راهنمای آموزشی SPSS 19

تهیه و تنظیم :

مهدی اکبرزاده

جلد اول

فهرست مطالب

پیشگفتار

با توجه به پیشرفت سریع علم و افزایش درک محققین از رشته خود، علوم مختلف در حال پیوند خوردن با یکدیگرند و محققین امروزی به خوبی می‌دانند که اشراف کامل آنها بر علم خود به تنهایی راه‌گشا نیست و در راه گسترش و تعمیم یک بحث در یک رشته خاص بدون شک باید با علوم دیگر نیز آشنا باشند. یکی از علومی که پیوند محکمی با سایر علوم دارد، علم "آمار" است. به‌طوری که برخی از محققین سرشناس بر این اعتقادند که دانستن علم خود، بدون دانستن علم آمار، کاری از پیش نمی‌برند و همواره شعار آنها این است که علم آمار، علم انجام علوم دیگر است. همان‌طور که رسیدن به مقصدی خاص مستلزم دانستن روش پیمودن مسیر است، طی نمودن مسیر تحقیق نیز بدون دانستن آمار و روش تحقیق امکان‌پذیر نیست.

همواره پس از اتمام مراحل اولیه پژوهش، پژوهشگر با انبوهی از داده‌ها سروکار دارد که می‌خواهد از آنها اطلاعات مورد نظر را استخراج کند. راه استخراج اطلاعات از داده‌های خام را علم آمار به ما می‌آموزد. به عبارتی فرآیند تحلیل آماری کمک می‌کند تا پژوهشگر بتواند از داده های اولیه، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد. اما انجام تحلیل آماری، با توجه به پیچیدگی فرمول‌ها و قضایای آماری، به‌طور دستی قابل انجام نیست و محقق باید از نرم‌افزارهای آماری کمک بگیرد، تا این کار را انجام دهد. لذا می‌توان گفت در به انجام رساندن صحیح یک پژوهش، روش تحقیق، علم آمار و نرم‌افزار آماری سه ضلع مثلث تحقیق را تشکیل می‌دهند که بدون دانستن حتی یکی از آنها، انجام پژوهش امکان‌پذیر نیست.

امروزه انواع نرم افزارهای آماری موجودند که قادرند تحلیل‌های آماری را انجام دهند، برخی از مهمترین این نرم افزارها عبارتند از : SPSS، S-plus، MINITAB، R، SAS و . . . . که البته بنا بر قابلیت‌های آنها، هر یک در رشته‌ای بخصوص رایج شده‌اند و هر کاربر با توجه به کاربرد آمار در رشته تحصیلی خود یکی را انتخاب می کند. برای مثال محققین در رشته کشاورزی از برنامه SAS بیشتر بهره می برند و یا افرادی که در رشته های مهندسی صنایع در حال تحقیق و پژوهش هستند از برنامه MINITAB برای انجام پروژه های صنعتی خود استفاده می کنند، یا محققین آماری از برنامه‌های R و S-plus استفاده می کنند.

اما چرا SPSS ؟

SPSS (Statistical Package for Social Sciences) یکی از تواناترین و جامع ترین نرم افزارهای آماری است که با توجه به سادگی کار و سایر خصوصیات بارز آن، امروزه پرکاربردترین نرم افزار آماری محسوب می‌شود. همان‌طور که از نام این نرم افزار مشخص است، از ابتدا این نرم افزار برای رشته علوم اجتماعی طراحی شده بود، اما به مرور زمان و با توجه به نیازهای مطرح شده در سایر علوم این نرم افزار کامل و

کامل‌تر شده است و در حال حاضر کلیه محققین در رشته های علوم اجتماعی، علـوم پزشـکی، علـوم تربیتـی، روان‌شناسی،کشاوزی و . . . در حال استفاده از این نرم افزار هستند و در این لحظه که شما تصمیم به فراگیـری این نرم افزار آماری گرفته اید، ویرایش ۱۹ این نرم افزار در کشورمان به‌طور کامل موجود بوده و ما نیز با همین ویرایش کار خواهیم کرد.

امید است تا با مطالعه این جزوه قسمتی از مشکلات شما محقق گرامی از میان برداشته شود. البته جای هیچ شکی نیست که این اثر خالی از اشکال نیست و نویسنده هرگونه اصلاح و پیشنهاد را با کمال میل پذیراست. لذا خالی از لطف نیست که موارد را از طریق ایمیل با نویسنده در میان بگذارید.

Email: akbarzadeh.ms@gmail.com

با تشکر و احترام

مهدی اکبرزاده

تابستان ۱۳۹۰

فصل اول

مروری بر نرم‌افزار SPSS

در این بخش قسمت‌های متفاوت نرم‌افزار SPSS را معرفی خواهیم کرد.

۱ـ شروع کار با SPSS

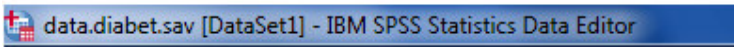
پس از نصب نرم‌افزار SPSS، برای اجرای آن به منوی Start بروید و بر آیکون IBM SPSS Statistics 19 کلیک کنید. در این قسمت به معرفی محیط نرم‌افزار و اعمال مقدماتی در آن، می‌پردازیم.

۱ـ۱ـ معرفی محیط نرم‌افزار

پس از اجرای نرم‌افزار، با پنجره‌ای تحت عنوان IBM SPSS Statistics 19 مواجه خواهید شد، که با انتخاب دکمه Cancel وارد پنجره اصلی SPSS خواهید شد. در این بخش به معرفی اجزای این محیط خواهیم پرداخت.

- **نوار عنوان:**

برنامه SPSS نیز مانند سایر نرم افزارهای تحت Windows دارای نوار عنوان است که در آن نام فایل SPSS را مشاهده می کنید. برای مثال فایل داده های زیر به نام data.diabet ذخیره شده‌است:



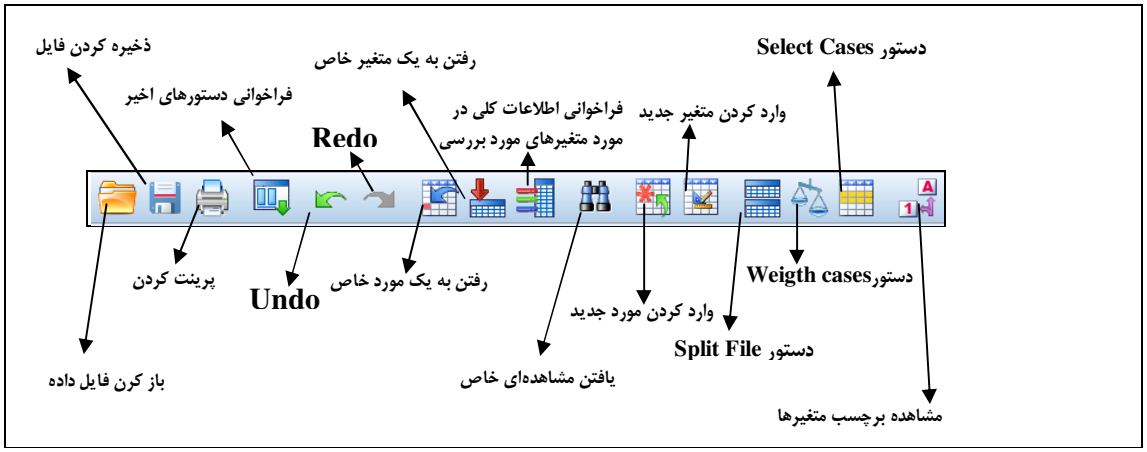
توجه داشته باشید که پسوند فایل‌های داده در این نرم‌افزار sav. و پسوند فایل‌های خروجی spv. است.

نوار منو:

همان‌طور که در شکل زیر مشاهده می کنید این نوار شامل منوهای زیادی می باشد که به مرور زمان با آنها آشنا می‌شوید (مهم‌ترین منوهای این نوار، منوهای Analyze و Graphs می باشند) :

نوار ابزار:

این نوار در زیر نوار منو قرار دارد. بر روی این نوار دکمه هایی تعبیه شده است که قابلیت دسترسی به برخی اعمال پرکاربرد را آسان‌تر می کند. به مرور زمان با دکمه های این نوار آشنا خواهید شد. در شکل زیر به اجمال دکمه های این نوار معرفی شده اند:



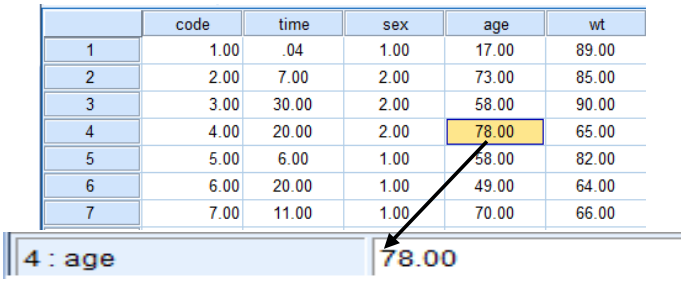
صفحه Data View

این صفحه مخصوص وارد کردن داده هاست. ما با این صفحه به دفعات زیاد بر خواهیم خورد. در صفحه بعد شکل این صفحه را آورده‌ایم.

	code	time	sex	age	wt	ht
1	1.00	.04	1.00	17.00	89.00	1.85
2	2.00	7.00	2.00	73.00	85.00	1.60
3	3.00	30.00	2.00	58.00	90.00	1.65
4	4.00	20.00	2.00	78.00	65.00	1.60
5	5.00	6.00	1.00	58.00	82.00	1.62
6	6.00	20.00	1.00	49.00	64.00	1.65
7	7.00	11.00	1.00	70.00	66.00	1.64
8	8.00	8.00	2.00	59.00	69.00	1.53
9	9.00	11.00	2.00	62.00	60.00	1.65
10	10.00	5.00	2.00	44.00	65.00	1.62
11	11.00	1.00	2.00	61.00	60.00	1.55

نوار آدرس:

در این نوار می‌توانید آدرس سلول فعال در صفحهٔ Data View را مشاهده کنید. برای مثال در شکل زیر سلول فعال، "مشاهده چهارم از متغیر age است که مشاهدهٔ ۷۸ برای آن ثبت شده‌است". در نوار آدرس هم می‌توان این بیان را مشاهده کرد:



	code	time	sex	age	wt
1	1.00	.04	1.00	17.00	89.00
2	2.00	7.00	2.00	73.00	85.00
3	3.00	30.00	2.00	58.00	90.00
4	4.00	20.00	2.00	78.00	65.00
5	5.00	6.00	1.00	58.00	82.00
6	6.00	20.00	1.00	49.00	64.00
7	7.00	11.00	1.00	70.00	66.00

4 : age 78.00

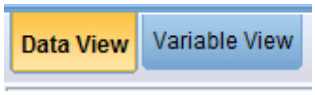
صفحه Variable View

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	code	Numeric	8	2		None	None	8	Right	Scale	Input
2	time	Numeric	8	2		None	None	8	Center	Ordinal	Input
3	sex	Numeric	8	2		None	None	8	Center	Nominal	Input
4	age	Numeric	8	2		None	None	8	Center	Ordinal	Input
5	wt	Numeric	8	2		None	None	8	Center	Ordinal	Input
6	ht	Numeric	8	2		None	None	8	Center	Ordinal	Input
7	BMI_1	Numeric	8	2		None	None	10	Right	Scale	Input
8	smoking	Numeric	8	2		None	None	8	Center	Nominal	Input
9	alcohol	Numeric	8	2		None	None	8	Center	Nominal	Input
10	HLP	Numeric	8	2		None	None	8	Center	Nominal	Input
11	HTN	Numeric	8	2		None	None	8	Center	Nominal	Input

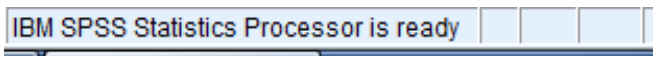
دکمه تعویض صفحه:

با استفاده از این دکمه، و با انتخاب گزینه موردنظر، می‌توانید از صفحه Data view به صفحه Variable view و بالعکس بروید.



نوار وضعیت :

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



نحوه باز کردن و ذخیره کردن :

در این نرم‌افزار نیز مانند سایر نرم‌افزارهای تحت سیستم عامل ویندوز می‌توان از منوی File و گزینه‌های Open و Save(یا Save AS…)، به ترتیب، برای ذخیره و باز کردن فایل انجام داد.

تمرین:

به عنوان تمرین فایل داده Employee Data.sav را از داده‌های آموزشی[1] نرم افزار SPSS که در رایانه شما ذخیره است را باز کرده و قسمت‌های مختلف آن را بررسی کنید. سپس آن را در مسیر دلخواه دیگری ذخیره کنید.

۱-۲- وارد کردن داده ها در SPSS :

داده‌های یک تحقیق شامل اطلاعات حاصل از اندازه‌گیری چند متغیر از تعدادی واحد آزمایشی خواهد بود. در اینجا متغیرها می‌توانند مواردی مانند وزن، قد، فشارخون، گروه خونی، تعداد تخت، نوع ژن و . . . باشند. همچنین یک واحد آزمایشی می‌تواند یک انسان، یک بیمارستان، یک حیوان یا . . . باشد. برای وارد کردن مشاهدات مربوط به متغیرهای واحدهای آزمایشی، درنظر داشته باشید که در صفحه Data view ستون‌ها، نماینده متغیرهای مورد بحث، سطرها نماینده واحدهای نمونه‌ای و در هر سطر مقادیر مشاهدات مربوطه ثبت خواهد شد.

فرم اطلاعاتی زیر برای بررسی وضعیت تغذیه ای بیماران دیابتی است. می‌خواهیم اطلاعات حاصل از این فرم را در نرم‌افزار SPSS وارد کنیم. توجه کنید که این فرم شامل ۱۱ متغیر است(چرا؟).

فرم جمع آوری اطلاعات

بررسی وضعیت تغذیه ای بیماران دیابتی

کد بیمار:

مدت بیماری از زمان شروع علائم(بر حسب ماه):

جنسیت: مرد☐ زن☐

سن: وزن: قد:

مصرف سیگار: بلی☐ خیر☐

مصرف الکل: بلی☐ خیر☐

میزان کلسترول: میزان چربی:

نوع رژیم دیابتیک: تزریق انسولین ☐ قرصهای خوراکی ☐

در شکل زیر اطلاعات مربوط به تعدادی بیمار، در نرم‌افزار SPSS وارد شده‌است. در نهایت در این بخش می‌خواهیم شما نیز بتوانید داده‌ها را اینگونه وارد SPSS کنید.

۱- این داده‌های آموزشی پس از نصب نرم‌افزار SPSS، در درایو C ذخیره می‌شوند.

	Code	Dur	Sex	Age	Weiht	Hight	Smoking	Alcohol	Chlo	Fat	Regime
1	1.00	5.00	Male	17.00	89.00	1.85	No	No	65.00	19.00	Insulin Injection
2	2.00	7.00	Female	73.00	85.00	1.60	No	No	46.00	33.00	Tablet
3	3.00	30.00	Female	58.00	90.00	1.65	No	No	48.00	28.00	Insulin Injection
4	4.00	20.00	Female	78.00	65.00	1.60	No	No	42.00	38.00	Insulin Injection
5	5.00	6.00	Male	58.00	82.00	1.62	No	No	44.00	28.00	Tablet
6	6.00	20.00	Male	49.00	64.00	1.65	No	No	54.00	22.00	Insulin Injection
7	7.00	11.00	Male	70.00	66.00	1.64	No	No	53.00	23.00	Insulin Injection
8	8.00	8.00	Female	59.00	69.00	1.53	No	No	40.00	33.00	Tablet
9	9.00	11.00	Female	62.00	60.00	1.65	No	No	33.00	34.00	Tablet
10	10.00	5.00	Female	44.00	65.00	1.62	No	No	46.00	31.00	Insulin Injection

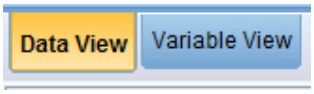
به‌طور کلی وارد کردن داده‌ها در SPSS شامل دو گام اساسی است:

۱- تعریف متغیرها ۲- ورود داده‌ها

ورود داده‌ها در SPSS، در قسمت Data View بسیار راحت است. با وارد کردن هر مشاهده مربوط به یک متغیر و فشردن دکمه Enter می‌توان داده‌های هر متغیر را وارد کرد. قدم اول که تعریف متغیرهاست و پایه و اساس گام بعدی، یعنی ورود داده‌ها، است.

تعریف متغیرها در SPSS:

برای تعریف متغیرها ابتدا روی عبارت Variable view در نوار زیر کلیک کنید و وارد صفحه Variable view شوید.



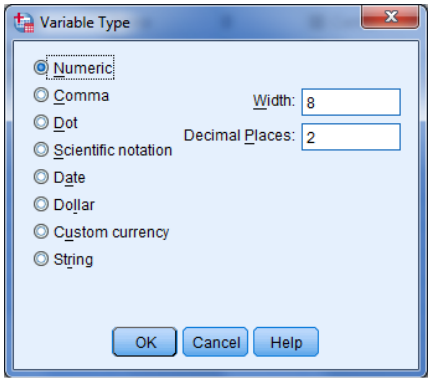
با رفتن به این پنجره، ستون‌های صفحه نرم‌افزار مشخصه‌های متغیر موردنظر را نشان می‌دهد، که تعریف یک متغیر به این نرم‌افزار با معرفی این مشخصه‌ها از آن متغیر به نرم‌افزار امکان‌پذیر است. در این بخش به معرفی این مشخصه‌ها می‌پردازیم.

مشخصه Name:

اولین و ابتدایی‌ترین مشخصه یک متغیر که باید به نرم‌افزار معرفی شود، نام آن متغیر است. نام متغیر مورد نظر را باید طوری انتخاب کنید که در مراجعات بعدی به فایل داده‌ها دچار سردرگمی نشویم. همچنین حواسمان باشد که دادن نام به متغیر در SPSS شامل یکسری محدودیت‌هایی است و در صورت رعایت نکردن آنها با پیام خطایی از سوی نرم‌افزار مواجه خواهید شد. برای مثال نام متغیر شامل کاراکتر "فاصله" نمی‌شود، یا با "عدد" شروع نمی‌شود. حال سعی کنید برای ۱۱ متغیر مربوط به فرم اطلاعاتی بیماران دیابتی نامی انتخاب کنید و آنها را وارد SPSS کنید.

مشخصه Type:

این مشخصه نوع کاراکتری را مشخص می‌کند که کاربر می‌خواهد از آن برای ورود داده‌های مربوط به متغیر موردنظر در صفحه Data view استفاده کند. در بین انواع حالت‌های این مشخصه، دو نوع Numeric و String بیشتر مورد استفاده است. نوع Numeric برای کاراکترهای عددی و String برای کاراکترهای حرفی استفاده می‌شود.



برای ورود داده‌های مربوط به متغیرهای عددی(مانند وزن و قد) تنها از کاراکترهای عددی استفاده می‌کنیم. لذا مشخصه Type برای این نوع متغیرها ناگزیر Numeric خواهد بود. برای ورود داده‌های متغیرهای کیفی مانند "نام بیمار" تنها می‌توان از کاراکترهای حرفی استفاده کرد. لذا مشخصه Type برای این نوع متغیرها ناگزیر String خواهد بود. اما برای ورود داده‌های متغیرهای کیفی مانند "جنسیت" هم می‌توان از کاراکترهای عددی و هم از کاراکترهای حرفی استفاده کرد. لذا مشخصه Type برای این نوع متغیرها هم می‌تواند Numeric باشد و هم می‌تواند String باشد. بدین صورت که اگر Type را Numeric انتخاب کردید، باید از کاراکترهای عددی (برای مثال) ۱ و ۲ استفاده می‌کنید و در تعریف مشخصه‌های بعدی این متغیر، نحوه این کدبندی (کد ۱ برای مردان و کد ۲ برای زنان) را به نرم‌افزار معرفی می‌کنید، و اگر Type آن‌را String انتخاب کردید، از کاراکترهای حرفی (برای مثال) male و female استفاده خواهید کرد.

حال سعی کنید مشخصه Type را بری متغیرهایی که در مرحله قبل تعریف کردید، مشخص کنید.

مشخصه Width و Decimals:

اگر نوع متغیر عددی باشد در این دو ستون تعداد کل ارقامِ عدد (Width) و تعداد رقم‌های اعشار (Decimals) را می‌توان تعیین کرد.

مشخصه Label:

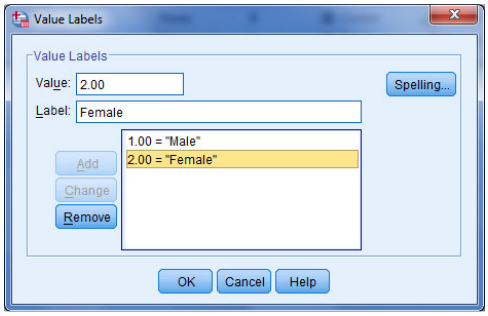
همان‌طور که گفتیم، دادن نام متغیرها در SPSS دارای محدودیت است و در صورتی که مایل باشیم می‌توانیم در این قسمت یک عنوان یا Label برای متغیر قرار بنویسیم. به عنوان مثال اگر متغیر مورد بررسی "طول مدت بیماری" باشد، می‌توانیم در قسمت Label تایپ کنیم: Duration of disease

توجه داشته باشید که اگر برای متغیر، عنوان تعریف شود در خروجی‌های SPSS به جای نام متغیر این عنوان نمایش داده خواهد شد در غیر این‌صورت نام متغیر به کار خواهد رفت.

مشخصه Values:

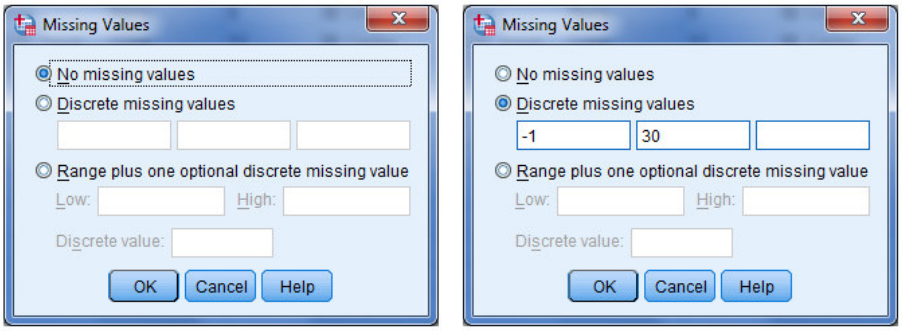
همان‌طور که گفتیم، اما برای ورود داده‌های متغیرهای کیفی مانند "جنسیت" هم می‌توان از کاراکترهای عددی و هم از کاراکترهای حرفی استفاده کرد و در صورتی که Type را Numeric انتخاب کردید، باید از کاراکترهای عددی (برای مثال) ۱ و ۲ استفاده می‌کنید. در این حالت برای تعریف نحوه این کدبندی (کد ۱ برای مردان و کد ۲ برای زنان) به نرم‌افزار از این خصوصیت استفاده می‌کنیم.

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

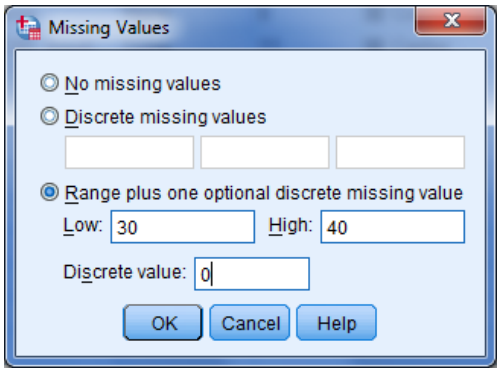


مشخصه Missing:

در این قسمت نحوه معرفی داده های گمشده به SPSS تعیین می‌شود. انتخاب گزینه اول به معنی عدم تعریف داده گمشده است. با انتخاب گزینه دوم میتوان چند مقدار مشخص را برای معرفی یک مقدار گمشده تعیین کرد. به عنوان مثال اگر متغیر مورد بررسی تعداد واحد باشد می‌توان از مقادیری که خارج از دامنه تغییرات‌اند چند مقدار برای معرفی مقادیر گمشده تعیین کرد(مثلاً ۳۰ یا ۱-).



با انتخاب گزینه سوم می‌توان یک بازه عددی را برای مشخص‌کردن مقادیر گمشده تعیین کرد. مثلاً اگر متغیر مورد بررسی "مدت زمان بستری بیمار" باشد می‌توان بازه ۳۰ تا ۴۰ ماه و عدد صفر را برای تعیین مقادیر گمشده تعیین کرد:



مشخصه Columns:

اندازه ستونی که محل وارد کردن داده‌های متغیر مورد بررسی است را در این قسمت وارد می‌کنیم.

مشخصه Align:

تعیین می‌کند که داده‌ها در سمت راست، چپ یا وسط سلول‌های صفحه Data view قرارگیرند.

مشخصه Measure:

در این قسمت مقیاس سنجش متغیرمورد نظر را تعیین می‌کنیم، و همان‌طور که گفتیم این مقیاس‌های سنجش عبارتند از فاصله‌ای، رتبه‌ای و اسمی.

مشخصه Role

توسط این مشخصه نقش متغیر موردنظر را در روش‌های آمار تحلیلی مشخص می‌کنیم. نقش یک متغیر می‌تواند به صورت زیر تعیین شود:

۱- نقش Input : برای متغیرهایی مانند پیشگو، مستقل یا توضیحی مورد استفاده قرار می‌گیرد.

۲- نقش Target: برای متغیرهای وابسته مدل استفاده میرشود.

۳- نقش Both: برای متغیرهایی که هم نقش Input و هم نقش Output را در مدل می‌گیرند.

۴- نقش None: برای متغیرهایی که نقش خاصی در مدل آماری ندارند.

۵- نقش Portion: برای متغیرهایی مورد استفاده قرار می‌گیرد که داده‌های نمونه را به مجموعه‌های آموزشی، آزمودنی و اعتبار سنجی، تقسیم می‌کند.

۶- نقش Split: برای متغیرهایی است که نقش هماهنگ‌کننده بین نرم‌افزارهای تحت SPSS را دارند. این متغیر را هیچگاه نمی‌توان در دستور Split file استفاده کرد.

توجه: به‌طور پیش‌فرض نقش تمام متغیرها Input است و تعیین نقش اصلی متغیر اثری بر نوع دستورهای SPSS نخواهد داشت و تنها در نمایش متغیرها در کادرهای گفتگوی نرم‌افزار تاثیر خواهد گذاشت.

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

تمرین[۲]:

۱- فرم اطلاعاتی بیماران دیابتی را در نظر بگیرید. تمام مشخصه‌های مربوط به هر یک از متغیرهای این فرم را در پنجره Variable View وارد کنید. سپس داده‌های مربوط به آن‌را در صفحه Data view وارد کنید.

۲- قسمتی از یک پرسش‌نامه تحقیقاتی در زیر آمده است. پنجره Variable view را برای متغیرهای مربوط به این پرسش‌نامه کامل کنید.

شماره پرسشنامه: . . .

مشخصات فردی

محل سکونت (مناطق بیستگانه شهر تهران): ☐

سال و محل تولد: ☐

وضعیت تاهل: مجرد ☐ متاهل ☐ مطلقه ☐ بیوه(شوهر فوت کرده) ☐

عنوان دقیق شغل: ☐

میزان دقیق سنوات خدمت، در شغل نامبرده: ☐ سال

میزان دقیق درآمدماهیانه شما: ☐ تومان

وضعیت استخدامی: ۱- رسمی ☐ ۲- پیمانی ☐ ۳- قراردادی ☐ ۴- شرکتی ☐

میزان تحصیلات: زیردیپلم ☐ دیپلم ☐ فوق دیپلم ☐ لیسانس ☐

فوق لیسانس ☐ دکترا ☐

در صورت داشتن مدرک تحصیلی دیپلم و بالاتر از آن،عنوان دقیق رشته تحصیلی

۲- توجه داشته باشید که برای حل برخی از تمرینات مستلزم مطالعه فصل‌های بعدی است.

۳- در هر یک از موارد زیر داده‌های داده شده را با تعریف صحیح متغیرها به SPSS وارد کنید.

الف) می خواهند مطالعه‌ای برای تأثیر نسبی دو نوع داروی موثر در افزایش خواب انجام دهند. به شش نفر که سرماخوردگی دارند در شب اول داروی A و در شب دوم داروی B داده می شود و میزان ساعات خواب آنها در هر شب ثبت می گردد . داده ها عبارتند از :

	۱	۲	۳	۴	۵	۶
داروی A	۴/۸	۴/۱	۵/۸	۴/۹	۵/۳	۷/۴
داروی B	۳/۹	۴/۲	۵/۰	۴/۹	۵/۴	۷/۱

ب) یکی از جنبه های مطالعه در اختلاف‌های جنسیت شامل مطالعه نحوه بازی میمون‌ها در خلال سال اول زندگی است. شش میمون نر و شش میمون ماده در گروه هایی از چهار خانواده در طی چندین جلسه مطالعه ده دقیقه‌ای مشاهده شدند. میانگین کل تعداد دفعاتی که هر میمون ، بازی با همسال دیگر را شروع کرده است ثبت می شود:

نرها	۳/۶۴	۳/۱۱	۳/۸۰	۳/۵۸	۴/۵۵	۳/۹۲
ماده ها	۱/۹۱	۲/۰۶	۱/۷۸	۲/۰۰	۱/۳۰	۲/۳۲

ج) اندازه های ریخت شناسی یک نوع خاص فسیلی که از حفاری در چهار محل مختلف به دست آمده است ، داده های زیر را به دست می دهند :

محل			
۱	۲	۳	۴
۱/۳۸	۱/۴۹	۳/۱۲	۱/۳۱
۱/۴۲	۱/۳۲	۲/۱۹	۱/۴۶
۱/۵۹	۱/۰۱	۲/۷۶	۱/۸۶

د) در پرتاب ۶۰ بار یک تاس مشاهده می شود که روهای مختلف تاس به صورت زیرظاهر می گردد. جدول فراوانی مربوط به این داده‌ها را رسم کنید.

روهای مختلف	فراوانی
۱	۱۵
۲	۷
۳	۴
۴	۱۱
۵	۶
۶	۱۷
جمع	۶۰

و) به منظور بررسی وجود یا عدم وجود رابطهٔ بین نوع شوک و زنده ماندن بیمار، از بیمارانی که به انواع شوک مبتلا می‌شوند، نمونه‌هایی انتخاب و آنها را بر حسب زنده ماندن و یا مردن در جدول زیر مرتب کرده‌اند.

نوع شوک	نتیجه	
	زنده	مرده
کاهش حجم خون	۷	۸
قلبی	۱۱	۱۱
عصبی	۱۰	۶
عفونی	۹	۷
غدد داخلی	۳	۵

داده‌های این جدول را در SPSS وارد کرده و جدول توافقی مربوط به آن‌را رسم کنید.

فصل دوم
آمار توصیفی در SPSS

۲-۱ مقدمه

در اصطلاح عامیانه آمار به معنای ثبت و نمایش اطلاعات عددی در مورد یک موضوع مثلاً ثبت و نمایش تعداد بیکاران، تعداد تصادفات رانندگی، میزان محصولات کشاورزی، میزان صدور نفت، جمعیت شهر تهران و غیره می باشد. ولی علم آمار امروزه دارای مفهومی بسیار وسیعتر از این کاربرد عامیانه است. مفاهیم عامیانه آمار زیر مجموعه ای از آمار مصطلح بین آمار دانان است. از نقطه نظر علمی، آمار به مجموعه روشهایی برای جمع آوری تنظیم و خلاصه کردن داده های عددی و غیر عددی و انجام استنباط و نتیجه گیری بوسیله تجزیه و تحلیل آنها، اطلاق می شود.

با بیان دیگر می توان گفت که آمار **عبارت است از هنر و علم جمع آوری، تعبیر، تجزیه و تحلیل داده ها و استخراج تعمیمهای منطقی در مورد پدیده های تحت بررسی.**

با توجه به تعاریف بالا می‌توان گفت یک فرآیند تحلیل آماری شامل دو بخش عمده است. اولین قدم نمایش دادن و خلاصه کردن داده ها می باشد تا توجه ما روی ویژگی‌های مهم داده ها متمرکز شود و جزئیات غیر ضروری کنار گذاشته شود. اما بخش دوم برای استخراج نکات کلی و استنباط هایی در مورد پدیده تحت مطالعه به کار می‌رود. بخش اول شامل روشهای آمار توصیفی و بخش دوم در برگیرنده روشهای موسوم به آمار استنباطی است. در این فصل به بررسی گام اول تحلیل آماری، یعنی آمار توصیفی، خواهیم پرداخت. آمار توصیفی به آن دسته از روش‌های آماری گفته می‌شودکه به پژوهشگر در طبقه‌بندی، خلاصه کردن، توصیف و تفسیر و برقراری ارتباط از طریق اطلاعات جمع آوری شده کمک می کند. نقش آمار توصیفی در فرآیند تحلیل آماری بسیار مهم و حیاتی است. آمار توصیفی با خلاصه کردن داده ها، ویژگیهای مهم آنرا نمایان می سازد تا ایده های لازم را در ذهن پژوهشگر برای مرحله دوم تحلیل آماری (آمار استنباطی) ایجاد کند.

مراحل اساسی **توصیف** داده ها عبارتست از:

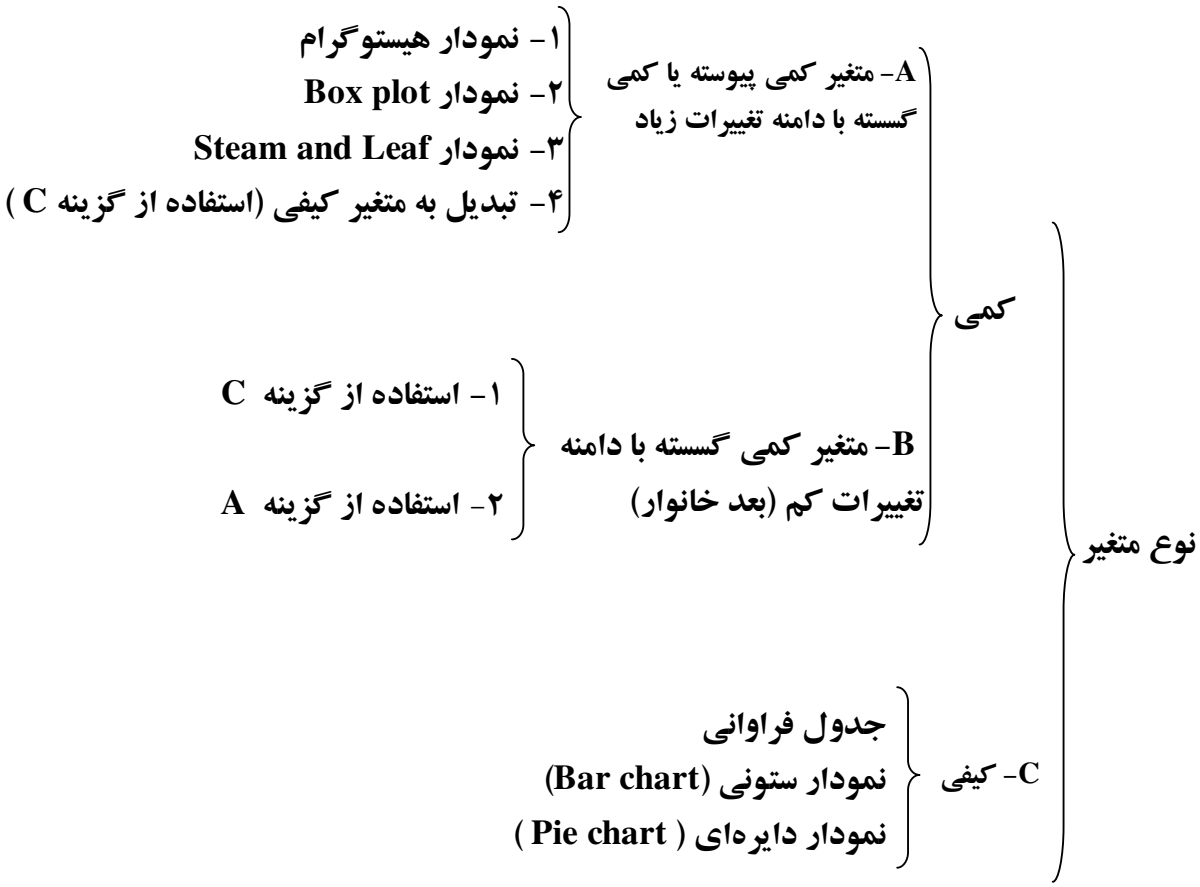
الف) خلاصه کردن و توصیف الگوی کلی

۱) فشرده کردن داده ها در قالب جدول های آماری

۲) نمایش آنها بوسیله نمودار

ب) محاسبه شاخصهای آماری

برای استفاده از مراحل مختلف آمار توصیفی می توان از چارت زیر بهره بگیرید:



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

۲-۲ خلاصه کردن و توصیف الگوی کلی:

یکی از روشهای خلاصه کردن و توصیف داده ها رسم یک نمودار آماری است. نوع نمودار مورد استفاده به نوع داده ها بستگی دارد و بسته به رسته ای بودن یا عددی بودن، نمودارهای مختلفی به کار برده می‌شود.

جداول فراوانی هم بسته به نوع متغیر، متفاوت خواهند بود، لذا مراحل فوق را برای انواع مختلف متغیر، جداگانه بررسی خواهیم کرد.

۲-۲-۱ داده های کیفی (Categorical Data):

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

-جدول فراوانی برای متغیرهای کیفی:

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

مثال: صنعتگری چهار نوع قطعه D,C,B,A را تولید می کند اگر 20 قطعه تولید شده توسط وی به ترتیب زیر باشند.

B,C,C,A,D,C,C,B,D,C,A,C,D,C,B,C,C,B,D,D

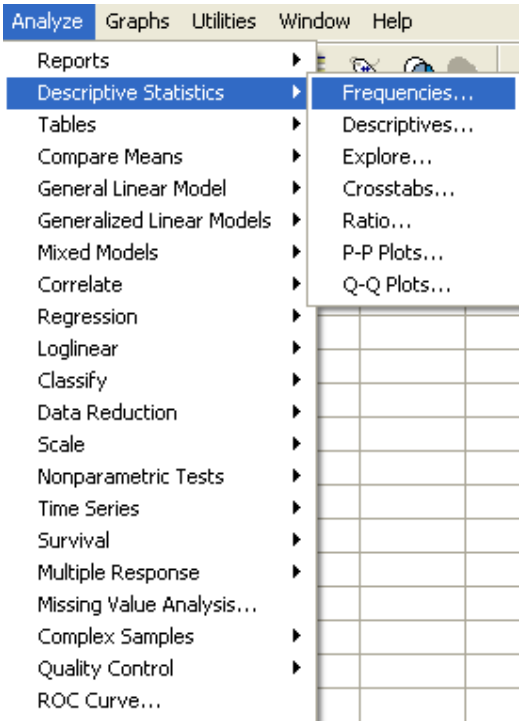
یک جدول فراوانی برای داده های فوق می سازیم. سعی کنید با کمک آن به سئوالات زیر پاسخ دهید:

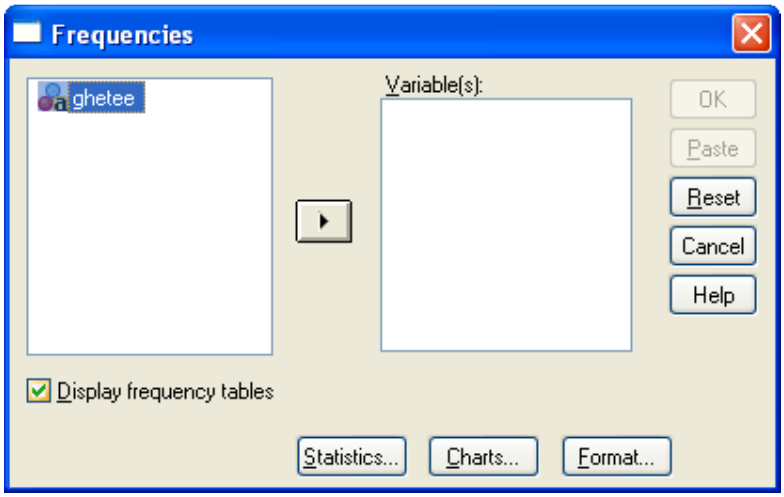
- چند عدد از قطعه C در این روز تولید شده است؟
- قطعات B,A چند درصد از کل تولید روزانه را در بر می گیرند؟

ابتدا داده ها را در محیط SPSS وارد می کنیم:

ghetee
b
c
c
a
d
c
c
b
d
c
a
c
d
c
b
c
c
b
d
d

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





متغیر مربوطه را انتخاب کرده روی دکمه  کلیک کنید

برای رسم جدول فراوانی، در کادر کنار عبارت Display frequency tables تیک بگذارید. و در نهایت دکمه  را کلیک کنید. صفحه جداگانه ای تحت عنوان Output view باز خواهد شد که خروجی های SPSS همواره در آن ظاهر خواهد شد. خروجی این مثال به صورت زیر خواهد بود:

Frequencies

Statistics

ghetee

N	Valid	20
	Missing	0

ghetee

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	a	2	10.0	10.0	10.0
	b	4	20.0	20.0	30.0
	c	9	45.0	45.0	75.0
	d	5	25.0	25.0	100.0
	Total	20	100.0	100.0	

جدول اول تعداد داده های موجود (Valid) و داده های گمشده (Missing) را نشان می‌دهد. و جدول دوم جدول فراوانی متغیر است. ستون اول مقادیر متغیر، ستون دوم فروانی هر مقدار، ستون سوم درصد فراوانی آن مقدار و ستون چهارم هم درصد فراوانی تجمعی می باشد.

از روی این جدول سعی کنید به سوالهای مطرح شده در صورت مثال جواب دهید.

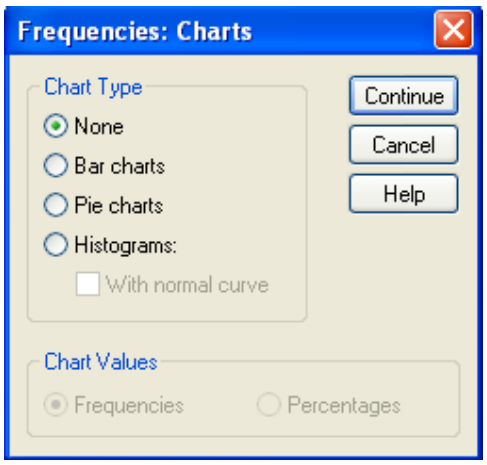
برای متغیرهای کیفی رتبه ای هم مراحل رسم جدول فراوانی به همین صورت خواهد بود.

- نمودارهای آماری برای متغیرهای کیفی:

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

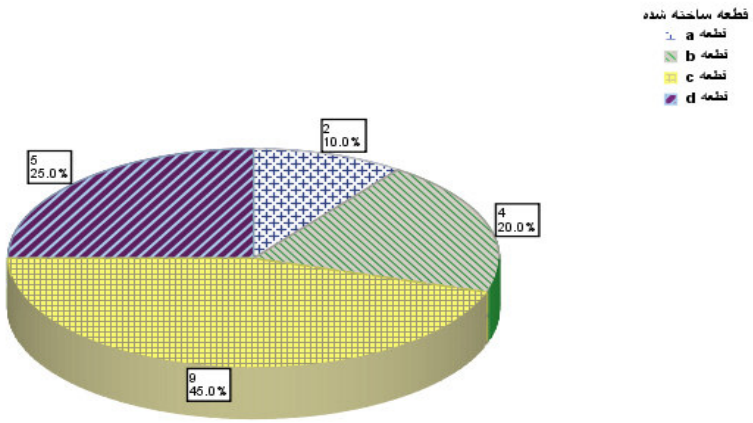
برای رسم این نمودارها می توان از دو راه زیر استفاده کرد:

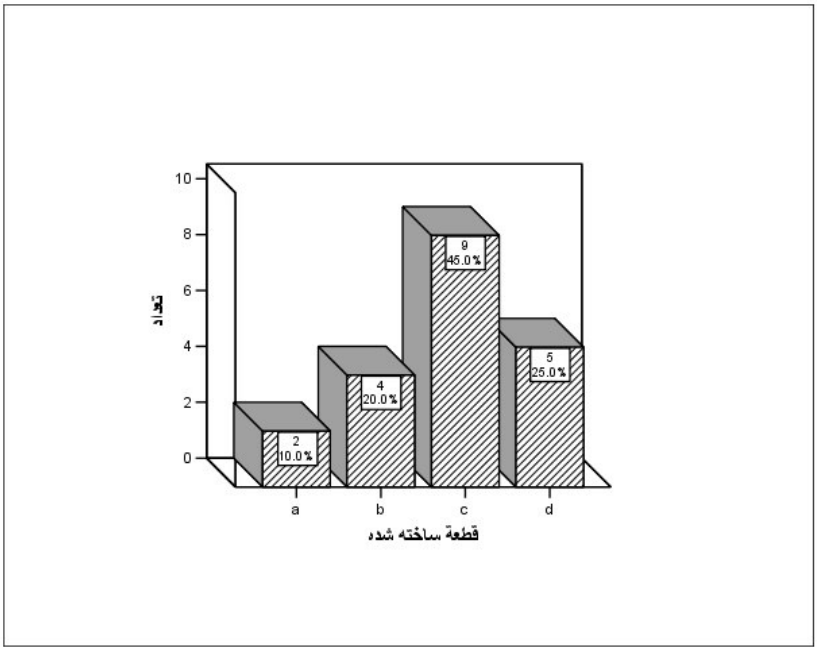
۱. از مسیری که برای رسم جدول فراوانی طی کردیم رفته روی دکمه Charts... کلیک کنید تا کادر گفتگوی زیر باز شود:



هر کدام از نمودارهای pie chart یا Bar chart را که مایل بودید انتخاب کرده و دکمه Continue را کلیک کنید. دکمه Ok را کلیک کنید.

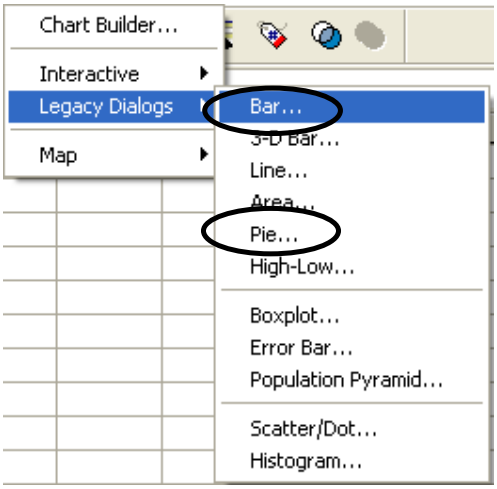
مسیر فوق را برای داده های مثال طی کنید و خروجی را ببینید:



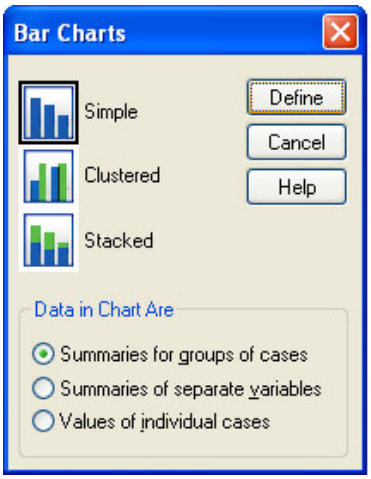


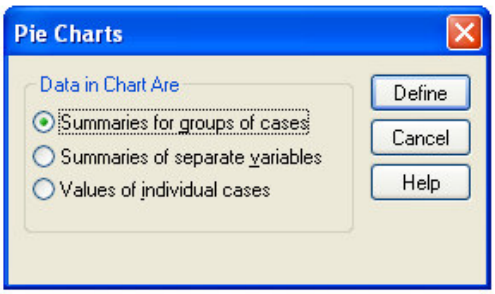
توجه کنید که نمودارهای دایره ای برای متغیرهای کیفی اسمی مناسبترند و نمودارهای میله ای برای متغیرهای کیفی رتبه ای.

۲- راه دوم استفاده از منوی Graph است:



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





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

۲-۲-۲ داده های عددی (Numerical Data)

داده های عددی دو نوع اند: *گسسته* و *پیوسته* مقادیر متغیرهای گسسته اعداد حاصل از شمارش می باشد برای مثال یک خانواده می تواند یک یا دو فرزند داشته باشد اما تعداد فرزندان خانواده نمی تواند عددی ما بین

این دو باشد. و در سمت مقابل، متغیرهای پیوسته فاقد واحدهای تفکیک پذیر می باشند. برای مثال وزن یک متغیر پیوسته است.

- متغیر عددی گسسته:

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

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

۰،۰،۷.۵.۳.۳.۴.۵.۳.۲.۸.۳.۳.۲.۴.۴.۳.۶.۷.۴.۵.۴.۶.۴.۵

۲.۳.۴.۲.۷.۳.۵.۴.۶.۲.۳.۲.۴.۵.۴.۸.۴.۳.۲.۲.۶.۴.۵.۷.۸

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

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

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

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

- متغیر عددی پیوسته:

-تبدیل داده های عددی پیوسته به کیفی

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

مثال: وزنهای ۴۰ قالب کره که به نزدیکترین عدد صحیح گرد شده اند به قرار زیر است:

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

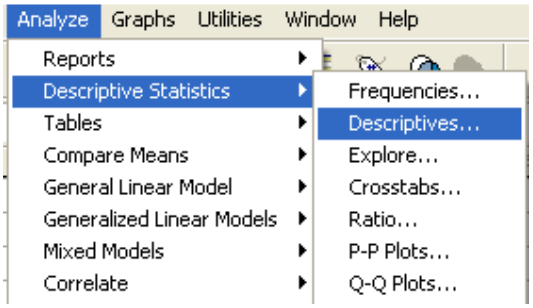
جدول فراوانی داده های فوق را رسم کنید.

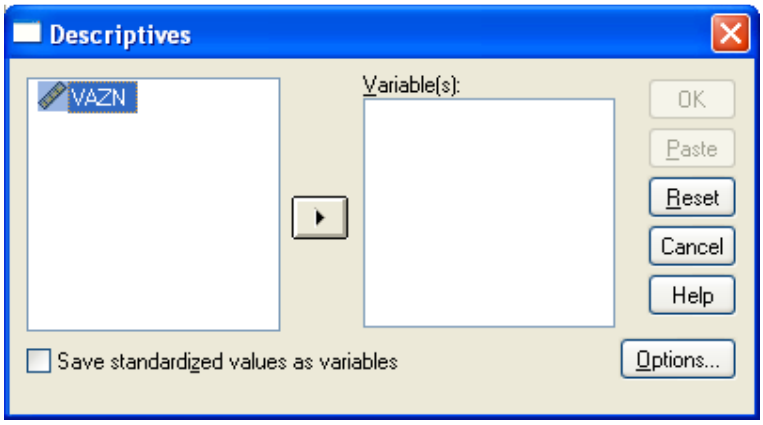
هرگاه داده های ما پیوسته باشد، داده ها را به تعدادی رده با طول مساوی تقسیم می کنیم و در هر رده فراوانی داده ها را می شماریم برای بدست آوردن تعداد رده ها در رسم جدول فراوانی به ترتیب زیر عمل کنید:

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

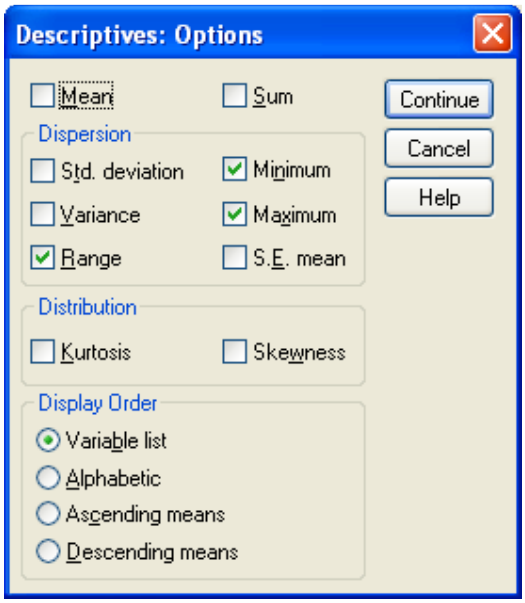
Range = Max - Min

برای محاسبه دامنه تغییرات، کمترین داده ها و بیشترین داده ها از مسیر زیر استفاده می کنیم:





روی ...Options کلیک کنید تا کادر زیر ظاهر شود.گزینه های مورد نظر را تیک بگذارید. تا خروجی به شکل زیر ظاهر شود:



Range, Maximum, Minimum را انتخاب کنید. دکمه Continue و سپس Ok را کلیک کنید.

۲. برای بدست آوردن تعداد رده ها یک قاعده عمومی وجود ندارد و معمولاً تعدادرده ها را بین ۵ تا ۲۵ رده اختیار می کنند. یک قاعده مفید استفاده از دستور استورگس Sturges است:

تعداد طبقات $m=1+3.322 \log (n)$ (n: تعداد کل داده ها)

چون حاصل یک عدد اعشاری خواهد بود. آن را به بزرگترین عدد صحیح گرد می کنیم.

در مثال بالا داریم:

m=1+3.322 log (40) =6.322

پس تعداد طبقات را ۷ می گیریم.

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

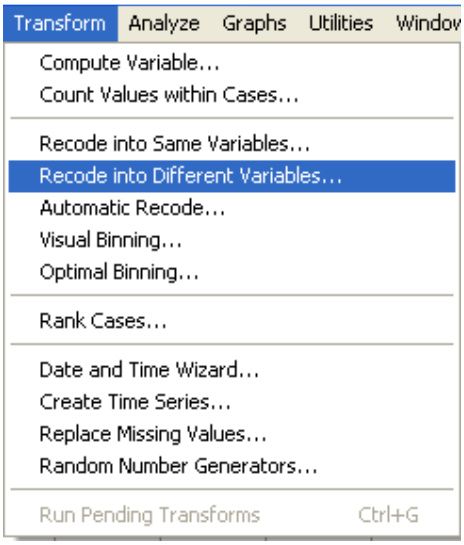
طول طبقه L=Range/m=4.574

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

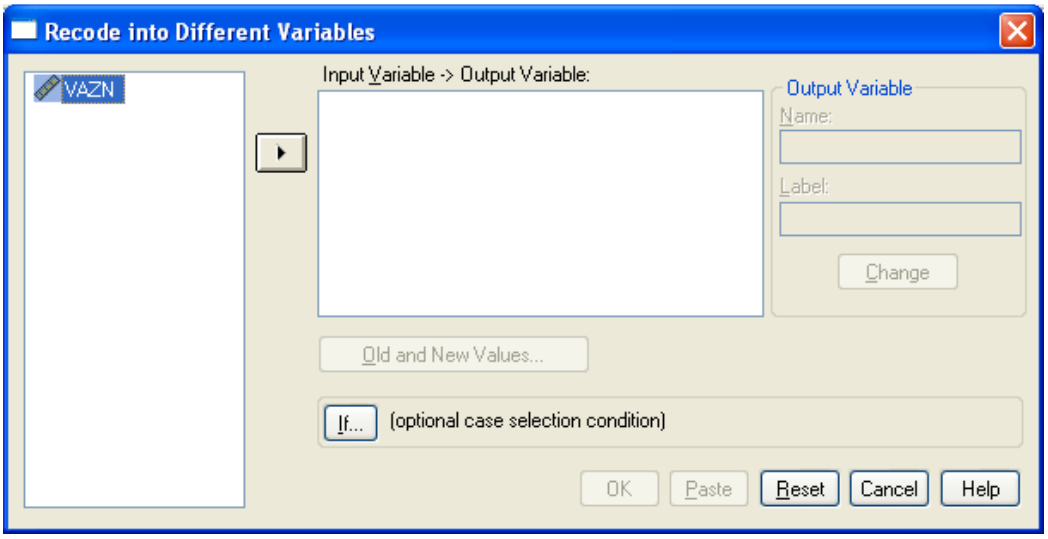
20.5-25.5	→ 1
25.5-30.5	→ 2
30.5-35.5	→ 3
35.5-40.5	→ 4
40.5-45.5	→ 5
45.5-50.5	→ 6
50.5-55.5	→ 7

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

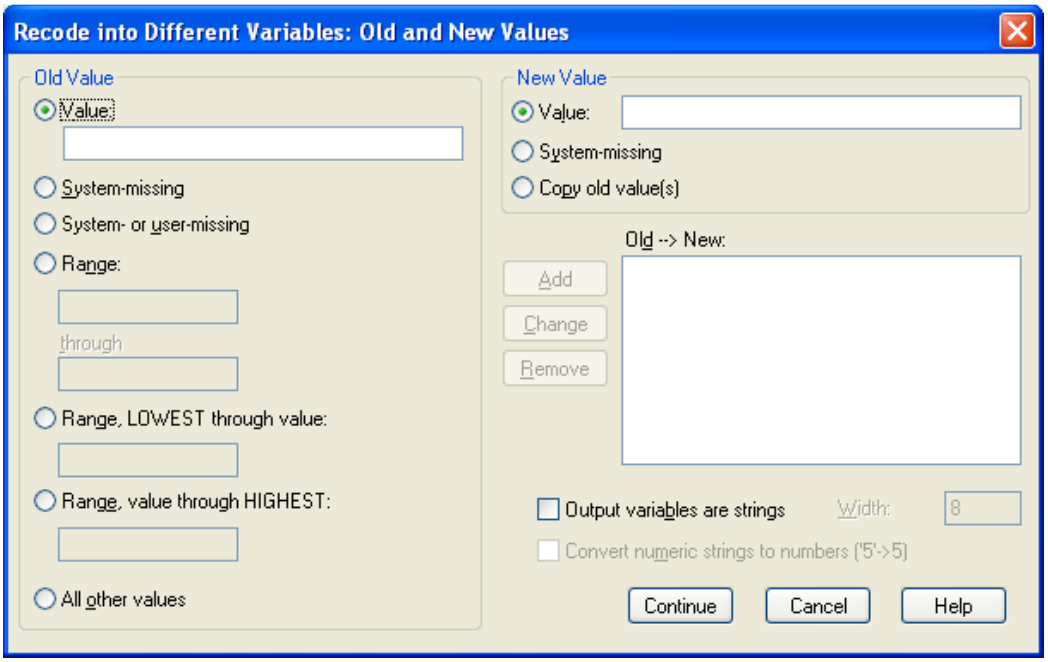
۴. حال متغیر جدیدی تعریف می کنیم که با توجه به واقع شدن داده در یکی از فواصل هفتگانه بالا یکی از مقادیر ۱ تا ۷ را بپذیرد. برای تعریف این متغیر مسیر زیر را طی کنید:



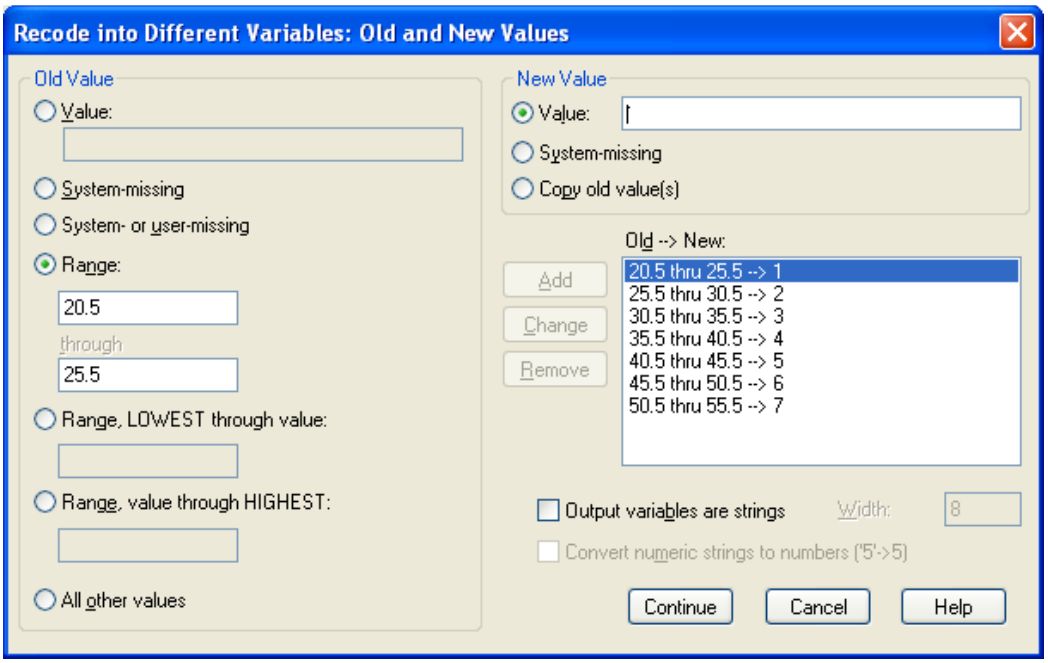
کادر گفتگوی زیر ظاهر می شود:



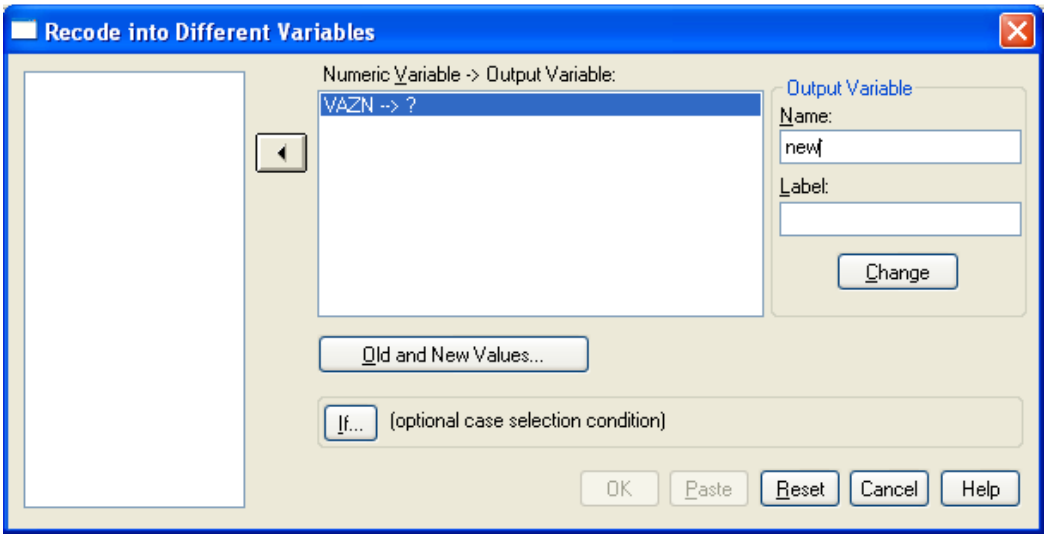
متغیر مورد نظر را انتخاب کرده دکمه  را فشار دهید. سپس روی  کلیک کنید کادر گفتگوی زیر باز می شود که در آن مقادیر متغیر جدید را با توجه به مقادیر متغیر اولیه مشخص می کنیم:



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



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

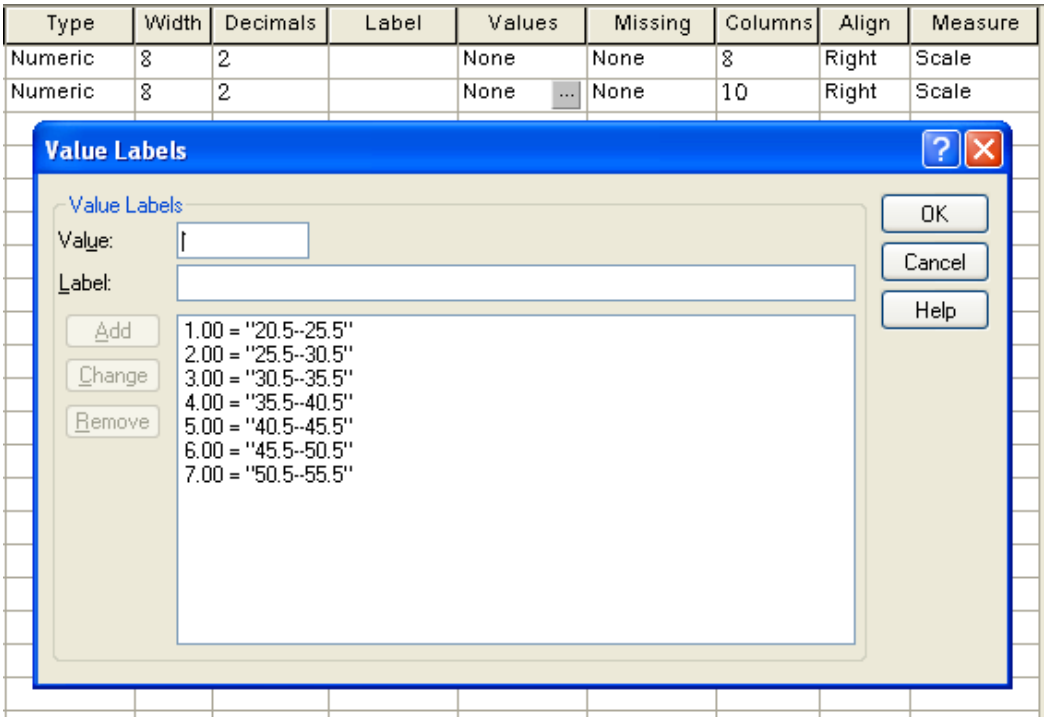


در پایان دکمه ok را کلیک کرده و نتیجه کار را به صورت زیر مشاهده کنید:

	VAZN	new	var
1	52.00	7.00	
2	47.00	6.00	
3	26.00	2.00	
4	31.00	3.00	
5	35.00	3.00	
6	36.00	4.00	
7	29.00	2.00	
8	30.00	2.00	
9	24.00	1.00	
10	38.00	4.00	
11	30.00	2.00	
12	26.00	2.00	
13	47.00	6.00	
14	50.00	6.00	
15	32.00	3.00	

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

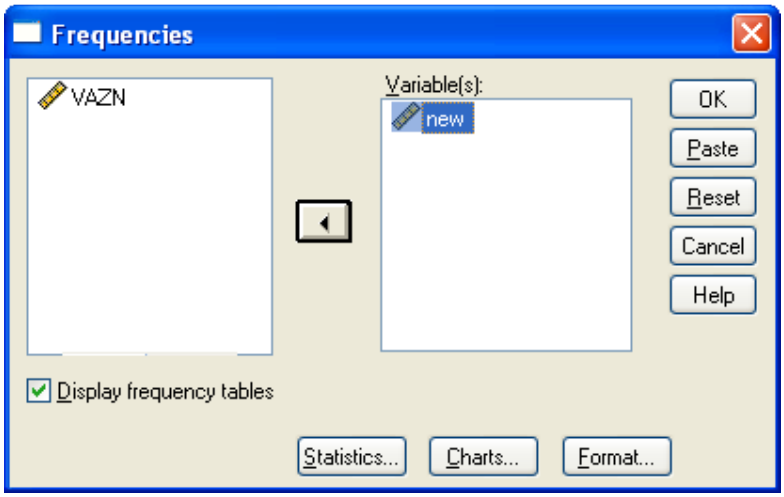
۵. به قسمت variable view بروید و در قسمت value برای مقادیر متغیر جدید، بر چسب‌هایی به صورت زیر تعریف کنید:



به قسمت Data view برگردید.

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

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



جدول فراوانی داده های مثال وزن قالبهای کره به صورت زیر خواهد بود:

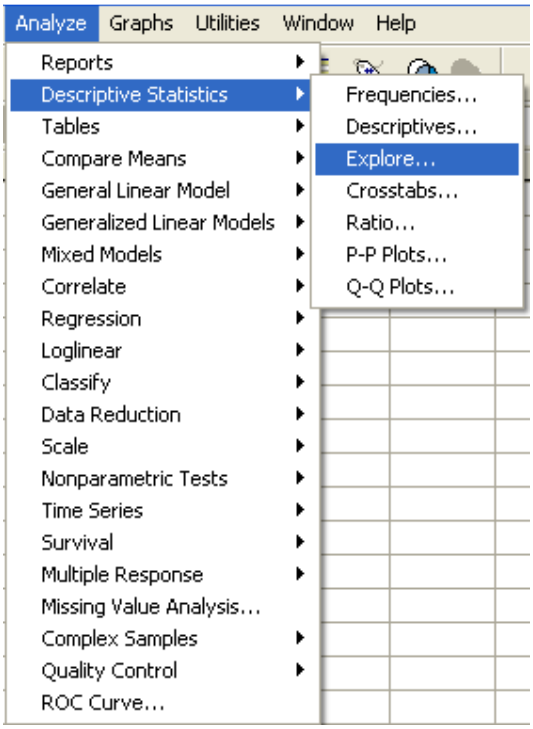
new

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

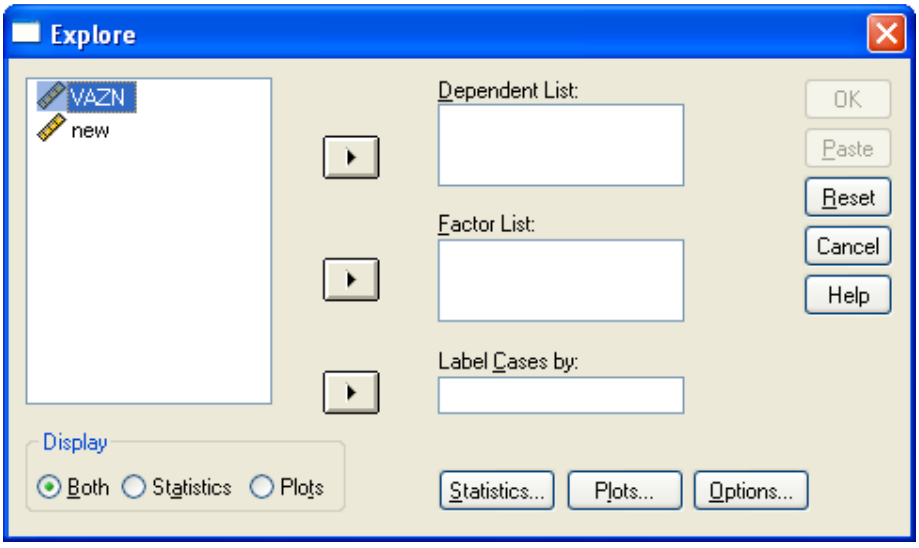
- رسم نمودار برای داده های کمی پیوسته:

برای داده های پیوسته نمودارهای مختلفی به کار می روند که هر کدام کاربردهای خاص خود را دارد.

ارجح ترین این نمودارها Histogram است. برای رسم هیستوگرام داده های پیوسته مسیر زیر را طی کنید:

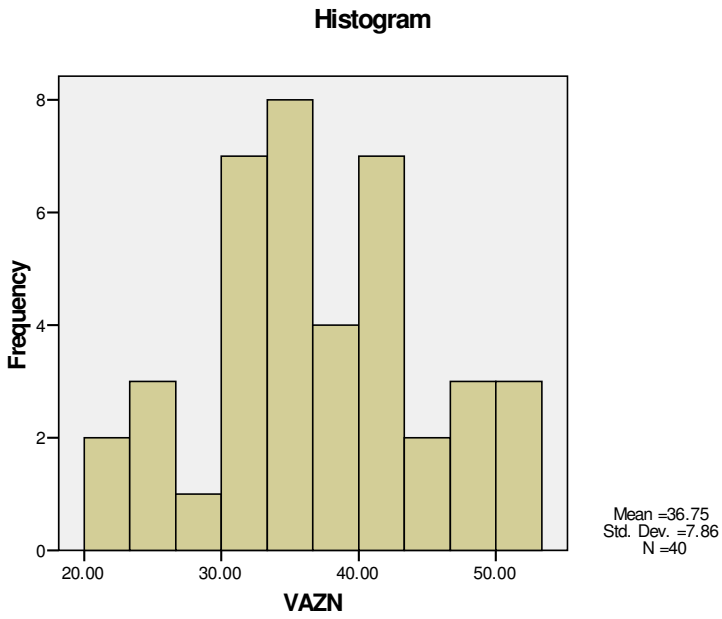


کادر گفتگوی زیر ظاهر خواهد شد:



در قسمت Display (سمت چپ پایین کادر بالا) plots را انتخاب کنید. روی دکمه ...Plot کلیک کنید در کادر گفتگو ظاهر شده Histogram را انتخاب کنید. Continue را کلیک کرده سپس ok را کلیک کنید.

هیستوگرام داده های مثال قبل به صورت زیر خواهد بود:



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

۲-۲-۳ محاسبه شاخصهای آماری:

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

شاخصهای آماری به دو دسته تقسیم میشوند: شاخصهائی که گرایش به مرکز یا مرکزیت داده ها را اندازه میگیرد(شاخصهای مرکزی) و شاخصهائی که برای اندازه گیری تغییر پذیری داده ها به کار میرود(شاخصهای پراکندگی.)

-شاخصهای مرکزی:

شاخصهای مرکزی مهم عبارتند از:

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

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

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

تعریف چندکها هم معادل میانه است،چندک مرتبهp ،مقداری است که تقریباً p۱۰۰ درصد داده ها از آن کمتریا مساوی آن و (1-p)۱۰۰ درصد داده ها از آن بیشرند. ساده ترین نوع چندکها ، چارکها (Quartiles) و دهکها هستند.

- چارک اول: مقداری است که یک چهارم داده ها از آن کمتر یا مساوی با آن هستند.
- چارک دوم: معادل میانه است.
- چارک سوم: مقداری است که سه چهارم داده ها از آن کمتر یا مساوی با آن هستند.

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

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

برای رفع مشکل داده های پرت، انواع دیگری از میانگین تعریف میشود که اثر اینگونه داده ها را کاهش میدهد.

شاخصهای پراکندگی:

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

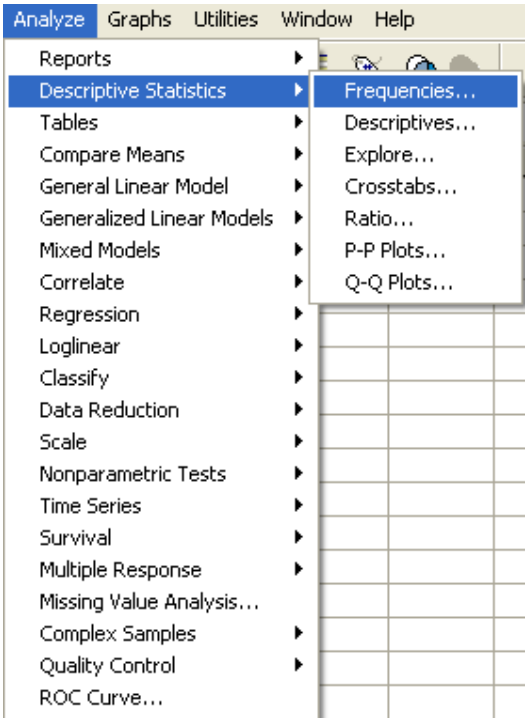
دامنه تغییرات(Range): تفاضل بزرگترین و کوچکترین داده را دامنه تغییرات می نامند.

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

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

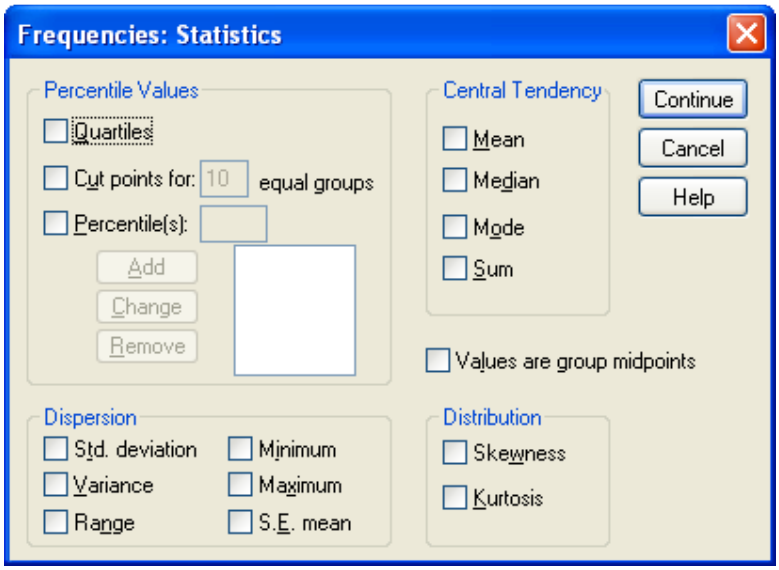
انحراف استاندارد میانگین (Standard Error of Mean):جذر حاصل تقسیم واریانس بر تعداد داده ها را انحراف استاندارد میانگین گویند.

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

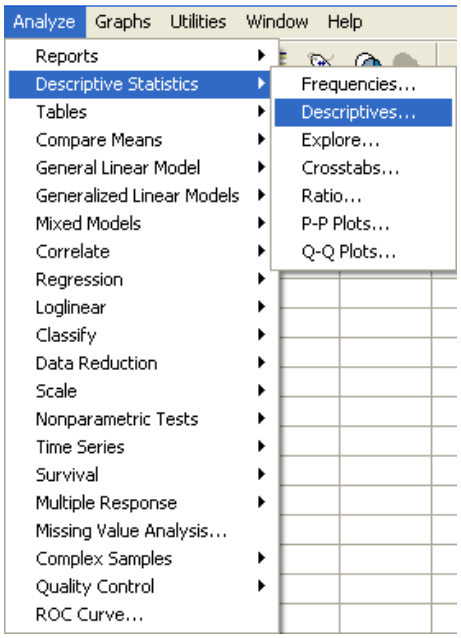


روی دکمه Statistics... کلیک کنید.کادر زیر باز خواهد شد. شاخصهای مرکزی را میتوانید از کادر Central Tendency و شاخصهای پراکندگی را از کادر Dispersion انتخاب کنید.

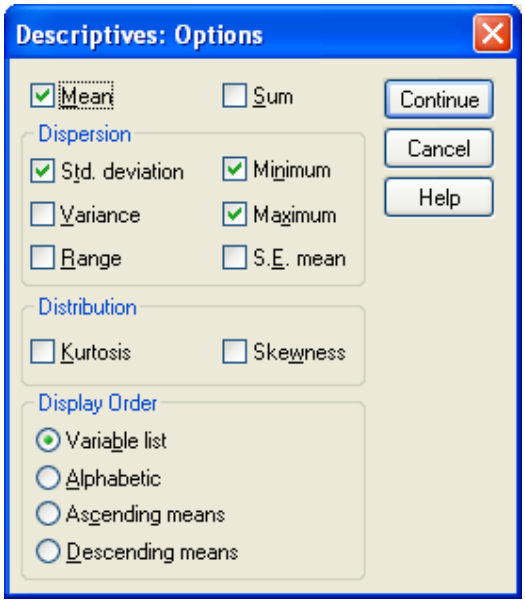
پس از انتخاب شاخصهای مورد نظر دکمه Continue و سپس OK را کلیک کنید.



و یا میتوان از مسیر زیر استفاده کرد:

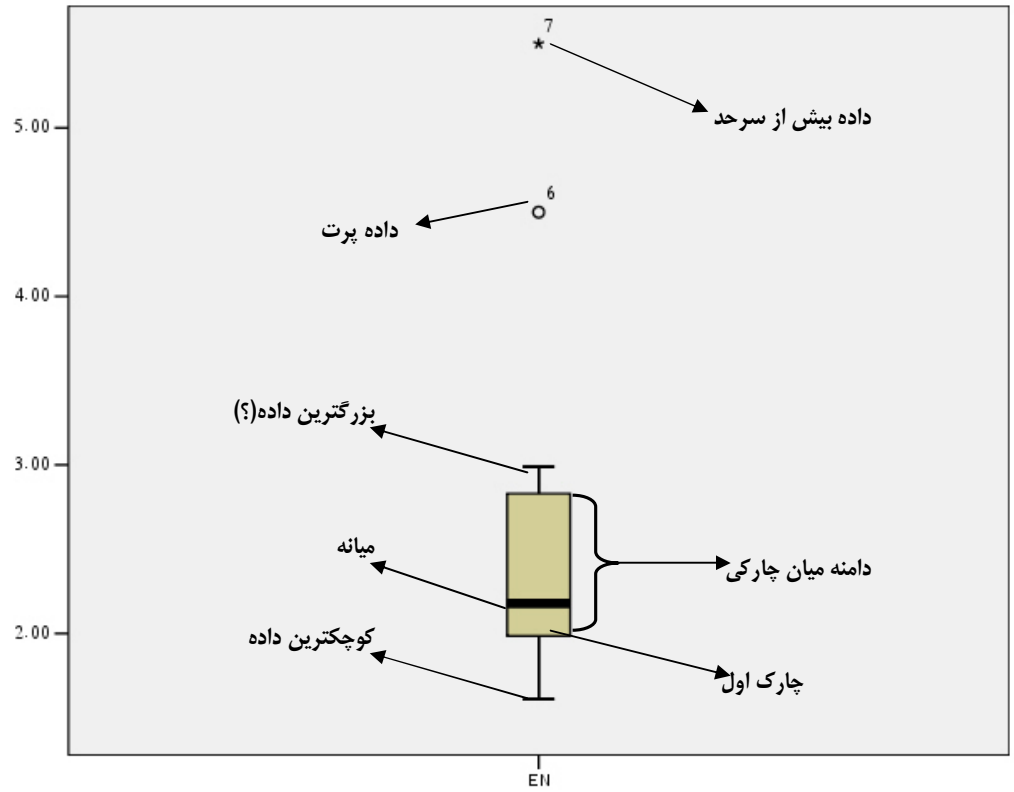


روی دکمه ...Option کلیک کنید و شاخصهای موردنظرتان را انتخاب کنید:



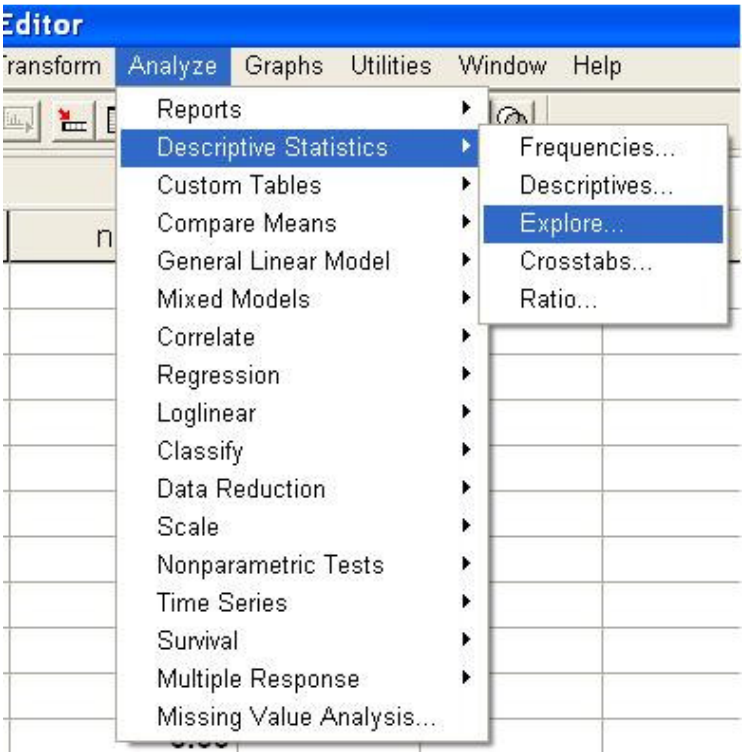
نمودار Box plot :

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



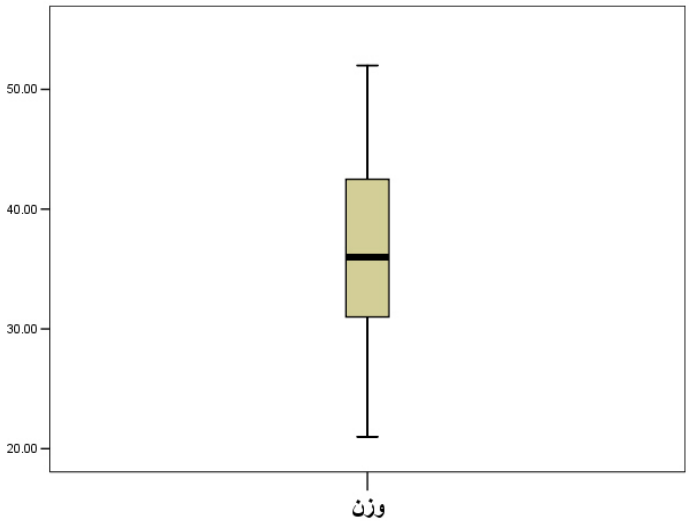
فاصله میان چارک اول و چارک سوم را دامنه میان چارکی می گویند.اگر داده ای بیش از ۱.۵ برابر دامنه میان چارکی از چارک اول یا چارک سوم فاصله داشته باشد، داده پرت محسوب میشود و به صورت یک نقطه خارج از نمودار نشان داده میشود و نمودار بدون احتساب آن رسم می شود.

برای رسم Box plot مسیر زیر را طی کنید:



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

نمودار جعبه ای داده های مثال وزن قالبهای کره ای در زیر رسم شده است. نحوه تقارن داده ها، میزان پراکندگی و وجود یا عدم وجود داده های پرت را بخوبی میتوان در نمودار مشاهده کرد:



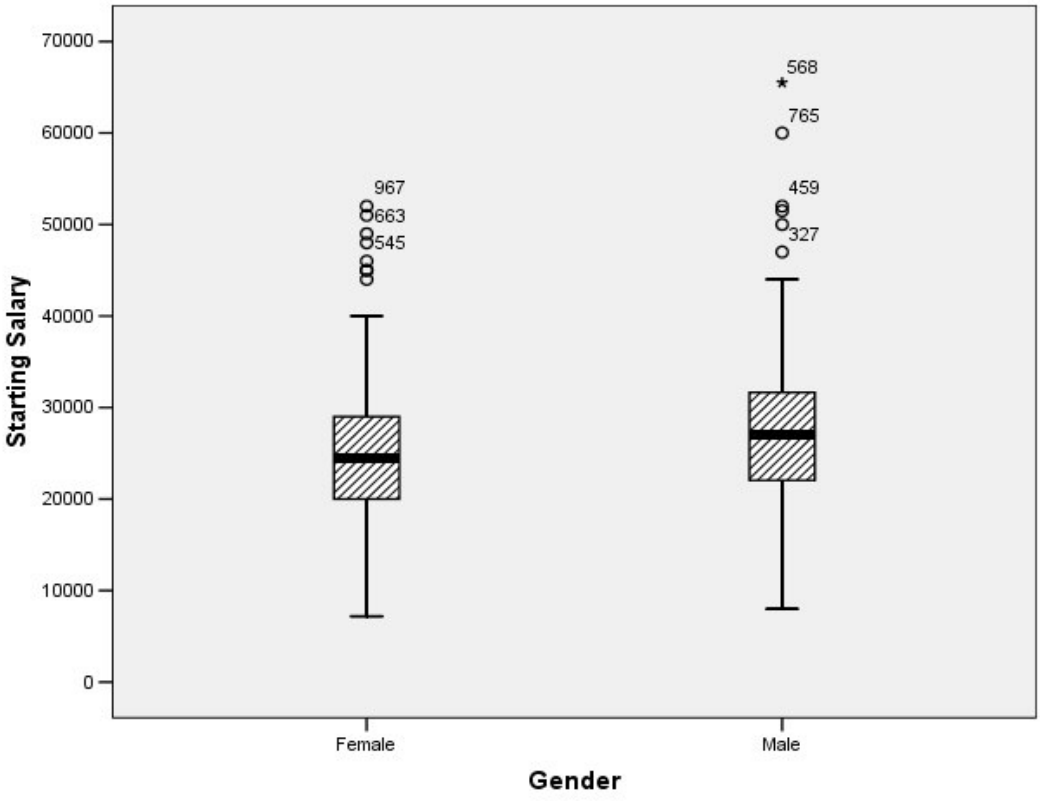
۲-۲-۴ چند نمودار مهم و کاربردی دیگر:

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

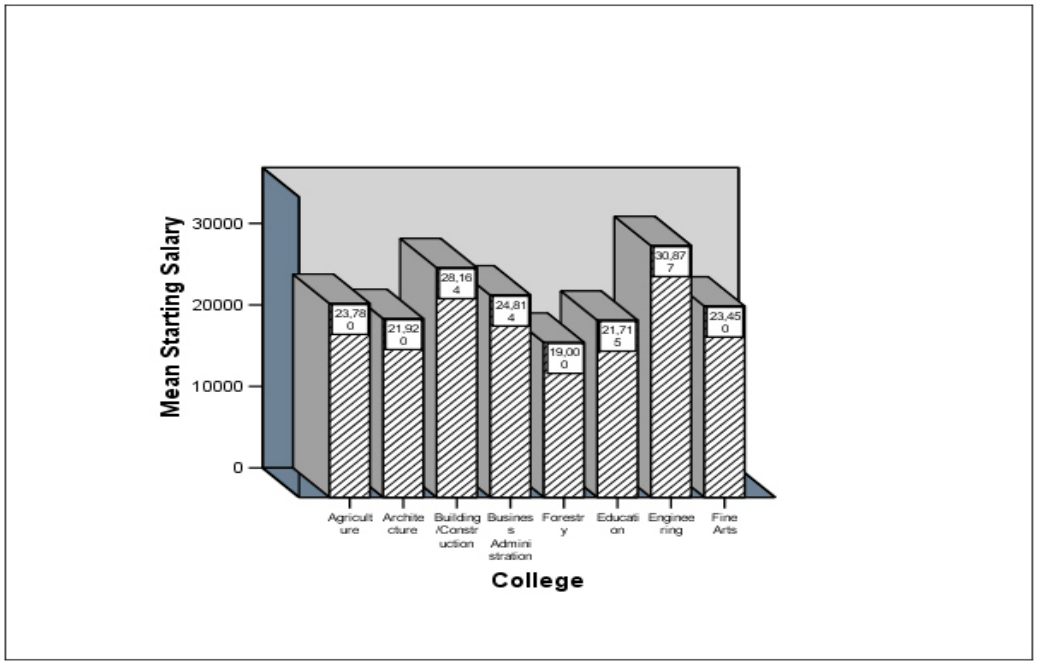
۱) زمانی که تنها یک متغیر عددی و یک متغیر رسته‌ای داشته باشیم:

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

الف) نمودار Box plot (برای دو جامعه):

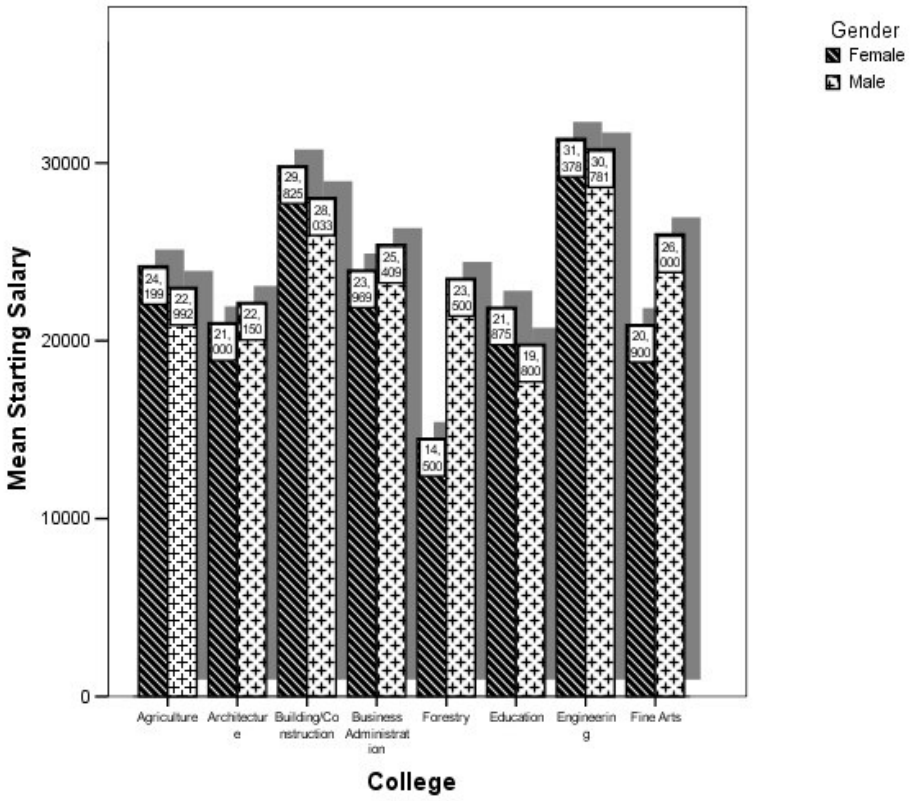


ب) نمودار ستونی (برای مقایسهٔ دو جامعه) :

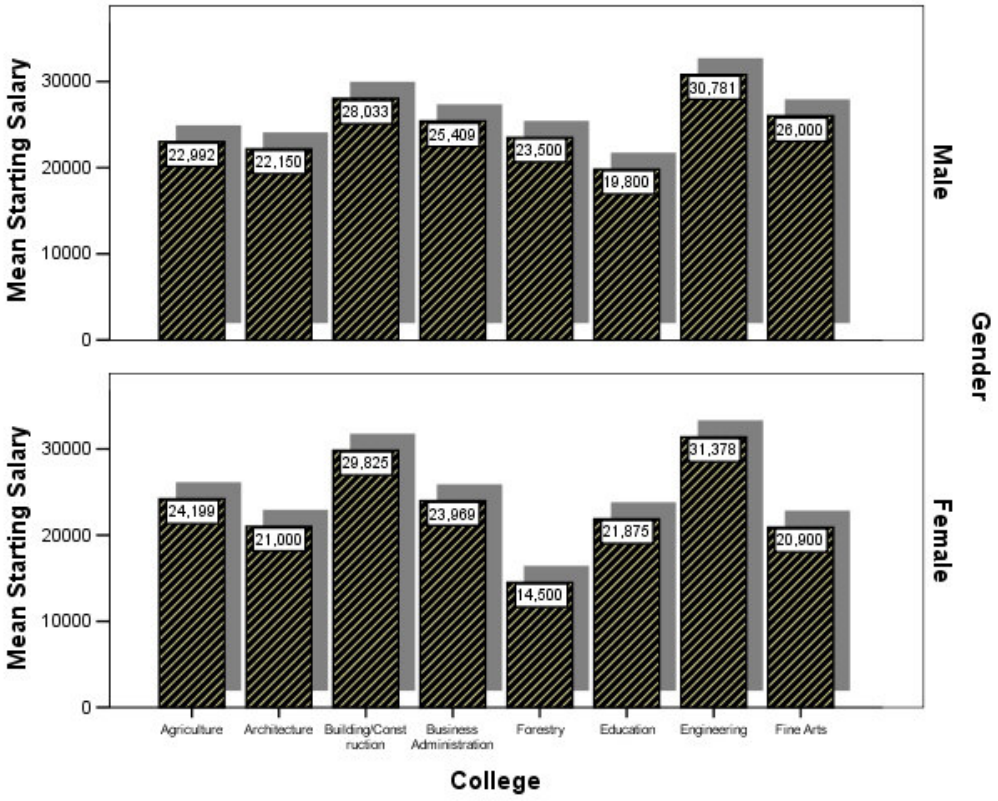


۲) **زمانی که تنها یک متغیر عددی و دو متغیر رسته‌ای داشته باشیم:**

الف) فرض کنید می‌خواهیم حقوق زنان و مردان را در هر دانشکده با هم مقایسه کنیم. در این حالت از **نمودار ستونی خوشه‌ای** استفاده می‌کنیم:



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



فصل سوم
دستورهایی برای دست‌کاری داده‌ها در SPSS

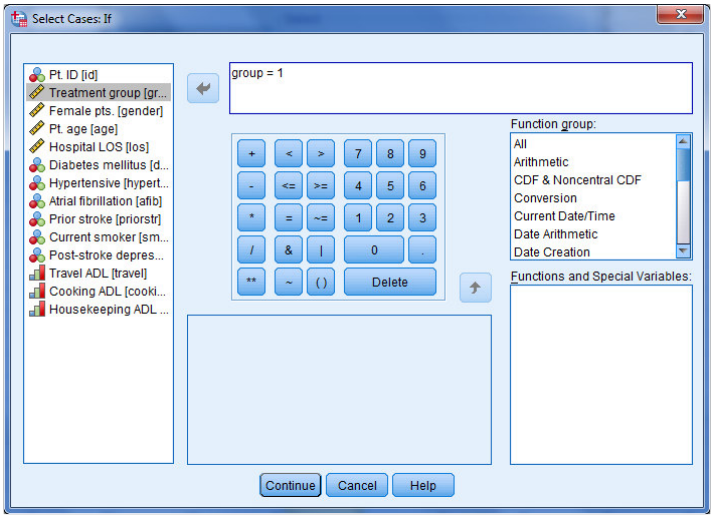
در این فصل به بررسی برخی دستورات در SPSS خواهیم پرداخت که برای کار با داده‌ها و اماده کردن آنها برای تحلیل‌های بعدی بسیار مهم است. به اعتقاد برخی نویسندگان، در حین کار با SPSS استفاده از این دستورها بسیار کمک‌کننده است و بدون وجود این دستورها، این نرم‌افزار قابل استفاده نیست.

۳-۱ دستور Select Cases

همان‌طور که از نام این دستور پیداست، توسط آن می‌توان موردهایی(افرادی) خاص را انتخاب کرد. برای مثال فرض کنید می‌خواهیم تحلیل موردنظر را بر روی افرادی از نمونه که دارای ویژگی خاصی(برای مثال، سیگاری بودن) هستند، اجرا کنیم. توسط این دستور به SPSS می‌گوئیم که تحلیل را فقط بر روی افراد سیگاری انجام بده.

برای بررسی این دستور ابتدا فایل داده "adl.sav" را از فایل‌های نمونه SPSS باز کنید. می‌خواهیم رابطه بین فشارخون (hypertns) و ابتلا به بیماری دیابت(diabetic) را فقط در گروه مورد(Treatment) داده‌ها بررسی کنیم. توجه کنید که گروه مورد در متغیر group با کد ۱ مشخص شده‌است.

می‌دانیم که برای بررسی این رابطه باید از دستور Crosstabs استفاده کرد، اما چون می‌خواهیم رابطه را فقط در گروه مورد انجام دهیم، ابتدا باید دستور Select Cases را اجرا کنیم. برای اجرای این دستور مسیر **Data>Select Cases** را انتخاب کنید و یا در نوار ابزار روی دکمه  کلیک کنید تا کادر مکالمه مربوطه باز شود. در این کادر گزینه دوم(If condition is satisfied) را انتخاب کنید و دکمه if را کلیک کنید، و متغیر group را وارد کادر سمت راست انداخته و عبارت "۱=" را تایپ کنید و گزینه Continue و سپس Ok را بزنید. در این حالت در فایل داده‌ها فقط افراد مورد انتخاب شده‌اند و افراد کنترل برای آنالیزها فیلتر شده‌اند.



حال با اجرای دستور Crosstabs رابطه بین فشارخون و دیابت را بررسی می‌کنیم. خروجی نرم‌افزار به صورت زیر است:

Diabetes mellitus * Hypertensive Crosstabulation

Count

Treatment group			Hypertensive		Total
			No	Yes	
Treatment	Diabetes mellitus	No	33	14	47
		Yes	1	6	7
	Total		34	20	54

Chi-Square Tests

Treatment group		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Treatment	Pearson Chi-Square	8.172[a]	1	.004		
	Continuity Correction[b]	5.950	1	.015		
	Likelihood Ratio	8.196	1	.004		
	Fisher's Exact Test				.008	.008
	Linear-by-Linear Association	8.021	1	.005		
	N of Valid Cases	54				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 2.59.
b. Computed only for a 2x2 table

همان‌طور که مشاهده می‌کنید، این همان خروجی دستور Crosstabs است با این تفاوت که رابطه بین دو متغیر فشار خون و دیابت فقط در افراد "مورد" بررسی شده است.

توجه:

در پنجره Select Cases:if، که نحوه انتخاب Caseها را تعیین می‌کنیم، گزینه‌های زیاد دیگری نیز برای انتخاب وجود دارد. برای مثال می‌توان با استفاده صفحه کلید این کادر تعیین کنیم که، برای مثال، "افراد با سن کمتر یا مساوی ۷۳ سال" انتخاب شوند. برای این‌کار باید در کادر تایپ شود: age<=30 . یا می‌توان از دستورهای منطقی &(and) و |(or) استفاده کرد.

تمرین:

- فایل "adl.sav" باز کنید و تمرینات زیر را انجام بدید.

۱- در بین افراد دیابتی، رابطه بین سیگاری بودن و فشارخون را بررسی کنید.

۲- در بین افراد بالای ۷۳ سال، رابطه بین سیگاری بودن و فشارخون را بررسی کنید.

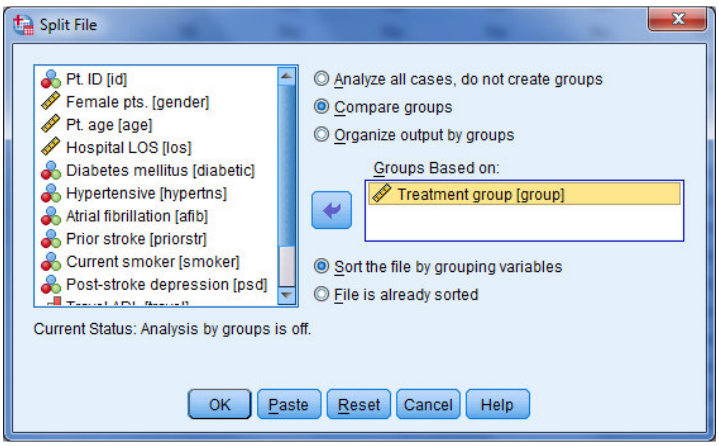
۳- رابطه بین سیگاری بودن و فشارخون را در بین افراد بالای ۷۳ سال که مبتلا به بیماری دیابت نیستند، بررسی کنید.

۴- رابطه بین سیگاری بودن و فشارخون را در بین افرادی که بالای ۷۳ سال دارند یا مبتلا به بیماری دیابت نیستند، بررسی کنید.

۳-۲ دستور Split File

همان‌طور که از نام این دستور پیداست، فایل داده‌ها را می‌شکند. به عبارتی با درنظر گرفتن یک متغر کیفی، مثل جنسیت یا مقطع تحصیلی، فایل داده‌ها تفکیک شده و تمام آنالیزها و خروجی‌ها به تفکیک طبقات این متغیر کیفی ارائه خواهد شد.

برای بررسی این دستور ابتدا فایل داده "adl.sav" را از فایل‌های نمونه SPSS باز کنید. می‌خواهیم رابطه بین فشارخونی بودن(hypertns) و ابتلا به بیماری دیابت(diabetic) را به تفکیک هر یک از گروه‌های تیماری(group) بررسی کنیم. می‌دانیم که برای بررسی این رابطه باید از دستور Crosstabs استفاده کرد، اما چون می‌خواهیم رابطه را به تفکیک یک متغیر سوم بررسی کنیم، ابتدا باید دستور Split file را اجرا کنیم. برای اجرای این دستور مسیر **Data>Split file** را انتخاب کنید و یا در نوار ابزار روی دکمه کلیک کنید تا کادر مکالمه موردنظر باز شود. در این کادر گزینه دوم(Compared groups) را انتخاب کنید و متغیر group را وارد کادر Groups Based On وارد کنید و گزینه Ok را بزنید.



در این حالت داده‌ها تفکیک شده‌اند. و حالا می‌توانید آنالیز موردنظر را انجام دهید.

بعد از اجرای دستور Crosstabs خروجی را به صورت زیر خواهید دید. همان‌طور که مشاهده می‌کنید، این همان خروجی دستور Crosstabs است با این تفاوت که رابطه بین دو متغیر فشار خون و دیابت به تفکیک یک متغیر سوم بررسی شده است.

Diabetes mellitus * Hypertensive Crosstabulation

Count

Treatment group			Hypertensive		Total
			No	Yes	
Control	Diabetes mellitus	No	32	6	38
		Yes	1	7	8
	Total		33	13	46
Treatment	Diabetes mellitus	No	33	14	47
		Yes	1	6	7
	Total		34	20	54

Chi-Square Tests

Treatment group		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Control	Pearson Chi-Square	16.763[a]	1	.000		
	Continuity Correction[b]	13.412	1	.000		
	Likelihood Ratio	15.600	1	.000		
	Fisher's Exact Test				.000	.000
	Linear-by-Linear Association	16.398	1	.000		
	N of Valid Cases	46				
Treatment	Pearson Chi-Square	8.172[c]	1	.004		
	Continuity Correction[b]	5.950	1	.015		
	Likelihood Ratio	8.196	1	.004		
	Fisher's Exact Test				.008	.008
	Linear-by-Linear Association	8.021	1	.005		
	N of Valid Cases	54				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.26.
b. Computed only for a 2x2 table
c. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 2.59.

توجه:

۱- این دستور به‌خصوص زمانی کاربرد دارد که بخواهیم اثر مخدوشگری یا اثرات متقابل را بررسی کنیم.

۲- بعد از اجرای دستور Split file، فایل SPSS همواره بصورت تفکیک شده است. و مادامی که فایل داده‌ها به این صورت است، در قسمت نوار وضعیت SPSS عبارت Split file on نوشته شده است. در صورتی که می‌خواهید از این حالت خارج شوید، دوباره مسیر بالا را طی کنید و در کادر مکالمه‌ای مذکور گزینه "Analysis all cases" را انتخاب کنید و ok را کلیک کنید. در این حالت فایل از حالت تفکیکی درآمده و می‌توانید تمام داده‌ها را تحلیل کنید.

۳- در صورتی که گزینه "Organize output by groups" را انتخاب کنید، خروجی فایل‌های را به صورت تفکیکی ارائه می‌دهد، اما نه کنار هم.

۴- این دستور ابتدا فایل داده‌ها را بر اساس متغیر تفکیکی مرتب می‌کند و سپس فایل داده‌ها را تفکیک می‌کند، و از آنجایی که عموماً فایل داده‌ها مرتب نیست، همواره گزینه " . . . Sort the file" را انتخاب کنید.

تمرین:

- فایل "adl.sav" باز کنید و تمرینات زیر را انجام بدید.

۱- رابطه بین فشارخون و دیابت را یک‌بار به‌صورت کلی و یک‌بار به تفکیک سیگاری بودن و نبودن بررسی کنید. آیا متغیر سیگاری بودن در بررسی رابطه بین فشارخون و دیابت، عامل مخدوشگر است؟

۲- رابطه بین فشارخون و دیابت را در بین افراد گروه کنترل که سیگاری هستند بررسی کنید.

۳-۳ دستور Weight Cases :

زمانی که داده‌ها از نوع فراوانی باشند، برای تعریف داده‌ها می‌توان از این دستور استفاده کرد. دستور را با یک مثال توضیح می‌دهیم.

مثال: در یک مطالعه ژنتیکی مربوط به ساختار کروموزوم ها ۲۸ نفر بر حسب نوع انحرافی که ساختار کروموزوم آنها از وضع طبیعی دارد و بر حسب این که والدینشان حامل این انحراف هستند یا نه رده بندی شده اند و در نتیجه داده های زیر به دست آمده است :

نوع انحراف از وضع طبیعی	حامل‌ها	
	یکی از والدین	هیچ‌یک از والدین
نوع ۱	۴	۱
نوع ۲	۳	۷

می‌خواهیم آزمون کنیم که "نوع انحراف از وضع طبیعی" مستقل از "حامل بودن والدین" است یا خیر.

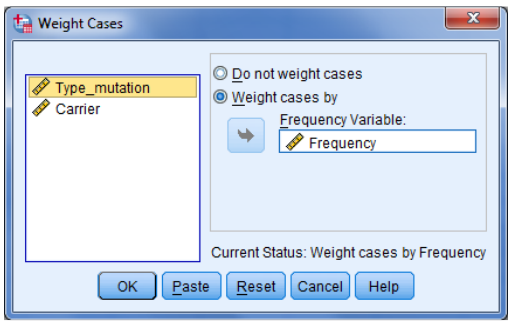
برای انجام این آزمون قائدتاً باید ابتدا داده‌ها را در SPSS کرد!! که البته در این قسمت تنها بر نحوه ورود داده‌ها بحث خواهیم کرد و نه بر نحوه انجام آزمون آماری آن. برای این کار دو راه وجود دارد. **راه اول** این‌که داده‌ها را به‌طور خام وارد SPSS کنیم. یعنی دو متغیر دو متغیر برای "نوع انحراف از وضع طبیعی" و "حامل بودن والدین" تعریف کنیم، آنها را کدبندی کرده و به‌صورت زیر وارد کنیم:

	Type_mutation	Carrier
1	Type 1	one parent
2	Type 1	one parent
3	Type 1	one parent
4	Type 1	one parent
5	Type 1	none of parent
6	Type 2	one parent
7	Type 2	one parent
8	Type 2	one parent
9	Type 2	none of parent
10	Type 2	none of parent
11	Type 2	none of parent
12	Type 2	none of parent
13	Type 2	none of parent
14	Type 2	none of parent
15	Type 2	none of parent

حال اگر فراوانی داده‌ها در هر یک از خانه‌های جدول زیاد باشد و بخواهیم از روش اول استفاده کنیم، ورود داده‌ها وقت زیادی خواهد گرفت. **راه دوم** این است که طبقات داده‌ها را به نرم‌افزار معرفی، برای آنها فراوانی تعریف کنیم و به فراوانی‌ها وزن دهیم. به عبارتی در این مثال که ۴ طبقه (به صورت ترکیبی) داریم، کافیست ۴ طبقه را به عنوان ۴ مشاهده تعریف کنیم و برای آنها فراوانی تعریف کنیم. به عبارتی داده‌ها را به صورت زیر تعریف کنیم. توجه کنید یک متغیر Frequency برای تعریف فراوانی‌ها به دو متغیر قبلی اضافه شده است.

	Type_mutation	Carrier	Frequency
1	Type 1	one parent	4.00
2	Type 1	none of parent	1.00
3	Type 2	one parent	3.00
4	Type 2	none of parent	7.00

بعد از این کار مسیر **Data>Weight Cases** طی کنید یا از نوار ابزار بر دکمه  کلیک کنید و در کادر مکالمه موردنظر متغیر Frequency را به عنوان متغیر وزنی تعریف کنید و گزینه ok را بزنید. با این کار نرم‌افزار دیگر می‌داند که افراد با انحراف نوع ۱ که فقط یکی از والدین‌شان هم به این عارضه دچارند، تعدادشان ۴ نفر است!!



حال می‌توانید دستور Crosstabs را اجرا کنید و خروجی به صورت زیر خواهد شد:

Type_mutation * Carrier Crosstabulation

Count

		Carrier		
		one parent	none of parent	Total
Type_mutation	Type 1	4	1	5
	Type 2	3	7	10
Total		7	8	15

توجه:

۱- مادامی که دستور Weight cases اجرا شده است، داده‌ها به صورت وزنی هستند. لذا برای خارج شده از این حالت مسیر **Data>Weight Cases** را طی کنید و گزینه "Do not weight cases" را انتخاب کنید.

۲- وقتی داده‌ها وزن دارند، در نوار وضعیت SPSS عبارت Weight On نوشته شده است.

۳- این دستور به‌خصوص زمانی کاربرد دارد که با داده‌هایی سروکار دارید که خام نیستند و تعداد آنها نیز زیاد است.

تمرین:

- فایل "adl.sav" باز کنید و تمرینات زیر را انجام بدید. (دستور Weight cases)

۱- در پرتاب ۶۰ بار یک تاس مشاهده می شود که روهای مختلف تاس به صورت زیرظاهر می گردد. جدول فراوانی مبوط به این داده‌ها را رسم کنید.

روهای مختلف	فراوانی
۱	۱۵
۲	۷
۳	۴
۴	۱۱
۵	۶
۶	۱۷
جمع	۶۰

۲- به منظور بررسی وجود یا عدم وجود رابطهٔ بین نوع شوک و زنده ماندن بیمار، از بیمارانی که به انواع شوک مبتلا می‌شوند. نمونه‌هایی انتخاب و آنها را بر حسب زنده ماندن و یا مردن در جدول زیر مرتب کرده‌اند.

نوع شوک	نتیجه	
	زنده	مرده
کاهش حجم خون	۷	۸
قلبی	۱۱	۱۱
عصبی	۱۰	۶
عفونی	۹	۷
غدد داخلی	۳	۵

داده‌های این جدول را در SPSS وارد کرده و جدول توافقی مربوط به آن‌را رسم کنید.

۳-۴ دستور Compute

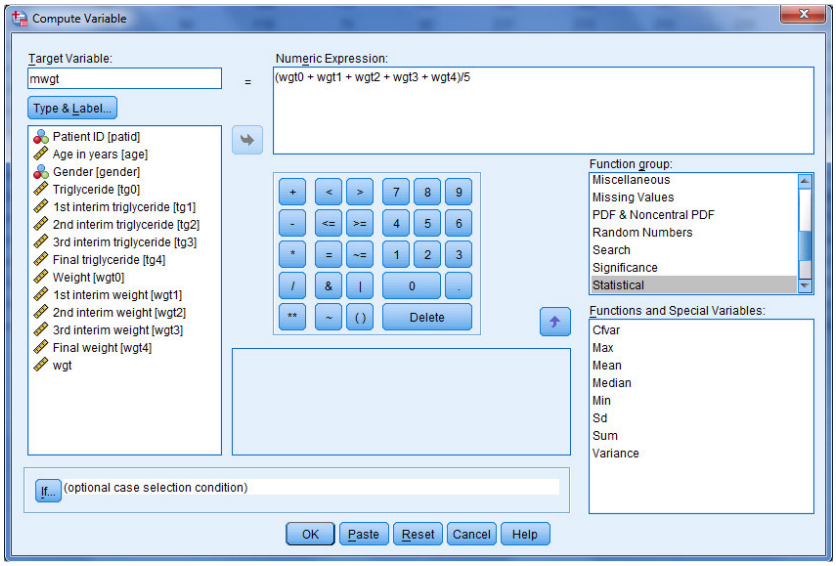
بسیاری اوقات پیش می‌آید که محقق می‌خواهد عبارتی را به‌طور سطری و برای هر یک از موردها(افراد) محاسبه کند. برای مثال فرض کنید قد و وزن نمونه را داریم و می‌خواهیم برای هر فرد به‌طور جداگانه BMI (یعنی $(\text{قد})^2/\text{وزن}$) را محاسبه کنیم. به عنوان مثالی دیگرفرض کنید در یک تست سنجش افسردگی که شامل ۱۵ سئوال ۵ گزینه‌ای (۱ تا ۵) است، می‌خواهیم امتیاز افسردگی هر شخص را به‌طور جداگانه محاسبه کنیم. در تمام این موارد دستور Compute می‌تواند به ما در محاسبه عبارت موردنظر کمک کند.

برای بررسی این دستور ابتدا فایل داده "dietstudy" را از فایل‌های نمونه SPSS باز کنید.

می‌خواهیم میانگین وزن افراد در ۵ نوبت پیگیری(wgt0 تا wgt4) را محاسبه کنیم و در متغیری به نام mwgt، در همین فایل داده‌ها ذخیره کنیم. برای اجرای این دستور مسیر **Transform>Compute variable** را انتخاب کنید تا کادر مکالمه Compute variable باز شود. در کادر سمت چپ (Target variable) عبارت mwgt را تایپ کنید و در کادر سمت راست (Numeric expression) عبارت

(wgt0 + wgt1 + wgt2 + wgt3 + wgt4)/5

را تایپ کنید. برای این قسمت می‌توانید از صفحه کلید پائین، یا گروه توابع (Function group) سمت راست هم استفاده کنید.



پس از این کار کلید Ok را کلیک کنید و به پنجره داده‌ها برگردید. خواهید دید که متغیری تحت عنوان mwgt به متغیرهای قبلی اضافه شده است که در واقع میانگین وزن هر فرد در ۵ دوره پیگیری است.

توجه:

۱- در کادر مکالمه Compute variable، با استفاده از تابع‌های موجود در قسمت Function group بسیاری از محاسبات آماری و ریاضی را می‌توان برای هر فرد به‌طور جداگانه انجام داد.

۲- این دستور به‌خصوص برای محققین رشته‌های پرستاری، روانشناسی و علوم اجتماعی، زمانی که با پرسش‌نامه سروکار دارند، کاربرد فراوانی دارد.

۳- فرض کنید بخواهیم برای گروهی از افراد نمونه یک فرمول و برای گروهی دیگر، فرمولی دیگر را محاسبه کنیم، در این حالت در کادر مکالمه Compute variable، می‌توان از دستور if استفاده کرد.

تمرین:

- فایل "Employee data.sav" باز کنید.

۱- فرض کنیدحقوق اولیه(salbegin) و حقوق جاری(salary) شخص iام را به‌ترتیب با X_i و Y_i نشان دهیم. مقادیر زیر را محاسبه کنید. (دستورCompute variables)

$$\sum_i X_i^2 - 474\overline{X} = \qquad \sum_i Y_i^2 - 474\overline{Y} =$$

$$\sum_i (X_i + Y_i)(X_i - Y_i) =$$

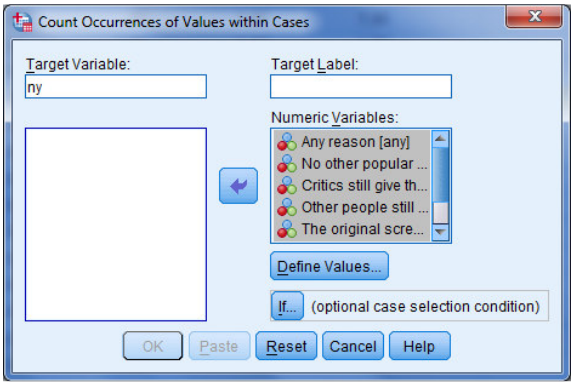
$$\frac{\sum_i (X_i + Y_i)(X_i - Y_i)}{\sqrt{\sum_i X_i^2 - 474\overline{X}}\sqrt{\sum_i Y_i^2 - 474\overline{Y}}} =$$

۲- هر یک از متغیرهای salbegin و salaty استاندارد کنید و میانگین و واریانس هر یک را بدست آورید.

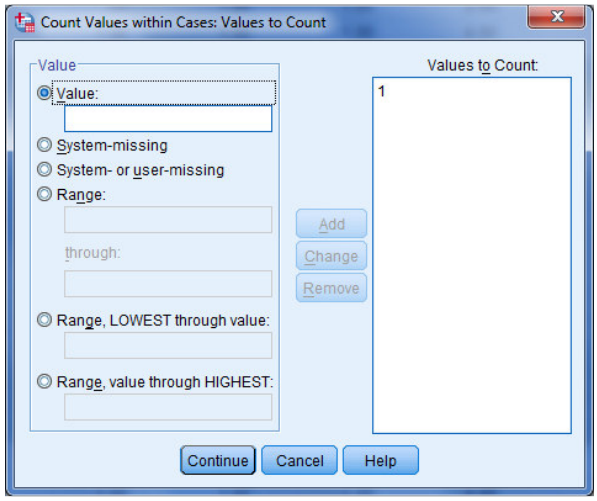
۳-۵ دستور Count Values:

یک تست سنجش افسردگی را در نظر بگیرید. ممکن است محقق بخواهد بداند هر فرد در پاسخ‌گویی چند بار گزینه ۳ را انتخاب کرده‌است. برای پاسخ به این سئوال باید از دستور Count values استفاده کرد.

برای بررسی این دستور ابتدا فایل داده "tv-survey.sav" را از فایل‌های نمونه SPSS باز کنید. این فایل مربوط به یک نظرسنجی در مورد برنامه‌های تلویزیونی می‌باشد، که شامل ۷ سئوال بلی و خیر است. می‌خواهیم بدانیم هر فرد به‌طور کلی چند مرتبه به این سئوالات جواب بلی(کد ۱) را داده است و جواب را در متغیری تحت با نام ny ذخیره کنیم. برای اجرای این دستور مسیر **Transform>Count values** را انتخاب کنید تا کادر مکالمه Count occurrence of values within cases باز شود. در کادر سمت چپ (Target variable) عبارت ny را تایپ کنید و در در کادر variable همه متغیرهای سمت چپ (متغیرهایی که می‌خواهیم در آنها عدد ۱ خوانده شود) را وارد کنید. در این پنجره کلید Define values را کلیک کنید تا پنجره Count values within cases: values to count باز شود. در این پنجره آن‌چه را که می‌خواهید شمارش شود را اضافه کنید. در این



مثال می‌خواهیم عدد ۱ شمارش شود. لذا عدد ۱ را در قسمت values تایپ کرده و عبارت Add را کلیک می‌کنیم، سپس دکمه Continue و Ok را کلیک می‌کنیم. حال به پنجره داده‌ها برگردید. خواهید دید که متغیری تحت عنوان ny به متغیرهای قبلی اضافه شده است که در واقع تعداد پاسخ‌های بله هر فرد است.



توجه:

۱- در کادر مکالمه‌ای Count values within cases: values to count می‌توان نحوهٔ شمارش را به‌صورت دیگری (غیر از مثال بالا) نیز تعریف کرد. همچنین می‌توان در این پنجره بیش از یک شرط شمارش تعریف کرد.

۲- هرگاه بخواهیم این شمارش را برای گروهی از افراد نمونه با گروه دیگر از افراد نمونه متفاوت باشد، باید از دستور if در کادر مکالمه Count occurrence of values within cases استفاده کرد.

تمرین:

- فایل "stoke_valid.sav" باز کنید و تمرینات زیر را انجام بدید. (دستورCount values)

۱- جدول فراوانی مربوط به تعداد سکته‌ها در سه نوبت پیگیری افراد (strok1، stroke2 و stroke3) را رسم کنید.

۲- تمرین (الف) را به تفکیک هر بیمارستان انجام داده و بگوئید، در کدام بیمارستان بیشترین تعداد سکته رخ داده است.

۳- آیا در بین کل این داده‌ها، داده گمشده‌ای وجود دارد.

۳-۶ دستور Recode :

در این قسمت به بررسی دو کاربرد دستور Recode می‌پردازیم.

۳-۶-۱ جدول فراوانی برای صفات کمی پیوسته(دستور Recode):

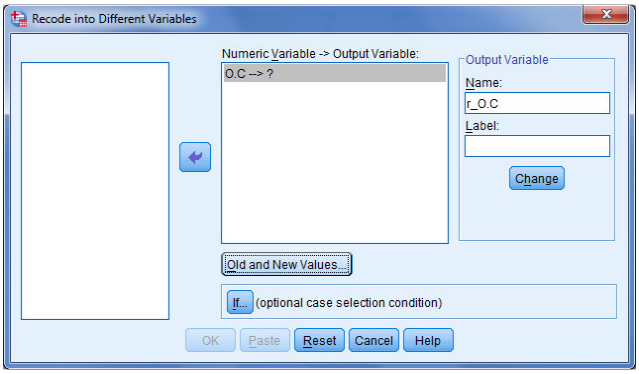
یک روش تعیین طبقات در فصل دوم ارائه شد. فرض کنید حدود طبقات را با این روش تعیین کردیم، دراین بخش می‌خواهیم این حدود طبقات را به نرم‌افزار معرفی کنیم تا داده‌ها را دسته‌بندی کند و به وسیلهٔ آن جدول فراوانی داده‌ها را رسم کنیم.

مثال۳) داده‌های مثال ۳ (تمرکز اوزن در ۴۰ شهر بزرگ) را در SPSS وارد کنید و نام متغیر را O.C بگذارید. برای دسته‌بندی داده‌ها با استفاده با SPSS، در **گام اول**، حدود طبقات را به آن تعریف می‌کنیم و متغیر دسته‌بندی شده را به عنوان متغیری جدید در صفحه داده‌ها ایجاد می‌کنیم. حدود طبقاتی که باید تعریف شوند، به این‌صورت است.

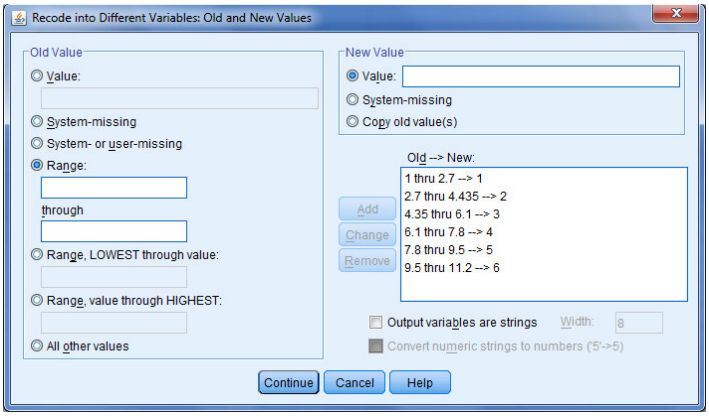
کد وارد شده	طبقات جدید	طبقات قدیم
۱	۱ـ۲/۷	۱ـ۲/۷
۲	۲/۷ـ۴/۳۵	۲/۷ـ۴/۴
۳	۴/۳۵ـ۶/۱	۴/۴ـ۶/۱
۴	۶/۱ـ۷/۸	۶/۱ـ۷/۸
۵	۷/۸ ـ۹/۵	۷/۸ ـ۹/۵
۶	۹/۵ـ۱۱/۲	۹/۵ـ۱۱/۲

توجه داشته باشید عدد ۴/۴ که یکی از کران‌هاست(کران بالای طبقه دوم و کران پائین طبقه سوم) در داده‌ها نیز وجود دارد. و از آنجایی که می‌خواهیم ۴/۴ در طبقه سوم باشد، در طبقه دوم و سوم به جای عدد ۴/۴ در کرانه‌ها، عدد ۴/۳۵ را درنظر گرفته‌ایم و اگر اعداد ۲/۷، ۱/۶، ۷/۸ یا ۹/۵ نیز در داده‌ها بودند، همین کار را انجام می‌دادیم.

برای دسته‌بندی داده‌ها مسیر **Transform>Recode into different variables** را انتخاب کنید تا کادر مکالمه Recode into different variables باز شود. در این کادر متغیر "مقدار تمرکز ازن (O.C)" را انتخاب و وارد پنجره سمت راست(Numeric Variable->Output variable) می‌کنیم. در قسمت سمت راست این کادر (Output Variable) و در قسمت Name، نام متغیر جدید، یعنی r_O.C، را تایپ می‌کنیم و سپس دکمه Old and New values را کلیک می‌کنیم تا کادر Old and New values: Recode into different variables باز شود. در این کادر و در قسمت Old values، گزینه Range را انتخاب کنید و اعداد ۱ و ۲/۷ (حد بالا و پائین طبقه اول) را در این جعبه تایپ کنید. در قسمت New values عدد ۱ را تایپ می‌کنیم و دکمه Add را می‌زنیم.

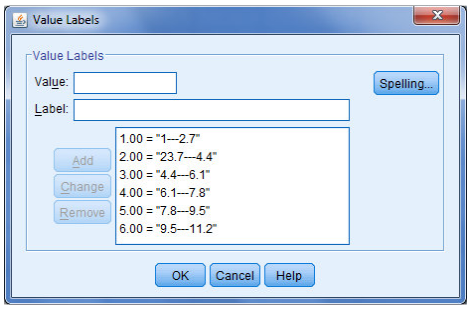


به همین ترتیب سایر طبقات را به نرم‌افزار معرفی می کنیم.



پس از اینکه معرفی طبقات به نرم‌افزار به اتمام رسید، دکمه‌های Continue، Change و Ok را به ترتیب کلیک می کنیم. حال به پنجره داده‌ها برگردید. خواهید دید که متغیری تحت عنوان r_O.C به متغیرهای قبلی اضافه شده است که در واقع همان متغیر O.C است، اما با کدبندی جدید.

گام دوم این است که به هر یک از طبقات متغیر جدید Valueای در خور،مانند شکل زیر، در نظر بگیریم (در پنجره Variable view).



حال با استفاده از دستور Frequency برای متغیر r_O.C جدول فراوانی رسم می کنیم.

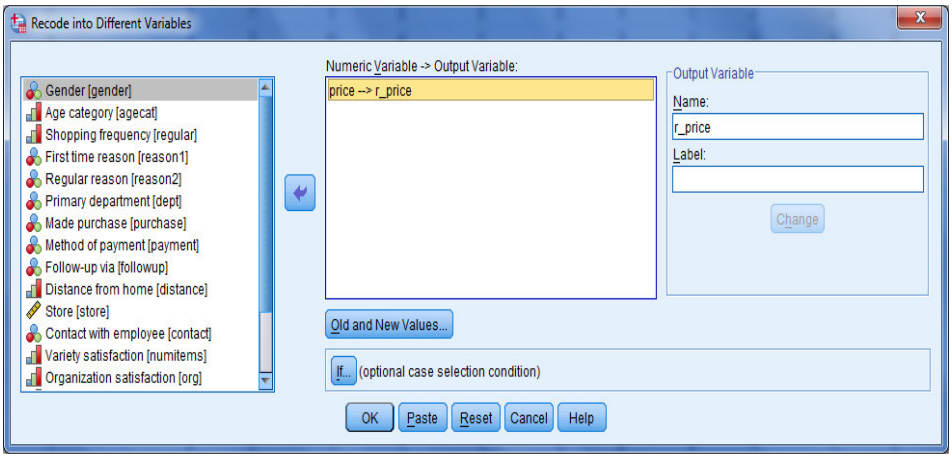
r_o.c

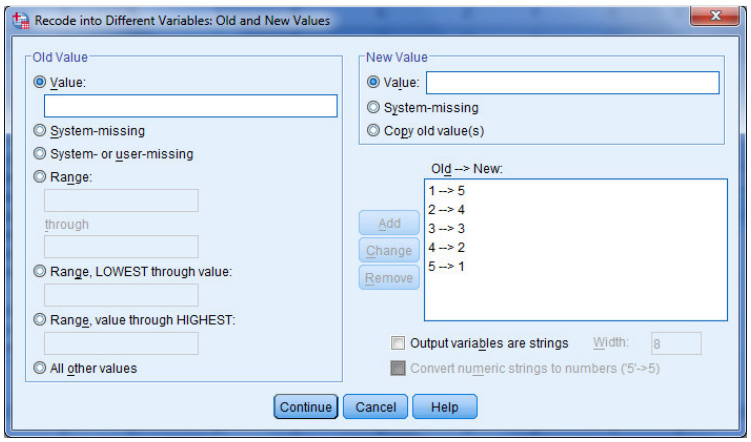
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1---2.7	6	15.0	15.0	15.0
	23.7---4.4	9	22.5	22.5	37.5
	4.4---6.1	13	32.5	32.5	70.0
	6.1---7.8	7	17.5	17.5	87.5
	7.8---9.5	3	7.5	7.5	95.0
	9.5---11.2	2	5.0	5.0	100.0
	Total	40	100.0	100.0	

۳-۶-۲ معکوس کردن امتیازات(دستور Recode):

همان‌طور که از نام این دستور پیداست، توسط آن می‌توان داده‌ها را دوباره کدبندی کرد. برای مثال فرض کنید در پرسش‌نامه‌ای ۲۰ سئوالی، ۱۸ سئوالات با نمره‌گذاری از ۱ تا ۵ (مستقیم) معنی می‌دهند، اما در ۲ سئوال جهت نمره‌گذاری ۵ تا ۱(معکوس) است. در این حالت بهتر است همه داده‌ها وارد شوند، اما جهت نمره گذاری آن دو سئوال را با استفاده از دستور Recode برعکس کرد.

برای بررسی این دستور ابتدا فایل داده "satisfy.sav" را از فایل‌های نمونه SPSS باز کنید. می‌خواهیم کدبندی متغیر price را معکوس کرده و آن‌را در متغیری جدید به نام r_price ذخیره کنیم. برای این کار مسیر **Transform>Recode into different variables** را انتخاب کنید تا کادر مکالمه Recode into different variables باز شود. در این کادر متغیر Price را انتخاب و وارد پنجره سمت راست می‌کنیم. در قسمت سمت راست این کادر (Output Variable) و در قسمت Name، نام متغیر جدید، یعنی r_price، را تایپ می‌کنیم و سپس دکمه Old and New values را کلیک می‌کنیم تا کادر Old and New values Recode into different variables: باز شود. در این کادر در قسمت Old values عدد ۱ و در قسمت New values عدد ۵ را تایپ می‌کنیم. و همین کار را برای کدهای ۲، ۳، ۴ و ۵ انجام می‌دهیم و آنها تبدیل به کدهای، به‌ترتیب، ۴، ۳، ۲ و ۱ می‌کنیم. سپس دکمه‌های Continue، change و Ok را به ترتیب کلیک می‌کنیم. حال به پنجره داده‌ها برگردید. خواهید دید که متغیری تحت عنوان r_price به متغیرهای قبلی اضافه شده است که در واقع همان متغیر price است، اما با کدبندی معکوس.





توجه:

۱- کاربرد دیگر دستور Recode را پیشتر در فصل دوم، در قسمت دسته‌بندی داده‌های کمی، بحث کرده‌ایم.

۲- هرگاه بخواهیم این دوباره کدبندی را برای گروهی از افراد نمونه با گروه دیگر از افراد نمونه متفاوت باشد، در کادر مکالمه Recode into different variables، باید از دستور if استفاده کرد.

۳- در منوی Transform، دو نوع دستور Recode تعبیه شده است. یک نوع را در بالا بررسی کردیم. حالت دیگر Recode into same . . . است که در آن متغیر جدیدی ایجاد نمی‌شود و اعداد دوباره کدبندی‌شده در همان متغیری اول ذخیره می‌شوند. لذا این امر باعث از بین رفتن اطلاعات اصلی می‌شود و استفاده از دستور Recode بدین‌گونه توصیه نمی‌شود.

فصل چهارم
آزمون فرض آماری

۴-۱ مقدمه

تحقیقات همواره با سوال و فرضیه شروع می شوند. بسیاری از تحقیقات از مرحله سوال گذشته و به مرحله فرضیه می رسند. فرضیه حدس زیرکانه درباره پارامتر جامعه است. به فنون آماری مناسب برای تحلیل صحت و یا نادرستی فرضیه ها، فنون «آزمون فرض آماری» (Hypothesis testing) گفته می شود که در این فصل آنها را بررسی می کنیم.

بطور کلی هدف «آزمون فرض آماری» تعیین این موضوع است که با توجه به اطلاعات بدست آمده از داده های نمونه، حدسی که درباره خصوصیتی از جامعه می زنیم به طور قوی تایید می شود یا نه. این حدس بنا به تحقیق، نوعا شامل ادعایی درباره مقدار یک پارامتر جامعه است. «در واقع هر حکمی درباره جامعه را یک فرض آماری می نامند که قابل قبول بودن آن باید بر مبنای اطلاعات حاصل از نمونه گیری از جامعه بررسی شود.»

چون ادعا ممکن است صحیح یا غلط باشد، بنابراین دو فرض مکمل در ذهن بوجود می آید:

۱- ادعا صحیح است (فرض H_0)

۲- ادعا غلط است (فرض H_1)

با بکار بردن اطلاعاتی که از مشاهدات نمونه بدست می آید، تصمیم گیرنده باید یکی از دو تصمیم یا استنباط را انتخاب کند:

۱- فرض H_1 را رد کند و نتیجه بگیرد که H_0 بوسیله داده ها تایید می شود.

۲- فرض H_1 را رد نکند و نتیجه بگیرد که داده ها H_0 را تایید نمی کند.

فرایند انتخاب یکی از دو تصمیم فوق را «آزمون فرض آماری» می نامند.

قبول و یا رد یک فرض آماری با اثبات و یا رد یک گزاره ریاضی متفاوت است، در ریاضی گزاره ای را اثبات و یا نفی می کنند و در هر حالت نتیجه اش که بدست می آید بدون هیچ شکی برقرار است ولی در مقابل، نتیجه حاصل از «آزمون فرض آماری» به وسیله تحلیل داده های تجربی حتمی و قطعی نیست. شیوه مناسب برای آزمون فرض دارای مراحل منطقی است. قبل از ذکر مراحل مورد نظر، به بیان مفاهیم و اصطلاحات استفاده شده در آزمون فرض می پردازیم. شیوه مناسب برای آزمون فرض آماری دارای مراحل منطقی است. قبل از ذکر مراحل مورد نظر، به بیان مفاهیم و اصطلاحات استفاده شده در آزمون پرداخته می شود. مهمترین مرحله در «آزمون فرض آماری» تبدیل «فرضیه پژوهشی» و نقیض آن، به «فرضیه های آماری» است. بنابراین مرحله فوق را با عنوان فرض صفر و فرض مقابل تشریح می کنیم. در این فصل به بیان برخی تعاریف در زمینه آمار استنباطی خواهیم پرداخت و سپس چهار آزمون آماری پرکاربرد در آمار کاربردی را معرفی خواهیم کرد.

۴-۲ فرض صفر و فرض مقابل

برای بحث درباره فرمول بندی مساله آزمون فرض آماری و حل آن، به معرفی پاره ای از تعاریف و مفاهیم نیاز داریم.

مثال ۱:

نظریه ای پیشنهاد می کند که محصول یک واکنش شیمیایی معنی دارای توزیع نرمال $X\sim N(\mu،۱۶)$ است. آزمایش گذشته نشان می دهد که اگر یک ماده معدنی به این محصول اضافه نشده باشد μ =۱۰ و در غیر اینصورت μ =۱۱ است. آزمایش ما عبارت است از انتخاب نمونه تصادفی به حجم n. بر اساس این نمونه سعی خواهیم کرد تصمیم بگیریم کدام مورد درست است؟

پاسخ:

با توجه به فرضیه ای که در صورت مساله بیان شده است دو فرض آماری زیر مطرح می شود:

$\begin{cases} \mu = ۱۰ \text{ : میانگین جامعه برابر عدد ۱۰ است} \\ \mu = ۱۱ \text{ : میانگین جامعه برابر عدد ۱۱ است} \end{cases}$

در اینجا عدد نامعلوم ، برابر بودن μ =۱۱ است. از دو حکم فوق، یکی را «فرض صفر» (Null Hypothesis) و دیگری را «فرض مقابل» (Alternative Hypothesis) می نامیم و آنها را به ترتیب و به اختصار با H_0 و H_1 نشان می دهیم. برای اینکه معلوم شود کدام فرض را باید فرض صفر نامید لازم است که تفاوت اساسی بین دو اصطلاح فوق به روشنی درک شود. قبل از اینکه ادعا کنیم حکمی معتبر است، باید شواهد کافی در تایید آن بدست آوریم. در نتیجه، شخص تحلیلگر باید حمل را غلط بداند مگر اینکه داده های بدست آمده خلاف آنرا تایید کنند. به عبارت دیگر باید «فرض صفر» را صحیح دانست و فقط وقتی آنرا رد کرد که داده ها برخلاف آن حکم کنند. تشابه زیادی بین این امر و محاکمه دادگاه وجود دارد که در آن هیات منصفه فرض «مجرم بودن» را اتخاذ می کنند مگر اینکه شواهد قانع کننده ای مجرم بودن منتهی را ثابت کنند و نه اینکه در اثبات بی گناهی او بکوشند.

با توجه به نکات فوق، می توان نتیجه گرفت که هرگاه بخواهیم یک ادعا را از طریق تایید آن بوسیله اطلاعات حاصل از نمونه آزمون کنیم، نفی آن ادعا را فرض صفر و خود آن را فرض مقابل می گیریم، بنابراین فرض صفر و فرض مقابل در فرضیه فوق باید به این صورت باشد.

$\begin{cases} H_0 : \mu = ۱۰ \\ H_1 : \mu = ۱۱ \end{cases}$

۴-۳ سطح معنی داری و خطاهای آماری

بعد از تعریف فرضیه های آماری، قدم بعدی مشخص کردن درجه ای برای معنی دار بودن تفاوت ها (α) و حجمی برای نمونه مورد بررسی (n) است. روش کار این است که فرض H_0 را به نفع فرض H_1 رد می کنیم

به شرط اینکه از یک آزمون آماری مقداری بدست آوریم که احتمال وقوع آن مقدار با توجه به H_0 برابر یا کمتر از یک احتمال بسیار کوچک باشد که با α نشان داده می شود. این احتمال وقوع کوچک را «سطح معنی داری» می گویند. مقادیری که معمولا برای α استفاده می شود بیشتر از ۰.۰۱ و ۰.۰۵ است. از آنجا که مقدار α در تعیین اینکه H_0 باید رد شود یا نه دخالت مستقیم دارد، الزام رعایت عینیت در تحقیق ایجاب می کند که α را پیش از شروع جمع آوری داده ها مشخص کنیم. سطح معنی داری که محقق برای تعیین α در تحقیق انتخاب می کند بر اساس تخمین او از اهمیت و یا درجه قابلیت کاربرد یافته هایش مبتنی است.طبیعی است که اگر تحقیق مثلا درباره آثار درمانی عمل جراحی روی مغز باشد، محقق باید α را خیلی کمتر در نظر بگیرد زیرا خطرهای رد کردن نادرست فرضیه صفر بسیار زیاد است.

هنگام اتخاذ تصمیم درباره H_0 ممکن است دو نوع خطا پیش آید:

۱- خطای نوع اول (Type I Error) : رد کردن فرض H_0 در حالی که درست است.

۲- خطای نوع دوم (Type I Error) : رد کردن فرض H_0 در حالی که غلط است.

احتمال وقوع خطای نوع اول با α بزرگتر شود، احتمال اینکه H_0 را به غلط رد کنیم یا به عبارت دیگر، احتمال اینکه مرتکب خطای نوع اول شویم، افزایش می یابد. خطای نوع دوم معمولا با β نشان داده می شود. α و β هم برای نشان دادن نوع خطاها و هم ارتکاب آن خطاها به کار می روند، یعنی:

α =P (خطای نوع اول) =P (رد کردن H_0 وقتی H_0 درست است)

β =P (خطای نوع اول) =P (رد کردن H_0 وقتی H_1 درست است)

احتمال α به مقدار مشخص پارامتر در دامنه ای بستگی دارد که H_0 آنرا در بر می گیرد و حال آنکه β به مقدار پارامتر در دامنه ای بستگی دارد که H_1 آنرا در برمی گیرد. این خطاها و احتمال آنها در رابطه با H_0 را می توان بطور زیر خلاصه کرد:

واضح است که بین α و β رابطه معکوس وجود دارد. بالا رفتن α، و مقدار β کاهش می یابد و برعکس. این رابطه در آمار به «بده – بستان» بین α و β معروف است. آنچه مسلم است، مجموع α و β الزاماً عدد یک نیست. واضح است که در هر استنباط آماری، احتمال وقوع یکی از این دو نوع خطا وجود دارد و لازم است که آزمون کننده به نوعی سازش که تعادل بین احتمال وقوع این دو نوع خطا را به حد مطلوب برساند دست یابد. آزمون های آماری مختلف ، احتمال تعادل های مختلفی را عرضه می کنند. در رسیدن به چنین تعادلی است که موضوع «توان آزمون» مطرح می شود. توان آزمون عبارت اس از احتمال رد کردن H_0 وقتی که H_0 حقیقتاً نادرست باشد، یعنی:

توان آزمون =۱- (احتمال وقوع خطای نوع دوم) =۱ - β

آنچه موجب کاهش خطای نوع اول و دوم و همچنین توان آزمون می شود، افزایش حجم نمونه است. منحنی های شکل ۱-۴ نشان می دهند که وقتی حجم نمونه (n) افزایش می یابد احتمال وقوع خطای نوع دوم (β) کاهش می یابد. در این شکل افزایش توان آزمون دو طرفه میانگین وقتی که حجم نمونه افزایش می یابد با یکدیگر مقایسه شده است. مشاهده می شود که وقتی حجم نمونه از ۴ به ۱۰،۲۰،۵۰ و ۱۰۰ افزایش می یابد چگونه توان آزمون نیز زیادتر می شود.

۴-۴ توزیع نمونه گیری آماره

صحت یک فرضیه آماری فقط با استفاده از نمونه ای n تایی از جامعه آماری و توزیع نمونه گیری آماره مشخص می شود. در بحث «آزمون فرض» تایید یا رد «فرضیه صفر» به توزیع نمونه گیری آماره بستگی دارد. واضح است که توزیع آماره متاثر از توزیع جامعه و شرایط برآورد و همچنین حجم نمونه است. جدول زیر نشان می دهد که از چه توزیع هایی برای آزمون هر پارامتر استفاده خواهد شد. متغیرهای استاندارد مورد استفاده به کمک توزیع نمونه گیری تعریف می شوند. اصطلاح «آماره آزمون» با پارامتر مورد آزمون تعریف می شود.

۴-۵ آزمون فرض یک طرفه و دو طرفه

بر اساس آنچه گفتیم، نتیجه می گیریم که H_0 را باید پذیرفت مگر آنکه دلیل محکمی بر رد آن وجود داشته باشد. این بدین معناست که فرض صفر همواره شامل سطح اطمینان $100(1-\alpha)$ درصد است، بنابراین H_1 در برگیرنده سطح معنی داری α است. مفهوم کاربردی این جمله آن است که رد یا قبول H_0 با سطح اطمینان دلخواه صورت خواهد گرفت. پس همواره H_1 به اندازه α در طرفین توزیع نمونه گیری تعریف خواهد شد. «یک طرفه » یا دو طرفه بودن آزمون فرض آماری به تعریف H_0 و یا H_1 بستگی دارد.

مثال ۲:

در فرضیه پژوهشی چنین آمده است: « میانگین مدت زمان کارکرد یک لامپ سالم کمتر از ۲۰۰ ساعت است» . فرضیه فوق طوری به فرضیه های آماری تبدیل می شود که H_0 نشان دهنده نقیض ادعا و H_1 نشان دهنده فرضیه پژوهشی است. بنابراین:

$$\begin{cases} H_0 : \mu \geq ۲۰۰ \\ H_1 : \mu < ۲۰۰ \end{cases}$$

واضح است که H_1 به اندازه α در سمت چپ توزیع نمونه گیری $\bar{X}$ تعریف خواهد شد. این نوع آزمون را «آزمون یک طرفه چپ» (left tailed Test) می خوانند (شکل ۲-۴).

مثال ۳:

فرض کنید در مثال فوق چنین آمده باشد:« میانگین مدت زمان کارکرد یک لامپ سالم بیشتر از ۲۰۰ ساعت است».

فرضیه فوق را می توان بصورت زیر به فرضیه آماری تبدیل کرد:

$$\begin{cases} H_0 : \mu \leq ۲۰۰ \\ H_1 : \mu > ۲۰۰ \end{cases}$$

به این نوع آزمون «آزمون یک طرف راست» (Right Tailed Test) گفته می شود (شکل ۲-۴).

مثال ۴:

مجددا فرض کنید در فرضیه پژوهشی چنین آمده است:« میانگین مدت زمان کارکرد یک لامپ سالم برابر ۲۰۰ ساعت است.» فرضیه آماری عبارت فوق چنین است:

$$\begin{cases} H_0 : \mu \neq ۲۰۰ \\ H_1 : \mu \neq ۲۰۰ \end{cases}$$

به چنین فرضیه هایی «آزمون دو طرف» (2 Tail Test) گویند (شکل ۴-۴).

لازم به ذکر است که اگر در شکل های فوق میانگین نمونه در «ناحیه بحرانی» قرار بگیرد، فرض صفر (H_0) را رد می کنیم.

۴-۶ مراحل کلی آزمون فرض آماری

از جمع بندی مبانی آزمون فرض می توان برای همه آزمون فرض های آماری، مراحل چهارگانه زیر را تدوین کرد. از این مراحل در طی فصول بعدی کتاب جهت تفسیر خروجی ها استفاده خواهد شد:

۱- تعریف فرضیه های آماری H_0 و H_1 (فرض ها) :

بر اساس قاعده ای که بیان شد، چنانچه فرضیه پژوهشی یا هدف، مرز مشخصی (=) داشته باشد، H_0 نشان دهنده ادعا خواهد بود، در غیر اینصورت نقیض آن در H_0 قرار خواهد گرفت، آنچه مسلم است فرض H_0 و H_1 مکمل یکدیگرند. با این توصیف H_0 گاهی بیان کننده ادعا و گاهی نقیض ادعا خواهد بود.

II- تعیین توزیع نمونه گیری آماره و نوع آماره آزمون (آماره آزمون):

توزیع نمونه گیری به شرایط تخمین پارامتر مورد ادعا بستگی دارد. بسته به اینکه فرضیه پژوهشی چه نوع پارامتری را بیان می کمند، توزیع نمونه گیری آماره و آماره آزمون تغییر خواهد کرد.

III- تعیین سطح زیر منحنی H_0و H_1 و محاسبه مقدار بحرانی (مقدار بحرانی) :

سطح زیر منحنی H_0و H_1به توزیع نمونه گیری و مقدار α بستگی دارد. یک طرفه یا دو طرفه بودن آزمون نیز بر سطح زیر منحنی فرضیه های آماری تاثیر مستقیم دارد. چنانچه گفته شد: H_0 در برگیرنده سطح اطمینان و H_1 سطحی برابر α خواهد بود. محاسبه مقدار استانداردی که تفکیک کننده H_0و H_1 بصورت عددی باشد از دیگر موارد ضروری این مرحله است. مقدار استاندارد بر اساس نوع آزمون و مقدار α از جداول آماری موجود استخراج می شود. این مقدار با توجه به علامت آن، «مقدار بحرانی» نامیده می شود. مقدار استاندارد و

جدول آماری مورد نیاز برای استخراج آن بر اساس آماره آزمون تعیین می شود. برای مثال اگر آماره آزمون از نوع Z باشد ف مقدار بحرانی بر اساس جدول استاندارد Z تعیین می شود و اگر از نوع F باشد بر اساس جدول F تعریف می شود.

IV. تصمیم گیری:

در این مرحله مقدار آماره آزمون محاسبه شده در مرحله دوم با مقدار بحرانی مرحله سوم مقایسه می شود. چنانچه آماره آزمون در ناحیه پذیرش H_0 قرار گیرد. گفته می شود که رد سطح اطمینان مورد نظر، دلیل کافی برای پذیرش H_0 وجود دارد. در غیر اینصورت فرض H_0 رد شده و H_1 آیا در سطح خطای α درصد پذیرفته می شود.

پس از تایید یا رد H_0، تحلیلگر باید بطور مشخص بیان کند که آیا فرضیه پژوهشی پذیرفته یا رد شده است. بدیهی است محقق هیچگاه ادعای اثبات یا عدم اثبات فرضیه پژوهشی یا فرضیه های آماری را ندارد، بلکه در تحلیل خود به لحاظ استقرار احتیاط را رعایت کند.

۴-۷ ماهیت P-Value

P-Value (P- مقدار) یک آزمون آماری، مقدار احتمالی است که میزان سازگاری داده های نمونه را با نتیجه H_0 اندازه می گیرد. این مقدار، خلاصه ای فشرده از یافته های نمونه ای را در یک آزمون آماری معرفی می کند، و غالباً در گزارش های منتشر شده نتایج آزمون آماری و در خروجی برنامه های کامپیوتری مورد استفاده واقع می شود.

برای یک آزمون یک طرفه مربوط به میانگین جامعه، P-Value بر حسب آماره آزمون استاندارد شده Z^* بصورت زیر تعریف می شود:

P-Value یک آزمون آماری یک طرفه برای μ عبارتست از احتمال آنکه اگر $\mu = \mu_0$ ، آماره آزمون استاندارد شده Z^* در جهت ناحیه رد می تواند کمتر از مقداری باشد که واقعاً مشاهده شده است.

آزمون P-Value

همان‌طور که قبلاً گفتیم، P-Value یک آزمون برای μ ، میزان سازگاری بین برآمد نمونه ای و مقدار μ_0 را که در H_0 مسلم فرض شده است اندازه می گیرد. یک P-Value بزرگ نشان می دهد که μ_0 موجه است، و بنابراین، باید H_0 نتیجه گرفته شود. در واقع برحسب اینکه P-Value بزرگتر یا کوچکتر از α (ناحیه بحرانی) باشد که برای اجرای آزمون مشخص شده است، می تواند مستقیما برای انتخاب بین H_0 و H_1 به کار رود. نتیجه ای که از آزمون مبتنی بر P-Value بدست می آید از نظر ریاضی هم ارز با نتیجه ای است که از قاعده تصمیم متناظر مبتنی بر آماره آزمون استاندارد شده حاصل می شود. قاعده تصمیم مبتنی بر P-Value بصورت زیر است:

اگر P-Value ≤ 0.05 باشد، گزینه H_0 را نتیجه بگیرید.

اگر P-Value ≥ 0.05 باشد، گزینه H_1 را نتیجه بگیرید.

با استفاده از یک P-Value یک طرفه یا دو طرفه، هر کدام که مقتضی باشد، این قاعده تصمیم، خواه آزمون یک طرفه یا دو طرفه باشد صادق است.

۴-۸ آزمون آماری برای میانگین جامعه – آزمون t تک‌نمونه‌ای

وقتی می‌خواهیم برای میانگین جامعه (μ) آزمونی انجام دهیم، در واقع علاقه‌مندیم آزمون کنیم که آیا میانگین جامعه برابر عدد مشخصی هست یا خیر؟ در این حالت فرض‌های آماری عبارتند از:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 = \mu \neq \mu_0 \end{cases}$$

از آنجایی که در عمل همواره واریانس جامعه نامعلوم است، آزمون مناسب، آزمون t تک‌نمونه‌ای می‌باشد.

مثال ۱) داده‌های زیر نمره‌های ۲۰ دانش‌آموز در درس آمار است:

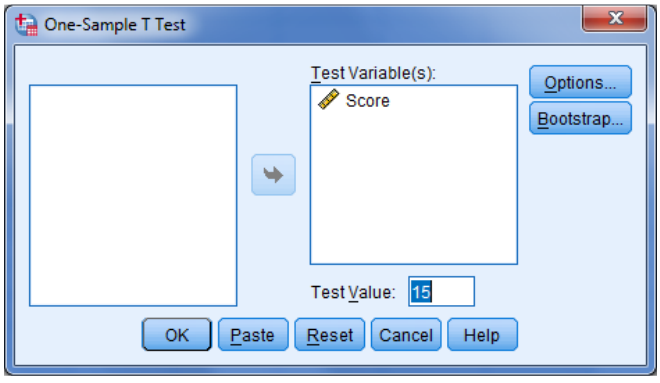
۱۷/۵، ۱۴، ۱۳/۵، ۱۶، ۱۴/۵، ۲۰، ۱۳، ۱۲، ۱۶/۵، ۱۹، ۱۶، ۱۰/۵، ۱۷، ۱۲، ۱۳، ۱۴/۵، ۱۷، ۱۱، ۱۹/۸، ۱۲/۵

آیا می‌توان گفت میانگین نمره‌های ریاضی دانش‌آموزان این کلاس برابر ۱۵ است؟

در واقع می‌خواهیم آزمون فرض زیر را انجام دهیم:

$$\begin{cases} H_0 : \mu = 15 \\ H_1 = \mu \neq 15 \end{cases}$$

برای انجام این آزمون مسیر Analyze>Compare Means>One-Sample T Test را طی کنید، تا کادر مکالمه‌ای One-SampleT Test باز شود. در این کادر متغیر "نمره" را در جعبه Test Variable وارد و در جعبه Test Values عدد ۱۵ را تایپ کنید.



سپس گزینه Ok را بزنید تا خروجی آزمون را ببینید.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Score	20	14.9500	2.82796	.63235

One-Sample Test

	Test Value = 15					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Score	-.079	19	.938	-.05000	-1.3735	1.2735

معرفی جداول: جدول اول (One-Sample Statistics)، حجم نمونه، شاخص‌های آماری میانگین، انحراف معیار و خطای معیار میانگین را نشان می‌دهد. جدول دوم (One-Sample Test) نتایج آزمون t تک‌نمونه‌ای را نشان می‌دهد. مقدار آماره t محاسبه‌شده برابر ۰/۰۷۹- شده است که می‌دانیم آماره مربوط به آن به‌صورت زیر است و دارای توزیع t با n-1 درجه آزادی(df) است.

$$T = \frac{\bar{x} - \mu_{0T}}{\frac{S}{\sqrt{n}}}$$

عبارت sig که مخفف significant levele، و به معنی "سطح معنی‌داری" است، نشان‌دهنده P-مقدار آزمون است، که در این آزمون برابر ۰/۹۴ شده‌است. همچنین مقدار Mean Difference نشان‌دهندهٔ $\bar{x} - \mu_0$ است، که در اینجا برابر است با ۰/۰۵-=۱۵-۱۴/۹۵.

در قسمت بعد، 95%Confidence Interval، فاصله اطمینان ۹۵٪ را برای پارامتر $\bar{x} - \mu$ را نشان می‌دهد، که در این مثال برابر (۱/۲۷، ۱/۳۷-) است. حال اگر بخواهیم فاصله اطمینان برای μ بدست آوریم، باید نامعادله $-1.37 < \bar{x} - \mu < 1.27$ را برای مجهول μ حل کنیم. لذا فاصله اطمینان برای μ برابر است با (۱۳/۶۸، ۱۶/۳۲).

تفسیر: همانطورکه می‌دانیم، از سه روش معادل می‌توان آزمون فرض را انجام داد که عبارتند از: ۱- ناحیه رد، ۲- P-مقدار، ۳- فاصله اطمینان. که در اکثر خروجی‌های SPSS نتایج بر اساس این سه روش ارائه می‌شود. ما نیز با استفاده از خروجی بالا از هر سه راه این فرضیه را در سطح معنی‌داری ۰/۰۵ آزمون می‌کنیم.

۱- **روش ناحیه رد:** با توجه به قاعده تصمیم‌گیری $|T| > t_{\frac{\infty}{2},(n-1)}$ برای رد فرض H_0: مقدار $t_{\frac{0.05}{2},(19)}$ برابر ۱/۷۳ است، و چون ۰/۰۷۹ کمتر از ۱/۷۳ است، لذا در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم.

۲- **روش P-مقدار:** همان‌طور که می‌دانیم قاعده در این روش به این‌صورت است که، "اگر P-مقدار کمتر از α باشد، فرض H_0 را در سطح معنی‌داری α رد می‌کنیم". در این مثال P-مقدار برابر ۰/۹۴ است و کمتر از ۰/۰۵ نیست، لذا در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم.

۳- **روش فاصله اطمینان:** در این‌جا فرضیه $H_0: \mu - 15 = 0$، که معادل است با $H_0: \mu = 15$، را بررسی می‌کنیم. بنابراین برای آزمون فرضیه فوق به روش فاصله اطمینان کافیست که چک کنیم آیا فاصله اطمینان $\mu - 15$ عدد صفر را شامل می‌شود؟ در صورتی که فاصله اطمینان شامل عدد صفر باشد، تصمیم به عدم رد فرض H_0 خواهیم گرفت و با توجه به فاصله اطمینان ۹۵٪ در این مثال، در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم.

مثال ۲) در مثال ۱ فرضیه $H_0: \mu = 12$ را آزمون کنید.

در این حالت همه مراحل بیان شده در مثال قبل را انجام خواهیم داد، با این تفاوت که عدد ۱۲ را در جعبه Test Values تایپ می‌کنیم و خروجی زیر را خواهیم دید:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Score	20	14.9500	2.82796	.63235

One-Sample Test

	Test Value = 12					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Score	4.665	19	.000	2.95000	1.6265	4.2735

از سه روش می‌توان نشان داد که فرضیه $H_0: \mu = 12$ در سطح معنی‌داری ۰/۰۵ رد می‌شود(؟).

توجه:

۱- از آنجایی که در عمل هیچ‌گاه واریانس جامعه معلوم نیست، در SPSS نیز آزمون از طریق آماره Z تعبیه نشده است.

۲- در این آزمون فرض بر این است داده‌ها از جامعه‌ای با توزیع نرمال هستند.

۳- در SPSS، گزینه انجام آزمون‌های یک‌طرفه تعبیه نشده‌است. اما در آزمون‌هایی که آماره آنها از توزیع t، که توزیعی متقارن است، پیروی می‌کند، P-مقدار آزمون یک‌طرفه نصف P-مقدار آزمون دوطرفه است(؟). لذا در این آزمون نیز می‌توان آزمون‌های یک‌طرفه را انجام داد.

۹-۴ آزمون آماری برای نسبت جامعه – آزمون دوجمله‌ای

مثال ۳) فرض کنید سکه ای را ۳۰ مرتبه پرتاب کرده اید و نتایج به صورت زیر بدست آمده است (۱= شیر ، ۰= خط) :

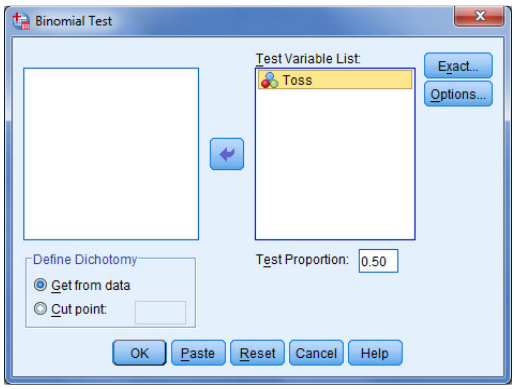
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1

می خواهیم آزمون کنیم که آیا سکه سالم است یا خیر ؟ می خواهیم فرض های زیر را آزمون کنیم:

$$\begin{cases} H_0 = p = 0.5 \\ H_1 = p \neq 0.5 \end{cases}$$

برای آزمون این فرض از آزمون دوجمله‌ای استفاده می‌کنیم.

برای انجام این کار داده‌های بالا را به عنوان متغیر Toss وارد کرده و مسیر Analyze>Nonparametric Tests>Legacy Diolog>Binomial را طی کنید تا کادر مکالمه‌ای Binomial Test باز شود.



متغیر Toss را به قسمت Test Variable List وارد کرده و مقدار Test Proportion را ۰/۵ وارد کنید و در قسمت Define Dichotomy گزینه Get from data را انتخاب کنید و سپس دکمه Ok را کلیک کنید تا خروجی زیر ظاهر شود :

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Toss	Group 1	1.00	12	.40	.50	.362
	Group 2	.00	18	.60		
	Total		30	1.00		

معرفی جداول: در ستون اول دو گروه جامعه را نشان می‌دهد و در ستون دوم آن‌ها را معرفی می‌کند. توجه داشته‌باشید که در SPSS هر گروهی که کدش اول وارد شود را به نام گروه اول(اصلی) می‌شناسد و نسبت P برای این گروه تعریف می‌شود. برای مثال اگر اولین عدد در وارد کردن داده‌های مربوط به متغیر Toss را "۱" وارد کنیم، عدد "۱" را به عنوان کد گروه اول می‌شناسد و نسبت P را برای این گروه (شیر بودن) و نسبت 1-P را برای گروه دیگر (خط بودن) درنظر می‌گیرد. در ستون سوم، N، تعداد افراد مشاهده‌شده در هر گروه را نشان می‌دهد و در ستون بعد نسبت مشاهده شده هر گروه را نشان می‌دهد. در ستون پنجم، Test Prop.، عدد مورد آزمون در فرض صفر را نشان می‌دهد که در اینجا ۰/۰۵ است. در ستون آخر، Exact Sig.(2-tailed)، P–مقدار دقیق را در حالت دوطرفه نشان می‌هد.

تفسیر: با توجه به اینکه P–مقدار دقیق دوطرفه بیشتر از ۰/۰۵ است، فرض صفر، سالم بودن سکه، را رد نمی‌کنیم.

مثال ۴) (مثال ۴ از فصل ۸) در یک نمونه‌ی تصادفی ۱۰۰۰۰ نفری از جامعه‌ی بزرگ، تعداد ۱۰۰ مورد از یک بیماری خاص مشاهده شده، در سطح اطمینانِ ۹۵ درصد آزمون کنید که:

الف) آیا نسبت مبتلایان به بیماری خاص در جامعه، از ۰/۰۲ کم‌تر است؟

ب) آیا می‌توان گفت نسبت مبتلایان به بیماری خاص در جامعه، برابر ۰/۰۲ است یا خیر؟

- این داده‌ها با استفاده از دستور weighte cases وارد کنید.

الف) تمام مراحل مربوط به مثال ۱ را در این‌جا انجام می‌دهیم، با این تفاوت که در قسمت Test Proportion از کادر مکالمه‌ای Binomial Test عدد ۰/۰۲ را جایگزین عدد ۰/۵ می‌کنیم و سپس دکمه Ok را می‌زنیم و خروجی را به‌صورت زیر خواهیم داشت.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
X	Group 1	1.00	100	.01	.02	.000[a]
	Group 2	2.00	9900	.99		
	Total		10000	1.00		

توجه داشته باشید که اگر مقدار Test Proportion از ۰/۵کمتر باشد، آزمون به صورت یک‌طرفه انجام می‌شود و فرض H_1 به‌صورت $P < 0.02$ خواهد بود. لذا با توجه به این‌که P–مقدار از ۰/۰۵ کمتر است، فرض صفر را رد نمی‌کنیم.

ب) به علت وجود تقارن توزیع نرمال و این‌که در این حالت SPSS آزمون دوطرفه را انجام نمی‌دهد، برای انجام آزمون دوطرفه P–مقدار آزمون یک‌طرفه را دوبرابر می‌کنیم. در این حالت باز هم می‌بینیم که P–مقدار از ۰/۰۵ کمتر است. بنابراین فرض صفر را رد نمی‌کنیم.

توجه:

آزمون‌هایی که در بالا انجام شده، از طریق توزیع نرمال استاندارد است. اما می‌دانیم که صورت مساله برای یک توزیع گسسته (دو جمله‌ای) مطرح شده‌بود. لذا نتایج آزمون‌های فوق همواره به صورت مجانبی(تقریبی) بوده و برای نمونه‌ای بزرگ صحیح است. اگر بخواهیم آزمون را در نمونه‌ای کوچک انجام دهیم، بهتر است از روش‌های شبیه‌سازی مونت‌کارلو استفاده کنیم. برای این‌کار می‌توانیم از گزینه Exact در کادر مکالمه‌ای Binomial Test استفاده کنیم و گزینه Monte Carlo را تیک زده و سپس دکمه‌های Continue و Ok را بزنیم.

۴-۱۰ آزمون اختلاف میانگین‌ها برای دو جامعه مستقل – آزمون t–دو نمونه مستقل

این آزمون زمانی به‌کار می‌رود که بخواهیم میانگین یک متغیر کمی را در بین دو گروه مستقل با هم مقایسه کنیم. برای مثال مقایسه فشارخون دو گروه افراد بعد از استفاده از داروهای A و B .

در آزمون t–دو نمونه مستقل فرضیات زیر را مورد بررسی قرار می‌دهیم:

$$\begin{cases} H_0 : \mu_2 - \mu_1 = 0 \\ H_1 : \mu_2 - \mu_1 \neq 0 \end{cases}$$

برای استفاده صحیح از این آزمون نیاز به اطلاع در مورد برابری و یا نابرابری واریانس‌های دو گروه خواهیم داشت. لذا ابتدا بایستی با استفاده از آزمون Levene برابری واریانس‌های دو گروه را مورد بررسی قرار داد. دو حالت می‌تواند پیش‌آید که هر دو حالت در SPSS به‌طور همزمان اجرا می‌شود.

۱- **برابری واریانس‌ها:** چنانچه P–مقدار حاصل از آزمون Levene بیشتر از α باشد می‌توان نتیجه گرفت که واریانس‌های دو گروه با هم برابر است. در این حالت برای بررسی معنی‌داری اختلاف بین میانگین‌های دو گروه، از آزمون "t–دو نمونه مستقل در حالت برابری واریانس‌ها" استفاده می‌کنیم.

۲- **نابرابری واریانس‌ها:** چنانچه P–مقدار حاصل از آزمون Levene کمتر یا مساوی α باشد می‌توان نتیجه گرفت که واریانس‌های دو گروه با هم برابر نیستند. در این حالت برای بررسی معنی‌داری اختلاف بین میانگین‌های دو گروه از آزمون "t–دو نمونه مستقل در حالت نابرابری واریانس‌ها" استفاده می‌کنیم.

مثال ۵) ده دانش‌آموز پسر و ده دانش‌آموز دختر به دلخواه انتخاب شده‌اند و وزن آن‌ها اندازه‌گیری شده‌است. داده‌ها به‌صورت زیرند:

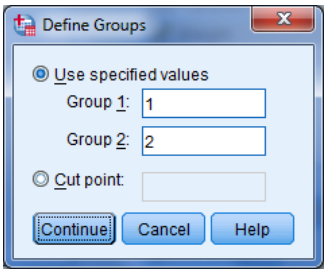
دختر	۵۴	۴۸	۶۵	۶۰	۴۵	۵۷	۴۵	۶۰	۶۳	۵۵
پسر	۷۵	۶۸	۸۰	۷۴	۷۲	۶۸	۷۰	۸۵	۹۰	۷۵

مایلیم بدانیم آیا وزن دانش‌آموزان دختر و پسر از لحاظ آماری با هم برابر است؟

ابتدا داده‌ها را به صورت زیر در SPSS وارد کنید. مردان با کد ۱ و زنان با کد ۲ مشخص شده‌اند.

	Sex	Weight
1	Male	75.00
2	Male	68.00
3	Male	80.00
4	Male	74.00
5	Male	72.00
6	Male	68.00
7	Male	70.00
8	Male	85.00
9	Male	90.00
10	Male	75.00
11	Female	54.00
12	Female	48.00
13	Female	65.00
14	Female	60.00
15	Female	45.00
16	Female	57.00
17	Female	45.00
18	Female	60.00
19	Female	63.00
20	Female	55.00

برای انجام این آزمون مسیر Analyze>Compare Means>Independent-Samples T Test را طی کنید، تا کادر مکالمه‌ای Independent-Samples T Test باز شود. در این کادر متغیر "Weight" را در جعبه Test Variable(s) و متغیر "Sex" را در جعبه Grouping Variable وارد کنید. روی دکمه Define Groups کلیک کنید تا کادر زیر باز شود:



مطابق تصویر مقادیری را که مشخص کننده گروه‌ها هستند، تعریف کنید (در این‌جا ۱ و ۲) و دکمه Continue و سپس Ok را کلیک کنید تا خروجی به صورت زیر ظاهر شود:

۱

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Weight	Male	10	75.7000	7.28850	2.30483
	Female	10	55.2000	7.20802	2.27938

۲ ۳

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	.015	.905	6.324	18	.000	20.50000	3.24157	13.68971	27.31029
	Equal variances not assumed			6.324	17.998	.000	20.50000	3.24157	13.68965	27.31035

معرفی جدول: این جدول شامل ۳ قسمت است:

۱- **Group Statistics :** شامل تعداد نمونه و شاخص‌های آماری به تفکیک هر یک از گروه‌هاست.

۲- **Levene's Test :** شامل آزمون Levene برای آزمون برابری واریانس‌ها (آمارهٔ F و P-مقدار)

۳- **t-test for Equality of Means :** نتایج آزمون t-دونمونه‌ای مستقل است که ستون‌های آن به‌ترتیب از چپ به راست عبارتند از: مقدار آمارهٔ t، درجه آزادی (df)، P- مقدارِ آزمون دوطرفه (sig(two-tailed)، اختلاف میانگین‌ها، یعنی $\overline{Y}-\overline{X}$ (Difference Mean)، خطای استاندارد اختلاف میانگین‌ها (Std. Error Difference) و فاصله اطمینان ۹۵٪ برای $\mu_2-\mu_1$ (95% Confidence Interval of the Difference). همه این نتایج برای دو حالت بیان شده‌اند. مقادیر خط اول برای حالت برابری واریانس‌ها و مقادیر خط دوم برای حالت نابرابری واریانس‌ها هستند.

تفسیر: P-مقدار آزمون Levene نشان می‌دهد که بین واریانس وزن دانش‌آموزان پسر و دختر اختلاف معنی‌داری وجود ندارد. لذا باید نتایج آزمون t را از خط دوم قسمت مربوطه گزارش کرد. در این‌جا نیز می‌توانیم از سه روش ناحیه رد، P-مقدار، فاصله اطمینان آزمون فرض را انجام دهیم.

۱- **روش ناحیه رد:** با توجه به قاعده تصمیم‌گیری $|T|\geq t_{\frac{\infty}{2},(n_1+n_2-2)}$ برای رد فرض H_0 : مقدار $t_{\frac{0.05}{2},(18)}$ برابر ۱/۷۳ است، و چون ۶/۳۲ بیشتر از ۱/۷۳ است، لذا در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را قبول نمی‌کنیم و معنی‌داری اختلاف را می‌پذیریم.

۲- **روش P-مقدار:** در این مثال P-مقدار ۰/۰۰۰ گزارش شده است و لذا کمتر از ۰/۰۵ است، بنابراین در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را قبول نمی‌کنیم و معنی‌داری اختلاف را می‌پذیریم.

۳- **روش فاصله اطمینان:** در این‌جا فرضیه $H_0:\mu_1-\mu_2=0$، بررسی می‌کنیم. بنابراین برای آزمون فرضیه فوق به روش فاصله اطمینان کافیست که چک کنیم آیا فاصله اطمینان $\mu_1-\mu_2$ عدد صفر را شامل می‌شود؟ با توجه به فاصله اطمینان ۹۵٪ در این مثال، در سطح معنی‌داری ۰/۰۵ فرض H_0، یعنی فرض برابری، را قبول نمی‌کنیم و معنی‌داری اختلاف را می‌پذیریم.

توجه:

۱- فرض صفر مطرح شده در این آزمون در واقع همان فرض صفر $\mu_2-\mu_1=k$ است که در آن k=0 است.

۲- در این آزمون نیز فرض بر این است داده‌ها، در هریک از گروه‌ها، از جامعه‌ای با توزیع نرمال هستند.

۳- در آزمون Levene فرض‌های زیر آزمون می‌شوند و آمارهٔ آزمون دارای توزیع F است:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_0 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

۴- همان‌طور که می‌بینید، این آزمون دوطرفه است. در خروجی این آزمون نیز می‌توان با نصف کردن P-مقدار، آزمون‌های یک‌طرفه را انجام داد.

۴-۱۱ آزمون اختلاف میانگین‌ها برای دو جامعه وابسته – آزمون t زوجی

همان‌طور که می‌دانید، وقتی می‌خواهیم میانگین یک متغیر کمی را در دو گروه وابسته، مقایسه کنیم، از آزمون t-زوجی استفاده خواهیم کرد.

در این حالت فرض‌های آماری عبارتند از:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 = \mu_1 - \mu_2 = 0 \end{cases}$$

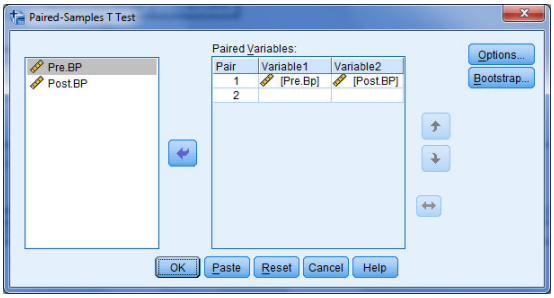
و از آنجایی که در عمل همواره واریانس جامعه نامعلوم است، آزمون مناسب، آزمون t-زوجی می‌باشد.

مثال ۶) مثال ۶ (فشار خون ده نفر) از فصل هشتم کتاب را درنظر بگیرید. در این می‌خواهیم بدانیم آیا رژیم غذایی جدید بر فشارخون بیماران تاثیر داشته است یا خیر؟

ابتدا داده‌ها را به‌صورت زیر در SPSS وارد کنید (فشار خون قبل از رژیم غذایی=PreBP، فشار خون بعد از رژیم غذایی=Post.BP)

	Pre.BP	Post.BP
1	170.00	140.00
2	170.00	160.00
3	140.00	150.00
4	140.00	160.00
5	170.00	150.00
6	160.00	130.00
7	160.00	110.00
8	140.00	140.00
9	170.00	160.00
10	180.00	180.00

برای انجام این آزمون مسیر Analyze>Compare Means>Paired-Samples T Test را طی کنید، تا کادر مکالمه‌ای Paired-Samples T Test باز شود. در این کادر متغیرهای "Pre.BP" و "Post.BP"را به ترتیب در جعبه Paired Variables، و در قسمت‌های Variable1 و Variable2 وارد کنید.



و سپس Ok را کلیک کنید تا خروجی به صورت زیر ظاهر شود:

۱

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre.BP	160.0000	10	14.90712	4.71405
	Post.BP	148.0000	10	19.32184	6.11010

۲

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Pre.BP & Post.BP	10	.270	.451

۳

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Pre.BP - Post.BP	12.00000	20.97618	6.63325	-3.00545	27.00545	1.809	9	.104

معرفی جدول: این جدول شامل ۳ قسمت است:

۱- **Paired Samples Statistics :** شامل تعداد نمونه و شاخص‌های آماری مربوط به فشار خون افراد در قبل و بعد از رژیم غذایی است.

۲- **Paired Samples Correlations :** ضریب همبستگی پیرسون و آزمون مربوط به آن است که در فصل دهم در مورد آن بحث خواهیم کرد.

۳- **Paired Samples Test:** نتایج آزمون t-زوجی است که ستون‌های آن به‌ترتیب از چپ به راست عبارتند از: میانگین اختلاف‌ها یا همان $\overline{D}$ (Mean) است، انحراف معیار اختلاف‌ها یا $\sqrt{S_D^2}$ (Std.Deviation) است، خطای استاندارد میانگین اختلاف‌ها(Std. Error Mean)، فاصله اطمینان ۹۵٪ برای μ_D (95% Confidence Interval of the Difference) است، مقدار آمارهٔ t، درجه آزادی (df)، P- مقدارِ آزمون دوطرفه ((sig(two-tailed).

تفسیر: در این‌جا نیز می‌توانیم از سه روش آزمون فرض را انجام دهیم.

۱- **روش ناحیه رد:** با توجه به قاعده تصمیم‌گیری $|T| \geq t_{\frac{\alpha}{2},(n-1)}$ برای رد فرض H_0 : مقدار $t_{\frac{0.05}{2},(9)}$ برابر ۲/۲۶ است، و چون ۱/۸۱ کمتر از ۲/۲۶ است، پس فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم، به عبارتی می‌توانیم با احتمال ۹۵ درصد قضاوت کنیم که رژیم غذایی جدید بر فشار خون تاثیر ندارد.

۲- **روش P-مقدار:** در این مثال P-مقدار ۰/۱۰ گزارش شده است و چون بیشتر از ۰/۰۵ است، پس فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم، به عبارتی می‌توانیم با احتمال ۹۵ درصد قضاوت کنیم که رژیم غذایی جدید بر فشار خون تاثیر ندارد.

۳- **روش فاصله اطمینان:** در این‌جا فرضیه $H_0 : \mu_1 - \mu_2 = 0$ را بررسی می‌کنیم. بنابراین برای آزمون فرضیه فوق به روش فاصله اطمینان کافیست که چک کنیم آیا فاصله اطمینان $\mu_1 - \mu_2$ عدد صفر را شامل می‌شود؟ با توجه به فاصله اطمینان ۹۵٪ در این مثال، پس فرض H_0، یعنی فرض برابری، را رد نمی‌کنیم، به عبارتی می‌توانیم با احتمال ۹۵ درصد قضاوت کنیم که رژیم غذایی جدید بر فشار خون تاثیر ندارد.

توجه:

۱- در آزمون t-زوجی، منظور از وابستگی دو جامعه، صرفاً یکسان بودن دو جامعه یا تکرار دوبار یک جامعه نیست. بلکه ممکن است دوجامعه به واسطه "جورشدن" نمونه‌ها نیز وابسته باشند.

۲- در این آزمون نیز فرض بر این است که داده‌ها از جامعه‌ای با توزیع نرمال هستند.

۳- آزمون معادل آزمون t-تک نمونه‌ای روی جامعه اختلاف‌ها با $\mu_0 = 0$ است.

۴- همان‌طور که می‌بینید، این آزمون در حالت دوطرفه اجرا شده‌است. در خروجی این آزمون نیز می‌توان با نصف کردن P-مقدار، آزمون‌های یک‌طرفه را انجام داد.

تمرین:

۱- اداره بهداشت یک شهر می‌خواهد تعیین کند که آیا میانگین تعداد باکتری‌ها در واحد حجم آب شهر از سطح ایمنی یعنی ۲۰۰ بیشتر است یا نه. پژوهشگران ۱۰ نمونه از آب، هر یک به حجم واحد، را گرد آوری کرده و دیده‌اند که تعداد باکتری‌ها عبارتند از :

۱۸۴,۱۹۸,۲۱۵,۱۹۰,۱۷۵,۱۸۰,۱۹۶,۱۹۳,۲۱۰,۲۰۷

آیا داده‌ها دلیلی بر نگرانی به‌دست می‌دهند؟

۲- از ۲۰ راننده تاکسی در یک شهر در مورد درآمد روزانهٔ آن‌ها سئوال شده‌است. داده‌های حاصل به‌صورت زیرند (بر حسب هزار تومان):

۱۶۰،۱۵۰،۱۴۵،۱۵۵،۱۶۰،۱۴۰،۱۳۵،۸۰،۱۱۰،۲۶۰،۱۲۰،۱۶۰،۱۵۰،۲۱۰،۲۵۰،۲۲۰،۱۳۰ ۱۳۰،۱۵۰،۱۷۰

سازمان تاکسیرانی این شهر اعلام کرده‌است که متوسط درآمد رانندگان تاکسی در این شهر ۱۷۵ هزار تومان است. آیا این ادعا صحیح است؟

۳- دو نوع اسباب بازی مکانیکی و غیرمکانیکی داریم. می‌خواهیم ببینیم که آیادرکودکان ۵ ساله تمایل به سمت هرکدام از این اسباب بازی‌ها یکسان است؟ بدین منظور از ۲۰۰ کودک ۵ ساله خواسته‌ایم که به دلخواه یکی از این اسباب‌بازی ها را انتخاب کنند و ۱۲۰ کودک اسباب‌بازی مکانیکی و ۸۰ کودک اسباب‌بازی غیرمکانیکی را انتخاب کردند. از این داده‌ها چه استنباطی دارید؟

۴- وزن ۲۰ بازیکن فوتبال که به تصادف از تیمهای فوتبال یک شهر انتخاب شده اند به صورت زیر است:

۸۱	۷۱/۵	۷۵	۶۴
۸۳	۷۰	۶۹/۵	۶۷
۹۴	۵۸/۵	۹۵	۹۰
۹۷	۸۸	۶۲	۵۹
۸۰	۶۱/۵	۶۰	۵۹/۵

آیا می‌توان گفت که :

الف -۶۰ درصد از بازیکنان فوتبال این شهر کمتر از $65kg$ وزن دارند؟

ب- ۴۰ درصد از بازیکنان فوتبال این شهر بیشتراز $60kg$ وزن دارند ؟

۵- به منظور مقایسه دو برنامه جهت آموزش کارگران صنعتی برای انجام کاری تخصصی، ۲۰ کارگر در آزمایشی شرکت داده می شوند. از بین آنها به طور تصادفی ۱۰ نفر را برای آموزش به وسیله روش ۱ انتخاب می کنند، ۱۰ نفر بقیه را با روش ۲ آموزش می دهند. بعد از تکمیل دوره آموزشی، همه کارگران در معرض یک آزمون زمان و حرکت قرار می گیرند که سرعت انجام یک کار تخصصی را ثبت می کند. داده های زیر به دست آمده اند:

روش۱	۱۵	۲۰	۱۱	۲۳	۱۶	۲۱	۱۸	۱۶	۲۷	۲۴
روش۲	۲۳	۳۱	۱۳	۱۹	۲۳	۱۷	۲۸	۲۶	۲۵	۲۸

الف) آیا از این داده ها می توان نتیجه گرفت که میانگین زمان کار بعد از آموزش با روش ۱ به طور معنی دار از میانگین زمان کار بعد از آموزش با روش ۲ کمتر است؟ (با $\alpha = ۰/۰۵$ آزمون کنید)

ب)فرضهایی را که برای توزیع های جامعه در نظر می گیرید بیان کنید.

ج)برای تفاضل میانگین جامعه های زمانهای کار بین دو روش، یک فاصله اطمینان ۹۵٪ بنا کنید.

۶- یک فرایند متداول تمیز کردن که در صنعت کمپوت سازی به کار می رود عبارت از شستن سبزیجات با حجم زیادی از آب جوش قبل از کمپوت کردن است. روش جدیدی به تازگی پیدا شده است که فرایند تمیز کردن با بخار نام دارد (steam blanching process =SBP) و انتظار می رود در این روش ویتامینها و مواد معدنی سبزیجات کمتر از دست برود و این به دلیل جاری بودن آب نسبت به بخار است . قرار است ده دسته لوبیا از مزارع مختلف برای مقایسه فرایند SBP و فرایند متعارف مورد استفاده قرار گیرند. نصف هر دسته از لوبیا ها را با فرایند متعارف و نصف دیگر هر دسته را با SBP تمیز می کنند . اندازه های محتوای ویتامینها در هر نیم کیلو لوبیای کمپوت شده عبارت اند از :

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
SBP	۳۵	۴۸	۶۵	۳۳	۶۱	۵۴	۴۹	۳۷	۵۸	۶۵
روش متعارف	۳۳	۴۰	۵۵	۴۱	۶۲	۵۴	۴۰	۳۵	۵۹	۵۶

الف)آیا داده ها برای این ادعا که در روش SBP نسبت به روش متعارف تمیز کردن ویتامینهای کمتری از دست می رود گواهی قوی فراهم می کنند ؟

ب) برای تفاضل بین میانگین های محتوای ویتامین در هر نیم کیلو با ا ستفاده از دو روش تمیز کردن یک فاصله اطمینان ۹۸٪ بسازید.

۷- مثال ۶ را با استفاده از آزمون t-تک‌نمونه‌ای حل کنید. (راهنمایی: با استفاده از دستور Compute متغیر D را بسازید و آزمون t-تک‌نمونه‌ای را اجرا کنید)

۸- ادعا شده است که یک برنامه ایمنی صنعتی در کاهش تضییع ساعات کار ناشی از نقص در ماشینهای کارخانه مؤثر است . داده های زیر مربوط به ضایع شدن ساعتهای کار هفتگی به واسطه نقص در ۶ دستگاه است که یکی قبل و دیگری بعد از اجرای برنامه ایمنی جمع آوری شده اند.

	۱	۲	۳	۴	۵	۶
قبل	۱۲	۲۹	۱۶	۳۷	۲۸	۱۵
بعد	۱۰	۲۸	۱۷	۳۵	۲۵	۱۶

هر فرضی که در نظر می گیرید بیان کنید. آنگاه آزمون فرضی مناسب را طرح نمایید و تعیین کنید که آیا داده ها از ادعای مطروحه حمایت می کنند یا نه .

۹- بر اساس یک دوره ورزش مرتب به ۸ دانش‌آموز طی یک‌ماه، می‌خواهیم بدانیم آیا این نوع ورزش تفاوتی معنی‌دار در وزن دانش‌آموزان یک مدرسه ایجاد کرده است یا نه؟ داده‌های زیر وزن دانش‌آموزان را در قبل از انجام دوره و بعد از آن نشان می‌دهد.

قبل از دوره	۴۸	۵۳	۳۵/۵	۴۶	۵۹	۶۱	۴۲	۴۸
بعد از دوره	۴۶/۵	۵۲	۳۶	۴۲/۵	۵۶/۵	۶۰	۴۰/۵	۴۷

در سطح ۰/۰۵ آزمون کنید که آیا وزن دانش آموزان قبل و بعد از یک ماه ورزش منظم یکسان است یا خیر؟

ضمیمه

تمرینات تکمیلی

هر یک از مثال‌های زیر را با استفاده از SPSS حل کرده و خروجی را تفسیر کنید.

۱-پزشک محققی می خواهد تعیین کند که آیا قرص معینی اثر جانبی نامطلوب تقلیل فشار خون را بر روی مصرف کننده دارد یا نه ؟ این مطالعه شامل ثبت فشار خون اولیه پانزده خانم هیجده ساله است. بعد از اینکه این خانمها برای مدت شش ماه مرتباً قرص را مصرف می کنند فشار خون آنها دوباره ثبت می شود. محقق مایل است درباره اثر قرص روی فشار خون به کمک مشاهدات داده شده در جدول زیر استنباطهایی انجام دهد.

جدول اندازه های فشار خون قبل و بعد از مصرف قرص آزمودنی

		۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
(X)	قبل	۷۰	۸۰	۷۲	۷۶	۷۶	۷۶	۷۲	۷۸	۸۲	۶۴	۷۴	۹۲	۷۴	۶۸	۸۴
(Y)	بعد	۶۸	۷۲	۶۲	۷۰	۵۸	۶۶	۶۸	۵۲	۶۴	۷۲	۷۴	۶۰	۷۴	۷۲	۷۴
D=(x-y)		۲	۸	۱۰	۶	۱۸	۱۰	۴	۲۶	۱۸	-۸	۰	۳۲	۰	-۴	۱۰

۲- به منظور مقایسه دو برنامه جهت آموزش کارگران صنعتی برای انجام کاری تخصصی، ۲۰ کارگر در آزمایشی شرکت داده می شوند. از بین آنها به طور تصادفی ۱۰ نفر را برای آموزش به وسیله روش ۱ انتخاب می کنند، ۱۰ نفر بقیه را با روش ۲ آموزش می دهند. بعد از تکمیل دوره آموزشی، همه کارگران در معرض یک آزمون زمان و حرکت قرار می گیرند که سرعت انجام یک کار تخصصی را ثبت می کند. داده های زیر به دست آمده اند:

زمان به دقیقه

روش۱	۱۵	۲۰	۱۱	۲۳	۱۶	۲۱	۱۸	۱۶	۲۷	۲۴
روش۲	۲۳	۳۱	۱۳	۱۹	۲۳	۱۷	۲۸	۲۶	۲۵	۲۸

الف) آیا از این داده ها می توان نتیجه گرفت که میانگین زمان کار بعد از آموزش با روش ۱ به طور معنی دار از میانگین زمان کار بعد از آموزش با روش ۲ کمتر است؟(با $\alpha = ۰/۰۵$ آزمون کنید)

ب)فرضهایی را که برای توزیع های جامعه در نظر می گیرید بیان کنید.

ج)برای تفاضل میانگین جامعه های زمانهای کار بین دو روش، یک فاصله اطمینان ۹۵٪ بنا کنید.

۳-گمان می رود روشی جدید در مقایسه با روش متعارفی که در حال حاضر برای آموزش خواندن به نو آموزان مورد استفاده است، به نحو بارزی قدرت خواندن را بهبود بخشد. برای آزمون این گمان، ۱۶ کودک به

طور تصادفی به دو گروه هشت نفری تقسیم می شوند. یک گروه با استفاده از روش متعارف و گروه دیگر با استفاده از روش جدید ، آموزش می بینند. نمرات کودکان در یک امتحان به صورت زیرند:

نمرات امتحان خواندن

روش متعارف	۶۵	۷۰	۷۶	۶۳	۷۲	۷۱	۶۸	۶۸
روش جدید	۷۵	۸۰	۷۲	۷۷	۶۹	۷۱	۷۱	۷۸

الف) داده ها را تجزیه و تحلیل کنید و درباره تفاضل میانگین جامعه های نمرات حاصل با استفاده از روشهای آموزشی متعارف و جدید استنباطی انجام دهید.

ب)فرضهایی را که برای تجزیه و تحلیل خود در نظر گرفته اید بیان کنید.

۴-جامعه شناسی مایل به مقایه نرخ باروری زنها در دو فرقه قبیله ای A و B در آفریقای غربی است. از هر فرقه نمونه ای تصادفی شامل ۱۰۰ زن در گروه سنی ۵۰ تا ۶۰ ساله را در بر می گیرند و تعداد کودکان متولد شده از هر کدام را ثبت می کند. توزیع های فراوانی زیر به دست آمده اند:

تعداد کودکان

		۰	۱	۲	۳	۴	۵	۶	۷	۸	جمع
فراوانی	A	۶	۱۴	۱۸	۲۵	۱۹	۱۱	۵	۲	۳	۱۰۰
	B	۰	۳	۸	۱۸	۳۰	۱۹	۱۵	۵	۲	۱۰۰

الف)میانگین و انحراف معیار را برای هر توزیع فراوانی محاسبه کنید.

ب) آیا داده ها بین میانگین های تعداد کودکان متولد شده از زنهای دو فرقه اختلاف معناداری را نشان می دهند؟

ج)برای تفاضل بین میانگین ها، یک فاصله اطمینان ۹۸٪ بنا کنید.

۵-برای مقایسه برتری یکی از دو روش حفظ کردن مطالب مشکل، آزمایشی انجام می شود ، برای این منظور ۹ زوج از دانش آموزان در آزمایش شرکت داده می شوند. دانش آموزان بر طبق ضریب هوشی (IQ) و زمینه علمی، زوج می شوند، و آنگاه به تصادف یکی از دو روش را بر روی یک عضو از هر زوج و روش دیگر را بر روی عضو دیگر زوج به کار می برند، سپس یک آزمون حافظه به تمام دانش آموزان داده می شود، و نمره های زیر را به دست می آورند:

زوجها

	۱	۲	۳	۴	۵	۶	۷	۸	۹
روش A	۹۰	۸۶	۷۲	۶۵	۴۴	۵۲	۴۶	۳۸	۴۳
روش B	۸۵	۸۷	۷۰	۶۲	۴۴	۵۳	۴۲	۳۵	۴۶

برای تعیین اینکه آیا اختلاف معنی داری در کارایی دو روش وجود دارد یا نه، آزمونی را در سطح α=۰/۰۵ انجام دهید.

۶- یک فرایند متداول تمیز کردن که در صنعت کمپوت سازی به کار می رود عبارت از شستن سبزیجات با حجم زیادی از آب جوش قبل از کمپوت کردن است. روش جدیدی به تازگی پیدا شده است که فرایند تمیز کردن با بخار نام دارد (steam blanching process =SBP) و انتظار می رود در این روش ویتامینها و مواد معدنی سبزیجات کمتر از دست برود و این به دلیل جاری بودن آب نسبت به بخار است . قرار است ده دسته لوبیا از مزارع مختلف برای مقایسه فرایند SBP و فرایند متعارف مورد استفاده قرار گیرند. نصف هر دسته از لوبیا ها را با فرایند متعارف و نصف دیگر هر دسته را با SBP تمیز می کنند . اندازه های محتوای ویتامینها در هر نیم کیلو لوبیای کمپوت شده عبارت اند از :

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
SBP	۳۵	۴۸	۶۵	۳۳	۶۱	۵۴	۴۹	۳۷	۵۸	۶۵
روش متعارف	۳۳	۴۰	۵۵	۴۱	۶۲	۵۴	۴۰	۳۵	۵۹	۵۶

الف)آیا داده ها برای این ادعا که در روش SBP نسبت به روش متعارف تمیز کردن ویتامینهای کمتری از دست می رود گواهی قوی فراهم می کنند ؟

ب) برای تفاضل بین میانگین های محتوای ویتامین در هر نیم کیلو با ا ستفاده از دو روش تمیز کردن یک فاصله اطمینان ۹۸٪ بسازید.

۷-می خواهند مطالعه ای برای تأثیر نسبی دو نوع داروی سرانه در افزایش خواب انجام دهند. به شش نفر که سرماخوردگی دارند در شب اول داروی A و در شب دوم داروی B داده می شود و میزان ساعات خواب آنها در هر شب ثبت می گردد . داده ها عبارتند از :

	۱	۲	۳	۴	۵	۶
داروی A	۴/۸	۴/۱	۵/۸	۴/۹	۵/۳	۷/۴
داروی B	۳/۹	۴/۲	۵/۰	۴/۹	۵/۴	۷/۱

الف)برای میانگین افزایش ساعت خواب ناشی از داروی B نسبت به داروی A یک فاصله اطمینان ۹۵٪ بسازید.

ب) در این مطالعه چه چیزی را و چگونه تصادفی می کنید؟ به طور خلاصه دلایل خود را برای تصادفی کردن بیان کنید.

۸- برای مقایسه دو روش بافتن طناب، پنج زوج آزمون اجرا می شود. هر دسته نمونه شامل کنف کافی برای بافتن دو رشته طناب است. اندازه های قدرت کشش طنابها عبارت اند از :

	۱	۲	۳	۴	۵
روش ۱	۱۴	۱۲	۱۸	۱۶	۱۵
روش ۲	۱۶	۱۵	۱۷	۱۶	۱۴

الف) داده ها را به عنوان پنج زوج مشاهدات در نظر بگیرید، و برای تفاضل میانگینهای قدرت کشش بین رشته طنابهای بافته شده با دو روش، یک فاصله اطمینان ۹۵٪ به دست آورید .

ب) با در نظر گرفتن داده ها به عنوان نمونه های تصادفی مستقل، محاسبه یک فاصله اطمینان ۹۵٪ را تکرار کنید.

ج) به طور خلاصه شرایطی را که تحت آنها هر نوع تجزیه و تحلیلی مناسب است مورد بحث قرار دهید.

۹- ادعا شده است که یک برنامه ایمنی صنعتی در کاهش تضییع ساعات کار ناشی از نقص در ماشینهای کارخانه مؤثر است . داده های زیر مربوط به ضایع شدن ساعتهای کار هفتگی به واسطه نقص در ۶ دستگاه است که یکی قبل و دیگری بعد از اجرای برنامه ایمنی جمع آوری شده اند.

دستگاه

	۱	۲	۳	۴	۵	۶
قبل	۱۲	۲۹	۱۶	۳۷	۲۸	۱۵
بعد	۱۰	۲۸	۱۷	۳۵	۲۵	۱۶

هر فرضی که در نظر می گیرید بیان کنید. آنگاه آزمون فرضی مناسب را طرح نمایید و تعیین کنید که آیا داده ها از ادعای مطروحه حمایت می کنند یا نه .

۱۰- برای تعیین اینکه افزودن یک ماده شیمیایی مخصوص به یک کود متعارف ، رشد گیاه را تسریع می کند یا نه، آزمایشی انجام می شود. برای این مطالعه ده مکان را در نظر می گیرند. در هر مکان دو گیاه روییده در جوار هم را تیمار می کنند : به یکی کود متعارف و به دیگری کود متعارف همراه با ماده شیمیایی افزایشی می

دهنـــد. رشـــد گیاهـــان بعـــد از چهـــار هفتـــه بـــر حســـب ســـانتیمتر انـــدازه گیـــری می شود، و داده های زیر به دست می آیند :

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
بدون ماده افزایشی	۲۰	۳۱	۱۶	۲۲	۱۹	۳۲	۲۵	۱۸	۲۰	۱۹
با ماده افزایشی	۲۳	۳۴	۱۵	۲۱	۲۲	۳۱	۲۹	۲۰	۲۴	۲۳

آیا داده ها این ادعا را که استفاده از ماده شیمیایی افزایشی موجب تسریع در رشد گیاه می شود تأیید می کنند؟ فرضهایی را که در نظر می گیرید بیان نمایید ، و آزمون فرض مناسب را پیشنهاد کنید.

۱۱- اندازه های قدرت فشار دست راست و دست چپ ده نفر نویسنده چپ دست را ثبت کرده اند:

اشخاص

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
فشار دست چپ	۱۴۰	۹۰	۱۲۵	۱۳۰	۹۵	۱۲۱	۸۵	۹۷	۱۳۱	۱۱۰
فشار دست راست	۱۳۸	۸۷	۱۱۰	۱۳۲	۹۶	۱۲۰	۸۶	۹۰	۱۲۹	۱۰۰

الف) آیا برای افرادی که با دست چپ می نویسند ، داده ها گواهی قوی فراهم می کنند که قدرت فشار بیشتری در دست چپ دارند؟

ب) برای تفاضل میانگین ها یک فاصله اطمینان ۹۰٪ بنا کنید.

۱۲-علی ادعا می کند که (طعمه معجزه آسای) او برای یک نوع ماهی نسبت به طعمه متداول قدیمی که ابراهیم از آن استفاده می کند دام مؤثرتری است. علی و ابراهیم در تابستان گذشته با یک قایق برای مدت ۱۲ روز به ماهیگیری رفتند و هر یک تعداد ماهیهای مورد نظر صید شده خود را روز به روز ثبت کردند :

روزها

	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
علی	۸	۲۷	۷	۹	۱۸	۱۵	۱۲	۱۹	۳	۱۲	۱۸	۱۲
ابراهیم	۱۳	۲۰	۲	۹	۱۹	۱۲	۱۰	۲۳	۰	۱۱	۱۵	۱۰

آیا این داده ها به طور کافی از ادعای علی حمایت می کنند تا ابراهیم را در تعویض دام قدیمی به (طعمه معجزه آسای) علی که قدری گرانتر است متقاعد کند؟

۱۳-برای مقایسه دقت دو نوع آشکار ساز جیوه در اندازه گرفتن غلظت جیوه هوا، آزمایشی انجام می شود. اواسط روز در ناحیه مرکزی شهری هفت اندازه غلظت جیوه با وسیله نوع A و شش اندازه با وسیله نوع B گرفته می شود. اندازه های ثبت شده بر حسب واحد میکروگرم در هر متر مکعب هوا عبارت اند از:

نوع A	۰/۹۵	۰/۸۲	۰/۷۸	۰/۹۶	۰/۷۱	۰/۸۶	۰/۹۹
نوع B	۰/۸۹	۰/۹۱	۰/۹۴	۰/۹۱	۰/۹۰	۰/۸۹	

الف) آیا داده ها گواهی قوی بر این است که وسیله نوع B غلظت جیوه در هوا را دقیقتر از وسیله نوع A اندازه می گیرد؟

ب) برای نسبت انحراف معیارهای اندازه های تهیه شده با نوع A و با نوع B یک فاصله اطمینان ۹۰٪ بسازید.

۱۴-یکی از جنبه های مطالعه در اختلافهای جنسیت شامل مطالعه نحوه بازی میمونها در خلال سال اول زندگی است. شش میمون نر و شش میمون ماده در گروه هایی از چهار خانواده در طی چندین جلسه مطالعه ده دقیقه ای مشاهده شدند. میانگین کل تعداد دفعاتی که هر میمون ، بازی با همسال دیگر را شروع کرده است ثبت می شود:

نرها	۳/۶۴	۳/۱۱	۳/۸۰	۳/۵۸	۴/۵۵	۳/۹۲
ماده ها	۱/۹۱	۲/۰۶	۱/۷۸	۲/۰۰	۱/۳۰	۲/۳۲

الف) نمودارها مشاهدات را رسم کنید.

ب) برای تفاضل میانگین جامعه ها یک فاصله اطمینان ۹۵٪ بنا کنید.

۱۵- ج. ج. تامسن (۱۸۵۶-۱۹۴۰) در حالی که مشغول تحقیق درباره طبیعت پایه ای اشعه کاتدی بود الکترون را کشف کرد. تامسن در تجارب آزمایشگاهی ، ذرات باردار شده با بار منفی را جدا کرد تا بتواند نسبت جرم به بار را تعیین نماید. به نظر می رسید که این نسبت برای انواع وسیعی از شرایط آزمایشی مقدار ثابتی است و مشخصه ای از ذرات جدید است. تامسن نتایج زیر را با دو لامپ اشعه کاتدی مختلف، با به کار بردن هوا به عنوان گاز به دست آورد:

لامپ ۱	۰/۵۷	۰/۳۴	۰/۴۳	۰/۳۲	۰/۴۸	۰/۴۰	۰/۴۰
لامپ ۲	۰/۵۳	۰/۴۷	۰/۴۷	۰/۵۱	۰/۶۳	۰/۶۱	۰/۴۸

الف) برای تفاضل میانگین های دو لامپ یک فاصله اطمینان ۹۵٪ بسازید. آیا به نظر می رسد که دو لامپ نتایج سازگاری را فراهم می کنند.

ب) با کاربرد دو مجموعه از اندازه های تامسن به عنوان یک نمونه به حجم ۱۴، برای میانگین نسبت جرم به بار، یک فاصله اطمینان ۹۹٪ بنا کنید.

آزمونهای ناپارامتری:

- معادل ناپارامتری آزمون فرض در مورد میانگین های دو جامعه :

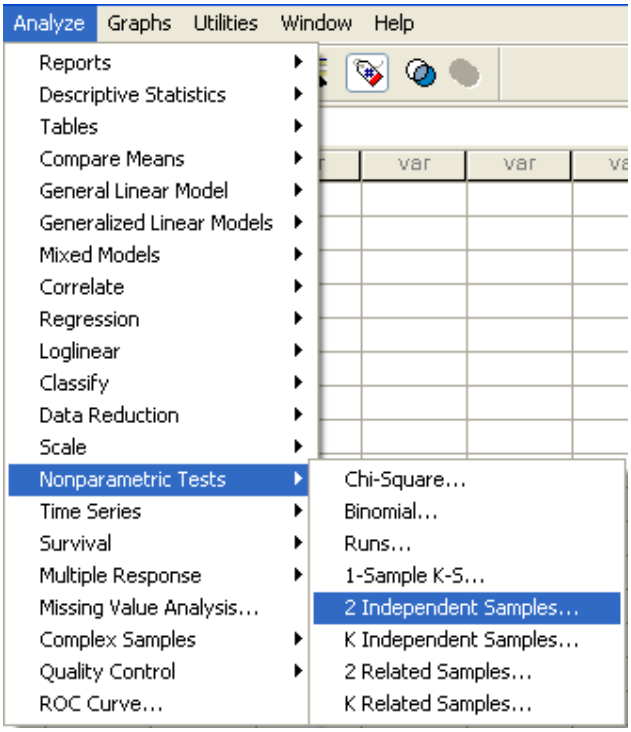
۱- آزمون من- ویتنی

زمانی که بخواهیم دو جامعه مستقل را به شیوه ناپارامتری آزمون و مقایسه کنیم, از آزمون من- ویتنی استفاده می کنیم .

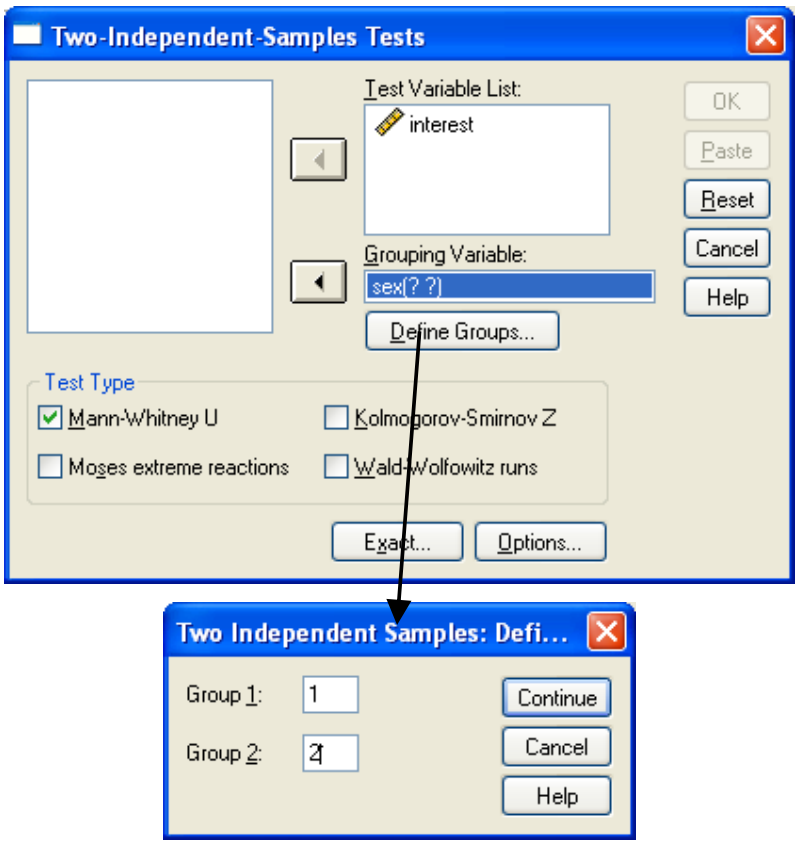
مثال : در یک بررسی می‌خواهیم ببینیم که آیا میزان علاقهٔ آقایان و خانمها به آشپزی یکسان است یا خیر؟ بدین منظور داده‌های زیر را از طریق پرسشنامه و دادن رتبه به میزان علاقهٔ افراد به آشپزی بدست آورده‌ایم:

خانمها	۵	۴	۴	۵	۳	۴	۵	۴	۳	۴
آقایان	۲	۳	۱	۲	۳	۵	۳	۴	۳	۲

برای انجام این مثال مسیر زیر را طی کنید:



متغیر مورد آزمون را به قسمت Test variable list ببرید و متغیری که گروهبندی را مشخص می کند به قسمت Grouping variable ببرید و مانند قبل گروهها را برای آن تعریف کنید.



بر دکمه **continue** کلیک کنید و در قسمت **Test type** گزینه **Mann-Whitney U** را تیک بزنید و پس از آن بر دکمه Ok کلیک کرده تا خروجی به صورت زیر ظاهر گردد.

Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
interest	۱,۰۰	۱۰	۱۳,۸۰	۱۳۸,۰۰
	۲,۰۰	۱۰	۷,۲۰	۷۲,۰۰
	Total	۲۰		

Test Statistics[b]

	interest
Mann-Whitney U	۱۷,۰۰۰
Wilcoxon W	۷۲,۰۰۰
Z	-۲,۵۷۷
Asymp. Sig. (۲-tailed)	.۰۱۰
Exact Sig. [۲*(۱-tailed Sig.)]	.۰۱۱[a]

a. Not corrected for ties.

b. Grouping Variable: sex

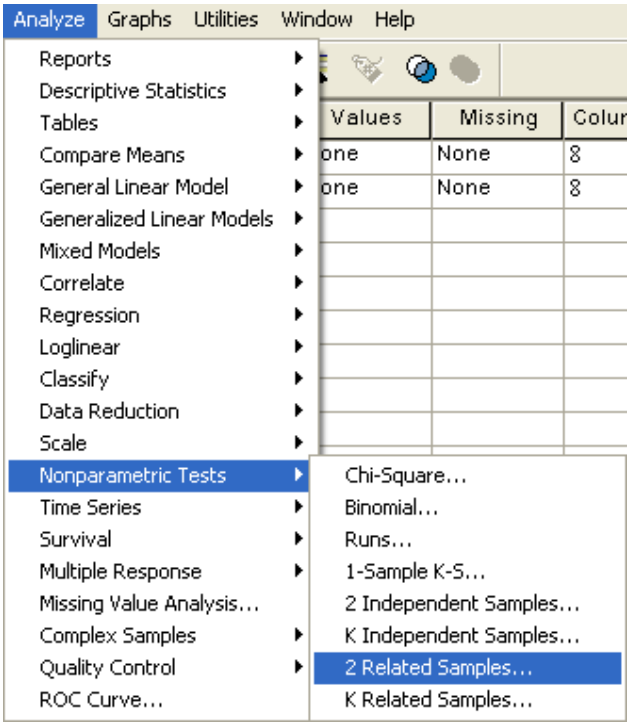
تذکر : در استفاده از آزمون من- ویتنی توجه داشته باشید که مقیاس اندازه گیری نمونه ها باید رتبه ای (با تعداد مقوله های کم) و یا فاصله ای باشد.

۲- آزمون های علامت, ویلکاکسون(آزمون رتبهٔ علامت دار) و مک نمار:

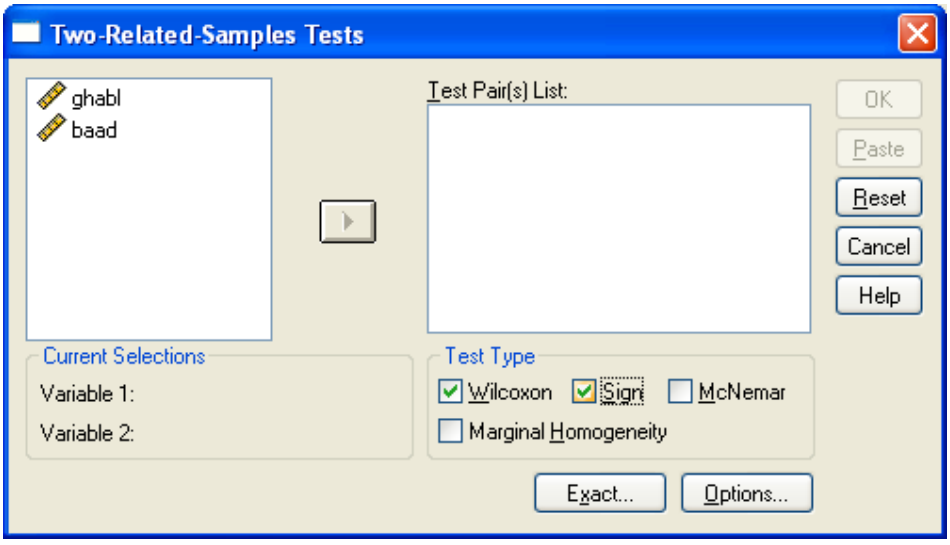
زمانی که بخواهیم میانگینهای نمونه های جفتی(زوجی) را از طریق آزمونهای ناپارامتری مقایسه و آزمون کنیم, از آزمون های **علامت, ویلکاکسون, و مک نمار** استفاده می کنیم. اما هر یک از این سه آزمون فرقهایی با یکدیگر دارند.

مثال (آزمون علامت و ویلکاکسون): می خواهیم مثال مربوط به وزن دانش آموزان (حاصل انجام یک ماده ورزش مرتب) را با استفاده از آزمون های ناپارامتری آزمون کنیم.

برای این کار مسیر زیر را طی کنید:



تا کادر مکالمه ای زیر باز شود .مانند حالت پارامتری دو متغیر ghabl و baad را انتخاب و آنها را به قسمت Test Pair(s) list منتقل کنید. در قسمت Test type گزینه های Wilcoxon و Sign را تیک دار کنید .



Ok را کلیک کنید تا نتایج بصورت زیر برای شما ظاهر شود :

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
baad – ghabl	Negative Ranks	7[a]	4.79	33.50
	Positive Ranks	1[b]	2.50	2.50
	Ties	0[c]		
	Total	8		

a. baad < ghabl

b. baad > ghabl

c. baad = ghabl

Test Statistics[b]

	baad – ghabl
Z	-2.200[a]
Asymp. Sig. (2-tailed)	.028

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Sign Test

Frequencies

		N
baad – ghabl	Negative Differences[a]	7
	Positive Differences[b]	1
	Ties[c]	0
	Total	8

a. baad < ghabl

b. baad > ghabl

c. baad = ghabl

Test Statistics[b]

	baad – ghabl
Exact Sig. (2-tailed)	.070[a]

a. Binomial distribution used.

b. Sign Test

تذکر: در مقایسه های زوجی در صورتی که دادهٔ پرت داشتیم (از نمودار پراکنش دو متغیر استفاده کنید) یا نمی توانستیم از آزمون t زوجی استفاده کنیم, می توانیم از روشهای ناپارامتری جهت آنالیز استفاده کنیم. توجه داشته باشید که آزمون علامت نسبت به داده های پرت کاملا ایمن است, در حالی که آزمون ویلکاکسون اینگونه نیست. (فرق آزمونهای علامت و ویلکاکسون)

تذکر: توجه داشته باشید که در آزمون های ویلکاکسون و علامت الزامی در مورد شکل توزیع نداریم ولی متغیر مورد بررسی می باید که پیوسته باشد.

مثال (آزمون مک نمار) :

یک محقق نظر افراد را در مورد موافق یا مخالف بودن افراد نسبت به یک کاندیدا را در قبل و بعد از سخنرانی مورد بررسی قرار می دهد. داده های حاصل به صورت زیر است:

بعد از انجام سخنرانی / قبل از انجام سخنرانی	موافق	مخالف
موافق	۲۰	۶۰
مخالف	۱۰	۱۰

می خواهیم بدانیم که آیا نظر افراد در دو وضعیت متفاوت است یا خیر؟

با استفاده از آزمون مک نمار و بدست آوردن خرجی زیر در SPSS متوانیم نتیجه بگیریم که نگرش افراد در قبل و بعد از سخنرانی متفاوت بوده است.(چرا؟)

McNemar Test

Crosstabs

before & after

before	after 1	after 2
1	20	60
2	10	10

Test Statistics[b]

	before & after
N	100
Chi-Square[a]	34.300
Asymp. Sig.	.000

a. Continuity Corrected

b. McNemar Test

تمرین :

در یک مطالعه مربوط به تعیین میزان توانایی تداعی پستانداران به جز انسان, ۱۹ میمون همسن به طور تصادفی به ۲ گروه ۹ و ۱۰ تایی تقسیم می شوند. دو گروه مزبور را برای به خاطر آوردن یک محرک مربوط به صدا, با دو روش آموزش مختلف تعلیم می دهند. نمرات میمون ها در آزمون بعد از آموزش عبارتند از :

روش۱	۱۶۷	۱۴۹	۱۳۷	۱۷۸	۱۷۹	۱۵۵	۱۶۴	۱۰۴	۱۵۱	۱۵۰
روش۲	۹۸	۱۲۷	۱۴۰	۱۰۳	۱۱۶	۱۰۵	۱۰۰	۹۵	۱۳۱	

آیا داده ها قویا تفاوت را در قدرت حافظه میمونهای تعلیم داده شده با دو روش نشان میدهد ؟

تطابق توزیع ها :

همانطور که در بخشهای پیش گفته شد یکی از اهداف بررسی یک جامعه این است که محقق بخواهد توزیع نمونه را با یک توزیع فرضی مانند توزیع نرمال بررسی کند. به عبارتی سئوال در مورد تطابق (Goodness-of-fit) به این گونه آزمونها در اصطلاح **آزمونهای نیکویی برازش** می گویند.

الف – داده های کمی (آزمون کلموگروف – اسمیرنف) :

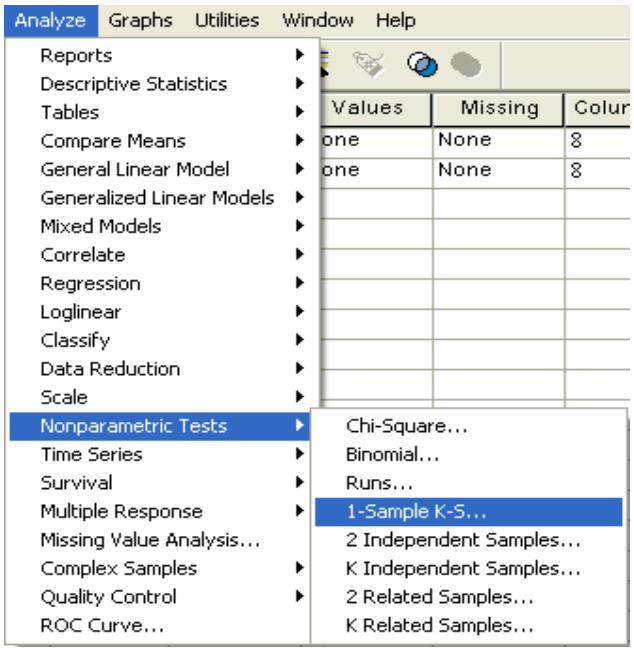
زمانی که برای داده های کمی بخواهیم آزمون نیکویی برازش را بکار ببریم از آزمون کلموگروف – اسمیرنف استفاده می کنیم . به عبارتی می خواهیم فرضهای زیر را آزمون کنیم :

$$\begin{cases} H_0 : F(x) = F_0(x) \\ H_1 : F(x) \neq F_0(x) \end{cases}$$

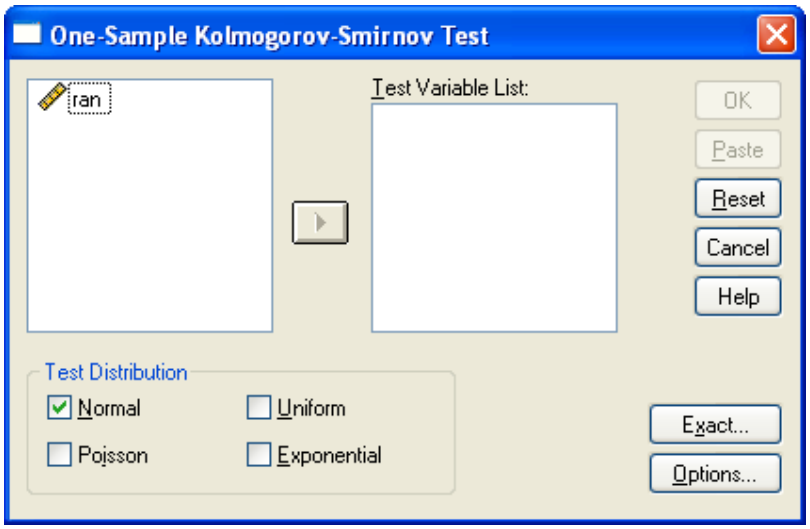
مثال : می خواهیم ببینیم که داده های زیر از توزیع نرمال پیروی می کنند یا خیر؟

	ran
1	3.50
2	5.33
3	3.87
4	8.19
5	8.72
6	7.05
7	6.37
8	8.76
9	2.81
10	-.46
11	5.70
12	9.25
13	5.11
14	5.14
15	10.38
16	1.86
17	7.17
18	3.69
19	10.69
20	10.44

برای انجام این کار مسیر زیر را طی می کنیم :



تا کادر محاوره ای زیر ایجاد شود :



متغیر ran را به قسمت Test variable list وارد کنید . گزینه Normal را تیک بزنید و سپس دکمه Ok را کلیک کنید تا خروجی را به صورت زیر ببینید :

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		RAN
N		20
Normal Parameters[a,b]	Mean	6.1788
	Std. Deviation	3.08160
Most Extreme Differences	Absolute	.095
	Positive	.073
	Negative	-.095
Kolmogorov-Smirnov Z		.424
Asymp. Sig. (2-tailed)		.994

a. Test distribution is Normal.

b. Calculated from data.

تذکر: آزمون کلموگروف – اسمیرنف از مقایسه تابع تجمعی احتمال مشاهدات و تابع احتمال توزیع مفروض, فرض پیروی مشاهدات از توزیع احتمالی خاص را پیروی می کند .

ب – داده های دو حالتی اسمی :

فرض کنید سکه ای را ۳۰ مرتبه پرتاب کرده اید و نتایج به صورت زیر بدست آمده است (۱= شیر , ۰= خط) :

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1

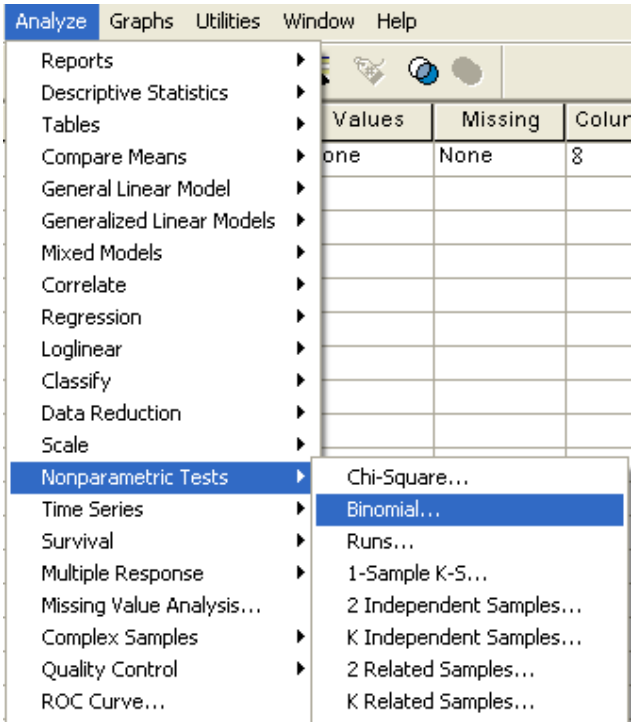
می خواهیم آزمون کنیم که آیا سکه سالم است یا خیر ؟ می خواهیم فرض های زیر را آزمون کنیم:

$H_0 = p = 0.5$

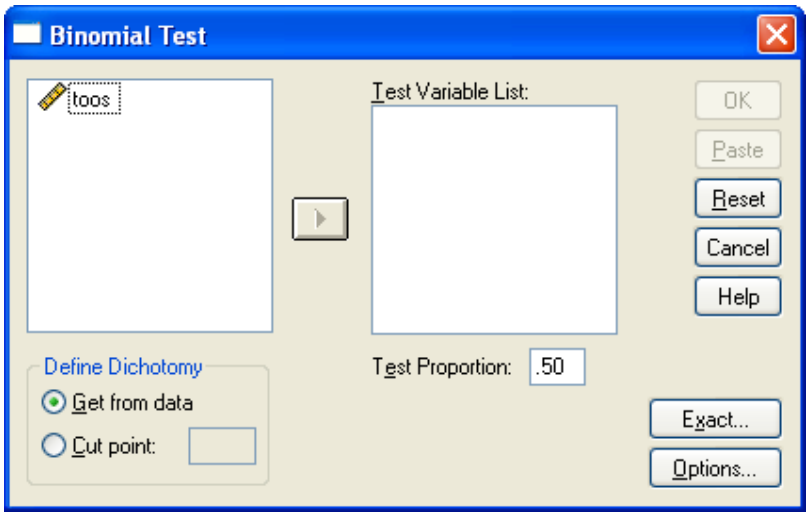
$H_1 = p \neq 0.5$

برای آزمون این فرض از آزمون دوجمله ای بهره می گیریم :

برای انجام این کار داده های بالا را به عنوان متغیر toss وارد کرده و مسیر زیر را طی کنید :



کادر مکالمه ای زیر را مشاهده می کنید :



متغیر toss را به قسمت Test variable list وارد کرده و مقدار Test proportion را ۰/۵ وارد کنید و در قسمت Define Dichotomy گزینه اول را انتخاب کنید و سپس دکمه Ok را کلیک کنید تا خروجی زیر ظاهر شود :

NPar Tests

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
TOOS	Group 1	1.00	18	.60	.50	.362[a]
	Group 2	.00	12	.40		
	Total		30	1.00		

a. Based on Z Approximation.

با توجه به این خروجی به سئوالی که در ابتدای بحث مطرح شد پاسخ دهید.

تمرین :

دو نوع اسباب بازی مکانیکی و غیر مکانیکی داریم.می خواهیم ببینیم که آیادرکودکان ۵ ساله تمایل به سمت هرکدام از این اسباب بازیها یکسان است؟ بدین منظور از ۲۰۰ کودک ۵ ساله خواسته ایم که به دلخواه یکی از این اسباب بازی ها را انتخاب کنند و نتایج به صورت زیر بدست آمده است:

مکانیکی	۱۲۰
غیرمکانیکی	۸۰

از این داده ها چه استنباطی دارید ؟

تمرین :

وزن ۲۰ بازیکن فوتبال که به تصادف از تیمهای فوتبال یک شهر انتخاب شده اند به صورت زیر است:

۸۱	۷۱/۵	۷۵	۶۴
۸۳	۷۰	۶۹/۵	۶۷
۹۴	۵۸/۵	۹۵	۹۰
۹۷	۸۸	۶۲	۵۹
۸۰	۶۱/۵	۶۰	۵۹/۵

آیا میتوان گفت که :

الف –۶۰ درصد از بازیکنان فوتبال این شهر کمتر از $65kg$ وزن دارند؟

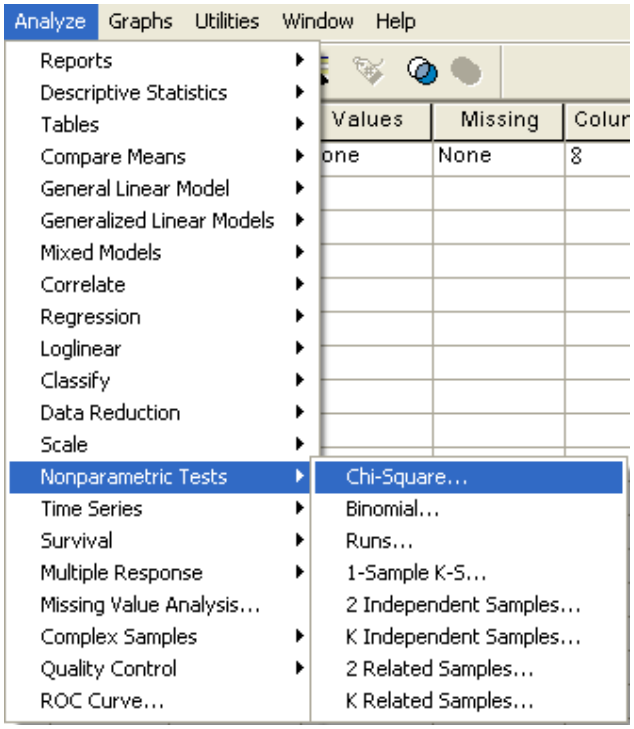
ب– ۴۰ درصد از بازیکنان فوتبال این شهر بیشتراز $60kg$ وزن دارند ؟

پ) آزمون تطابق توزیع با سه طبقه یا بیشتر (آزمون Chi – Square) :

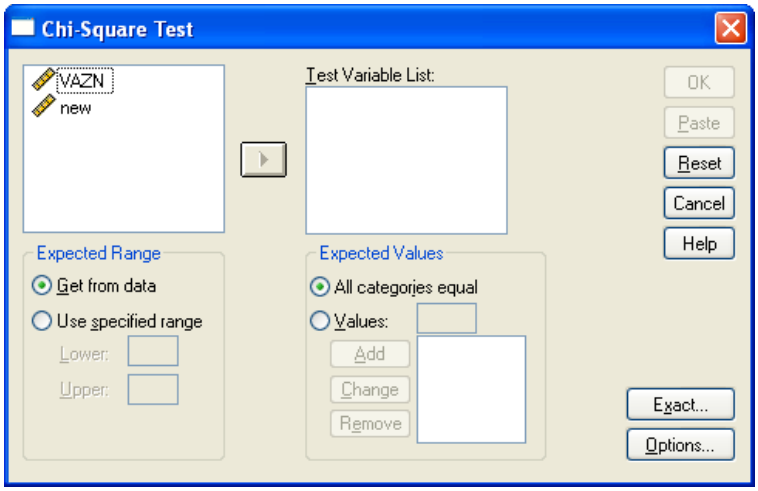
فرض کنید می خواهیم آزمون کنیم که آیا نسبت با سوادی در گروه هایی با درآمدهای مختلف یکسان است یا خیر؟ یا توزیع نمرات دانشجویان یک درس در بازه های مختلف یکسان است یا خیر؟ و به طور کلی می خواهیم توزیع نسبت مقادیر یک متغیر در رده های مختلف را با توزیع از پیش تعیین شده ای مقایسه کنیم . برای این کار آزمون Chi – Square را به کار می بریم.

مثال : داده های صفحهٔ ۲۰ جزوه را در نظر بگیرید. می خواهیم آزمون کنیم که آیا اوزان کره ها در ۷ رده وزنی ساخته شده بطور یکسانی توزیع شده اند یا خیر ؟

برای این کار مسیر زیر را طی کنید :



تا کادر مکالمه ای زیر باز شود :



متغیرnew را به قسمت Test variable list وارد کرده و Ok را کلیک کنید تا خروجی را به صورت زیر مشاهده کنید :

Chi-Square Test

Test Statistics

	NEW
Chi-Square[a]	7.950
df	6
Asymp. Sig.	.242

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.7.

در قسمت Expected Range می توانید برد داده ها, و در قسمت Expected Values نیز می توانید نسبت رده را خودتان تعیین کنید .

آزمون تصادفی بودن (گردش) :

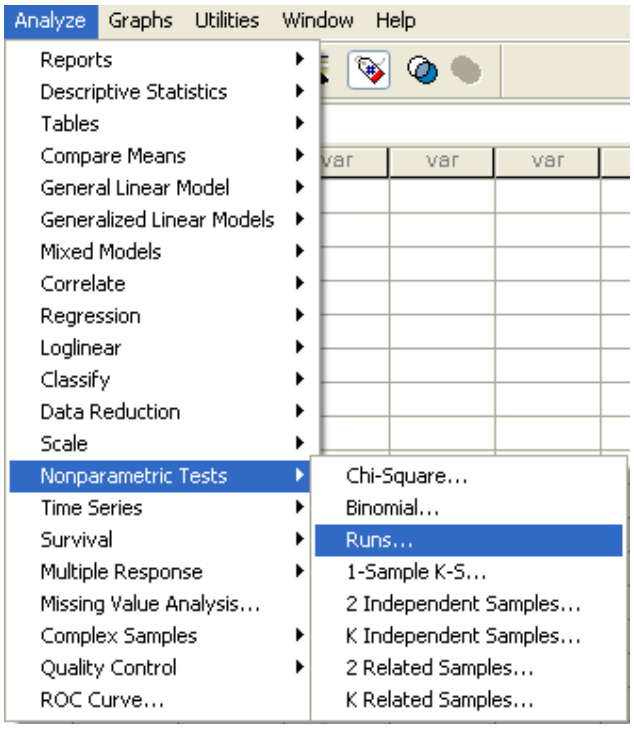
در بسیاری از آنالیزهای آماری یکی از فرضهای اساسی "تصادفی بودن مشاهدات"است. برای مثال در بررسی مانده های یک مدل آماری , برقراری فرض تصادفی بودن (استقلال مشاهدات) حیاتی است.انحراف از فرض تصادفی بودن می تواند به دلیل وجود روندهای افزایشی یا کاهشی, رفتارهای دوره ای یا افزایش تغییرپذیری و برخی علل دیگر باشد.

مثال : فرض کنید نتایج ۲۰ بار پرتاب یک سکه بصورت زیر است (۱= شیر , ۰= خط) :

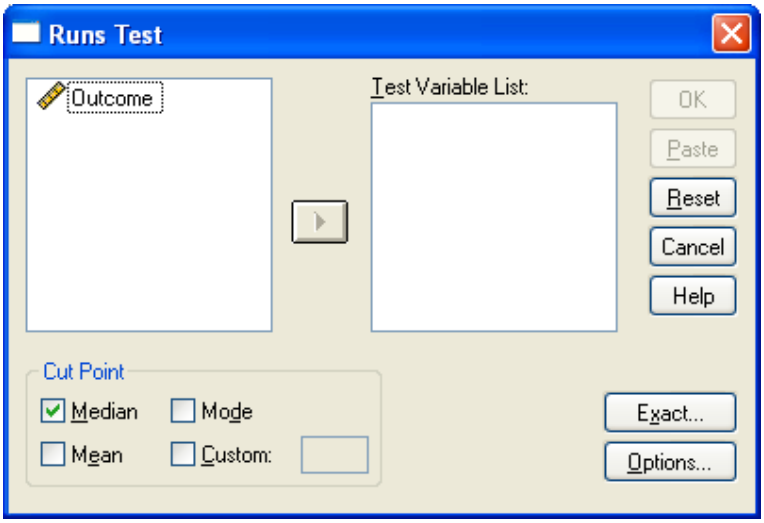
0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1

می خواهیم بدانیم که آیا این نتایج تصادفی است یا خیر ؟

برای انجام این کار ابتدا داده ها را به عنوان متغیر Outcome وارد SPSS کنید و سپس مسیر زیر را طی کنید :



تا کادر محاوره ای زیر باز شود :



در کادر محاوره ای بالا متغیر Outcome را به قسمت Test variable list برده و در قسمت Cut point خط مبنای تشخیص نوسانات تصادفی را انتخاب کنید (برای مثال میانه را انتخاب کنید) و دکمه Ok را کلیک کنید و خروجی را به صورت زیر مشاهده کنید :

NPar Tests

Runs Test

	OUTCOME
Test Value[a]	.50
Cases < Test Value	10
Cases >= Test Value	10
Total Cases	20
Number of Runs	20
Z	3.905
Asymp. Sig. (2-tailed)	.000

a. Median

با توجه به این خروجی به سئوالی که در ابتدای بحث مطرح شد پاسخ دهید .

با انتخاب هر خط مبنایی, کوچکتر از آن, کد **صفر** و به مقادیر بزرگتر یا مساوی با آن, کد **یک** خواهند داشت و با ساختن دنباله مشاهدات با مقادیر صفر و یک آزمون گردش اجرا می شود .

تمرین :

تعداد دانشجویانی را که ساعت ۸ صبحِ ۲۰ روز تحصیلی, یک اتوبوس از خوابگاه به دانشگاه می برد, اعداد زیر می باشد :

۲۱	۲۴	۲۳	۱۵	۳۲	۲۸	۲۶	۱۷	۲۰	۲۸
۳۰	۲۴	۱۳	۳۵	۲۶	۲۱	۱۹	۲۹	۲۷	۱۸

آیا این داده ها یک نمونه تصادفی اند؟

ارتباط و رابطه بین متغیرها

در تجزیه و تحلیل آماری بجز اطلاعاتی که در مورد هر متغیر بدست می آوریم، علاقه‌مندیم تا نحوهٔ ارتباط متغیرها را شناسایی کنیم:

– آیا رابطه ای بین قد شخص با وزن او وجود دارد؟

– آیا معدل دانش آموزان رابطه ای با شغل والدین دارد؟

– آیا رنگ پوست یک مجرم تاثیری در رای دادگاه علیه وی دارد؟

–آیا بعد از یک عمل جراحی شدت ضربان قلب با مقدار داروی بیهوشی مصرف شده ارتباط دارد؟

–آیا مقدار محصول یک مزرعه با نوع کود مصرفی در آن ارتباط دارد؟

در حقیقت در اکثر بررسی های آماری با سئوالاتی ازاین دسته به وفور برخورد خواهیم کرد. ولی با تأملی در متغیرهای به کار رفته در هر مورد می بینیم که نمی توان تمامی سئوالات فوق را در یک دسته قرار داد. در واقع بسته به نوع متغیرهایی که مایل به کشف رابطه بین آنها هستیم، روش های آماری مختلفی را به کار خواهیم گرفت. برای شروع کار فرض می کنیم تنها با ۲ متغیر روبرو هستیم، بسته به نوع دو متغیر نحوه انجام آنالیزهای آماری را شرح می دهیم:

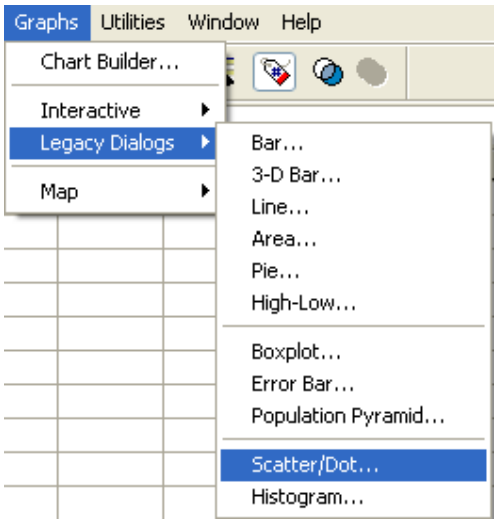
۱–ارتباط بین متغیرهای عددی:

نمودار پراکنش وسیله ی مناسبی برای کشف رابطه بین متغیرهای عددی است. قبل از بدست آوردن ضریب همبستگی پیرسون (برای دو متغیر عددی) بایستی که نمودار پراکنش داده ها را رسم کنیم.

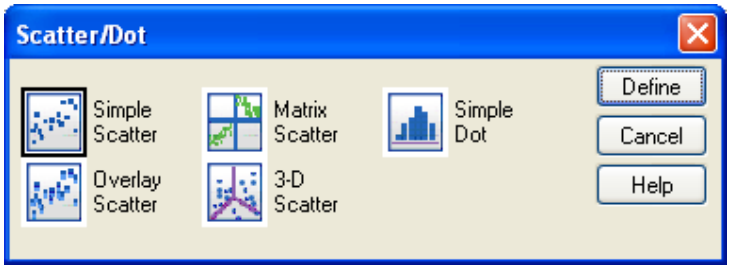
مثال (*نمودار پراکنش (Scatter plot) و بررسی معنی داری رابطه*): جهت بررسی رابطهٔ بین ضریب هوشی فرد و نمره ریاضی از ۱۰ نفر، این صفات در جدول زیر ثبت شده است :

نمره ریاضی	۳۱	۲۵	۲۰	۲۴	۱۸	۲۸	۱۸	۲۰	۱۶	۲۷
ضریب هوشی	۱۲۰	۱۱۲	۱۱۰	۱۲۳	۱۰۵	۱۲۵	۱۱۴	۱۱۳	۱۰۶	۱۲۷

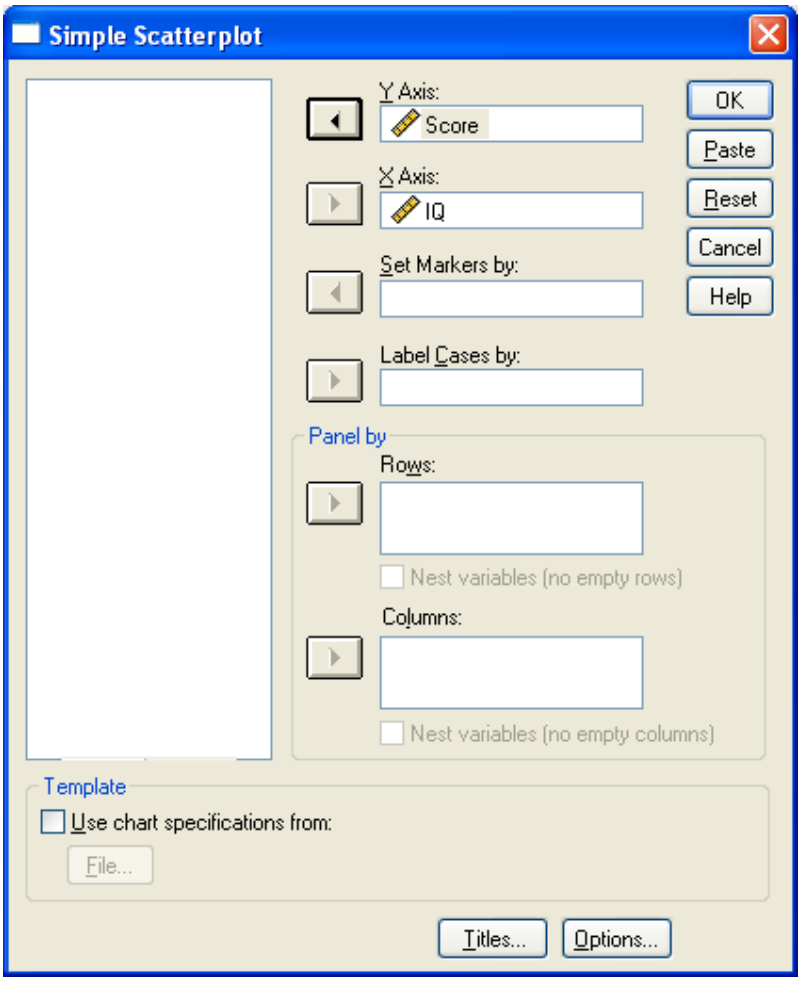
ابتدا داده ها را در دو ستون، در محیط SPSS وارد کنید.حال مسیر زیر را برای رسم نمودار پراکنش طی کنید تا کادر مربوطه باز شود. متغیر ها را به صورت زیر نامگذاری و وارد کنید :



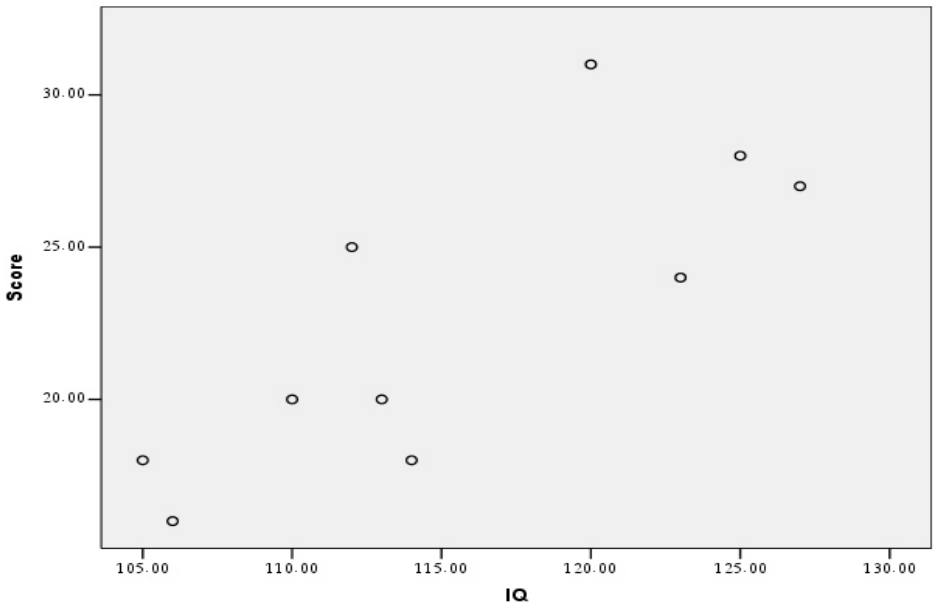
در کادر زیرگزینه Simple Scatter را انتخاب کنید و دکمه Define را کلیک کنید:



تا کادر مکالمه ای زیر باز شود:



دکمه Ok را کلیک کنید. نمودار رسم شده به صورت زیر خواهد بود:



ملاحظه می کنید که با افزایش مقادیر IQ، مقادیر متناظر Score هم افزایش می یابند و بـه نظـر می رسد مقادیر IQ وScore با هم رابطه دارند. با اینکه نمودار پراکنش وسیله خوبی برای بیـان رابطـه بین دو متغیر عددی است، ولی علاقه مندیم بوسیله یک شاخص آماری شدت این رابطه را بیان کنیم.

شاخص آماری که برای نشان دادن سطح ارتباط دو متغیر عددی به کار می رود ضریب همبستگی است که عددی است بین ۱- و ۱. هر چه ضریب همبستگی به صفر نزدیکتر باشد. نشان دهنده عدم وجود رابطه بین دو متغیر است ضریب همبستگی مثبت، رابطه هم جهت را نشان می‌دهد و ضریب همبستگی منفی، مبین رابطه عکس است. در SPSS از ۳ نوع ضریب همبستگی استفاده می شود:

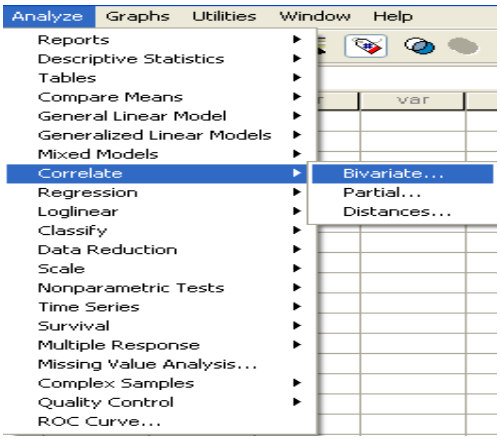
۱) *ضریب همبستگی ساده (Pearson)* : در حالتی که هر دو متغیر دارای مقیاس فاصله ای باشند به کار می رود.

۲) *ضریب همبستگی Spearman* : به جای مقادیر متغیرها، ضریب همبستگی Spearman با رتبه های دو متغیر محاسبه می شود. ضریب همبستگی Spearman در مواردی که حداقل یکی از متغیرها ordinal باشد، به کار می رود.

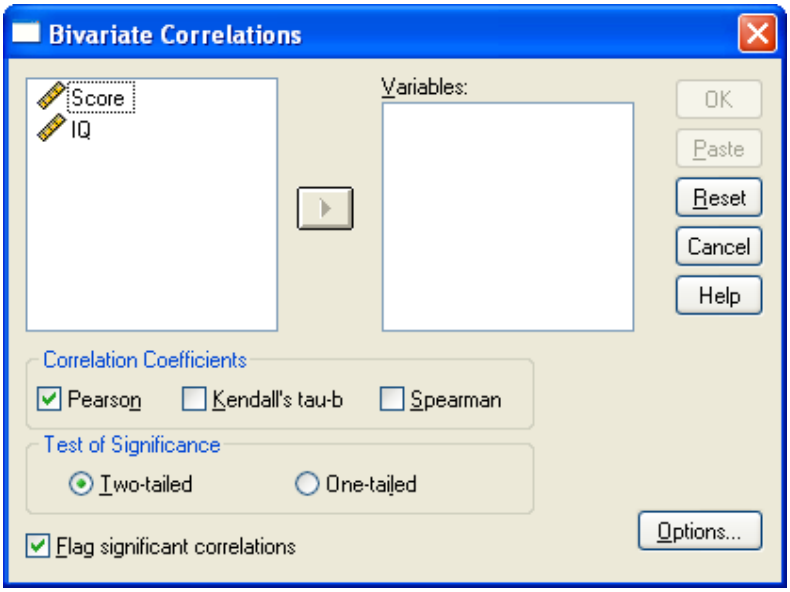
۳) *ضریب همبستگی kendall's tau-b* : همانند ضریب همبستگی Spearman یک ضریب همبستگی ناپارامتری است و هنگامی که مقیاس حداقل یکی از متغیرها ordinal باشد به کار خواهد رفت.

برای داده های IQ و Score مثال قبل ضریب همبستگی ساده (Pearson) و Spearman را محاسبه کنید.

برای محاسبه ضریب همبستگی ساده مسیر زیر را طی کنید:



در کادر ظاهر شده، هر دو متغیر IQ وScore را انتخاب کنید (توجه کنید که در قسمت Correlation coefficient، ضریب همبستگی Pearson انتخاب شده باشد):



سپس دکمه ok را فشار دهید. خروجی به صورت زیر خواهد بود:

Correlations

Correlations

		IQ	Score
IQ	Pearson Correlation	1	.795**
	Sig. (2-tailed)		.006
	N	10	10
Score	Pearson Correlation	.795**	1
	Sig. (2-tailed)	.006	
	N	10	10

**. Correlation is significant at the 0.01 level

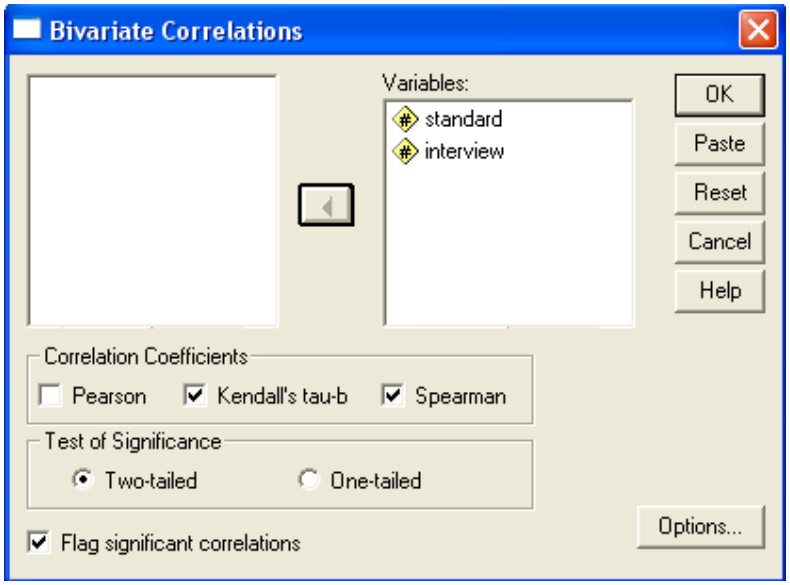
مشاهده می کنید که ضریب همبستگی IQ وIQ، همچنین Score و Score برابر یک است و ضریب همبستگی IQ و Score هم ۰.۷۹۵ است در ضمن آزمون معناداری ضریب همبستگی هم انجام شده است که صفر آزمون به صورت $H_0 : \rho = 0$ می باشد. اگر فرض H_0 رد می شود به این معنی است که ضریب همبستگی صفر نیست (اصطلاحاً معنادار است.)

مثال: مصاحبه کننده ای که مسئول استخدام تعداد زیادی ماشین نویس است می خواهد قدرت رابطهٔ بین رتبه های داده شده بر مبنای یک مصاحبه و نمرات یک آزمون استعداد را تعیین کند. داده های ۶ متقاضی عبارتند از :

رتبهٔ مصاحبه	۵	۲	۳	۱	۶	۴
نمرهٔ استاندارد	۴۷	۳۲	۲۹	۲۸	۵۶	۳۸

ضریب همبستگی اسپیرمن و کندال-تاو را بدست می آوریم و معنی داری رابطهٔ را نیز بررسی میکنیم:

ابتدا متغیرها را با نام standard وinterview در SPSS وارد کنید. در کادر مکالمه ای قبل متغیرها را وارد کرده و گزینه های kendall's tua-b و spearman را تیک دار کنید.



حال گزینهٔ Ok را کلیک کنید تا خروجی به صورت زیر نمایش داده شود:

Correlations

			standard	interview
Kendall's tau_b	standard	Correlation Coefficient	1.000	.867*
		Sig. (2-tailed)	.	.015
		N	6	6
	interview	Correlation Coefficient	.867*	1.000
		Sig. (2-tailed)	.015	.
		N	6	6
Spearman's rho	standard	Correlation Coefficient	1.000	.943**
		Sig. (2-tailed)	.	.005
		N	6	6
	interview	Correlation Coefficient	.943**	1.000
		Sig. (2-tailed)	.005	.
		N	6	6

*. Correlation is significant at the 0.05 level (2-tailed).

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

خروجی را تفسیر کنید(؟)

تمرین: محققی مایل است بداند آیا بین سن افراد و قدرت بدنی افراد رابطه ای خطی وجود دارد یا نه؟ داده های جمع آوری شده توسط وی به صورت زیر است:

سن	قدرت بدنی
۱۰	۱۲
۱۵	۴۵
۲۰	۷۵
۲۵	۷۰
۳۰	۶۳
۳۵	۶۰
۴۰	۵۵
۴۵	۵۰
۵۰	۴۰
۵۵	۳۰
۶۰	۲۸
۶۵	۲۵
۷۰	۲۰
۷۵	۱۵
۸۰	۵

آیا بین سن افراد و قدرت بدنی افراد رابطه ای خطی وجود دارد؟

۲-ارتباط بین متغیرهای رسته ای:

در این بخش با سوالاتی بدین صورت روبرو هستیم:

– آیا رابطه ای بین جنسیت و سیگاری بودن افراد وجود دارد؟ یعنی به عنوان مثال مردان بیش از زنان گرایش به سیگار داشته باشند؟

– آیا گرایش سیاسی یک شخص (دمکرات – جمهوری خواه) با گرایش مذهبی او (مسیحی – یهودی – مسلمان – غیره)رابطه ای دارد؟

– جداول توافقی:

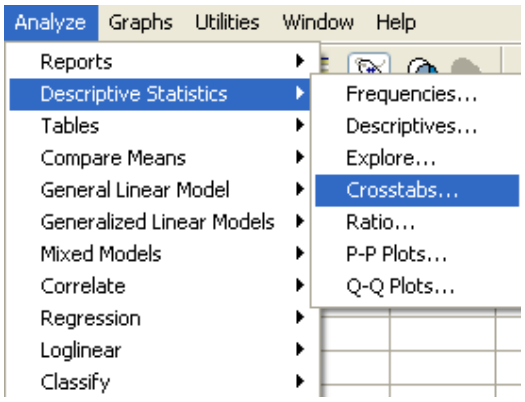
اولین قدم در آنالیز مسائلی به این صورت، رسم جدول توافقی است. جدول توافقی دو متغیر رسته ای، به تعداد رسته های متغیر اول سطر و به تعداد رسته های متغیر دوم، ستون دارد و در هر خانه این جدول تعداد فراوانیهای مربوط به آن حالت قرار می گیرد (به جای فراوانی، ممکن است درصد فراوانی نوشته شود)

داده های زیر نتایج یک بررسی بر روی ۳۰ نفر است که جنسیت و سیگاری بودن (یا نبودن) پرسیده شده است:

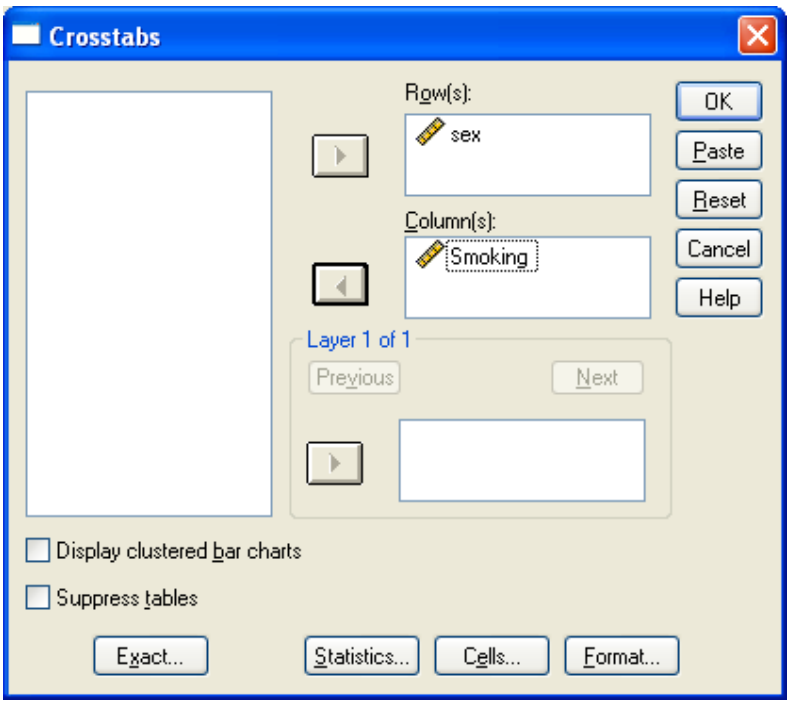
	sex	Smoking
1	Female	No
2	Female	No
3	Female	No
4	Female	Yes
5	Female	No
6	Female	Yes
7	Female	No
8	Female	No
9	Female	No
10	Female	No
11	Female	No
12	Female	Yes
13	Female	No
14	Female	No
15	Male	Yes

	sex	Smoking
16	Male	No
17	Male	No
18	Male	Yes
19	Male	Yes
20	Male	Yes
21	Male	No
22	Male	No
23	Male	Yes
24	Male	Yes
25	Male	No
26	Male	Yes
27	Male	Yes
28	Male	No
29	Male	Yes
30	Male	Yes

برای رسم جدول توافقی داده های فوق مسیر زیر را طی کنید:



تا کادر زیر باز شود :



مطابق تصویر، متغیر sex را به قسمت Row و متغیر Sigari را به قسمت Column(s) ببرید و دکمه Ok را کلیک کنید. خروجی به صورت زیر خواهد بود:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SEX * SIGARI	30	100.0%	0	.0%	30	100.0%

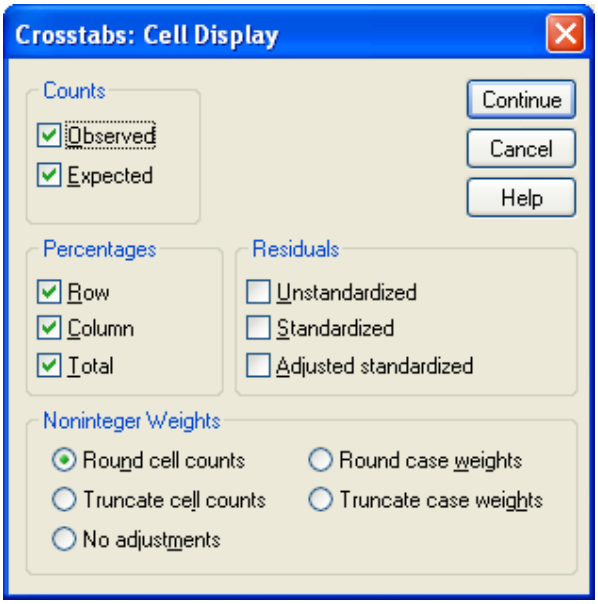
SEX * SIGARI Crosstabulation

Count

		SIGARI		Total
		بلی	خیر	
SEX	زن	3	11	14
	مرد	10	6	16
Total		13	17	30

جدول اول تعداد داده های موجود، گمشده و کل داده ها را نشان می دهد و جدول دوم، جدول توافقی دو متغیر فوق است که تعداد مردان سیگاری، زنان سیگاری، مردان غیرسیگاری و زنان غیرسیگاری را نشان می دهد.

برای مشاهده درصدها و مقادیر موردانتظار در کادر اصلی Crosstabs روی دکمه Cells... کلیک کنید تا کادر زیر باز شود:



مطابق تصویر بالا، مقادیر موردانتظار، درصدهای سطری، ستونی و کل را انتخاب کنید.

Continue و سپس Ok را کلیک کنید.

جدول توافقی حاصل به صورت زیر است:

sex * Smoking Crosstabulation

			Smoking		Total
			Yes	No	
sex	Female	Count	3	11	14
		Expected Count	6.1	7.9	14.0
		% within sex	21.4%	78.6%	100.0%
		% within Smoking	23.1%	64.7%	46.7%
		% of Total	10.0%	36.7%	46.7%
	Male	Count	10	6	16
		Expected Count	6.9	9.1	16.0
		% within sex	62.5%	37.5%	100.0%
		% within Smoking	76.9%	35.3%	53.3%
		% of Total	33.3%	20.0%	53.3%
Total		Count	13	17	30
		Expected Count	13.0	17.0	30.0
		% within sex	43.3%	56.7%	100.0%
		% within Smoking	100.0%	100.0%	100.0%
		% of Total	43.3%	56.7%	100.0%

رسم جدول توافقی بدون داشتن داده های خام:

فرض کنید در مثال بالا به جای اطلاعات هر ۳۰ نفر تنها تعداد مردان سیگاری، زنان سیگاری، مردان غیر سیگاری و زنان غیر سیگاری را داشته باشیم. داده ها را به ترتیب زیر در SPSS وارد می کنیم:

sex	smoking	frequency
Male	Yes	10.00
Male	No	6.00
Female	Yes	3.00
Female	No	11.00

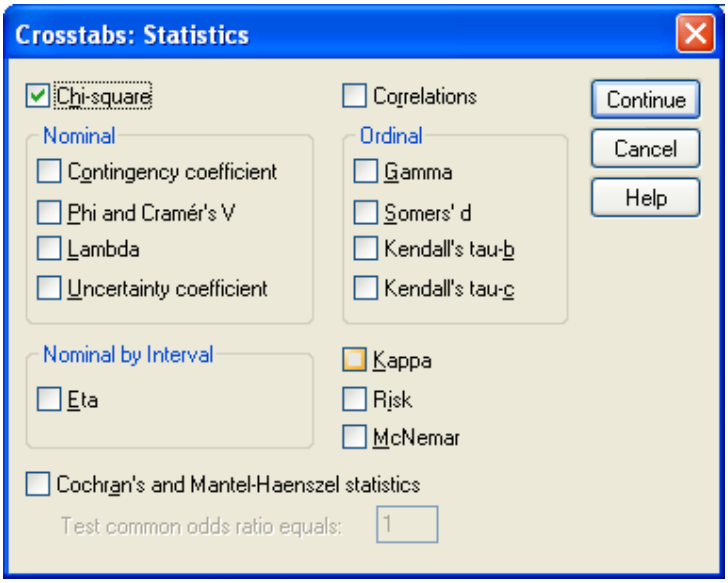
سپس به داده ها بر حسب متغیر freq وزن می دهیم و مراحل رسم جدول توافقی را طی می کنیم .

آزمون استقلال متغیرها:

پس از تحلیل توصیفی داده ها و خلاصه کردن در قالب جدول توافقی، علاقه مندیم بدانیم آیا بین متغیرها رابطه ای وجود دارد یا نه؟ دو متغیر در صورتی که مستقل باشند، هیچ رابطه ای با هم ندارند. مثلاً اگر دو متغیر مثال قبلی مستقل باشند. سیگاری بودن شخص رابطه ای با جنسیت او ندارد یا به عبارتی جنسیت تاثیری در گرایش مردم به سیگار ندارد. ولی در صورتیکه دو متغیر مستقل نباشند، با هم رابطه دارند. برای آزمون استقلال دو متغیر از *آزمون خی دو* استفاده خواهیم کرد.

منتها استفاده از آزمون خی دو شرایطی دارد مه داده ها بایستی انرا دارا باشند. این شرایط را باید در جدول توافقی داده عها جستجو کرد. با توجه به جدول توافقی حاصل، آیا می توانیم از آزمون خی دو استفاده کنیم؟؟؟ . . .

برای انجام آزمون خی دو در کادر Crosstabs روی دکمه ...Statistics کلیک کنید مطابق شکل گزینه chi-square را انتخاب کنید Continue و سپس ok را کلیک کنید:



نتیجه آزمون استقلال برای داده های مثال قبل به صورت زیر خواهد بود:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.129[b]	1	.024		
Continuity Correction[a]	3.593	1	.058		
Likelihood Ratio	5.336	1	.021		
Fisher's Exact Test				.033	.028
Linear-by-Linear Association	4.958	1	.026		
N of Valid Cases	30				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.07.

نتایج آزمون خی دو پیرسون در سطر اول آمده است. فرض صفر در این آزمون استقلال متغیرهاست و در صورتیکه فرض صفر رد شود، نشان دهنده وجود رابطه بین دو متغیر خواهد بود.

تمرین ۱ :

محققی مایل است بداند آیا بین جنسیت افراد و چپ دست بودن-راست دست بـودن افـراد رابطـه ای وجود دارد یا نه؟ بدین منظور داده های زیر را جمع آوری کرده است:

	دختر	پسر
چپ دست	۲	۱۰
راست دست	۱۱	۱

به نظر شما این رابطه معنی دار است یا خیر؟

تمرین ۲ :

میخواهیم بدانیم آیا رابطهٔ بین گروه خونی افراد و فشار خون آنها رابطـه ای وجـود دارد یـا نـه؟ بـدین منظور داده های زیر را جمع آوری کرده ایم:

	A	B	O	AB
+	۲۵	۱۲	۱۷	۸
–	۸	۱۷	۱۲	۲۵

به نظر شما این رابطه معنی دار است یا خیر؟

۳- مقایسه چند جمعیت (آنالیز واریانس)

با کاربرد آزمون t برای مقایسه میانگین دوجامعه آشنا شدیم، آنالیز واریانس به ما کمک مـی کنـد تا میانگین های چند جامعه را با هم مقایسه کنیم. متغیری که قرار است میانگین آن در چند زیر جامعه مقایسه شود متغیر وابسته و متغیر یا متغیرهایی که باعث تقسیم جامعه به چند زیر جامعه می شـوند، متغیر(های) مستقل نامیده می شوند. متغیر مستقل معمولاً از نوع اسمی است. جواب سـئوالات زیـر بـا آنالیز واریانس داده می شود:

- آیا درآمد نهایی فیلم های امسال به نوع فیلم بستگی دارد یا نه؟
- آیا شغل والدین در معدل دانش آموزان یک مدرسه تاثیر دارد یا نه؟
- آیا چهار روش مختلف آموزش ریاضی در دانـش آمـوزان چهـار کـلاس یـک مدرسـه، تـاثیر یکسانی دارد یا خیر؟

در سئوالات فوق می بینید که متغیرهای مستقل باعث تقسیم جامعه به چند زیر گروه یا زیر جامعه می شوند و هدف آزمون این مساله است که آیا میانگین متغیر وابسته در زیر گروهها یکسان است یا خیر. به عبارتی فرض صفر در آنالیز واریانس، یکسان بودن میانگین تمام زیر گروههاست.

در آنالیز واریانس فرض بر این است که داده ها از توزیع نرمال آمده اند و واریانس داده ها در زیر گروهها یکسان است. در واقع درستی فرض های فوق، برای معنا دار بودن آنالیز واریانس لازم است. لذا همواره قبل از انجام آنالیز واریانس فرض های فوق را آزمون خواهیم کرد مخصوصاً فرض تساوی واریانس ها.

اگر شکی در نرمال بودن داده ها داشته باشیم یا تعداد داده ها کم باشد می توان به جای آنالیز واریانس از معادل ناپارامتری آن استفاده کرد.

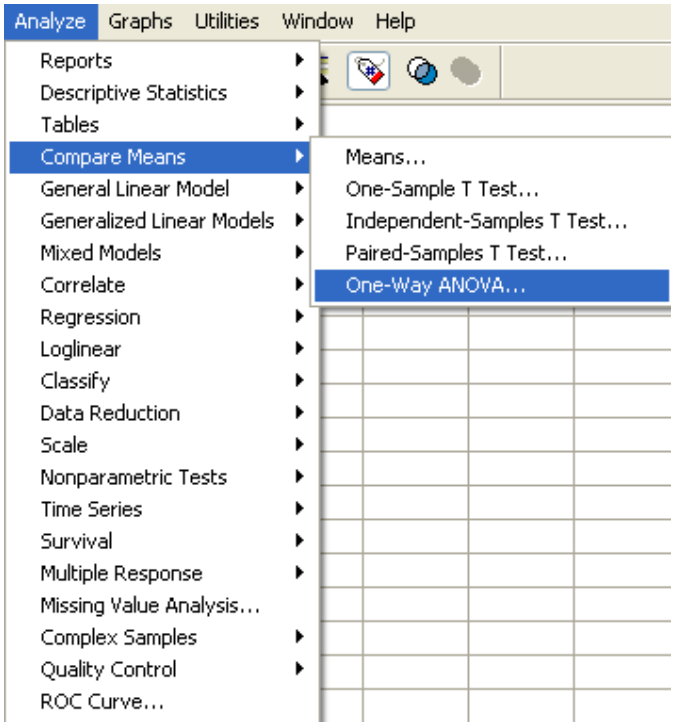
مثال: داده های زیر مقدار نیکوتین ۲۰ سیگار است که به طور تصادفی از میان ۳ نوع سیگار مختلف انتخاب شده اند، آیا هر سه نوع سیگار، مقدار نیکوتین یکسان دارند؟

	nikotin	sigar
1	74.0	1
2	73.5	1
3	74.5	1
4	74.0	1
5	73.5	1
6	74.0	1
7	64.5	2
8	65.0	2
9	64.0	2
10	65.0	2
11	64.5	2
12	64.5	2
13	64.0	2
14	65.0	2
15	65.5	3
16	65.0	3
17	64.5	3
18	66.0	3
19	64.5	3
20	65.5	3

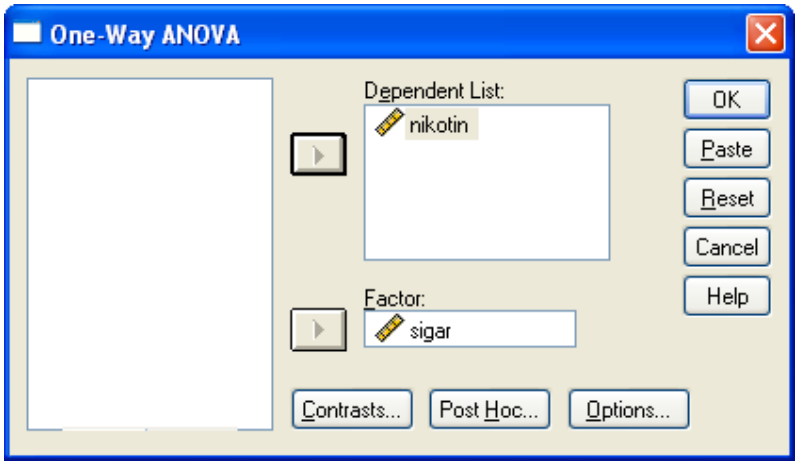
✓ **به نحوه وارد کردن داده ها توجه کنید.**

در این مسأله مقدار نیکوتین (Nikotin) متغیر وابسته و نوع سیگار (cigar) متغیر مستقل است.

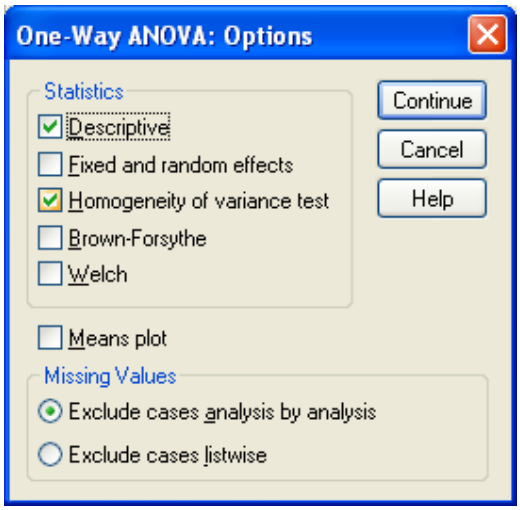
برای انجام آنالیز واریانس مسیر زیر را طی کنید:



تا کادر زیر باز شود:



متغیرها را مطابق شکل در قسمتهای مربوطه وارد کنید. سپس روی دکمه ...Options کلیک کنید تـا کادر زیر باز شود:



گزینه های Homogeneity of variance test و Descriptive را انتخاب کنید سپس دکمه Continue و سپس OK را کلیک کنید. خروجی به صورت زیر خواهد بود:

Descriptives

NIKOTIN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	6	73.917	.3764	.1537	73.522	74.312	73.5	74.5
2	8	64.563	.4173	.1475	64.214	64.911	64.0	65.0
3	6	65.167	.6055	.2472	64.531	65.802	64.5	66.0
Total	20	67.550	4.3070	.9631	65.534	69.566	64.0	74.5

Test of Homogeneity of Variances

NIKOTIN

Levene Statistic	df1	df2	Sig.
1.506	2	17	.250

ANOVA

NIKOTIN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	348.690	2	174.345	788.174	.000
Within Groups	3.760	17	.221		
Total	352.450	19			

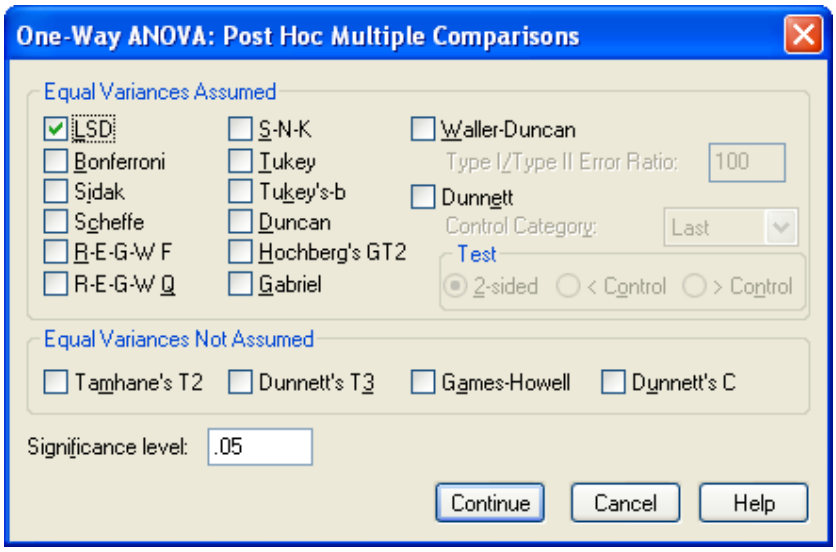
در کادر Descriptive آماره های توصیفی برای زیر گروهها به صورت جداگانه محاسبه شده است.

در کادر دوم آزمون تساوی واریانس زیر گروهها انجام شده است که همگنی واریانسها را نشان می دهد.

نتیجه نهایی آنالیز واریانس در کادر ANOVA آمده است. با توجه به سطح معناداری آزمون چه نتیجه ای می گیریم؟

نتیجه آنالیز واریانس مشخص می کند که آیا میانگین زیر گروهها یکسان است یا خیر؟ اگر نتیجه آنالیز واریانس متفاوت بودن میانگین ها را نتیجه دهد، علاقه مندیم بدانیم این تفاوت کلی ناشی از کدام تفاوتهای جزئی است. به عبارتی دیگر کدام تفاوت های دوگانه باعث ایجاد تفاوت کلی شده است. به عنوان مثال در مساله بالا نتیجه آنالیز واریانس تفاوت در میانگین نیکوتین ۳ نوع سیگار مورد آزمایش را نشان می‌دهد، حال علاقه مندیم با تحلیل بیشتر داده ها، بدانیم این تفاوت کلی از کجا نشأت گرفته است، لذا میانگین ها را دو بدو مقایسه می‌کنیم.

در کادر آنالیز واریانس روی دکمه Post Hoc کلیک کنید تا کادر زیر باز شود:



گزینه LSD را انتخاب کنید Continue و سپس Ok را کلیک کنید.

ملاحظه می کنید در خروجی جدولی به صورت زیر اضافه شده است:

Post Hoc Tests

Multiple Comparisons

Dependent Variable: NIKOTIN
LSD

(I) SIGAR	(J) SIGAR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	9.354*	.2540	.000	8.818	9.890
	3	8.750*	.2715	.000	8.177	9.323
2	1	-9.354*	.2540	.000	-9.890	-8.818
	3	-.604*	.2540	.029	-1.140	-.068
3	1	-8.750*	.2715	.000	-9.323	-8.177
	2	.604*	.2540	.029	.068	1.140

*. The mean difference is significant at the .05 level.

در این جدول آزمون تساوی میانگین زیر گروهها دوبه دو انجام شده است.

آنالیز واریانس دوطرفه :

در SPSS برای طرح های پیچیده تر از آنالیز واریانس یک طرفه از یک روند آماری به نام مدل عمومی خطی(GLM) استفاده می کنیم.

در این بخش به آنالیز واریانس دو طرفه (یک نوع مدل خطی) می پردازیم :

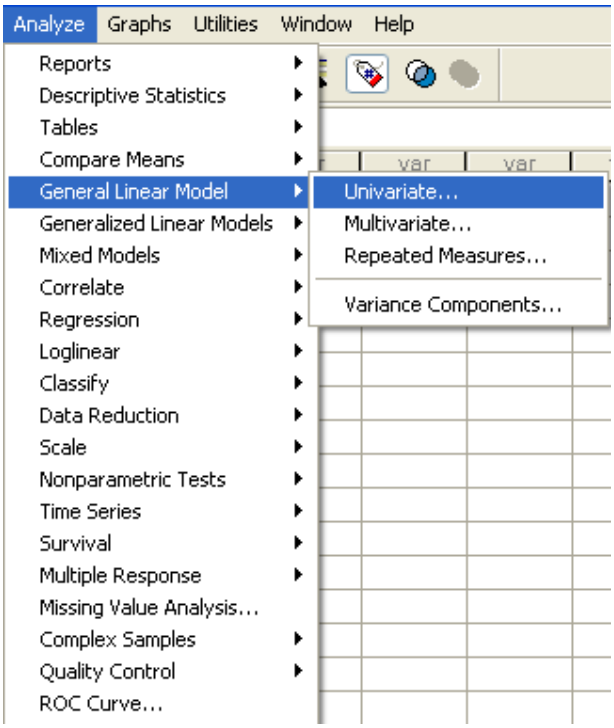
فرض کنید بجای بررسی تاثیر یک فاکتور بر متغیر پاسخ بخواهیم اثر دو فاکتور را بر متغیر پاسخ بررسی کنیم, در این صورت با یک طرح دو عاملی روبرو خواهیم بود .

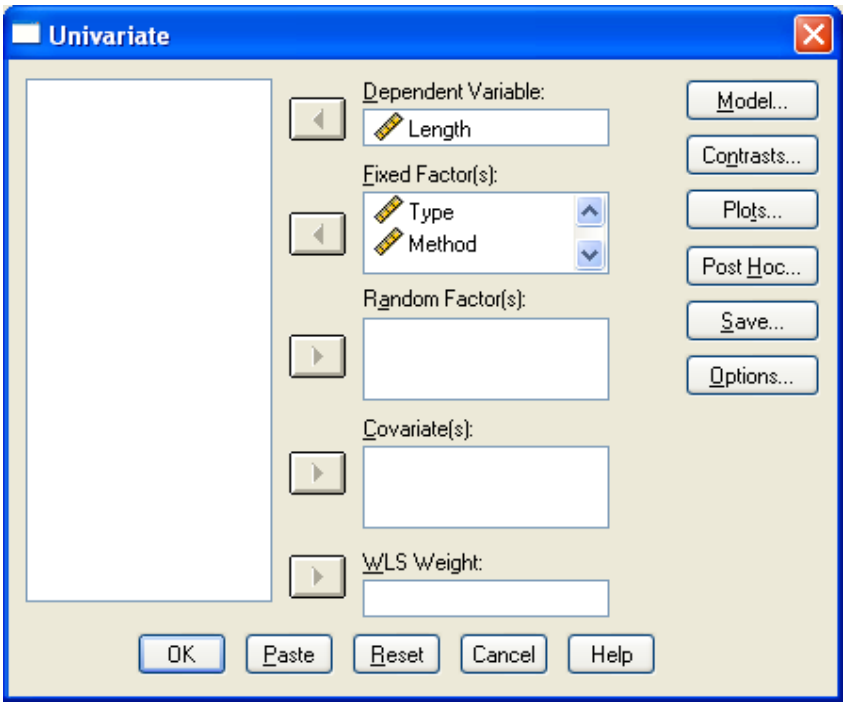
مثال :

می خواهیم بدانیم که آیا روش درمان یک بیماری خاص و نوع داروی مصرفی جهت درمان آن بیماری در طول مدت درمان آن بیماری تاثیری دارد یا خیر؟ برای ارزیابی این فرضیه ۶۰نفر از افراد مبتلا به این بیماری خاص را مورد آزمایش قرار داده ایم. داده ها به صورت زیرند :

ابتدا باید داده هارا در SPSS وارد کنیم. به نحوه داده وارد کردن توجه کنید.(با نامهای method ،type و length و سپس مسیر زیر را طی کنید ،تا کادر مکالمه ای مربوط به آن باز شود و متغیرها را به صورت نشان داده شده وارد کنید:

		نوع دارو			
		۱	۲	۳	۴
روش درمان	۱	۶	۷	۸	۸
		۵	۸	۷	۹
		۴	۲	۱۰	۱۰
		۷	۳	۵	۱
		۲	۵	۶	۶
	۲	۳	۴	۲	۳
		۲	۶	۵	۴
		۱	۵	۶	۵
		۴	۳	۳	۷
		۵	۲	۴	۱
	۳	۴	۶	۵	۸
		۵	۸	۹	۶
		۲	۵	۲	۳
		۷	۴	۳	۹
		۱	۹	۱	۱





توجه داشته باشید که هر دو فاکتور method و type تثبیت شده اند . اگر چنانچه هر کـدام از آنها تصادفی بودند آنگاه آن را به قسمت Random Factor(s) وارد می کردیم.

پس از زدن کلید Ok خروجی را به صورت زیر مشاهده می کنید :

Between-Subjects Factors

		N
Method	1.00	20
	2.00	20
	3.00	20
Type	1.00	15
	2.00	15
	3.00	15
	4.00	15

Tests of Between-Subjects Effects

Dependent Variable: lenght

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	97.333[a]	11	8.848	1.508	.160
Intercept	1421.067	1	1421.067	242.227	.000
Method	48.433	2	24.217	4.128	.022
Type	20.933	3	6.978	1.189	.324
Method * Type	27.967	6	4.661	.795	.579
Error	281.600	48	5.867		
Total	1800.000	60			
Corrected Total	378.933	59			

a. R Squared = .257 (Adjusted R Squared = .087)

آیا هیچ نشانی از اینکه یکی از این عوامل در طول مدت درمان موثر باشند وجود دارد؟
آیا دو عامل اثر متقابل دارند؟

تمرین ۱)

مهندسین برق برای اندازه گیری روشنایی لامپ تصویر تلویزیون, جریان لازم برای بدست آوردن سطح روشنایی مشخص را ملاک قرار میدهند. فرض کنید میخوا هیم اثر نوع شیشه ونوع فسفر, در روشنایی لامپ تصویر را بررسی کنیم. در این مسئله متغیر پاسخ, مقدار جریان لازم است. داده های بدست آمده بصورت زیراست :

		نوع فسفر		
		۱	۲	۳
نوع شیشه	۱	۲۸۰ ۲۹۰ ۲۸۵	۳۰۰ ۳۱۰ ۲۹۵	۲۹۰ ۲۸۵ ۲۹۰
	۲	۲۳۰ ۲۳۵ ۲۴۰	۲۶۰ ۲۴۰ ۲۳۵	۲۲۰ ۲۲۵ ۲۳۰

آیا هیچ نشانی از اینکه یکی از این عوامل در روشنایی موثر باشد, وجود دارد؟
آیا دو عامل اثر متقابل دارند؟

تمرین۲)

از چهار شیمیدان خواسته شده است که درصد الکل متیلیک یک ترکیب خاص شیمیایی را تعیین کنند. هر یک از این شیمیدانها سه اندازه گیری انجام داده و نتایج زیر را گزارش کرده اند :

		درصد الکل متیلیک		
شیمیدان	۱	۸۴/۹۹	۸۴/۰۴	۸۴/۳۸
	۲	۸۵/۱۵	۸۵/۱۳	۸۴/۸۸
	۳	۸۴/۷۲	۸۴/۴۸	۸۵/۱۶
	۴	۸۴/۲۰	۸۴/۱۰	۸۴/۵۵

آیا گزارش شیمیدانها تفاوت معنی دار دارد؟ (در سطح $\alpha = 0.05$ آزمون را انجام دهید.)

تمرین ۳)

می خواهیم بدانیم که آیا روش درمان یک بیماری خاص در طول مدت درمان آن بیماری تـاثیری دارد یا خیر؟ برای ارزیابی این فرضیه ۲۵ نفر از افراد مبتلا به بیماری خاص به ۴ گروه تقسیم و هر کـدام از گروها با یک روش مورد درمان قرار گرفتند . داده های حاصل به صورت زیرند:

	روشهای درمان			
	۱	۲	۳	۴
طول مدت درمان	۸	۷	۹	۵
	۷	۵	۷	۷
	۶	۸	۵	۸
	۸	۴	۸	۷
	۷	۵	۸	۵
	۹	۶	۷	۸
				۷

آیا داده ها دلیلی بر تفاوت طول مدت درمان در روشهای درمان متفاوت است یا خیر؟

تمرین ۴)

آزمایشی برای تعیین اثر دمای کوره یا اثرهای موقعیت کوره بر چگالی یک آنودکربنی اجرا شده است .داده ها را در زیر نشان داه شده است.

		دما		
		۸۰۰	۸۲۵	۸۵۰
موقعیت	۱	۵۷۰	۱۰۶۳	۵۶۵
		۵۶۵	۱۰۸۰	۵۱۰
		۵۸۳	۱۰۴۳	۵۹۰
	۲	۵۲۸	۹۸۸	۵۲۶
		۵۴۷	۱۰۲۶	۵۳۸
		۵۲۱	۱۰۰۴	۵۳۲

داده ها را تحلیل کنید ؟

رگرسیون:

در بحث ضریب همبستگی چگونگی بررسی رابطه میان متغیرهای عددی و محاسبه ضریب همبستگی را بیان کردیم. ولی اغلب در مسائلی که با متغیرهای عددی روبرو هستیم، علاقه مندیم علاوه بر بررسی ساده رابطه دو متغیر مقادیر یک متغیر (متغیر وابسته) را به صورت تابعی از یک یا چند متغیر مستقل دیگر بنویسیم. روشهای تحلیل رگرسیون به ما کمک می کنند تا چگونگی این رابطه را تعیین کنیم.

-رگرسیون خطی ساده:

در ساده ترین حالت فرض می کنیم متغیر وابسته تنها با یک متغیر مستقل در ارتباط باشد و رابطه این دو به صورت زیر باشد:

$$y = \alpha + \beta x + \varepsilon$$

یعنی y ها به صورت خطی وابسته به مقادیر x اند. اگر صحت چنین رابطه ای تأیید شود، توانسته ایم داده ها را به صورت بسیار خوبی خلاصه کنیم. علاوه بر این برای مقادیر مختلف x می توانیم برآوردی از مقادیر y مورد انتظار ارائه دهیم.

داده های زیر مخارج تبلیغات فروش هفتگی ۱۲ فروشگاه مختلف در یک شهر است.

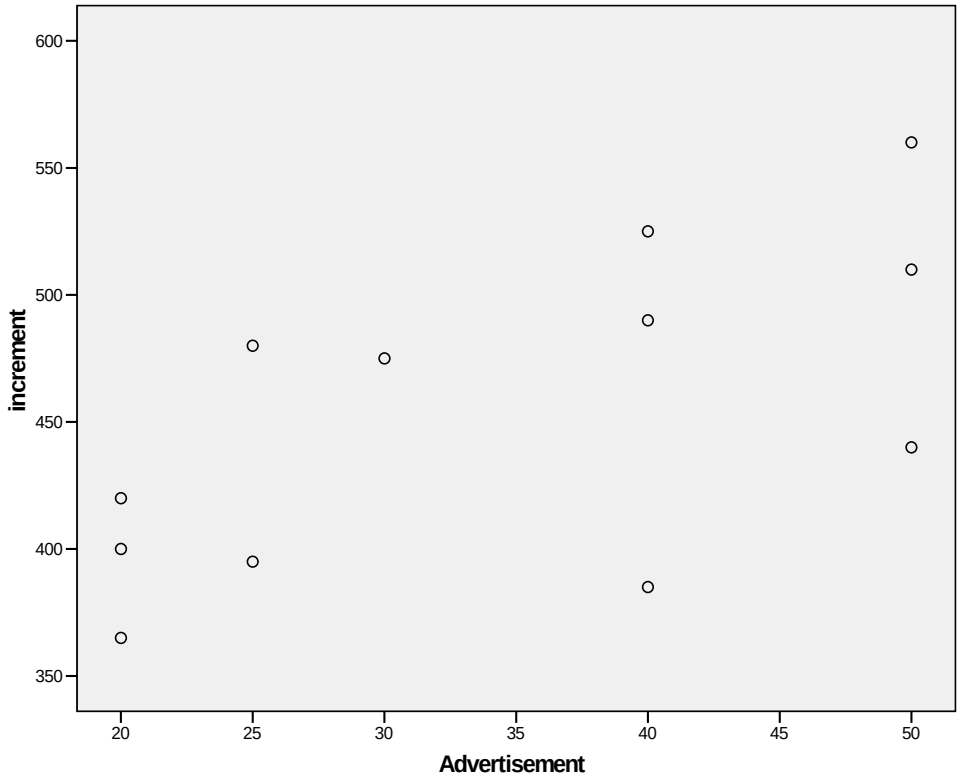
مخارج تبلیغ	۴۰	۲۰	۲۵	۲۰	۳۰	۵۰	۴۰	۲۰	۵۰	۴۰	۲۵	۵۰
فروش	۳۸۵	۴۰۰	۳۹۵	۳۶۵	۴۷۵	۴۴۰	۴۹۰	۴۲۰	۵۶۰	۵۲۵	۴۸۰	۵۱۰

رابطه خط رگرسیونی بین دو متغیر را در صورت وجود بیابید. (رابطه خطی)

نخست داده ها را در دو ستون در محیط SPSS وارد می کنیم:

Advertisement	increment
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

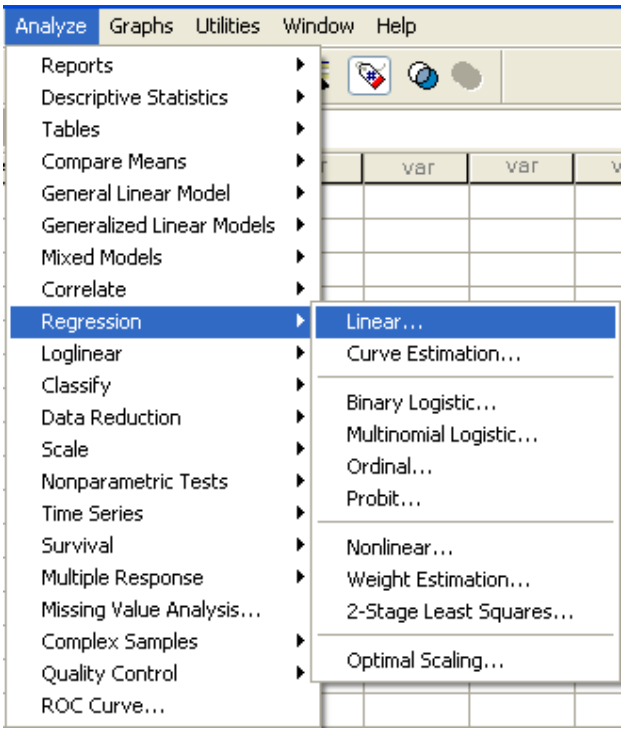
قدم اول در تحلیل رگرسیونی ساده همانند بخش ضریب همبستگی، رسم نمودار پراکنش است. نمودار پراکنش می تواند وجود یا عدم وجود رابطه خطی بین متغیرها را به خوبی نشان دهد. نمودار پراکنش را برای دو متغیر رسم کنید. نمودار پراکنش به صورت زیر خواهد شد:



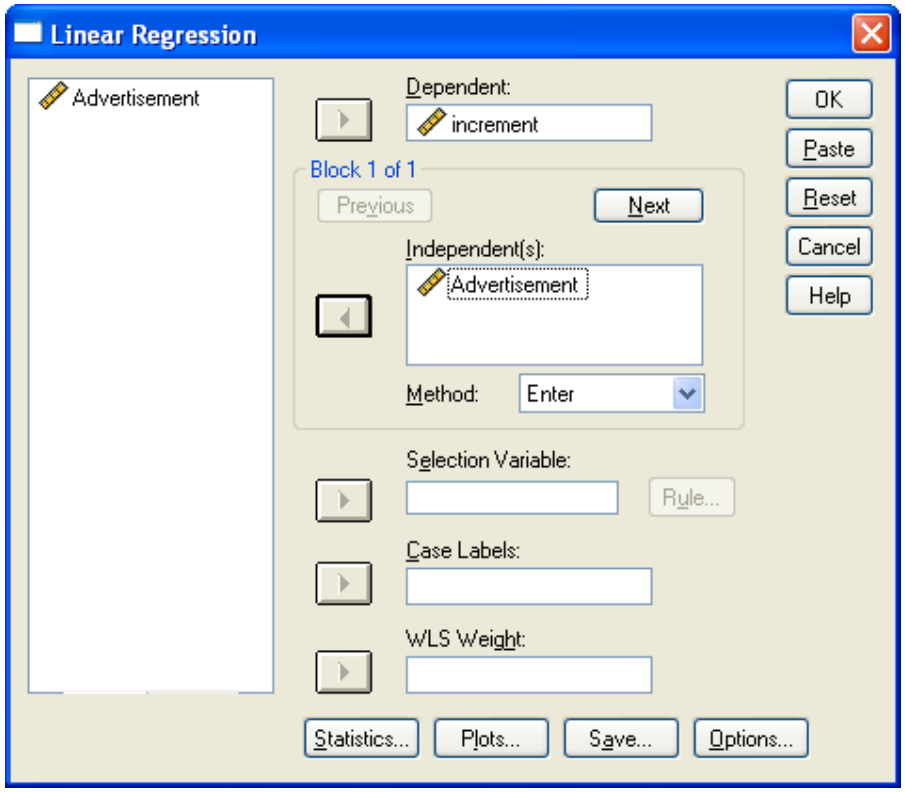
نمودار پراکنش نشاندهنده رابطه نسبی دو متغیر است. یعنی به نظر می رسد y با x رابطه خطی دارد.

مرحلهٔ بعد پیدا کردن برآورد ضریبهای معادله خط رگرسیون (β,α) است.

برای انجام تحلیل رگرسیونی به صورت کامل مسیر زیر را طی کنید:



تا کادر زیر باز شود :



متغیر increment را در قسمت Dependent و متغیر Advertisement را در قسمت Independents وارد کنید و دکمه Ok را فشار دهید، خروجی به صورت زیر خواهد بود:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	TABLIGH[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: FOROOSH

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.635[a]	.403	.343	50.226

a. Predictors: (Constant), TABLIGH

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17030.04	1	17030.044	6.751	.027[a]
	Residual	25226.21	10	2522.621		
	Total	42256.25	11			

a. Predictors: (Constant), TABLIGH

b. Dependent Variable: FOROOSH

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	343.706	44.766		7.678	.000
	TABLIGH	3.221	1.240	.635	2.598	.027

a. Dependent Variable: FOROOSH

جدول دوم، حاوی اطلاعاتی در مورد کارایی مدل است. هر چه مقدار (Adjust R Square) یا R Square به یک نزدیکتر باشد مدل کاراتر است. در این مثال مقدار R Square چندان زیاد نیست.

جدول سوم، جدول اصلی تحلیل رگرسیونی است. آزمون صورت گرفته به آزمون معناداری خط رگرسیون مشهور است. اگر نتیجه آزمون رد فرض H_0 باشد (sig. کمتر از 0.05 باشد) خط رگرسیون معنادار است.(که در این مورد رگرسیون با معنی است.)

جدول چهارم، جدول ضرایب خط رگرسیونی است و آزمون های صورت گرفته هم، آزمون برابری این ضرایب با صفر می باشد که در صورت رد شدن آزمون، ضرایب مخالف صفر خواهند بود (یا به عبارتی این ضرایب در خط رگرسیون وجود دارند و بی معنی نیستند.)

- **رگرسیون چند متغیره خطی:**

اگرتعداد متغیرهای مستقل بیش از یکی باشد. مثلاً دو متغیر باشد رابطه خط رگرسیون به صورت زیر خواهد بود:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

و هدف پیدا کردن برآورد ضرایب فوق است.

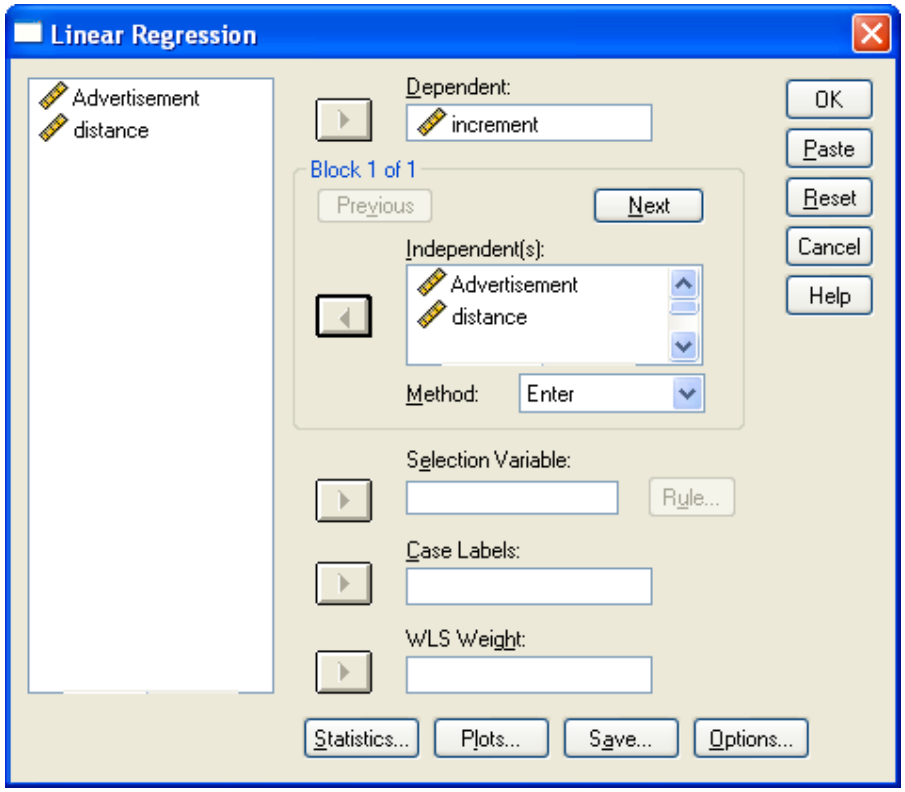
فرض کنید در مساله قبل طبق تجارب گذشته حدس می زنیم که میزان فروش هفتگی علاوه بر هزینه تبلیغ به فاصله فروشگاه تا میدان مرکزی شهر هم بستگی دارد. داده های زیر فاصله ۱۲ فروشگاه مربوطه از مرکز شهر بر حسب کیلومتر است:

۶/۳	۰/۲	۹/۴	۵/۳	۳/۴	۲/۶	۷/۱	۸/۳	۲/۴	۰/۷	۳/۱	۴/۲

داده ها را در ستون جدیدی وارد کنید:

Advertisement	increment	distance
40	385	4.20
20	400	3.10
25	395	.70
20	365	2.40
30	475	8.30
50	440	7.10
40	490	2.60
20	420	3.40
50	560	5.30
40	525	9.40
25	480	.20
50	510	6.30

مسیری که باید طی کنیم همان مسیر قبلی است اما این بار در کادر تحلیل رگرسیونی متغیر مستقل دوم را هم اضافه می کنیم:



خروجی در این حالت به صورت زیر خواهد بود:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	FASELE, TABLIGH[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: FOROOSH

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.650[a]	.422	.294	52.079

a. Predictors: (Constant), FASELE, TABLIGH

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17846.45	2	8923.223	3.290	.085[a]
	Residual	24409.80	9	2712.201		
	Total	42256.25	11			

a. Predictors: (Constant), FASELE, TABLIGH

b. Dependent Variable: FOROOSH

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	344.412	46.436		7.417	.000
	TABLIGH	2.736	1.560	.539	1.754	.113
	FASELE	3.592	6.546	.169	.549	.597

a. Dependent Variable: FOROOSH