آشنایی با نرمافزار **SPSS** برای تحلیل آمارگیریها



تدوین و تألیف: شهاب جولانی

۱. آشنایی با بستهی نرمافزاری SPSS

بستهی نرمافزاری SPSS که مخفف عبارت Statistical Package for the Social Science است، یک بستهی نرم افزاری برای تحلیل دادهها میباشد و به وفور در علوم اجتماعی و رفتاری مورد استفاده قرار می گیرد. نسخههای متفاوتی از بستهی نرمافزاری SPSS برای سیستم عاملهای ویندوز، یونیکس و مکینتاش وجود دارد. در این دست نویس نسخهی مربوط به سیستم عامل ویندوز شرح داده می شود و تحلیلها بر اساس آخرین نسخه موجود (شماره ۱۳) ارائه می شوند.

بستهی نرمافزاری SPSS امکانات یک صفحهی گسترده را برای ورود و مشاهدهی فایل دادهها –Data Editor بستهی نرمافزاری Output View نشان داده – فراهم میکند. خروجی مربوط به روشهای آماری در پنجرهای جداگانه به نام Output View نشان داده میشود. این پنجره شامل جدولها و نمودارهایی است که می وان آنها را در سایر نرم افزارها نیز مورد استفاده قرار داد.

پس از اجرای بستهی نرمافزاری SPSS پنجرهی ویرایش گر دادهها بر روی صفحهی نمایش ظاهر می شود. علاوه بر آن کلیهی منوهای اصلی نرمافزار، که مجموعهی قابلیتهای کلی آن را نمایش می دهد، بر روی یک نوار ابزار نشان داده می شود (شکل ۱.۱ را ببینید).

> File Edit View Data Transform Analyze Graphs Utilities Window Help شکل ۱.۱- نوار ابزار در بستهی نرمافزاری SPSS

نوار ابزار شامل ۱۰ منوی اصلی است که هر یک بخشی از فعالیتهای نرمافزار را بر عهده دارند. این منوها عبارتاند از:

- منوی فایل (File): ساختن فایل جدید دادهها، فراخوانی و ذخیرهسازی انواع فایلها در این نرمافزار و سایر نرمافزارها (مثلاً SAS یا EXCEL) و چاپ گزارشها و نمودارها
 - ۲. منوی ویرایش (Edit): انجام انواع ویرایشها مانند کپی، حذف یا الصاق و جستجو در فایل دادهها
- ۳. منوی نمایش (View): تغییرات در صفحهی نمایش دادهها شامل تغییر فونت، سفارشی کردن نوار ابزار و غیره
- ۴. منوی دادهها (Data): تغییر مشخصههای متغیرها، اضافه کردن متغیرها یا نمونه های جدید، مرتب کردن، جدا کردن، ادغام کردن، خرد کردن، وزن دادن و انتخاب بخشی از دادهها در یک فایل
 - ۵. منوی تبدیل (Transform): تبدیل، رتبه بندی و کدگذاری دادهها و جانهی مقادیر گمشده
- ۶. منوی تحلیل (Analyze): انواع تحلیلهای آماری مانند آمار توصیفی، مقایسه یمیانگینها،
 رگرسیون، تحلیل واریانس و غیره
- ۷. منوی نمودارها (Graphs): نمودارهایی مانند نمودار میلهای، دایرهای و خطی، هیستوگرام، نمودار نرمال، نمودار پراکنش و غیره
- ۸. منوی امکانات (Utilities): اطلاعاتی درباره یفایل داده، سیستم مدیریت فایل خروجی، تغییر شکل
 ظاهری پنجرهها و ویرایش نوار ابزار
 - ۹. منوی پنجره (Window): مرتب کردن، گزینش و کنترل خصوصیات پنجرههای نرمافزار

۱۰. منوی کمک رسانی (Help): راهنماییهای کلی در مورد نحوهی استفاده از نرمافزار و بخشهای مختلف آن، امکان جستجوی یک عنوان خاص، نمایش ترکیب فرمانهای SPSS و خلاصهای از تعاریف اصطلاحات و فرمانها

۱٫۱ مدیریت دادهها و فایلها

۱٫۱٫۱. وارد کردن دادهها

به طور کلی برای فعال کردن اغلب بخشهای عملیاتی، ضروری است فایل دادهای در محیط SPSS فعال باشد. به همین دلیل است که پنجرهی SPSS Data Editor به عنوان پنجرهی اصلی در اولین بار ورود به نرمافزار مشاهده می شود. این صفحه شامل دو پنجره است:

- Data View قسمتی است که برای ورود دادهها در نظر گرفته شده است. با ورود دادهها از طریق صفحه کلید، این مقادیر به ترتیب در سطرهای مختلف متغیرها قرار می گیرند. برای اغلب تحلیلهای آماری در بستهی نرمافزاری SPSS فرض شده است که سطرها نشان دهندهی شمارهی مشاهدات و ستونها متغیرها هستند.
- Variable View قسمتی است که مشخصه های متغیرها شامل نوع و ساختار در آن تعریف می شوند.
 این قسمت یک صفحه ی گسترده است که شامل ده مشخصه به شرح زیر می باشد:
- ۱. نام (Name): نام هر متغیر در این ستون مشخص می شود که حداکثر شامل ۸ کاراکتر است.
- ۲. نوع متغیر (Type): ستون دوم نوع متغیر را تعریف می کند که با توجه به ماهیت داده، می تواند شامل قالبهای مختلف دادههای عددی، تاریخ یا پول باشد. انواع مختلف قالبهای متغیرها به شرح زیر است:
- مددی (Numeric): اختصاص به ورود مقادیر عددی عادی برای متغیرها دارد و بر اساس اندازهی واقعی داده و تعداد ارقام اعشار آن تعیین می شود. هر عدد به همان صورت که وارد می شود در حافظه ذخیره می شود اما با محدودیت ۸ کاراکتر و دو رقم اعشار (گزینه های پیش فرض برای اندازهی واقعی داده و تعداد ارقام اعشار آن) به نمایش در می آید. به عنوان مثال عدد ۱.۲۳۴ به صورت ۱.۲۳ و عدد ۱۲۳۴۵۶۷۸۹ به صورت ۹۰+ E ۲۰۳ به نمایش در می آید.
- ویرگول (Camma): در این نمایش هر متغیر عددی به صورت سه رقم سه رقم از
 سمت راست با یک علامت ویرگول از سایر ارقام جدا می شود. مانند ۱۲,۳۴۵,۶۷۸
- o نقطه (Dot): در این نمایش هر متغیر عددی به صورت سه رقم سه رقم از سـمت راست با یک علامت نقطه از سایر ارقام جدا می شود. مانند ۱۲.۳۴۵.۶۷۸
- ماد علمی (Scientific Notation): هر متغیر عددی به صورت مضربی از توانهای
 ۱۰ نوشته می شود مانند عدد ۷۸ که به صورت ۸.E +۰۰۱ به نمایش در می آید.
- تاریخ (Data): این قالب برای ثبت زمان دقیق جمع آوری داده ها و همچنین ورود
 اطلاعات مربوط به تاریخ یا زمان در نظر گرفته شده است. پس از انتخاب این

- می شود که با انتخاب نوع مناسب، می توان زمان ها را در متغیر وارد کرد. ۰ دلار (Dollar): این قالب برای ثبت مقادیر قیمت ها بر حسب واحد دلار است و با
- انتخاب شکل مناسب بر اساس پهنای رقم نمایش و تعداد ارقام اعشار، ورود دادهها صورت می گیرد.
- سفارشی (Custom Currency): با انتخاب این گزینه امکان انتخاب چندین قالب
 سفارشی بر حسب تعداد کاراکتر و رقم اعشار فراهم می شود.
- رشتهای (String): این نوع متغیر مخصوص اطلاعاتی است که به صورت رشتهای
 از کاراکترها (مانند اسامی خاص) در متغیر وارد می شوند. تعداد کاراکترهای یک
 رشته به صورت پیش فرض ۸ است که می توان آن را کم یا زیاد کرد.

Variable Type		? 🗙
 Numeric Comma Dot Scientific notation Date Dollar Custom currency String 	dd.mm.yy yyddd q Q yyyy q Q yy mmm yyyy mmm yy ww WK yyyy ww WK yy dd-mmm-yyyy hh:mm	OK Cancel Help

شکل ۱.۲- پنجرهی Variable Type

- ۳. پهنا (Width): پهنا یا اندازه واقعی دادههای وارد شده توسط این مشخصه تعیین می شود. مقدار پهنای پیش فرض برای دادههای عددی ۸ کاراکتر است که می توان آن را کم یا زیاد کرد.
- ۴. رقم اعشار (Decimals): شمارهی رقمهای سمت راست اعشار برای دادههای وارد شده شرح
 داده می شود، که آن ها را می توان کم یا زیاد کرد.
- ۵. برچسب (Label): یک برچسب متناسب با نام متغیر تعریف می کند که در مقایسه با نام متغیر محدود به ۸ کاراکتر نمی باشد. این ستون عموماً برای نوشتن معنی متغیرها متناسب است که در صفحه یخروجی تحلیل دادهها از آن استفاده می شود.
- ۶. مقدار (Values): برای متغیرهای گروهی (دستهبندی شده)، یک کد عددی به ازای هر دسته نسبت داده می شود. این امکان مخصوص متغیرهایی است که از نوع عددی (Numeric) هستند. به وسیلهی کلیک کردن روی Values پنجرهی Value labels باز خواهد شد که توسط آن برچسبها برای کدهای دستهها نسبت داده می شوند. شکل ۱۰۳ جزئیات این پنجره را نشان می دهد و به عنوان مثال، برای متغیر جنسیت می توان عدد صفر را به زنان و عدد یک را به مردان نسبت داد.

Value Labels	? 🛛
Value Labels Value: Value Label: Add Change Remove	OK Cancel Help

شكل ۲.۲- پنجرهی Value Labels

- ۲. کد مقادیر گمشده (Missing): اگر کدهای دیگری مانند ۹- به غیر از کد پیشفرض ۲.
 ۲. کد مقادیر گمشده در نظر گرفته شوند (علامت «.») در بستهی نرمافزاری SPSS به عنوان مقادیر گمشده در نظر گرفته شوند توسط این ستون مشخص خواهند شد.
- ۸. ستونها (Columns): پهنای ستون متغیرها را در صفحه گستردهی Data view مشخص میکند که مقدار آن به صورت پیشفرض ۸ کاراکتر است، وقتی پهنای دادهها بیشتر از پهنای ستون متغیرها باشد، فقط قسمتی از دادهی وارد شده مشاهده میشود که با افزایش مقدار پهنای ستون تمام آن قابل رؤیت خواهد بود.
- ۹. صف بندی (Align): صف بندی ورود دادهها را مشخص میکند که میتواند به صورت
 چپچین، راستچین و یا وسطچین تعریف شود.
- ۱۰. اندازه (Measure): مقیاس اندازه گیری متغیرها را مشخص می کند که به سه صورت مقیاس پیوسته یا فاصلهای (Scale)، ترتیبی (Ordinal) و اسمی (Nominal) تقسیم بندی می شود.
 با توجه به نوع دادهها مقیاسهای متفاوتی را می توان تعریف کرد. به عنوان مثال برای دادههای عددی، مقیاس پیوسته، برای دادههای گروهی ترتیبی، مقیاس ترتیبی و برای دادههای گروهی که ترتیب می است.

اولین قدم در اجرای هر گونه تحلیل، ورود دادههای خام است. در ادامه اقدام به ورود دادهها در بستهی نرمافزاری SPSS از طریق صفحهی کلید می کنیم. مثلاً فرض کنید می خواهیم یک فایل داده با دو متغیر به صورت زیر بسازیم:

وزر
٣
٩
۲ ۱

ابتدا عدد ۲۲ را در بالاترین خانهی سمت چپ (سطر اول و ستون اول) تایپ کرده و سپس کلید Enter را فعال می کنیم. در این صورت نام پیشفرض VAR00001 در بالای ستون اول ظاهر می شود. به همین ترتیب اعداد ۲۱ و ۲۵ در سطرهای دوم و سوم وارد می شوند. با انتقال نشانگر به بالای ستون دوم، مقادیر عددی برای متغیر دوم نیز وارد می شود و این عمل تا ورود تمام اطلاعات تکرار خواهد شد. شکل۱.۴ داده های وارد شده در صفحهی گستردهی Data Editor را نشان می دهد.

🛅 Unt	titled - SP	SS Data E	ditor	
File Ed	dit View I	Data Trans	form An	alyze Grap
e	6 🖷		ي ا	M
3 : V.	AR0000:	2	61	
	VAR0000 1	VAR0000 2	var	var
1	22.00	53.00		
2	21.00	49.00		
3	25.00	61.00		
4				
5				

شکل ۱.۴- نمونه ای از ورود متغیرها در پنجرهی New Data

بستهی نرمافزاری SPSS نام و مشخصات هر متغیر را به صورت پیشفرض تعیین میکند. به عبارت دیگر ######VAR برای نام و دو رقم اعشار به عنوان مشخصات یک متغیر عددی مورد استفاده قرار میگیرد. در بسیاری از موارد، نیاز به تغییر نام و همچنین مشخصات متغیر مورد نظر وجود دارد که این عمل توسط پنجرهی Variable View امکان پذیر است.

۲٫۱٫۱. ذخیره و فراخوانی فایل دادهها

از منوی File در نوار ابزار می توان یک مجموعه ی داده را به وسیله ی گزینه های Save as یا ... Save as ذخیره نمود. فایل داده ها در این بسته ی نرم افزاری با پسوند sav.* ذخیره می شوند. ذخیره ی فایل داده ها با قالب های دیگر از جمله اکسل (xls*) یا متن (dat.*) نیز امکان پذیر است. برای فراخوانی یک فایل از پیش ذخیره شده کافی است از منوی File گزینه ی Open را انتخاب کرده و فایل مورد نظر را از آدرسی که در آن ذخیره شده است فراخوانی یک در آن ذخیره شده است فراخوانی ی که در آن ذخیره شده است فراخوانی یک فایل از پیش فراخوانی نور شده کافی است از می می مورد نظر را از آدرسی که در آن ذخیره شده است فراخوانی کرد.

برای فراخوانی فایلهایی غیر از قالب SPSS (فایلهایی با پسوندهای xls ،dat و یا txt) کافی است زیـر منـوی Open فعال شود و نوع فایل در قسمت Files of type تعریف شود. به عنوان مثال فرض کنید برای فراخـوانی دادههایی با کدهای ASCII (با پسوند dat) باید گزینهی (Data (*.dat) انتخاب شود.

۳,۱,۱. مرتب کردن یک فایل داده

یک فایل داده را میتوان بر حسب متغیرهای آن فایل مرتب کرد. برای این منظور باید گزینهی Sort از طریق منوی Data در نوار ابزار فعال شود. در صورت انتخاب این فرمان، پنجرهی آن مطابق با شکل۱۰۵ ظاهر میشود. از فهرست متغیرهای موجود (جعبهی سمت چپ شکل۱۰۵)، میتوان متغیر یا متغیرهایی را برای مرتب کردن انتخاب کرد. با قرار دادن متغیرهای مورد نظر در جعبهی :Sort by این امر تحقق مییابد. در این حالت تمام متغیرهای موجود در فایل بر اساس متغیر یا متغیرهای منتخب در جعبهی :Sor by مرتب میشوند. بخش Sort Order چگونگی مرتب کردن دادهها را نشان میدهد. مرتب کردن میتواند به صورت افزایشی (Ascending) یا کاهشی (Descending) باشد.

Sort Cases		
🚸 Graduate [graduate]	Sort by:	ОК
Starting Salary [salary] Degree Famed [degree	College [college] Ascend	Paste
Graduation Date [grad		Reset
	Sort Order	Cance
	 Ascending Descending 	Help

شکل۵۰۲- پنجرهی Sort

۲٫۱ ویرایش دادهها

غالباً ویرایش دادهها مربوط به اعمال یک سری تغییرات روی مشاهدات است. این تغییرات کمک میکننـد تـا دادهها برای انجام تحلیلهای خاص آماری مهیا شوند. برخی از این تغییرات عبارتند از:

برای اضافه کردن یک مشاهدهی جدید بین دو مشاهده در یک فایل داده، ابت دا نشانگر صفحه را در زیر سطری که قرار است مشاهده جدید وارد شود قرار میدهیم و سپس از نوار ابزار منوی Data و بعد گزینهی Insert Case را انتخاب میکنیم. یک سطر تازه برای ورود دادههای جدید اضافه میشود. برای حذف یک سری از مشاهدهها نشانگر صفحه روی سطر مورد نظر قرار داده میشود (با کلیک کردن بر روی شمارهی سطر این امر حاصل میشود). از نوار ابزار منوی Edit انتخاب شده و سپس گزینه ی انتخاب میشود.

۲٫۲٫۱. اضافه و حذف کردن یک متغیر

امکان اضافه کردن یک متغیر جدید (یک ستون) به مجموعه متغیرهای یک فایل در بستهی نرمافزاری SPSS وجود دارد. برای اضافه کردن یک متغیر جدید بین دو متغیر به ترتیب زیر عمل می شود:

 نشانگر صفحه در قسمت راست متغیری که قرار است متغیر جدید در آنجا تعریف شود قرار داده می شود.

از نوار ابزار منوی Data و سپس گزینهی Insert Variable انتخاب می شود.
 برای حذف یک متغیر فقط کافی است روی نام آن متغیر کلیک کرده و گزینهی Clear از طریق منوی Edit در نوار ابزار فعال شود.

۳,۲,۱. ایجاد متغیرهای جدید

در بسیاری از موارد محاسبهی متغیرهای جدید به منظور اجرای محاسبه های مورد نظر، لازم و ضروری هستند. ایجاد متغیرهای جدید با اعمال محاسبات و گزارههای منطقی توسط سایر متغیرها انجام می شود. به

عنوان مثال، فرض کنید فایل دادههای شما شامل نمرههای میان ترم و پایان ترم تعدادی از دانـشجوها باشـد اما شما علاقهمند به نمرهی نهایی این دانشجوها هستید. برای ساختن متغیر جدید به طریقهی زیر عمل میکنیم:

از نوار ابزار منوی Transform و سپس عبارت Compute را انتخاب کرده تا پنجرهی آن باز شود.

- ۱. نام متغیر جدید (متغیری که قرار است ایجاد شود) را در جعبهی Target Variable تایپ شده و برای ایجاد آن از توابع ریاضی یا آماری^۱ استفاده می شود. به عنوان مثال، برای محاسبهی نمره نهایی دانشجویان ابتدا باید نمره میان ترم را از لیست سمت چپ پائین انتخاب کرده و آن را در قسمت محاسبات قرار داد. سپس بر روی علامت «+» کلیک کرده و در نهایت نمره یایان ترم است، محاسبه کنیم. به این ترتیب نمره ی کل دانشجو که حاصل نمره های میان ترم و پایان ترم است، محاسبه خواهد شده و خواهد شد. (شکل۱۰۶)
- ۲. پس از کامل شدن عبارت ریاضی مورد نظر، بر روی گزینهی Ok کلیک کرده تا متغیر جدید محاسبه شود.

Compute Variable	
Target Variable: Numeric Expression: Total = Type & Label Mid + Final Image: Write Term [Mid] Image: Write Term [Mid]	
Final [Final] + ≤ > 7 8 9 . < = > = 4 5 6 . < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = = = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < < = > = 1 2 3 . < . < . < . < . < . < . <p< td=""><td>hmetic F & Noncentral CDF nversion rrent Date/Time te Arithmetic te Creation te Extraction ctions and Special Variables:</td></p<>	hmetic F & Noncentral CDF nversion rrent Date/Time te Arithmetic te Creation te Extraction ctions and Special Variables:
OK Paste Reset Cancel Help	

شكل Compute پنجرهى Compute

ایجاد یک متغیر جدید با در نظر گرفتن یک شرط اضافی، از طریق کلید ...If در پنجرهی Compute امکان پذیر است (شکل۱۰۶ را ببینید). مثلاً فرض کنید نمرهی نهایی دانشجویانی مورد نظر است که نمرهی پایان

^۱ - علاوه بر توابع جبری ساده مانند جمع، تفریق، ضرب و تقسیم بسیاری از توابع ریاضی مفید مانند قدر مطلق، جذر و توان دوم و همچنین توابع آماری نظیر مجموع، میانگین و انحراف معیار در این قسمت قابل محاسبه هستند.

ترم آنها بیشتر از ۷/۵ باشد. در این صورت ابتدا کلید ...If را فعال کرده تا پنجرهی مربوط بـه آن بـاز شـود. سپس گزینهی Include if cases satisfies condition را انتخاب میکنیم. شرط مورد نظر با عبارت

Final > 7.5

در جعبهی عبارت منطقی جای می گیرد. این شرط باعث می شود تا فقط مشاهداتی در محاسبهی متغیر جدید شرکت داده شوند که در شرط مورد نظر (نمرهی پایان ترم بیشتر از ۷/۵) صدق کنند. به عبارت دیگر مقدار نمرهی نهایی (با متغیر Total) فقط برای مشاهداتی که نمرهی پایان ترم آنها بیش از ۷/۵ است، محاسبه می شود و سایر مقادیر این متغیر که در شرط مورد نظر صدق نمی کنند، مقادیر گمشده محسوب می شوند.

در صورتی که در جریان اجرای فرمان نام متغیر هدف با یکی از نامهای قبلی یکسان باشد (خواه از متغیرهای موجود یا متغیرهای منطقی جدید) پیامی مبنی بر تایید جایگذاری دادههای جدید به جای دادههای قدیم ظاهر می شود. که تنها در صورت تایید جایگذاری، متغیر قدیم حذف می شود.

۴,۲,۱. کدگذاری متغیرها

کدگذاری متغیرها – متغیرهای عددی یا گروهی– یکی از تکنیکهای مفید است که می توان مقادیر یک یا چند متغیر را به طور همزمان یا تک تک به مقادیری کدگذاری شده تغییر داد. این کدها ممکن است معانی خاصی در مسائل تحقیقی داشته باشند. بهعنوان مثال، فرض کنید یک فایل داده شامل تعداد نفرهای خانوار، درآمد و سایر مشخصههای سرپرست خانوار باشد که بر اساس استانی که در آن زندگی می کنند جمع آوری شده است. متغیر استان با مقادیر ۱ تا ۲۸ نشان دهنده ی محل زندگی این خانوارها بر حسب شماره ی استان می باشد. برای تحلیل توزیع درآمد خانوارها بر اساس این که در شمال یا جنوب کشور زندگی می کنند نیاز به کدگذاری متغیر «استان» است. همچنین ممکن است علاقه مند به دسته بندی یک متغیر عددی – پیوسته یا باشد و ما علاقهمند به دسته بندی آن به گروههای جوان (زیر ۳۵ سال)، میان سال سن سرپرست خانوارها کهن سال (۴۶ سال و بیشتر) هستیم.

برای کدگذاری متغیرها دو حالت وجود دارد:

ک*دگذاری مقادیر در متغیر یکسان* که باعث از بین رفتن تمام مقادیر اصلی متغیر مورد نظر می شود. به عبارت دیگر مقادیر کدگذاری شده در همان متغیر جایگزین می شوند.

کدگذاری مقادیر در یک متغیر متفاوت که مقادیر کدگذاری شده در یک متغیر جدید ثبت میشود و مقادیر متغیر اصلی بدون تغییر باقی میمانند.

به منظور کدگذاری در متغیر یکسان مراحل زیر را باید انجام داد:

 از نوار ابزار منوی Transform، سپس گزینه Recode و در نهایت گزینه Into same variable را انتخاب کرده تا پنجره ی آن باز شود.

- ۲. نام متغیر مورد نظر را برای کدگذاری انتخاب کرده و آن را در قسمت Numeric variables box قرار دهید.
 - ۳. روی گزینهی Old and new values کلیک کرده تا پنجرهی مربوط به آن باز شود. (شکل ۱.۷)
- ۴. با مشخص کردن مقادیر قدیم و جدید (Old and new values) و کلیک کردن بر روی گزینهی Add مقادیر آمادهی کدگذاری می شوند.
- ۵. به وسیله کلیک کردن بر روی گزینهی Continue و سپس گزینهی Ok متغیر کدگذاری شده ایجاد می شود.

Old Value C Value:	New Value	C System-missing
C System-missing	Old> New:	
 System- or user-missing Range: 	Add	vest thru 35> 1
36 through 64	Change 65 t	hru Highest> 3
C Range: Lowest through	Remove	
C Range:		
through highest		
C All other values	Continue C	ancel Help

شکل ۸.۱- پنجرهی Recode: Old and New Values

مراحل کدگذاری در یک متغیر متفاوت مشابه حالت قبل است با این تفاوت که گزینهی Into different مراحل کدگذاری در یک متغیر متفاوت مشابه حالت قبل است با این حالت، پس از انتخاب متغیر مورد variable نظر برای کدگذاری باید نام یک متغیر جدید برای مقادیر کدگذاری شده نیز ثبت شود.

۵,۲,۱. گزینش مشاهدات

در برخی از کاربردها به جای استفاده از تمام مشاهدات تنها بخش کوچکتری از آنها مورد نیاز است. به عنوان مثال، حالتی را در نظر بگیرید که فایل دادهها شامل وزن و قد ۲۰۰ نفر است. متوسط وزن افرادی که بیشتر از ۷۰ کیلوگرم هستند مورد نیاز است. پس در این مسئله تنها قسمتی از مشاهدات مورد استفاده قرار میگیرند. برای انتخاب بخشی از مشاهدات (یعنی برخی از سطرهای یک فایل)، دستور Select Cases در نظر گرفته شده است که از طریق منوی Data در نوار ابزار فعال میشود (شکل ۱۰۸ را ببینید). درصد مشاهدات گزینش شده میتواند بین صفر تا صد درصد تغییر کند که تغییر این درصد به قید تعیین شده و محدودهی اعمال آن بر روی مشاهدات بستگی دارد. مطابق با شکل ۱۸۸ گزینههای مختلفی برای

۲. گزینهی All Cases: با انتخاب این گزینه کلیهی مشاهدات یک فایل مورد استفاده قرار می گیرند.
 این گزینه حالت پیش فرض این بخش است.

Select Cases		
 weight ige 	 Select All cases If condition is satisfied If Random sample of cases Sample Based on time or case range Range Use filter variable: Use filter variable: Filtered Deleted 	
Current Status: Do not filt	er cases	
	OK Paste Reset Cancel Help	

شکل ۸.۱- پنجرهی Select Cases

۲. گزینه Satisfied از مشاهدات با در نظر گرفتن این گزینه انتخاب بخشی از مشاهدات بر اساس یک شرط خاص انجام می شود. در صورت انتخاب گزینه فعال شده ی ... If، پنجره ی آن به فرم شکل۱۰۹ باز می شود. نحوه ی عمل این جعبه تقریباً مشابه دستور Compute (شکل۱۰۶) است. در صورت برقراری شرط خاصی بر روی متغیرهای فایل، مشاهداتی که این شرایط را داشته باشند، انتخاب می شوند. مثلاً برای انتخاب افرادی که وزن بیشتر از ۷۰ کیلوگرم دارند، شرط 70 مرط دارد کرد.

Select Cases: If		
Image: weightImage: Image: WeightImage: Image: Image	weight > 70	< >
	+ < > 7 8 9 - <= >= 4 5 6 × = ~= 1 2 3 / & 1 0 . × ~ () Delete Functions: ▲ ABS(numexpr) ANY(test,value,value,) ARSIN(numexpr) CDFNORM(zvalue) CDF.BERNOULLI(q,p)	
	Continue Cancel Help	

شکل۵۰۹- پنجرهی Select Case: If

۳. گزینهی Random Sample of Cases؛ این گزینه امکان انتخاب یک نمونهی تصادفی ساده از روی تمام مشاهدات را فراهم میسازد. در صورت انتخاب این گزینه کلید ...Sample برجسته شده کـه در صورت فعال کردن، پنجره آن مطابق با شکل ۱.۱۰ باز می شود.

Select Cases:	Random Sample	
Sample Size C Approximately Exactly 30	35 % of all cases cases from the first 200	cases
	Continue Cancel	Help

شكل ۱.۱۰- پنجرهى Select Case: Random Sample

در بخش Sample Size اطلاعات مربوط به حجم نمونه قرار دارد. گزینه ی Sample Size در امکان انتخاب درصد معینی از کل مشاهدات را فراهم می کند. این درصد می تواند بین ۱ تا ۹۹ در تغییر باشد. در صورت انتخاب این گزینه، درصد تعیین شده ای از مشاهدات به طور تصادفی انتخاب می شوند. گزینه ی Exactly یک نمونه ی تصادفی n تایی از جامعه ای به حجم N انتخاب می کند (یعنی انتخاب می شوند. گزینه ی کنمونه ی تصادفی n تایی از جامعه ای به حجم N انتخاب می کند (یعنی انتخاب می شوند ی تنا ۹۹ در تصادفی انتخاب می شوند. گزینه ی Exactly یک نمونه ی تعادفی n تایی از جامعه ای به حجم N انتخاب می کند (یعنی انتخاب n سطر از یک فایل که دارای N سطر است). لازم به ذکر است که این انتخاب بر اساس N سطر اول صورت می گیرد. مثلاً اگر علاقه مند به انتخاب یک نمونه ی تایی از جامعه ای مربوط وارد می کنیم.

 ۶. گزینهی Based on Time or Case Renge: انتخاب قسمتی از مشاهدات بر اساس تغییرات مقادیر در دامنهای معین توسط این گزینه انجام می شود. شکل ۱.۱۱ پنجره ی این فرمان را در صورتی که کلیدRange انتخاب شده باشد، نشان می دهد. عدد مشاهده ی انتخابی اول را در جعبه ی First Case و عدد مشاهده ی انتخابی آخر در جعبه ی Last Case وارد می شود. این گزینه سبب انتخاب مشاهدات از شماره ی مشاهده ی اول تا شماره ی مشاهده ی آخر (شـمارههای مـورد نظـر) و عـدم انتخاب سایر مشاهدات می شود.

🔲 Select Ca	ses: Rang	e	×
	First Case	Last Case	Continue
Observation:	<u> </u>	1	Cancel
			Help

شكل ۱.۱۱- پنجرهی Select Case: Range

۵. گزینه Use Filter Variable این گزینه سبب انتخاب مشاهدات بر حسب مقادیر یکی دیگر از متغیرهای همان فایل است که این متغیر باید گروهی باشد. با فرض این که کدهای دسته بندی شامل صفر است، مشاهده ی متناظر با آن به عنوان نمونه انتخاب نمی شود. اگر کدهای متغیر مورد نظر رقمی بزرگتر از صفر داشته باشند، کلیه ی مشاهدات بودن توجه به مقدار کد انتخاب می شوند.
 پس از انتخاب روش گزینش مشاهدات و تایید شرطه ای مورد نظر، در قسمت Are کروه ی بارتد با نرش این که که مای می بود.
 پس از انتخاب روش گزینش مشاهدات و تایید شرطه ای مورد نظر، در قسمت Unselected Cases Are روش از انتخاب می بود. ای می بارت در شرط مدق نمی کنند، تعیین می کنیم. گزینه های این جعبه عبارتند (شکل ۱۰۸) وضعیت مشاهداتی را که در شرط صدق نمی کنند، تعیین می کنیم. گزینه های این جعبه عبارتند از:

- Filtered در این حالت مشاهداتی که انتخاب نشدهاند با یک خط مورب بر روی شمارهی مـشاهده از سایر مشاهدات متمایز می شوند و متغیری به نـام \$-Filter ایجـاد مـی شـود کـه بـه مـشاهدات انتخابی کد یک و به سایر مشاهدات کد صفر نسبت می دهد.
- Deleted: انتخاب این گزینه منجر به حذف مشاهداتی خواهد شد که در شرط مورد نظر صدق نمی کنند.

۲٫۲ پنجرهی خروجی

پس از اجرای یک فرمان در SPSS، نتایج در پنجرهای به نام Output view نمایش داده می شود که به راحتی توسط گزینهی Print از طریق منوی File می توان نتایج را چاپ کرد.

۲. روشهای آمار توصیفی

تحلیل دادهها معمولاً با محاسبهی آمارههای خلاصه شده مانند میانگین، میانه و انحراف معیار آغاز شده و با رسم نمودارهای گرافیکی مانند هیستوگرام، نمودار میلهای و نمودار شاخه و برگ ساختار دادهها شرح داده میشوند. هدف اصلی در این مرحله توصیف خواص عمومی توزیع دادهها است. همچنین رفتار غیر عادی هر مشاهده که ممکن است باعث بروز مشکلات در تحلیلهای آتی شود، توسط این آمارهها یا نمودارها مشخص میشود. به عنوان مثال، پژوهشگران برای بررسی عوامل مؤثر در بروز حملهی قلبی نیازمند داشتن اطلاعاتی از قبیل سن، جنسیت، میزان فعالیتهای روزانه (ورزش) و سابقهی سایر بیماریها هستند. برای مطالعه ی تأثیر این متغیرها در بروز حملهی قلبی، اطلاعات به دست آمده به صورت خام نمیتوانند برای نتیجه گیری به کار برده شوند اما با کاهش حجم آنها به وسیلهی تنظیم جداول مناسب، رسم نمودارها و محاسبهی معیارهای لازم میتوان نتیجه گیری را سادهتر نمود.

روشهای مورد استفاده برای هدف فوق را آمار توصیفی مینامند. پس آمارههای توصیفی میتوانند یک تصویر کلی اما مفید از تمام دادهها ایجاد کنند. بهعنوان مثال، فرض کنید نمونهای از بیمارانی که دچار حملهی قلبی شدهاند، تهیه شده است. سوالهای زیر برخی از کمیتهای مورد علاقه دربارهی این دادهها را بیان میکند:

- چند درصد از بیمارانی که دچار حمله قلبی شدهاند، زن هستند؟
 - پراکندگی سن بیماران چگونه است؟
- در بین بیمارانی که دچار حمله قلبی شدهاند، متوسط زمان ورزش روزانه چقدر است؟

به طور کلی در مطالعات آماری، شرح دادههای جمع آوری شده برای متغیرها بستگی به ماهیت آنها دارد. بدین منظور ابتدا انواع متغیرها را توضیح داده و سپس روشهای بررسی دادهها را با توجه به نوع آنها شرح میدهیم. معمولاً دو نوع متغیر در آمار مد نظر هستند:

- متغیرهای گروهی (کیفی): متغیرهایی که نمیتوان آنها را اندازه گیری کرد و با اعداد و ارقام نشان داد اما میتوان آنها را به دستهها یا طبقات مختلف تقسیم بندی نمود. متغیرهایی نظیر جنسیت، رنگ پوست، نژاد و گروه خونی از این قبیل هستند. متغیرهای گروهی خود به دو گروه متغیرهای اسمی و ترتیبی تقسیم میشوند:
- متغیرهای اسمی آنهایی هستند که ترتیب برای دستهها یا طبقات مهم نیست مانند گروه خونی (A، B، A و O) یا جنسیت (مرد و زن)
- متغیرهای ترتیبی آنهایی هستند که ترتیب برای دستهها یا طبقات حائز اهمیت است
 مانند مقاطع تحصیلی (دیپلم، فوق دیپلم، لیسانس و ...) یا اندازهی اتومبیل (کوچک،
 متوسط و بزرگ)
 - ۲. متغیرهای عددی (کمی):

متغیرهایی مانند تعداد فرزندان یک خانواده یا تعداد حوادث در یک کارخانه در روزهای مختلف که نتیجهی شمارش هستند و یا متغیرهایی مانند سن یا میزان تولید پسته که نتیجهی اندازه گیری باشند. متغیرهای عددی نیز خود به دو گروه متغیرهای گسسته و پیوسته تقسیم میشوند:

- متغیرهای گسسته آنهایی هستند که قابل شمارش بوده و میتوانند مقادیر معینی را اختیار کنند و هیچ اندازهای بین آنها مفهوم ندارد مانند تعداد دانش آموزان یک کلاس یا تعداد درختان بارور یک استان
- متغیرهای پیوسته آنهایی هستند که قابل شمارش نبوده و بتوانند هر عددی را در یک
 فاصله یا بازه اختیار کنند مانند حجم شربتهای مولتی ویتامین یا وزن موشهای
 آزمایشگاهی

نمای کلی تقسیمبندی متغیرها را میتوان به صورت شکل ۲.۱ نشان داد.



شکل ۲.۱– تقسیم بندی انواع متغیرها

در پایان لازم به یادآوری است که یک متغیر کمی را میتوان به سادگی به یک متغیر کیفی تبدیل نمود. به عنوان مثال سن، یک متغیر کمی است که میتوان با تقسیم بندی آن به صورت « زیر ۵۰ سال»، « بین ۵۱ تا ۷۰ سال» و « ۲۱ سال و بیشتر» آن را به یک متغیر کیفی تبدیل کرد.

۱٫۳ آمار توصیفی دادههای کیفی

در مطالعات آماری بخش عمدهای از دادهها مربوط به دادههای کیفی هستند. آمار توصیفی این گونه دادهها که شامل تنظیم جداول توزیع فراوانی و رسم نمودارهای میلهای و دایرهای است، در ادامه شرح داده میشود.

۱,۱,۲. جدول توزيع فراواني

در جدول توزیع فراوانی مقادیر یک متغیر به چند دسته تقسیم بندی می شوند و سپس فراوانی مقادیری که درون هر دسته قرار می گیرند شمار ش شده تا یک جدول توزیع فراوانی تشکیل شود. سایر کمیت ها مانند درصد فراوانی، فراوانی تجمعی و درصد فراوانی تجمعی با استفاده از مقادیر فراوانی محاسبه خواهند شد. در ادامه از داده های مربوط به مطالعه ی عوامل مؤثر در حمله یقلبی، به منظور تشکیل یک جدول توزیع فراوانی استفاده می کنیم. در این مطالعه داده ها برای سن، جنسیت، فشار خون، فعالیت های ورزشی، سابقه ی حمله یقلبی، سابقه یی ماری دیابت و مدت زمان بستری شدن برای بیمارانی که دچار حمله یقلبی شده اند،

- ۱. از نوار ابزار منوی Analysis و سپس گزینهی Descriptive Statistics را انتخاب کنید.
- ۲. از زیر منوی باز شده گزینهی ...Frequencies را انتخاب کرده تا پنجرهی Frequency باز شود (شکل ۲۰۲) (شکل ۲۰۲)
- ۳. در قسمت (Variable(s) نام متغیر یا متغیرهایی را که میخواهیم توزیع فراوانی آنها تـشکیل شـود، قرار میدهیم. (به عنوان مثال متغیر جنسیت)

کنید.	کلیک	Ok	گزینهی	روى	١.
-------	------	----	--------	-----	----

🛞 Age in years [age]		Variable(s):	OK
 Blood pressure [bp] Physically active [active] 		W Gender [gender]	Paste
History of myocardial in			Reset
 History of diabetes [dia Log-length of stay [Inlo 		es -	Cancel
· · · · · · ·			Help
	6		
Display frequency tables	2		

شکل۲.۲- پنجرهی Frequency

توزیع فراوانی تولید شده به وسیلهی بستهی نرمافزاری SPSS در جدول (۲.۱) نـشان داده شـده اسـت. ایـن شکل فقط جدولهای تشکیل شده در پنجرهی Output را نشان میدهد. جدول «Statistics» نشان دهنـدهی تعداد مشاهدات معتبر و غیر معتبر (گمشده) برای متغیر جنسیت است. جدول «Gender» نیز بیـانگر جـدول توزیع فراوانی میباشد که شامل فراوانی، درصد فراوانی، درصد فراوانی معتبر و درصد فراوانی تجمعی است.

Ger	nder	
Ν	Valid	178
	Missing	0

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	79	44.4	44.4	44.4
	Female	99	55.6	55.6	100.0
	Total	178	100.0	100.0	

جدول ۲۰۱ - توزیع فراوانی برای متغیر جنسیت

۲,۱,۲ نمودارهای آماری

رهیافت دیگری برای سازماندهی دادهها وجود دارد که از طریق نمودارهای آماری به دست میآید. در مقایسه با جدولهای فراوانی، نمودارهای آماری شکل توزیع دادهها را بهتر نشان میدهند و تفسیر بـر اسـاس آنهـا راحت تر است. برای دادههای کیفی نمودارهای میلهای و نمودار دایرهای یا کلوچـهای متـداول تـرین نمودارهـا میباشد که در ادامه به چگونگی رسم آنها پرداخته میشود:

<u>نمودار میلهای (Bar Chart)</u> یک نمودار مفید برای دادههای گروهی است. در نمودار میلهای طول هر میله نشان دهندهی فراوانی یا درصـد فراوانی برای دستههای مختلف یک متغیر است. این نمودار را میتوان بـر اسـاس فراوانـی تجمعـی یـا درصـد فراوانی تجمعی نیز رسم کرد. برای رسم یک نمـودار میلـهای از طریـق منـوی Graph بـه صـورت زیـر عمـل میشود:

- ابتدا از نوار ابزار منوی Graphs و سپس گزینهی Bar... و انتخاب کرده تا پنجرهی آن مطابق با شکل ۲.۳ باز شود.
- ۲. به منظور رسم نمودار میله ای برای یک متغیر حالت Simple را انتخاب کرده و عبارت Summaries
 ۲. به منظور رسم نمودار میله ی برای یک متغیر حالت Simple کایک
 ۲. می کنیم (گزینه ی پیش فرض). سپس روی عبارت Define کلیک می کنیم.
- ۳. در پنجرهی فعال شده، متغیر گروهی مورد نظر را در قسمت Category Axis قرار میدهیم
 (بهعنوان مثال متغیر فشار خون که به ۳ گروه دستهبندی شده است را وارد میکنیم).
- ۴. با فعال کردن هر کدام از گزینههای زیر نمودار میلهای را میتوان براساس فراوانی، فراوانی تجمعی،
 درصد فراوانی و درصد فراوانی تجمعی رسم کرد:
 - N of cases
 - Cum. N
 - % of cases
 - cum. % -
 - پس از انتخاب نوع نمودار روی گزینه Ok کلیک می کنیم تا نمودار مورد نظر ایجاد شود.



شکل ۲.۳- پنجرهی Bar Charts

یک نمودار میلهای شبیه شکل ۲.۴ در پنجرهی Output باید ظاهر شود. اطلاعات شرح داده شده در این نمودار یک نسخهی گرافیکی از جدول توزیع فراوانی برای متغیر فشار خون است. بیشترین فراوانی مربوط به بیمارانی است که فشار خون معمولی دارند (۷۳/۶ درصد) و کمترین آن مربوط به فشار خون بالا است (۷/۸ درصد). از این نمودار نیز میتوان تعداد بیماران در هر گروه را حساب کرد.

پس از ساختن یک نمودار، به منظور بالا بردن کارایی آن برای ارائهی گزارش یا به دست آوردن اطلاعات اضافی، تعدیل کردن آن ضروری به نظر میرسد. در بستهی نرمافزاری SPSS میتوان این کار را به وسیلهی فعال کردن ویرایشگر نمودار (Chart Editor) انجام داد. برای این منظور فقط کافی است روی نمودار مورد نظر دو بار کلیک کرده تا ویرایشگر نمودار فعال شود.



شکل۲.۴- نمودار میلهای برای متغیر فشار خون

نمودار دایرمای (Pie Chart): نمودار دیگری که برای توصیف داده های گروهی بسیار مفید است، نمودار دایرمای یا کلوچهای است. نمودارهای دایرمای مانند سایر نمودارها مشخص کننده ی چگونگی پراکندگی داده ها هستند. برای به دست آوردن نمودارهای دایرمای ابتدا از نوار ابزار منوی Graphs، سپس گزینه ی ...op را انتخاب می کنیم. در پنجره Pie Charts با توجه به اینکه چه نوع نموداری مورد نیاز است یکی از سه حالت موجود را فعال کرده و روی گزینه و Define کلیک می کنیم. پس از وارد کردن متغیر مورد نظر، نوع نمودار را که می تواند بر اساس تعداد یا درصد باشد، انتخاب کرده و روی گزینه ی Ok کلیک می کنیم تا نمودار دایرمای تشکیل شود. شکل ۲.۵ نمودار دایرمای را برای متغیر فشار خون نشان می دهد. این نمودار نیز نسخه ی گرافیکی دیگری از جدول توزیع فراوانی است.



شکل۲.۵- نمودار دایرهای برای متغیر فشار خون

۲٫۲ آمار توصیفی دادههای کمی

در ادامه به بررسی روشهای آماری برای دادههای کمی میپردازیم. در دادههای کیفی بیشتر بر اساس جداول توزیع فراوانی و نمودارها اقدام به خلاصهسازی اطلاعات میکنند اما برای دادههای کمی وضعیت بـه گونـهای دیگر است، به عبارت دیگر روشهای خلاصهسازی این نوع دادهها کمی متفاوت است.

با استفاده از جداول توزیع فراوانی و نمودارها می توان تا حدودی اطلاعات نهفته از دادهها را مختصر و محسوس کرد. با این حال سعی می شود تا این اطلاعات به صورت یک یا چند عدد معقول در آورده شوند تا هم بتوان یک ایده یکلی درباره ی خصوصیت مورد مطالعه بدست آورد و هم نتیجه ی مطالعات را به سادگی گزارش کرد. آماردانان داده ها را معمولاً به صورت یک یا چند عدد به نام معیار خلاصه می کنند که این معیارها قسمتی از اطلاعات موجود در داده ها را منعکس می سازند. ما همواره در تصمیم گیریها از این گونه معیارها استفاده می کنیم. برای فهم بیشتر موضوع فرض کنید فردی از سفر خود می گوید و شما قسمتهایی از صحبتهای او را می شنوید. وی می گوید « در آنجا اکثر افراد سیاه چرده بودند» در واقع او نمی گوید به ترتیب افرادی را که دیدهام سیاه، سفید، زرد و یا ... بودهاند بلکه با خلاصه سازی نتایج مشاهداتی که داشته است در قالب کلمه ی « اکثر»، مشاهدات را در یک معیار خلاصه کرده است. با این جمله ی وی شما می توانید تصور کنید، جایی که وی به آنجا سفر کرده است یا از کشورهای آفریقایی و یا از کشورهای عربی خاور میانه بوده است. حال وی در ادامه می گوید « بسیاری از مردان پارچه هایی سفید بر سر می گذاردند » با این جمله شما می توانید دریابید که ایس کشور از کشورهای عربی بوده است. یعنی با افزایش معیارهای خلاصه ای که گفته می شود درک ما از جامعه بیشتر و بیشتر می شود. لذا این دو معیار نقش اساسی در شناسائی یک جامعه ی ناآشنا ایفا می کنند.

در این قسمت برای دادههای کمی جداول فراوانی، نمودارها و معیارهای خلاصهسازی به طور مجزا تـشریح خواهند شد.

۱,۲,۲ جدول توزيع فراواني

در این حالت جدول توزیع فراوانی به راحتی توسط همان فرآیندی که در بخش ۲٫۱٫۱ توضیح داده شد، قابـل دستیابی است اما شرایط ساختن این جداول به گونهای دیگر اسـت و بـا توجـه بـه ماهیـت دادههـای کمـی (گسسته یا پیوسته) تعیین میشوند.

برای دادههای گسسته، جدول توزیع فراوانی بر اساس ردههای متغیر گسسته محاسبه می شود. به عنوان مثال تعداد فرزندان یک متغیر گسسته است و جدول توزیع فراوانی بر اساس ردههای این متغیر یعنی تعداد فرزندان محاسبه می شود. این جدول زمانی مفید است که دامنه یتغییرات متغیر مورد مطالعه کم باشد. برای دادههای پیوسته، ابتدا دادهها را به چند رده دسته بندی می کنیم. به عبارت دیگر یک متغیر کمی به یک متغیر کیفی تبدیل می شود. تعداد طبقات بایستی به گونه ای انتخاب شود که از یک طرف اطلاعات زیادی در داخل هر طبقه ناپدید نشود و از سوی دیگر تعداد طبقات آنقدر زیاد نباشد که تحلیل داده ها مشکل شود. روش های تجربی متعددی برای تعیین تعداد طبقات پیشنهاد شده است که از جمله ی آن می توان به فرمول

$K=1+T/T \log(n)$

اشاره کرد که در آن n تعداد مشاهدات و K تعداد طبقات میباشد. پس از تعیین تعداد طبقات، ابتدا دادهها را دستهبندی کرده (با استفاده از روش کدبندی) و سپس جدول توزیع فراوانی را رسم میکنیم. به عنوان مثال، در مطالعهی عوامل مؤثر بر حملهی قلبی، جدول توزیع فراوانی برای متغیر سن در جدول ۲.۲ به نمایش در آمده است.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	age <=50	8	4.5	4.5	4.5
	51< age<=70	130	73.0	73.0	77.5
	age>=71	40	22.5	22.5	100.0
	Total	178	100.0	100.0	

CatAge

جدول ۲۰۲ - توزیع فراوانی برای متغیر سن

۲٫۲٫۲. نمودارهای آماری

برای دادههای گسسته، نمودار میلهای مناسب ترین نمودار است که همانند نمودار میلهای برای دادههای کیفی رسم میشود. تنها تفاوت در این است که به جای گروههای مختلف دادهی کیفی، از ردهها در محور افقی استفاده میشود. به عنوان مثال، برای متغیر تعداد فرزندان یک نمودار میلهای مطابق با شکل ۲.۶ رسم شده است. نمودار دایرهای نیز برای این دادهها به سادگی رسم میشود.



شکل۲.۶- نمودار میلهای برای متغیر تعداد فرزندان

برای توصیف دادههای پیوسته به جای استفاده از نمودار میلهای از یک نمودار گرافیکی دیگر که نشان دهندهی توزیع فراوانی نیز میباشد، استفاده میشود. این نمودار به هیستوگرام (Histogram) معروف است که متداولترین نمودار برای ترسیم دادههای پیوسته میباشد. برای ساختن هیستوگرام به صورت زیر عمل میشود:

- ۱. از نوار ابزار منوی Graphs و سپس گزینهی ...Histogram را انتخاب می کنیم.
- ۲. در پنجرهی هیستوگرام متغیر سن را در قسمت :Variable قرار میدهیم. اگر بخواهیم در نمودار مرودار مربوطه منحنی نرمال نیز رسم شود، گزینهی Display Normal Curve را فعال میکنیم.
- ۳. برای اضافه کردن عنوان (Title) و یا نوشتن زیرنویس (Footnote) در هیستوگرام می توان از گزینه ی
 Titles استفاده کرد.
- ۴. در انتها بر روی گزینه Ok کلیک کرده تا هیـستوگرام رسـم شـود. شـکل ۲.۷ هیـستوگرام وزن بیماران دچار حمله یقلبی را نشان میدهد.

محور افقی (x-axis) نشان دهنده یسن بیماران در مطالعه ی مورد بررسی است که مقادیر آن از ۴۰ تا ۹۰ در تغییر است. محور عمودی (y-axis) نیز نشان دهنده ی تعداد بیماران برای هر رده است. به عبارت دیگر هر ستون نشان دهنده ی تعداد بیماران برای هر دارند.



شکل۲.۷- نمودار هیستوگرام برای متغیر سن

تعداد ردهها در هیستوگرام به طور خودکار، توسط بستهی نرمافزاری SPSS مشخص میشود اما میتوان با کم و زیاد کردن تعداد ردهها هیستوگرام را تغییر داد. به عنوان مثال میتوان تعداد ردهها در مثال قبل را به ۷ رده کاهش داد. این عمل از طریق ویرایشگر نمودار قابل انجام است. شکل ۲.۸ هیستوگرام وزن بیماران دچار حملهی قلبی را با ساختار جدید نشان میدهد.



شکل ۲.۸- نمودار هیستوگرام برای متغیر سن با تعداد هفت رده

روش دیگری که برای توصیف دادههای کمی به کار برده می شود استفاده از نمودار شاخه و برگ & Stem (دوش دیگری که برای توصیف دادههای کمی به کار برده می شود استفاده از نمودار شاخه و برگ & مودار العام است. رسم نمودار شاخه و برگ به دلیل از دست ندادن اطلاعات به نمودارهای فراوانی ترجیح داده می شود. برای رسم نمودار شاخه و برگ، هر مقدار کمی بدست آمده را به دو بخش تقسیم نموده و یک بخش را شاخه و بخش دیگر را برگ می نامیم. آنگاه برای هر شاخه برگهای مربوطه را به طور جداگانه در مقابل آن می نوان مثال، شکل (۲۰۹) این مقابل آن می نویسیم. شکل حاصل را نمودار شاخه و برگ دادهها می نامند. به عنوان مثال، شکل (۲۰۹) این مقابل آن می نویسیم. شکل حاصل را نمودار شاخه و برگ دادهها می نامند. به عنوان مثال، شکل (۲۰۹) این نمودار را برای متغیر سن در افرادی که دچار حملهی قلبی شده اند، نشان می دهد. با مقابل آن می نویسیم. شکل حاصل را نمودار شاخه و برگ دادهها می نامند. به عنوان مثال، شکل (۲۰۹) این نمودار را برای متغیر سن در افرادی که دچار حمله قلبی شده اند، نشان می دهد. با مقابل آن می نویسیم. شکل حاصل را نمودار شاخه و برگ داده می نامند. به عنوان مثال، شکل (۲۰۹) این نمودار را برای متغیر سن در افرادی که دچار حمله و برگ خاده می نامند. می دهد. دامودار را برای متغیر سن در افرادی که دچار حمله و برگ ضمن داشتن می دهد. با مقایسه این نمودار و هیستوگرام دیده می شود که در هیستوگرام با ایجاد طبقات دیگر دستیابی به مقادیر انفرادی مشاهدات امکان پذیر نیست اما در نمودار شاخه و برگ ضمن داشتن ایده ی تصویری از نحوه ی توزیع داده ها به تک تک اطلاعات نیز دستر سی داریم.

Age in years Stem-and-Leaf Plot

Frequency	Stem & Leaf
6.00	4. 789999
18.00	5. 00111122333444444
21.00	5. 555566666777778888999
40.00	6.000000011111122222233333344444444444444
42.00	6. 55555566666666666777777777888888888899999999
20.00	7.000000000111223334
19.00	7. 5555666777888899999
9.00	8.000112344
3.00	8.555
Stem width:	10
Each leaf:	1 case(s)

شکل۲.۹- نمودار شاخه و برگ برای متغیر سن

۳,۲,۲. معیارهای تمرکز و پراکندگی

علاوه بر نمودارهای آماری، خصوصیات اصلی دادهها توسط شاخصهای عددی نیز میتوانند خلاصه شوند. این شاخصها به دو دستهی تمرکز یا مکان (Location) و پراکندگی یا تغییرپذیری (Variability) تقسیم بندی میشوند. در یک مجموعه داده، معیارهای تمرکز میزان تمایل به مرکزیت را نشان میدهند در حالی که چگونگی پراکنش آنها توسط معیارهای پراکندگی اندازه گیری میشود. به عبارت دیگر، معیارهای پراکندگی چگونگی پخششدن دادهها حول معیارهای تمرکز را نشان میدهند. در این قسمت روشهای محاسبهی معیارهای تمرکز و پراکندگی توسط بستهی نرمافزاری SPSS شرح داده می شود اما قبل از آن به صورت خلاصه معیارهای خلاصه سازی را تعریف می کنیم.

غالباً مجموعهای از دادهها بهوسیلهی یک عدد بهعنوان نمایندهی دادهها معرفی میشوند. بـرای نمونـه فـرض کنید، متوسط وزن بیماران حملهی قلبی ۶۵ کیلـوگرم اسـت. ایـن نـوع انـدازه را یکـی از معیارهـای تمرکـز مینامند. بهطور کلی یک معیار تمرکز، مرکز ثقل یک هیستوگرام یا یـک منحنـی توزیـع را نـشان مـیدهـد. معیارهای تمرکز متعددی تعریف شدهاند که به برخی از آنها در زیر اشاره شده است:

 ۱. نما (Mode): آن مشاهدهای که بیشترین فراوانی را دارا میباشد نما نام دارد که یک لغت فرانسوی و به معنای متداول ترین است. این معیار برای متغیرهای گروهی یا متغیرهای گسسته مناسب میباشد. اگر بیشترین فراوانی در دو یا چند رده اتفاق بیافتد، می گوییم نمونه دو مدی (Bi Modal) یا چند مدی (Multi Modal) است.

- ۲. میانه (Median): اگر دادههای یک متغیر را از کوچک به بزرگ مرتب کنیم آنگاه مشاهدهی وسط این دادهها میانه است. بهعبارت دیگر، نقطهای که حداقل نیمی از دادهها از آن کمتر باشند میانه نام دارد. این معیار یکی از معیارهای مهم تمرکز میباشد.
- ۳. میانگین (Mean): مجموع تمام مقادیر یک متغیر تقسیم بر تعداد آنها میانگین داده ها را میسازد که به آن میاگین حسابی نیز می گویند. این معیار متداول ترین معیار، برای نشان دادن تمرکز در متغیرهای عددی است.

وقتی که مقادیر میانگین، میانه و نما تقریباً شبیه به هم باشند می توان انتظار داشت که داده ها به صورت متقارن حول مرکز پراکنده شدهاند یا به عبارت دیگر، توزیع آنها به فرم یک توزیع متقارن (مثلا نرمال) نزدیک است. حال اگر داده ها به صورت تقریبی حول یک نقطهی مرکزی پراکنده نباشند آنگاه درک یک مفهوم به نام چولگی (Skew ness) مهم به نظر می رسد. پس توزیع های چوله آنهایی هستند که متقارن نمی باشند. میزان عدم تقارن منحنی را چولگی می نامند که می نواند به می را زمان می رسد. پس توزیع مان در از می نام نواند به نام نباشند آنگاه درک یک مفهوم به نام چولگی (Skew ness) مهم به نظر می رسد. پس توزیع های چول آنهایی هستند که متقارن نمی باشد. میزان عدم تقارن منحنی را چولگی می نامند که می تواند به دو صورت چپ (منفی) و راست (مثبت) ظاهر شود. شکل ۲۰۱۰ یک توزیع چوله به سمت راست را نشان می دهد. به دلیل این که در توزیع های چوله، میانگین به سمت مقادیر دورافتاده تغییر می کند، استفاده از میانه به عنوان معیار تمرکز مناسب به نظر می رسد چون این معیار تحت تا ثیر مقادیر خیلی کوچک یا بزرگ قرار نمی گیرد.

- ۴. چارکها (quartiles): چارکها اندازههایی هستند که دادههای مرتب شده را به چهار قسمت مساوی تقسیم میکنند. معمولاً چارک اول، دوم و سوم را با 91, 22, 23 نشان داده که بهترتیب ۵۲، ۵۰ و ۷۵ درصد از دادههای مرتب شده زیر آنها قرار دارند. واضح است که چارک دوم همان میانه است.
- ۵. دهکها (Deciles) و صدکها (Percentiles): دهکها اندازههایی هستند که دادههای مرتب شده را به ده قسمت مساوی تقسیم میکنند و اندازههایی که دادهها را به صد قسمت مساوی تقسیم کنند، صدکها نامیده می شوند.

این معیارها برای تعیین وضعیت قرار گرفتن یک مشاهده در مقایسه با سایر مشاهدات برای یـک مجموعـه از دادههای نمونهای یا جامعه بکار برده میشوند. در بحث آمارهای توصیفی چارکها، دهـکهـا و صـدکهـا بـا عنوان کلی چندک (Quantile) نام برده میشوند.

معیارهای تمرکز مانند میانگین، میانه و نما توصیف کنندهی وضعیت کامل توزیع دادهها نیستند. به عبارت دیگر دو مجموعهی داده، که دارای میانگینهای یکسان هستند ممکن است پراکندگی متفاوت داشته باشند و زمانی میتوان توزیع دادهها را دقیقاً توصیف نمود که علاوه بر شناخت معیار مناسبی برای مرکزیت آن، معیاری را هم برای پراکندگی آن تعیین نموده باشیم. معیارهای پراکندگی متداول عبارتند از:

- برد (Range): ساده ترین اندازه ی پراکندگی برای یک مجموعه داده برد است که از اختلاف مقادیر ماکسیمم و مینیمم در یک مجموعه داده به دست میآید. برد نیز همانند میانگین تحت تاثیر داده های پرت قرار می گیرد و در چنین حالت هایی یک معیار مناسب برای پراکندگی نیست.
- ۲. برد چارکها (Interquartile range): تفاوت بین چارک اول و سوم، برد چارکها را تشکیل میدهد. این معیار در حقیقت برد پیراستهی دادهها است که تحت تاثیر دادههای دورافتاده قرار نمی گیرد.



شکل ۲.۱۰- نمودار هیستوگرام برای یک متغیر که شکل توزیع آن چوله است

- ۳. واریانس (Variance): هدف اصلی معیارهای پراکندگی نشان دادن مقدار پراکندگی دادهها از میانگین توسط یک اندازهی عددی است. بهترین راه حل مجموع توان دوم تفاضل مشاهدهها از میانگین است. اگر این معیار به تعداد مشاهدهها تقسیم شود یک معیار مهم برای پراکندگی محاسبه میشود که به واریانس معروف است.
- ۴. انحراف معیار (Standard Deviation): مفیدترین معیار پراکندگی جذر واریانس دادهها است که به آن انحراف معیار می گویند و اندازهی انحراف معیار نشاه دهندهی این است که مشاهدهها تا چه مقدار در اطراف میانگین قرار دارند. یک اندازهی کم (زیاد) برای انحراف معیار مجموعهای از دادهها نشان دهندهی این واقعیت است که دادهها در دامنهی کوچکی (بزرگی) حول میانگین پراکنده شدهاند.

میزان کشیدگی یا پخی منحنی فراوانی را نسبت به منحنی نرمال استاندارد، برجستگی (Kurtosis) آن مینامند. برحسب آن که ضریب برجستگی مثبت یا منفی باشد منحنی فراوانی نسبت به منحنی نرمال استاندارد کشیده یا پخ است. اگر این ضریب نزدیک صفر باشد برجستگی منحنی فراوانی طبیعی یا معمولی است.

با توجه به این که توزیع فراوانی از مفاهیم اساسی در تئوری آمار بوده و اساس بخش عمدهای از روشهای آمار استنباطی را تشکیل میدهد، معرفی معروفترین توزیع فراوانی، توزیع نرمال استاندارد که بهصورت زنگی شکل بوده و از لحاظ تقارن، کشیدگی و پخی متناسب است، ضروری به نظر میرسد. در عمل بهندرت متغیری وجود دارد که منحنی فراوانی آن کاملاً نرمال استاندارد باشد. اغلب منحنی فراوانی دادهها نامتقارن، کشیده و یا پخ هستند. میزان نرمال نبودن را با دو معیار چولگی و برجستگی می سنجند.

معیارهای تمرکز و پراکندگی با استفاده از شیوههای زیر در بستهی نرمافزاری SPSS قابل محاسبه هستند:

شيوهي اول: Descriptive procedure

ابتدا از نوار ابزار منوی Analysis و سپس از زیرمنوهای باز شده Descriptive statistics و ...Descriptive و انتخاب انتخاب کرده تا پنجرهی مربوط به آن باز شود (شکل ۲۰۱۱ را ببینید). متغیر یا متغیرهای مورد نظر را انتخاب کرده و برای دستیابی به آمارههای توصیفی بر روی گزینهی ...Options کلیک میکنیم (شکل ۲۰۱۲). آمارههای میانگین، انحراف معیار، ماکسیمم و مینیمم به صورت پیشفرض انتخاب شدهاند. سایر آمارههای توصیفی، بسته به نیاز، میتوانند انتخاب یا حذف شوند. با کلیک کردن روی گزینهی Continue و سپس اجرا کردن دستور، آمارههای موردنظر محاسبه خواهند شد.

Descriptives		
 Gender [gender] Blood pressure [bp] Physically active [active History of myocardial inf History of diabetes [diat 	Variable(s):	OK Paste Reset Cancel Help
Save standardized values as vari	ables .	Options

شکل ۲.۱۱- پنجرهی Descriptives

Descriptives: Op	tions	
🔽 Mean	🔲 Sum	Continue
Dispersion	= w :	Cancel
Variance	Minimum	Help
□ Valiance	S.E. mean	
Distribution]
🗆 Kurtosis	🔲 Skewness	
Display Order		1
 Variable list 		
C Alphabetic		
C Ascending mea	ans	
C Descending m	eans	

شکل Descriptives: Options پنجرهی -۲.۱۲

شیوهی دوم Frequencies procedure

در این قسمت از دستور جدول توزیع فراوانی استفاده میکنیم (آمارههای توصیفی نیز از همین دستور محاسبه میشوند). برای این منظور پس از بازکردن پنجرهی Frequencies (شکل۲۰۲)، بر روی گزینهی Statistics کلیک کرده تا پنجرهی مربوط به آمارههای توصیفی ظاهر شود (شکل۲۰۱۳). مطابق با شکل۲.۱۳، این جعبه شامل چهار قسمت به شرح زیر است:

- تمایل بهمرکز: میانگین، میانه، نما و مجموع
 - پراکندگی: انحراف معیار، واریانس، برد، مینیمم، ماکسیمم و خطای معیار
 - چندکها: چارکها، دهکها و صدکها
 - معیارهای توزیع: چولگی و کشیدگی

با فعال کردن هر کدام از آمارههای فوق، معیار مورد نظر محاسبه خواهد شد.

Frequencies: Statistics		δ
Percentile Values Quartiles Cut points for: Percentile(s): Add Change	Central Tendency Mean Median Mode Sum	Continue Cancel Help
Bemove Dispersion Std. deviation Variance Maximum Range S.E. mean	 Values are group n Distribution Skewness Kurtosis 	nidpoints

شكل-۲.۱۳ پنجرهی Frequencies: Statistics

برخی از نمودارهای آماری مانند هیستوگرام یا نمودار میلهای توسط این دستور نیز قابل محاسبه است. برای این منظور در پنجرهی Frequencies بر روی گزینهی Chart کلیک کرده تا پنجرهی مربوط به آن باز شود (شکل۲۰۱۴). با در نظر گرفتن ماهیت دادهها که به صورت گروهی یا عددی هستند، می توان نمودار مورد نظر را رسم کرد.

requencies: Charts	×	
Chart Type	Continue	
None	Cancel	
 Bar charts Pie charts 	Help	
C Histograms:		

شکل۲.۱۴ پنجرهی Frequencies: Charts

نکته: نمودارهای میلهای و دایرهای میتوانند بر اساس فراوانی مشاهدات یـا درصـد فراوانـی مـشاهدات رسـم شوند.

شیوهی سوم Explore procedure

این شیوه آمارهای خلاصه شده و نمودارهای مفید آماری را که اساس هر تحلیل هستند، محاسبه می کند. دلیلهای زیادی از جمله غربال کردن دادهها، تشخیص دادههای دورافتاده و چک کردن فرضیهها برای استفاده از این روش وجود دارد. غربال کردن دادهها ممکن است نشان دهندهی دادههای غیرمعمول یا دورافتاده باشد. بررسی و کنکاش دادهها به شما کمک می کند که آیا تکنیکهای آماری که برای دادهها در نظر گرفته اید، مناسب هستند یا نه؟ این بررسیها ممکن است نشان دهندهی این باشند که دادهها نیاز به تبدیل دارند و یا آزمونهای ناپارامتری مورد نیاز است. برای استفاده از این شیوه به مورت زیر عمل می کنیم:

۱. از نوار ابزار منوی Analysis و از زیر منوهای بعدی گزینه های Descriptive Statistics و Explore از نوار ابزار منوی در منوهای بعدی گزینه می و از منود.

Gender [gender] Blood pressure [bp] Physically active [active]		Dependent List: Age in years [age] Age in years [age]	OK Paste
 History of myocardial in 			Reset
History of diabetes [dia		Factor List:	Cance
	\rightarrow		Help
	×	Label Cases by:	
Display			
Both C Statistics C Pl	lots	Statistics Plots Opt	ions

شکل۲.۱۵- پنجرهی Explore

- ۲. متغیرهای پیوستهای که آمارههای توصیفی برای آنها نیاز است را در قسمت Dependent List قـرار میدهیم.
- ۳. آمارههای توصیفی یا نمودارهای آماری را میتوان بهترتیب با گزینههای Statistics و Plots بهصورت تک تک یا همزمان فعال کرد.
- ۴. پنجرهی Statistics برای آمارهه ای توصیفی، برآوردگره ای استوار، شناس ایی نقاط دورافت اده و چندکها مورد استفاده قرار می گیرد.
- ۵. پنجرهی Plots برای نمایش نمودارهای شاخه و برگ و جعبهای (باکس ویسکر)، هیستوگرام،
 ۵. تست نرمال بودن و نمودار نرمال و تست همسانی واریانس به همراه نمودار آن تهیه شده است.
- ۶. با انتخاب هر کدام از کمیتهای مورد نظر گزینه Continue را کلیک کرده و نهایتاً دستور را اجرا می کنیم. جدول ۲.۳ آمارههای توصیفی را برای سن بیماران در مثال مربوط به حادثه یقلبی نشان می دهد.

	Cases					
	Valid Missing		Total			
	Ν	Percent	Ν	Percent	Ν	Percent
Age in years	178	100.0%	0	.0%	178	100.0%

Case Processing Summary

			Statistic	Std. Error
Age in years	Mean		65.25	.663
	95% Confidence	Lower Bound	63.94	
	Interval for Mean	Upper Bound	66.56	
	5% Trimmed Mean		65.14	
	Median		65.00	
	Variance		78.359	
	Std. Deviation		8.852	
	Minimum		47	
	Maximum		85	
	Range		38	
	Interquartile Range		11	
	Skewness		.165	.182
	Kurtosis		464	.362

Descriptives

جدول ۲.۳ - آماره های توصیفی برای متغیر سن

در پایان یکی از مفیدترین نمدارهای آماری که برای دادههای پیوسته رسم می شود را معرفی می کنیم. این نمودار به نمودار جعبهای یا باکس ویسکر معروف است. نمودار باکس ویسکر نموداری است که به کمک معیارهای مرکزی و پراکندگی، موقعیت مجموعه دادهها را به شکلی بسیار گویا و مفید ارائه می دهد. این نمودار ابتدا توسط آمار شناس معروف توکی ارائه شد که با استفاده از یک مستطیل (باکس) و دو خط در دو طرف مستطیل (ویسکر) و به وسیلهی میانه، چارکهای اول و سوم و مقادیر ماکسیمم و مینیمم رسم می شود. با استفاده از این نمودار می توان اطلاعات زیر را در مورد دادهها کسب نمود.

- اگر میانه نزدیک وسط مستطیل (باکس) باشد توزیع داده ا تقریباً متقارن است.
- اگر میانه در طرف چپ وسط مستطیل باشد توزیع چوله به راست و اگر میانه در طرف راست وسط مستطیل قرار گیرد توزیع چوله به چپ است.
- اگر خطوط دو طرف مستطیل (ویسکرها) تقریباً برابر باشند توزیع دادهها به توزیع متقارن نزدیک بوده و در صورت نامساوی بودن خطوط توزیع دارای چولگی است.
- در مقایسه نمودار باکس_ویسکر دو مجموعه از دادهها میتوان پراکندگی آنها را با توجه به طول مستطیلهای نمودار با یکدیگر مقایسه نمود. مستطیلی که طول بزرگتری دارد دارای پراکندگی بیشتر میباشد.
- دادههای دورافتادهی ضعیف و دادههای دورافتادهی قـوی را بـا اسـتفاده از نمـودار بـاکســ ويـسکر می توان تعیین نمود.

در ادامه نمودار باکس ویسکر سنهای بیماران قلبی را رسم نموده که شکل ۲.۱۶ آن را نشان میدهد.



شکل۲.۱۶– نمودار باکس – ویسکر برای متغیر سن

۳.طرحهای آمارگیری نمونهای

هدف اصلی در علم آمار، پی بردن به خصوصیات یک جامعهی هدف معین بر اساس دادههای جمع آوری شده است. معمولاً مطالعهی ویژگیهای یک جامعه، وقتی که بسیار بزرگ باشد، مستلزم صرف هزینه و وقت زیادی میباشد و در بسیاری از مواقع، این امر اصولاً امکان پذیر نیست. در این گونه موارد، زیر مجموعهای از اجزای جامعه را طبق قاعده و ضوابط خاصی انتخاب می کنند و نتایج به دست آمده را به کل جامعه تعمیم می دهند. بخش انتخاب شده از جامعه را نمونه یا نمونهی آماری می نامند. به عبارت دیگر، نمونه بخشی از جامعه است که برای مطالعهی خصوصیات جامعه به روش صحیح انتخاب می شود.

از آنجایی که هدف از مطالعه ینمونه ای تعمیم نتایج به کل جامعه است، انتخاب صحیح نمونه بسیار اهمیت دارد. به عبارت دیگر، نمونه گیری مصداق ضرب المثل «مشت نمونه ی خروار است» می باشد اما باید توجه داشت که هر مشتی نمونه ی خروار نیست. در این ارتباط روش های نمونه گیری و طرح آزمایش ها از مجموعه روش های آماری هستند که استفاده از آن ها در رسیدن به نتایج صحیح و قابل اطمینان ضروری است. به عنوان مثال اگر هدف یک مطالعه برآورد متوسط در آمد افراد جامعه باشد، نمی توان نمونه ی انتخاب شده را فقط از قشر کارمندان و یا فقط از قشر افراد با مشاغل آزاد انتخاب نمود، بلکه بایستی با شناخت جامعه، تعداد نمونه ی مناسب را به نسبت جمعیت اقشار مختلف جامعه به دست آورده و اطلاعات لازم را جمع آوری نمود. برای بررسی خصوصیات یک جامعه، اگر تمام اجزای آن جامعه بررسی شوند می گوییم سرشماری انجام گرفته است اما یک جامعه آنقدر بزرگ است که نمی توان برای آن سرشماری انجام داد. بنابراین اکثر مطالعات آماری، مطالعه های نمونه ای هستند. سرشماری ها هر چند سال یک مرتبه برای شناخت جامعه و تهیه چار چوب مناسب جهت مطالعات نمونه ای انجام می گیرند. در این مورد می توان از سرشماری نفوس در ایران که هر ۱۰ سال یک مرتبه انجام می گیرد و یا سرشماری های که در سال یک مرتبه برای شماری نوس می ای متاخت جامعه و تهیه

به طور کلی دو نوع روش نمونه گیری وجود دارد که به نمونه گیری احتمالاتی و غیر احتمالاتی معروف هستند. نمونه گیری احتمالاتی، فرایندی است که در آن احتمال انتخاب هر کدام از واحدهای جامعه از قبل معلوم و مشخص است اما نمونه گیری غیر احتمالاتی، روشی است که در آن نمی توان به واحدهای جامعه، احتمال مشخصی را برای انتخاب شدن نسبت داد و در نتیجه دقت نمونه گیری میسر نیست. در واقع نمونه گیری غیر احتمالاتی هر چند ممکن است نتایج مفیدی در بر داشته باشد اما به دلیل این که امکان ارزیابی نتایج بر اساس قوانین احتمال میسر نیست، معمولاً قابل اعتماد نمیباشد. در اینجا ذکر این نکته ضروری است که نتایج یک آمار گیری نمونه ای هر گز با قطعیت و حتمیت همراه نیست، زیرا اگر میزان خطاهای غیر نمونه ای ناچیز باشد، نتایج تحت تاثیر خطایی تحت عنوان خطای نمونه گیری قرار دارد که ناشی از انتخاب بخشی از جامعه به جای کل جامعه میباشد. میزان خطای نمونه گیری را میتوان با افزایش حجم نمونه تقلیل داد. لکن نظر به اینکه زیاد کردن حجم نمونه روی عوامل متعددی از جمله حجم عملیات میدانی و هزینه آمار گیری تأثیر دارد، لذا در این زمینه باید عوامل مورد اشاره و دیگر عوامل مربوط، مد نظر قرار گیرد.

نمونه گیری از یک جامعه ممکن است به دو طریق انجام شود: نمونه گیری بدون جایگذای و نمونه گیری با جایگذاری. در روش اول یعنی نمونه گیری بدون جایگذاری، هر کدام از واحدها فقط یک بار شانس انتخاب شدن دارد و در حقیقت وقتی یک واحد انتخاب شد، دیگر در معرض انتخاب شدن قرار نخواهد گرفت. در روش دوم یعنی نمونه گیری با جایگذاری که در موارد خاصی مورد استفاده قرار می گیرد، هر واحد جامعه پس از انتخاب، به جامعه برگردانده می شود و مجدداً در معرض انتخاب قرار می گیرد و بدین ترتیب در نمونه انتخابی، ممکن است برخی از واحدها تکراری باشند.

انتخاب روش نمونه گیری برای یک طرح آماری، از مهمترین بخشهای یک طرح نمونه گیری است که توسط آمارشناسان حرفهای انجام می گیرد. در واقع، یک آمارشناس مجرب، با آگاهی کافی از روشهای مختلف نمونه گیری و شناخت ویژگیها و قابلیتهای هر کدام از این روشها میتواند یک روش نمونه گیری مناسب را برای آمار گیری مورد نظر مشخص نماید.

در نمونه گیری احتمالاتی، اگر شانس انتخاب شدن برای همهی اعضای جامعه برابر و انتخاب هر عضو جامعه تاثیری در انتخاب اعضای دیگر نداشته باشد، آن گاه نمونهی انتخاب شده تصادفی ساده است. به عنوان مثال، برای بررسی اندازهی قد دانشجویان بیست سالهی یک شهر، انتخاب مثلاً ۱۵۰ نفر از بین این جمعیت به طور تصادفی یا انتخاب ۲۰۰ خانوار از خانوارهای یک شهر، برای تعیین هزینهی خانوارها، مثالهایی از نمونههای تصادفی ساده هستند.

یکی دیگر از روشهای نمونه گیری که بسیار سریع و راحت اقدام به انتخاب نمونهها میکند، روش نمونه گیری سیستماتیک است. در این روش ابتدا واحدهای جامعه از اول تا آخر (از ۱ تا N) شماره گذاری شده و سپس فاصلهی نمونه گیری (k=N/n) مشخص می شود. در ادامه یک عدد به تصادف بین ۱ تا k انتخاب شده تا شماره ی اولین واحد نمونه به دست آید. با اضافه کردن متوالی k به شماره ی اولین واحد نمونه، شماره ی سایر واحدهای نمونه نیز به دست می آید.

در ادامه روشهای نمونه گیری با طبقه بندی و خوشهای را که به طور معمول در آمار گیریها مورد استفاده قرار می گیرند، شرح داده می شوند:

نمونه گیری تصادفی با طبقهبندی:

در این روش نمونه گیری، ابتدا جامعه به زیر جامعه هایی همگن تقسیم بندی شده که هیچ گونه عنصر مشتر کی ندارند و مجموعهی عناصر آن ها برابر با مجموع کل جامعه است. هر زیرجامعه را یک طبقه مینامند و وقتی طبقات مشخص شد اگر از هر طبقه، نمونه ای به روش نمونه گیری تصادفی ساده گرفته شود، شیوه کلی نمونه گیری را نمونه گیری تصادفی با طبقه بندی مینامند. معمولاً در نمونه گیری با طبقه بندی، هدف از تقسیم جامعه به طبقات مختلف، گروهبندی جامعه به زیرجامعههای همگن میباشد و با این کار میتوان با تعداد نمونه کمتری نسبت به نمونه گیری تصادفی ساده، بر آوردهایی با دقت موردنظر به دست آوریم. در اینجا طبقه همگن به این معناست که واحدهای هر طبقه از نظر متغیرهای مورد بررسی با یک دیگر تشابه زیادی دارند و بدیهی است که طبقات مختلف باید با یکدیگر تفاوت زیادی داشته باشند.

نمونهگیری تصادفی خوشهای:

تاکنون شیوههایی را به عنوان روشهای نمونه گیری معرفی کردیم که استفاده از آنها مستلزم وجود فهرستی از واحدهای جامعه تحت عنوان چارچوب نمونه گیری است تا نمونه تصادفی از آن انتخاب شود. وقتی حجم جامعه خیلی بزرگ نباشد و چارچوب مورد نیاز برای انتخاب واحدهای نمونه در دسترس باشد می توان روشهای مذکور را بدون اشکال مورد استفاده قرار داد. اما وقتی حجم جامعه بزرگ می شود، این شیوهها با مشکلاتی همراه می باشند. یکی از این مشکلات عدم وجود فهرست دقیق، کامل و بهنگام از واحدهای جامعه است. مشکلاتی همراه می باشند. یکی از این مشکلات عدم وجود فهرست دقیق، کامل و بهنگام از واحدهای جامعه است. مشکلاتی همراه می باشند. یکی از این مشکلات عدم وجود فهرست دقیق، کامل و بهنگام از واحدهای جامعه است. مشکلاتی مواردی، مراجعه به واحدهای نمونه، بسیار زیاد می باشد. در چنین مواردی، جامعه را به قسمتهای ا مغرافیایی، هزینه مراجعه به واحدهای نمونه، بسیار زیاد می باشد. در چنین مواردی، جامعه را به قسمتهایی از که اصطلاحاً خوشه نامیده می شود و مبنای تشکیل خوشهها در اغلب اوقات، تقسیم بندی های جغرافیایی از قبیل استان، شهرستان، دهستان، شهر و بلوک یا آبادی است، تقسیم بندی می کنند. خوشهها باید کل جامعه را پوشش دهند و فاقد همپوشانی باشند. به بیان دیگر هر کدام از عناصر جامعه باید به یکی و فقط یکی از این خوشهها تعلق داشته باشد. با استفاده از چارچوبی که بر پایه روش فوقالذکر حاصل می شود و نیز با استفاده از مونه های که دارای اطلاعات کمکی مفیدی باشند، می توان طرحهای نمونه گیری پیچیده ای از قبیل نمونه گیریهای خوشهای یک مرحلهای، دو مرحلهای و یا سه مرحلهای را به کار برد.

یک فرضیهی موروثی در شیوههای تحلیلی بستههای نرمافزاری این است که دادهها نشاندهندهی یک نمونهی تصادفی ساده بدون جایگذاری از جامعهی مورد مطالعه هستند. این فرضیه در بسیاری از تحقیقها غیر منطقی به نظر میرسد و تحلیل دادهها بر اساس آن منجر به استنباطهای نادرست دربارهی جامعهی مورد بررسی خواهد بود. بستهی نرمافزاری SPSS امکان انتخاب و تحلیل یک نمونهی تصادفی را مطابق با یک طرح نمونه گیری پیچیده (نمونه گیری خوشهای یک یا چند مرحلهای، نمونه گیری با طبقهبندی، نمونه گیری با احتمال متغیر و غیره)، در قالب شیوهی Complex Samples فراهم می کند. شیوهی قرار می گیرد: موارد زیر مورد استفاده قرار می گیرد:

- طراحی و اجرای آمار گیری های نمونه ای

تحلیل دادههای نمونه گیری شده بر اساس یک طرح نمونه ای خاص

بدون توجه به این موضوع که کدام یک از موارد فوق مورد نیاز است، اطلاعات طرح نمونهای باید درون یک فایل به نام Plan File ذخیره شود. این فایل شامل مشخصات طرح نمونهای مورد نظر میباشد که برای استفادههای بعدی کاربرد دارد. Plan File ها به دو دستهی زیر تقسیم می شوند:

- Sampling Plan .۱ : معرفی مشخصات یک طرح نمونه گیری پیچیده برای انتخاب نمونه ها از جامعه ی مورد مطالعه، توسط این فایل انجام می شود.
- ۲. Analysis Plan: این فایل شامل مشخصات یک طرح نمونه گیری پیچیده برای تحلیل داده های نمونه گیری شده است. ساختار نمونه، روش های برآورد پارامترها در هر مرحله و منابع مورد نیاز متغیرها مانند وزن نمونه ای از طریق این فایل قابل دسترسی است.

به منظور سهولت در مراحل ساخت، اصلاح و اجرای هر کدام از فایلهای فـوق، یـک راهنمـای قـدم بـه قـدم (Wizard) برای این فایلها وجود دارد. در ادامه این فایلها توضیح داده میشوند.

۱٫۳ فایلهای Sampling Plan

برای ساختن یک فایل Sampling Plan، ابتدا از نوار ابزار منوی Analyses و سپس از زیر منوی فعال شده عبارت Complex Samples را انتخاب می کنیم. در ادامه، عبارت Select a Sample برای ایجاد فایل انتخاب می شود. شکل ۳.۱ پنجره ی راهنمای قدم به قدم برای این فایل را نشان می دهد. پیش از ساختن این فایل، فقط کافیست، داشته باشیم:

- .۱ یک جامعه آماری مناسب
- ۲. یک فهرست از واحدهای نمونه گیری
 - ۳. یک طرح نمونهای مناسب در ذهن

	What would you like to do?		
	Design a sample		
<u> </u>	Choose this option if you have not created a plan file. You will have the option to draw the sample.	File: p\SPSS\sampledata.csplarj	Browse
▲ <u>●</u>	C Edit a sample design		
	Choose this option if you want to add, remove, or modify stages of an existing plan. You will have the option to draw the sample.	File:	Browse
IS IL	🔿 🔿 Draw a sample		
X ares Hill	Choose this option if you already have	File:	Browse

شکل ۳.۱- پنجرهی راهنمای جادویی برای Sampling Plan Files

گزینه ی Design a sample (گزینه ی پیش فرض) را انتخاب کرده و برای ذخیره ی مشخصات طرح نمونه ای مورد نظر، یک نام مناسب در قسمت :File معرفی می کنیم. با کلیک کردن روی گزینه ی Next به مرحله ی بعد رفته و در صورتی که طرح نمونه گیری به صورت خوشه ای یا طبقه بندی باشد، آن ها را در مکان های مربوط قرار می دهیم (شکل ۲۰۰ را ببینید). در صورتی که طرح نمونه گیری جاری، قسمتی از یک طرح نمونه گیری بزرگ تر باشد، می توان وزن های نمونه ای در مرحله ی پیشین (یعنی در مرحله ی ی مربوط قرار می دهیم (شکل ۲۰۰ را ببینید). در صورتی که طرح نمونه گیری به صورت خوشه می یا طبقه بندی باشد، آن ها را در مکان های مربوط قرار می دهیم (شکل ۲۰۰ را ببینید). در صورتی که طرح نمونه گیری جاری، قسمتی از یک طرح نمونه گیری بزرگ تر باشد، می توان وزن های نمونه ای در مرحله ی پیشین (یعنی در مرحله ی نمونه گیری بزرگ تر باشد، می توان وزن های نمونه ای در مرحله ی پیشین (یعنی در مرحله ی نمونه گیری بزرگ تر) را به عنوان وزن های مرحله ی جاری بکار برد. این وزن ها در صورت وجود در جعبه ی Bueght قرار داده می شوند.

/ariables:		Stratify By:
🔗 Units [Units]		
🚺 Gender [gender]	•	
Graduation Date [gra		
Starting Salary [salary]		Clusters:
	•	
		Input Sample Weight:
	•	
	Stage Label:	

شکل ۲.۲- پنجرهی Sampling Wizard:Design Variables

در مرحلهی بعد، بر اساس طرح تعریف شده، روش نمونه گیری مناسب انتخاب می شود (شکل ۳.۳). انواع ایـن روش ها عبار تند از:

- تصادفی سادہ (Simple Random Sampling)
 - (Simple Systematic) سیستماتیک ساده
 - دنبالهای ساده (Simple Sequential)
- (Probability Proprtional to Size) با احتمال متناسب
 - سیستماتیک با احتمال متناسب (PPS Systematic)
 - دنبالهای با احتمال متناسب (PPS Sequential)
 - بروویر با احتمال متناسب (PPS Brewer)
 - مورفى با احتمال متناسب (PPS Murthy)
 - سامپ فورد با احتمال متناسب (PPS Sampford)

valiables.	Method	
🔗 Units [Units]	Type: Simple Random Sampling	•
🚺 Gender [gender]	Simple Random Sampling	
🛐 Graduation Date [gra	Simple Systematic	-
🔗 Starting Salary [salary]	Simple Sequential	
<i></i>	PP5 PPS Sustematic	
	PPS Sequential	
	PPS Brewer	
	PPS Murthy	
	Measure of <u>EPPS Sampford</u>	_
	Read from variable:	
	C Count data records	
		-

شکل ۳.۳- پنجرهی Sampling Wizard: Sampling method

برخی از انواع روشهای نمونه گیری با احتمال متناسب (با علامت PPS) تنها زمانی که متغیری به عنوان خوشه تعریف شده باشد، قابل دسترس هستند و این روشها فقط در حالت نمونه گیری یک مرحلهای میتوانند به کار گرفته شوند. اگر روش نمونه گیری با احتمال متناسب انتخاب شده باشد، باید مقادیر احتمالهای متناسب با هر واحد نمونه گیری از طریق جعبهی (MOS) Measure of Size معرفی شوند. این مقادیر میتوانند از طریق تعریف یک متغیر یا شمارش از روی دادهها حساب شوند (شکل ۳.۳ را ببینید). در نهایت، برخی از انواع روشهای نمونه گیری میتوانند به صورت با جایگذاری (با علامت WR) و یا بدون جایگذاری (با علامت WOR) انجام شوند.

مرحلهی بعد انتخاب حجم نمونه است که میتواند بر اساس تعداد (Counts) و یا نسبت (Proportions) تعیین شود. شکل ۳.۴ پنجرهی Sample Size را در راهنمای قدم به قدم انتخاب یک طرح نمونه گیری، نشان می دهد. اگر گزینهی Counts انتخاب شود، یک عدد صحیح مثبت به عنوان حجم نمونه در جعبهی Value: می دهد. اگر گزینهی Counts انتخاب شود، یک عدد صحیح مثبت به عنوان حجم نمونه در جعبهی counts وارد خواهد شد و در صورتی که گزینهی Propertions انتخاب شود، باید یک عدد بین صفر تا یک به عنوان حجم نمونه در جعبهی counts وارد خواهد شد و در صورتی که گزینهی Propertions انتخاب شود، باید یک عدد بین صفر تا یک به عنوان حجم نمونه در نظر گرفته شود. تنها در حالتی که روش نمونه گیری با جایگذاری است، مقدار value میتواند هر مقدار مثبتی باشد. اگر روش نمونه گیری با طبقهبندی باشد و حجم نمونه در داخل طبقات یکسان نباشد، میتوان عبارت solut میتوان عبارت counts میتوان را فعال کرده و حجم نمونه در داخل طبقات را از طریق میتوان عبار شدهی عال شدهی asult شدهی نمود.

🖉 Inite [] Inite]	Units:	Counts
Graduation Date (gra		Counts
College [college]		Proportions
Starting Salary [salary]	•	Value: The size value applie to each stratum.
	С	Unequal values for strata:
	c	Read values from variable:
	•	
	bdie	imum Mavimum

شكل ۲.۴- پنجرهى Sampling Wizard: Sample Size

مرحلهی بعدی مربوط به ذخیره کردن برخی از متغیرها پس از اجرای روش نمونه گیری است. این متغیرها عبارتند از: حجم جامعه، نسبت نمونهای، حجم نمونه و وزن نمونهای(عکس احتمال ورود به نمونه) . برخی دیگر از متغیرها مانند احتمال ورود به نمونه و وزنهای تجمعی به صورت خودکار تولید و ذخیره می شوند.

پس از عبور از مرحلهی Output Variables، به آخرین مرحله که خلاصهای از مشخصات طرح نمونه گیری تعریف شده را به نمایش می گذارد، میرسیم. به عنوان مثال، شکل ۳.۵ مشخصات یک طرح نمونه گیری خوشهای با طبقهبندی را با استفاده از روش تصادفی ساده و بدون جایگذاری نشان میدهد. در این قسمت می توان طرح را به یک طرح نمونه گیری دو مرحله ای نیز تبدیل کرد. این عمل به وسیله ی انتخاب گزینه ی Ves, add stage 2 now انجام می شود. پس از انتخاب این گزینه و کلیک کردن روی گزینه ی Next، مجدداً پنجره های مربوط به معرفی مشخصات مرحله ی دوم طرح نمونه گیری فعال می شوند. ذکر این نکته ضروری است که بسته ی نرمافزاری SPSS طرحه ای نمونه گیری چند مرحله ای را تا مرحله ی سوم پیش می برد. به عبارت دیگر قابلیت اجرای طرحه ای آمار گیری تا سه مرحله را دارد.

پس از اتمام معرفی مشخصات طرح نمونه گیری (یک یا چند مرحلهای)، انتخاب نمونهها از واحدهای جامعه صورت خواهد گرفت. این عمل به وسیلهی کلیک کردن روی گزینهی Finish انجام می شود. در صورتی که روش نمونه گیری با جایگذاری (WR) انتخاب شده باشد، ابتدا باید نام یک فایل خروجی داده را برای نمونه هایی که با جایگذاری انتخاب می شوند مشخص کرده و بعداً اقدام به انتخاب نمونه کرد. این عمل از طریق گزینهی External File در پنجرهی Output File از راهنمای قدم به قدم انتخاب یک طرح نمونه گیری، امکان پذیر است.

Stage	Label	Strata	Clusters	Size	Method
1	(None)	gender	college	(Varies By Stratum)	Simple Random Sampling (WOR)
	20.00	- 10 2 10 - 201			
File: C:\l	Documents	and Settings\jo	lani\Desktop\SF	°SS\sampleda	ta.csplan
File: C:\\ Do you	Documents want to add	and Settings\jo stage 2?	lani\Desktop\SF	PSS\sampleda	ta.csplan
File: C:\ Do you O Ye	Documents want to add s, add stage	and Settings\jo stage 2? : 2 now	lani\Desktop\SF	®SS∖sampleda ● No, do not	ta.csplan add another stage now

شكل ۲.۵- پنجرەى Sampling Wizard: Plan Summary

۲,۳ فایلهای Analysis Plan

با فرض این که دادههای نمونه گیری شده در دسترس است، مشخصات طرح نمونه گیری متناظر با آن باید معلوم شده تا بر اساس آن تحلیلها انجام شود. این مشخصات در یک فایل به نام Analysis Plan ذخیره می شود. راهنمای قدم به قدم Analysis Plan به ما کمک می کند تا مراحل ساخت و یا اصلاح این فایل به سادگی انجام شود. ساختن این گونه فایل ها وقتی که امکان دسترسی به طرح نمونهای دادههای نمونه گیری شده وجود ندارد، بسیار مفید است.

برای ساختن یک فایل Analysis Plan، از نوار ابزار منوی Analysis و از زیر منوی فعال شده عبارت Complex Samples را انتخاب میکنیم. گزینهی Prepare for Analysis، امکانات ایجاد و اصلاح در یک فایل Analysis Plan را فراهم میکند. شکل۳.۶ راهنمای قدم به قدم را برای این فایل نشان میدهد. گزینه کی Create a plan file (گزینه کی پیش فرض) را انتخاب کرده و یک نام برای ذخیره ی مشخصات طرح نمونه گیری معرفی می کنیم و از طریق کلید Next به مرحله ی بعد می رویم. در این مرحله (مرحله ی معرفی متغیرها) طبقات و یا خوشه ها را معرفی می کنیم. همچنین متغیری که نشان دهنده ی وزن نمونه ای می باشد را در جعبه ی Sample Weight قرار می دهیم. بودن معرفی کردن وزن های نمونه ای امکان ساختن یک فایل Analysis Plan وجود ندارد و بدین ترتیب نمی توان هیچ گونه تحلیلی انجام داد. پس وزن های نمونه ای که همان عکس احتمال ورود به نمونه می باشند، برای تحلیل داده های نمونه گیری شده بسیار ضروری هستند. در این مرحله می توان با کلیک کردن روی کلید Finish می توان طرح مورد نظر را ذخیره نمود و یا با انتخاب کلید Next به مرحله ی بعدی رفت.



شکل ۸.۶- پنجرهی راهنمای جادویی برای Analysis Plan Files

در مرحلهی بعد چگونگی انتخاب واحدهای نمونهای با توجه به این که تکرار واحدها، مشخص میشوند. به عبارت دیگر، روش انتخاب، با جایگذاری و یا بدون جایگذاری است. همچنین در روش نمونه گیری بدون جایگذاری، آیا احتمالهای انتخاب یکسان هستند یا متفاوت؟ برای این مرحله سه حالت در اختیار کاربر قرار می گیرد:

- ۱. نمونه گیری با جایگذاری (WR)
- ۲. نمونه گیری بدون جایگذاری و با احتمال یکسان (Equal WOR)
- ۳. نمونه گیری بدون جایگذاری و با احتمال متغیر (Unequal WOR)

با انتخاب کلید Next به مرحلهی بعد رفته که شامل اطلاعات مربوط به احتمال ورود و یا حجم جامعه است. شکل۳.۷ پنجرهی این مرحله را نشان میدهد.

در صورت انتخاب گزینهی Population Size از طریق جعبهی Units (شکل ۳.۷ را ببینید) حجم جامعه را معرفی میکنیم. اندازهی جامعه میتواند ثابت و یا نسبت به طبقات متفاوت باشد. در صورتی که جحم طبقات یکسان نیست گزینهی :Unequal values for strata انتخاب میشود. در صورتی که حجم نمونهها در غالب یک متغیر معرفی شده باشند، آنها را از طریق جعبهی :Read values from variable نشان میدهیم.

🖉 Inite [] Inite]	Units	Population Sizes	-
Gender [gender]		Population Sizes	
Graduation Date Igra		Inclusion Probabilities	
College [college]	(•		
	С	Unequal values for strata:	
		Define	
	C	Read values from variable:	
	•		

شکل ۸.۲- پنجرهی Analysis Preparation Wizard: Size

پس از تعیین حجم جامعه به مرحلهی آخر رفته که شامل خلاصهای از مشخصات طرح نمونه گیری می باشد. در صورتی که روش نمونه گیری دو مرحلهای باشد با انتخاب گزینه ی Yes, add stage 2 now، مشخصات مرحلهی دوم را ثبت می کنیم و در غیر این صورت مشخصات را در فایل مربوط ذخیره می کنیم.

در صورت انتخاب گزینهی Population Size از طریق جعبهی Units (شکل ۳.۷ را ببینید) حجم جامعه را معرفی می کنیم. اندازهی جامعه می تواند ثابت و یا نسبت به طبقات متفاوت باشد. در صورتی که جحم طبقات یکسان نیست گزینهی :Unequal values for strata انتخاب می شود.

پس از تعیین حجم جامعه به مرحلهی آخر رفته که شامل خلاصهای از مشخصات طرح نمونه گیری میباشد. در صورتی که روش نمونه گیری دو مرحلهای باشد با انتخاب گزینهی Yes, add stage 2 now، مشخصات مرحلهی دوم را ثبت می کنیم و در غیر این صورت مشخصات را در فایل مربوط ذخیره می کنیم.

به منظور اجرای تحلیل نمونه گیری، ابتدا باید مشخصات طرح نمونه گیری را معرفی کنیم. این مشخصات از طریق یک فایل (Sampling Plan File یا Sampling Plan File یا معریف میشود. برای ایجاد جداول فراوانی یا آمارههای توصیفی و برخی دیگر از تحلیلهای پیشرفته ی آماری شبیه رگرسیون، می توان از Complex Samples کرد. دستور مای توان از مارههای خلاصه شده مانند میانگین، مجموع، انحراف معیار و ... از طریق کلید Statistics در پنجره ی Descriptive فعال میشود (شکل ۳.۸ را ببینید).

anabies.		Manager	OK
🔗 Units [Units]		Starting Salary [salary]	Paste
🗿 Gender (gender)			Reset
📔 College (college)	4		Cancel
			Help
	Þ	Subpopulations:	

شکل ۳.۸- پنجرهی Complex Samples Descriptives