیف قابلیت اعتماد خوبی دارد. شما باید ماهیت سازه مورد سنجش را که قرار است طیف آن را بسنجد در

---

<sup>1</sup> temporal stability

این نوع قابلیت اعتماد مورد توجه قرار دهید. طیفی که طراحی شده است تا وضعیتی خاص را بسنجد در چند هفته پایدار نمی ماند. بنابراین ممکن است قابلیت اعتماد یک طیف پایین به دست آید. با این حال در مورد وضعیتهای شخصی پایدار می توان امیدوار بود که در آزمون-بازآزمون همبستگی های بالا مشاهده شود.

دومین ویژگی قابلیت اعتماد که می توان مورد بررسی قرار داد سازگاری درونی است. این میزان نشان دهنده آن است که گویه هایی که سازنده طیف هستند تا چه حد همگی یک خصیصه پنهان را می سنجد ( یعنی میزان هم بسته بودن گویه ها با هم ). سازگاری درونی را می توان به شیوه های گوناگونی سنجید. رایج ترین آماره مورد استفاده برای این کار ضریب آلفای کرونباخ است ( که در SPSS در دسترس است و در فصل ۹ درباره آن بحث خواهد شد). این آماره ارائه دهنده متوسط همبستگی میان تمام گویه های تشکیل دهنده سازه است. مقادیر آن بین صفر و یک هستند و هرچه این مقدار بیشتر باشد نشان دهنده آن است که قابلیت اعتماد بیشتر خواهد بود.

بسته به ماهیت و هدف طیف، سطوح متفاوتی از قابلیت اعتماد مورد نیاز است، اما پیشنهاد شده است به طور کلی حداقل آن ۰/۷ باشد. مقدار آلفای کرونباخ به تعداد گویه های موجود در طیف وابسته است. هنگامی که تعداد گویه ها در طیف اندک است (کمتر از ده گویه ) مقدار آن میتواند بسیار کم باشد. در چنین وضعیتی شاید بهتر باشد میانگین همبستگی بین گویه های محاسبه و گزارش شود. میانگین همبستگی بین گویه ای بهینه بین ۰/۲ تا ۰/۴ است.

### اعتبار

اعتبار یک طیف مربوط به آن است که تا چه حد آن طیف همان چیزی را که قرار بوده بسنجد، سنجیده است. متأسفانه هیچ شاخص واضح و مطلقی برای بررسی اعتبار طیف وجود ندارد. اعتباریابی طیف شامل جمع آوری شواهد تجربی با توجه به کاربرد آن است. مهمترین انواع اعتبار که با آنها مواجه خواهید شد عبارتند از اعتبار محتوا، اعتبار معیار، و اعتبار سازه.

اعتبار محتوا مربوط به کفایت نمونه گیری سنجه یا طیف از جهان یا دامنه مورد نظر است. اعتبار معیار مربوط به رابطه میان نمرات طیف و برخی معیارهای مشخص و قابل اندازه گیری است. اعتبار سازه مربوط به آزمون یک سازه است، نه بر اساس یک معیار بلکه با توجه به فرضیه های به دست آمده از نظریه ها که درباره ماهیت متغیر یا سازه مورد نظر است. اعتبار سازه با بررسی رابطه آن با دیگر سازه ها مورد بررسی قرار می گیرد، هم آنهایی که مربوط هستند (اعتبار همگرا) و هم آنهایی که مربوط نیستند (اعتبار متمایز<sup>۲</sup>). برای مطالعه بیشتر درباره انواع اعتبار می توانید به کتابهای روش تحقیق مراجعه کنید.

---

<sup>2</sup> discriminant validity

کتابها و مقالات خوب بسیاری وجود دارند که می توانند به انتخاب طیفهای مناسب کمک کنند. اگر بخواهید خود طیفی را طراحی کنید هم برخی از این کتابها مفید خواهند بود. می توانید مراجعی را که در آخر فصل معرفی شده اند بررسی کنید.

### آماده سازی یک پرسشنامه

در بسیاری از پژوهشها لازم است که اطلاعات از سوژه ها یا پاسخگویان جمع آوری شوند. این امر ممکن است در بردارنده جمع آوری اطلاعات جمعیت شناختی از سوژه ها قبل از وارد کردن آنها در آزمایش باشد. ازسوی دیگر ممکن است بخواهیم یک پیمایش گسترده را در نمونه ای از جمعیت مورد مطالعه انجام دهیم. پرسشنامه ای که ضعیف برنامه ریزی و طراحی شده است نمی تواند به خوبی به سوالات پژوهش پردازد. در تهیه یک پرسشنامه باید توجه کنید که قصد دارید چگونه از اطلاعات استفاده کنید. باید بدانید که از کدام آماره ها می خواهید استفاده کنید. بسته به تکنیک آماری که می خواهید استفاده کنید ممکن است سوال خاصی را به شیوه خاصی پرسید و یا اشکال خاصی را برای پاسخها در نظر بگیرید. در بخشهای بعدی برخی از عواملی را که نیاز است در طراحی و ساخت پرسشنامه به آنها توجه شود ارائه شده اند.

این بخش تنها به طراحی اولیه پرسشنامه می پردازد.

### انواع سوال

اغلب سوالات را می توان در دو گروه دسته بندی کرد: بسته و باز. در یک سوال بسته پاسخهایی به پاسخگو ارائه شده اند که او باید یکی از آنها را انتخاب کند. از آنها خواسته شده است که با علامت خاصی یکی از آنها مشخص کند. گزینه ها می توانند به سادگی بله/خیر یا مرد/زن باشند و یا گزینه های متفاوتی را در بر داشته باشند:

بالاترین دوره تحصیلی که گذرانده اید کدام است (لطفا علامت بزنید)؟

۱- ابتدائی و راهنمائی                      ۲- دبیرستان اما تمام نکرده ام

۳- اتمام دبیرستان                      ۴- دوره های آموزشی دیده ام

۵- دانشگاه (لیسانس و پایینتر)                      ۶- دانشگاه (بالا تر از لیسانس)

سوالات بسته را معمولاً به سادگی می توان به صورت عدد درآورد که برای SPSS قابل فهم است. به عنوان مثال بله می تواند ۱ باشد، و خیر می تواند ۲ باشد. مرد می تواند ۱ باشد و زن می تواند ۲ باشد. در سوالی که در بالا درباره تحصیلات پرسیده شد عدد مربوط به گزینه انتخاب شده وارد SPSS می شود. به عنوان مثال اگر پاسخگو گزینه دانشگاهی (لیسانس و کمتر) را انتخاب کند کد ۵ را به آن



خواهیم داد. دادن عدد به هر کدام از پاسخها در پرسشنامه کدگذاری را ساده می کند. برای ورود داده ها کد را انتخاب کنید و در خود پرسشنامه و یا کنار صفحه قرار دهید.

گاهی اوقات ممکن است شما نتوانید تمام پاسخهای پاسخگویان را حدس بزنید در چنین هنگامی استفاده از سوالات باز لازم خواهد بود. حسن چنین سوالاتی آن است که پاسخگویان آزاد هستند تا آنطور که می خواهند پاسخ دهند و به گزینه هایی که پژوهشگر در مقابل او قرار داده محدود نمی شوند. به عنوان مثال:

در حال حاضر چه چیز هایی بیش از همه در زندگی به شما استرس وارد می کنند؟

.....

پاسخهایی که به سوالات باز داده شده است را می توان در تعدادی مقوله محدود کرد به صورتی که بتوان آنها را وارد SPSS کرد. این مقولات معمولاً بعد از آنکه محدوده پاسخهای واقعی داده شده به سوالات بررسی شد به دست می آید. برخی پاسخهای احتمالی را هم می توان از پژوهشهای قبلی انجام شده در زمینه مورد نظر به دست آورد. به هر کدام از این مقولات عددی تخصیص داده می شود (به عنوان مثال کار=۱؛ امور مالی=۲؛ روابط اجتماعی=۳ و مانند اینها) و این اعداد وارد SPSS می شوند. جزئیات بیشتر در این زمینه در بخش مربوط به آماده کردن کدنامه ارائه شده است.

گاهی اوقات ترکیبی از سوالات باز و بسته نتیجه بهتری دارد. این کار به آن معنی است که پاسخهایی در اختیار پاسخگویان گذارده می شود و مقوله ای اضافی هم با نام «غیره» اضافه می شود که اگر پاسخ مورد نظر پاسخگو در میان گزینه های ارائه شده نبود، می تواند گزینه غیره را انتخاب کند. یک با دو خط جا هم برای نوشتن توضیحات مورد نظر در اختیار پاسخگو قرار داده می شود. استفاده از ترکیب سوالات باز و بسته در حوزه هایی که پژوهشهای کمتری در آنها انجام شده است لازم است چرا که به پژوهشگر می فهماند که آیا مقولاتی که او به عنوان پاسخ در نظر گرفته است مناسب هستند یا خیر و می تواند به خوبی پاسخهایی که پاسخگویان در ذهن دارند پوشش دهند یا نه.

### شکل پاسخ

هنگام پرسیدن یک سوال شما باید درباره شکل پاسخها هم تصمیم بگیرید. شکل پاسخ انتخابی شما می تواند در تحلیل های آماری شما تاثیر گذار باشند. برخی تحلیلها مثل همبستگی نیاز به نمرات پیوسته دارند، از کم تا زیاد، که دامنه مقادیر آن هم زیاد باشند. اگر از پاسخگویان خواسته اید که سن خود را با انتخاب یکی از مقولات مشخص کنند (کمتر از ۳۰ سال؛ بین ۳۱ تا ۵۰ و بیش از

۵۰) چنین داده ای را نمی توان برای تحلیل همبستگی به کار برد. پس اگر می خواهید بین سن و اعتماد به نفس همبستگی بگیرید باید سوالات خود را به گونه ای پرسید که پاسخگویان سن واقعی خود را بگویند.

تلاش کنید تا آنجا که امکان دارد گزینه های پاسخ متنوع تر باشند. شما همواره می توانید تعداد گزینه ها را بعداً در در SPSS کاهش دهید (فصل ۸ را ببینید). هرگز تنها از پاسخگویان نپرسید که با گزاره ای موافق هستند یا مخالف. از طیف لیکرت استفاده کنید که می تواند از بسیار موافق تا بسیار مخالف باشد.

بسیار مخالف	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	بسیار موافق
-------------	---	---	---	---	---	---	---	---	---	----	-------------

چنین طیف پاسخی امکان نمره دهی گسترده تری را به شما می دهد و تحلیل های آماری بیشتری را در اختیار شما قرار می دهد. شما باید درباره تعداد رتبه های پاسخ تصمیم بگیرید (مثلاً از یک تا ده).

هر نوع پاسخی را که انتخاب کردید باید راهنمایی های واضحی برای پاسخگویی ارائه دهید. آیا می خواهید پاسخگویان در مربعی تیک بزنند، دور عددی خط بکشند، یا اینکه در خطی علامت بزنند؟ ممکن است برای بسیاری از پاسخگویان این اولین پرسشنامه آنها باشد. فکر نکنید که آنها می دانند چگونه باید پرسشنامه را پر کنند. راهنمایی های واضحی ارائه کنید و اگر امکان دارد مثالی بزنید. و همواره با همان کسانی که مورد پژوهش شما هستند آزمون پایلوت را اجراء کنید. قبل اینکه پرسشنامه خود را به تعداد نمونه تکثیر کنید، تمام ابهامات آن را از بین ببرید.

در طراحی پرسشها همواره توجه کنید که پاسخگو سوال شما را چگونه خواهد فهمید و تمام جوابهای ممکن را در ذهن آورید. به عنوان مثال ممکن است شما بخواهید بدانید که آیا پاسخگویان سیگار می کشند یا خیر. می توانید چنین پرسید:

آیا سیگار می کشید؟ (لطفاً دور پاسخ خط بکشید)	بله	خیر
----------------------------------------------	-----	-----

پاسخگویان ممکن است پرسند منظور شما سیگار، سیگارت یا ماری جوانا است؟ آیا این سوال به شما می گوید که آنها چه اندازه سیگار می کشند؟ (دو سه سیگار، در مقابل دو یا سه بسته سیگار) یا چگونه سیگار می کشند (هر روز یا تنها در موقعیتهای خاص)؟ در اینجمل منظور آن است

که به تمام سوالات خود توجه کنید، و ببینید چه اطلاعاتی آنها به شما می دهند و چه اطلاعاتی ممکن است از دست برود.

### جمله بندی سوالات

نوشتن سوالات واضح و روشن خود یک هنر است. متأسفانه قوانین مشخصی برای راهنمایی در انجام این فرایند وجود ندارد. اما کارهایی وجود دارد که می توانید برای بهبود سوالات خود انجام دهید. اوپنهیم نکاتی را پیشنهاد داده است که در تنظیم سوالات خود باید از آنها اجتناب کنید. اجتناب کنید از:

- سوالهای طولانی و پیچیده
  - سوالاتی که دارای دو محتوا هستند.
  - کلمات و اصطلاحات فرهنگی خاص
  - سوالات جهت دهنده
  - بسیار منفی
  - اصطلاحات تخصصی و اختصارات
  - کلماتی که دارای دو معنی هستند.
  - کلماتی که بار احساسی دارند.
- در جای مناسب، خوب است که گزینه ای به عنوان «نمی دانم» یا «کاربردی ندارد» اضافه شود. برای توضیحات اضافه به کتابهای روش تحقیق مراجعه کنید.

## فصل دوم

### تهیه کدنامه

قبل از اینکه بتوانید اطلاعات را از پرسشنامه، مصاحبه یا آزمایش وارد SPSS کنید لازم است که کدنامه‌ای تهیه شود. این فصل خلاصه‌ای از راهنمایی‌هایی است که برای تبدیل اطلاعات به دست آمده از سوژه به چیزی که برای SPSS قابل فهم باشد، لازم است. لازمه تهیه کدنامه آن است که درباره نکات زیر تصمیم بگیرید و آنها را ثبت کنید:

- تعریف کردن و برجسب گزاردن برای متغیرها و
- تخصیص یک عدد برای هر کدام از پاسخها.

تمام این اطلاعات باید در یک دفترچه یا فایل کامپیوتری ثبت شود. این کدنامه باید در جایی امن نگاه داری شود. هیچ چیز بدتر از آن نیست که بعد از مدتی به فایل داده‌های خود مراجعه کنید و کلمات اختصاری را فراموش کرده باشید.

در کدنامه باید تمام متغیرهای پرسشنامه را لیست کنید. نام‌های اختصاری برای متغیرها و کدگذاری پاسخها که در SPSS آورده شوند. در این فصل مثالهای ساده‌ای برای نشان دادن مراحل مختلف آورده شده‌اند.

در اولین ستون جدول ۱-۲ اسم متغیر آورده شده است. در ستون دوم نام اختصاری متغیر و همان نامی که در SPSS آورده شده است آمده است و در ستون سوم جزئیات کدگذاری پاسخها آورده شده است.

### نام متغیرها

هر سوالی در پرسشنامه شما باید نامی منحصر به فرد داشته باشد. برخی از این نامها به روشنی اطلاعات را مشخص می‌کند (به عنوان مثال جنس، سن و غیره). سوالات دیگر مانند آنهایی که سازنده یک طیف است، ممکن است بوسیله کلمه اختصاری مشخص شوند (مانند خوشبینی ۱، خوشبینی ۲، خوشبینی ۳، خوشبینی ۴ که برای مشخص کردن گویه‌هایی که طیف خوش بینی را می-سازند به کار رفته‌اند).

متغیر	نام متغیر در SPSS	راهنمای کدگذاری
شماره شناسایی	ID	شماره اختصاص یافته به هر

پرسشنامه		
جنس	Sex	۱=مرد
		۲=زن
سن	Age	سن پاسخگو بر حسب سال
وضعیت تاهل	Marital	۱=مجرد
		۲=رابطه دائمی
		۳=ازدواج اول
		۴=ازدواج مجدد
		۵=طلاق/جدایی
		۶=بیوه
گویه های	Op1 تا Op6	شماره ای را که دور آن
خوشبینی ۱ تا ۶		خط کشیده شده وارد کنید
		از ۱ (بسیار مخالف) تا ۶
		(بسیار موافق)

نکاتی وجود دارند که در نام گذاری متغیرها در SPSS باید در نظر داشت. این قوانین در کادر زیر آورده شده است.

قوانین نام گذاری متغیرها
نام متغیر باید:
• منحصر به فرد باشد ( به این خاطر که در مجموعه داده ها هر متغیر باید نام متفاوتی داشته باشد)؛
• باید با یک حرف شروع شود نه عدد یا علامت
• نباید شامل جای خالی و برخی کاراکترها باشد (! * ? " )
• نباید شامل دستورات مورد استفاده در SPSS باشد (all, ne, eq, to, le, lt, by, or, gt, and, , not, ge, with).

اولین متغیر در هر داده‌ای باید شماره شناسایی پرسشنامه باشد که شماره منحصر به فردی است که مشخص کننده هر کیس است. قبل از شروع ورود داده ها به هر کدام از پرسشنامه ها یا رکورد داده ها شماره ای را اختصاص دهید. شماره را به روشنی روی صفحه اول آن بنویسید. این کار به شما

این امکان را می دهد که اگر در داده ها به مشکلی برخوردید به آسانی بتوانید به پرسشنامه بازگردید و موضوع را بررسی کنید.

### کد گذاری پاسخها

قبل ورود به SPSS هر پاسخی باید عددی مختص به خود داشته باشد. برخی از پاسخها ممکن است از قبل چنین شکلی داشته باشند (مانند سن که بر حسب سال است) اما برخی دیگر مانند جنس باید تبدیل به عدد شوند (مانند ۱=مرد؛ ۲=زن). اگر شما در سوالات خود از اعداد برای پاسخها استفاده کرده اید (مانند مثال تحصیلات در فصل اول)، این کار راحت تر خواهد بود. اگر چنین نکرده اید شیوه ای را برای تبدیل انتخاب کنید و انجام دهید. به عنوان مثال اولین پاسخ را ۱ و دومی را ۲ و الی آخر.

وضعیت تاهل شما اکنون چگونه است؟

مجرد      دارای رابطه      متاهل      جدا شده

برای کد گذاری سوال بالا اگر کسی مجرد باشد کد ۱ می گیرد. اگر متاهل باشد کد ۲ می گیرد و الی آخر.

### کد گذاری سوالات باز

برای کد گذاری سوال های باز که در آنها پاسخگویان می توانند به دلخواه پاسخ دهند، کد گذاری اندکی پیچیده تر است. به عنوان مثال این سوال را در نظر بگیرید: در حال حاضر اصلی ترین منبع استرس در زندگی شما چیست؟ برای کد گذاری این سوال شما باید تمام پرسشنامه ها را مرور کنید و به دنبال جوابهای مشترک بگردید. ممکن است پاسخگویان بسیاری را بیابید که منبع استرس خود را کار، مسائل مالی، روابط عشقی، سلامت یا کمبود وقت بدانند.

در کدنامه این پاسخها را تحت متغیر استرس و اختصاص شماره ای به هر کدام از پاسخها وارد می کنید (۱=کار؛ ۲=مسائل مالی و غیره). شما ممکن است که کد عددی دیگری را به عنوان پاسخ وارد کنید که شامل پاسخهایی باشد که در مقولات دیگر جای نگرفته است (۹=دیگر). هنگامی که داده ها را برای هر پاسخگو وارد می کنید آنها را با آنچه در کدنامه وجود دارد تطبیق دهید و شماره مناسب برای متغیر استرس وارد می کنید.

هنگامی که کدنامه خود را ساختید آماده ورود داده ها هستید. دو چیز است که اول باید انجام دهید:

۱- با فضای SPSS آشنا شوید، چگونگی باز کردن و بستن فایلها. همچنین با پنجره های مختلف آن آشنا شوید.

۲- فایل داده ای را با استفاده از اطلاعاتی که در کدنامه خود فراهم کرده اید، بسازید.

در فصل سوم ساختار SPSS مرور شده است و در فصل ۴ رویه های مورد نیاز برای ورود ساختن فایل و ورود داده ها بررسی شده است.

## فصل سوم

### آشنایی با SPSS

چند نکته است که قبل از هر چیز باید بدانید. SPSS با تعدادی پنجره متفاوت کار می کند که هر کدام کارهای متفاوتی را انجام می دهد. قبل از اینکه بتوانید به آن پنجره ها دست بیابید، باید یک فایل داده را باز کنید. بنابراین در این فصل می آموزیم که چگونه فایلی را باز کنیم یا ببندیم و اینکه چگونه فایل داده ای را بسازیم. سپس پنجره های مختلف را بررسی خواهیم کرد.

### آغاز کار با SPSS

- راههای متفاوتی برای شروع کار با SPSS وجود دارد:
- راحتترین کار کلیک کردن بر آیکون SPSS در دسک تاپ است.
  - همچنین می توانید از منوی Start و سپس programs گزینه spss for windows را انتخاب کنید.
- به هر طریقی که spss را شروع کنید کادر محاوره ۱-۳ مشاهده می شود. البته می توانید در پایین کادر با فعال کردن گزینه مورد نظر، دیگر هنگام باز کردن spss آن را مشاهده نکنید. اگر بخواهید فایلی را باز کنید ok را انتخاب می کنید.

### کار کردن با فایل های داده

نسخه های قبل از 15 spss به شما اجازه نمی دهد که بیش از یک فایل را همزمان باز کنید و اگر بخواهید دو فایل داده را همزمان باز داشته باشید باید دو بار spss را باز کنید. اما در نسخه 15 spss می توان همزمان بیش از یک فایل را باز کرد. همواره هر تغییراتی در فایل خود را ذخیره کنید اگر نه با بستن فایل و ذخیره نکردن، زحمات شما از دست می رود.





شکل ۳-۱- کادر محاوره هنگام باز کردن SPSS

### باز کردن فایلی که از قبل وجود دارد

اگر بخواهید فایلی را که از قبل وجود دارد، باز کنید (مثلا یکی از فایل‌هایی که به عنوان مثال در خود spss وجود دارد یا فایل‌هایی که در پایگاه‌های اینترنتی وجود دارند)، بر روی open an existing data source در شکل بالا کلیک کنید. با این کار می‌توانید در دایرکتوری‌های رایانه خود فایل‌هایی را ذخیره کرده اید بیابید. هنگامی که فایل خود را مشخص کردید بر روی آن کلیک کنید و سپس open را انتخاب کنید. به یاد داشته باشید که تمام فایل‌های spss دارای پسوند sav هستند. فایل داده‌ای را که انتخاب کرده اید در پنجره‌ای به نام data editor باز خواهد شد (درباره این پنجره بعداً بحث خواهد شد).

### نکته

اگر فایل داده‌های شما بر روی فلاپی دیسک است برای آنکه سرعت کار بیشتر شود و احتمال خرابی کمتر باشد بهتر است قبل شروع کار با spss فایل داده‌ها را بر روی کامپیوتر خود کپی کنید و با آن کار کنید. هنگامی که کار خود را تمام کردید فایل خود را بر روی فلاپی دیسک خود کپی کنید.

### ذخیره کردن فایل داده‌ها

هنگامی که برای اولین بار فایلی می‌سازید و یا در فایلی که از قبل موجود بوده است تغییری ایجاد می‌کنید باید ذخیره کردن را به یاد داشته باشید. این کار به صورت اتوماتیک صورت نمی‌گیرد

و شما باید خود این کار را انجام دهید. اگر فایل خود را هر چند دقیقه ذخیره نکنید با بروز مشکلی یا قطع برق و خاموش شدن رایانه تمام زحمات شما به هدر می رود. اگر در حال وارد کردن داده ها هستید هر چند دقیقه حتما عمل ذخیره سازی را انجام دهید.

برای آنکه فایل خود را ذخیره کنید به داخل منوی File بروید و Save را انتخاب کنید. همچنین می توانید دکمه ای را که شکل آن شبیه فلاپی دیسک است، در بالا سمت چپ پنجره، انتخاب کنید تا تغییرات ذخیره شود.

توجه کنید که کلیک کردن بر این دکمه تغییرات ایجاد شده توسط شما را بر روی همان فایلی که باز است ذخیره می کند.

هنگامی که برای اولین بار می خواهید فایلی را ذخیره کنید از Save as باید استفاده کنید. در اینجا از شما خواسته می شود که نامی برای آن انتخاب کنید و مکان ذخیره کردن آن را مشخص کنید.

### **باز کردن یک فایل داده دیگر**

اگر کار شما با یک فایل داده تمام شد و خواستید فایل دیگری را باز کنید تنها بر روی File و سپس Open کلیک کنید و دایرکتوری را که فایل شما در آنجا ذخیره شده است مشخص کنید. بر روی فایل دلخواه کلیک کنید و سپس open را انتخاب کنید. با این کار در spss 15 فایل جدید باز می شود در حالیکه فایل قبلی هم باز است اما در نسخه های قبلی، با باز شدن فایل جدید فایل قبلی بسته می شود.

### **ساختن یک فایل داده جدید**

هنگامی که spss باز می شود یکی گزینه هایی که در اختیار شما قرار می دهد Type in data است یعنی وارد کردن داده ها. این گزینه به شما امکان می دهد که فایل داده های جدیدی بسازید. با انتخاب این گزینه Spss صفحه سفیدی در اختیار شما می گذارد تا متغیرهای خود را تعریف کنید و داده های خود را وارد کنید (درباره این موضوع بعدا بیشتر گفته خواهد شد). اگر با یک فایل در حال کار هستید و می خواهید فایل جدیدی را شروع کنید می توانید از File سپس New و بعد از آن Data را انتخاب کنید.

قبل از آنکه بتوانید ورود داده ها را شروع کنید و آن را تحلیل کنید باید درباره پنجره ها و کادرهای گفتگویی مورد استفاده در spss توضیح داده شود. در بخش بعدی در این باره بحث خواهد شد.

## پنجره های SPSS

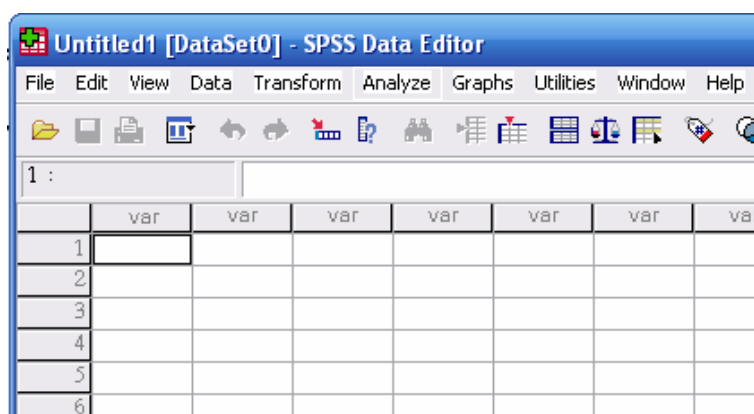
پنجره های اصلی که در spss از آنها استفاده خواهید کرد عبارتند از: ویرایش داده ها ( data editor)، نمایشگر (viewer)، ویرایشگر جدول (pivot table editor)، ویرایشگر نمودار ( chart editor) و ویرایشگر برنامه نویسی (syntax editor). این پنجره ها به طور خلاصه در اینجا توضیح داده شده اند اما در بخشهای بعدی توضیح مفصل داده خواهد شد.

هنگامی که تحلیل داده ها را شروع می کنید همزمان برخی از این پنجره ها باز خواهند بود. برای برخی دانشجویان این امر گیج کننده است. اما هنگامی که مسلط شوید این امر برای شما کاملاً ساده خواهد بود. Data Editor همیشه باز خواهد بود چون داده های شما در آن قرار دارد. هنگامی که تحلیل داده های خود را آغاز کنید پنجره Viewer هم همواره باز خواهد بود چون نتایج تحلیل های شما در آنجا نمایش داده خواهد شد و ترتیب نمایش نتایج هم آن هم مطابق ترتیب انجام تحلیل های شما خواهد شد.

بنابراین پنجره های مختلف مانند کاغذهای روی میز شما خواهد بود و در هر زمان یکی از آنها روی میز خواهد بود. دکمه مربوط به هر کدام از پنجره ها در پایین دسکتاپ قابل مشاهده است. برای تغییر پنجره ها تنها کافی است بر روی دکمه مربوطه کلیک کنید. برای دیدن پنجره ها می توانید منو window را هم کلیک کنید. در آنجا لیست تمام پنجره های باز را خواهید دید و می توانید پنجره-ای را که مورد نظر شماست انتخاب کنید.

## پنجره ویرایش داده ها (Data Editor window)

Data Editor نشان دهنده محتویات فایل داده شماست. شما برای باز کردن، ذخیره کردن و بستن فایل های موجود، بوجود آوردن فایل داده های جدید، ورود داده ها، اعمال تغییرات در فایل داده-ها، و اجرای تحلیل های آماری از آن استفاده خواهید کرد. اگر بخواهید هر تغییری در فایل داده های خود اعمال کنید، یا ذخیره کنید باید این پنجره را باز کنید و روی بقیه پنجره ها باشد. چنین گزینه-هایی در منوی پنجره های دیگر وجود ندارد. (شکل ۲-۳ را نگاه کنید)

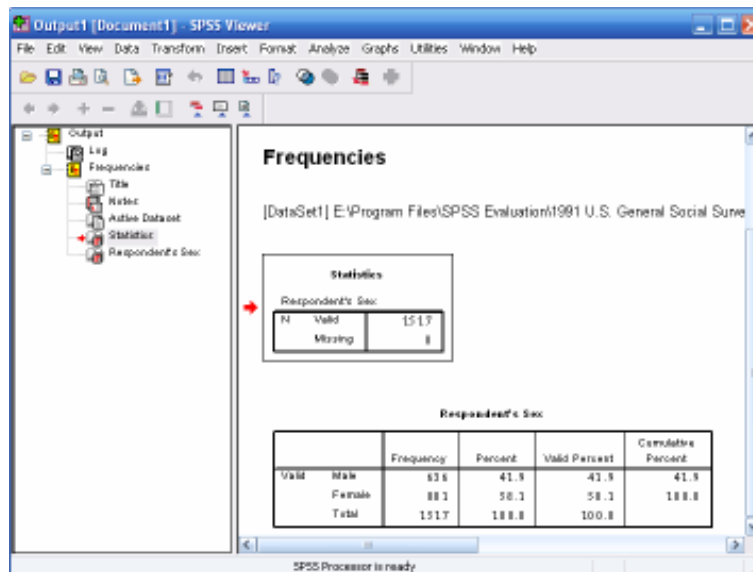


شکل ۲-۳- پنجره Data Editor

### پنجره نمایشگر (Viewer)

هنگامی که تحلیل خود را شروع می کنید، پنجره Viewer به طور اتوماتیک باز خواهد بود (شکل ۳-۳ را ببینید). این پنجره نمایشگر نتیجه تحلیل هایی است که شما انجام داده اید (نمودارها و جدولها). در این پنجره می توانید در خروجی ها تغییر ایجاد کنید، حذف کنید، کپی کنید، و حتی آنها را به word منتقل کنید. هنگامی که خروجی را ذخیره می کنید در فایل جداگانه ای با پسوند spo. ذخیره می شود به این طریق فایلی را که خروجی ها در آن ذخیره شده است می توان از دیگر فایلها تشخیص داد.

صفحه viewer دارای دو بخش اصلی است. در سمت چپ قسمتی وجود دارد که لیست کاملی از تحلیل هایی که انجام داده اید را در اختیار شما قرار می دهد. با استفاده از این قسمت می-توانید به آسانی قسمتهای مختلف خروجی را بیابید. (این خروجی با چند تحلیل به آسانی بسیار طولانی می شود). کافی بر روی نام قسمتی که می خواهید آن را ببینید کلیک کنید تا آن را در سمت راست ببینید. در قسمت سمت راست پنجره viewer می توانید نتایج تحلیلهایی را که انجام داده اید ببینید که شامل جداول و نمودارهاست.



شکل ۳-۳- خروجی نمونه در SPSS

### ذخیره کردن خروجی

برای ذخیره کردن خروجی ها پنجره viewer باید روبروی شما باشد و باید مانند آنچه برای ذخیره کردن در پنجره data editor انجام دادید عمل کنید. یعنی در بالای پنجره viewer از منو File گزینه save را انتخاب کنید. نامی را برای فایل انتخاب کنید و سپس جایی را که می خواهید فایل شما در آنجا ذخیره شود مشخص کنید و سپس بر روی save کلیک کنید. بهتر است برای نام گذاری خروجی ها از نام های اختصاری استفاده کنید که به خوبی معرف فایلی باشد که با آن کار کرده اید و زمان انجام تحلیل باشد. به عنوان مثال نام family6mordad85.spo نام خروجی است که من در تاریخ ۶ مرداد ۱۳۸۵ از فایل family گرفته ام. همچنین بهتر است نام فایلها و تحلیل هایی را که روی آنها انجام داده اید در دفترچه ای یادداشت کنید تا بعدا به آسانی بتوانید به آنچه دنبالش می گردید دست بیایید. هنگامی که کار تحلیل فایل خود را شروع می کنید به زودی با انبوهی از فایلهای خروجی که در هربار کار کردن با فایل داده ذخیره کرده اید مواجه می شوید. برای پیشگیری از سردرگمی و دوباره کاری بهتر است تحلیل های خود را به خوبی ذخیره و سازمان دهی کنید.

### پرینت گرفتن از خروجی ها

شما می توانید با استفاده از قسمت سمت چپ پنجره viewer قسمتهای خاصی از خروجی را برای چاپ کردن انتخاب کنید. برای این کار باید قسمت های مورد نظر خود را هایلایت کنید. اولین قسمت مورد نظر را انتخاب کنید سپس کلید کنترل (Ctrl) را روی صفحه فشار دهید و نگاه دارید و سپس قسمتهای دیگری را که مورد نظر شماست انتخاب کنید. برای چاپ کردن File را باز کنید و از

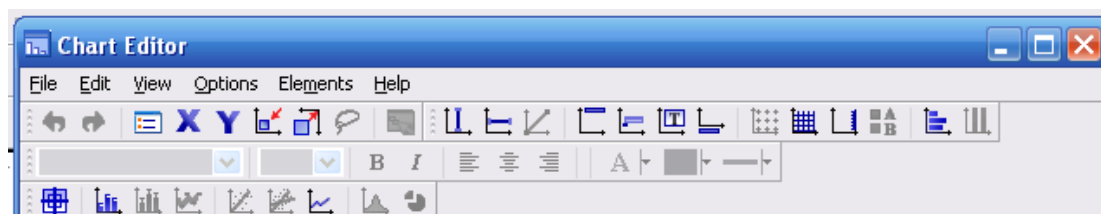
آنجا Print را انتخاب کنید. spss از شما خواهد پرسید که می خواهید تمام خروجی را چاپ کنید یا فقط قسمتهای انتخاب شده را.

### پنجره ویرایش جدول اصلی (Pivot Table Editor window)

ظاهر جدول هایی را که شما در Viewer می بینید می توانید با توجه به نیاز خود تغییر دهید. این کار با استفاده از پنجره Pivot Table Editor انجام می شود. برای مشاهده این پنجره باید بر روی جدول خود دوبار کلیک کنید. با استفاده از این پنجره می توانید نما، اندازه نوشته ها و فونت آنها و غیره را تغییر دهید. همچنین می توانید جای سطر و ستون را در جدول تغییر دهید.

### پنجره ویرایش نمودار (Chart Editor window)

هنگامی که از spss می خواهید نموداری برای شما رسم کند مثلاً هیستوگرام، نمودار ستونی یا نمودار پراکندگی آن را در پنجره viewer نمایش می دهد. اگر بخواهید تغییراتی در نوع یا شیوه نمایش بدهید باید از Chart Editor استفاده کنید. این پنجره با دوبار کلیک بر روی نمودار ظاهر می شود.



شکل ۳-۴- نمای Chart Editor window

در این پنجره شما می توانید ظاهر و نوع نمودار را مشخص کنید، فونتها، رنگها و الگوها و خطوط را تغییر دهید. شکل ۳-۴ را ببینید.  
در فصل هفتم ساختن نمودارها و استفاده از Chart Editor توضیح داده شده است.

### پنجره ویرایش برنامه (Syntax Editor window)

در نسخه های اولیه spss تمام دستورات باید به صورت برنامه به زبان خاص spss نوشته می - شد. با اینکه در نسخه های جدید دیگر برنامه نوشته نمی شود و تمام کارها در کادرهای محاوره ای

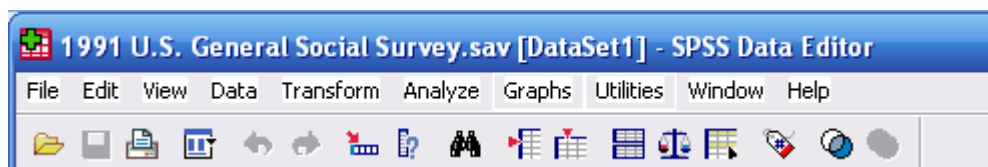
انجام می شود اما spss این برنامه ها را برای تمام کارهایی که شما در آن انجام می دهید ارائه می دهد. باید گفت تقریباً تمام کارهایی که دانشجوی کارشناسی احتیاج دارد در منوها وجود دارد اما گاهی اوقات که می خواهیم کنترل بیشتری بر آنچه انجام می شود داشته باشیم، برنامه نویسی به کار می آید. این کار با استفاده از Syntax Editor انجام می شود. در این پنجره می توانید دستورات را بنویسید یا آنها کپی کنید و دستوراتی را که spss ارائه داده است ویرایش کنید. مثالی برای این کار در فصل ۱۱ ارائه شده است.

## منوها

در تمام پنجره هایی در که در بالا توصیف شد منوهای وجود دارد. منوها هر کدام با دکمه کوچکی مشخص شده اند. به شکلهای ۳-۶ و ۳-۷- نگاه کنید.

برخی از منوها میان پنجره ها مشترک هستند. برخی هم مختص پنجره های خاصی هستند. این موضوع دانشجویانی را که تازه کار با spss شروع کرده اند گیج می کند. به عنوان مثال اگر شما بخواهید در داده های خود تغییری دهید (مثلاً متغیر جدیدی بسازید، متغیرهای موجود را کدگذاری مجدد کنید و یا داده ها را ذخیره کنید) باید از پنجره Data Editor این کار کنید. اگر بخواهید نتایج یا خروجی را ذخیره کنید یا تغییراتی در آنها بوجود آورید باید پنجره Viewer باز باشد. بنابراین این مهم است که توجه کنید کدام پنجره روی صفحه قرار دارد.

همانطور که شکلهای ۳-۶ و ۳-۷ و مشاهده می کنید بسیاری از گزینه ها در هر دو پنجره مشترک است اما برخی از آنها هم در پنجره دیگر قابل مشاهده نیستند مانند Data و Transform که مختص Data Editor هستند. در بیشتر اوقات شما می توانید با پنجره Viewer کارهای خود را انجام دهید و تنها زمانی به Data Editor مراجعه کنید که می خواهید تغییراتی در داده های خود بوجود آورید.



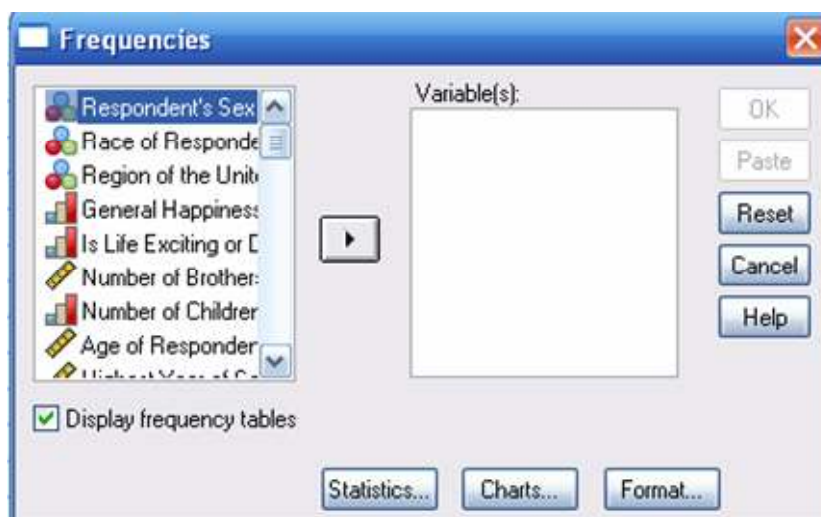
شکل ۳-۶- منوهای Data Editor



شکل ۳-۷- منوهای پنجره Viewer

### کادر های گفتگو (Dialogue Boxes)

معمولا هنگامی که گزینه ای را در یک منو انتخاب می کنید از شما خواسته می شود که اطلاعات بیشتری ارائه دهید. این کار از طریق Dialogue Box انجام می شود. به عنوان مثال هنگامی که از Spss می خواهید Frequencies بگیرد، Dialogue Box زیر ظاهر می شود و از شما می خواهد که مشخص کنید از کدام متغیر برای این کار می خواهید استفاده کنید.



شکل ۳-۸- نمونه کادر گفتگویی در SPSS

با استفاده از این Dialogue Box شما می توانید Dialogue Box های بیشتری را باز کنید که با استفاده از آنها می توانید مشخص کنید، می خواهید کدام آماره ها نمایش داده شوند، مایل هستید چه نموداری رسم شود و می خواهید خروجی به چه شکلی ارائه شود. بر حسب تحلیلی که شما می-خواهید انجام دهید گزینه های متفاوتی در Dialogue Box ها وجود دارد اما اساس کار با Dialogue Box ها یکی است که در زیر توضیح داده شده است.



## انتخاب متغیرها در کادرهای گفتگو

برای اینکه مشخص کنید از کدام متغیرها می خواهید استفاده کنید باید متغیرهای مورد نظر را در لبست ارائه شده در Dialogue Box هایلایت کنید، سپس بر روی دکمه‌ای که تصویر فلش روی آن قرار دارد کلیک کنید تا به کادر خالی که با نام Variable(s) روبروی آن قرار دارد وارد شود. برای انتخاب کردن متغیرهای می توانید آنها یک به یک انتخاب کرده و دکمه فلش را بزنید و یا اینکه گروهی را از متغیرهای را انتخاب کنید و سپس دکمه فلش را بزنید. اگر متغیرهای مورد نظر شما همه پشت سرهم قرار دارند می توانید اولی را انتخاب کنید، کلید shift را روی صفحه کلید نگاه دارید و با دکمه فلش رو به پایین که روی صفحه کلید قرار دارد، آنهایی را که می خواهید هایلایت کنید و سپس دکمه فلش روی Dialogue Box را بزنید تا همه متغیرهای انتخاب شده به قسمت Variable(s) منتقل شوند.

اگر متغیرهایی را که می خواهید انتخاب کنید در لیست پراکنده شده اند، شما باید اولین متغیر دلخواه را انتخاب کنید، سپس کلید ctrl را روی صفحه کلید نگاه دارید، نوار کنار جعبه حرکت دهید تا به متغیر دلخواه بعدی برسید و آن را انتخاب کنید و بقیه متغیرهای دلخواه را نیز می توانید این چنین انتخاب کنید. هنگامی که تمام متغیرهای مورد نظرتان را هایلایت کردید بر دکمه فلش را بزنید تا آنها به جعبه کناری منتقل شوند و در تحلیل وارد شوند.

برای خارج کردن یک متغیر از جعبه متغیرهای انتخاب شده کافی این همین فرآیند را به طور معکوس انجام دهید. بر روی متغیری در جعبه Variables که می خواهید حذف شود کلیک کنید و سپس بر روی کلید فلش کلیک کنید و این چنین متغیر مورد نظر به جعبه اولیه برگردانده می شود. اگر به جهت فلش روی دکمه دقت کنید می بینید که بسته جعبه ای که شما می خواهید متغیرها را از آن منتقل کنید تغییر می کند.

### دکمه های کادر گفتگو Dialogue box

در بیشتر Dialogue box ها شما تعدادی دکمه استاندارد می بینید (OK، Paste، Reset، Cancel و Help؛ شکل ۳-۸ را ببینید). کاربرد هر کدام از این دکمه ها چنین است:

- OK: هنگامی بر روی این دکمه کلیک کنید که متغیرهای خود را انتخاب کرده اید و آماده انجام تحلیل مورد خود توسط SPSS هستید.
- Paste: این دستور برای انتقال دستورهای برنامه نویسی است که SPSS برای تحلیلی که شما خواسته ساخته است به پنجره Syntax Editor (ویرایشگر برنامه نویسی) است. این کار هنگامی مفید است که شما می خواهید یک تحلیل را چندین بار تکرار کنید، و یا اینکه می خواهید تغییراتی در دستورات SPSS انجام دهید.

- **Reset**: این دکمه برای پاک کردن Dialogue box از تمام دستوراتی است که شما قبلاً انجام داده اید است. با انتخاب این دکمه Dialogue box آماده دریافت دستورات و تنظیمات جدید است.

- **Cancel**: با انتخاب این دکمه Dialogue box مورد نظر بسته می شود و هیچکدام از دستورات و تنظیماتی را که شما در آن Dialogue box اعمال کرده اید اجرا نمی شود.

- **Help**: با انتخاب این دکمه SPSS اطلاعاتی درباره تکنیک یا رویه ای که می خواهید اجرا کنید به شما می دهد.

با اینکه در شکل ۳-۸ از Frequencies برای نشان دادن ویژگی های Dialogue box ها استفاده شده است اما بقیه Dialogue box ها هم مشابه هستند و چنین ویژگی هایی دارند. هر چند هر Dialogue box گزینه های خاص مربوط به تحلیلی که انجام می دهد را هم دارد. این دکمه ها Dialogue box های دیگری را باز می کند و به شما امکان می دهد ویژگی های تحلیلی را که می خواهید انجام شود مشخص کنید.

### بستن SPSS

هنگامی که کار شما با SPSS تمام شد و می خواهید آن را ببندید کافی است از منو File در سمت چپ بالای صفحه Exit را انتخاب کنید. SPSS از شما خواهد پرسید که آیا می خواهید فایل داده ها و فایل خروجی خود را ذخیره کنید یا خیر. فایل داده ها دارای پسوند sav. و فایل خروجی دارای پسوند spo. است.

### استفاده از Help

اگر هنگام استفاده از SPSS نیاز به کمک داشتید و یا با برخی از گزینه ها آشنا نبودید می توانید از منو Help استفاده کنید. در نوار مربوط به منوها در بالای صفحه Help را انتخاب کنید. در این منو گزینه های مختلفی وجود دارد. می توانید درباره موضوع خاصی جستجو کنید، آموزش ببینید (tutorial)، و یا از مشاور آماری کمک SPSS استفاده کنید. این گزینه آخر امکان جالبی است که به دانشجویان و پژوهشگران راهنمایی های لازم را ارائه می کند. این امکان شما را قدم به قدم راهنمایی می کند تا تصمیم درست را اتخاذ کنید و آماره مناسب را انتخاب کنید. البته این امکان نمی تواند جایگزین کتابهای آماری باشد اما می تواند راهنمایی های مناسبی به شما ارائه دهد. این امکان همچنین شما را یاری می کند تا برخی از خروجی های خود را درک و تفسیر کنید.

در هر Dialogue box هم یک گزینه Help وجود دارد که به شما کمک می کند تکنیک آماری را که انتخاب کرده اید بهتر بشناسید. همچنین اگر بر روی دکمه های موجود در Dialogue box یک بار کلیک راست کنید کادر کوچکی ظاهر می شود که توضیحاتی خلاصه درباره آن دکمه ارائه می کند.

## بخش دوم

### تهیه فایل داده

تهیه و آماده سازی فایل داده برای تحلیل دارای مراحل است. که شامل ساختن فایل داده و وارد کردن اطلاعات به دست آمده از پژوهش به شیوه ای است که در کدنامه تعیین شده است (که در فصل دوم به آن اشاره شد). سپس باید خطاهای احتمالی رخ داده در هنگام ورود داده ها را برطرف ساخت. قسمت دوم این کتاب به این گام می پردازد. در فصل چهارم شیوه ساختن یک فایل داده و ورود داده ها مورد بحث قرار گرفته اند و در فصل ۵ بررسی و تمیز کردن داده ها توضیح داده شده اند.

## فصل چهارم

### ساختن فایل داده ها و ورود داده ها

در این فصل چگونگی ساختن یک فایل داده و ورود داده ها توضیح داده خواهند شد. این فرآیند دارای مراحل است:

- گام اول: اگر لازم است تنظیمات SPSS و چگونگی نمایش داده ها و خروجی ها بررسی و تنظیم کنید. (در اغلب موارد برای تازه کارها، این کار چندان الزامی نیست)
- گام دوم: تعیین ساختار فایل داده ها بوسیله تعریف متغیرها.
- گام سوم: ورود داده ها که شامل مقادیری می شوند که از پاسخگویان جمع آوری شده اند.

این مراحل را با استفاده از فایل داده survey.sav توضیح خواهیم داد. کدنامه ای که برای ساختن این فایل داده به کار رفته است در ضمیمه موجود است. فایل داده ها را می توان از نرم افزارهای دیگر (مانند Excel) هم وارد Spss کرد. این کار می تواند فرآیند ورود داده ها تسهیل کند به خصوص برای آنهایی که به نرم افزارهایی مانند Excel دسترسی بیشتری دارند. راهنمایی های مربوط به برای چگونگی استفاده از Excel در ورود داده در پایان این فصل ارائه شده است.

### تغییر در تنظیمات SPSS

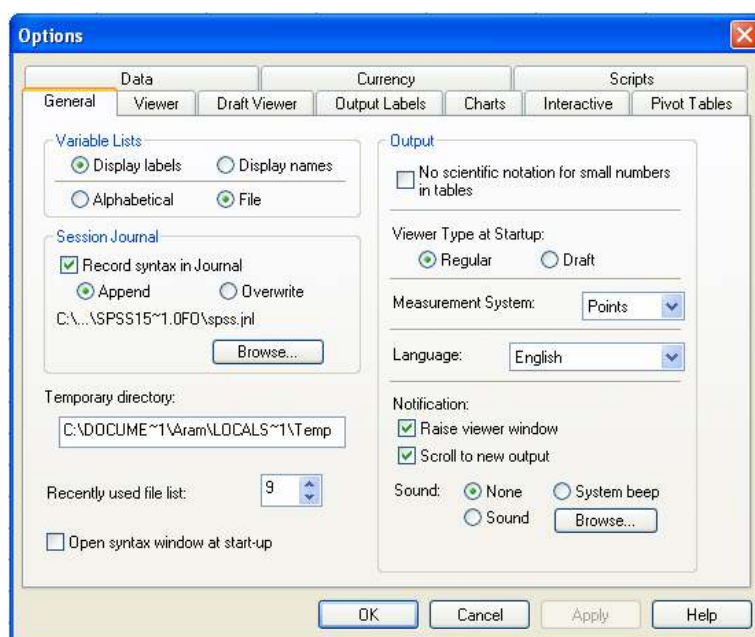
بهر آن است که قبل از ساختن فایل داده ها تنظیمات مربوط به چگونگی نمایش داده ها و خروجی ها در SPSS بررسی شوند و به شکل دلخواه تغییر یابند. این تنظیمات به شما این امکان را می دهد که انتخاب کنید متغیرهای شما چگونه نمایش داده شوند، اندازه نمودارها چقدر باشد، شکل

جدولهای خروجی چگونه باشد و بسیاری از دیگر امکانات. در ابتدا ممکن است برخی از این تنظیمات گیج کننده باشند، اما هنگامی که داده های خود را وارد کردید و برخی تحلیل ها را انجام دادید خود احساس نیاز خواهید کرد که به این بخش باز گردید.

اگر در آزمایشگاه از رایانه استفاده می کنید و کسان دیگر جز شما هم به SPSS دسترسی دارند بهتر از چگونگی این تنظیمات مطلع باشید چرا که ممکن است با تغییراتی ناخواسته در ظاهر SPSS مواجه شوید و بخواهید آن را به حالت اول باز گردانید.

بر روی منوی Edit در نوار ابزار بالای صفحه کلیک و در آن Options را انتخاب کنید.

تصویری که در شکل ۴-۱ نشان داده شده ظاهر خواهد شد.



شکل ۴-۱- مثال گزینه Options

در این شکل گزینه های بسیاری وجود دارند که بسیاری از آنها نیاز به تغییر ندارند. در ذیل برخی از گزینه های کلیدی توضیح داده شده اند. برای انتخاب هر گزینه کافی بر روی آن کلیک کنید. قبل از اینکه تمام تغییرات موردنظر خود را انجام دهید بر روی OK کلیک نکنید.

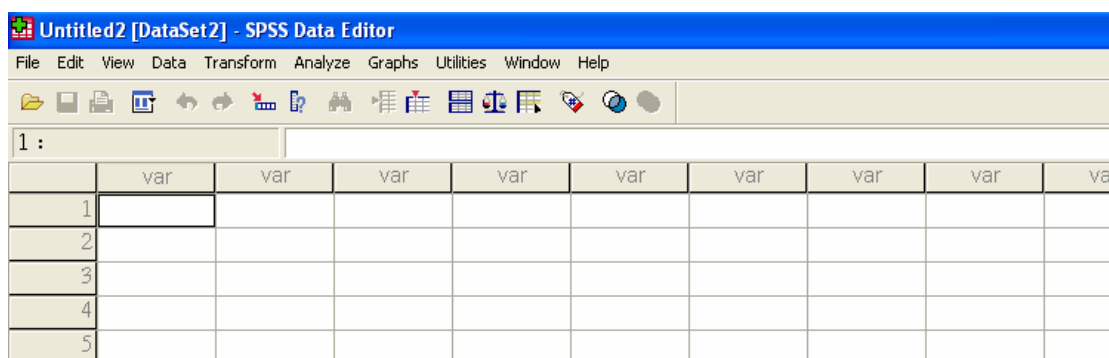
## دکمه General

هنگامی که می خواهید تحلیل های خود را انجام دهید می توانید از SPSS بخواهید که متغیرها را به ترتیب الفبائی نشان دهد و یا به همان ترتیبی که در فایل داده ها وجود دارند. معمولاً ترتیب موجود در فایل داده مرجع است چرا که متغیرهای مشابه، مثلاً آنهایی که یک طیف را می سازند و یا آنهایی که خود یک طیف هستند کنار هم قرار می گیرند. ترتیب موجود در فایل همان ترتیب متغیرها در کدنامه است و این هم مزیت دیگری برای این انتخاب است. برای اینکه انتخاب کنید متغیرها بر

اساس ترتیب موجود در فایل در تحلیلها ظاهر شوند کافی است در بخش Variable Lists بر روی File کلیک کنید. در بخش Output Notification دقت کنید که هر دو گزینه Raise Viewer و Window و Scroll to new output فعال باشند.

### تعریف متغیرها

قبل از اینکه بتوانید داده های خود را وارد کنید باید متغیرهای خود را به SPSS معرفی کنید. به این کار تعریف متغیرها می گویند. این کار در پنجره Data Editor انجام می شود (شکل ۲-۴ را ببینید). از نسخه های ۱۰ و ۱۱ به بعد این پنجره دارای دو قسمت اصلی است. یکی Data View که داده ها در آن مشاهده می کنید و دیگری Variable View که متغیرها و ویژگی های آنها را در آن مشاهده می کنید.



شکل ۲-۴- پنجره Data Editor

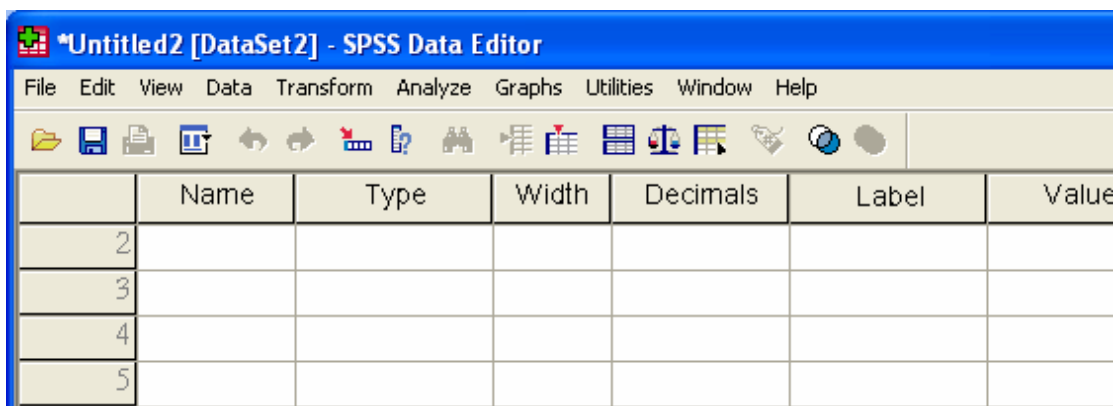
می توانید بوسیله دکمه کوچکی که در پایین سمت چپ صفحه وجود دارد از یکی از این دو به دیگری بروید.

در Variable View مشاهده می کنید که هر کدام از ستونها نام var دارند (شکل ۲-۴ را ببینید). این سرستونها با نام متغیرهایی که شما در کدنامه خود تهیه کرده اید جایگزین خواهند شد. در اول هر ردیف شماره هایی ۱ و ۲ و ۳... را مشاهده می کنید. اینها شماره هایی هستند که SPSS به هر کیس یعنی به هر ردیف اختصاص می دهد. این شماره بر شماره هایی که شما با نام شماره شناسایی به کیس های خود اختصاص داده اید لزوما یکی نیستند چرا که ممکن است شماره تغییراتی در فایل داده خود بدهید و ترتیب کیسها برهم بخورد (به عنوان مثال کیسها بر حسب یک متغیر مرتب (Sort) کنید).

## شیوه تعریف متغیرها

برای آنکه متغیرهای سازنده فایل داده خود را تعریف کنید در ابتدا باید بر روی دکمه Variable View در پایین سمت چپ صفحه کلیک کنید تا این صفحه ظاهر شود. در این صفحه (شکل ۳-۴ را ببینید) هر متغیر در یک ردیف وجود دارد و مشخصات آن روبروی آن ذکر شده است. (مشخصاتی چون name, type, width, decimals, label).

حال شما باید هر متغیر را بر اساس ویژگی‌هایی که برای آن در کدنامه ذکر شده است تعریف کنید. برخی از ویژگی‌ها را شما باید مشخص کنید (مانند نام متغیر name) و برخی دیگر را خود SPSS به طور خودکار در نظر می‌گیرد. شما می‌توانید اگر نیاز داشتید این مقادیر پیش فرض را تغییر دهید.



	Name	Type	Width	Decimals	Label	Value
2						
3						
4						
5						

شکل ۳-۴- نمایشی از Variable View

اطلاعات کلیدی را که شما باید مشخص کنید در زیر توضیح داده شده‌اند. تیرهای این توضیحات منطبق با سرستونهای صفحه Variable View است. هنگامی که بیشتر بر SPSS مسلط شدید برخی از این کارها را می‌توانید به شیوه‌ای ساده‌تر انجام دهید که بعداً توضیح داده خواهند شد. اما در ابتدا باید با اصول کار آشنا شوید.

## نام متغیر Name

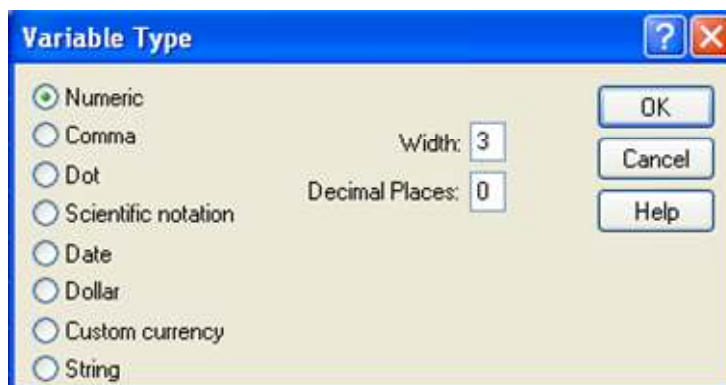
در این ستون باید نامی را که به هر متغیر اختصاص داده‌اید تا بازشناخته شود وارد کنید. نام متغیرها باید در کدنامه شما مشخص شده باشد. نام متغیرها باید دارای شرایطی باشد که در فصل دوم بیان شد. در یک فایل داده نام هر متغیر باید منحصر به فرد باشد.

## نوع متغیر Type

آنچه پیش فرض نوع متغیر در SPSS است، متغیر عددی (Numeric) است. در اغلب موارد این پیش فرض برای کار شما مناسب است. در برخی موارد هم ممکن است گزینه‌های دیگر مناسب



باشند. اگر نیاز به تغییر این ویژگی داشتید بر روی آن کلیک کنید و سپس بر روی دکمه کوچکی در همان قسمت ظاهر شده و روی آن سه نقطه وجود دارد کلیک کنید. این کار گزینه های قابل انتخاب را به شما نشان می دهد. در پنجره ای که به شما نشان داده شده می توانید تعداد رقم های اعشاری را که متغیر شما می تواند بگیرد را نیز مشخص کنید.



### عرض متغیر Width

مقدار پیش فرض برای Width برابر ۸ است. برای اکثر داده ها این مقدار کافی است. اگر متغیر قرار است مقادیر بسیار بزرگی را بگیرد می توانید این مقدار را بیشتر کنید در غیر اینصورت مقدار مناسبی است.

### رقم های اعشار Decimals

مقدار پیش فرض برای رقم های اعشار ( که قبلا در قسمت Options آن را تنظیم کردیم) صفر است. اگر متغیر شما مقادیر اعشاری به خود می گیرد می توانید این مقدار را با توجه به نیاز های خود تغییر دهید. اگر تمام متغیرهای شما می توانند مقادیر اعشاری بگیرند از قسمت Options این گزینه را متناسب با نیاز خود تغییر دهید. این کار بسیار کمتر از تغییر یک به یک متغیرها وقت شما را می گیرد.

### برچسب متغیر Label

ستون Label به شما این امکان را می دهد که توضیحات مفصلتری نسبت به نام متغیر درباره متغیر خود ارائه دهید. این توضیحات در خروجی تحلیل ها هم ذکر می شوند و به کار گرفته می شوند. به عنوان مثال ممکن است شما بخواهید به متغیر Model این برچسب را بزنید: معدل پاسخگو در آخرین ترم تحصیل.

## مقادیر متغیر Values

در ستون Values شما می توانید معنای مقادیری که متغیر شما خواهد گرفت و شما آنها را در کدنامه خود ذکر کرده اید بیاورید. به عنوان مثال برای مشخص کردن مقادیر متغیر Sex باید مراحل زیر را انجام دهید:

۱. بر روی دکمه ای که سمت راست خانه مورد نظر قرار دارد و تصویر سه نقطه روی آن است کلیک کنید. با این کار کادری باز می شود که مربوط به Values است.
۲. بر روی قسمتی که نام آن Value است کلیک کنید و بنویسید ۱.
۳. بر روی قسمتی که نام Value Label است کلیک کنید و بنویسید مرد.
۴. بر روی دکمه Add کلیک کنید. این چنین در قسمتی که نام آن Summary box است خواهید دید: مرد = 1.
۵. برای زن هم این کار را تکرار کنید: برای Value مقدار ۲ را وارد کنید و برای Value label بنویسید: زن و دکمه Add را بزنید.
۶. هنگامی که تمام مقادیر مورد نظر را وارد کردید دکمه Continue را فشار دهید.

## داده های گمشده Missing

گاهی اوقات پژوهشگران مقادیر خاصی را به داده های گمشده خود نسبت می دهند. این کار اجباری نیست. اگر مقادیر خاصی را به عنوان داده های گمشده تعریف نکنید، SPSS به طور خود کار خانه های خالی داده را به عنوان داده گمشده در نظر می گیرد. بنابراین اگر می خواهید هنگامی که با داده های گمشده مواجه شدید خانه مورد نظر را خالی بگذارید لازم نیست در این قسمت چیزی را تغییر دهید.

## ستونها Columns

به عنوان پیش فرض در SPSS عرض ستونها ۸ کاراکتر در نظر گرفته شده است. این میزان در اغلب موارد کافی اما اگر نیاز شما را بر آورده نمی کرد می توانید آن را تغییر دهید. اگر می خواهید فایل داده شما کوچکتر شود و بر روی صفحه بهتر نمایش داده شود می توانید عرض ستونها را کم کنید. فقط مطمئن شوید عرض ستون انتخابی شما به اندازه کافی است.

## محل قرار گیری Align

این قسمت شیوه قرار گیری داده ها در درون خانه ها را مشخص می کند این که چپ چین، راست چین یا وسط چین باشند. پیش فرض SPSS راست چین است و نیازی نیست که آن را تغییر دهید.

## سطح سنجش Measure

ستونی که نام آن Measure است مربوط به سطح سنجش متغیر شماست. یعنی شما باید مشخص کنید که سطح سنجش هر کدام از متغیرهای شما چیست. پیش فرض سطح سنجش Scale است که شامل سطوح سنجش فاصله ای و نسبی می شود. اگر متغیر شما مقوله ای (اسمی) است مانند جنس، می توانید Nominal را انتخاب کنید و اگر رتبه ای است می توانید Ordinal را انتخاب کنید.

## میان بر انتخابی

اگر تعداد متغیرهای شما خیلی زیاد باشند فرایندی که در بالا ذکر شد بسیار زمانبر و آزاردهنده خواهد بود. میانبرهایی وجود دارد برای اینکه بتوانید این فرآیند را کوتاه تر انجام دهید. اگر تعدادی متغیر دارید که مشخصات یکسان دارند (مشخصاتی مثل نوع، عرض، تعداد ارقام اعشاری و غیره) می توانید برای متغیر اول این ویژگی را تعریف کنید و سپس برای دیگر متغیرها آن را کپی کنید.

شیوه کپی کردن ویژگی های یک متغیر به دیگر متغیرها

۱. در Variable View بر روی خانه ای که می خواهید آن را در دیگر متغیرها کپی کنید کلیک کنید. (به عنوان مثال Width).
۲. در منوی Edit گزینه Copy را انتخاب کنید.
۳. بر روی خانه که می خواهید این ویژگی به آن منتقل شود کلیک کنید.
۴. از منوی Edit گزینه Paste را انتخاب کنید.

شیوه کپی کردن ویژگی های یک متغیر در تعدادی متغیر دیگر

۱. در Variable View بر روی خانه ای که می خواهید آن را در دیگر متغیرها کپی کنید کلیک کنید. (به عنوان مثال Width).
۲. در منوی Edit گزینه Copy را انتخاب کنید.

۳. بر روی خانه‌ای که می‌خواهید این ویژگی به آن منتقل شود کلیک کنید و با نگاه داشتن کلید سمت چپ موس خود سعی کنید دیگر خانه‌های مورد خود را نیز هایلایت کنید.
۴. از منوی Edit گزینه Paste را انتخاب کنید.

### وجود آوردن تعدادی متغیر که همه آنها ویژگی‌های مشترکی دارند

اگر در داده‌های شما طیفی وجود دارد که از تعدادی آیتم تشکیل شده که هر کدام یک متغیر هستند و همگی ویژگی‌های مشترکی دارند می‌توانید در یک گام تمام این متغیرها و ویژگی‌های آنها را بسازید. شیوه کار در کادر ذیل توضیح داده شده است. سازه مورد نظر خوشبینی (Optimism) است که قبلاً به آن اشاره شد و آیتم‌های آن op1 تا op6 نام دارند:

۱. در Variable View ویژگی‌های اولین متغیر را با استفاده از توضیحاتی که قبل از این داده شد تعریف کنید (op1). به این ترتیب که برچسب مقادیر (Value labels) آن چنین باشند: بسیار مخالف=۱، مخالف=۲، بی‌نظر=۳، موافق=۴ و بسیار موافق=۵.
۲. در Variable View بر روی شماره ردیف این متغیر کلیک کنید. با این کار کل ردیف یعنی متغیر و تمام ویژگی‌های آن باید هایلایت شوند.
۳. در منوی Edit گزینه Copy را انتخاب کنید.
۴. بر روی ردیف بعدی که خالی است کلیک کنید.
۵. در منوی Edit گزینه Paste Variable را انتخاب کنید.
۶. در کادری که با خواهد شد تعداد متغیرهایی که می‌خواهید اضافه شوند را مشخص کنید که در اینجا ۵ است. پیشوند نام متغیر (که در اینجا op است) و شماره‌ای را که می‌خواهید متغیرها از آن آغاز شوند (که در اینجا ۲ است) وارد کنید و سپس بر روی OK کلیک کنید. با این کار ۵ متغیر جدید ساخته خواهد شد (که عبارتند از op2, op3, op4, op5, op6).

برای اینکه مجموعه‌ای از دیگر متغیرها بسازید کافی همین فرآیند را تکرار کنید. به عنوان مثال اگر بخواهیم آیتم‌های مربوط به سازه اعتماد به نفس (Self-esteem) را بسازیم همین کار را برای تعریف متغیرهای sest1 تا sest10 انجام خواهیم داد. توجه کنید که این شیوه تنها برای ساختن

متغیرهایی مناسب است که ویژگی های آنها کاملاً با هم یکسان است و اگر سطح سنجش متغیرها یا برچسب مقادیر آنها متفاوت باشد این شیوه مناسب نیست.

## ورود داده ها

هنگامی که متغیرهای خود را تعریف کردید و برچسب مناسب به مقادیر آنها اختصاص داده شد (البته اگر لازم باشد) آماده ورود داده ها هستید. کدنامه خود را آماده در دسترس داشته باشید.

### فرآیند ورود داده ها

۱. برای اینکه بتوانید داده ها را وارد کنید باید Data View روی صفحه باشد. برای اینکار بر روی کلید کوچکی که در پایین سمت چپ صفحه قرار دارد و نام Data View روی آن نوشته شده کلیک کنید. صفحه سفیدی را مشاهده خواهید کرد که بر سرستونها آن نام متغیرهایی که شما تعریف کردید نوشته شده است.
۲. بر روی اولین خانه خالی در سطر اول ستون اول کلیک کنید. با این کار حاشیه آن خانه خالی پررنگ می شود تا نشان دهد این خانه اکنون آماده دریافت داده است.
۳. شماره مورد نظر را وارد کنید. اگر متغیر اول شما شماره پرسشنامه است، این شماره ۱ خواهد بود که نشان دهنده کیس یا پرسشنامه اول است.
۴. دکمه فلش به سمت راست را بر روی صفحه کلید فشار دهید. با این کار کursor به خانه دوم خواهد رفت و آماده دریافت شماره مربوط به متغیر دوم در پرسشنامه اول خواهد بود.
۵. این چنین تمام داده های مربوط به کیس اول یا پرسشنامه اول را وارد کنید و مطمئن شوید که داده ها را در خانه مناسب وارد کرده اید.
۶. برای اینکه به ستون اول باز گردید دکمه Home را بر روی صفحه کلید خود بزنید. و حال برای کیس دوم داده ها وارد کنید.
۷. اگر اشتباهی داده ای وارد کردید و خواستید آن را تغییر دهید بر خانه مورد نظر کلیک کنید. عدد مورد نظر در قسمت بالای صفحه ظاهر خواهد شد، مقدار درست را جایگزین کنید و به کار خود ادامه دهید.

بعد از اینکه متغیرهای خود را تعریف کردید و داده های خود را وارد کردید Data Editor

باید شبیه شکل ۴-۴ باشد.

survey.sav [DataSet1] - SPSS Data Editor						
File Edit View Data Transform Analyze Graphs Utilities Window Help						
1 : id 1						
	id	sex	age	marital	child	educ
1	1	1	45	4	1	4
2	2	2	21	1	2	5
3	3	2	42	4	1	2
4	4	2	47	4	1	2
5	5	2	48	4	2	3

شکل ۴-۴- مثالی برای داده ها در Data Editor

### تغییر در فایل داده ها

بعد از اینکه داده های خود را وارد کردید ممکن است بخواهید تغییراتی در فایل خود ایجاد کنید. در اینجا به برخی از تغییرات پرداخته خواهد شد.

اگر بخواهید یک کیس را حذف کنید کافی است بر روی قسمت سایه خورده در سمت چپ کیس کلیک کنید تا تمام ردیف مربوط به کیس های لایت شود آنگاه با فشردن دکمه Delete بر روی صفحه کلید کیس مورد نظر حذف خواهد شد.

اگر بخواهید بین دو کیس موجود یک کیس اضافه کنید کافی است موس را روی قسمت سایه خورده کیسی که می خواهید کیس جدید بالای آن باشد بگذارید و دکمه کلیک راست را فشار دهید و گزینه Insert case را انتخاب کنید.

اگر بخواهید یک متغیر را حذف کنید کافی است بر روی سر ستون متغیر مربوطه کلیک کنید تا تمام ستون مربوط به متغیر های لایت شود آنگاه با فشردن دکمه Delete روی صفحه کلید متغیر مورد نظر حذف خواهد شد.

اگر بخواهید یک متغیر جدید بین دو متغیر موجود ایجاد کنید کافی است موس را روی سرستونی که می خواهید متغیر جدید قبل از آن باشد قرار دهید و راست کلیک را فشار دهید و گزینه Insert variable را انتخاب کنید.

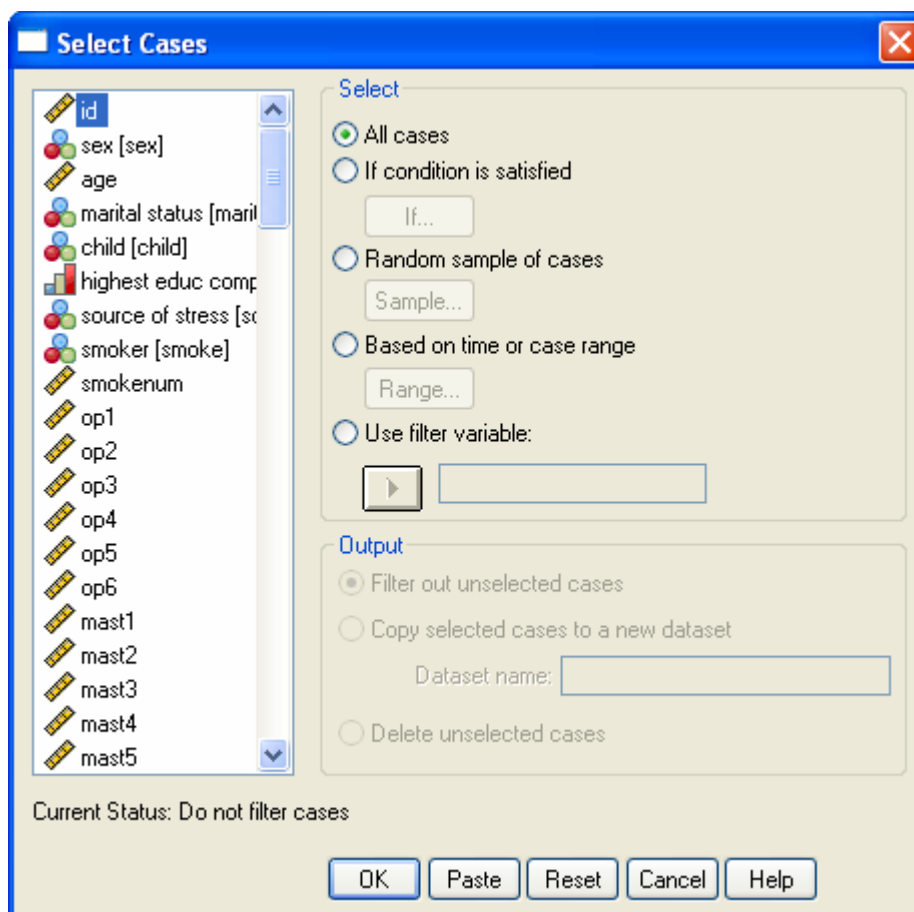
اگر بخواهید یک متغیر را جابجا کنید کافی است آن را انتخاب و سپس Copy و بعد از آن در جای مورد نظر Paste کنید.

برای آنکه کیس های خود را بر اساس یک متغیر مرتب کنید کافی است متغیر مورد نظر را انتخاب کنید، دکمه راست کلیک را فشار دهید و انتخاب کنید که می خواهید کیس ها بر اساس مقادیر نزولی (Sort descending) یا مقادیر صعودی (Sort ascending) مرتب شوند.

اگر بخواهید کیس های خود را به دو دسته تقسیم کنید که تحلیل های شما جداگانه بر روی آنها انجام شود باید از Split File استفاده کنید. برای اینکار در منوی Data گزینه Split File را انتخاب کنید. سپس Compare groups را انتخاب کنید و متغیری را می خواهید بر اساس آن کیسهای شما جدا شوند مشخص کنید (به عنوان مثال جنس) و OK را بزنید. با اینکار هنگامی که تحلیلی را انجام می دهید به طور جداگانه برای گروههای مختلف انجام خواهد (مثلا زنان و مردان).

برای اینکه Split File انجام شده را غیر فعال کنید باید دوباره کادر گفتگوی آن را باز کنید و گزینه اول را انتخاب کنید (که این جمله کنار آن نوشته شده است: Analyze all cases, do not create groups) و سپس OK را بزنید.

گاهی اوقات ممکن است بخواهید زیر مجموعه ای از داده های خود انتخاب کنید یعنی تعدادی از کیسها را جدا کنید و تحلیل های خود را تنها بر روی آنها انجام دهید. برای اینکار از منوی Data گزینه Select Cases را انتخاب کنید. در اینجا گزینه های متفاوتی وجود دارند که شیوه های متفاوتی برای انتخاب بخشی از کیسها در اختیار شما قرار می دهد. اگر گزینه If condition is satisfied را انتخاب کنید آنگاه می توانید متغیری را وارد کنید و بگوئید اگر شرط خاصی در مورد آن متغیر ارضاء شد آنگاه کیس را انتخاب کن. به عنوان مثال هنگامی که IF... را انتخاب کردید و پنجره مربوطه باز شد بنویسید: sex=1. یعنی شما از spss می خواهید هنگامی که متغیر جنس مساوی یک است (یعنی مردان) کیس مورد نظر را انتخاب کن. این چنین مردان انتخاب می شوند و زنان کنار گذاشته می شوند. همچنین شما می توانید از SPSS بخواهید زنان را فیلتر کند (یعنی این کیسها در فایل داده ها وجود داشته باشند اما در تحلیل ها دخالت داده نشوند) و یا اینکه آنان را حذف کند (یعنی از فایل داده ها حذف شوند). لازم به ذکر است که با انجام Select File به شیوه Filter به طور خودکار متغیر جدیدی ساخته می شوند که Filter نام دارد و مشخص کننده کیسهای موجود و فیلتر شده است.



همانطور که ملاحظه می کنید گزینه های دیگری هم برای شیوه انتخاب کردن وجود دارند. اگر Random Sample of cases را انتخاب کنید، می توانید نسبت دلخواهی از کیس ها را به طور تصادفی انتخاب و از بقیه جدا کنید. گزینه Based on time or case range کیس ها را بر اساس شماره آنها در فایل داده ها جدا می کند و گزینه Use filter variable هم زمانی استفاده می شود که شماره خود متغیری را به عنوان متغیر فیلتر در اختیار دارید و می خواهید از آن استفاده کنید (توجه کنید که برای آنکه توضیحات را به خوبی بفهمید حتما باید به طور آزمایشی این کارها را انجام دهید).



## فصل ۵

### بررسی و تمیز کردن داده ها

مرحله بسیار مهم قبل از آنکه تحلیل داده های خود را شروع کنید، تصحیح اشتباهات احتمالی در ورود داده ها است. اگر این مرحله را انجام ندهید ممکن است که تمام زحمات شما در مراحل بعدی تلف شود. چرا که وجود چند داده خطا می تواند تحلیل های شما را از اعتبار ساقط کند و یا منجر به نتایجی غیرمنتظره و نامعقول شود. بنابراین با حوصله بسیار باید این مرحله را به طور کامل انجام دهید.

تمیز کردن داده ها دارای سه قدم مهم است:

- قدم اول: بررسی وجود خطاها. شما باید هر کدام از متغیرها را بررسی کنید که آیا داده های خارج از دامنه در آن وجود دارد یا خیر. یعنی داده هایی که خارج از محدوده مورد نظر هستند.
- گام ۲: پیدا کردن خطاها در فایل داده. باید کیسی که خطای مشاهده شده در یک متغیر مربوط به آن است بیابید.
- گام ۳: داده مورد نظر را تصحیح کنید.

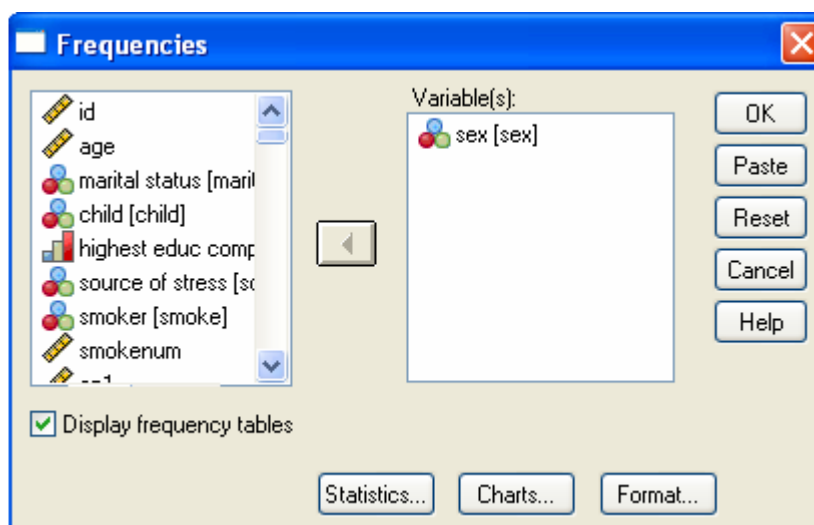
### گام اول: بررسی متغیرها

فرض کنید که می خواهید ببینید در ورود داده های مربوط به متغیر جنس خطا وجود دارد یا خیر. انتظار داریم که تمام داده های مربوط به این متغیر ۱ یا ۲ باشند چرا که خود هنگام تعریف متغیر گفته ایم ۱ عدد مربوط به مردان و ۲ عدد مربوط به زنان است. بنابراین هر عددی غیر از ۱ و ۲ که مشاهده شود، خطا خواهد بود.

برای بررسی وجود خطاها باید از آماره های توصیفی و جداول فراوانی استفاده کرد.

اگر متغیر شما مقوله ای است (هم متغیرهای اسمی و هم متغیرهای ترتیبی)، در Analyze گزینه Descriptive statistics و سپس Frequencies را انتخاب کنید. متغیری را که می خواهید

بررسی کنید وارد کنید و در گزینه Statistics آماره های ماکزیمم و مینیمم را انتخاب کنید و OK را بزنید.



خروجی شما باید چنین باشد:

**Statistics**

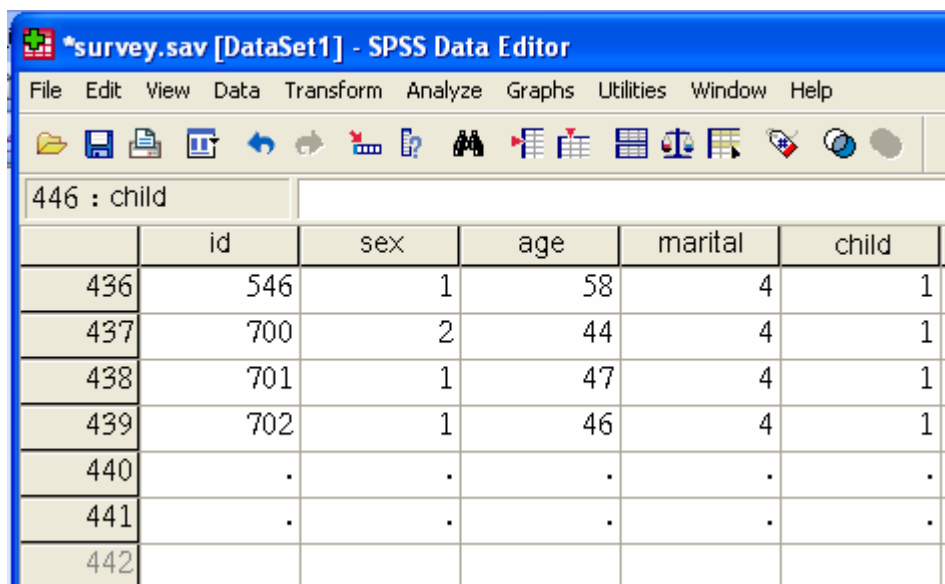
sex		
N	Valid	43 9
	Missing	0
Minimum		1
Maximum		2

sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALES	18 5	42.1	42.1	42.1
	FEMALES	25 4	57.9	57.9	100.0
	Total	43 9	100.0	100.0	

جدول اول آماره های توصیفی که انتخاب کرده ایم ارائه داده است که شامل ماکزیمم و مینیمم می شوند. انتظار داریم که ماکزیمم داده ها برای متغیر جنس، ۲ و مینیمم آن ۱ باشد. جدول را بررسی می کنیم و می بینیم چنین است. پس تا اینجا خطایی ملاحظه نکرده ایم. در این جدول تعداد داده های گمشده (Missing) هم ارائه شده است که در اینجا صفر است یعنی همه به سوال جنس پاسخ داده اند. اگر تعداد داده های گمشده شما زیاد و بیش از حد انتظار بود باید به دنبال علت بگردید.

ممکن است در انتهای کیسهای موجود در فایل داده ها، تعدادی سطر خالی فعال شده باشند و به کیسهای شما اضافه شده باشند. شکل زیر را ببینید:



	id	sex	age	marital	child
436	546	1	58	4	1
437	700	2	44	4	1
438	701	1	47	4	1
439	702	1	46	4	1
440	.	.	.	.	.
441	.	.	.	.	.
442					

در این شکل ملاحظه می کنید که کیسهای ۴۴۰ و ۴۴۱ فعال شده اند (رنگ سررديف آنها پررنگ شده است) در حالیکه هیچ داده ای برای آنها وارد نشده اند و اصلاً جزء کیسهای ما نیستند. در اینجا باید آنها را حذف کنیم (Delete).

جدول دوم جدول فراوانی متغیر جنس است. ما انتظار داریم تنها دو نوع پاسخ در این جدول ببینیم، مردان (Males) و زنان (Females). با مشاهده این جدول می بینیم که چنین است، پس خطایی برای متغیر جنس نیافته ایم. اگر جدول مشابه شکل زیر بود:

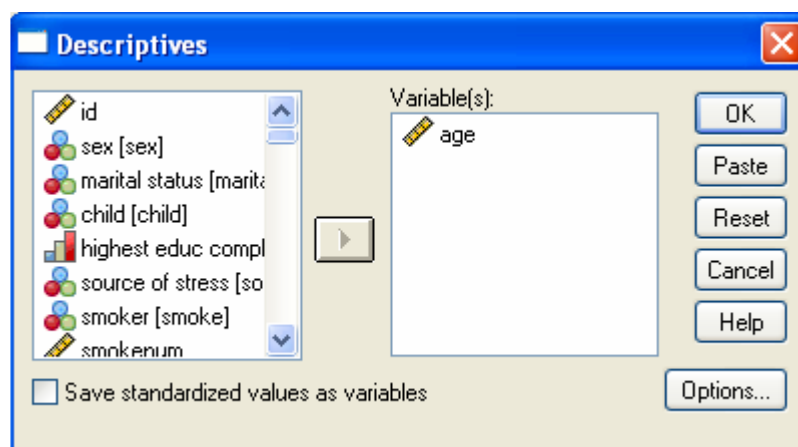
sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALES	184	41.9	41.9	41.9
	FEMALES	254	57.9	57.9	99.8
	3	1	.2	.2	100.0
	Total	439	100.0	100.0	

آنگاه در می یافتم که خطایی در داده ها وجود دارد، چرا که ما غیر مردان و زنان پاسخ دیگری هم داریم که مقدار آن ۳ است و ما انتظار چنین مشاهده ای را نداشته ایم. پس با توجه به جدول بالا در متغیر جنس یک داده (چون فراوانی آن ۱ است) وجود دارد که به خطا ۳ وارد شده است.

اگر داده های شما پیوسته باشند (متغیرهای فاصلهای و نسبی) همین فرآیند برای بررسی متغیرها باید انجام شود اما با این تفاوت که در اینجا دیگر جدول فراوانی کمکی به ما نخواهد کرد و

تنها باید از آماره های توصیفی استفاده کنیم. برای آنکه آماره های توصیفی را به تنهایی بگیریم ابتدا در منوی Analyze گزینه Descriptive statistics انتخاب و در آن Descriptives بر می گزینیم.



متغیر مورد نظر را وارد می کنیم و از گزینه Options... آماره هایی را هم که می خواهیم محاسبه شوند انتخاب می کنیم. ما مینیمم، ماکزیمم، میانگین و انحراف معیار را انتخاب کرده ایم. خروجی چنین خواهد بود:

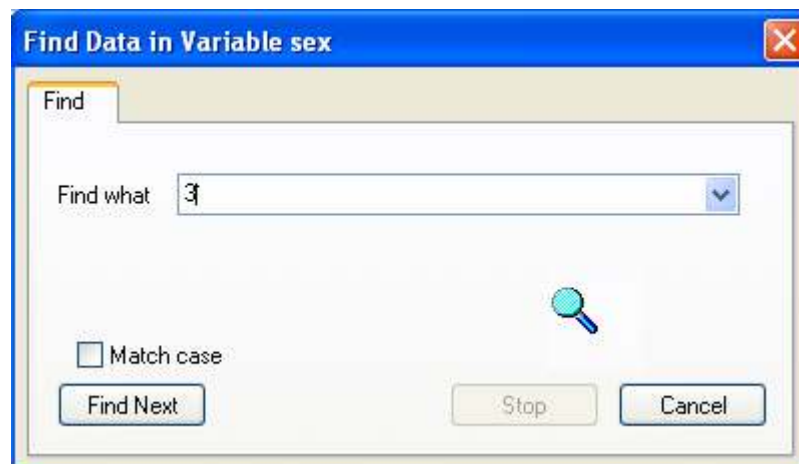
**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
age	439	18	82	37.44	13.202
Valid N (listwise)	439				

حال ماکزیمم و مینیمم را بررسی کنید. اگر مقادیر آنها غیرمنطقی باشند نشان دهنده وجود خطاست. به عنوان مثال اگر ماکزیمم سن برابر ۱۲۳ باشد و یا هنگامی که مینیمم آن ۱۰- باشد، نشان دهنده خطاست. میانگین و انحراف معیار را هم بررسی می کنیم. اگر متغیر مورد بررسی یک طیف باشد که قبلا هم در جامعه آماری ما اجراء شده است می توانیم میانگین خود را با میانگین پژوهش قبلی مقایسه کنیم و اگر تفاوت فاحشی دیدیم باید در صحت داده های خود شک کنیم.

### گام دوم: یافتن محل خطاها

حال اگر دریافتیم مقداری به خطا در متغیری وارد شده (به عنوان مثال عدد ۳ برای جنس وارد شده) در این مرحله باید کیسی را که داده خطا مربوط به اوست بیابیم. برای این کار ابتدا متغیر مورد نظر را هایلایت می کنیم (در اینجا متغیر جنس) این کار بوسیله کلیک کردن روی سرستون در پنجره Data View صورت می گیرد. سپس در منوی Edit گزینه Find را انتخاب می کنیم که کادر زیر ظاهر می شود:



عددی را به دنبال آن هستیم وارد می کنیم و دکمه Find Next را می زنیم. با این کار SPSS کیس مورد نظر را را نشان خواهد داد. می توانیم این کار را تکرار کنیم تا دیگر کیسها هم شناسایی شوند.

### گام سوم: تصحیح مقادیر خطا

حال باید شما مقدار صحیح را به جای خطا وارد کنید. برای پیدا کردن مقدار صحیح می توانید با توجه به شماره پرسشنامه به پرسشنامه مربوطه مراجعه کنید و مقدار صحیح را بیابید. گاهی اوقات که دسترسی به پرسشنامه های اصلی وجود ندارد، پژوهشگران ترجیح می دهند که مقدر خطا را حذف کنند تا داده مربوطه به عنوان گمشده (Missing) در نظر گرفته شود.

## فصل ششم

### آماره های توصیفی

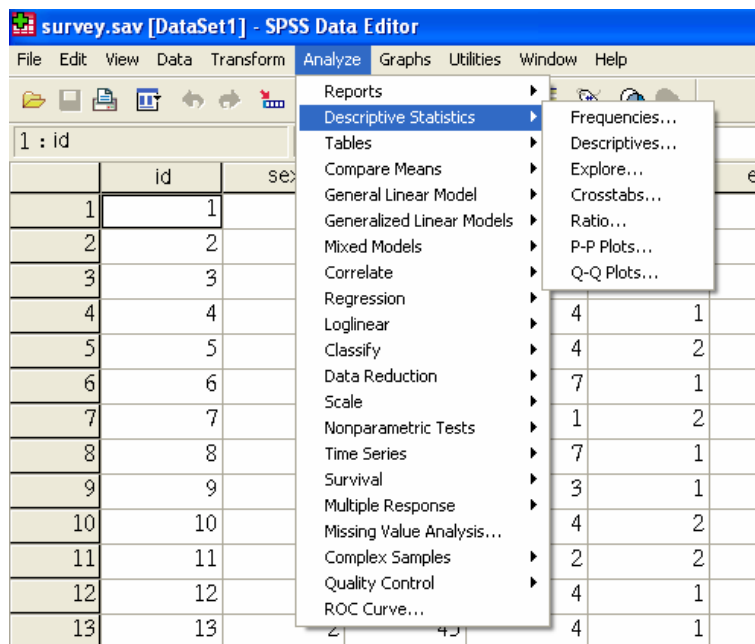
هنگامی که مطمئن شدید خطاهای موجود در خطاها را یافته اید و آنان را رفع کرده اید، می-توانید مرحله توصیفی تحلیل داده های خود را آغاز کنید. آماره های توصیفی کاربردهایی دارند که به شرح زیر هستند:

- توصیف ویژگی های نمونه در بخش روش گزارش پژوهش شما.
- برای بررسی مفروضاتی که لازمه استفاده از برخی از تکنیکهای آماری است.
- برای پاسخ دادن به برخی از سوالهای پژوهش.

فرآیندی که برای پیدا کردن خطاهای موجود در داده ها در فصل قبل بیان شد هم می توانند اطلاعاتی را درباره ویژگی های نمونه به شما بدهد تا در بخش روش گزارش پژوهش خود ذکر کنید. در پژوهش هایی که اطلاعات آنها از انسانها جمع آوری شده اند خوب است که حجم نمونه، تعداد و درصد مردان و زنان در نمونه، دامنه تغییرات و میانگین سن، میزان تحصیلات و دیگر اطلاعات زمینه‌ای ذکر شوند.

قبل از انجام بسیاری از تحلیل های آماری (مانند t-test، ANOVA، همبستگی) باید مطمئن شوید که مفروضات مربوط به این تحلیل ها زیر پا نگذاشته اید (به این مفروضات به تفصیل در بخش چهارم و پنجم کتاب پرداخته شده است). برای بررسی این مفروضات معمولاً باید آماره های توصیفی را به دست آورید. این آماره های توصیفی عبارتند از میانگین، انحراف معیار، دامنه نمرات، چولگی و کشیدگی.

آماره های توصیفی را در SPSS را می توان به شیوه های گوناگونی به دست آورد، با استفاده از گزینه های Descriptives، Frequencies یا Explore. این رویه ها در منوی Analyse و گزینه Descriptive Statistics وجود دارند. اما این رویه ها بسته به اینکه متغیر مورد نظر مقوله ای یا پیوسته باشد، با هم متفاوت هستند. اگر متغیر مورد نظر شما مقوله ای باشد برخی آماره ها (مانند میانگین، انحراف معیار) مناسب نخواهند بود. در دو بخشی که ارائه خواهند شد رهیافتهای متفاوت برای متغیرهای مقوله ای و پیوسته ارائه شده اند.



### متغیرهای مقوله ای

برای آنکه برای متغیرهای مقوله آماره های توصیفی را به دست آوریم باید به سراغ گزینه Frequencies برویم. به این ترتیب خواهیم فهمید که چند نفر هر پاسخ را انتخاب کرده اند (به عنوان مثال چند نفر مرد و چند نفر زن). محاسبه میانگین، انحراف معیار و مانند آنها برای متغیرهای مقوله ای مانند جنس، وضعیت تاهل اصلاً تناسبی ندارد.

رویه گرفتن آماره های توصیفی برای متغیرهای مقوله ای

۱. در منوی Analyse گزینه Descriptive Statistics را انتخاب و پس از آن Frequencies را برگزینید.
۲. متغیری مقوله ای را که مایل هستید توصیف کنید (مانند جنس) انتخاب کنید و وارد کادر مربوطه سازید.
۳. بر روی دکمه Statistics کلیک کنید. در بخش مربوط به آماره های پراکندگی (Dispersion) بیشترین (Maximum) و کمترین (Minimum) را انتخاب کنید. دکمه های Continue و سپس OK را فشار دهید.

خروجی این رویه چنین خواهد بود:

### Statistics

sex		
N	Valid	439
	Missing	0
Minimum		1
Maximum		2

### sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALES	185	42.1	42.1	42.1
	FEMALES	254	57.9	57.9	100.0
	Total	439	100.0	100.0	

### تفسیر خروجی Frequencies

از خروجی بالا ما در می یابیم که ۱۸۵ مرد (۴۲/۱ درصد) و ۲۵۴ زن (۵۷/۹ درصد) در این نمونه وجود دارند و کل آنها ۴۳۹ نفر هستند. توجه به تعداد پاسخگویان در زیر گروههای گوناگون نمونه شما مهم است. در برخی تحلیل ها (مانند ANOVA) بهتر است که تعداد افراد در گروهها تقریباً مساوی باشند. هنگامی که اندازه ها بسیار نابرابر است، به خصوص در زمانی که حجم نمونه شما کم است برخی تحلیل ها نمی توان انجام داد.

### متغیرهای پیوسته

برای متغیرهای پیوسته مانند سن استفاده از Descriptives آسانتر است. چون آماره های تلخیصی مانند میانگین، میانه، انحراف معیار را در اختیار شما می گذارد. ارائه جدول فراوانی برای متغیرهای پیوسته معمولاً چندان موضوعیتی ندارد چرا که چنین جدولی طولانی و کم محتوا خواهد بود. شما می توانید آماره های مورد نظر برای همه متغیرهای پیوسته را در یک قدم به دست آورید و لازم نیست این کار را متغیر به متغیر انجام دهید. کافی است متغیرهای مورد نظر خود را وارد قسمت Variables کنید. البته اگر تعداد متغیرهای شما زیاد باشد خروجی شما طولانی خواهد شد. به این ترتیب شاید بهتر باشد برای هر چند تا متغیر با هم آماره های توصیفی را بگیرید.



رویه به دست آوردن آماره های توصیفی برای متغیرهای پیوسته:

۱. در منوی Analyse گزینه Descriptive Statistics را انتخاب و پس از آن Descriptives را برگزینید.
۲. تمام متغیرهایی را که می خواهید برای آنها آماره های توصیفی را به دست آورید وارد قسمت Variables کنید. مانند سن، کل استرس ادراک شده و غیره.
۳. در گزینه Options آماره های mean (میانگین)، standard deviation (انحراف معیار)، بیشترین (Maximum)، کمترین (Minimum)، چولگی (skewness) و کشیدگی (kurtosis) را انتخاب کنید.
۴. دکمه های Continue و سپس OK را فشار دهید.

خروجی این رویه چنین خواهد بود:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
age	439	18	82	37.44	13.202	.606	.117	-.203	.233
total perceived stress	433	12	46	26.73	5.848	.245	.117	.182	.234
total optimism	435	7	30	22.12	4.429	-.494	.117	.214	.234
total mastery	436	8	28	21.76	3.970	-.613	.117	.285	.233
total PCOISS	430	20	88	60.63	11.985	-.401	.118	.257	.235
Valid N (listwise)	425								

### تفسیر خروجی Descriptives

در جدول بالا اطلاعات خواسته درباره هر متغیر ارائه شده است. به عنوان مثال به متغیر سن توجه کنید. داده های مربوط به سن ۴۳۹ پاسخگو وجود دارد. دامنه تغییرات سن پاسخگویان بین ۱۸ تا ۸۲ بوده است که میانگین آن ۳۷/۴۴ و انحراف معیار آن ۱۳/۲۰ بوده است. این اطلاعات می تواند در بخش روش گزارش پژوهش برای توصیف نمونه مورد استفاده قرار گیرد.

بخشی از آماره های توصیفی درباره توزیع متغیر اطلاعاتی به ما می دهند (چولگی و کشیدگی). این اطلاعات زمانی مورد نیاز خواهد بود که شما می خواهید از تحلیل های پارامتری استفاده کنید (به عنوان مثال t-test و آنالیز واریانس). مقدار چولگی شاخصی برای میزان تقارن توزیع است. کشیدگی هم شاخصی برای میزان بلندی قله توزیع (بالا تر یا پایین تر از توزیع نرمال بودن) است. اگر توزیع کاملاً نرمال باشد مقادیر کشیدگی و چولگی صفر خواهند شد که البته در علوم اجتماعی اتفاقی نادر است.

اگر مقدار شاخص چولگی مثبت باشد نشان دهنده وجود چولگی مثبت توزیع متغیر مورد نظر است (یعنی داده ها بیشتر حول مقادیر کمتر انباشته شده اند) و مقادیر منفی شاخص چولگی هم نشان

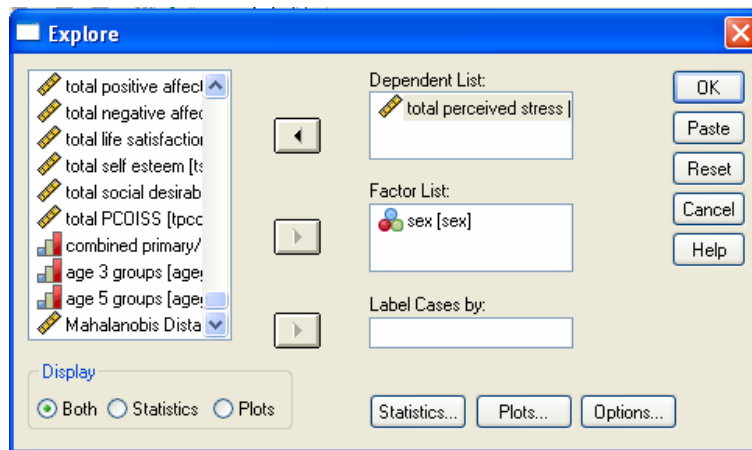
دهنده آن است که توزیع داده چولگی منفی است (یعنی داده ها بیشتر در سمت راست توزیع مجتمع شده اند). کشیدگی مثبت نشان دهنده آن است که داده ها در مرکز مجتمع شده اند و قله بلندی را ساخته اند که دنباله های آن باریک اند. مقادیر منفی برای شاخص کشیدگی هم نشان دهنده آن است که داده ها کمتر در مرکز مجتمع شده اند و قله کوتاهی در توزیع وجود دارد که دارای دنباله های ضخیم است. اگر نمونه به اندازه کافی بزرگ باشد چولگی مشکلی برای تحلیل بوجود نخواهد آورد. کشیدگی موجب کمتر تخمین زدن واریانس خواهد شد که البته این امر هم با بزرگ بودن نمونه کمتر خواهد شد (حجم نمونه بیشتر از ۲۰۰ باشد).

آزمونهای برای بررسی چولگی و کشیدگی وجود دارند، اما این نسبت به بزرگ بودن نمونه بسیار حساس هستند یعنی با بزرگ بودن حجم نمونه مشکل حل خواهد شد. پیشنهاد شده است که برای بررسی توزیع، شکل آن را ببینیم مثلاً هیستوگرام رسم کنیم. شیوه بررسی دقیقتر نرمال بودن توزیع در بخش بعدی توضیح داده شده است.

### بررسی نرمال بودن

بسیاری از تکنیکهایی که در بخشهای چهارم و پنجم این کتاب ارائه شده اند فرض می کنند که توزیع نمرات متغیر وابسته نرمال است. توزیع نرمال متقارن و زنگی شکل است و بیشترین فراوانی آن در میانه توزیع قرار دارد و در دو سوی آن داده ها به یکسان توزیع شده اند. نرمال بودن توزیع را تاحدی می توان با محاسبه چولگی و کشیدگی بررسی کرد. اما تکنیکهای دیگری هم برای انجام این کار در گزینه Explore در Descriptive Statistics وجود دارد که توضیح داده خواهند شد.

در این مثال ما می خواهیم نرمال بودن متغیر استرس ادراک شده (Total perceived stress) را بررسی کنیم. این کار را برای زنان و مردان به طور جداگانه انجام داده ایم (برای این کار از قسمت Factor List در کادر Explore استفاده کرده ایم). کاری که ما انجام می دهیم مقدمه ای برای انجام t-test برای بررسی تفاوت استرس ادراک شده در میان زنان و مردان است (برای توضیحات بیشتر درباره نمونه های مستقل در t-test به فصل پانزدهم مراجعه کنید). اگر می خواهید نرمال بودن متغیر استرس ادراک شده را به طور کلی بررسی کنید، آنچه در زیر درباره استفاده از Factor List گفته شده است انجام ندهید.



شیوه بررسی نرمال بودن توزیع با استفاده از Explore:

۱. در منوی Analyse گزینه Descriptive Statistics را انتخاب و پس از آن Explore را برگزینید.
۲. متغیری را که می خواهید نرمال توزیع آن را بررسی کنید انتخاب کنید و در قسمت Dependent List Box وارد کنید. در اینجا ما متغیر استرس ادراک شده ( Total perceived stress ) را وارد کرده ایم.
۳. متغیری را که می خواهید بر اساس داده های شما گروه بندی یا تقسیم شود وارد قسمت Factor List کنید. در اینجا ما متغیر جنس ( Sex ) را وارد کرده ایم.
۴. در قسمتی که Display نام دارد گزینه Both ( یعنی هردو ) را انتخاب کنید. با این کار نمودارها و آماره ها هردو نمایش داده خواهند شد.
۵. دکمه Plots را فشار دهید و سپس در قسمت Descriptives آن Histogram را انتخاب کنید. Normality plots with tests را هم انتخاب کنید و Continue را انتخاب کنید.
۶. دکمه Options را فشار دهید. در قسمت Missing Values گزینه Exclude cases pairwise را انتخاب کنید. دکمه Continue و سپس Ok را بزنید.

خروجی این کار چنین خواهد بود:

### Descriptives

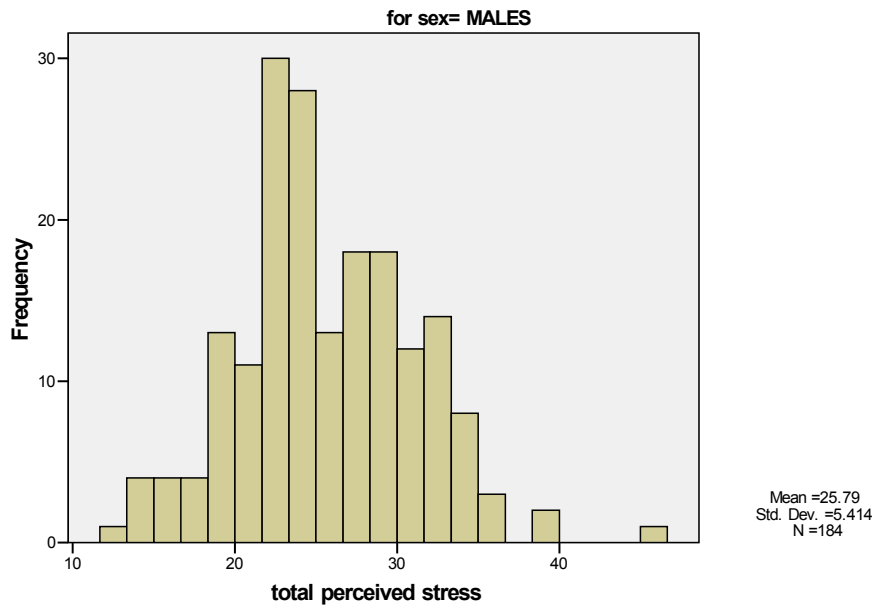
sex				Statistic	Std. Error
total perceived stress	MALES	Mean		25.79	.399
		95% Confidence Interval for Mean	Lower Bound	25.00	
			Upper Bound	26.58	
		5% Trimmed Mean		25.74	
		Median		25.00	
		Variance		29.315	
		Std. Deviation		5.414	
		Minimum		13	
		Maximum		46	
		Range		33	
		Interquartile Range		8	
		Skewness		.271	
		Kurtosis		.393	
					.179
					.356
	FEMALES	Mean		27.42	.384
		95% Confidence Interval for Mean	Lower Bound	26.66	
			Upper Bound	28.18	
		5% Trimmed Mean		27.35	
		Median		27.00	
		Variance		36.793	
		Std. Deviation		6.066	
		Minimum		12	
		Maximum		44	
		Range		32	
		Interquartile Range		7	
		Skewness		.173	
		Kurtosis		.074	
					.154
					.307

### Tests of Normality

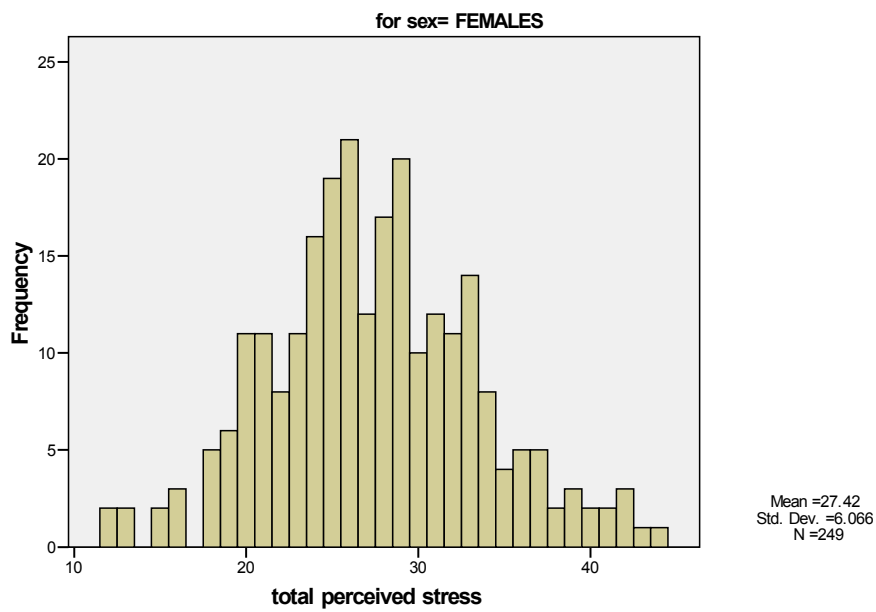
sex		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
total perceived stress	MALES	.074	184	.015	.987	184	.096
	FEMALES	.064	249	.015	.992	249	.176

a. Lilliefors Significance Correction

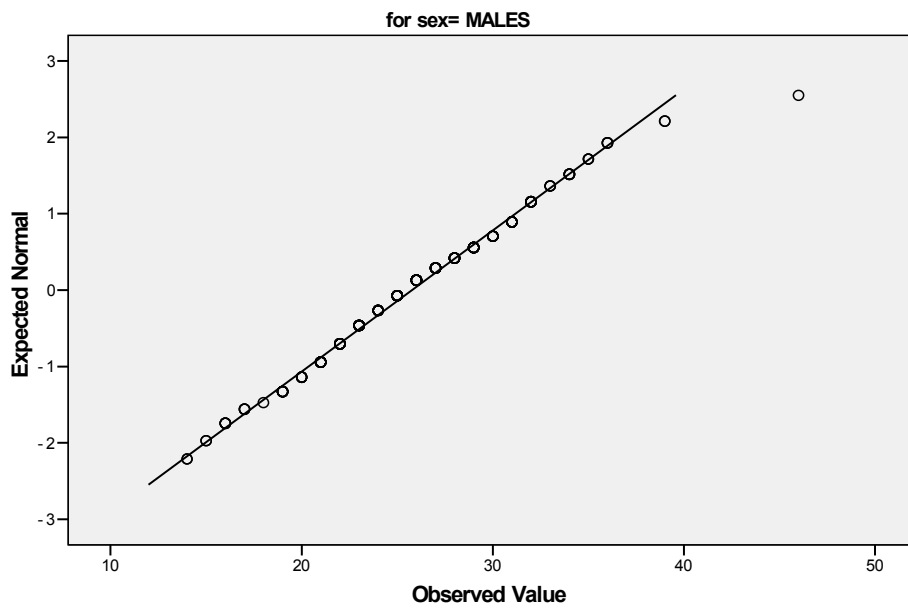
### Histogram



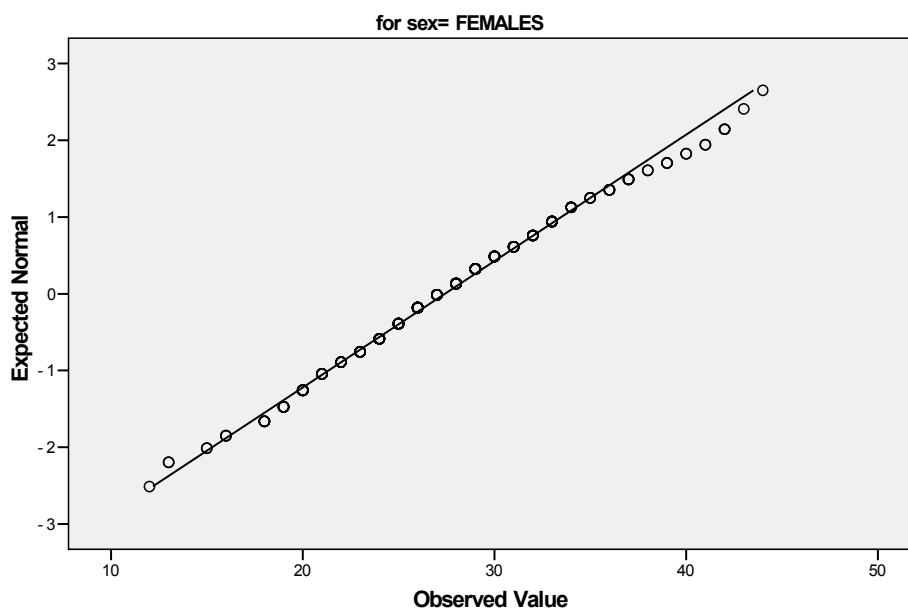
### Histogram



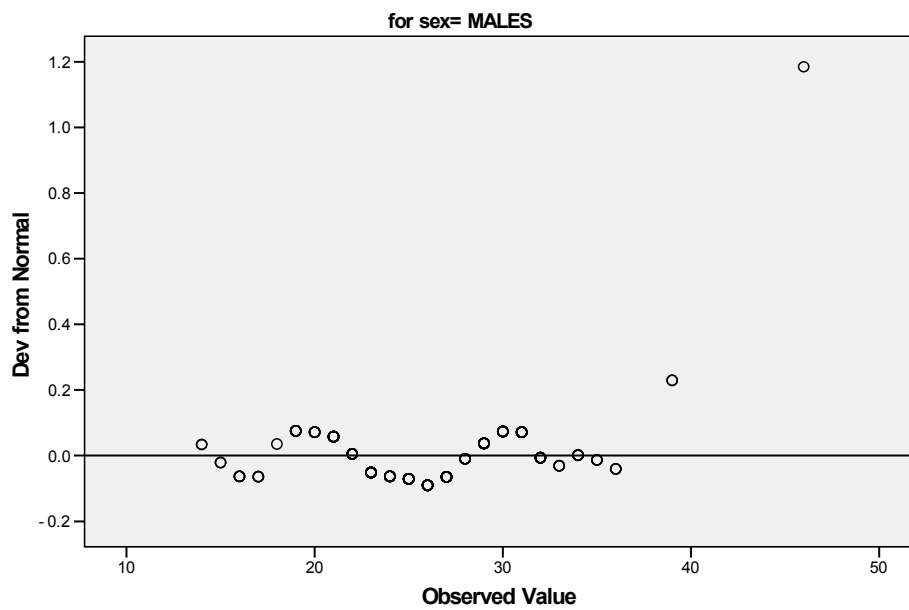
Normal Q- Q Plot of total perceived stress



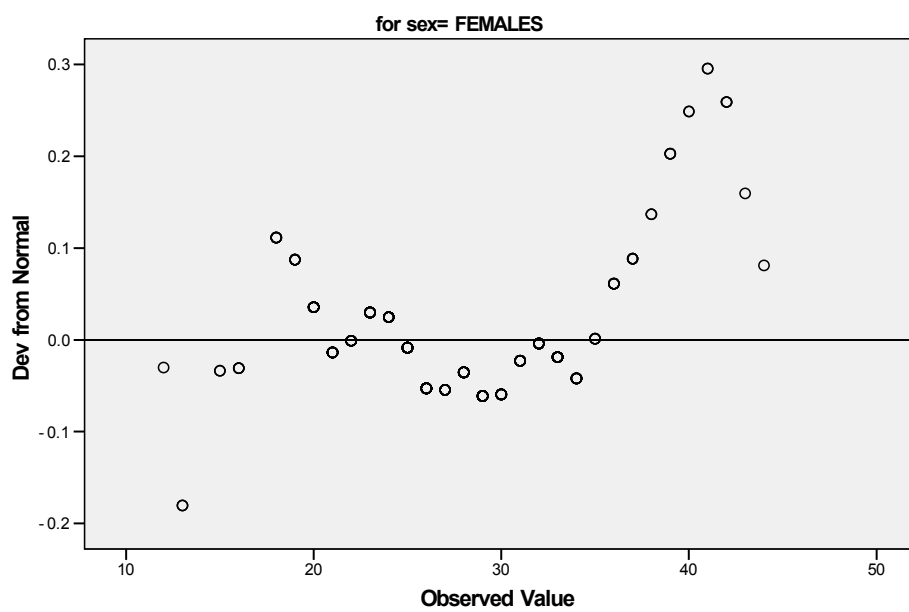
Normal Q- Q Plot of total perceived stress

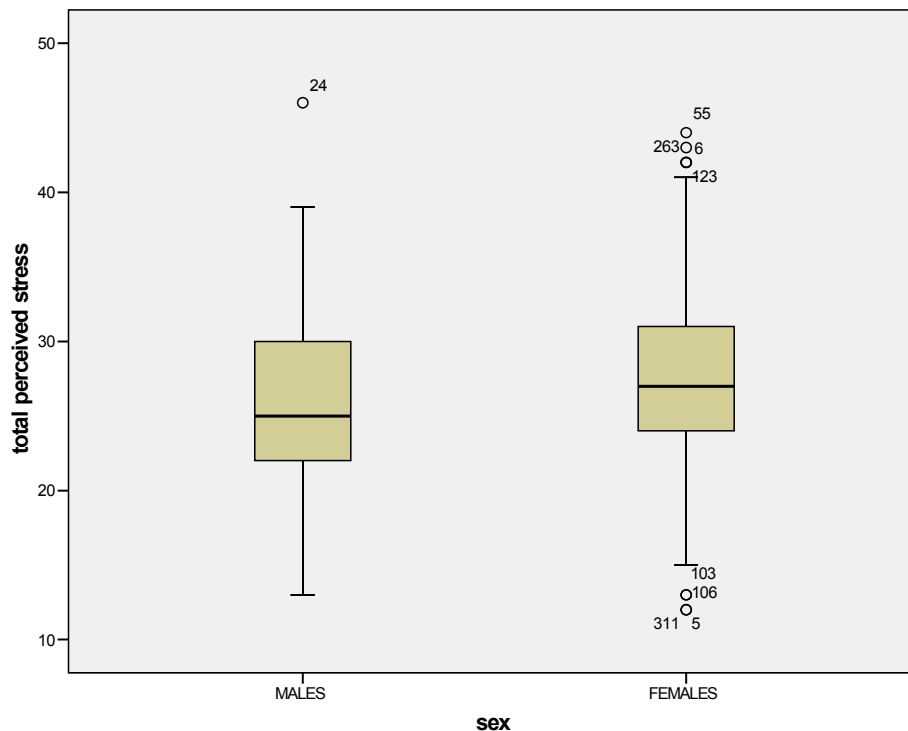


Detrended Normal Q- Q Plot of total perceived stress



Detrended Normal Q- Q Plot of total perceived stress





### تفسیر خروجی های بدست آمده از Explore

خروجی های بدست آمده در این قسمت بسیار طولانی شد و اگر ندانید که به دنبال چه هستید تفسیر آن ها سخت خواهد بود. در اینجا قدم به قدم پیش خواهیم رفت:

- در جدولی که Descriptives نام دارد، آماره های توصیفی و دیگر اطلاعات مربوط به متغیرهای شما ارائه شده اند. در اینجا این جدول به دو بخش تقسیم شده است که مربوط به دو گروه مردان و زنان است. اگر شما در قسمت Factor List متغیری را وارد نکنید، این اطلاعات برای کل داده ها یک جا داده خواهد شد و به تفکیک زنان و مردان نخواهد بود. شما برخی از این اطلاعات را می شناسید (میانگین، میانه، انحراف معیار، مینیمم، ماکزیمم و غیره). آنچه که شما احتمالاً با آن آشنا نیستید، میانگین پیراسته ۵٪ (5 Trimmed Mean) است. SPSS برای محاسبه این آماره ۵ درصد داده ها را از بالا و پایین داده ها حذف می کند و سپس میانگین داده های باقیمانده را محاسبه می کند. اگر مقدار میانگین پیراسته را با میانگین داده های اصلی مقایسه کنید و تفاوت زیادی را مشاهده کنید آنگاه می توانید نتیجه بگیرید که در داده های شما، مقادیر دورافتاده (چه مقادیر کم و چه زیاد) تاثیر زیادی دارند. به عبارت دیگر اگر میانگین پیراسته از میانگین اصلی بسیار کمتر باشد، آنگاه نتیجه می گیریم که تعدادی اندک از مقادیر بسیار کوچک، نسبت به دیگر مقادیر، در داده ها وجود دارند که میانگین اصلی را به شدت تحت تاثیر حضور خود قرار می دهند و



اگر میانگین پیراسته به دست آمده بسیار بزرگتر از میانگین اصلی باشد آنگاه می-توانیم نتیجه که تعداد اندکی مقادیر بسیار بزرگ نسبت به بقیه داده ها وجود دارند که میانگین اصلی را تحت تاثیر حضور خود قرار داده اند. اگر داده های دورافتاده تاثیر زیادی بر آماره های شما داشتند، آنها را بیشتر مورد بررسی قرار دهید چرا که ممکن است بقیه تحلیلهای شما هم را به شدت تحت تاثیر قرار دهند.

- مقادیر چولگی (Skeness) و کشیدگی (Kortosis) هم در بخشی از این جدول داده شده اند. این دو آماره درباره توزیع داده ها، اطلاعاتی را به ما می دهند (در بخش قبل درباره معنی مقادیر این آماره ها بحث شده است).
- در جدولی که Tests of Normality نام دارد نتایج آماره کولموگروف-اسمیرنوف (Kolmogrov-Smirniv) داده شده اند. این آماره نرمال بودن توزیع را بررسی می کند. اگر نتیجه غیر معنی دار باشد (یعنی مقدار معنی داری آن بیش از ۰/۰۵ باشد) معنی آن این است که توزیع نرمال است. در مورد این داده ها مقدار معنی داری برای هر گروه ۰/۰۱۵ به دست آمده است که نشان دادن تخطی از فرض نرمال بودن است. در نمونه های بزرگتر چنین نتیجه ای کاملاً معمول است.
- شکل واقعی توزیع داده ها در گروه (زنان و مردان) را می توان در هیستوگرام ها (Histogram) مشاهده کرد. در این مثال برای هر دو گروه، داده ها تا حد قابل قبولی توزیعی نرمال دارند. این موضوع را می توان با بررسی نمودارهای احتمال نرمال هم مورد بررسی قرار داد (نمودارهایی که Normal Q-Q Plots نام دارند). در این نمودارهای مقادیر مشاهده شده در مقابل مقادیر مورد انتظار در توزیع نرمال رسم شده اند. خط تقریباً مستقیم نشان دهنده نرمال بودن توزیع است.
- نمودارهایی که Detrended Normal Q-Q نام دارند، حاصل رسم مقادیر مشاهده شده در مقابل انحراف مقادیر از توزیع نرمال هستند. در این نمودار نباید دسته بندی خاصی در داده ها مشاهده شود و اکثر آنها باید حول خط صفر توزیع شده باشند.
- نمودار آخری که در خروجی ارائه شده اند، نموداری جعبه (box plot) توزیع داده ها در دو گروه است. مستطیل نشان دهنده فاصله میان چارک اول و چارک سوم است (که ۵۰ درصد داده ها در آن فاصله قرار دارند). دو بال بالا و پایین مستطیل هم نشان دهنده مقادیر مینیمم و ماکزیمم داده ها است.

## بررسی داده های دورافتاده

بسیاری از تحلیل های آماری که در این کتاب بررسی می شوند نسبت به داده های دورافتاده (outliers) حساس هستند. داده های دورافتاده مقادیری هستند که بسیار بالاتر یا پایینتر از اکثریت داده ها قرار دارند. تکنیکهایی که در بخش قبل توضیح داده شدند را می توان برای بررسی داده های دورافتاده هم به کار برد. در این زیر رهیافت دیگری هم به این منظور توضیح داده شده است. قبلا این رهیافت را برای بررسی داده های اشتباه به کار برده ایم.

## شیوه تشخیص داده های دورافتاده

۱. در داخل منوی Analyze، گزینه Descriptive Statistics و سپس گزینه Explore را انتخاب کنید.
  ۲. در قسمت Display مطمئن شوید که گزینه Both انتخاب شده است. انتخاب این گزینه موجب می شود که Spss هم آماره ها و هم نمودارها را ارائه دهد.
  ۳. بر روی متغیر مورد نظر خود کلیک کنید (به عنوان مثال متغیر استرس ادراک شده (tpstress) و آن را به قسمت Dependent List منتقل کنید.
  ۴. بر روی متغیر شماره شناسایی (id) کلیک کنید و آن را به قسمت Label cases منتقل کنید. این کار موجب می شود که در خروجی نقاط دورافتاده با شماره شناسایی شان مشخص شوند.
  ۵. بر دکمه Statistics کلیک کنید. بر روی Outliers کلیک کنید و سپس continue را انتخاب کنید.
  ۶. بر روی دکمه Plots کلیک کنید. Histogram را انتخاب کنید. اگر خواستید می توانید نمودار ساقه و برگ (Stem and Leaf) را هم انتخاب کنید.
  ۷. دکمه Options را انتخاب کنید. گزینه Exclude cases pairwise را انتخاب کنید و سپس Continue بعد از آن OK را بزنید.
- خروجی که از این رویه حاصل می شود، چنین خواهد بود:

### Descriptives

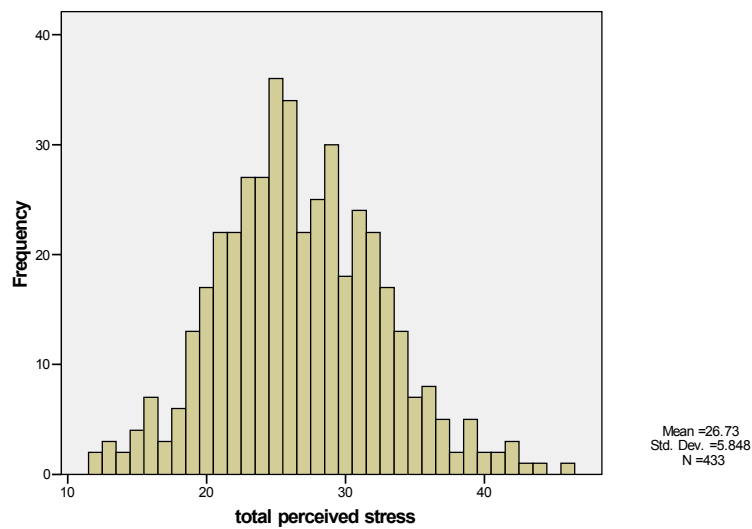
			Statistic	Std. Error
total perceived stress	Mean		26.73	.281
	95% Confidence Interval for Mean	Lower Bound	26.18	
		Upper Bound	27.28	
	5% Trimmed Mean		26.64	
	Median		26.00	
	Variance		34.194	
	Std. Deviation		5.848	
	Minimum		12	
	Maximum		46	
	Range		34	
	Interquartile Range		8	
	Skewness		.245	.117
	Kurtosis		.182	.234

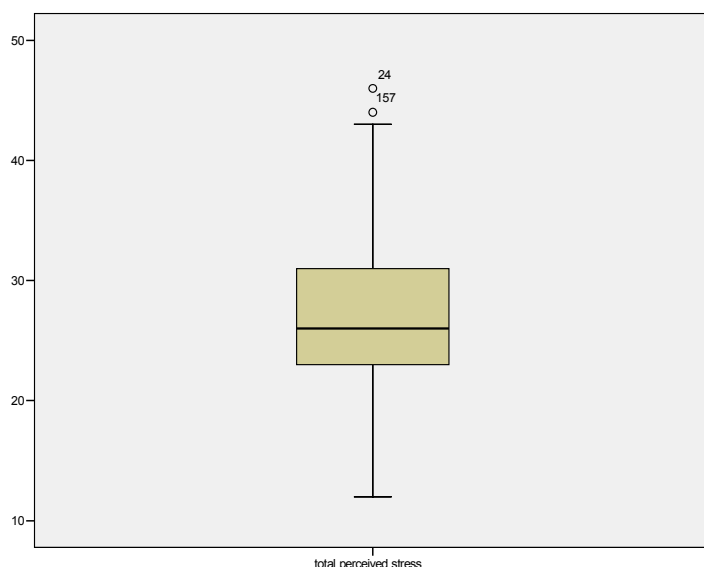
### Extreme Values

			Case Number	id	Value
total perceived stress	Highest	1	24	24	46
		2	130	157	44
		3	55	61	43
		4	6	6	42
		5	123	144	42 <sup>a</sup>
	Lowest	1	311	404	12
		2	5	5	12
		3	239	301	13
		4	106	127	13
		5	103	119	13

a. Only a partial list of cases with the value 42 are shown in the table of upper extremes.

### Histogram





## تفسیر خروجی های Explore

این خروجی اطلاعات بسیار مفیدی درباره توزیع متغیر به شما می دهد.

- اول به هیستوگرام نگاه کنید. به دنباله های توزیع بنگرید. آیا داده هایی خارج از محدوده وجود دارند؟ اگر چنین است آنها پتانسیل دورافتاده بودن را دارند. اگر توزیع شبیه منحنی یکنواختی است آنگاه احتمالاً داده های دورافتاده ای وجود ندارند.

- دوم به نمودار جعبه ای بنگرید. داده هایی را که Spss به عنوان دورافتاده شناخته است با دایره هایی کوچک که شماره ای به همراه آنهاست نشان داده شده اند. این شماره ها همان شماره شناسایی کیس است که به شما کمک می کند آن را پیدا کنید. Spss نقاطی را به عنوان دورافتاده در نظر می گیرد که به اندازه بیش از ۱/۵ برابر طول جعبه ها از لبه آنها فاصله داشته باشند. نقاطی که بیش از ۳ برابر طول جعبه ها از لبه آنها فاصله داشته باشند با علامت ستاره (\*) نشان داده می شوند. این نقاط بسیار دورافتاده هستند و در مثال ما چنین نقاطی وجود ندارند، اما نقاط دورافتاده وجود دارند که شماره شناسایی آنها ۲۴ و ۱۵۷ است. اگر چنین نقاطی در داده های خود یافتید باید تصمیم بگیرید که با آنها چه کنید.

- مهم است که بررسی شود نقاط دورافتاده خطا هستند یا اینکه نقاطی خاص هستند. بررسی کنید که آیا مقدار به دست آمده در دامنه تغییرات ممکن برای آن متغیر قرار

دارد یا خیر. گاهی اوقات بهتر است به پرسشنامه مربوطه مراجعه شود تا مطمئن شویم که مقدار مورد نظر به خطا وارد نشده است. اگر دریافتید که مقدار مورد نظر خطا بوده است آن را تصحیح کنید و نمودار جعبه ای را دوباره رسم کنید. اگر مطمئن شدید که نقطه دورافتاده مورد نظر خطا نیست بلکه موردی خاص است آنگاه باید تصمیم بگیرید که با آن چه کنید. برخی آماردانان پیشنهاد کرده اند که تمام نقاط دورافتاده از داده ها حذف شوند. برخی از آنها پیشنهاد کرده اند که مقدار داده های دورافتاده تغییر داده شود و مقادیری جایگزین آنها شوند که در داده ها وجود دارد.

- اطلاعاتی که در جدول Descriptives به شما داده شده است کمک می کند که دریابید مقادیر دورافتاده برای شما تا چه حد مشکل ساز هستند. آماره ای که برای این بررسی مناسب است میانگین پیراسته است. همانطور که گفتیم برای محاسبه این آماره، Spss ۵ درصد از بیشترین مقادیر داده ها و ۵ درصد از مقادیر کمترین داده ها را حذف کرده است و سپس از روی داده های باقیمانده میانگین را محاسبه کرده است. اگر میانگین پیراسته را با میانگین اصلی مقایسه کنید آنگاه در خواهید یافت که نقاط دورافتاده تا چه حد موثر بوده اند. اگر این دو میانگین تفاوت فاحشی داشتند آنگاه لازم است که شما این داده ها را بیشتر بررسی کنید (در مثال ما میانگین داده ها ۲۶/۷۳ و میانگین پیراسته ۲۶/۶۴ بوده است). از آنجایی که این دو مقدار در این مثال چندان متفاوت نیستند ما آنها را از داده های خود حذف نمی کنیم.
- اگر تصمیم می گرفتیم که داده های دورافتاده را حذف کنیم آنگاه لازم می شد که جدول Extreme values را هم بررسی کنیم. این جدول کمترین و بیشترین مقادیر مربوط به متغیر، به همراه شماره شناسایی آنها (متغیر id) را در اختیار ما می گذارد. این جدول برای پیدا کردن مقادیر دورافتاده به ما کمک می کند. هنگامی که کیس مورد نظر را پیدا کردیم لازم است به فایل داده ها بازگردیم و مقدار متغیر را در کیس مورد نظر تغییر دهیم.

## فصل هفتم

### استفاده از نمودارها برای توصیف و بررسی داده ها

با اینکه آماره های به دست آمده در فصول قبل اطلاعات بسیار خوبی درباره نمونه و متغیرهای ما به دست می دهد اما برخی از ویژگی ها به طور بصری بسیار بهتر مشخص می شوند. Spss نمودارهای بسیار گوناگونی را در اختیار شما قرار می دهد. در این فصل روشهای اساسی برای رسم نمودارهای زیر ارائه خواهند شد:

- هیستوگرام
- نمودار میله ای (بار چارت)
- نمودار پراکندگی (scatterplots)
- نمودار جعبه ای (box charts)
- نمودارهای خطی (line charts)

خودتان با آزمایش به بررسی هریک از نمودارها و ویژگی های آنها پردازید. در این بخش تنها مقدمه کوتاهی ارائه خواهد شد تا شما بتوانید کار خود را آغاز کنید. نمودارهایی که در این فصل رسم شده اند همگی با استفاده از فایل survey.sav هستند.

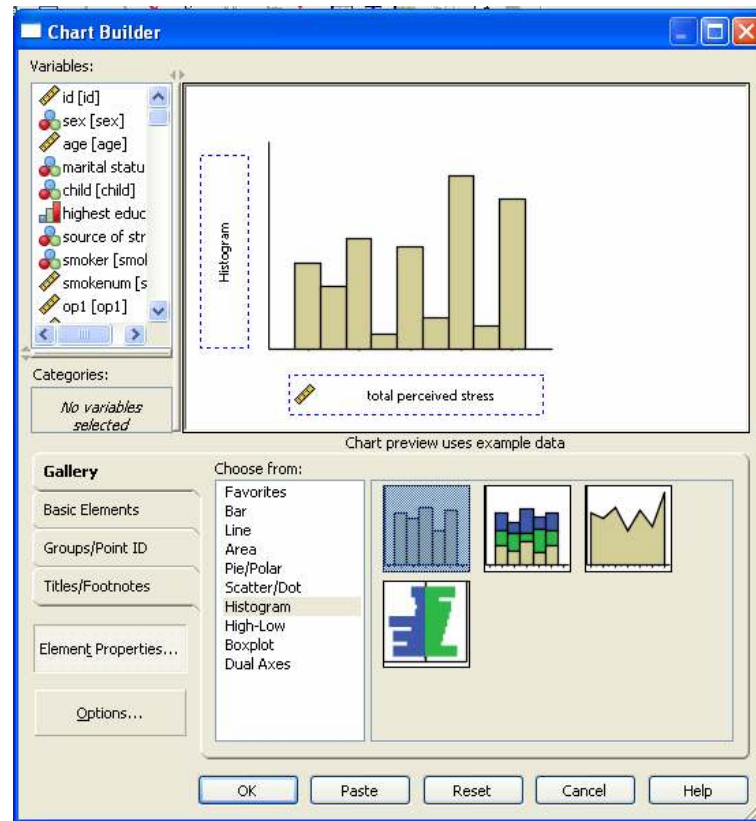
در پایان فصل راهنمایی های برای ایجاد تغییرات دلخواه در ظاهر نمودارها ارائه شده است تا بتوانید نمودار خود را آنگونه که دلخواهتان است ارائه کنید. این کار در ارائه گزارش پژوهش بسیار کاربرد دارد. نمودارهای spss را می توان به آسانی به نرم افزار Word منتقل کرد. شیوه این کار در پایان همین فصل توضیح داده شده است.

### هیستوگرام

هیستوگرام برای نمایش توزیع یک متغیر پیوسته به کار می رود (مانند سن، استرس ادراک شده).

## شیوه رسم یک هیستوگرام

۱. از منوی بالای صفحه، ابتدا Graph و سپس Chart Builder را انتخاب کنید. این گزینه به شما امکان می دهد که به راحتی نمودار مورد نظر خود را انتخاب کنید. صفحه ای مانند شکل زیر ظاهر خواهد شد.



شکل ۷-۱

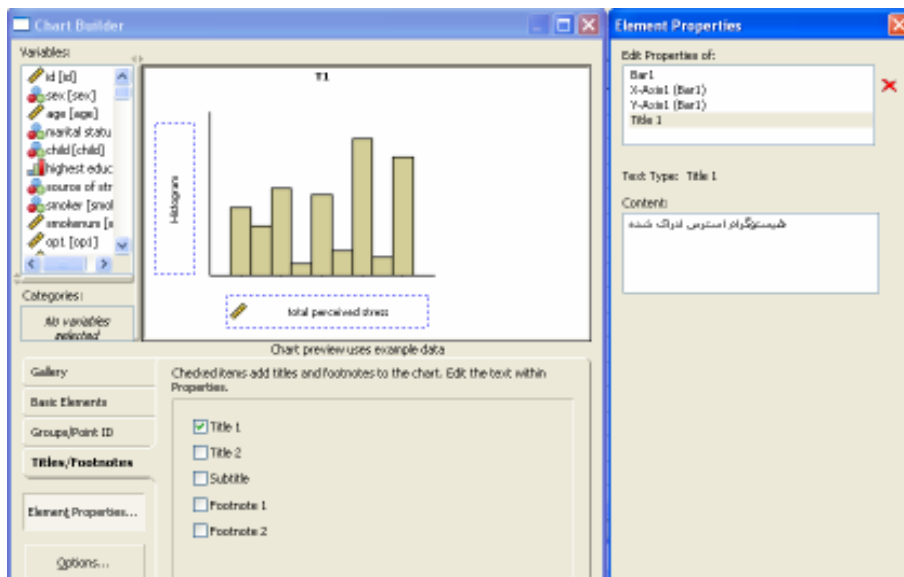
۲. بر روی گزینه Histogram کلیک کنید. و از میان انواع هیستوگرام، مدل ساده آنرا انتخاب کنید (مانند شکل بالا).
۳. بر روی متغیر مورد نظر خود کلیک کنید و آن را به قسمتی که تصویر نمودار در آن در محور Xها درگ کنید (یعنی بر روی آن کلیک کنید، دکمه موس را فشرده نگه دارید و در این حال متغیر مورد نظر را در محور Xها قرار دهید). این متغیر باید پیوسته باشد. (به عنوان مثال استرس ادراک شده tpstress). با این کار پنجره کوچکتر دیگری نیز ظاهر می شود که مانند شکل زیر است.



شکل ۷-۲

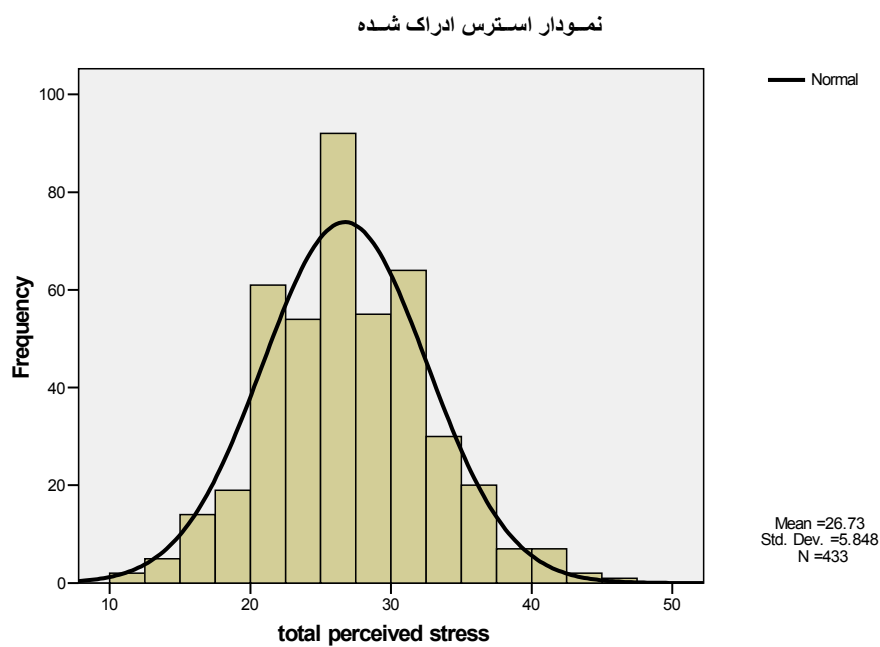
۴. گزینه Display normal curve را انتخاب کنید. این گزینه نمودار نرمالی بر روی توزیع شما رسم خواهد که نشان می دهد نمودار نرمال برای چنین توزیعی چگونه شکلی دارد.

۵. اگر مایل هستید که نمودار شما عنوان داشته باشد بر روی گزینه Titles/Footnotes (عنوانها/ پانویسها) کلیک کنید و عنوان دلخواه خود را در قسمت مشخص شده تایپ کنید. ( به عنوان مثال «هیستوگرام استرس ادراک شده» ). مانند شکل زیر





۶. بر روی گزینه OK کلیک کنید.  
خروجی حاصل چنین شکلی خواهد داشت:



تفسیر خروجی هیستوگرام

## فصل هشتم

### دستکاری در داده ها

هنگامی که داده های خود را وارد کردید و صحت آن هم بررسی شد حال باید در داده های خام آنچنان تغییر ایجاد کنید که به شکلی در آیند که برای آزمون فرضیه های شما قابل استفاده باشند. بسته فایل داده ها، متغیرهای مورد نظر و نوع سوالات تحقیق این فرآیند شامل مراحل زیر خواهد بود:

- جمع کردن گویه هایی که با هم یک طیف را می سازند تا نمره ای کلی برای طیف که از تعدادی گویه تشکیل شده است به دست آید. مانند طیف اعتماد به نفس (Self-esteem)، خوش بینی (optimism)، استرس ادراک شده و غیره. Spss این کار را با سرعت و دقت بسیار انجام می دهد و نیاز به محاسبه با دست نیست.
- تغییر شکل توزیع متغیرهایی که برای تحلیل آنها نیاز است که توزیع آنها نرمال باشند.
- تبدیل متغیرهای پیوسته (مانند سن) به متغیرهای مقوله ای (مانند جوان، میانسال و سالمند) برای انجام برخی تحلیل ها مانند آنالیز واریانس.

شیوه اعمال تغییرات در داده ها در ذیل توضیح داده شده اند.

#### محاسبه نمره کل برای طیفها

قبل از آنکه تحلیلهای آماری را بر روی داده ها انجام دهید باید نمره کل را برای طیفهایی که از تعدادی گویه تشکیل شده اند به دست آورید. این کار دو مرحله دارد:

- مرحله اول: معکوس کردن کدگذاری گویه هایی که معنی منفی دارند.
- مرحله دوم: جمع کردن گویه هایی که سازنده طیفها و خرده طیفها هستند.

قبل از اینکه این کار را انجام دهید لازم است که طیفها و سنجه هایی را که در پژوهش خود به کار می برید به خوبی درک کنید و بفهمید. این کار به این خاطر مهم است که شما باید بفهمید که کدام گویه ها معکوس کدگذاری شده اند تا قبل از جمع کردن آنها با هم، کدگذاری آنها را تغییر دهید. همچنین برخی طیفها از چند خرده طیف تشکیل شده اند و نباید همه آنها را با هم یکجا جمع کرد. انجام درست اینکار اهمیت زیادی دارد به این خاطر که صحت تحلیل های بعدی شما به صحت انجام این کار بستگی دارد.

### مرحله اول: معکوس کردن کدگذاری گویه های منفی

در برخی از طیفها برخی گویه ها را منفی پیرسیده اند تا از سوگیری پاسخگویان جلوگیری شود. مثال این موضوع طیف سنجش خوش بینی (Optimism) در مثال ماست. سوال اول مثبت پیرسیده شده است (یعنی نمره بالاتر نشان دهنده خوش بینی است): «در موقعیتهای نامشخص، من انتظار بهترین حالت را دارم». اما گویه دوم منفی عنوان شده است (نمره بالاتر نشان دهنده خوش بینی کم است): «اگر قرار باشد اتفاق بدی برای من بیافتد خواهد افتاد». بنابراین قبل از اینکه نمره کل را برای طیف خوش بینی محاسبه کنیم باید نمرات مربوط به سوالاتی را که منفی پیرسیده اند معکوس کنیم. باید مطمئن شویم که در تمام گویه ها نمره بالاتر نشان دهنده خوش بینی بیشتر است. شیوه معکوس کردن کدگذاری گویه های ۲، ۴ و ۶ از طیف خوش بینی در ذیل ارائه شده است. در این طیف از طیف لیکرت پنج تایی استفاده شده است. بنابراین نمرات این طیف بین ۱ (بسیار مخالف) و ۵ (بسیار موافق) خواهند بود.

قبل از اینکه تحلیل خود را شروع کنید به دو نکته زیر توجه کنید:

- از فایل داده های خود یک کپی تهیه کنید زیرا اگر اشتباه کنید داده های اصلی خود را از دست خواهید داد. پس حتما نسخه های دیگری از فایل اصلی خود ذخیره کنید.
- در منوی Edit سپس Options و بعد از آن Data مطمئن شوید که گزینه Calculate values immediately فعال است. این گزینه برای آن است که تمام دستورات شما بلافاصله اجرا شوند. (برای توضیحات بیشتر به فصل ۵ مراجعه کنید).

### شیوه معکوس کردن نمرات گویه های یک طیف

۱. در پنجره Data Editor، از منوی Transform گزینه Recode سپس Into same variable را انتخاب کنید.

۲. گویه هایی را که می خواهید نمرات آنها معکوس شوند انتخاب کنید ( op2, op4, op6) و آنها را به قسمت Variables منتقل کنید.
۳. بر روی دکمه Old and new variables کلیک کنید.
  - در قسمت Old value عدد ۱ را وارد کنید.
  - در قسمت New value عدد ۵ را وارد کنید. (این کار موجب می شود که تمام متغیرهایی که در ابتدا ۱ بوده اند تبدیل به ۵ شوند).
۴. دکمه Add را بزنید. این کار موجب می شود که در قسمتی که Old→New نام دارد چنین چیزی ظاهر شود: 1→5
۵. برای نمرات باقیمانده هم همین کار را انجام دهید. به عنوان مثال:
  - در Old value عدد ۲ و در New value عدد ۴ را وارد کنید و Ok را بزنید.
  - در Old value عدد ۳ و در New value عدد ۳ را وارد کنید و Ok را بزنید.
  - در Old value عدد ۴ و در New value عدد ۲ را وارد کنید و Ok را بزنید.
  - در Old value عدد ۵ و در New value عدد ۱ را وارد کنید و Ok را بزنید.
 همیشه یکبار دیگر مقادیری را که به عنوان مقادیر قدیم و جدید وارد کرده بررسی کنید تا اشتباهی رخ نداده باشد. همیشه تمام گویه ها ۵ تایی نیستند. ممکن است بیشتر یا کمتر باشند. مطمئن شوید که تمام نمرات ممکن را معکوس کرده اید.
۶. هنگامی که کاملاً از صحت کاری که انجام داده اید مطمئن شدید آنگاه Continue و Ok را انتخاب کنید.

هنگامی که معکوس کردن کدها در گویه های مورد نظر را انجام دادید لازم است که این نکته را در کدنامه خود یادداشت کنید. می توانید برای آنکه از صحت کار خود اطمینان یابید، چند کیس را قبل و بعد از Recode بررسی کنید.

### مرحله دوم: جمع کردن گویه و محاسبه نمره کلی طیف

بعد از اینکه کدگذاری گویه های منفی معکوس شد آنگاه می توانید نمره کل طیف را برای پاسخگویان محاسبه کنید. این کار را زمانی انجام دهید که فایل داده های شما تکمیل شده است، یعنی تمام متغیرها و کیس ها وارد شده اند. نکته مهم آن است که spss مانند نرم افزارهایی مثل Excel که محاسبات را با ورود داده های جدید دوباره انجام می دهند، نیست.

## شیوه محاسبه نمره کل برای یک طیف

۱. در پنجره Data Editor، از منوی بالای صفحه گزینه Transform و سپس Compute را انتخاب کنید.
۲. در قسمتی که Target Value نام دارد، نام متغیری جدیدی را که می خواهید نمرات کل طیف در آن باشند وارد کنید (خوب است که اول نام متغیرهای خود را با T شروع کنید تا نشان دهنده مجموع بودن آن باشد و هنگام انجام تحلیل ها بتوان به ترتیب الفبایی متغیرهایی که نمره کل طیف در آنهاست آسانتر یافت).
- نکته مهم آن است که مراقب باشید به طور اتفاقی از نام متغیری که در فایل داده های شما وجود دارد استفاده نکنید، در غیر این صورت مقادیر جدید جایگزین مقادیر اصلی در آن متغیر خواهد شد و شما داده های خود را از دست خواهید داد.
۳. بر روی دکمه Type & Label کلیک کنید. در قسمت Label توضیح متغیر جدید را بنویسید (به عنوان مثال خوشبینی کل) و بر روی Continue کلیک کنید.
۴. از لیست متغیرهایی که در سمت چپ پنجره Compute قرار دارند، متغیر مربوط به گویه اول را انتخاب کنید (op1).
۵. بر روی دکمه ای که علامت ► روی آن است کلیک کنید تا این متغیر به قسمت Numeric Expression منتقل شود.
۶. در قسمتی که مانند ماشین حساب است دکمه + را بزنید.
۷. گویه دوم را هم به قسمت Numeric Expression منتقل کنید. و دوباره دکمه + بزنید. این کار را تا زمانی که تمام گویه های مربوط به طیف در قسمت Numeric Expression با هم جمع شوند انجام دهید.
۸. عبارت محاسباتی که در قسمت Numeric Expression ظاهر می شود باید این چنین باشد:  $op1+op2+op3+op4+op5+op6$
۹. دوباره آنچه انجام داده اید بررسی کنید تا خطایی وجود نداشته باشد، آنگاه کلید OK را بزنید.

با انجام این کار متغیر جدیدی که آن را TOPTIM نامیدیم در انتهای متغیرهای شما اضافه می‌شود و مقادیر جدید در درون آن قرار می‌گیرند. نمره ای که برای هر کیس در این متغیر جدید قرار داده شده حاصل جمع متغیرهای op1 تا op6 است. اگر هر کدام از این متغیرها برای یک کیس، داده گمشته باشد، حاصل جمع آنها هم داده گم شده و جای آن خالی خواهد بود.

همواره جزئیات مربوط به متغیرهای جدیدی را که می‌سازید در کدنامه خود یادداشت کنید. نام متغیر، توضیحات مربوطه و نیز شیوه ساختن آن را با جزییات در کدنامه مشخص کنید. همچنین مطلوب است دامنه تغییرات ممکن برای این متغیر جدید را هم در کدنامه خود ثبت کنید. این کار به شما کمک می‌کند که مقادیر خارج از محدوده مورد انتظار را راحت تر بیابید. این کار همچنین در بررسی توزیع متغیر جدید به شما کمک می‌کند. به عنوان مثال اینکه آیا میانگین در وسط قرار گرفته است یا خبر؟

بعد از اینکه متغیر جدید را ساختید باید با استفاده از Descriptives صحت مقادیر آن را بررسی کنید (فصل ۵ را ببینید):

- به پرسشنامه بازگردید و بررسی کنید که دامنه تغییرات ممکن برای این متغیر چه حدودی است. اگر طیف شما از ده سوال تشکیل شده باشد و طیف لیکرت شما چهار گزینه داشته باشد آنگاه کمترین مقدار ممکن برای طیف ۱۰ و بیشترین مقدار ممکن ۴۰ خواهد بود. چرا که اگر کسی به همه گویه‌ها ۱ بدهد نمره کل او ۱۰ و اگر به همه گویه‌ها ۴ بدهد نمره کل او ۴۰ خواهد بود.

- با استفاده از منو Descriptives مطمئن شوید که هیچ داده‌ای خارج از دامنه تغییرات ممکن قرار ندارد.

- اگر برای متغیری که سنجیده‌اید، پیشینه‌ای در منابع وجود دارد، میانگین به دست آمده برای طیف خود را با مقادیری که دیگران به دست آورده‌اند مقایسه کنید. اگر تفاوت فاحشی را مشاهده می‌کنید باید به دنبال دلیل این همه تفاوت باشید. شاید در معکوس کردن کد گویه‌ها و یا جمع کردن آنها دچار خطا شده باشید. شما می‌توانید تحلیل‌های دیگری هم برای بررسی توزیع نمرات به دست آمده انجام دهید:

- توزیع داده‌ها با استفاده مقادیر چولگی و کشیدگی مورد بررسی قرار دهید (فصل ششم را ببینید)

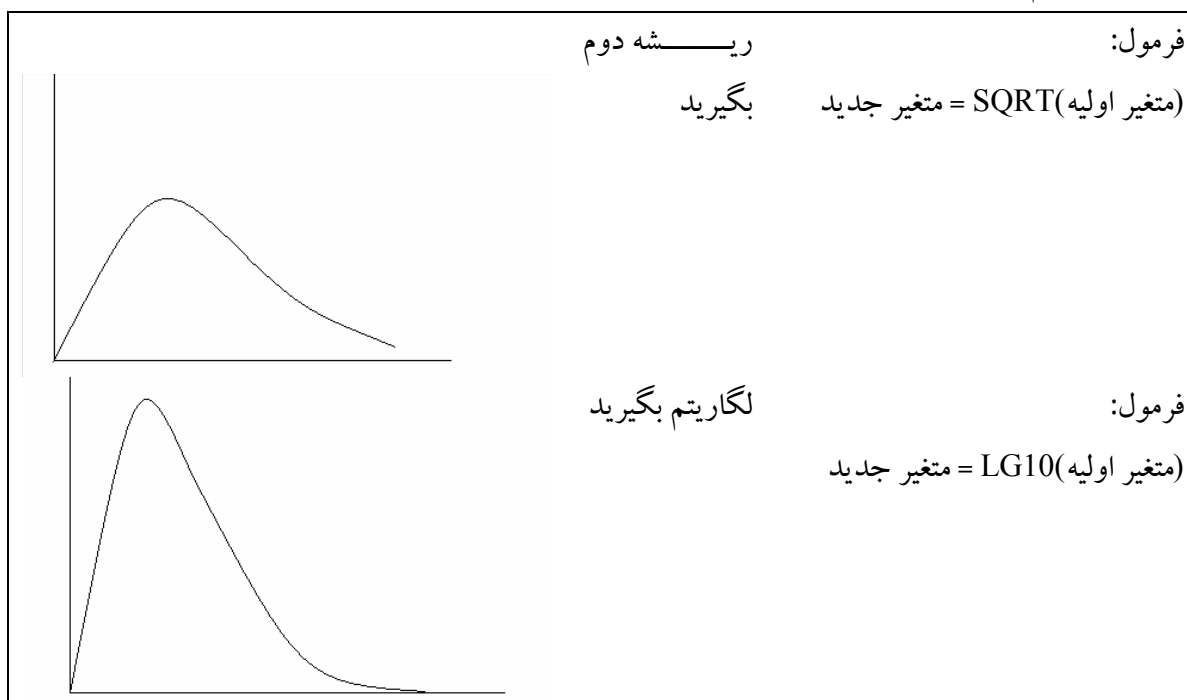
- هیستوگرام را برای متغیر جدید بکشید تا ببینید که آیا داده‌ها نرمال توزیع شده‌اند یا خیر؟ و یا اینکه به تبدیل نیاز دارند.

## تبدیل توزیع داده ها

اغلب هنگامی که توزیع نمرات یک طیف را بررسی می کنید (به عنوان مثال طیف اعتماد به نفس یا عصیت) می بینید که متاسفانه توزیع داده ها به خوبی شبیه منحنی نرمال نیست. گاهی اوقات داده ها چولگی مثبت دارند، یعنی بیشتر پاسخگویان نمرات کم گرفته اند (مانند توزیع متغیر افسردگی)؛ گاهی اوقات توزیع داده ها چولگی منفی دارد یعنی بیشتر داده ها مقادیر بالا دارند (مانند اعتماد به نفس). بیشتر تحلیل های آماری بر اساس نرمال بودن توزیع متغیر به کار می کنند، پس چه باید کنیم؟

یکی از گزینه ها آن است که استفاده از آمار پارامتری (مانند همبستگی پیرسون و آنالیز واریانس) را به کنار بگذاریم و و از جایگزینهای ناپارامتری آنها (رو اسپیرمن، کروسکال-والیس) استفاده کنیم. Spss برخی از بهترین تکنیکهای ناپارامتری را در اختیار شما می گذارد. در باره این تکنیکها در فصل ۲۱ بحث خواهد شد. اما متاسفانه این تکنیکها چندان قوی نیستند یعنی در زمانی که رابطه ای وجود دارد ممکن است نتوانند وجود آن را نشان دهند.

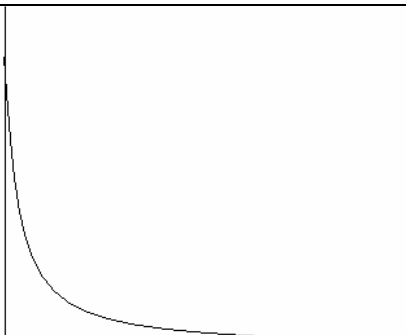
یک راه دیگر هنگامی که با توزیع غیرنرمال مواجه می شوید آن است که توزیع خود را تبدیل کنید. یعنی بر اساس فرمولهای خاصی داده های خود را تغییر دهید تا توزیع آنها توزیعی نرمال شود. تبدیلهای گوناگونی وجود دارند که بسته به نوع توزیعی که شما دارید از آنها باید استفاده کنید. در منابع دیدگاههای متضادی درباره این تکنیک وجود دارد، برخی به شدت آن را توصیه کرده اند و برخی هم آن را رد کرده اند.



فرمول:

معکوس کنید

(متغیر اولیه) / ۱ = متغیر جدید

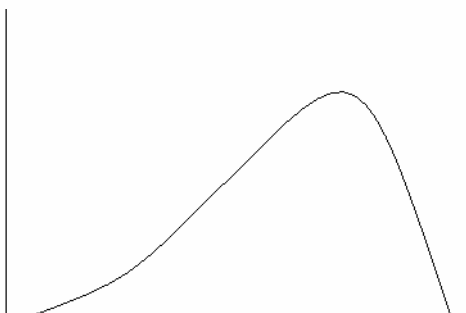


فرمول:

تصویر کنید و

(متغیر اولیه-K) = SQRT(متغیر جدید) سپس معکوس

که در آن K بزرگترین مقدار متغیر  
به علاوه ۱ است.

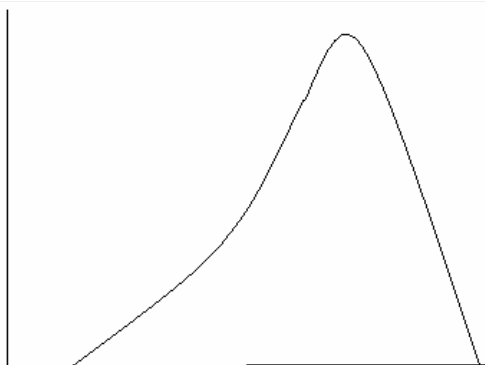


فرمول:

تصویر کنید و

(متغیر اولیه-K) = LG10(متغیر جدید) سپس لگاریتم

که در آن K بزرگترین مقدار متغیر  
به علاوه ۱ است.

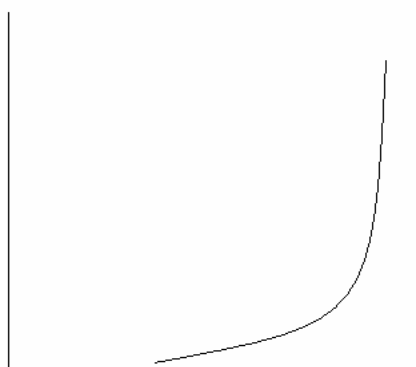


فرمول:

تصویر کنید و

(متغیر اولیه-K) / ۱ = متغیر جدید سپس معکوس

که در آن K بزرگترین مقدار متغیر  
به علاوه ۱ است.



شکل ۸-۱



در جدول ۸-۱ برخی از توزیعهایی که با ممکن است به آنها برخورد و شیوه تبدیل آنها به توزیع نرمال ارائه شده است. شما باید تشخیص دهید که کدام تبدیل بهتر می تواند به رفع مشکل شما کمک کند.

### **تبدیل یک متغیر پیوسته (فاصله ای) به متغیری طبقه بندی شده (ترتیبی کمی)**

گاهی اوقات (به عنوان مثال هنگامی که می خواهید آنالیز واریانس انجام دهید) ممکن است بخواهید که نمونه خود را به گروههای مساوی بر حسب نمره مربوط به یک متغیر تقسیم کنید (به عنوان مثال سه گروه که نمره آنها کم، متوسط و زیاد است). در این بخش شیوه انجام این کار توضیح داده می شود. به عنوان مثال متغیر age (سن) در فایل survey.sav را به سه گروه تقریباً مساوی تقسیم می کنیم. به طور مشابهی می توان از تکنیک تقسیم کردن بر اساس میانه (median split) استفاده کرد که در آن میانه نقطه برش و تقسیم کیس ها به دو گروه است.

اول باید توزیع متغیر پیوسته مورد نظر را بررسی کنیم و نقاطی که برش از آنها کیسهای ما را به گروه تقریباً مساوی سن کم، سن متوسط و سن زیاد تقسیم می کند، مشخص کنیم. آنگاه متغیر جدیدی خواهیم ساخت که ما آن را agegp نامیده ایم. این متغیر جدید تنها سه مقدار می پذیرد (۱=کم، ۲=متوسط، ۳=زیاد).

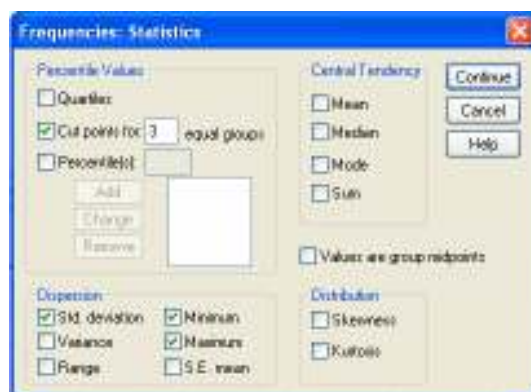
### **گام اول: محاسبه نقاط برش**

#### **شیوه محاسبه نقاط برش برای گروه بندی داده ها**

۱. از منوی Analyze، گزینه Descriptive statistics و سپس Frequencies را انتخاب کنید.

۲. متغیری را که قصد دارید (مانند سن) گروه بندی کنید انتخاب و به قسمت Variables منتقل کنید.

۳. بر روی دکمه Statistics کلیک کنید. و سپس گزینه Cut points for Equal groups را انتخاب کنید. در جای خالی عدد ۳ وارد کنید چون می-خواهیم سه گروه تقریباً برابر بسازیم. دکمه Continue و سپس OK را بزنید.



خروجی حاصل جدول زیر خواهد بود:

Statistics		
age		
N	Valid	439
	Missing	0
Std. Deviation		13.202
Minimum		18
Maximum		82
Percentiles	33.33333333	29.00
	66.66666667	44.00

در جدول بالا به قسمتی که Percentiles (صدکها) نام دارد توجه کنید. اگر شما بخواهید سه گروه بسازید آنگاه دو نقطه برش وجود خواهد داشت که یکی از آنها صدک ۳۳/۳۳ است و دیگری صدک ۶۶/۶۷. مقادیری که برای این دو صدک محاسبه شده است (اولی ۲۹ و دومی ۴۴) برای تبدیل متغیر سن به متغیری طبقه بندی شده که تنها دارای سه مقدار (کم، متوسط و زیاد) است به کار می-روند. دسته اول شامل کیسهایی خواهد بود که سن آنها از کمترین مقدار تا ۲۹ است (سن کم). دسته دوم کیسهایی خواهند بود که سن آنها بین ۳۰ تا ۴۴ است و دسته سوم کیسهایی خواهند بود که سن آنها ۴۵ سال و بیشتر است.

### گام دوم: ساختن متغیر جدید مقوله ای (agegp)

حال می خواهیم متغیر جدید مقوله ای خود را که تنها دارای سه مقدار است بسازیم. هریک از کیسهای نمونه ما در یکی از این سه دسته (کم، متوسط و زیاد) قرار خواهند گرفت.

## شیوه تبدیل یک متغیر پیوسته به متغیری گروه بندی شده

۱. در پنجره Data Editor از منوی Transform گزینه Recode و سپس Into Different Variable را انتخاب کنید. این گزینه موجب می شود که تغییرات شما در متغیری جدید ذخیره شود (عکس گزینه Into Same Variable که تغییرات را در همان متغیر اصلی انجام می دهد و جایگزین می کند).
۲. متغیر مورد نظر خود را به قسمت Input Variable → Output Variable منتقل کنید.
۳. در قسمت Output Variable Name نام متغیر جدید را که قرار است ساخته شود وارد کنید. ما نام agegp3 را انتخاب کرده ایم. پسوند gp را به نام متغیر اضافه کرده ایم تا یادآور گروه بندی شده بودن متغیر جدید باشد (gp مخفف grouped است). مراقب باشید نام متغیر جدید از قبل در میان متغیرهای موجود شما وجود نداشته باشد.
۴. بر روی دکمه Change کلیک کنید.
۵. در قسمت Label توصیفی را که برای متغیر جدید در نظر دارید وارد کنید. به عنوان مثال «سن سه گروهی».
۶. بر روی گزینه Change کلیک کنید.
۷. بر روی دکمه Old and New Values کلیک کنید.
۸. بر روی قسمتی که Range: Lowest through..... نام دارد کلیک کنید:
  - در این قسمت مقدار صدک ۳۳/۳۳ را وارد کنید (در این مثال ۲۹).
  - در قسمتی که New Value نام دارد عدد را وارد کنید. یعنی شما از spss خواسته اید از کمترین مقدار تا ۲۹ را به عنوان ۱ در نظر بگیرد. دکمه Add را بزنید.
۹. بر روی دکمه که Range: ..... Through..... نام دارد کلیک کنید.
  - در جای خالی اول مقداری را وارد کنید که اندکی بیشتر از حد بالای طبقه قبل است ( به عنوان مثال ۳۰).
  - در جای خالی دوم عدد مربوط به صدک ۶۶/۶۷ را وارد کنید که در این مثال ۴۴ است.

- در قسمت New Value عدد ۲ وارد کنید. یعنی شما از spss خواسته اید تمام مقادیر بین ۳۰ تا ۴۴ را در یک گروه قرار دارد که عدد متناظر آن ۲ است.

- ۱۰. بر روی قسمتی که Range: ..... through highest نام دارد کلیک کنید.
- در جای خالی مقداری را که اندکی بیشتر از حد بالای طبقه قبلی است وارد کنید. در مثال ما عدد ۴۵ مناسب است.
- در قسمت New Value عدد ۳ را وارد کنید. این یعنی شما از spss خواسته اید تمام مقادیری را که مقدار آنها ۴۵ یا بیش از آن است در گروهی قرار دهد که عدد متناظر آن ۳ است.

۱۱. حال که تمام گروهها را مشخص کردید در قسمتی که Old→New نام دارد دوباره بررسی کنید که تغییرات را درست وارد کرده اید. در این قسمت باید عبارت زیر مشاهده شود:

Lowest thru 29 → 1  
30 thru 44 → 2  
45 thru highest → 3

- ۱۲. بر روی Continue سپس OK کلیک کنید.
- ۱۳. بعد از اینکه متغیر جدید ساخته شد با برچسب های مربوط به مقادیر (Value Labels) را هم مشخص کنید. در صفحه Variable View در قسمت Value Labels به شیوه ای که قبلا آموختیم برچسبهای مقادیر را چنین وارد کنید: مقدار ۱ برچسب ۱۸-۲۹؛ مقدار ۲ برچسب ۳۰-۴۴؛ مقدار ۳ برچسب ۴۵+.

برای اینکه از صحت کاری که انجام دادید مطمئن شوید می توانید برای متغیر جدید جدول فراوانی رسم کنید. انتظار دارید که در این جدول تنها سه مقدار برای این متغیر مشاهده کنید.

## فصل نهم

### بررسی قابلیت اعتماد طیف

هنگامی که از طیف برای سنجیدن یک مفهوم استفاده می کنید مهم است که قابلیت اعتماد طیف مورد نظر را بررسی کرده باشید. دیدگاههای مختلفی درباره قابلیت اعتماد وجود دارد که درباره آنها قبلاً بحث کرده ایم. یکی از این دیدگاه ها، سازگاری درونی طیف است. یعنی تا چه حد گویه-هایی را که طیف را می سازند با هم همبستگی دارند. آیا همه آنها یک سازه را می سنجدند؟ یکی از رایج ترین شاخصهای سازگاری درونی ضریب آلفای کرونباخ است. آلفای کرونباخ برای یک طیف خوب ۰/۷ یا بیشتر است. ضریب آلفای کرونباخ به تعداد گویه ها هم بستگی دارد. هر چه تعداد گویه-ها بیشتر باشند، مقدار آلفای کرونباخ هم احتمالاً بیشتر خواهد بود. احتمال اینکه در طیف های کوچک (آنهايي که تعداد گویه های آنها کمتر از ۱۰ عدد است) مقدار آلفا کم باشد (حدود ۰/۵) زیاد است. در چنین حالتی مناسبتر آن است که میانگین همبستگی بین گویه ای گزارش شود. مقدار مناسب برای میانگین همبستگی بین گویه ای را بین ۰/۲ تا ۰/۴ دانسته اند.

قابلیت اعتماد به نمونه ای که طیف در آن استفاده می شود هم بستگی دارد. بنابراین لازم است که قابلیت اعتماد طیف مورد استفاده در نمونه خود را بررسی کنید و نتایج آن را در قسمت روش تحقیق گزارش پژوهش خود ذکر کنید. اگر در طیف شما گویه هایی وجود دارند که منفی پرسیده شده اند باید قبل از محاسبه آلفای کرونباخ کدگذاری آنها را معکوس کنید. یعنی تمام گویه های طیف، هنگام محاسبه آلفای کرونباخ باید هم جهت کدگذاری شده باشند. مشروح این کار در فصل ۸ توضیح داده شد. اگر طیف خود را از مقاله یا پژوهش دیگری گرفته اید به متن اصلی مراجعه کنید و بررسی کنید که آیا طیف از تعدادی خرده طیف تشکیل شده است یا خیر. اگر طیف دارای خرده طیفهایی در درون خود باشد، آنگاه نباید همه گویه ها را با هم در تحلیل قابلیت اعتماد وارد کرد. بلکه گویه های مربوط به هر خرده طیف باید جداگانه از نظر قابلیت اعتماد بررسی شوند.

## مثال

حال می خواهیم با استفاده از فایل survey.sav به بررسی یک مثال پردازیم. قصد ما این است که قابلیت اعتماد را برای طیف رضایت از زندگی بررسی کنیم. این طیف از ۵ گویه ساخته شده است. نام این گویه ها عبارتند از: lifsat1, lifsat2, lifsat3, lifsat4, lifsat5.

### شیوه بررسی قابلیت اعتماد یک طیف

نکته مهم: قبل از آنکه کار خود را شروع کنید مطمئن شوید که تمام گویه های شما هم جهت کد گذاری شده اند. اگر چنین نکنید آنگاه نتایج تحلیل شما نادرست خواهد بود.

۱. از منوی بالای صفحه Analyze و سپس Reliability Analysis را انتخاب کنید.

۲. تمام گویه هایی را که سازنده طیف مورد نظر شما هستند (در مثال ما lifsat1, lifsat2, lifsat3, lifsat4, lifsat5) انتخاب و در قسمت Items وارد کنید.

۳. در قسمت Model گزینه Alpha را انتخاب کنید.

۴. بر روی دکمه Statistics کلیک کنید. در قسمت Descriptives for گزینه های Item، Scale، و Scale if item deleted را انتخاب کنید.

۵. دکمه Continue و سپس OK را انتخاب کنید.

خروجی چنین خواهد بود:

Reliability Statistics

Cronbach's Alpha	N of Items
.890	5

Item Statistics

	Mean	Std. Deviation	N
lifsat1	4.37	1.528	436
lifsat2	4.57	1.554	436
lifsat3	4.69	1.519	436
lifsat4	4.75	1.641	436
lifsat5	3.99	1.855	436

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
lifsat1	18.00	30.667	.758	.861
lifsat2	17.81	30.496	.752	.862
lifsat3	17.69	29.852	.824	.847
lifsat4	17.63	29.954	.734	.866
lifsat5	18.39	29.704	.627	.896

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
22.38	45.827	6.770	5

### تفسیر خروجی های Reliability

- خروجی این تکنیک اطلاعاتی درباره طیف مورد استفاده به شما می دهد.
- اولین چیزی که باید بررسی کنید آن است که ببینید آیا تعداد گویه ها درست است. همچنین ببینید میانگین به دست آمده در حدود دامنه تغییرات ممکن برای گویه-هاست یا خیر. خطای کدگذاری اغلب در این قسمت خود را نشان می دهد.
- مهمترین مقدار برای ما مقدار آلفا است. این مقدار همان ضریب آلفای کرونباخ است که در مثال ما ۰/۸۹ به دست آمده است. این مقدار بیش از ۰/۷ است پس می-توانیم بگوییم که این طیف برای نمونه ما دارای قابلیت اعتماد است.
- در جدول Item-Total Statistics ستونی که برای ما اهمیت دارد Corrected Item-Total Correlation (همبستگی تصحیح شده گویه با کل طیف) است. این مقدار به شما نشان می دهد که هر گویه چه مقدار با کل طیف همبستگی دارد. مقادیر کم (کمتر از ۰/۳) نشان دهنده آن است که گویه مورد نظر چیزی غیر از سازه مورد نظر ما می سنجد. اگر مقدار آلفای کرونباخ شما کم باشد (کمتر از ۰/۷) آنگاه باید به فکر آن باشید که گویه های نامناسب را حذف کنید. در ستون Cronbach's Alpha if Item Deleted اثر حذف هر یک از گویه ها نشان داده شده است. این ستون به شما می گوید که اگر هر کدام از گویه ها حذف شود، مقدار آلفا برای گویه های باقیمانده چقدر خواهد بود. اگر مقدار آلفا برای گویه های باقیمانده بیشتر

از آلفای موجود باشد آنگاه می توانید تصمیم بگیرید که گویه مورد نظر را حذف کنید.

### ارائه نتایج تحلیل قابلیت اعتماد

معمولا چنین است که شما باید نتایج بررسی قابلیت اعتماد طیفهای خود را در بخش روش پژوهش و زیر عنوان ابزار سنجش ارائه دهید. اگر از طیف مورد استفاده شما قبلا پژوهشگران دیگری هم استفاده کرده اند، تحلیل قابلیت اعتماد آنها را به طور خلاصه ارائه دهید و آنگاه در یک جمله خلاصه کار خود را توضیح دهید. به عنوان مثال:

به گفته محمدی و حسینی (۱۳۸۲) طیف رضایت از زندگی سازگاری درونی خوبی دارد و مقدار آلفای کرونباخ آن ۰/۸۵ است. در این پژوهش مقدار آلفای کرونباخ برای این طیف ۰/۸۹ بوده است.



## فصل دهم

### انتخاب آزمون صحیح

یکی از دشوارترین و شاید دلهره آور ترین قسمتهای پژوهش برای دانشجویان انتخاب تکنیک مناسب برای تحلیل داده هاست. محاسبه همبستگی و دیگر تکنیکهای آماری را شاید بسیاری از دانشجویان بدانند اما اینکه چه زمانی باید از هر کدام از این تکنیکها استفاده کنند، بحثی جداگانه است که بسیاری در این زمینه دچار مشکل هستند. در یک پژوهش ممکن است شما بسته به نوع سوالات پژوهشی خود و ماهیت داده هایی که در اختیار دارید، از تکنیکهای گوناگونی استفاده کنید. بنابراین لازم است که شما اطلاعات مبنایی را درباره تکنیکهای گوناگون، نوع سوالاتی را که هر کدام از آنها می توانند پاسخ دهند و مفروضات و ملزومات آنها بدانید.

بنابراین باید در کتابهای آماری جستجو کنید و ویژگی های هر تکنیک آماری را به خوبی مطالعه کنید. شما باید مقالات مهم در حوزه مورد پژوهش خود را بررسی کنید و ببینید آنها از چه تکنیکهایی استفاده کرده اند. بنابراین مهم است که شما بدانید دیگر پژوهشگران در حوزه پژوهش شما چگونه داده های خود را تحلیل کرده اند. به دنبال مقالاتی باشید که با جزئیات و کامل تکنیکهای آماری خود را توضیح داده اند. این مقالات را در پوشه ای به عنوان مراجع کار خود جمع آوری و مرتب کنید و همواره در دسترس خود داشته باشید. از این مقالات می توانید برای انتخاب چگونگی ارائه تحلیلهای خود در گزارش پژوهش هم استفاده کنید.

### مروری بر تکنیکهای آماری گوناگون

این بخش در دو قسمت ارائه شده است. ابتدا به تکنیکهایی پرداخته ایم که به بررسی رابطه میان متغیرها می پردازند (مانند رابطه میان سن و خوشبینی)، در قسمت دوم تکنیکهایی بیان شده اند که به مقایسه میان گروهها می پردازند (مانند تفاوت در نمره خوش بینی در جنسیت های مرد و زن). این تفکیک از آن جهت انجام شده است که در بیشتر کتابهای آماری چنین تفکیکی رایج است و بسیاری از دانشجویان این تکنیک ها را چنین آموخته اند. اما می دانیم که شباهتهای زیادی میان تکنیکهای

آماري مختلف وجود دارد كه البته در نگاه اول ممكن است به چشم نيايند. بحث كامل در مورد اين شباهتها فراتر از مجال اين كتاب است و مي توانيد براي اين بحث به كتابهاي آماري مراجعه كنيد. در اينجا تنها توصيفي اوليه براي برخي از تكنيكهاي آماري ارائه شده است، تا ذهني اوليه درباره هر يك از اين تحليلها به دست آوريد.

### **بررسی روابط میان متغیرها**

گاهی اوقات شما در پژوهش خود دنبال بررسی تفاوتهای بین گروهی نیستید، بلکه می خواهید بدانید شدت رابطه میان برخی متغیرها چه اندازه است. تکنیکهای بسیاری برای انجام این کار وجود دارند.

### **همبستگی پیرسون**

همبستگی پیرسون هنگامی به کار می رود كه می خواهید رابطه میان دو متغیر پیوسته را بررسی كنید. این تكنيك جهت رابطه و شدت (قدرت) آن را در اختیار شما می گذارد. همبستگی مثبت به معنای آن است كه هنگامی كه متغیری افزایش می یابد، متغیر دیگر هم افزایش می یابد و هنگامی كه یکی کاهش می یابد دیگری هم کاهش می یابد به عبارت دیگر متغیرها هم جهت با هم تغییر می کنند. همبستگی منفی به آن معنی است كه هنگامی كه یکی از متغیرها افزایش می یابد دیگری کاهش می یابد، و هنگامی كه اولی کاهش می یابد دیگری افزایش می یابد، به عبارت دیگر متغیرها خلاف جهت یکدیگر تغییر می کنند. درباره این تكنيك در فصل ۱۱ بحث خواهد شد.

### **همبستگی جزئی**

همبستگی جزئی بسط همبستگی پیرسون است و به شما امکان می دهد كه اثر متغیرهای مداخله گر دیگری را كنترل كنید. همبستگی جزئی اثر متغیرهای مداخله گر را حذف می كند و به شما امکان می دهد كه تصویر صحیح تری از رابطه میان دو متغیر به دست آورید. همبستگی جزئی در فصل ۱۲ مورد بحث قرار خواهد گرفت.

### **رگرسیون چندگانه**

رگرسیون چندگانه هم بسط همبستگی است و هنگامی به کار می رود كه شما می خواهید توانایی پیش بینی تعدادی از متغیرهای مستقل را بر روی يك متغیر پیوسته وابسته بررسی كنید. انواع گوناگون رگرسیون چند متغیره به شما كمك می كند تا بهترین ترکیب متغیرهای مستقل برای پیش بینی متغیر وابسته را بیابید.

## تحلیل عاملی

تحلیل عاملی به شما می‌کند تا مجموعه بزرگی از متغیرها یا گویه‌های طیف را فشرده کنید و تعداد کمتری بعد یا عامل به دست آورید که بحث درباره آنها آسانتر است. در واقع کار تحلیل عاملی آن است که متغیرهای شما را خلاصه می‌کند یعنی متغیرهایی را که با هم همبستگی بالایی دارند و تشکیل دسته‌ای را می‌دهند در کنار هم قرار می‌دهد تا با هم یک عامل را تشکیل دهند. این تکنیک معمولاً هنگام ساختن و بهبود دادن طیفها و سنجها مورد استفاده قرار می‌گیرد تا ساختار پنهان موجود در آنها مشخص شود. در فصل چهاردهم در این باره بحث شده است.

## خلاصه

تمام تحلیل‌هایی که در بالا توضیح داده شدند، به بررسی رابطه میان متغیرهای پیوسته می‌پردازند. اگر متغیرهای شما مقوله‌ای هستند، می‌توانید از آزمون کی دو برای وابستگی یا استقلال برای بررسی رابطه آنها استفاده کنید (به عنوان مثال اگر می‌خواهید ببینید آیا کسانی که از انجام یک آزمایش کناره‌گیری کرده‌اند از جنس خاصی هستند یا خیر). در چنین وضعیتی آنچه برای شما مهم است تعداد کیسها در هر مقوله است (مردان/زنانی که آزمایش را انجام داده‌اند/ نداده‌اند) و نه نمره آنها از یک طیف.

برخی تکنیکهایی که شما باید درباره آنها اطلاعاتی داشته باشید و در این کتاب درباره آنها بحث مفصلي نشده است، عبارتند از:

- **تحلیل تابع تمایز** که هنگامی استفاده می‌شود که شما می‌خواهید توانایی پیش‌بینی مجموعه‌ای از متغیرهای مستقل بر یک سنج مقوله‌ای وابسته را بیابید. به عنوان مثال شما می‌خواهید بدانید کدام متغیر به بهترین شکل عضویت در گروه را پیش‌بینی می‌کند. در اینجا متغیر وابسته معمولاً معیار واضحی دارد (قبولی/مردودی، انصراف/ادامه کار).
- **رگرسیون لجستیک** که مشابه تحلیل تابع تمایز است اما با این تفاوت که در اینجا متغیرهای مستقل یا پیش‌بین می‌توانند پیوسته یا مقوله‌ای باشند و یا ترکیبی از هر دو. متغیر وابسته متغیری مقوله‌ای است.
- **همبستگی کانونی** هنگامی استفاده می‌شود که شما می‌خواهید رابطه میان دو مجموعه از متغیرها را بررسی کنید. به عنوان مثال یک پژوهشگر علاقمند است بداند که چگونه مجموعه‌ای از متغیرهای جمعیت‌شناختی با مجموعه‌ای از متغیرهایی که مربوط به بهزیستی و تطبیق هستند، رابطه دارند.

- **مدلسازی معادلات ساختاری** تقریباً جدیدتر از بقیه است و تکنیکی پیچیده تر است که به شما امکان می دهد مدل های گوناگون مربوط به رابطه مجموعه ای از متغیرها را بررسی کنید. اساس این تکنیک رگرسیون چندگانه و تحلیل عاملی است و به شما امکان می دهد اهمیت هر متغیر مستقل در مدل را بیابید و برازش کلی مدل با داده های خود را بررسی کنید. SPSS مدلسازی معادلات ساختاری را انجام نمی دهد اما نرم افزاری قابل اضافه شدن به آن است که این کار را انجام می دهد. این نرم افزار AMOS نام دارد. نرم افزار رایج تر برای انجام مدلسازی معادلات ساختاری Lisrel نام دارد.

### بررسی تفاوت های میان گروهی

این دسته شامل تکنیک هایی می شود که به شما کمک می کنند بررسی کنید که آیا تفاوت معنی دار آماری میان چند گروه وجود دارد یا خیر. بیشتر این تحلیل ها میانگین یک متغیر در هر گروه را بر اساس یک یا چند متغیر وابسته با هم مقایسه می کنند. آماره های گوناگون اما مرتبطی در این دسته جای دارند. تکنیک های اصلی در اینجا به طور خلاصه توضیح داده شده اند.

### T-test ها

تی تست زمانی به کار می روند که شما دو گروه (به عنوان مثال زنان و مردان) یا دو مجموعه داده دارید (مثلاً قبل و بعد از انجام آزمایش) و می خواهید نمره آنها از یک متغیر پیوسته را با هم مقایسه کنید. دو نوع اصلی تی تست وجود دارد. تی تست های نمونه جفت شده (که آن را سنجه های تکراری هم می نامند). این نوع تی تست زمانی به کار می رود که شما می خواهید نمرات کیسهای خود را در زمان یک با نمرات همان کیسها در زمان ۲ مقایسه کنید (معمولاً هنگامی که مداخله یا آزمایش مورد نظر انجام گرفت). این نمونه ها به هم وابسته هستند چون کیسهای شما در زمان های ۱ و ۲ یکسان بوده اند. تی تست های نمونه های مستقل هنگامی مورد استفاده قرار می گیرند که شما دو گروه متفاوت (مستقل) از افراد (مردان و زنان) را می خواهید با هم مقایسه کنید. در چنین موقعیتی شما اطلاعات را تنها در یک زمان جمع آوری می کنید، اما این اطلاعات مربوط به دو گروه متفاوت است. تی تستها در فصل ۱۵ مورد بحث قرار گرفته اند.

### آنالیز واریانس یک طرفه

آنالیز واریانس یک طرفه مانند تی تست است، اما هنگامی استفاده می شود که شما نمرات دو گروه یا بیشتر را از یک متغیر پیوسته می خواهید با هم مقایسه کنید. یک طرفه نامیده می شود چون

شما تنها به تاثیر یک متغیر مستقل بر متغیر وابسته خود توجه می کنید. آنالیز واریانس یک طرفه (ANOVA) به شما می گوید که آیا گروههای شما با هم متفاوت هستند یا خیر. اما نمی گوید که تفاوت در کجا وجود دارد (مثلا بین گروه ۱ و گروه ۳ و یا بین گروه ۲ و گروه ۳). شما می توانید آزمونهای بعد وقوع (post-hoc) را انجام دهید تا دریابید که کدام گروهها به طور معنی داری با هم تفاوت دارند. همچنین می توانید با استفاده از مقایسه های برنامه ریزی شده، انتخاب کنید که گروههای خاصی با هم مقایسه شوند و همه گروهها را یکجا با هم مقایسه نکنید. مانند تی تستها، دو نوع اصلی هم برای آنالیز واریانس وجود دارد: آنووا سنجه های تکراری (افراد یک سان در بیش از دو موقعیت متفاوت)، و آنووا بین گروهی (یا نمونه های مستقل) که در آن شما میانگین نمرات را در دو گروه یا بیشتر با هم مقایسه می کنید. آنووا یک طرفه در فصل ۱۶ توضیح داده شده است.

### **آنالیز واریانس دو طرفه**

آنالیز واریانس دو طرفه به شما امکان می دهد که تاثیر دو متغیر مستقل بر یک متغیر وابسته را ببیند. مزیت استفاده از آنالیز واریانس دو طرفه آن است که شما می توانید در آن اثر تعامل را بررسی کنید. به عنوان مثال هنگامی که شما فرض می کنید خوشبینی با سن افزایش پیدا می کند، اما تنها در میان مردان. این تکنیک تاثیرات اصلی را هم آزمون می کند، که اثر کلی هر متغیر مستقل است (مانند سن و جنس). دو نوع آنووا دو طرفه متفاوت وجود دارد: آنووا بین گروهی (هنگامی که گروههای متفاوت هستند) و آنووا سنجه های تکراری (هنگامی که افراد یکسانی در موقعیت های مختلف مورد سنجش قرار گرفته اند). در برخی از طرحهای پژوهشی بین گروهی و سنجه های تکراری با هم ترکیب می شوند. این طرحهای تحقیق را ترکیب میان-بین طرحی (Mixed Between-Within Designs) یا نمودار تقسیم شده (Split Plot) می نامند. آنالیز واریانس دو طرفه در فصل ۱۷ و طرحهای ترکیبی هم در فصل ۱۸ مورد بحث قرار گرفته است.

### **آنالیز واریانس چند متغیری**

آنالیز واریانس چند متغیری (MANOVA) هنگامی استفاده می شود که شما می خواهید گروههای خود را از نظر چند متغیر وابسته که به هم مربوطند با هم مقایسه کنید. به عنوان مثال تاثیر تیمارهای مختلف بر تعدادی از متغیرهای وابسته (مانند عصبیت، افسردگی و علائم فیزیکی). آنووا چند متغیری را می توان با طرحهای فاکتوریال یک طرفه، دو طرفه یا بیشتر به کار برد که در آنها یک، دو یا بیشتر متغیر وابسته وجود داشته باشند. مانووا در فصل ۱۹ مورد بررسی قرار گرفته است.

## آنالیز کوواریانس

آنالیز کوواریانس (ANCOVA) هنگامی به کار می رود که شما می خواهید مداخله متغیری (هم تغییری) را کنترل کنید. این کار زمانی مفید است که فکر می کنید گروههای شما از نظر متغیر دیگری با هم متفاوت هستند و تاثیرات آن متغیر می تواند تاثیر گروهبندی بر متغیر وابسته را تحت تاثیر خود قرار دهد. برای آنکه بتوانید تاثیر متغیر وابسته را با خلوص بیشتری ببینید، آنکووا تاثیر متغیر هم تغییر را حذف می کند. آنالیز کوواریانس را می توان به عنوان بخشی از یک طرح یک طرفه، دو طرفه یا چند متغیری به کار برد. آنکووا در فصل بیستم مورد بررسی قرار گرفته است.

## خلاصه ویژگی های تکنیکهای مختلف آماری

هدف	نمونه سوال پژوهشی	آماره پارامتری	جایگزین ناپارامتری	متغیر مستقل	متغیر وابسته	ویژگی های مهم
بررسی روابط	چه رابطه ای میان جنسیت و ترک تحصیل وجود دارد؟	ندارد	کی دو فصل ۲۱	یک متغیر مقوله ای مثل جنسیت	یک متغیر مقوله ای مثل ترک/تمام تحصیل	تعداد کیسها در مقوله مورد توجه است و نه نمره آنها.
	آیا رابطه ای میان سن و خوشبینی وجود دارد؟	همبستگی پیرسون (r) فصل ۱۱	همبستگی رتبه ای اسپیرمن (rho) فصل ۲۱	دو متغیر پیوسته مثل سن و خوشبینی		یک نمونه که با دو متغیر سنجیده شده است، یا سنجه های زمان ۱ و ۲.
	بعد کنترل پاسخ های مطلوب اجتماعی، آیا هنوز بین خوش بینی و رضایت از زندگی رابطه وجود دارد؟	همبستگی جزئی فصل ۱۲	ندارد	دو متغیر وابسته و یک متغیر وابسته که می خواهید کنترل شود. خوش بینی و رضایت از زندگی با کنترل مطلوبیت اجتماعی		یک نمونه با نمراتی از دو سنجه متفاوت یا سنجه های یکسان در زمانهای ۱ و ۲
مقایسه بین گروهی	تا چه حد واریانس نمرات رضایت از زندگی را می توان بوسیله اعتماد به نفس، استرس ادراک شده و خوشبینی توضیح داد؟	رگرسیون چندگانه فصل ۱۳	ندارد.	مجموعه ای از دو یا بیشتر متغیر مستقل پیوسته اعتماد به نفس، استرس ادراک شده، خوش بینی	یک متغیر وابسته پیوسته رضایت از زندگی	یک نمونه که تمام متغیرها در آن سنجیده شده است.
	ساختار مکنون در گویه هایی که با هم یک طیف را می سازند کدام است؟	تحلیل عاملی فصل ۱۴	ندارد	مجموعه ای از گویه های پیوسته و مرتبط مانند گویه هایی که طیف تاثیر مثبت و منفی را می سازند		یک نمونه، سنجه های چندگانه
	آیا مردان بیش از زنان ترک تحصیل می کنند؟	ندارد	کی دو فصل ۲۱	یک متغیر وابسته مقوله ای مانند جنسیت	یک متغیر مقوله وابسته مانند ترک/تمام تحصیل	شما می خواهید تعداد افراد در هر دسته را بررسی کنید و نه نمره آنها را.
	آیا مردان از زنان خوشبین تر هستند؟	تی تست نمونه های مستقل فصل ۱۵	یو من ویتنی فصل ۲۱	یک متغیر مستقل مقوله ای (دو سطحی) مثل جنسیت	یک متغیر وابسته پیوسته مانند نمره خوش بینی	دو گروه: تفاوت افراد در هر گروه.
	آیا میزان عصبيت در پاسخگویان بین زمانهای ۱ و ۲ تغییر کرده است؟	تی تست نمونه های جفت شده فصل ۱۵	آزمون رتبه نشاندار و یلکاکسون <sup>۳</sup> فصل ۲۱	یک متغیر مستقل مقوله ای (دو سطحی) زمان ۱/ زمان ۲	یک متغیر وابسته پیوسته مانند نمره عصبيت	افراد یکسان در موقعیتهای متفاوت
	آیا میان نمره خوشبینی کسانی که در گروههای سنی کمتر از ۳۵ سال، ۳۶ تا ۴۹ و بالای ۵۰ سال هستند تفاوت وجود دارد؟	آنسوا بین گروهی یک طرفه فصل ۱۶	کروسکال والیس فصل ۲۱	یک متغیر وابسته مقوله ای (سه سطح یا بیشتر) مانند گروه سنی	یک متغیر پیوسته مانند نمره خوشبینی	سه گروه یا بیشتر که در هر گروه افراد متفاوتی قرار دارند.

<sup>3</sup> Wilcoxon Signed-Rank test

	دارد؟				
سه گروه یا بیشتر: افراد یکسان در موقعیت های متفاوت	یک متغیر وابسته پیوسته نمره عصبيت	یک متغیر مقوله ای مستقل (سه سطح یا بیشتر) زمان ۱/زمان ۲/زمان ۳	آزمون فریدمن فصل ۲۱	آنووا سنج های تکراری یک طرفه فصل ۱۶	آیا عصبيت پاسخگویان در زمانهای ۱ و ۲ و ۳ تغییر کرده است؟
دو گروه یا بیشتر برای هر متغیر مستقل: افراد گوناگون در هر گروه	یک متغیر وابسته پیوسته مانند نمره خوش بینی	دو متغیر مستقل مقوله ای (دو سطحی یا بیشتر) گروه سنی، جنس	ندارد	آنووا بین گروهی دو طرفه فصل ۱۷	آیا نمره خوشبینی زنان و مردان در گروههای سنی کمتر از ۳۵ سال، بین ۳۶ تا ۴۹ و بیش از ۵۰ سال متفاوت است؟
دو گروه یا بیشتر با افرائ متفاوت در هر گروه که هر کدام در دو موقعیت یا بیشتر اندازه گیری شده اند.	یک متغیر وابسته پیوسته نمره آزمون ترس از آمار	یک متغیر مستقل بین گروهی (دو سطح یا بیشتر) یک متغیر مستقل درون گروهی (دو سطح یا بیشتر) نوع مداخله زمان است	ندارد	آنووا ترکیبی بین گروهی - درون گروهی فصل ۱۸	کدام شیوه (مهارتهای ریاضی/وجود آوردن اعتماد) در کاهش ترس افراد از آمار موثرتر است که در سه دوره زمانی اندازه گیری شده است؟
	دو یا بیشتر متغیر پیوسته وابسته مانند عصبيت، افسردگی و استرس ادراک شده	دو یا بیشتر متغیر مقوله ای وابسته (دو سطحی یا بیشتر) گروه سنی، جنس	ندارد	آنووا چند متغیری (مانووا) فصل ۱۹	آیا بین زنان و مردان، در سه گروه سنی مختلف نوع تطبیق آنها (عصبيت، افسردگی و استرس ادراک شده) تفاوت وجود دارد؟
	یک متغیر وابسته پیوسته نمره ترس از امتحان آمار در زمان ۲	دو تا یا بیشتر متغیر مقوله ای مستقل (دو سطح یا بیشتر) یک متغیر هم تغییر پیوسته نوع مداخله آزمون ترس از آمار در زمان ۱	ندارد	آنوالیز کوواریانس (آنکووا) فصل ۲۰	آیا تفاوت معنا میان نمره ترس از آمار در میان افرادی که در گروه ریاضی هستند با کسانی که در گروه اعتمادسازی وجود دارند، با کنترل نمره آنها در زمان ۱؟



## فصل ۱۱

### همبستگی

همبستگی برای توصیف قدرت و جهت رابطه خطی بین دو متغیر به کار می رود. آماره گوناگونی وجود دارند که در SPSS بسته به سطح سنجش در دسترس هستند. در این فصل همبستگی گشتاوری پیرسون مورد بحث قرار می گیرد. ضریب همبستگی پیرسون برای متغیرهای فاصله ای (پیوسته) طراحی شده است. البته از آن می توان هنگامی که شما یک متغیر پیوسته (مانند نمرات اعتماد به نفس) و یک متغیر مقوله ای دو حالتی دارید، هم استفاده کرد (مانند جنسیت: مرد/زن). همبستگی رتبه ای اسپیرمن (که برای داده های ترتیبی یا رتبه بندی شده طراحی شده است) در فصل تکنیکهای ناپارامتری مورد بررسی قرار گرفته است (فصل ۲۱ را ببینید).

SPSS دو نوع همبستگی را برای شما محاسبه می کند. ابتدا همبستگی دو متغیری ساده را محاسبه می کند و با نام همبستگی درجه صفر هم شناخته می شود. SPSS همچنین به شما امکان می دهد که رابطه بین دو متغیر را هنگامی که متغیر دیگری کنترل شود بررسی کنید. این یکی همبستگی جزئی نام دارد. در این فصل به همبستگی ساده دو متغیری پرداخته می شود و در فصل ۱۲ به همبستگی جزئی پرداخته خواهد شد.

ضریب همبستگی پیرسون ( $r$ ) تنها مقادیری بین  $-1$  و  $+1$  می تواند داشته باشد. علامت مثبت نشان دهنده آن است که همبستگی مثبت وجود دارد (هنگامی که یکی از متغیرها افزایش می یابد دیگری هم افزایش می یابد) و هنگامی که علامت منفی است نشان دهنده آن همبستگی منفی وجود دارد (هنگامی که یکی افزایش می یابد دیگری کاهش می یابد). مقدار مطلق ضریب همبستگی (بدون در نظر گرفتن علامت آن) قدرت همبستگی را نشان می دهد. اگر مقدار همبستگی دقیقاً ۱ یا  $-1$  باشد یعنی می توان با دانستن یکی از مقادیر دیگری را دقیقاً تعیین کرد. نمودار پراکنندگی چنین رابطه ای دقیقاً یک خط راست است. از سوی دیگر اگر مقدار ضریب همبستگی صفر باشد نشان دهنده آن است که هیچ رابطه ای میان این دو متغیر وجود ندارد و دانستن مقدار یکی از متغیرها هیچ کمکی به تعیین مقدار متغیر دیگر نمی کند. نمودار پراکنندگی چنین رابطه ای نشان دهنده آن است نقاط بدون هیچ الگوی خاصی به صورت دایره ای پراکنده شده اند.

استفاده از همبستگی نکاتی دارد که باید به آن پرداخت. این نکات عبارتند از: تاثیر روابط غیرخطی، نقاط دورافتاده، محدودیت دامنه تغییر متغیرها، همبستگی در مقابل علیت و معنی داری آماری و معنی داری عملی.

در مورد تاثیر رابطه غیر خطی باید گفت که در محاسبه ضریب همبستگی پیرسون فرض می شود که رابطه میان دو متغیر خطی است. ممکن است رابطه ای غیر خطی اما قوی میان دو متغیر وجود داشته اما ضریب همبستگی پیرسون نمی تواند این رابطه را نشان دهد. پس باید قبل از آنکه به سراغ ضریب همبستگی برویم، نمودار پراکندگی برای دو متغیر رسم کنیم، اگر رابطه خطی مشاهده کردیم به سراغ همبستگی پیرسون می رویم و اگر چنین نبود، از الگوهای رابطه غیرخطی میان دو متغیر باید استفاده کرد البته از بحث فعلی ما خارج است.

نقاط دورافتاده گاهی می توانند به مقدار همبستگی شما آسیب زیادی بزنند بنابراین باید قبل از آنکه همبستگی بگیرید، آنها را مشخص کنید و درباره شان تصمیم بگیرید (یا آنها را حذف کنید و یا مقادیری جایگزین آنها کنید).

### مثال

برای بررسی چگونگی انجام تکنیک همبستگی پیرسون و تفسیر آن به مثالی خواهیم پرداخت که داده های آن مربوط به فایل داده survey.sav است. این فایل برای بررسی عوامل موثر تعادل روانی و بهزیستی افراد تهیه شده است. در این مثال می خواهیم به بررسی رابطه میان احساس کنترل افراد بر احساساتشان و میزان استرس ادراک شده آنان پردازیم. هر دو این متغیرها طیف هستند.

متغیر احساس افراد از کنترل آنان بر احساسات خود tpcoiss نام دارد. این متغیر حاصل جمع ۱۸ گویه است (pc1 تا pc18). نمره ای که افراد از این طیف می گیرند بین ۱۸ تا ۹۰ خواهد بود (پاسخها لیکرت ۵ تایی بوده اند). متغیر استرس ادراک شده هم سازه ای است که از جمع ۱۰ گویه ساخته شده است (pss1 تا pss10). نمرات افراد از این طیف هم بین ۱۰ تا ۵۰ خواهد بود.

### تحلیل اولیه برای انجام همبستگی

قبل از آنکه تحلیل همبستگی را انجام دهیم بهتر آن است که نمودار پراکندگی را رسم کنیم. با این کار می توانیم مفروضات خطی بودن و هم پراشی متغیرها را بررسی کنیم. بررسی نمودار پراکندگی ذهنیت بهتری به شما درباره ماهیت رابطه بین متغیرها می دهد.

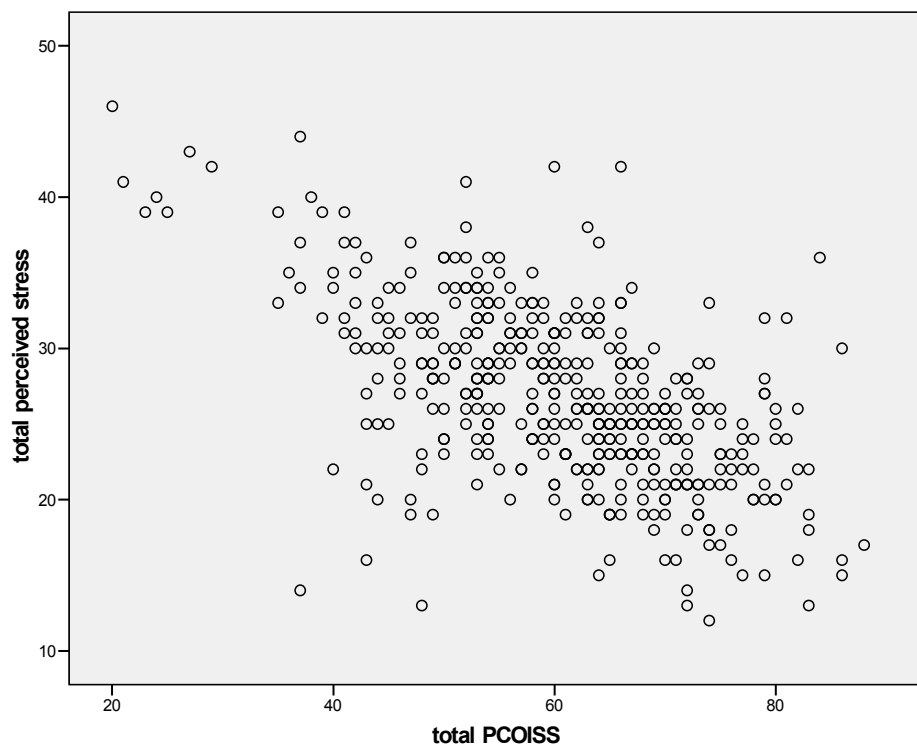
### شیوه رسم نمودار پراکندگی (scatterplot)

۱. از منوی بالای صفحه ابتدا Graphs سپس Scatter را انتخاب کنید.

۲. گزینه Simple را انتخاب و سپس Define را بزنید.

۳. بر روی متغیر اول کلیک کنید و آن را به قسمت محور Y ها منتقل کنید (این متغیر بر روی محور عمودی قرار خواهد گرفت). معمولاً متغیر وابسته را در محور Y ها می گذارند (در مثال ما استرس ادراک شده).
۴. متغیر دوم را به قسمت محور X ها منتقل کنید. (این متغیر بر روی محور افقی قرار خواهد گرفت). معمولاً متغیر مستقل را در این قسمت قرار می دهند. (در مثال ما PCOISS).
۵. اگر می خواهید برای نمودار خود عنوانی بنویسید، آن را در قسمت Titles وارد کنید و سپس Continue و OK را بزنید.

خروجی به دست آمده از این رویه چنین خواهد بود.



### تفسیر خروجی نمودار پراکندگی (scatterlot)

با استفاده نمودار پراکندگی می توان برخی ویژگی های رابطه میان این دو متغیر را بررسی

کرد.

## گام اول: بررسی نقاط دورافتاده

در نمودار پراکندگی به دنبال داده های دورافتاده باشید. آنهایی که مقادیر بسیار کم یا بسیار زیاد دارند. یا اینکه از جمع داده ها دورافتاده اند. داده های بسیار دورافتاده ارزش آن را دارند که بررسی شوند- اینکه آیا داده ها به درستی وارد شده اند، آیا می توان این داده ها را خطا دانست؟ دورافتاده می توانند تحلیل ما را بسیار تحت تاثیر خود قرار می دهند، پس باید بررسی شوند. برخی از متون آماری پیشنهاد داده اند که دورافتاده ها را حذف کنیم و برخی دیگر پیشنهاد داده اند که مقدار آنها تغییر داده شود و داده نزدیکی که دورافتاده نیست جایگزین شود (فصل ۶ را ببینید).

## گام دوم: بررسی توزیع نقاط داده ها

توزیع داده می تواند چند نکته را به شما بگوید:

- آیا نقاط در تمام صفحه پخش شده اند؟ این نشان دهنده همبستگی بسیار ضعیفی است.
- آیا نقاط حول یکدیگر توزیع شده اند و شکلی شبیه مداد را ساخته اند؟ این نشان دهنده همبستگی بسیار قوی است.
- آیا می توانید خطی از میان جمع اصلی داده ها بکشید، یا اینکه یک منحنی بهتر می تواند از میان آنها عبور کند؟ اگر منحنی مشاهده شود (نشان دهنده آن است که همبستگی بین دو متغیر خطی نیست)، پس نباید از همبستگی پیرسون استفاده کرد چرا که همبستگی پیرسون فرض می کند رابطه خطی است.
- شکل خوشه داده ها چگونه است؟ اگر در ابتدا باریک و سپس ضخیم است نشان دهنده آن است که شرط هم پراشی به خوبی اضاء نمی شود.

## گام سوم: مشخص کردن جهت همبستگی میان متغیرها

نمودار پراکندگی به شما نشان می دهد که همبستگی میان دو متغیر مثبت است یا منفی. خطی که از میان داده ها می توان کشید، اگر از چپ به راست پایین بیاید آنگاه همبستگی منفی و اگر بالا برود همبستگی مثبت خواهد بود. در این مثال ما همبستگی منفی نسبتاً قوی داریم.

## شیوه محاسبه همبستگی گشتاوری پیرسون

۱. از منوی بالای صفحه ابتدا Analyze و سپس Correlate و بعد از آن Bivariate را انتخاب کنید.

۲. دو متغیر مورد نظر خود را انتخاب کنید و آن‌ها را به قسمت Variables منتقل کنید (در مثال ما tpstress و pcoiss). شما در این قسمت می‌توانید متغیرهای بسیاری را وارد کنید. در خروجی ماتریسی ارائه خواهد شد همبستگی میان هر جفت متغیری را به شما خواهد داد. اگر متغیرهای زیادی را وارد کنید این ماتریس بسیار بزرگ خواهد بود.

۳. در قسمتی که Pearson نام دارد چک کنید که گزینه دو دنباله (2 tail) فعال باشد. این گزینه به آن معنی است که شما هیچ پیش‌بینی درباره مثبت بودن و منفی بودن همبستگی ندارید. اگر دلایلی دارید که جهت همبستگی را مشخص می‌کند می‌توانید گزینه one-tail را انتخاب کنید.

۴. بر روی دکمه Options کلیک کنید.
- در قسمت Missing Values گزینه Exclude cases pairwise را انتخاب کنید.
  - در قسمت Options می‌توانید میانگینها و انحراف معیارها را هم ببخشید. بر روی دکمه Continue و سپس OK کلیک کنید.

خروجی این رویه در شکل زیر نشان داده شده است

Correlations			
		total perceived stress	total PCOISS
total perceived stress	Pearson Correlation	1	-.581**
	Sig. (2-tailed)		.000
	N	433	426
total PCOISS	Pearson Correlation	-.581**	1
	Sig. (2-tailed)	.000	
	N	426	430

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

### تفسیر خروجی همبستگی

خروجی Correlation جدولی است که در آن ضریب همبستگی برای تمام زوج متغیرهایی که شما خواسته‌اید ارائه شده‌اند. اگر شما متغیرهای زیادی را انتخاب کرده باشید این جدول بسیار بزرگ خواهد بود. برای هر جفت متغیر، مقدار ضریب همبستگی، تعداد کیسها و مقدار معنی داری ارائه شده-

اند. نکاتی در این خروجی وجود دارند که باید به آنها توجه شود. این جا این نکات را بررسی خواهیم کرد.

### گام اول: بررسی اطلاعات داده شده درباره نمونه

اولین چیزی که در جدول خروجی همبستگی می بینیم تعداد کیسهایی است که در همبستگی شرکت کرده اند (N). آیا این تعداد درست هستند؟ اگر تعداد کیسها بسیار کم شده اند، یعنی تعداد داده های گم شده شما بسیار زیاد هستند به دنبال دلیل باشید. شاید فراموش کرده اید که گزینه Exclude cases pairwise را فعال کنید. استفاده از شیوه Exclude cases listwise (گزینه دیگری که در پنجره Correlation وجود دارد) برای حذف داده ها، هر کیسی را که در هر متغیری داده گمشده دارد حذف می کند. این کار ممکن است گاهی تعداد کیسهای شما را بسیار کاهش می دهد. در این مثال ۴۲۶ کیس در محاسبه همبستگی حضور داشته اند. روشی که ما برای حذف داده های گمشده انتخاب کردیم، کیسهایی را حذف می کند که در یکی از این دو متغیر داده گمشده دارد.

### گام دوم: مشخص کردن جهت همبستگی

اگر در کنار عدد داده برای ضریب همبستگی در جدول علامت منفی وجود داشته باشد، نشانه آن است که همبستگی بین دو متغیر منفی بوده است. تفسیر این همبستگی به شیوه نمره دهی متغیرها بستگی دارد. همواره این موضوع را در پرسشنامه خود بررسی کنید و به یاد داشته باشید که قبل از انجام همبستگی باید کدگذاری گویه هایی را که منفی پرسیده شده اند، معکوس کنید. این موضوع بسیار مهم است و بسیاری از دانشجویان در این زمینه دچار اشتباه می شوند. در مثال ما ضریب همبستگی منفی (۰/۵۸۱-) نشان دهنده آن است که هر چه افراد احساس کنترل بیشتری بر احساسات خود دارند، استرس کمتری را تجربه می کنند.

### گام سوم: تعیین قدرت همبستگی

سومین چیزی که باید به آن توجه کرد اندازه ضریب همبستگی پیرسون (r) است. این مقدار می تواند بین ۱ و ۱- باشد و نشان دهنده میزان قدرت همبستگی بین دو متغیر است. مولفان مختلف نظرات متفاوتی درباره تفسیر میزان قدرت همبستگی ارائه کرده اند. یکی از پیشنهاد ها چنین است:

ضعیف	-0.29 تا -0.10	یا	0.29 تا 0.10	$r =$
میانه	-0.49 تا -0.30	یا	0.49 تا 0.30	$r =$
قوی	-1.0 تا -0.50	یا	1.0 تا 0.50	$r =$

در مثال که مقدار همبستگی ۰/۵۸۱- بود، همبستگی قوی است.

#### گام ۴: محاسبه ضریب تعیین

برای به دست آوردن ایده ای درباره اینکه چه میزان از واریانس دو متغیر مشترک است می توان ضریب تعیین را محاسبه کرد. تنها کاری که برای محاسبه ضریب تعیین باید انجام داد به توان دو رساندن ضریب همبستگی است. برای آنکه آن را در مقیاس درصد بیان کنید هم کافی ضریب تعیین محاسبه شده در ۱۰۰ ضرب کنید. به عنوان مثال اگر ضریب همبستگی بین دو متغیر برابر ۰/۲ باشد، آنگاه ضریب تعیین برابر ۰/۴ و بیان درصدی آن ۴ درصد خواهد بود. این نشان دهنده آن است که واریانس مشترک بین دو متغیر بسیار کم است.

در مثال ما ضریب همبستگی ۰/۵۸۱- بود که ضریب تعیین آن می شود ۳۳/۷۶ درصد. یعنی ۳۳/۷۶ درصد از واریانس بین دو متغیر مشترک است. یعنی کنترل ادراک شده نزدیک به ۳۴ درصد واریانس متغیر استرس ادراک شده را تبیین می کند. این میزان در حوزه علوم اجتماعی بالا به حساب می آید.

#### گام پنجم: بررسی سطح معنی داری

نکته بعدی که باید به آن توجه کرد سطح معنی داری است (که با عنوان sig. 2 tailed شده است). این قسمت بسیار مشکل است و باید با احتیاط با آن برخورد کرد. معنی داری  $r$  بسیار متأثر از حجم نمونه است. در نمونه کوچک ( $N=30$ ) ممکن است شما همبستگی های متوسط داشته باشید که معنی داری آماری را در سطح  $p<0.05$  نداشته باشند. اما در نمونه های بزرگ (مثلاً  $N=100$ ) ممکن است همبستگی های خیلی ضعیف معنی دار باشند. در این موارد بسیاری از مولفان گفته اند که مقدار معنی داری باید گزارش شود اما نباید به آن توجه کرد و باید به مقدار قدرت همبستگی توجه کرد.

#### ارائه نتایج حاصل از همبستگی

نتایج حاصل شده از مثال بالا باید به چنین شکلی در گزارش پژوهش ارائه داد:

رابطه میان کنترل ادراک شده بر حالات درونی (که به وسیله متغیر PCOISS سنجیده شد) و استرس ادراک شده (که بوسیله متغیر tpstress سنجیده شد) با استفاده از ضریب همبستگی گشتاوری پیرسون بررسی شد. تحلیل اولیه ای انجام شد تا عدم تخطی داده ها از مفروضات نرمال بودن، خطی بودن و هم پراشی داده ها بررسی شوند. همبستگی منفی بین متغیرها وجود داشت [ $r=-.58$ ,  $n=426$ ,  $p<.0005$ ] که بر اساس آن مقادیر بالای کنترل ادراک شده با سطوح پایین مقادیر استرس ادراک شده همبستگی دارد.

## فصل ۱۲

### همبستگی جزئی

همبستگی جزئی مشابه همبستگی گشتاوری پیرسون است، جز آنکه به شما امکان می دهد یک متغیر اضافه را کنترل کنید. این متغیر معمولاً آن متغیری است که شما فکر می کنید بر هر دو متغیر تاثیر می گذارد. با حذف آماری تاثیر متغیرهای مداخله گر می توانید تصویری روشن تر و صحیح تر از رابطه میان این دو متغیر ارائه می دهد.

مداخله یک متغیر زمانی رخ می دهد که آن متغیر حداقل تاحدی بر دو متغیری که شما می خواهید رابطه میان آنها بررسی کنید تاثیر می گذارد. این موضوع می تواند به طور تصنعی تاثیر فاحشی بر میزان ضریب همبستگی میان دو متغیر بگذارد. اگر متغیر مداخله گر کنترل شود آنگاه از رابطه میان دو متغیر مطمئن تر خواهیم بود.

### مثال

برای توضیح چگونگی استفاده از همبستگی جزئی همان مثال فصل ۱۱ را بررسی خواهیم کرد تنها با این تفاوت که یک متغیر اضافه را کنترل خواهیم کرد. ما می خواهیم رابطه میان متغیرهای کنترل بر حالات درونی و استرس ادراک شده با کنترل متغیری که آن را سوءگیری مطلوبیت اجتماعی می-نامیم است. این متغیر مربوط به آن است که مردم معمولاً تمایل دارند خود را مثبت تر از آنچه هستند نشان دهند. این متغیر بوسیله طیف سنجش مطلوبیت اجتماعی مارلو-کراون سنجیده شده است. این متغیر tmarlow نامیده شده است.

### روش انجام همبستگی جزئی

۱. از منوی بالای صفحه گزینه Analyze و سپس Correlate بعد از آن Partial را انتخاب کنید.

۲. دو متغیری را که می خواهید همبستگی آنها را محاسبه کنید وارد قسمت Variables کنید. (متغیرهای PCOISS و tpstress).



۳. متغیری را که می خواهید کنترل کنید وارد قسمتی کنید که Controlling for نام دارد. در مثال ما tmarlow.

۴. انتخاب کنید که آزمون یک طرفه را می خواهید یا دو طرفه (در مثال ما دو طرفه انتخاب شده است).

۵. در گزینه Options:

- در قسمت Missing Values بر روی Exclude cases pairwise کلیک کنید.
- در قسمت Statistics بر روی گزینه Zero Order Correlation کلیک کنید.
- این کار به شما دو ماتریس همبستگی می دهد: ماتریس همبستگی اصلی بدون کنترل متغیر مزاحم و ماتریس همبستگی بعد از کنترل متغیر مزاحم.
- ۶. بر روی Continue و سپس OK کلیک کنید.

خروجی این چنین خواهد بود:

Correlations

Control Variables			total perceived stress	total PCOISS	total social desirability
-none-a	total perceived stress	Correlation	1.000	-.581	-.228
		Significance (2-tailed)	.	.000	.000
		df	0	424	426
	total PCOISS	Correlation	-.581	1.000	.295
		Significance (2-tailed)	.000	.	.000
		df	424	0	425
	total social desirability	Correlation	-.228	.295	1.000
		Significance (2-tailed)	.000	.000	.
		df	426	425	0
total social desirability	total perceived stress	Correlation	1.000	-.552	
		Significance (2-tailed)	.	.000	
		df	0	423	
	total PCOISS	Correlation	-.552	1.000	
		Significance (2-tailed)	.000	.	
		df	423	0	

a. Cells contain zero-order (Pearson) correlations.

### تفسیر خروجی همبستگی جزئی

در خروجی شما دو ماتریس خواهید داشت:

- ماتریس اول همان همبستگی های گشتاوری معمولی پیرسون است. اینجا همبستگی های بین تمام جفت هایی که می توان با سه متغیری که ما وارد کردیم ساخت محاسبه شده است. ملاحظه می کنید همبستگی میان tpcoiss و tpstress همان ۰/۵۸- است.
- جدول دوم مربوط به همبستگی میان دو متغیر هنگامی است که متغیر tmarlow کنترل شود. ملاحظه می شود که با کنترل این متغیر همبستگی جزئی دو متغیر tpcoiss و tpstress شده است ۰/۵۵-.

شما باید این دو همبستگی را با هم مقایسه کنید تا بفهمید آیا متغیر کنترل شده تاثیری بر رابطه میان دو متغیر داشته است یا خیر. در این مثال کنترل tmarlow تاثیر اندکی بر رابطه میان tpcoiss و tpstress داشته است (از ۰/۵۸- به ۰/۵۵-). این نشان می دهد که رابطه ای که میان tpcoiss و tpstress مشاهده می شود تحت تاثیر tmarlow نیست.

### ارائه نتایج مشاهده شده در همبستگی جزئی

نمونه ای از ارائه نتیجه همبستگی جزئی می تواند چنین باشد:

برای بررسی رابطه میان کنترل بر حالات درونی (که با tpcoiss سنجیده شده است) و استرس ادراک شده (که با tpstress سنجیده شده است) از همبستگی جزئی استفاده شده است و متغیر مطلوبیت اجتماعی کنترل شده است. تحلیل اولیه ای برای بررسی نرمال بودن، خطی بودن و هم پراشی انجام شد. همبستگی قوی و منفی میان این دو متغیر وجود داشت [ $r=-.55$ ,  $n=55$ ,  $p<.0005$ ]. بررسی همبستگی درجه صفرم ( $r=-.58$ ) نشان داد که متغیر tmarlow تاثیر اندکی بر رابطه قوی میان tpcoiss و tpstress داشته است.

## فصل ۱۳

### رگرسیون چند گانه

در این فصل به طور خلاصه شیوه انجام رگرسیون چند گانه در SPSS توضیح داده خواهد شد. لازم است که شما به توضیحاتی که در اینجا ارائه شده است اکتفا نکنید و به کتابهای آماری مراجعه کنید و مبانی نظری و مفروضات انجام رگرسیون را مطالعه کنید.

رگرسیون تنها یک تکنیک نیست بلکه مجموعه ای از تکنیکهایی است که به بررسی رابطه میان یک متغیر وابسته پیوسته و تعدادی متغیر (معمولاً پیوسته) مستقل می پردازد. اساس رگرسیون چند متغیره همبستگی است. اما مزیت رگرسیون چند متغیری آن است که تعدادی متغیر را با هم وارد مدل می کند و موجب می شود که مدل ما بیشتر به زندگی واقعی شبیه باشد. البته باید بدانید که نمی توان به سادگی هر متغیری را وارد رگرسیون چند متغیری کرد و انتظار داشت که نتیجه خوبی به دست آید. برای متغیرهایی که وارد مدل می کنید باید توضیح نظری داشته باشید. همچنین ترتیب وارد کردن متغیرها به مدل هم مهم است و برای آن هم باید دلیل و توضیحی منطقی داشته باشید.

با رگرسیون چند متغیری می توان به مجموعه ای از سوالات پژوهشی جواب داد. این تکنیک به شما می گوید مجموعه ای از متغیرها تا چه حد می توانند به پیش بینی یک متغیر وابسته پردازند. به عنوان مثال ممکن است شما بخواهید بدانید که مجموعه ای از متغیرهای زمینه ای تا چه حد می توانند تبیین کننده موفقیت تحصیلی دانش آموزان باشند. علاوه بر این رگرسیون چند متغیری می تواند به شما بگوید که افزودن متغیری خاص به مدل چه تاثیری بر توانایی پیش بینی مدل خواهد داشت. همچنین با این تکنیک می توان به کنترل آماری متغیری در مدل پرداخت. برخی از انواع اصلی سوالات پژوهشی که رگرسیون چند متغیری می تواند به آنها پاسخ گوید چنین اند:

- مجموعه ای از متغیرهای مستقل تا چه حد می توانند به پیش بینی یک متغیر وابسته پردازند.
- کدام متغیر در میان مجموعه ای از متغیرها بهترین پیش بینی کننده متغیر وابسته است.
- آیا هنگامی که اثر چند متغیر دیگر کنترل شوند، باز هم متغیر خاصی می تواند پیش بینی کننده خوبی برای متغیر وابسته باشد.

## انواع اصلی رگرسیون چند متغیره

رگرسیون چند متغیری انواع گوناگونی دارد که بسته به نوع سوال پژوهشی که می خواهید به آنها پاسخ گوید، می توانید از آنها استفاده کنید. این انواع اصلی عبارتند از:

- استاندارد یا همزمان
- سلسله مراتبی یا تناوبی
- گام به گام

البته مولفان مختلف اسامی گوناگونی را برای نام بردن انواع رگرسیون گند متغیری به کار برده اند که ممکن است گیج کننده باشد. اما به هر حال انواع اصلی رگرسیون چند متغیری چنین است.

### رگرسیون چندگانه استاندارد

در رگرسیون چندگانه استاندارد، تمام متغیرهای مستقل (پیش بینها) همزمان وارد مدل می-شوند. هر متغیر با توانایی پیش بینی آن ارزیابی می شود. این روش رایج ترین روش انجام رگرسیون چند متغیری است و زمانی انجام می شود که شما می خواهید بدانید چند متغیر در کنار هم تا چه حد توانایی پیش بینی یک متغیر وابسته را دارند.

### رگرسیون چندگانه سلسله مراتبی

در رگرسیون سلسله مراتبی ( که آن را تناوبی هم می گویند) متغیرهای به ترتیبی وارد مدل می شوند که خود محقق تعیین می کند. این ترتیب ورود متغیرها به مدل را محقق بر اساس مبانی نظری خود مشخص می کند. هر متغیر (یا بلوکی از متغیرها) در هر مرحله وارد مدل می شوند، تا مشخص شود که هر کدام از آنها چه میزان به قدرت مدل می افزایند. این کار به آن معنی است که متغیرهایی که از قبل وارد مدل شده اند، کنترل شده اند تا مشخص شود که متغیر مستقل جدید چه میزان می تواند به پیش بینی متغیر وابسته پردازد. به عنوان مثال اگر شما بخواهید بدانید که اگر سن کنترل شود، خوش بینی تا چه حد می تواند به پیش بینی متغیر رضایت از زندگی پردازد، آنگاه سن را در بلوک اول وارد می کنید و خوش بینی را در بلوک دوم. به این ترتیب ابتدا سن وارد مدل می شود و سپس خوش بینی. هنگامی که همه متغیرها وارد شدند آنگاه می توان به بررسی قدرت پیش بینی مدل پرداخت و سهم هر بلوک از پیش بینی را مشخص کرد.

### رگرسیون چندگانه گام به گام

در این روش محقق تعدادی متغیر را به SPSS می دهد آنگاه خود نرم افزار آنها را وارد مدل می کند، ترتیب ورود متغیرها به مدل را خود نرم افزار، بر اساس معیارهای آماری تعیین می کند. این

روش سه نوع دارد: انتخاب پیشرو (forward)، حذف پسرو (backward) و رگرسیون گام به گام (stepwise). البته این روشها هر کدام عیوبی دارند و برخی از مولفان توصیه کرده اند که از آنها استفاده نشود. اما اگر خواستید از این روشها استفاده کنید رگرسیون گام به گام از بقیه کامل تر است.

## مثال

برای نشان دادن چگونگی استفاده از رگرسیون چندگانه، مثالی خواهیم زد که داده های آن مربوط به فایل survey.sav می شود. ما می خواهیم به بررسی عوامل موثر میزان ادراک استرس نزد افراد بپردازیم. تحقیقات پیشین می گویند هر چه فرد احساس کنترل بیشتری بر زندگی خود داشته باشند، آنگاه احساس استرس کمتری خواهند داشت. احساس کنترل فرد بر زندگی با دو متغیر سنجیده شده است. اول احساس کنترل فرد بر رویدادهای زندگی (tmast) و نیز احساس کنترل فرد حالات درونی خویش (tpcoiss). حال می خواهیم بدانیم این دو متغیر تا چه حد می توانند به پیش بینی متغیر وابسته ما بپردازند.

سعی کنید که شما هم در رایانه این مراحل را انجام دهید چرا که انجام دادن بهترین راه برای یادگیری است و صرفا با خواندن کتاب نمی توانید مهارت های خود را افزایش انجام دهید. پس شما هم همین مراحل را انجام دهید و سعی کنید خروجی ها تفسیر کنید.

## رگرسیون چندگانه استاندارد

در این مثل ما می خواهیم به این دو سوال پاسخ دهیم:

سوال ۱: دو متغیری که مربوط به کنترل هستند (tmast، tpcoiss) تا چه حد می توانند استرس ادراک شده (tpsterss) را پیش بینی کنند؟ چه مقدار از واریانس استرس ادراک شده بوسیله این دو متغیر تبیین می شود؟

سوال ۲: بهترین پیش بینی کننده متغیر استرس ادراک شده کدام است: کنترل بر رویدادهای خارجی (tmast) یا کنترل بر حالات درونی (tpcoiss).

برای پاسخگویی به این سوالات از رگرسیون استاندارد استفاده می کنیم. یعنی دو متغیر مستقل را یکجا وارد مدل می کنیم.

## شیوه انجام رگرسیون چندگانه استاندارد

۱. از منوی بالای صفحه ابتدا Analyze سپس Regression و بعد از آن Linear را انتخاب کنید.

۲. متغیر وابسته خود را (tpstress) در قسمت Dependent وارد کنید.
۳. متغیرهای مستقل خود (tomast و tpcoiss) را در قسمت Independent وارد کنید.
۴. در قسمت Method (روش) گزینه Enter (یعنی روش استاندارد انجام رگرسیون چندگانه) را انتخاب کنید.
۵. دکمه Statistics را انتخاب کنید.
  - گزینه های Estimates، fit، Model، Descriptives و Colinearity diagnostics را انتخاب کنید.
  - در قسمتی که Residuals (باقیمانده ها) نام دارد بر روی گزینه های Casewise diagnostics و Outliers outside 3 standard deviation کلیک کنید.
  - Continue را بزنید.
۶. دکمه Options را بزنید و در آن در قسمت Missing values گزینه Exclude cases pairwise را انتخاب کنید.
۷. دکمه Plots را بزنید.
  - گزینه ZRESID\* را به قسمت محور Y منتقل کنید.
  - گزینه ZPRED\* را به قسمت محور X منتقل کنید.
  - در قسمتی که عنوان آن Standardized Residual Plots نام دارد گزینه Normal probability plot را انتخاب کنید.
۸. بر روی دکمه Save کلیک کنید.
  - در قسمتی که Distances نام دارد گزینه Mahalanobis را انتخاب کنید (این گزینه دورافتاده های چندمتغیری را برای شما مشخص می کند).
  - Continue را بزنید.
۹. OK بزنید.

خروجی به دست آمده چنین خواهد بود:

### Correlations

		total perceived stress	total PCOISS	total mastery
Pearson Correlation	total perceived stress	1.000	-.581	-.612
	total PCOISS	-.581	1.000	.521
	total mastery	-.612	.521	1.000
Sig. (1-tailed)	total perceived stress	.	.000	.000
	total PCOISS	.000	.	.000
	total mastery	.000	.000	.
N	total perceived stress	433	426	433
	total PCOISS	426	430	429
	total mastery	433	429	436

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.684 <sup>a</sup>	.468	.466	4.274

a. Predictors: (Constant), total mastery, total PCOISS

b. Dependent Variable: total perceived stress

### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6806.728	2	3403.364	186.341	.000 <sup>a</sup>
	Residual	7725.756	423	18.264		
	Total	14532.48	425			

a. Predictors: (Constant), total mastery, total PCOISS

b. Dependent Variable: total perceived stress

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	50.971	1.273		40.035	.000		
	total PCOISS	-.175	.020	-.360	-8.660	.000	.729	1.372
	total mastery	-.625	.061	-.424	-10.222	.000	.729	1.372

a. Dependent Variable: total perceived stress

### Casewise Diagnostics<sup>a</sup>

Case Number	Std. Residual	total perceived stress	Predicted Value	Residual
359	-3.475	14	28.85	-14.849

a. Dependent Variable: total perceived stress

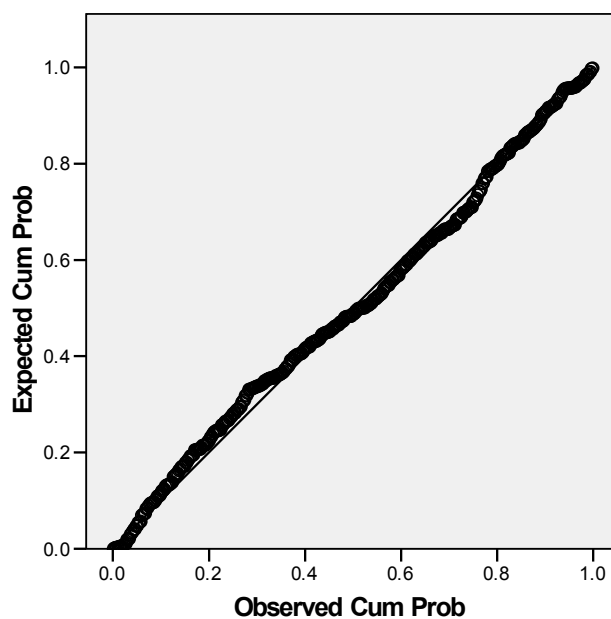
### Residuals Statistics<sup>a</sup>

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	18.03	41.31	26.74	4.001	429
Std. Predicted Value	-2.174	3.644	.002	1.000	429
Standard Error of Predicted Value	.207	.800	.341	.111	429
Adjusted Predicted Value	18.04	41.39	26.75	4.009	426
Residual	-14.849	12.612	-.002	4.268	426
Std. Residual	-3.475	2.951	.000	.999	426
Stud. Residual	-3.514	2.969	.000	1.003	426
Deleted Residual	-15.190	12.765	-.001	4.306	426
Stud. Deleted Residual	-3.562	2.997	-.001	1.006	426
Mahal. Distance	.004	13.897	1.993	2.234	429
Cook's Distance	.000	.094	.003	.008	426
Centered Leverage Value	.000	.033	.005	.005	429

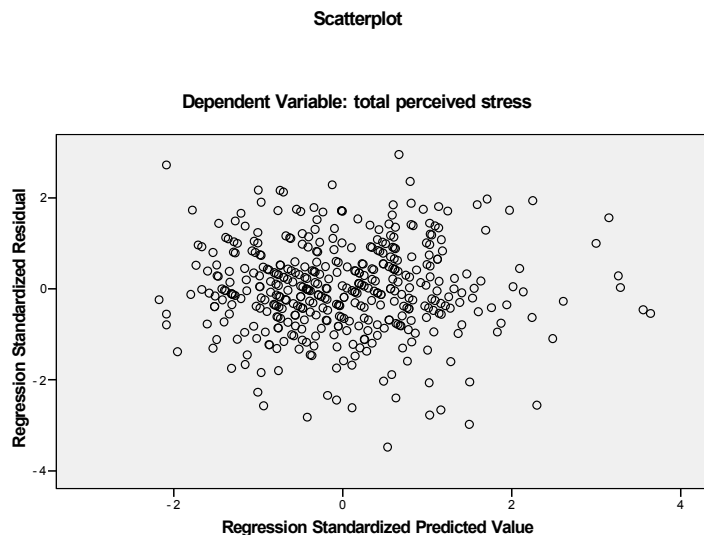
a. Dependent Variable: total perceived stress

### Normal P-P Plot of Regression Standardized Residual

Dependent Variable: total perceived stress







### تفسیر خروجی رگرسیون چندگانه استاندارد

همانطور که می بینید خروجی های زیادی بعد از انجام رگرسیون چندگانه استاندارد حاصل شده است که می تواند گیج کننده باشد. برای آنکه تفسیر ساده تر باشد ابتدا از این خروجی ها برای پاسخ دادن به سوال اولی که مطرح کردیم می پردازیم.

### گام اول: بررسی مفروضات رگرسیون

هم خطی چندگانه - جدولی که Colrelations نام دارد همبستگی های میان متغیرهای موجود در مدل ارائه شده اند. باید متغیرهای مستقل شما حداقلی از همبستگی را با متغیر وابسته داشته باشند (ترجیحا ضریب همبستگی بین آنها بالای ۰/۳ باشد). در این مثال هر دو متغیر ما همبستگی قوی با متغیر وابسته دارند (همبستگی میان tpcioss و tmast و متغیر وابسته به ترتیب ۰/۶۱- و ۰/۵۸- است). همچنین همبستگی میان متغیرهای وابسته هم نباید خیلی بالا باشد (از ۰/۷ بیشتر نباشد). اگر با چنین وضعیتی روبه رو شدید یا باید یکی از متغیرها را حذف کنید و یا آنها را با هم ترکیب کنید. در مثال ما همبستگی میان متغیرهای مستقل ۰/۵۲ است که از ۰/۷ کمتر است پس مشکلی نیست.

SPSS همچنین شاخصی را برای تشخیص هم خطی در اختیار شما می گذارد. این شاخص در جدولی است که Coefficients نام دارد. در آخرین ستون این جدول به مقادیر Tolerance توجه کنید. این مقدار از فرمول  $1-R^2$  برای هر متغیر محاسبه می شود. اگر این مقدار خیلی کم باشد معنی آن این است که همبستگی چندگانه با دیگر متغیرها بالاست و این به معنای هم خطی چندگانه است و در مثال ما این مقدار به اندازه کافی بزرگ است (۰/۷۲۹ برای هر متغیر) پس نگران هم خطی متغیرها نیستیم.

دورافتاده ها، نرمال بودن، خطی بودن، هم پراشی، استقلال باقیمانده ها- یکی از راههای بررسی این مفروضات توجه به نمودار پراکندگی باقیمانده و نمودار احتمال نرمال باقیمانده های استاندارد شده رگرسیون است که رسم کردیم. این نمودارها در انتهای خروجی ها ارائه شده اند. در نمودار احتمال نرمال شما انتظار دارید که نقاط به خوبی روی خط نرمال بودن که به صورت قطری رسم شده است توزیع شده باشند که نمودار ما تقریباً چنین است. در نمودار پراکندگی باقیمانده های استاندارد شده انتظار داریم که توزیع داده ها به صورتی مستطیلی و کاملاً پراکنده باشد و بیشتر آنها در مرکز نمودار حول خط صفر باشند. اگر چنین توزیعی را مشاهده نکنیم آنگاه باید در استقلال باقیمانده- ها شک کنیم.

از نمودار پراکندگی می توانیم دورافتاده ها را هم بررسی کنیم. برخی مولفان دورافتاده ها را داده هایی دانسته اند که مقدار استاندارد شده آنها بیش از  $3/3$  یا کمتر  $-3/3$  است. در نمونه های بزرگ وجود داده های دورافتاده چندان نامعمول نیست. اگر تعداد آنها اندک بود لازم نیست کاری برای آنها انجام دهید.

دورافتاده ها را می توان با بررسی فاصله های ماحالانوبیس (Mahalanobis distances) هم بررسی کرد. این مقادیر را SPSS محاسبه کرده است. این مقادیر در خروجی مشاهده نمی شوند بلکه در فایل داده، با نام متغیری که نام Mah\_1 نام دارد، ذخیره شده اند. برای اینکه دریابید کدام داده های شما دورافتاده اند باید مقدار بحرانی کی دو را محاسبه کنید که در کتابهای آماری به ازای تعداد متغیرهای مستقل ارائه شده اند. هنگامی که تعداد متغیرهای مستقل ۲ است آنگاه مقدار کی دو بحرانی ۱۳/۸۲ خواهد بود. پس اگر مقدار متغیر Mah\_1 کیسی بیشتر از ۱۳/۸۲ بود آن داده دورافتاده است.

## گام دوم: ارزیابی مدل

به جدولی که Model Summary نام دارد نگاه کنید. مقداری که R Square نام دارد را ببینید. این مقدار به شما می گوید که چه میزان از واریانس متغیر وابسته به وسیله متغیرهای مستقل تبیین می شود. در مثال ما این مقدار ۰/۴۶۸ است. اگر به صورت درصد بیان کنیم آنگاه می توانیم بگوییم که ۴۶/۸ درصد از واریانس متغیر وابسته به وسیله دو متغیر مستقل موجود در مدل تبیین می شود. این مقدار قابل توجهی است.

ملاحظه می کنید که مقدار ضریب تعیین اصلاح شده (Adjusted R Square) هم در جدول ارائه شده است. هنگامی که حجم نمونه کوچک است آنگاه ضریب تعیین بیش از مقدار واقعی برآورد می شود و به همین خاطر در چنین موقعیتی از ضریب تعیین اصلاح شده استفاده می شود. اگر حجم نمونه شما کوچک است ضریب تعیین اصلاح شده را ببینید در غیر این صورت ضریب تعیین را ملاحظه کنید.

برای بررسی معنی دار بودن مدل باید جدولی را که ANOVA نام دارد ببینید. در این جدول فرضیه صفر مبنی بر اینکه مقدار R چندگانه در جامعه برابر صفر است، مورد آزمون قرار می گیرد. در مثال ما مدل معنی دار است ( $\text{Sig}=.000$  که به آن معناست که  $p<.0005$ ).

### گام سوم: ارزیابی هر کدام از متغیرهای مستقل

در این مرحله می خواهیم بدانیم کدام یک از متغیرهای مستقل به پیش بینی متغیر وابسته کمک می کنند. این مساله را می توان در جدولی که Coefficients نام دارد بررسی کرد. به ستونی نگاه کنید که Beta نام دارد. برای مقایسه متغیرها با هم باید قسمت Standardized Coefficients را ببینید و نه Unstandardized Coefficients را. این استاندارد کردن به آن معنی است که ضرایب هر کدام از متغیرها چنان تغییر یافته اند که قابل مقایسه با هم شده اند. اگر بخواهیم معادله رگرسیون را بنویسیم آنگاه باید از قسمت Unstandardized Coefficients ستونی را نگاه کنیم که عنوان آن B است.

حال که می خواهیم ببینیم کدام متغیر در پیش بینی متغیر وابسته مهمتر است ستونی Beta را می بینیم و به دنبال بیشترین مقدار می گردیم. متغیری را که بیشتری مقدار Beta را دارد (بدون توجه به علامت آن) مهمترین متغیر در پیش بینی (تبیین) متغیر وابسته است. در مثال بیشترین بتا از آن متغیر tmast و برابر  $0/424-$  است. بتای متغیر tpcoiss برابر  $0/36-$  است که نسبتاً کمتر است.

برای هر کدام از متغیرها به ستون Sig نگاه کنید. این ستون به شما می گوید که آیا متغیر مورد نظر حضورش کمک معنی داری به تبیین متغیر وابسته می کند یا خیر. اگر مقدار Sig اگر  $0/05$  کمتر باشد آنگاه حضورش در سطح اطمینان ۹۵٪ برای تبیین متغیر وابسته معنی دار است. در مثال حضور هر دو متغیر مستقل در مدل معنی دار است.

نتایجی که در بالا ارائه پاسخهای سؤالی را که مطرح کردیم برایمان فراهم کرد. مدل ما که شامل دو متغیر مستقل (tmast و tpcoiss) است  $46/8$  درصد از واریانس متغیر وابسته (tpstress) را تبیین می کند (سوال ۱). از بین دو متغیر وابسته متغیر tmast بیشترین تاثیر را در پیش بینی متغیر وابسته دارد ( $\text{beta}=-.42$ ) البته حضور متغیر tpcoiss هم در مدل معنی دار است.

### رگرسیون سلسله مراتبی

گفتیم که هدف از انتخاب مدل سلسله مراتبی برای انجام رگرسیون آن است که بتوانیم متغیرهای مستقل را به ترتیب دلخواه وارد مدل کنیم. یعنی در هر مرحله متغیرهایی را که ما مشخص کرده ایم وارد مدل شوند.

برای اینکار در پنجره رگرسیون بلوک‌هایی در نظر گرفته شده است. شما باید متغیرهای مورد نظر خود را در قسمت Independents وارد کنید و دکمه Next را بزنید تا بلوک بعدی ظاهر شود. در بلوک جدید متغیرهایی را که می‌خواهید در مرحله بعد وارد مدل شوند، وارد می‌کنید. نتیجه در خروجی شامل چند مدل خواهد بود که هر مرحله با ورود یا خروج متغیرهایی که شما وارد کرده‌اید تغییر می‌کند.

### رگرسیون گام به گام

هنگامی که شما هیچ ایده نظری برای ترتیب ورود متغیرها به مدل ندارید می‌توانید این کار را به SPSS واگذارید. هنگامی که شما متغیرهای مستقل را در قسمت Independents وارد کردید و متغیر وابسته را هم در قسمت Dependent وارد کردید و در قسمت Method گزینه Forward را انتخاب کردید آنگاه SPSS یک به یک و به ترتیب اهمیت متغیرهای مستقل را وارد مدل خواهد. شرط ورود متغیرها به صورت پیش فرض در SPSS آن است که مقدار معنی داری آنها (Sig) کوچکتر یا مساوی ۰/۰۵ باشد (شما می‌توانید این مقدار را تغییر دهید).

اگر روش Backward را انتخاب کنید آنگاه ابتدا تمام متغیرها وارد مدل می‌شوند و سپس یک به یک اگر مقدار معنی داری آنها بیشتر از ۰/۱۰ بود از مدل خارج می‌شوند (این مقدار را هم شما می‌توانید تغییر دهید).

اگر در قسمت روش Stepwise را انتخاب کنید آنگاه هم برای ورود و هم برای خروج متغیرها معیار وجود خواهد داشت. معیار ورود متغیرها آن است که معنی داری آنها کوچکتر یا مساوی ۰/۰۵ باشند و شرط خروج متغیرهای از مدل آن است که مقدار معنی داری آنها بیشتر از ۰/۱۰ باشد.

### مثال

از آنجایی که در عمل روش گام به گام بیشتر از روشهای پیشرو و پسرو مورد استفاده قرار می‌گیرد، پس در این مثال از روش Stepwise استفاده خواهیم کرد. در این مثال متغیر وابسته ما همان tpstress است و متغیرهای وابسته ما عبارتند از tmarlow (مطلوبیت اجتماعی)، age (سن)، tpcoiss و tmast.

به روش مثال قبل متغیرهای مستقل و وابسته را وارد می‌کنیم تنها در قسمت Method به جای روش Enter روش Stepwise را انتخاب می‌کنیم. خروجی‌های حاصل چنین خواهد بود:

**Model Summary<sup>c</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.612 <sup>a</sup>	.374	.373	4.632
2	.684 <sup>b</sup>	.468	.466	4.274

- a. Predictors: (Constant), total mastery  
b. Predictors: (Constant), total mastery, total PCOISS  
c. Dependent Variable: total perceived stress

**ANOVA<sup>c</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5436.925	1	5436.925	253.449	.000 <sup>a</sup>
	Residual	9095.559	424	21.452		
	Total	14532.48	425			
2	Regression	6806.728	2	3403.364	186.341	.000 <sup>b</sup>
	Residual	7725.756	423	18.264		
	Total	14532.48	425			

- a. Predictors: (Constant), total mastery  
b. Predictors: (Constant), total mastery, total PCOISS  
c. Dependent Variable: total perceived stress

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	46.337	1.252		37.010	.000		
	total mastery	-.901	.057	-.612	-15.920	.000	1.000	1.000
2	(Constant)	50.971	1.273		40.035	.000		
	total mastery	-.625	.061	-.424	-10.222	.000	.729	1.372
	total PCOISS	-.175	.020	-.360	-8.660	.000	.729	1.372

- a. Dependent Variable: total perceived stress

### تفسیر خروجی های رگرسیون چندگانه به روش گام به گام

در جدولی که Model Summary نام دارد شما میزان قدرت هر مدل برای تبیین متغیر وابسته را ملاحظه می کنید. دو مدل وجود دارد. مدل اول دارای ضریب تعیین ۰/۳۷۴ و مدل دوم دارای ضریب تعیین ۰/۴۶۸ است. در جدول ANOVA می توانید معنی دار بودن مدلها را بررسی کنید.

ملاحظه می کنید که هر دو مدل معنی دار هستند. در روش Stepwise تمام مدل هایی که SPSS به شما ارائه می دهد معنی دار هستند.

برای بررسی میزان تاثیر گذاری هر کدام از متغیرهای وابسته، معنی دار بودن متغیرهای مستقل در مدل و نیز نوشتن معادله رگرسیون برای مدل باید جدول Coefficients را ببینید. ملاحظه می کنید که در مدل اول یک متغیر مستقل (tmast) حضور دارد و در مدل دوم دو متغیر مستقل (tmast و tpcoiss) حضور دارند. از آنجایی که مدل دوم ضریب تعیین بیشتری دارد پس مدل مورد علاقه ما مدل دوم است. در این مدل هم هر دو متغیر حضورشان معنی دار است اما تاثیر متغیر tmast بیشتر از تاثیر متغیر tpcoiss است (بتای اولی ۰/۴۲۴- و دومی ۰/۳۶۰- است).

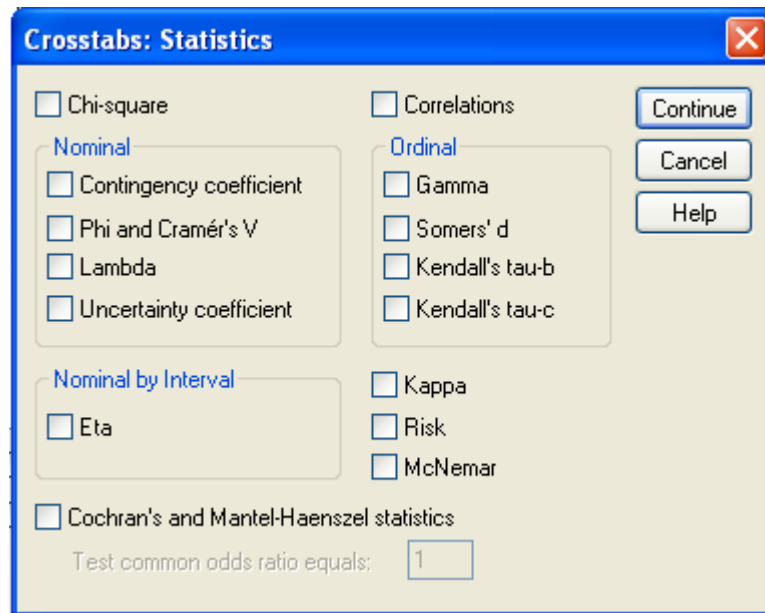
## آمار ناپارامتری

در فصلهای قبل ما تکنیکهای گوناگونی را بررسی کردیم که در دو دسته کلی جای می گرفتند: تکنیکهایی که رابطه میان متغیرها را بررسی می کنند و تکنیکهایی که به مقایسه بین گروهی می پردازند. گفتیم که تمام این تکنیک ها فرض می کنند که توزیع متغیر وابسته نرمال است. در بسیاری از اوقات این فرض در داده ها وجود ندارند. یعنی متغیر وابسته شما نرمال توزیع نشده است. گفتیم که در این مواقع شما دو راه می توانید در پیش بگیرید. اول اینکه توزیع داده های خود را «تبدیل» کنید به توزیعی نرمال. شیوه انجام این کار در فصل هشتم توضیح داده شد. اما تبدیل توزیع داده ها هم مشکلات خود را دارد و همیشه به آسانی امکان پذیر نیست. راه دیگر آن است که برای آزمون فرضیه های خود به سراغ تکنیکهای آماری برویم که نرمال بودن توزیع متغیر وابسته در مفروضات آنها نیست. این تکنیکهای آماری را آزمونهای ناپارامتری (Non-parametric tests) نامیده اند. این آزمون ها را آزمونهای بدون مفروض (assumption-free) و یا آزمونهای بدون توزیع (distribution-free) هم نامیده اند. برخی گفته اند توان آزمونهای ناپارامتری کمتر از آزمونهای پارامتری است اما می دانیم که همواره چنین نیست. این آزمونها به جای اینکه با خود نمرات متغیرها کار کنند، رتبه آنها را در نظر می گیرند. به همین خاطر می توانیم هنگامی که متغیر وابسته ما ترتیبی است هم از این تکنیکها استفاده کنیم. پس ما در دو موقعیت به سراغ آزمونهای ناپارامتری می رویم: اول زمانی که توزیع متغیر وابسته ما نرمال نباشد (می دانید قبل از اینکه تحلیل خود را شروع کنید باید توزیع متغیر وابسته را بررسی کنید؛ این کار در منوی Explore انجام می شود و درباره آن قبلاً بحث کرده ایم) دوم زمانی که متغیر وابسته ما ترتیبی باشد.

در SPSS در دو منو می توانید آزمونهای ناپارامتری را بیابید:

- در منوی Analyze در گزینه Nonparametric Tests.
- در منوی Analyze سپس Descriptive Statistics و سپس Crosstabs.

در این فصل ما به آزمونهای موجود در منوی Crosstabs خواهیم پرداخت. در پنجره مربوط به Crosstabs اگر دکمه Statistics را بزنید پنجره جدیدی باز می شود که تعدادی آزمون در آنجا وجود دارد.



همانطور که در شکل بالا ملاحظه می کنید در کادر بالا آزمون ها پدر چند دسته کلی جال داده شده اند. ابتدا کی دو (Chi-square) و در زیر آن آماره هایی وجود دارند که برای بررسی رابطه میان دو متغیر اسمی به کار می روند. این آماره ها عبارتند از ضریب توافق (Contingency coefficient)، فی و وی کرامر (Phi and Cramer's V)، لاندا (Lambda) و ضریب عدم اطمینان (Uncertainty coefficient).

در سمت راست گزینه همبستگی ها (Correlations) وجود دارد و در زیر آماره هایی وجود دارند که برای بررسی رابطه میان متغیر های ترتیبی به کار می روند. این آماره ها عبارتند از گاما (Gamma)، دی سامرز (Somers' d)، تاو بی کندال (Kendall's tau-b) و تاو سی کندال (Kendall's tau-c).

در پایین کادر هم انا (Eta) قرار دارد که برای بررسی رابطه میان متغیرهای فاصله ای و رتبه ای به کار می رود. در کنار آنها کاپا (kappa)، ضریب خطر (Risk) و مک نمار (McNemar) قرار دارند. در آخر هم آماره های کوکران و مانتل-هنتزل (Cochran and Mantel-Haenszel) را مشاهده می کنید. حال به توضیح هر کدام از آماره هایی که در بالا نام برده شد می پردازیم.

کی دو و جدول توافقی با متغیرهای اسمی

برای جدولهایی که دو سطر و دو ستون دارند اگر کی دو را انتخاب کنید آنگاه در خروجی کی دو پیرسون (Pearson chi-square)، کی دو نسبت درستنمایی (likelihood-ratio chi-square)، آزمون دقیق فیشر (Fisher's exact test) و تصحیح یتس برای کی دو (Yates' corrected chi-square) که با نام ضریب پیوستگی (continuity correction) ارائه می شود را خواهید داشت. آزمون دقیق فیشر زمانی بررسی می شود که برای یکی از خانه های جدول توافقی فراوانی مورد انتظار کمتر از



۵ باشد. تصحیح یتس برای کی دو در تمام جدولهای دو در دو محاسبه می شود. اگر جدول شما بیش از دو سطر و ستون داشته باشد آنگاه اگر کی دو را انتخاب کنید در خروجی کی دو پیرسون و کی دو نسبت درستنمایی را خواهید داشت. اگر هر دو متغیر کمی باشند، کی دو همان همبستگی خطی-خطی (linear-by-linear association) خواهد شد. اگر هر دو متغیر شما مقوله ای (اسمی) هستند نباید به این آماره توجهی کنید. لازم به ذکر است که برای داده های تکراری نباید از آزمون کی دو استفاده کنیم. (منظور از داده های تکراری آن است که شما یک متغیر را در دو زمان در یک گروه اندازه گیری کرده اید به عنوان مثال قبولی یا عدم قبولی یک گروه ۳۰ نفره از دانشجویان در یک آزمون، قبل و بعد از یک دوره آموزشی). همچنین فراوانی های مورد انتظار در هر خانه نباید از ۵ کمتر باشد. هر چند توصیه شده است که اگر حجم نمونه بزرگ باشد ایرادی ندارد که تا فراوانی مورد انتظار در ۲۰ درصد خانه کمتر از باشد. اما در هر صورت وجود چنین خانه هایی در جدول از توان آزمون می کاهد. در جدول های بزرگ (جدول هایی که تعداد سطر و ستون آن ها بالاست) فراوانی های مورد انتظار در خانه های جدول نباید از ۱ کمتر باشد.

درباره نسبت درستنمایی (Likelihood ratio) باید گفت که این آماره بدیلی برای کی دو به شمار می رود و مبتنی بر نظریه بیشترین درستنمایی است. ایده کلی این نظریه آن است که داده هایی جمع آوری می کنید و مدلی می سازید که در آن احتمال به دست آمدن داده های مشاهده شده بیشینه است. سپس در فرضیه صفر خود مدل را با احتمال به دست آمدن آن داده ها مقایسه می کنید. بنابراین آماره این آزمون مبتنی است بر مقایسه میان فراوانی های مشاهده شده و فراوانی هایی که توسط مدل پیش بینی می شوند. این آماره هم مانند آماره کی دو پیرسون دارای توزیع کی دو است و دارای همان درجه آزادی است. هنگامی که حجم نمونه بزرگ است این آماره تقریباً با کی دو پیرسون برابر می شود اما برای نمونه های کوچک ترجیح داده می شود که از این آماره به جای کی دو پیرسون استفاده شود.

حال با استفاده از فایل survey.sav به بررسی یک مثال می پردازیم.

مثال

فرضیه پژوهشی ما آن است که میان سیگاری بودن و جنسیت رابطه وجود دارد. برای بررسی این فرضیه، با توجه به اینکه هر دو متغیر اسمی هستند ناچاریم که به سراغ کی دو و آماره های مربوط به آن برویم.

شیوه محاسبه کی دو برای جدول توافقی

- در منوی Analyze گزینه Descriptive Statistics سپس Crosstab را انتخاب می کنیم.

- متغیر sex (جنسیت) را در قسمت column(s) (ستون) و متغیر smoke (سیگاری بودن) را در قسمت row(s) (سطر) وارد می کنیم. با توجه به اینکه هر دو متغیر ما دو مقوله ای هستند (مرد/زن و سیگاربودن/سیگاری نبودن) انتظار داریم که جدول توافقی ما دو در دو باشد.
  - اگر دکه Cells را فشار دهید و در قسمت Counts هر دو گزینه Observed (فراوانی مشاهده شده) و expected (فراوانی مورد انتظار) را انتخاب کنید آنگاه در خروجی فراوانی های مورد انتظار را هم خواهیم دید.
- خروجی چنین خواهد بود:

smoker \* sex Crosstabulation

			sex		Total
			MALES	FEMALES	
smoker	YES	Count	33	52	85
		Expected Count	35.9	49.1	85.0
	NO	Count	151	200	351
		Expected Count	148.1	202.9	351.0
Total		Count	184	252	436
		Expected Count	184.0	252.0	436.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.494 <sup>b</sup>	1	.482	.541	.282
Continuity Correction <sup>a</sup>	.337	1	.562		
Likelihood Ratio	.497	1	.481		
Fisher's Exact Test					
Linear-by-Linear Association	.493	1	.483		
N of Valid Cases	436				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 35.87.

#### تفسیر خروجی کی دو

- جدول توافقی مورد انتظار ما جدول اولی است که در خروجی ارائه شده است. این جدول دو در دو است و چهار خانه دارد. فراوانی های مشاهده شده و فراوانی های مورد انتظار را می توانید در این جدول ببینید. فراوانی های مورد انتظار برای محاسبه

مقدار کی دو مورد استفاده قرار می گیرند. برای بررسی چگونگی محاسبه آماره کی دو می توانید به کتابهای آمار مقدماتی مراجعه کنید.

- جدول دوم آزمون های مربوط به کی دو را در خود جای داده است. همچنین درجه آزادی مورد نظر و مقدار معنی داری هم در جدول آمده است. اولین آماره همان کی دو پیرسون است. ملاحظه می کنید که مقدار آن در اینجا  $0/494$  و مقدار معنی داری آن هم  $0/482$  است. این مقدار معنی داری از  $0/05$  بیشتر است یعنی میان جنیست و سیگاری بودن رابطه مشاهده نمی شود. اما چون جدول ما دو در دو است باید به جای کی دو پیرسون به ضریب پیوستگی (Continuity Coefficient) یا همان تصحیح یتس برای کی دو توجه کنیم. ملاحظه می کنید که ضریب پیوستگی  $0/337$  و مقدار معنی داری آن  $0/562$  است. یعنی تصحیح یتس برای کی دو هم نشان می دهد که رابطه ای میان جنسیت و سیگاری بودن وجود ندارد. آماره بعدی نسبت درستیابی است (Likelihood ratio) است که مقدار معنی آن هم بیش از  $0/05$  است. نیازی نداریم که به آزمون دقیق فیشر (Fisher's exact test) توجه کنیم چون فراوانی های مورد انتظار ما در تمام خانه ها بیش از ۵ است. همچنین همبستگی خطی-خطی را هم نمی خوانیم چون متغیرهای ما کمی نیستند و هر دو کیفی هستند.

اما سوالی وجود دارد که کی دو نمی تواند به آن پاسخ دهد و آن عبارت است از اینکه ما می خواهیم بدانیم که شدت رابطه میان متغیرهای اسمی در جدول توافقی با هم چقدر است. کی دو تنها به ما می گوید که آیا رابطه میان دو متغیر اسمی در جدول توافقی قابل تعمیم به جامعه است یا خیر. نی دانیم که کی دو هم مانند دیگر آماره ها به شدت به حجم نمونه وابسته است. یعنی اگر حجم نمونه بالا باشد مقدار معنی داری کی دو به احتمال زیاد کمتر از  $0/05$  خواهد شد. حال پس آنچه برای ما مهم است این است که بدانیم قدرت رابطه میان دو متغیر ما چه میزان است تا بتوانیم قضاوت درست تری داشته باشیم. آماره های ضریب توافق (Contingency coefficient)، فی و وی کرامر (Phi and Cramer's V)، لاندای (Lambda) و ضریب عدم اطمینان (Uncertainty coefficient) به همین منظور تعریف شده اند. در زیر به توضیح کاربرد و معنای هر کدام از این آماره ها می پردازیم.

#### ضریب توافق (Contingency coefficient)

ضریب توافق معیاری است برای بررسی همبستگی میان دو متغیر اسمی و با استفاده از کی دو محاسبه می شود. مقدار این آماره بین صفر و ۱ است اما هیچ وقت یک نمی شود. صفر بودن آن به این

معناست که هیچ همبستگی میان سطر و ستون در جدول توافقی وجود ندارد و اگر مقدار آن نزدیک به یک باشد به معنای وجود همبستگی قوی میان متغیرهای سطر و ستون است. مقدار این آماره به تعداد سطر و ستون جدول توافقی بستگی دارد. لازم به ذکر است که وی کرامر بر این آماره ترجیح دارد.

فی و وی کرامر (Phi and Cramer's V)

فی هم به بررسی رابطه میان دو متغیر اسمی پردازد و برای جداول دو در دو مورد استفاده قرار می گیرد. فی حاصل تقسیم مقدار کی دو بر حجم نمونه و جذر گرفتن از آن است. V کرامر هم با استفاده از کی دو محاسبه می شود. در جداولی از وی کرامر استفاده می کنیم که یکی از متغیرهای دو در دو و دیگری بیش از دو مقوله داشته باشد.

لاندا (Lambda)

به آماره لاندا گودمن و کروسکال هم می گویند و سنجه ای است برای بررسی رابطه میان دو متغیر اسمی. اساس لاندا بر نسبت کاهش در خطای پیش بینی متغیر وابسته با دانستن متغیر مستقل است. اگر لاندا یک باشد به این معنی است که متغیر مستقل به طور کامل می تواند متغیر وابسته را پیش بینی کند. اگر مقدار آن صفر باشد به آن معنی است که متغیر مستقل هیچ کمکی به پیش بینی متغیر وابسته نمی کند.

ضریب عدم اطمینان (Uncertainty coefficient)

این آماره نشان دهنده کاهش نسبی خطای پیش بینی متغیر وابسته با دانستن متغیر مستقل است. بع عنوان مثال اگر مقدار ضریب عدم اطمینان برابر  $0/83$  باشد به آن معنی است که با دانستن متغیر مستقل به اندازه  $83$  درصد از خطای پیش بینی متغیر وابسته کاسته می شود.

متغیرهای ترتیبی

در جدولهایی که هر دو متغیر در آنها ترتیبی هستند اگر گزینه Correlations را انتخاب کنید آنگاه ضریب همبستگی ترتیبی که رو اسپیرمن (Spearman rho) نام دارد محاسبه خواهد شد. این ضریب همبستگی میان رتبه ها را محاسبه می کند. اگر داده های شما کمی باشند ضریب همبستگی مناسب I پیرسون خواهد بود.

اگر بخواهید شدت همبستگی میان دو متغیر موجود در سطر و ستون را انتخاب کنید می توانید از آماره های گاما، تاو بی کندال و تاو سی کندال استفاده کنید. اگر بخواهید با استفاده از متغیر موجود در سطر به پیش بینی متغیر موجود در ستون پردازید باید از d سامرز استفاده کنید.

گاما (Gamma)

آماره ای است مقارن که همبستگی متغیرهای ترتیبی را بررسی می کند. مقدار گاما بین  $1-$  و  $1$  تغییر می کند. اگر مقدار گاما (بدون توجه به علامت آن) نزدیک به  $1$  باشد به معنای وجود

همبستگی قوی میان دو متغیر است. مقادیر نزدیک به صفر نشان دهنده آن است رابطه میان دو متغیر بسیار ضعیف است.

دی سامرز

دی سامرز هم برای بررسی همبستگی میان دو متغیر رتبه ای به کار می رود و مقدار آن هم بین ۱- و ۱ تغییر می کند. برای این آماره هم مقادیر نزدیک به یک نشان دهنده وجود رابطه قوی و مقادیر نزدیک به صفر نشان دهنده وجود رابطه بسیار ضعیف است. دی سامرز بسط گاما است با این تفاوت که جفت ناپیوسته در متغیر وابسته را هم در محاسبه دخالت می دهد (برای جزئیات بیشتر درباره محاسبه این آماره ها می توانید به کتابهای آماری مراجعه کنید).

تاو بی کندال (Kendall's tau-b)

آماره هم مشابه گاما است. برای بررسی رابطه میان متغیرهای ترتیبی به کار می رود. مقدار آن بین ۱- و ۱ است. اما خود مقادیر ۱- و ۱ تنها در جدولهای توافقی مربعی (سطر و ستون برابر باشند) به دست می آید.

تاو سی کندال (Kendall's tau-c)

این آماره هم مشابه تاو بی کندال است با این تفاوت که پیوند ها در نظر نمی گیرد. مقدار این آماره هم بین ۱- و ۱ است و خود مقادیر ۱- و ۱ تنها در جدولهای توافقی مربعی به دست می آیند توصیه شده است هنگامی که حجم داده ها نسبتاً کم و رتبه های یکسان فراوان است به جای همبستگی اسپیرمن از این آماره استفاده شود.

حال به بررسی آنچه گفته شد با یک مثال می پردازیم.