



قدم به قدم، همراه دانشجو...

WWW.GhadamYar.Com

جامع ترین و به روزترین پرتال دانشجویی کشور (پرتال دانش)  
با ارائه خدمات رایگان، تحصیلی، آموزشی، رفاهی، شغلی و...  
برای دانشجویان

- (۱) راهنمای ارتقاء تحصیلی. (کاردانی به کارشناسی، کارشناسی به ارشد و ارشد به دکتری)
- (۲) ارائه سوالات کنکور مقاطع مختلف سالهای گذشته، همراه پاسخ، به صورت رایگان
- (۳) معرفی روش های مقاله و پایان نامه نویسی و ارائه پکیج های آموزشی مربوطه
- (۴) معرفی منابع و کتب مرتبط با کنکورهای تحصیلی (کاردانی تا دکتری)
- (۵) معرفی آموزشگاه ها و مراکز مشاوره تحصیلی معتبر
- (۶) ارائه جزوات و منابع رایگان مرتبط با رشته های تحصیلی
- (۷) راهنمای آزمون های حقوقی به همراه دفترچه سوالات سالهای گذشته (رایگان)
- (۸) راهنمای آزمون های نظام مهندسی به همراه دفترچه سوالات سالهای گذشته (رایگان)
- (۹) آخرین اخبار دانشجویی، در همه مقاطع، از خبرگزاری های پربازدید
- (۱۰) معرفی مراکز ورزشی، تفریحی و فروشگاه های دارای تخفیف دانشجویی
- (۱۱) معرفی همایش ها، کنفرانس ها و نمایشگاه های ویژه دانشجویی
- (۱۲) ارائه اطلاعات مربوط به بورسیه و تحصیل در خارج و معرفی شرکت های معتبر مربوطه
- (۱۳) معرفی مسائل و قوانین مربوط به سربازی، معافیت تحصیلی و امریه
- (۱۴) ارائه خدمات خاص ویژه دانشجویان خارجی
- (۱۵) معرفی انواع بیمه های دانشجویی دارای تخفیف
- (۱۶) صفحه ویژه نقل و انتقالات دانشجویی
- (۱۷) صفحه ویژه ارائه شغل های پاره وقت، اخبار استخدامی
- (۱۸) معرفی خوابگاه های دانشجویی معتبر
- (۱۹) دانلود رایگان نرم افزار و اپلیکیشن های تخصصی و...
- (۲۰) ارائه راهکارهای کارآفرینی، استارت آپ و...
- (۲۱) معرفی مراکز تایپ، ترجمه، پرینت، صحافی و ... به صورت آنلاین
- (۲۲) راهنمای خرید آنلاین ارزی و معرفی شرکت های مطرح
- (۲۳) .....



WWW.GhadamYar.Ir

WWW.PortaleDanesh.com

WWW.GhadamYar.Org

۰۹۱۲ ۳۰۹۰۱۰۸

باما همراه باشید...

۰۹۱۲ ۰۹ ۰۳ ۸۰۱



# راهنمای آموزش SPSS 19

تهیه و تنظیم :  
مهدی اکبرزاده

جلد اول

## فهرست مطالب

۱	پیشگفتار.....
۳	فصل اول.....
۳	مروری بر نرم افزار SPSS.....
۳	۱- شروع کار با SPSS.....
۳	۱-۱- معرفی محیط نرم افزار.....
۳	نوار منو:.....
۳	نوار ابزار:.....
۴	صفحه Data View.....
۴	نوار آدرس:.....
۵	صفحه Variable View.....
۵	دکمه تعویض صفحه:.....
۵	نوار وضعیت:.....
۵	نحوه باز کردن و ذخیره کردن:.....
۵	تمرین:.....
۶	۱-۲- وارد کردن داده ها در SPSS:.....
۷	تعریف متغیرها در SPSS:.....
۷	مشخصه Name:.....
۷	مشخصه Type:.....
۸	مشخصه Width و Decimals:.....
۸	مشخصه Label:.....
۹	مشخصه Values:.....
۹	مشخصه Missing:.....
۱۰	مشخصه Columns:.....
۱۰	مشخصه Align:.....
۱۰	مشخصه Measure:.....
۱۰	مشخصه Role:.....
۱۴	فصل دوم.....
۱۴	آمار توصیفی در SPSS.....
۱۴	۱-۲ مقدمه.....
۱۵	۲-۲ خلاصه کردن و توصیف الگوی کلی:.....
۱۵	۱-۲-۲ داده های کیفی (Categorical Data):.....
۲۱	۲-۲-۲ داده های عددی (Numerical Data):.....
۳۱	۳-۲-۲ محاسبه شاخصهای آماری:.....

۳۱	.....شاخصهای مرکزی:
۳۲	.....شاخصهای پراکندگی:
۳۵	.....Box plot نمودار
۳۸	.....۴-۲-۲ چند نمودار مهم و کاربردی دیگر:
۴۲	.....فصل سوم
۴۲	.....دستورهایی برای دست کاری داده ها در SPSS
۴۲	.....۱-۳ دستور Select Cases
۴۴	.....تمرین:
۴۴	.....۲-۳ دستور Split File
۴۵	.....تمرین:
۴۶	.....۳-۳ دستور Weight Cases:
۴۸	.....تمرین:
۴۹	.....۴-۳ دستور Compute
۵۰	.....تمرین:
۵۰	.....۵-۳ دستور Count Values:
۵۲	.....۶-۳ دستور Recode:
۵۲	.....۱-۶-۳ جدول فراوانی برای صفات کمی پیوسته (دستور Recode):
۵۴	.....۲-۶-۳ معکوس کردن امتیازات (دستور Recode):
۵۶	.....فصل چهارم
۵۶	.....آزمون فرض آماری
۵۶	.....۱-۴ مقدمه
۵۷	.....۲-۴ فرض صفر و فرض مقابل
۵۷	.....۳-۴ سطح معنی داری و خطاهای آماری
۵۹	.....۴-۴ توزیع نمونه گیری آماره
۵۹	.....۵-۴ آزمون فرض یک طرفه و دو طرفه
۶۰	.....۶-۴ مراحل کلی آزمون فرض آماری
۶۱	.....۷-۴ ماهیت P-Value
۶۱	.....آزمون P-Value
۶۲	.....۸-۴ آزمون آماری برای میانگین جامعه - آزمون t تک نمونه ای
۶۴	.....۹-۴ آزمون آماری برای نسبت جامعه - آزمون دو جمله ای
۶۶	.....۱۰-۴ آزمون اختلاف میانگین ها برای دو جامعه مستقل - آزمون t- دو نمونه مستقل
۷۰	.....۱۱-۴ آزمون اختلاف میانگین ها برای دو جامعه وابسته - آزمون t زوجی
۷۳	.....تمرین:
۷۶	.....ضمیمه
۷۶	.....تمرینات تکمیلی

## 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 راه گشا نیست و در راه گسترش و تعمیم یک بحث در یک رشته خاص بدون شک باید با علوم دیگر نیز آشنا باشند. یکی از علمی که پیوند محکمی با سایر علوم دارد، علم "آمار" است. به طوری 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 داده‌های 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، به طور 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 نرم‌افزارهای آماری موجودند که قادرند تحلیل‌های آماری را انجام دهند، برخی از مهمترین این نرم‌افزارها 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 می‌کند. برای مثال محققین در رشته کشاورزی از برنامه SAS بیشتر بهره می‌برند و یا افرادی که در رشته‌های مهندسی صنایع در حال تحقیق و پژوهش هستند از برنامه MINITAB برای انجام پروژه‌های صنعتی خود استفاده می‌کنند، یا محققین آماری از برنامه‌های R و S-plus استفاده می‌کنند.

### اما چرا SPSS ؟

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

کامل تر شده است و در حال حاضر کلیه محققین در رشته های علوم اجتماعی، علوم پزشکی، علوم تربیتی، روان شناسی، کشاورزی و ... در حال استفاده از این نرم افزار هستند و در این لحظه که شما تصمیم به فراگیری این نرم افزار آماری گرفته اید، ویرایش ۱۹ این نرم افزار در کشورمان به طور 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 هر گونه 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 نرم افزار SPSS

در این بخش قسمت های متفاوت نرم افزار SPSS را معرفی خواهیم کرد.

#### ۱- شروع کار با SPSS

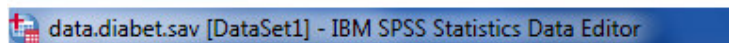
پس از نصب نرم افزار SPSS، برای اجرای آن به منوی Start بروید و بر آیکون IBM SPSS Statistics 19 کلیک کنید. در این قسمت به معرفی محیط نرم افزار و اعمال مقدماتی در آن، می پردازیم.

#### ۱-۱- معرفی محیط نرم افزار

پس از اجرای نرم افزار، با پنجره ای 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:

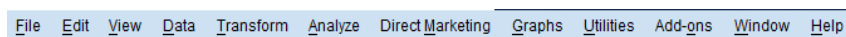
برنامه SPSS نیز مانند سایر نرم افزارهای تحت Windows دارای نوار عنوان است که در آن نام فایل SPSS را مشاهده می کنید. برای مثال فایل داده های زیر به نام data.diabet ذخیره شده است:



توجه داشته باشید که پسوند فایل های داده در این نرم افزار sav. و پسوند فایل های خروجی spv. است.

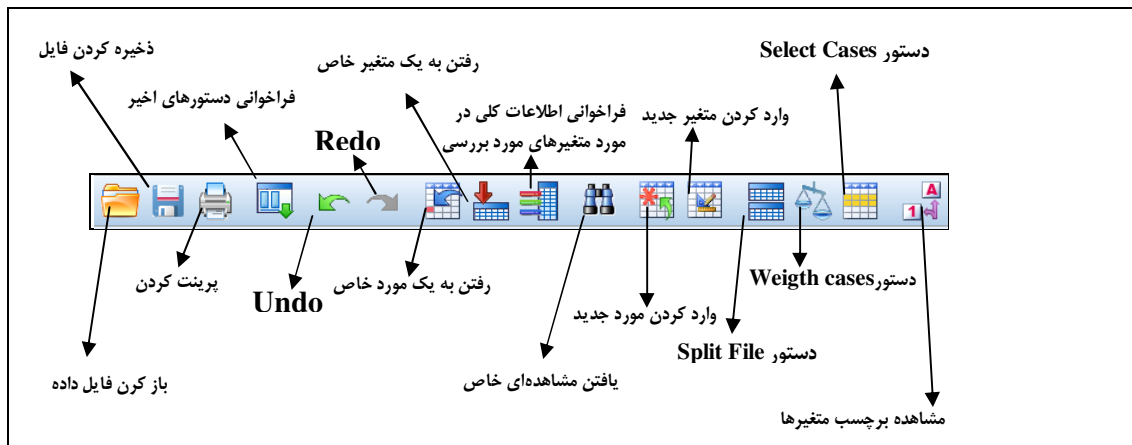
#### نوار منو:

همان طور که در شکل زیر مشاهده می کنید این نوار شامل منوهای زیادی می باشد که به مرور زمان با آنها آشنا می شوید (مهم ترین 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 پر کاربرد را آسان تر 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 آورده ایم.

	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 می توانید آدرس سلول فعال در صفحه Data View را مشاهده کنید. برای مثال در شکل زیر سلول فعال، "مشاهده چهارم از متغیر age است که مشاهده ۷۸ برای آن ثبت شده است". در نوار آدرس هم می توان 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

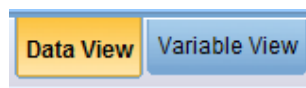
### صفحه Variable View

این صفحه نیز همان طور 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، می توانید 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.

IBM SPSS Statistics Processor is ready

### نحوه باز کردن و ذخیره کردن :

در این نرم افزار نیز مانند سایر نرم افزارهای تحت سیستم عامل ویندوز می توان از منوی File و گزینه های 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<sup>۱</sup> 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 اندازه گیری 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، در نظر 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 نرم افزار SPSS وارد کنیم. توجه کنید که این فرم شامل ۱۱ متغیر است (چرا؟).

<b>فرم جمع آوری اطلاعات</b>		
<b>بررسی وضعیت تغذیه ای بیماران دیابتی</b>		
کد بیمار:		
مدت بیماری از زمان شروع علائم (بر حسب ماه):		
جنسیت: مرد <input type="checkbox"/> زن <input type="checkbox"/>		
سن:	وزن:	قد:
مصرف سیگار: بلی <input type="checkbox"/> خیر <input type="checkbox"/>		
مصرف الکل: بلی <input type="checkbox"/> خیر <input type="checkbox"/>		
میزان کلسترول:		
میزان چربی:		
نوع رژیم دیابتیک: تزریق انسولین <input type="checkbox"/> قرصهای خوراکی <input type="checkbox"/>		

در شکل زیر اطلاعات مربوط به تعدادی بیمار، در نرم افزار SPSS وارد شده است. در نهایت در این بخش می‌خواهیم شما نیز بتوانید داده‌ها را اینگونه وارد SPSS کنید.

۱- این داده‌های آموزشی پس از نصب نرم افزار 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

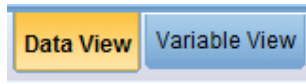
به طور کلی وارد کردن داده ها در SPSS شامل دو گام اساسی است:

۱- تعریف متغیرها ۲- ورود داده ها

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

### تعریف متغیرها در SPSS:

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



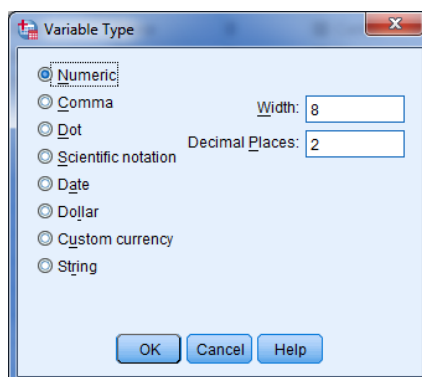
با رفتن به این پنجره، ستون های صفحه نرم افزار مشخصه های متغیر موردنظر را نشان می دهد، که تعریف یک متغیر به این نرم افزار با معرفی این مشخصه ها از آن متغیر به نرم افزار امکان پذیر است. در این بخش به معرفی این مشخصه ها می پردازیم.

### مشخصه Name:

اولین و ابتدایی ترین مشخصه یک متغیر که باید به نرم افزار معرفی شود، نام آن متغیر است. نام متغیر مورد نظر را باید طوری انتخاب کنید که در مراجعات بعدی به فایل داده ها دچار سردرگمی نشویم. همچنین حواسمان باشد که دادن نام به متغیر در SPSS شامل یکسری محدودیت هایی است و در صورت رعایت نکردن آنها با پیام خطایی از سوی نرم افزار مواجه خواهید شد. برای مثال نام متغیر شامل کاراکتر "فاصله" نمی شود، یا با "عدد" شروع نمی شود. 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 می کند که کاربر می خواهد از آن برای ورود داده های مربوط به متغیر مورد نظر در صفحه Data view استفاده کند. در بین انواع حالت های این مشخصه، دو نوع Numeric و String بیشتر مورد استفاده است. نوع Numeric برای کاراکترهای عددی و String برای کاراکترهای حرفی استفاده می شود.



برای ورود داده های مربوط به متغیرهای عددی (مانند وزن و قد) تنها از کاراکترهای عددی استفاده می کنیم. لذا مشخصه Type برای این نوع متغیرها ناگزیر Numeric خواهد بود. برای ورود داده های متغیرهای کیفی مانند "نام بیمار" تنها می توان از کاراکترهای حرفی استفاده کرد. لذا مشخصه Type برای این نوع متغیرها ناگزیر String خواهد بود. اما برای ورود داده های متغیرهای کیفی مانند "جنسیت" هم می توان از کاراکترهای عددی و هم از کاراکترهای حرفی استفاده کرد. لذا مشخصه Type برای این نوع متغیرها هم می تواند Numeric باشد و هم می تواند String باشد. بدین صورت که اگر Type را Numeric انتخاب کردید، باید از کاراکترهای عددی (برای مثال) ۱ و ۲ استفاده می کنید و در تعریف مشخصه های بعدی این متغیر، نحوه این کدبندی (کد ۱ برای مردان و کد ۲ برای زنان) را به نرم افزار معرفی می کنید، و اگر Type آن را String انتخاب کردید، از کاراکترهای حرفی (برای مثال) male و female استفاده خواهید کرد. حال سعی کنید مشخصه Type را برای متغیرهایی که در مرحله قبل تعریف کردید، مشخص کنید.

**مشخصه Width و Decimals:**

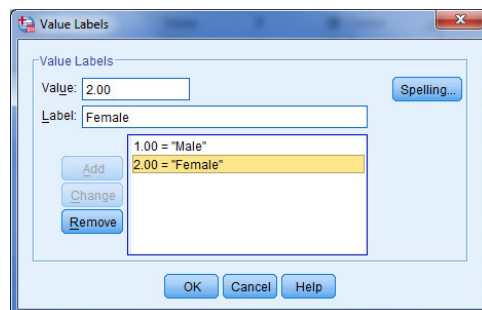
اگر نوع متغیر عددی باشد در این دو ستون تعداد کل ارقام عدد (Width) و تعداد رقم های اعشار (Decimals) را می توان تعیین کرد.

**مشخصه Label:**

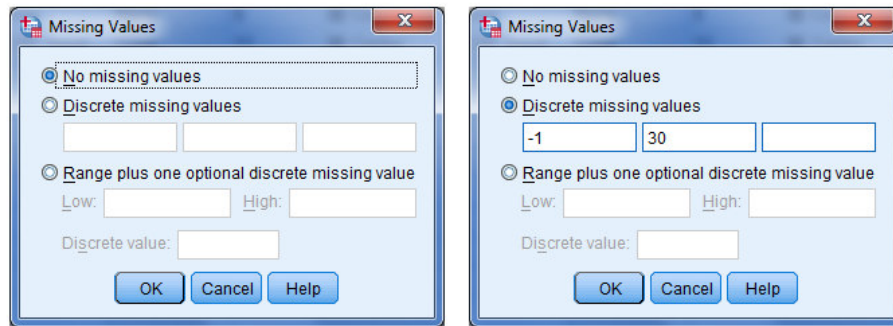
همان طور که گفتیم، دادن نام متغیرها در SPSS دارای محدودیت است و در صورتی که مایل باشیم می توانیم در این قسمت یک عنوان یا Label برای متغیر قرار بنویسیم. به عنوان مثال اگر متغیر مورد بررسی "طول مدت بیماری" باشد، می توانیم در قسمت Label تایپ کنیم: Duration of disease. توجه داشته باشید که اگر برای متغیر، عنوان تعریف شود در خروجی های SPSS به جای نام متغیر این عنوان نمایش داده خواهد شد در غیر این صورت نام متغیر به کار خواهد رفت.

**مشخصه Values:**

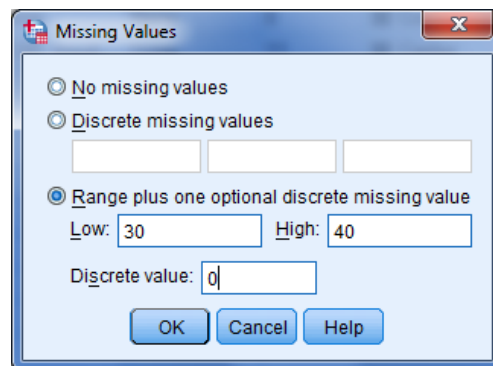
همان طور که گفتیم، اما برای ورود داده های متغیرهای کیفی مانند "جنسیت" هم می توان از کاراکترهای عددی و هم از کاراکترهای حرفی استفاده کرد و در صورتی که Type را Numeric انتخاب کردید، باید از کاراکترهای عددی (برای مثال) ۱ و ۲ استفاده می کنید. در این حالت برای تعریف نحوه این کدبندی (کد ۱ برای مردان و کد ۲ برای زنان) به نرم افزار از این خصوصیت استفاده می کنیم. 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 می شود. انتخاب گزینه اول به معنی عدم تعریف داده گمشده است. با انتخاب گزینه دوم می توان چند مقدار مشخص را برای معرفی یک مقدار گمشده تعیین کرد. به عنوان مثال اگر متغیر مورد بررسی تعداد واحد باشد می توان از مقادیری که خارج از دامنه تغییرات اند چند مقدار برای معرفی مقادیر گمشده تعیین کرد (مثلاً ۳۰ یا -۱).



با انتخاب گزینه سوم می‌توان یک بازه عددی را برای مشخص کردن مقادیر گم شده 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 گم شده 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 قرار گیرند.

#### مشخصه Measure:

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

#### مشخصه Role:

توسط این مشخصه نقش متغیر مورد نظر 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 می‌شود.

- ۳- نقش Both: برای متغیرهایی که هم نقش Input و هم نقش Output را در مدل می گیرند.
- ۴- نقش None: برای متغیرهایی که نقش خاصی در مدل آماری ندارند.
- ۵- نقش Portion: برای متغیرهایی مورد استفاده قرار می گیرد که داده های نمونه را به مجموعه های آموزشی، آزمودنی و اعتبار سنجی، تقسیم می کند.
- ۶- نقش Split: برای متغیرهایی است که نقش هماهنگ کننده بین نرم افزارهای تحت SPSS را دارند. این متغیر را هیچگاه نمی توان در دستور Split file استفاده کرد.
- توجه:** به طور پیش فرض 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 نرم افزار تاثیر خواهد گذاشت. بعد از تعریف تمامی متغیرهای لازم با کلیک روی عبارت Data view به محیط وارد کردن داده ها برگشته و داده ها را وارد کنید.

### تمرین ۲:

- ۱- فرم اطلاعاتی بیماران دیابتی را در نظر بگیرید. تمام مشخصه های مربوط به هر یک از متغیرهای این فرم را در پنجره Variable View وارد کنید. سپس داده های مربوط به آن را در صفحه Data view وارد کنید.
- ۲- قسمتی از یک پرسش نامه تحقیقاتی در زیر آمده است. پنجره Variable view را برای متغیرهای مربوط به این پرسش نامه 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 درآمد ماهیانه شما:  تومان

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

میزان تحصیلات: زیر دیپلم ☐ 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 فصل های 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 اختلاف های جنسیت شامل مطالعه نحوه بازی میمون ها 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

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 زیر ظاهر می گردد. جدول فراوانی مربوط به این داده ها 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 رابطه بین نوع شوک و زنده ماندن بیمار، از بیمارانی که به انواع شوک مبتلا می شوند، نمونه هایی انتخاب و آنها را بر حسب زنده ماندن و یا مردن در جدول زیر مرتب کرده اند.

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

داده های این جدول را در SPSS وارد کرده و جدول توافقی مربوط به آن را 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 می توان 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 ویژگی های 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 می رود. 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 روش های آماری گفته می شود که به پژوهشگر در طبقه بندی، 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 آنها 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 بگیرد:



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 می شود.

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

## 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 A, B, C, D 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:

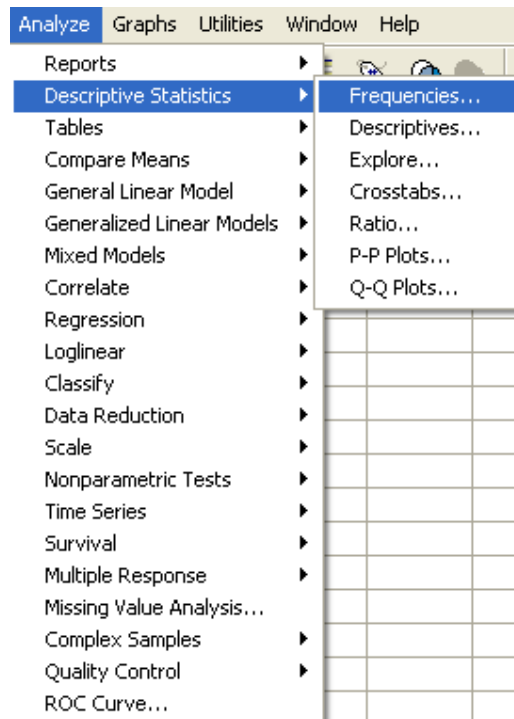
– چند عدد از قطعه C در این روز تولید شده است؟

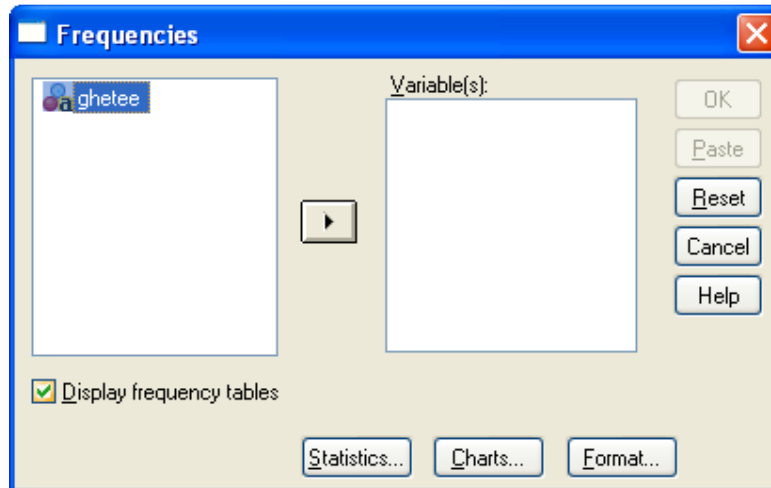
– قطعات A, B چند درصد از کل تولید روزانه را در بر می گیرند؟


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 می دهد. و جدول دوم جدول فراوانی متغیر است. ستون اول مقادیر متغیر، ستون دوم فراوانی 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.

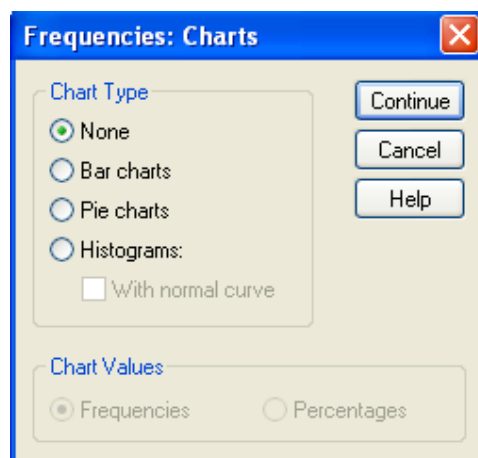
برای متغیرهای کیفی رتبه ای هم مراحل رسم جدول فراوانی به همین صورت خواهد بود.

### - نمودارهای آماری برای متغیرهای کیفی:

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

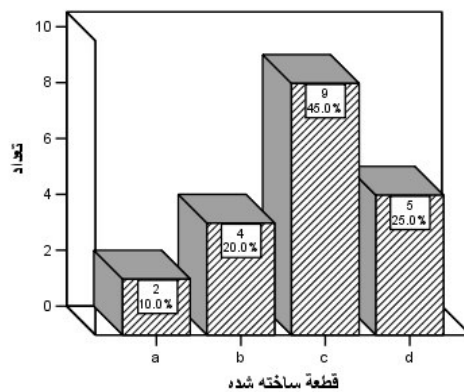
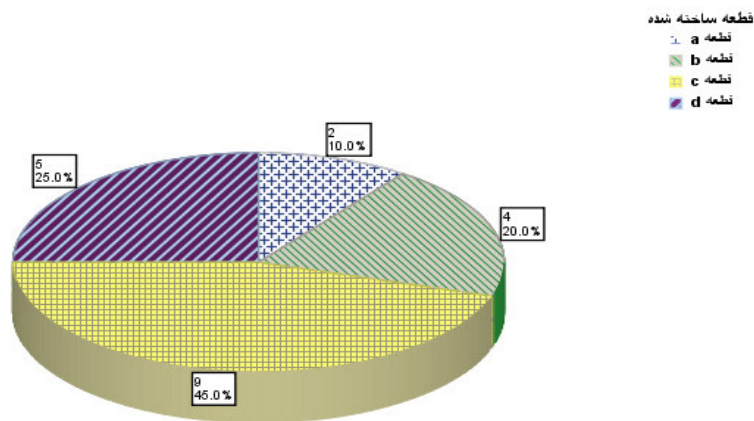
برای رسم این نمودارها می توان از دو راه زیر استفاده کرد:

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



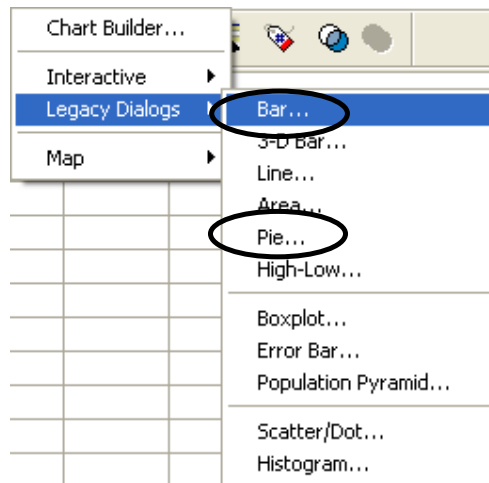
هر کدام از نمودارهای pie chart یا Bar chart را که مایل بودید انتخاب کرده و دکمه Continue را کلیک کنید. دکمه Ok را کلیک کنید.

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

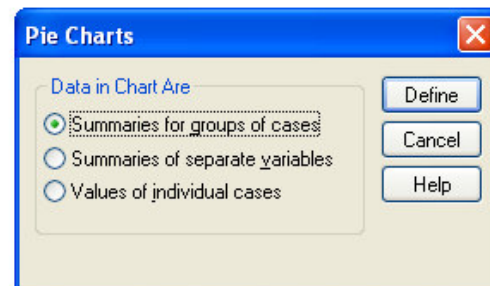
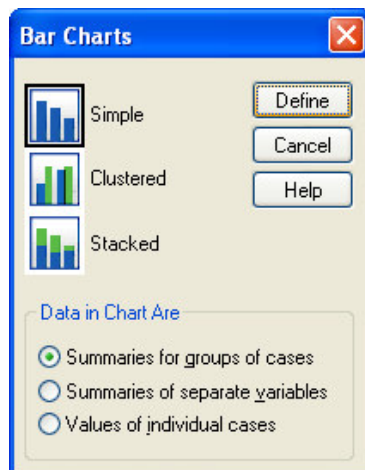


توجه کنید که نمودارهای دایره ای برای متغیرهای کیفی اسمی مناسبترند و نمودارهای میله ای برای متغیرهای کیفی رتبه ای.

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 گسسته 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 انتخاب شده به صورت زیر باشد:

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، 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.

همچون 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:** وزنه های 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

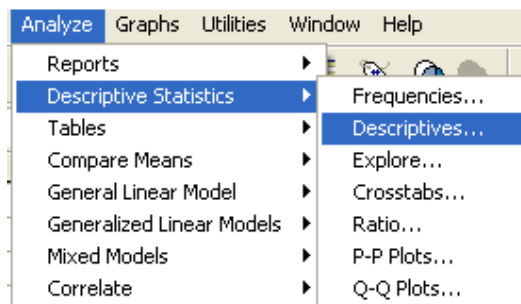
جدول فراوانی داده های فوق را رسم کنید.

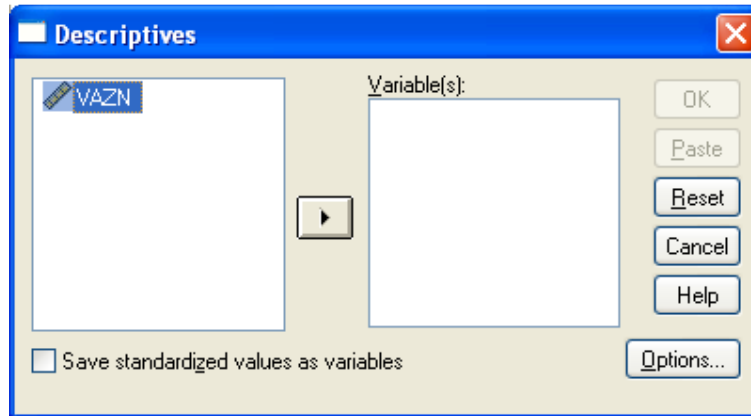
هرگاه داده های ما پیوسته باشد، داده ها را به تعدادی رده با طول مساوی تقسیم می کنیم و در هر رده فراوانی داده ها را می شماریم برای بدست آوردن تعداد رده ها در رسم جدول فراوانی به ترتیب زیر عمل کنید:

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

$$\text{Range} = \text{Max} - \text{Min}$$

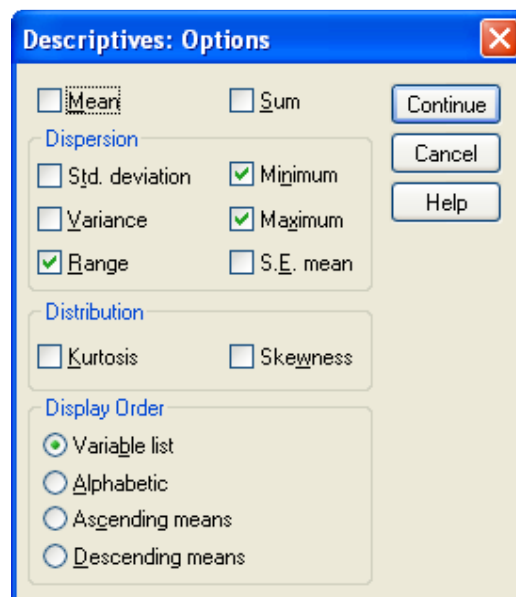
برای محاسبه دامنه تغییرات، کمترین داده ها و بیشترین داده ها از مسیر زیر استفاده می کنیم:





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 تعداد رده ها را بین ۵ تا ۲۵ رده

اختیار می کنند. یک قاعده مفید استفاده از دستور استورگس Sturges است:

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

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 می گیرد. طول هر طبقه را هم از فرمول زیر محاسبه می کنیم:

$$L=\text{Range}/m=4.574$$

طول طبقه

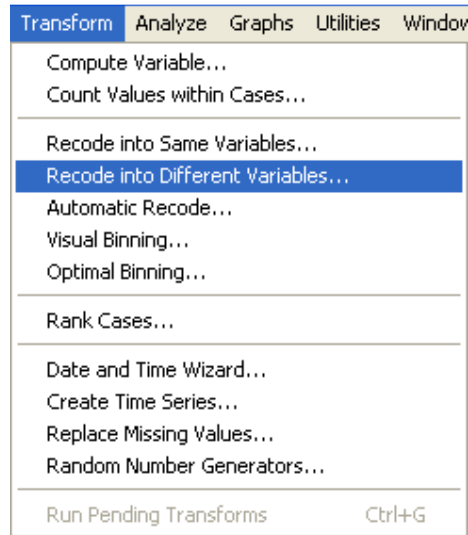
پس طول هر طبقه را ۵ در نظر می گیریم. حال باید ۷ طبقه به طول ۵ بسازیم. طبقات مورد نظر به صورت زیر خواهد بود:

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

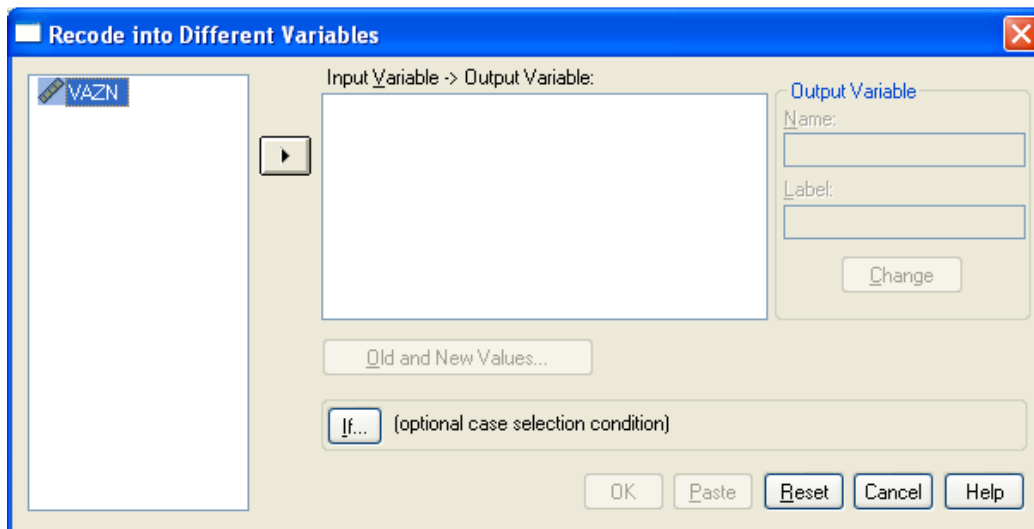
انتخاب حدود طبقات به صورت فوق باعث می شود که هر عدد دقیقاً در یک دسته قرار گیرد.



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:

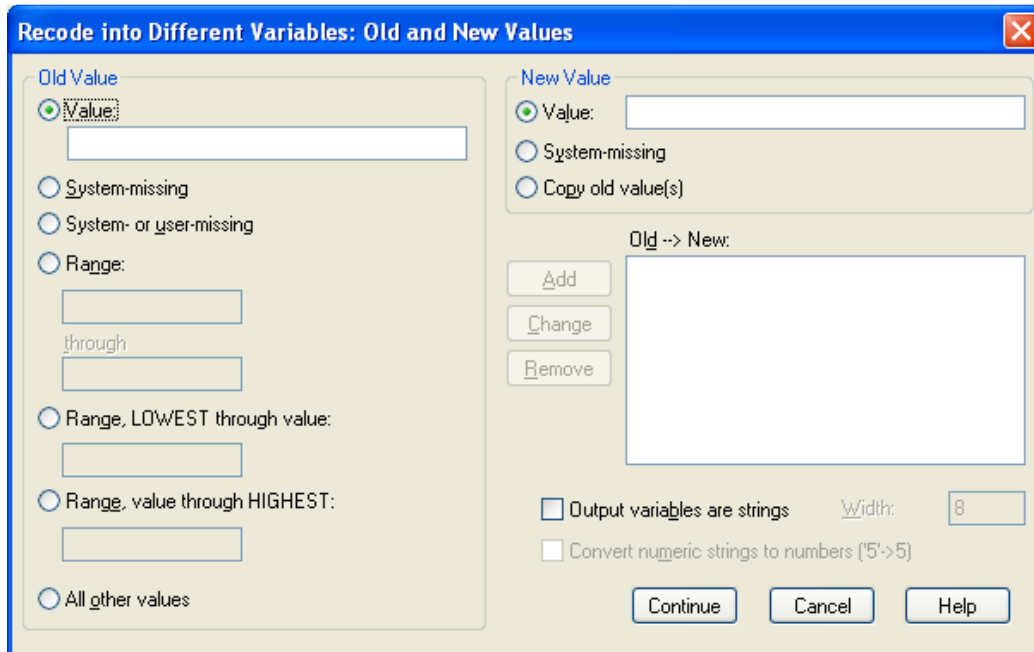




کادر گفتگوی زیر ظاهر می شود:



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:



**Recode into Different Variables: Old and New Values**

**Old Value**

- ☒ Value: [ ]
- ☐ System-missing
- ☐ System- or user-missing
- ☐ Range: [ ] through [ ]
- ☐ Range, LOWEST through value: [ ]
- ☐ Range, value through HIGHEST: [ ]
- ☐ All other values

**New Value**

- ☒ Value: [ ]
- ☐ System-missing
- ☐ Copy old value(s)

**Old -> New:**

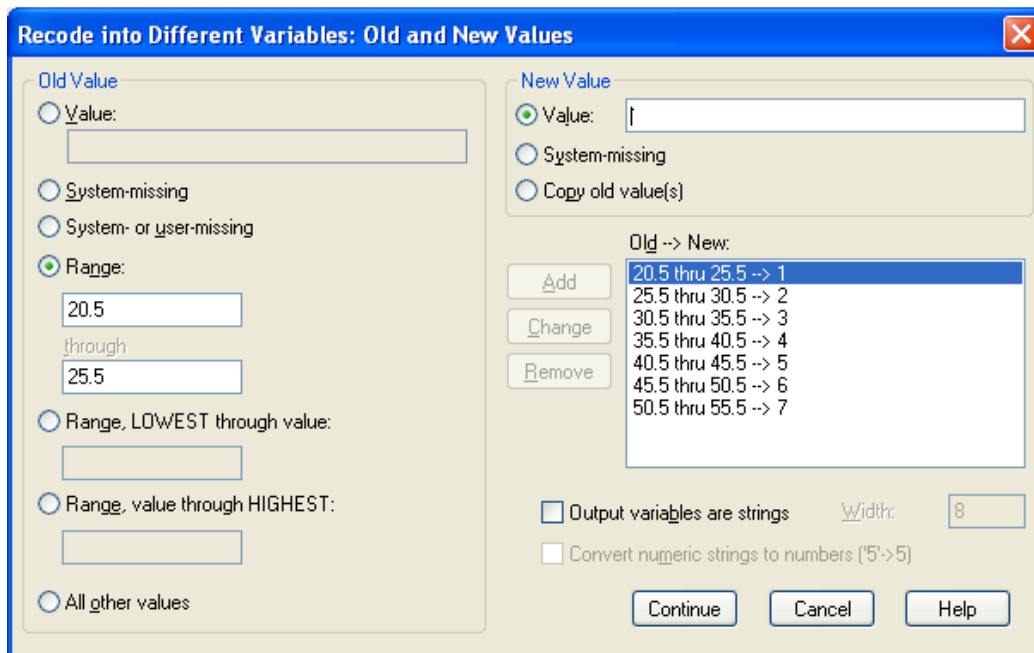
[Add] [Change] [Remove]

☐ Output variables are strings Width: [8]

☐ Convert numeric strings to numbers ('5'>5)

[Continue] [Cancel] [Help]

همانند تصویر زیر در سمت چپ کادر گفت و گوی بالا حدود طبقات را وارد کرده، 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.



**Recode into Different Variables: Old and New Values**

**Old Value**

- ☐ Value: [ ]
- ☐ System-missing
- ☐ System- or user-missing
- ☒ Range: [20.5] through [25.5]
- ☐ Range, LOWEST through value: [ ]
- ☐ Range, value through HIGHEST: [ ]
- ☐ All other values

**New Value**

- ☒ Value: [1]
- ☐ System-missing
- ☐ Copy old value(s)

**Old -> New:**

[Add] [Change] [Remove]

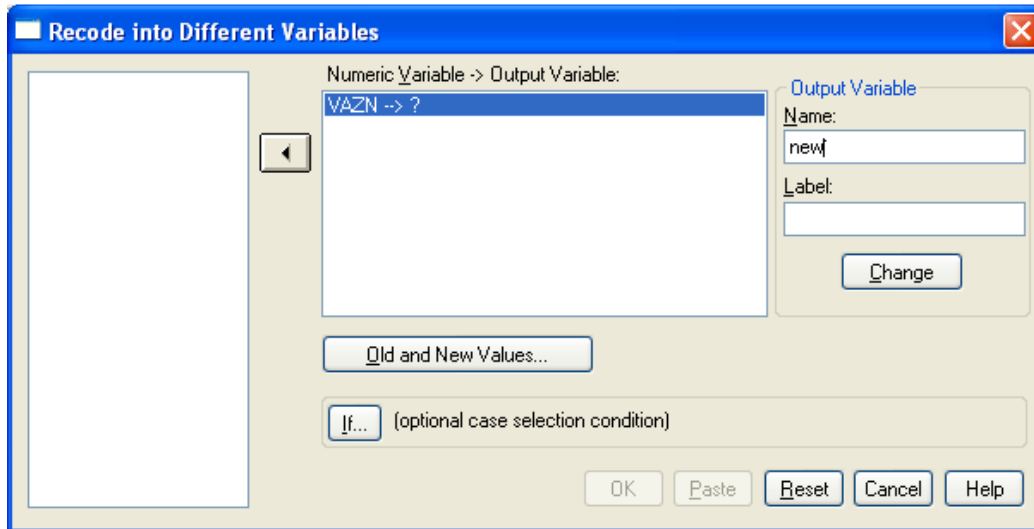
20.5 thru 25.5 -> 1  
25.5 thru 30.5 -> 2  
30.5 thru 35.5 -> 3  
35.5 thru 40.5 -> 4  
40.5 thru 45.5 -> 5  
45.5 thru 50.5 -> 6  
50.5 thru 55.5 -> 7

☐ Output variables are strings Width: [8]

☐ Convert numeric strings to numbers ('5'>5)

[Continue] [Cancel] [Help]

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.

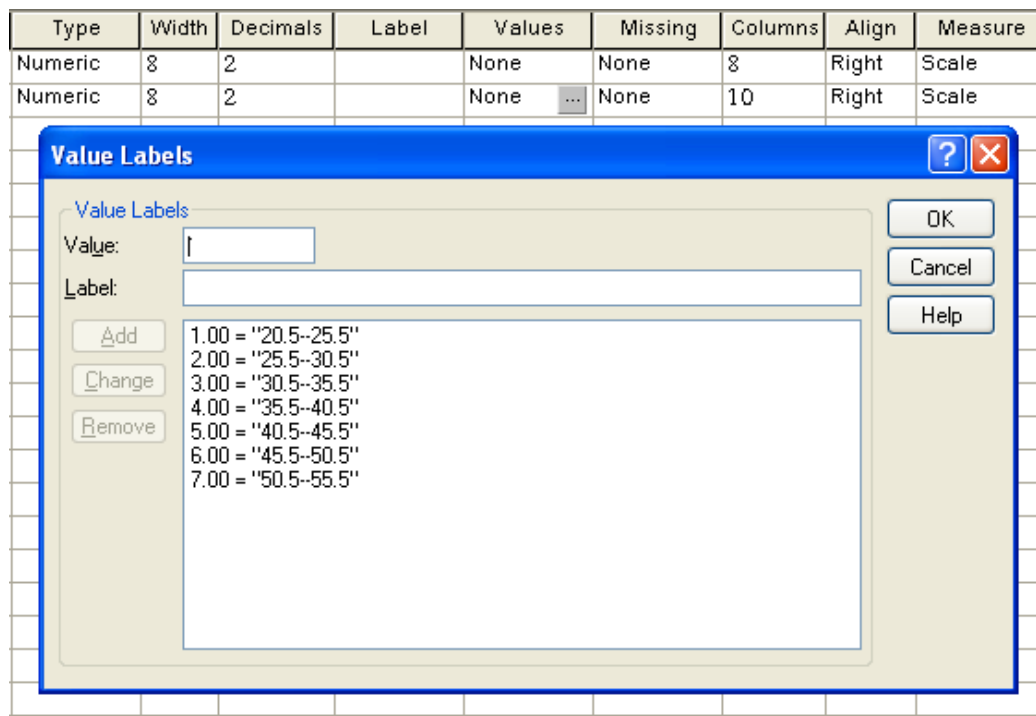


در پایان دکمه OK را کلیک کرده و نتیجه کار را به صورت زیر مشاهده کنید:

	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:



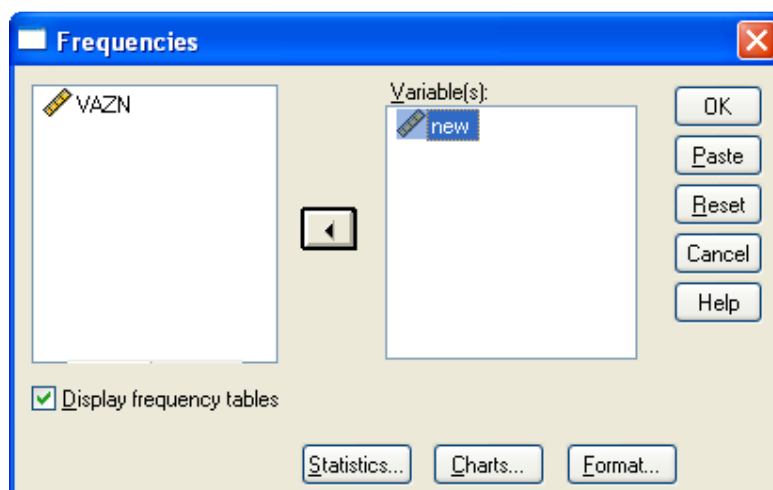
به قسمت Data view برگردید.

اگر برای مقادیر یک متغیر، بر چسب تعریف شده باشد، نمایش مقادیر آن، به دو صورت خواهد بود:

مقادیر متغیر و بر چسب‌های مقادیر. با کلیک روی دکمه زیر می‌توانید 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).



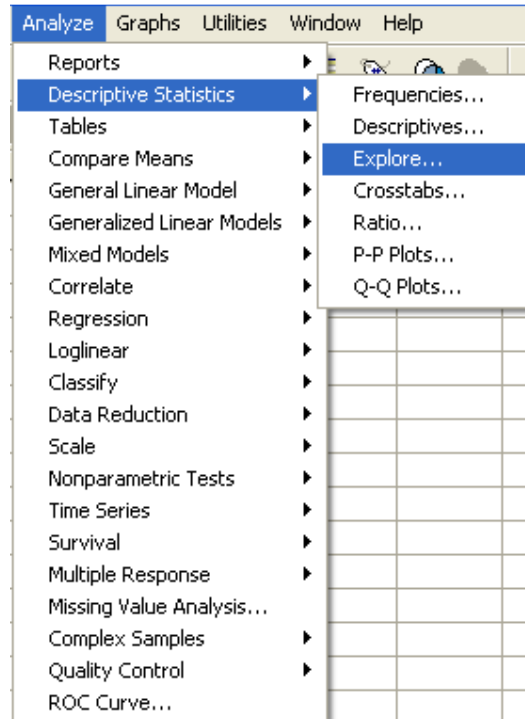
جدول فراوانی داده های مثال وزن قالبهای کره به صورت زیر خواهد بود:

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	

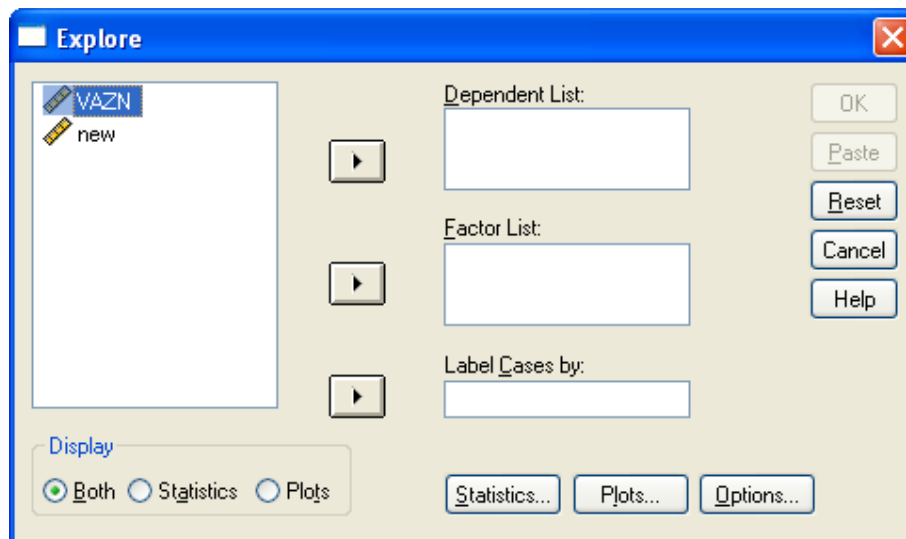
- رسم نمودار برای داده های کمی پیوسته:

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:



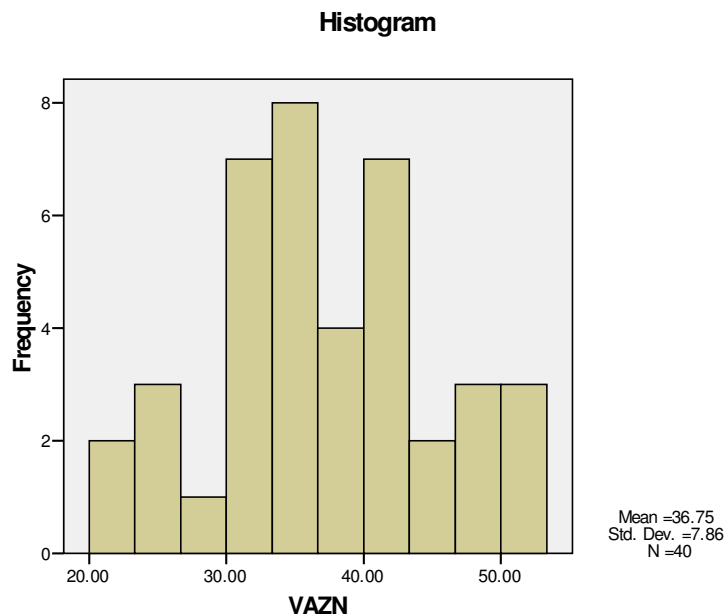
کادر گفتگوی زیر ظاهر خواهد شد:



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 بسیار زیاد شده است نمودار جعبه ای یا Box plot است که به آن بعد از بحث شاخصهای آماری خواهیم پرداخت.

### 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: شاخصهایی که گرایش به مرکز یا مرکزیت داده ها را اندازه میگیرد (شاخصهای مرکزی) و شاخصهایی 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 پر کاربرد 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 مرتبه  $p$ ، مقداری است که تقریباً  $100p$  درصد داده ها از آن کمتری مساوی آن و  $100(1-p)$  درصد داده ها از آن بیشترند. 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 در صورتیکه 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 شاخصهایی که گرایش داده ها را به یک مقدار مرکزی نشان میدهد، علاقه مند به شاخصهایی 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.

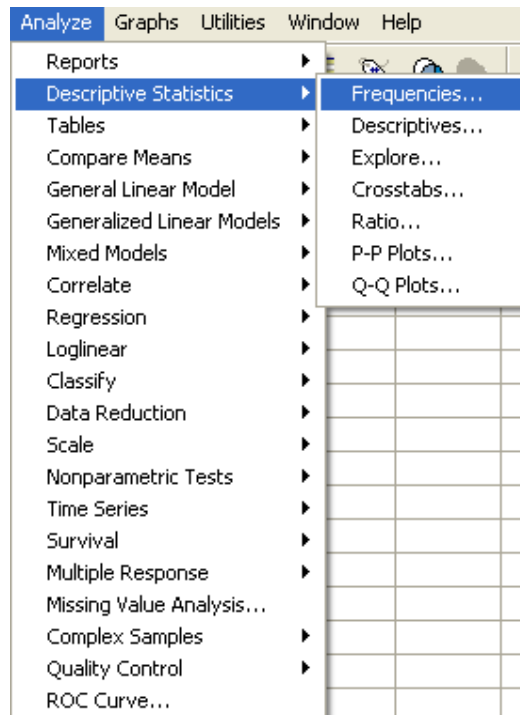
**واریانس (Variance):** میانگین مربعات تفاضل داده ها از میانگین را واریانس گویند.

**انحراف معیار (Standard deviation):** 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.

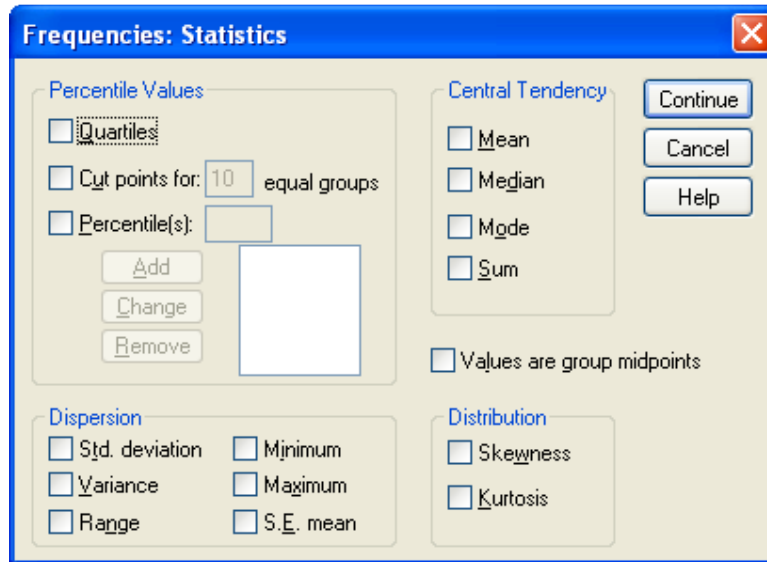
برای محاسبه شاخصهای بالا، ابتدا مسیر زیر را طی کنید:



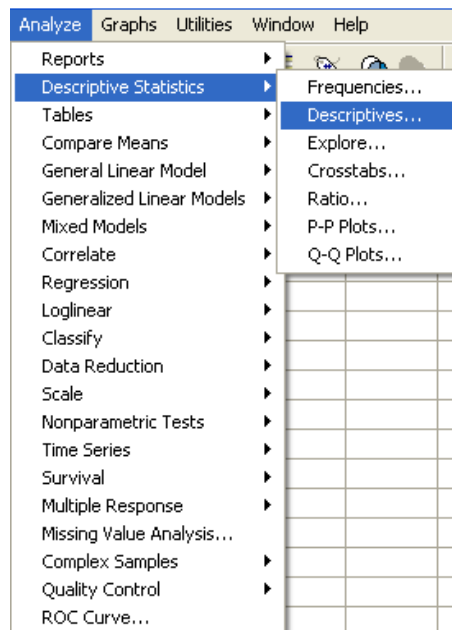
روی دکمه **Statistics...** کلیک کنید. کادر زیر باز خواهد شد. شاخصهای مرکزی را میتوانید از کادر

Central Tendency و شاخصهای پراکندگی را از کادر Dispersion انتخاب کنید.

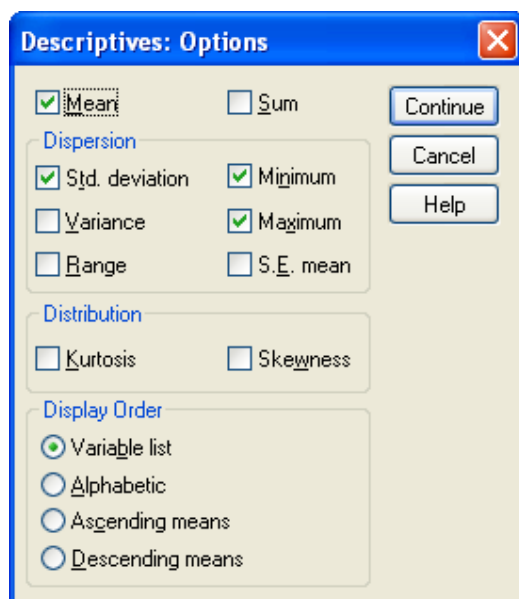
پس از انتخاب شاخصهای مورد نظر دکمه Continue و سپس OK را کلیک کنید.



و یا میتوان از مسیر زیر استفاده کرد:

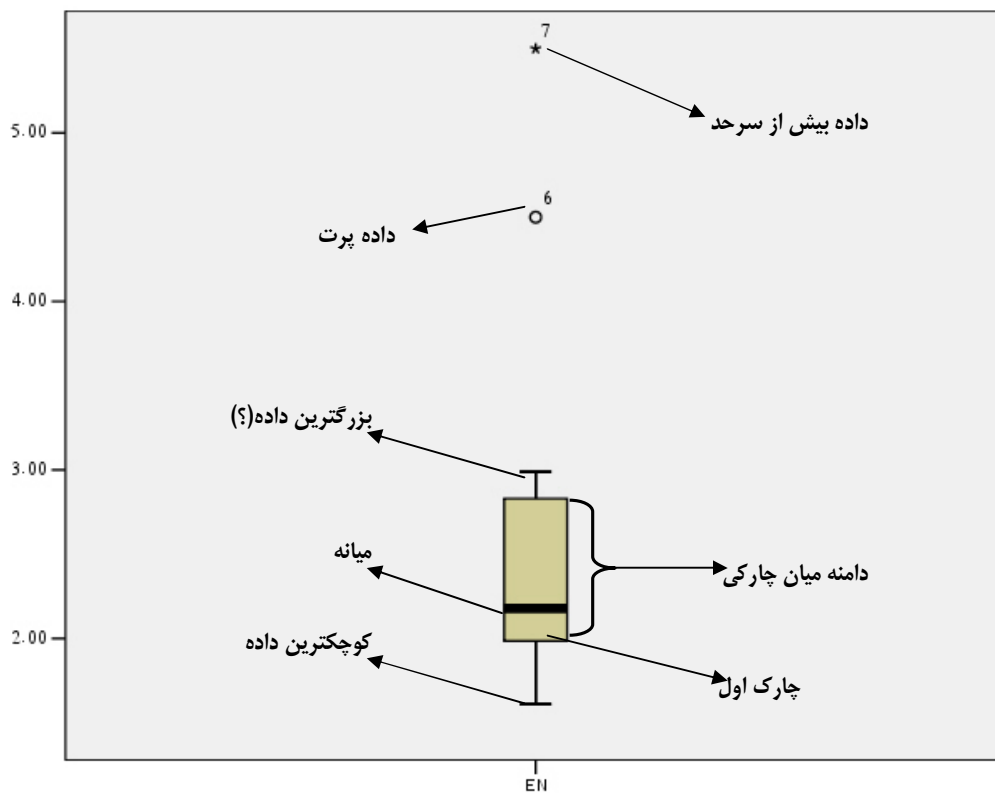


روی دکمه Option... کلیک کنید و شاخصهای موردنظرتان را انتخاب کنید:



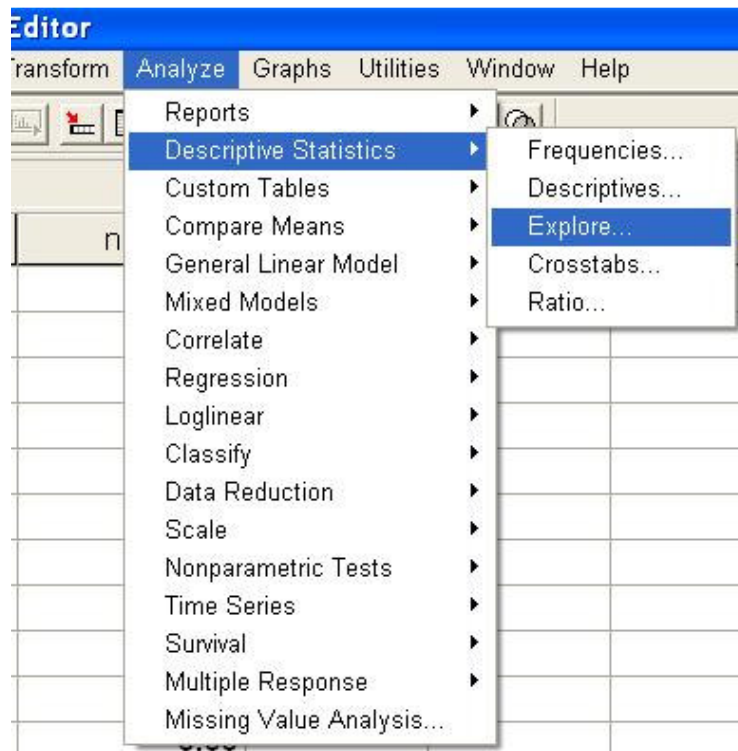
### نمودار Box plot :


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:



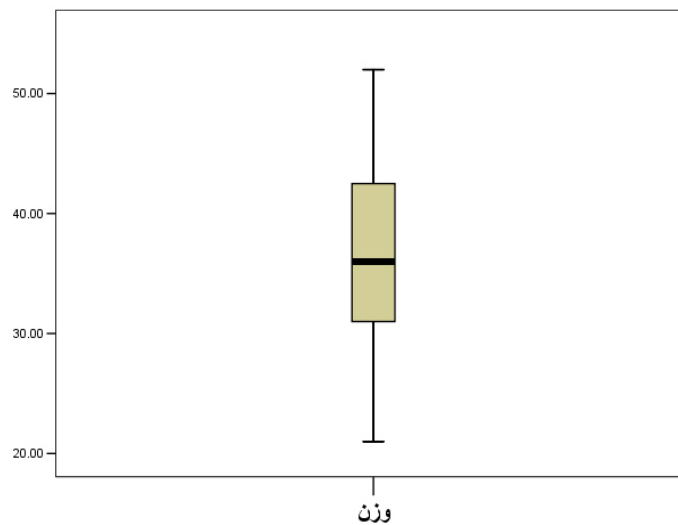
فاصله میان چارک اول و چارک سوم را دامنه میان چارکی می گویند. اگر داده ای بیش از ۱.۵ برابر دامنه میان چارکی از چارک اول یا چارک سوم فاصله داشته باشد، داده پرت محسوب میشود و به صورت یک نقطه خارج از نمودار نشان داده میشود و نمودار بدون احتساب آن رسم می شود.

برای رسم Box plot مسیر زیر را طی کنید:



روی  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:



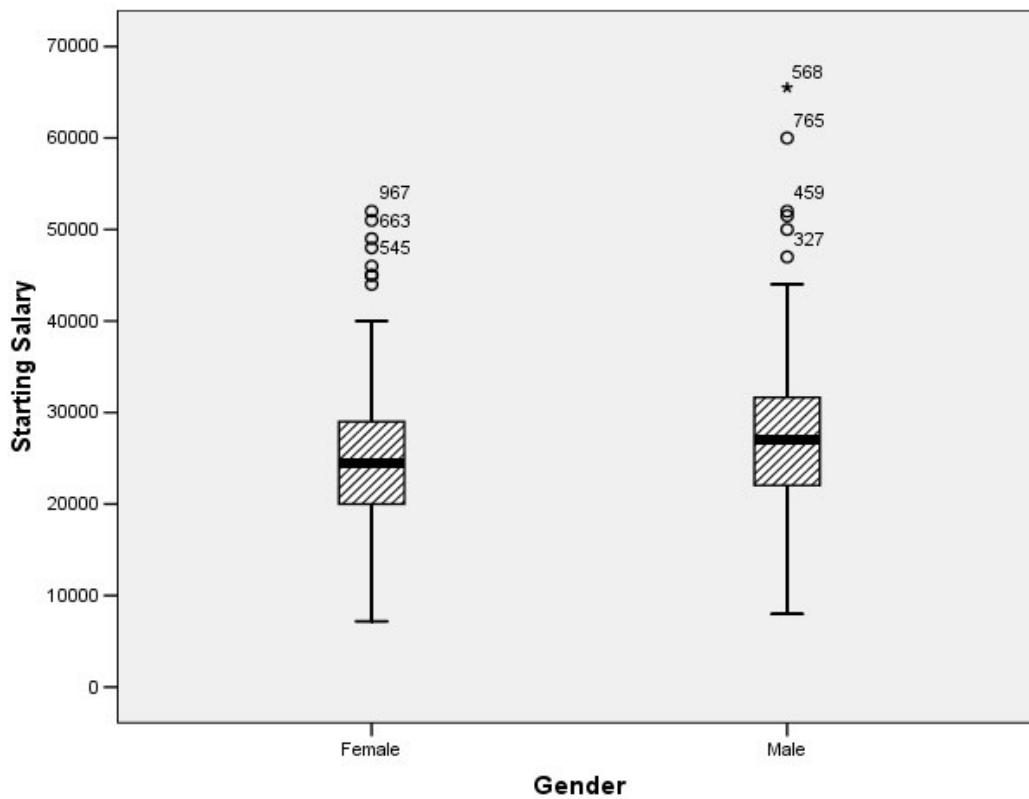
## ۲-۲-۴ چند نمودار مهم و کاربردی دیگر:

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 نحوه رسم آنها را نیز در سر کلاس با هم بررسی می‌کنیم.

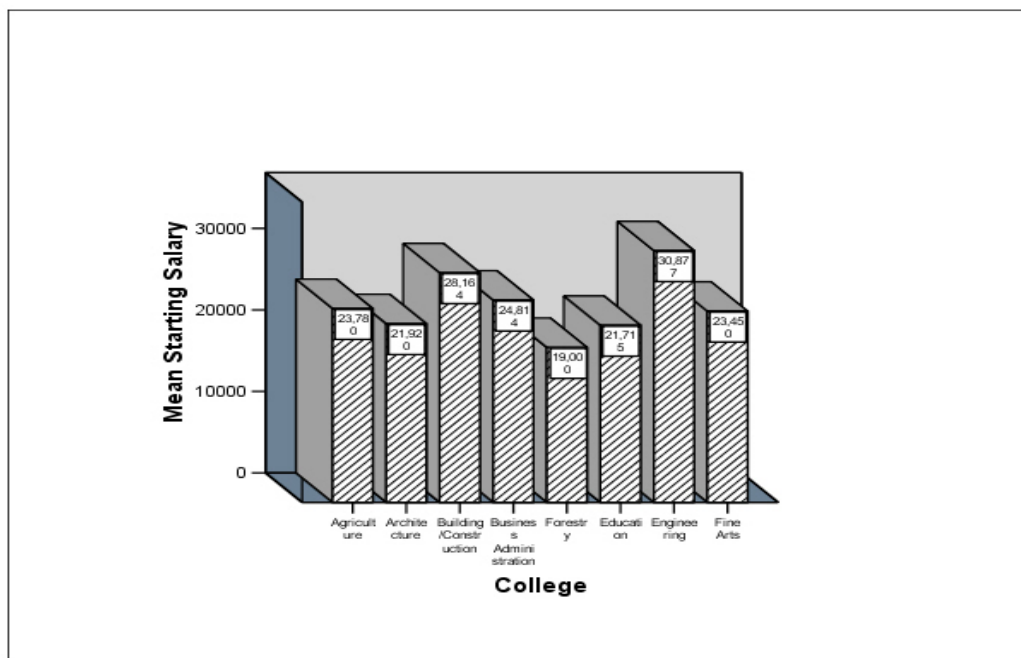
### ۱) زمانی که تنها یک متغیر عددی و یک متغیر رسته‌ای داشته باشیم:

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 مقایسه دو جامعه ) :

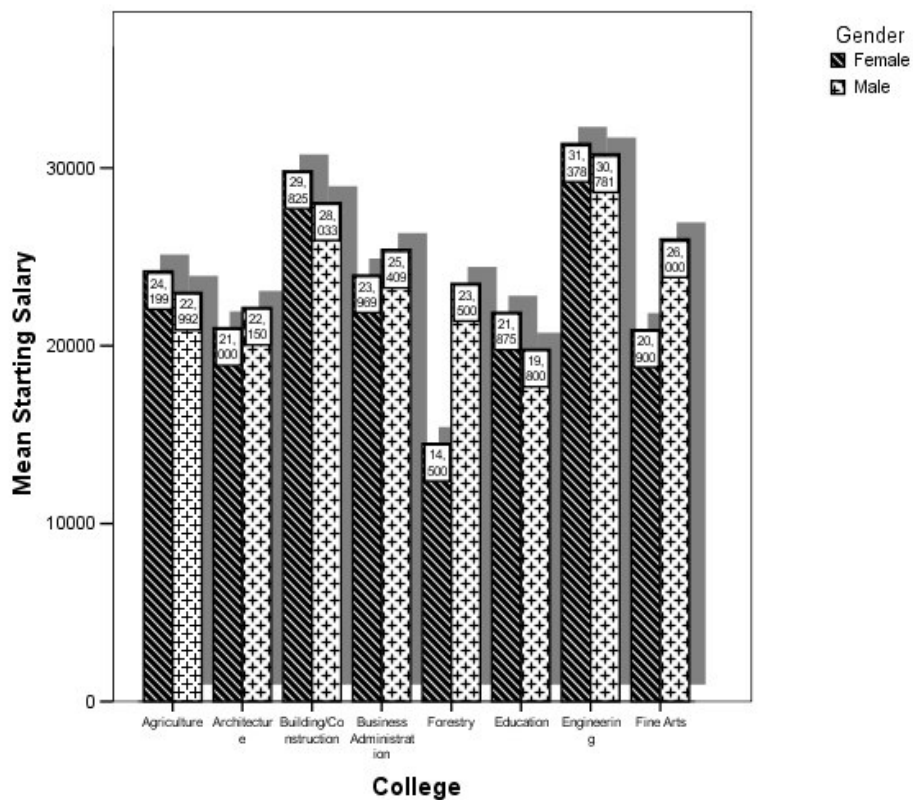




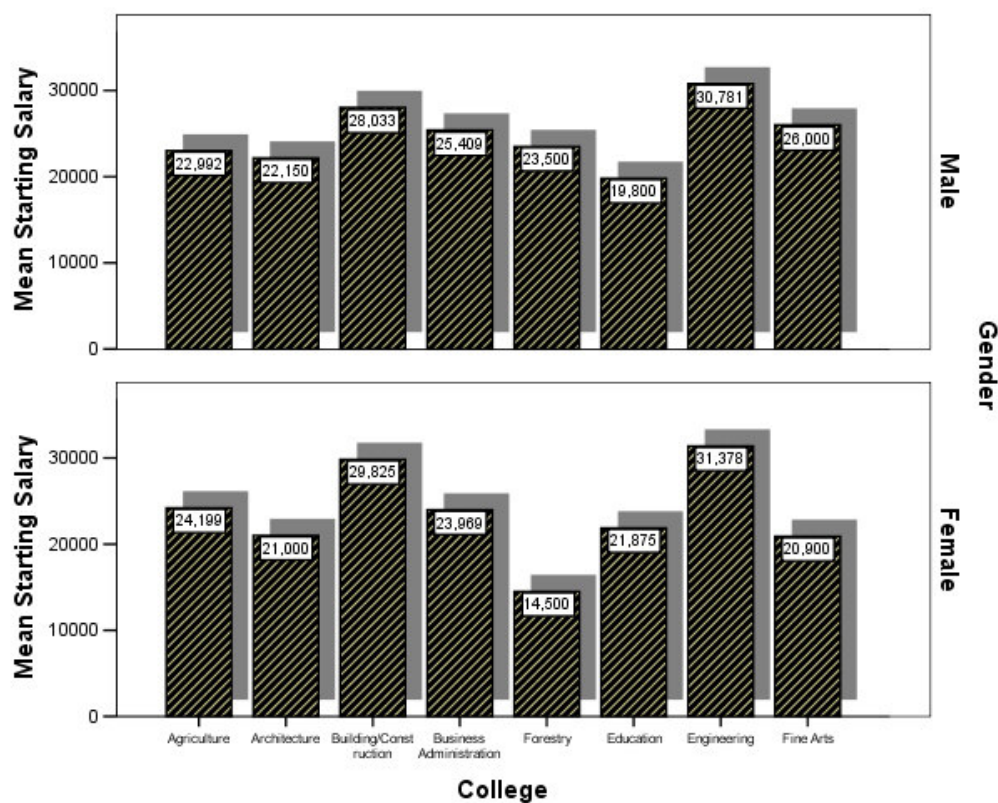
## ۲) زمانی که تنها یک متغیر عددی و دو متغیر رسته‌ای داشته باشیم:

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


### دستورهای برای دست کاری داده ها در SPSS

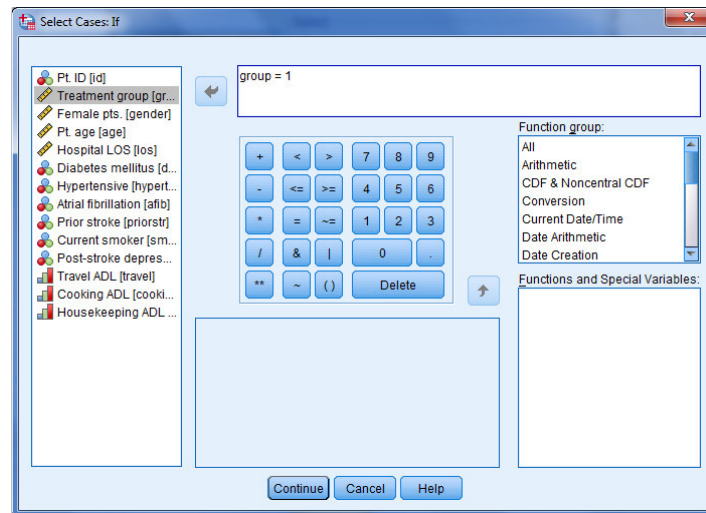
در این فصل به بررسی برخی دستورات در SPSS خواهیم پرداخت که برای کار با داده ها و آماده کردن آنها برای تحلیل های بعدی بسیار مهم است. به اعتقاد برخی نویسندگان، در حین کار با SPSS استفاده از این دستورها بسیار کمک کننده است و بدون وجود این دستورها، این نرم افزار قابل استفاده نیست.

#### ۳-۱ دستور Select Cases

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

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

می دانیم که برای بررسی این رابطه باید از دستور Crosstabs استفاده کرد، اما چون می خواهیم 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 داده ها فقط افراد مورد انتخاب شده اند و افراد کنترل برای آنالیزها فیلتر شده اند.



حال با اجرای دستور Crosstabs رابطه بین فشارخون و دیابت را بررسی می کنیم. خروجی نرم افزار به صورت زیر است:

Diabetes mellitus \* Hypertensive Crosstabulation

Count			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 <sup>a</sup>	1	.004		
	Continuity Correction <sup>b</sup>	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

همان طور که مشاهده می کنید، 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 Case ها را تعیین می کنیم، گزینه های زیاد دیگری نیز برای انتخاب وجود دارد. برای مثال می توان با استفاده صفحه کلید این کادر تعیین کنیم که، برای مثال، "افراد با سن کمتر یا مساوی ۷۳ سال" انتخاب شوند. برای این کار باید در کادر تایپ شود:  $age \leq 30$ . یا می توان از دستورهای منطقی & (and) و | (or) استفاده کرد.

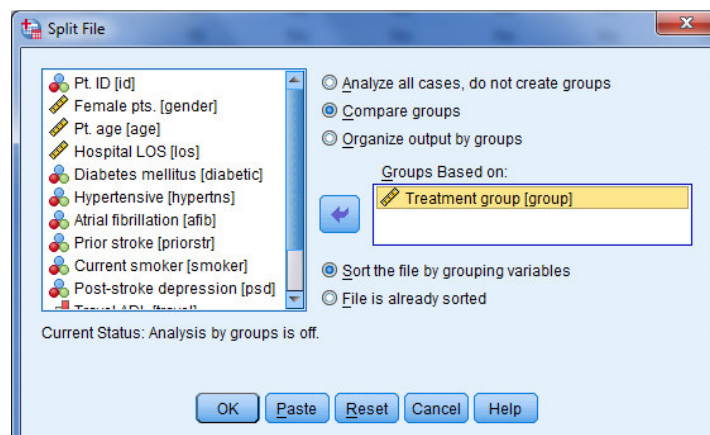
## تمرین:

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

## ۲-۳ دستور Split File

همان طور که از نام این دستور پیداست، فایل داده ها را می شکند. به عبارتی با در نظر گرفتن یک متغیر کیفی، مثل جنسیت یا مقطع تحصیلی، فایل داده ها تفکیک شده و تمام آنالیزها و خروجی ها به تفکیک طبقات این متغیر کیفی ارائه خواهد شد.

برای بررسی این دستور ابتدا فایل داده "adl.sav" را از فایل های نمونه SPSS باز کنید. می خواهیم رابطه بین فشارخونی بودن (hypertns) و ابتلا به بیماری دیابت (diabetic) را به تفکیک هر یک از گروه های تیماری (group) بررسی کنیم. می دانیم که برای بررسی این رابطه باید از دستور Crosstabs استفاده کرد، اما چون می خواهیم 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 مورد نظر 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 داده ها تفکیک شده اند. و حالا می توانید آنالیز مورد نظر را انجام دهید.

بعد از اجرای دستور Crosstabs خروجی را به صورت زیر خواهید دید. همان طور که مشاهده می کنید، 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 <sup>a</sup>	1	.000	.000	.000
	Continuity Correction <sup>b</sup>	13.412	1	.000		
	Likelihood Ratio	15.600	1	.000		
	Fisher's Exact Test					
	Linear-by-Linear Association	16.398	1	.000		
	N of Valid Cases	46				
Treatment	Pearson Chi-Square	8.172 <sup>c</sup>	1	.004	.008	.008
	Continuity Correction <sup>b</sup>	5.950	1	.015		
	Likelihood Ratio	8.196	1	.004		
	Fisher's Exact Test					
	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 به خصوص 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 دهید.

۱- رابطه بین فشارخون و دیابت را یکبار به صورت 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 :


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

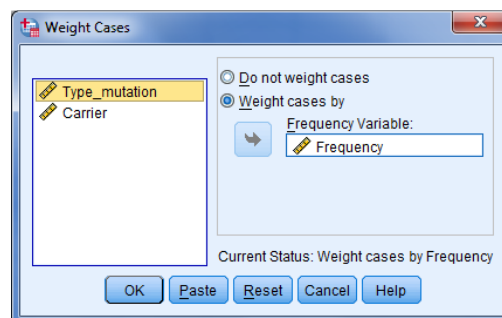
می‌خواهیم آزمون کنیم که "نوع انحراف از وضع طبیعی" مستقل از "حامل بودن والدین" است یا خیر. برای انجام این آزمون قانداً باید ابتدا داده‌ها را در SPSS کرد!! که البته در این قسمت تنها بر نحوه ورود داده‌ها بحث خواهیم کرد و نه بر نحوه انجام آزمون آماری آن. برای این کار دو راه وجود دارد. **راه اول** این که داده‌ها را به طور 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 به صورت 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، کافیت 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		Total
		one parent	none of parent	
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 به خصوص 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 دهید. (دستور Weight cases)
- ۱- 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.

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

۲- به منظور بررسی وجود یا عدم وجود رابطه بین نوع شوک و زنده ماندن بیمار، از بیمارانی که به انواع شوک مبتلا می شوند. نمونه هایی انتخاب و آنها را بر حسب زنده ماندن و یا مردن در جدول زیر مرتب کرده اند.

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

داده های این جدول را در SPSS وارد کرده و جدول توافقی مربوط به آن را رسم کنید.

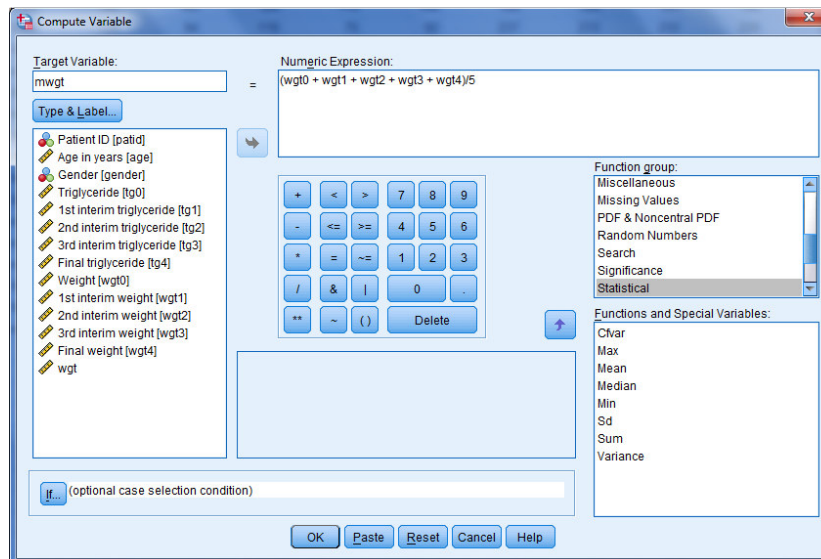
### ۳-۴ دستور Compute

بسیاری اوقات پیش می آید که محقق می خواهد عبارتی را به طور سطری و برای هر یک از موردها (افراد) محاسبه کند. برای مثال فرض کنید قد و وزن نمونه را داریم و می خواهیم برای هر فرد به طور جداگانه BMI (یعنی  $^2$  (قد/وزن) را محاسبه کنیم. به عنوان مثالی دیگر فرض کنید در یک تست سنجش افسردگی که شامل ۱۵ سؤال ۵ گزینه ای (۱ تا ۵) است، می خواهیم امتیاز افسردگی هر شخص را به طور جداگانه محاسبه کنیم. در تمام این موارد دستور Compute می تواند به ما در محاسبه عبارت مورد نظر کمک کند.

برای بررسی این دستور ابتدا فایل داده "dietstudy" را از فایل های نمونه SPSS باز کنید. می خواهیم میانگین وزن افراد در ۵ نوبت پیگیری (wgt0 تا wgt4) را محاسبه کنیم و در متغیری به نام mwgt، در همین فایل داده ها 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 می توانید 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 کنید حقوق اولیه (salbegin) و حقوق جاری (salary) شخص نام را به ترتیب با  $X_i$  و

$Y_i$  نشان دهیم. مقادیر زیر را محاسبه کنید. (دستور Compute variables)

$$\sum_i X_i^2 - 474\bar{X} = \quad \sum_i Y_i^2 - 474\bar{Y} =$$

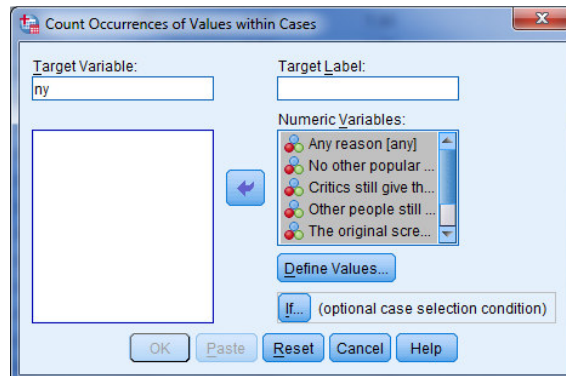
$$\sum_i (X_i + Y_i)(X_i - Y_i) =$$

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

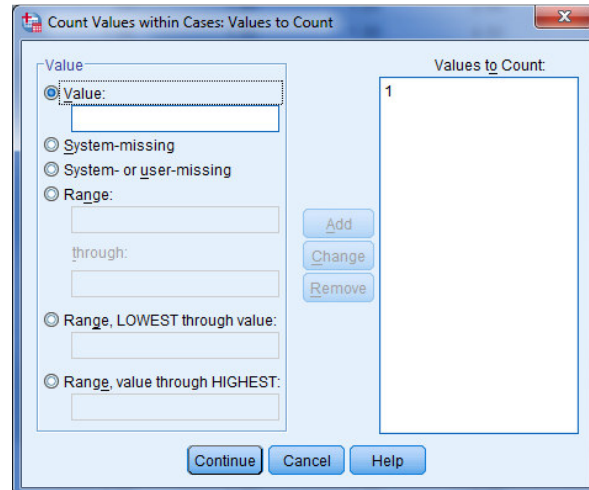
۲- هر یک از متغیرهای salbegin و salaty استاندارد کنید و میانگین و واریانس هر یک را بدست آورید.

### ۳-۵ دستور Count Values:

یک تست سنجش افسردگی را در نظر بگیرید. ممکن است محقق بخواهد بداند هر فرد در پاسخ گویی چند بار گزینه ۳ را انتخاب کرده است. برای پاسخ به این سؤال باید از دستور Count values استفاده کرد. برای بررسی این دستور ابتدا فایل داده "tv-survey.sav" را از فایل های نمونه SPSS باز کنید. این فایل مربوط به یک نظرسنجی در مورد برنامه های تلویزیونی می باشد، که شامل ۷ سؤال بلی و خیر است. می خواهیم بدانیم هر فرد به طور کلی چند مرتبه به این سؤالات 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 می خواهیم 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 آن چه را که می خواهید شمارش شود را اضافه کنید. در این



مثال می خواهیم عدد ۱ شمارش شود. لذا عدد ۱ را در قسمت values تایپ کرده و عبارت Add را کلیک می کنیم، سپس دکمه Continue و Ok را کلیک می کنیم. حال به پنجره داده ها برگردید. خواهید دید که متغیری تحت عنوان ny به متغیرهای قبلی اضافه شده است که در واقع تعداد پاسخ های بله هر فرد است.

**توجه:**

- ۱- در کادر مکالمه‌ای Count values within cases: values to count می‌توان نحوه شمارش را به صورت 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. (دستور Count values)
- ۱- جدول فراوانی مربوط به تعداد سکتها در سه نوبت پیگیری افراد (stroke1 و stroke2 و stroke3) را رسم کنید.
- ۲- تمرین (الف) را به تفکیک هر بیمارستان انجام داده و بگوئید، در کدام بیمارستان بیشترین تعداد سکت 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 می پردازیم.

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

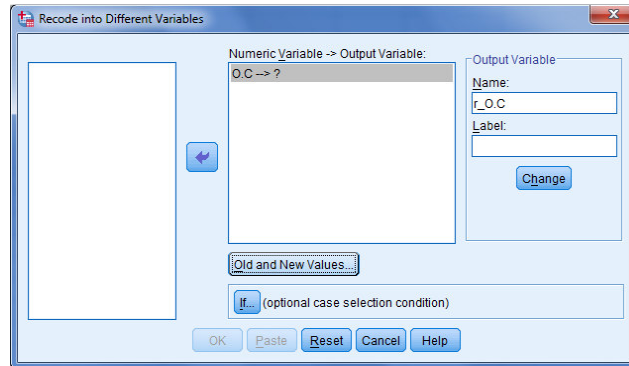
یک روش تعیین طبقات در فصل دوم ارائه شد. فرض کنید حدود طبقات را با این روش تعیین کردیم، در این بخش می خواهیم این حدود طبقات را به نرم افزار معرفی کنیم تا داده ها را دسته بندی کند و به وسیله آن جدول فراوانی داده ها را رسم کنیم.

**مثال ۳)** داده های مثال ۳ (تمرکز اوزن در ۴۰ شهر بزرگ) را در SPSS وارد کنید و نام متغیر را O.C بگذارید. برای دسته بندی داده ها با استفاده از SPSS، در **گام اول**، حدود طبقات را به آن تعریف می کنیم و متغیر دسته بندی شده را به عنوان متغیری جدید در صفحه داده ها ایجاد می کنیم. حدود طبقاتی که باید تعریف شوند، به این صورت است.

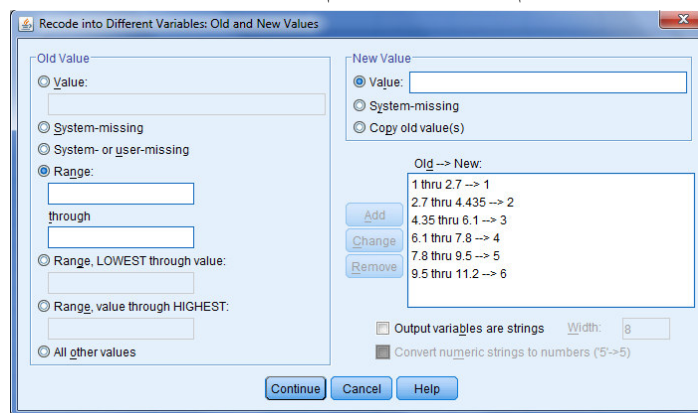
کد وارد شده	طبقات جدید	طبقات قدیم
۱	۱-۲/۷	۱-۲/۷
۲	۲/۷-۴/۳۵	۲/۷-۴/۴
۳	۴/۳۵-۶/۱	۴/۴-۶/۱
۴	۶/۱-۷/۸	۶/۱-۷/۸
۵	۷/۸-۹/۵	۷/۸-۹/۵
۶	۹/۵-۱۱/۲	۹/۵-۱۱/۲

**توجه داشته باشید** عدد ۴/۴ که یکی از کران هاست (کران بالای طبقه دوم و کران پائین طبقه سوم) در داده ها نیز وجود دارد. و از آنجایی که می خواهیم ۴/۴ در طبقه سوم باشد، در طبقه دوم و سوم به جای عدد ۴/۴ در کرانه ها، عدد ۴/۳۵ را در نظر گرفته ایم و اگر اعداد ۲/۷، ۱/۶، ۷/۸ یا ۹/۵ نیز در داده ها بودند، همین کار را انجام می دادیم.

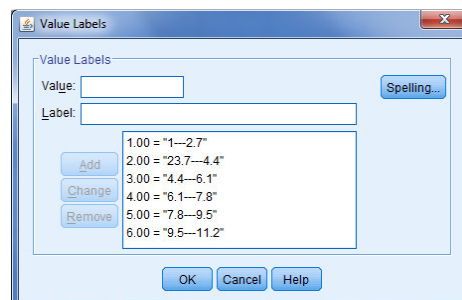
برای دسته بندی داده ها مسیر **Transform>Recode into different variables** را انتخاب کنید تا کادر مکالمه **Recode into different variables** باز شود. در این کادر متغیر "مقدار تمرکز ازن (O.C)" را انتخاب و وارد پنجره سمت راست (Numeric Variable->Output variable) می کنیم. در قسمت سمت راست این کادر (Output Variable) و در قسمت Name، نام متغیر جدید، یعنی O.C\_x را تایپ می کنیم و سپس دکمه Old and New values را کلیک می کنیم تا کادر Old and New values: 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 می کنیم و دکمه Add را می زنیم.



به همین ترتیب سایر طبقات را به نرم افزار معرفی می کنیم.



پس از اینکه معرفی طبقات به نرم افزار به اتمام رسید، دکمه های Continue، Change و Ok را به ترتیب کلیک می کنیم. حال به پنجره داده ها 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 Value ای در خور، مانند شکل زیر، در نظر بگیریم (در پنجره Variable view).



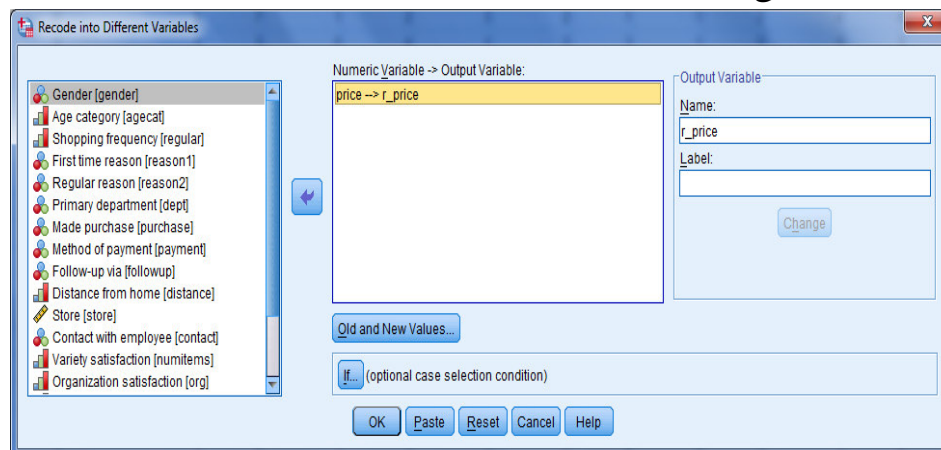
حال با استفاده از دستور Frequency برای متغیر r\_O.C جدول فراوانی رسم می کنیم.

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	

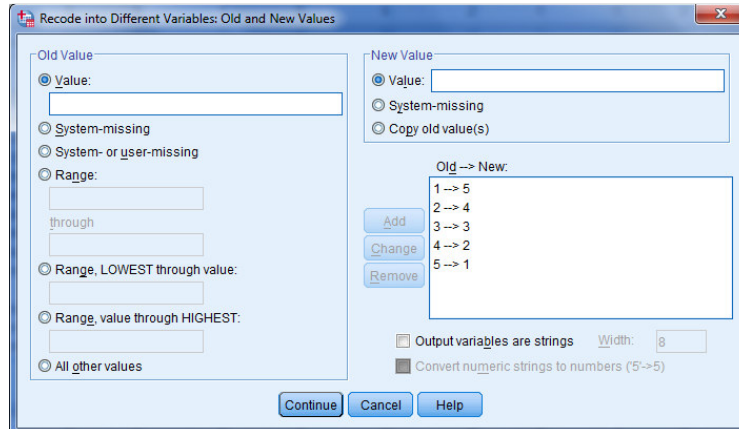
### ۳-۶-۲ معکوس کردن امتیازات (دستور Recode):

همان طور که از نام این دستور پیداست، توسط آن می توان داده ها را دوباره کدبندی کرد. برای مثال فرض کنید در پرسش نامه ای ۲۰ سوالی، ۱۸ سوالات با نمره گذاری از ۱ تا ۵ (مستقیم) معنی می دهند، اما در ۲ سوال جهت نمره گذاری ۵ تا ۱ (معکوس) است. در این حالت بهتر است همه داده ها وارد شوند، اما جهت نمره گذاری آن دو سوال را با استفاده از دستور Recode برعکس کرد.

برای بررسی این دستور ابتدا فایل داده "satisfy.sav" را از فایل های نمونه SPSS باز کنید. می خواهیم کدبندی متغیر price را معکوس کرده و آن را 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 می کنیم. در قسمت سمت راست این کادر (Output Variable) و در قسمت Name، نام متغیر جدید، یعنی r\_price، را تایپ می کنیم و سپس دکمه Old and New values را کلیک می کنیم تا کادر Old and New values: باز شود. در این کادر در قسمت Old values عدد ۱ و در قسمت New values عدد ۵ را تایپ می کنیم. و همین کار را برای کدهای ۲، ۳، ۴ و ۵ انجام می دهیم و آنها تبدیل به کدهای، به ترتیب، ۴، ۳، ۲ و ۱ می کنیم. سپس دکمه های Continue، change، و Ok را به ترتیب کلیک می کنیم. حال به پنجره داده ها 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 کدبندی شده در همان متغیری اول ذخیره می‌شوند. لذا این امر باعث از بین رفتن اطلاعات اصلی می‌شود و استفاده از دستور Recode بدین گونه 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.

## ۲-۴ 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  $X \sim N(\mu, 16)$  است. آزمایش گذشته نشان می دهد که اگر یک ماده معدنی به این محصول اضافه نشده باشد  $\mu = 10$  و در غیر اینصورت  $\mu = 11$  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:

$$\begin{cases} \mu = 10 & \text{میانگین جامعه برابر عدد ۱۰ است} \\ \mu = 11 & \text{میانگین جامعه برابر عدد ۱۱ است} \end{cases}$$

در اینجا عدد نامعلوم، برابر بودن  $\mu = 11$  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.

$$\begin{cases} H_0: \mu = 10 \\ H_1: \mu = 11 \end{cases}$$

## ۳-۴ سطح معنی داری و خطاهای آماری

بعد از تعریف فرضیه های آماری، قدم بعدی مشخص کردن درجه ای برای معنی دار بودن تفاوت ها ( $\alpha$ ) 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  $\alpha$  نشان داده می شود. این احتمال وقوع کوچک را «سطح معنی داری» می گویند. مقادیری که معمولاً برای  $\alpha$  استفاده می شود بیشتر از ۰.۰۱ و ۰.۰۵ است. از آنجا که مقدار  $\alpha$  در تعیین اینکه  $H_0$  باید رد شود یا نه دخالت مستقیم دارد، الزام رعایت عینیت در تحقیق ایجاب می کند که  $\alpha$  را پیش از شروع جمع آوری داده ها مشخص کنیم. سطح معنی داری که محقق برای تعیین  $\alpha$  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 مثلاً درباره آثار درمانی عمل جراحی روی مغز باشد، محقق باید  $\alpha$  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  $\alpha$  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 معمولاً با  $\beta$  نشان داده می شود.  $\alpha$  و  $\beta$  هم برای نشان دادن نوع خطاها و هم ارتکاب آن خطاها به کار می روند، یعنی:

(رد کردن  $H_0$  وقتی  $H_0$  درست است)  $= P$  (خطای نوع اول)  $= \alpha$

(رد کردن  $H_0$  وقتی  $H_1$  درست است)  $= P$  (خطای نوع اول)  $= \beta$

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

واضح است که بین  $\alpha$  و  $\beta$  رابطه معکوس وجود دارد. بالا رفتن  $\alpha$ ، و مقدار  $\beta$  کاهش می یابد و برعکس. این رابطه در آمار به «بده - بستان» بین  $\alpha$  و  $\beta$  معروف است. آنچه مسلم است، مجموع  $\alpha$  و  $\beta$  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:

$$1 - \beta = 1 - (\text{احتمال وقوع خطای نوع دوم}) = 1 - \beta$$

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 ( $\beta$ ) 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 چگونگی 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 متأثر 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 برگرفته سطح معنی داری  $\alpha$  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  $\alpha$  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:

$$\begin{cases} H_0: \mu \geq 200 \\ H_1: \mu < 200 \end{cases}$$

واضح است که  $H_1$  به اندازه  $\alpha$  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:

$$\begin{cases} H_0: \mu \leq 200 \\ H_1: \mu > 200 \end{cases}$$

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:

$$\begin{cases} H_0: \mu \neq 200 \\ H_1: \mu \neq 200 \end{cases}$$

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 کند، توزیع نمونه گیری آماره و آماره آزمون تغییر خواهد کرد.

#### III- تعیین سطح زیر منحنی $H_0$ و $H_1$ و محاسبه مقدار بحرانی (مقدار بحرانی):

سطح زیر منحنی  $H_0$  و  $H_1$  به توزیع نمونه گیری و مقدار  $\alpha$  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  $\alpha$  خواهد بود. محاسبه مقدار استاندارد که تفکیک کننده  $H_0$  و  $H_1$  بصورت عددی باشد از دیگر موارد ضروری این مرحله است. مقدار استاندارد بر اساس نوع آزمون و مقدار  $\alpha$  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 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  $H_1$  آيا در سطح خطای  $\alpha$  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 (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  $\mu$  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  $\mu$ ، میزان سازگاری بین برآمد نمونه ای و مقدار  $\mu_0$  را که در  $H_0$  مسلم فرض شده است اندازه می گیرد. یک P-Value بزرگ نشان می دهد که  $\mu_0$  موجه است، و بنابراین، باید  $H_0$  نتیجه گرفته شود. در واقع برحسب اینکه P-Value بزرگتر یا کوچکتر از  $\alpha$  (ناحیه بحرانی) باشد که برای اجرای آزمون مشخص شده است، می تواند مستقیماً 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  $P\text{-Value} \leq \alpha$  باشد، گزینه  $H_0$  را نتیجه بگیرید.

اگر  $P\text{-Value} \geq \alpha$  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.

#### ۸-۴ آزمون آماری برای میانگین جامعه - آزمون t تک نمونه‌ای

وقتی می‌خواهیم برای میانگین جامعه ( $\mu$ ) 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 تک نمونه‌ای می‌باشد.

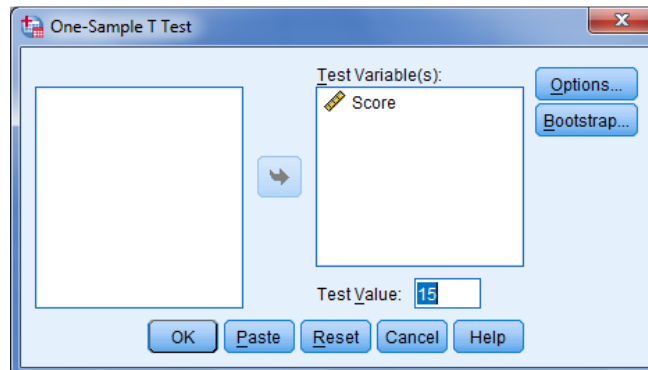
**مثال ۱)** داده‌های زیر نمره‌های ۲۰ دانش آموز در درس آمار است:

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

آیا می‌توان گفت میانگین نمره‌های ریاضی دانش آموزان 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  $\text{Analyze} > \text{Compare Means} > \text{One-Sample T Test}$  را طی کنید، تا کادر مکالمه‌ای One-Sample T 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					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Score	-.079	19	.938	-.05000	-1.3735	1.2735

**معرفی جداول:** جدول اول (One-Sample Statistics)، حجم نمونه، شاخص های آماری میانگین، انحراف معیار و خطای معیار میانگین را نشان می دهد. جدول دوم (One-Sample Test) نتایج آزمون t تک نمونه ای را نشان می دهد. مقدار آماره t محاسبه شده برابر ۰/۰۷۹- شده است که می دانیم آماره مربوط به آن به صورت 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 معنی داری" است، نشان دهنده P-مقدار آزمون است، که در این آزمون برابر ۰/۹۴ شده است. همچنین مقدار Mean Difference نشان دهنده  $\bar{x} - \mu_0$  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 می دهد، که در این مثال برابر (۱/۲۷، -۱/۳۷) است. حال اگر بخواهیم فاصله اطمینان برای  $\mu$  بدست آوریم، باید نامعادله  $1.27 < \bar{x} - \mu < -1.37$  را برای مجهول  $\mu$  حل کنیم. لذا فاصله اطمینان برای  $\mu$  برابر است با (۱۳/۶۸، ۱۶/۳۲).

**تفسیر:** همانطور که می دانیم، از سه روش معادل می توان آزمون فرض را انجام داد که عبارتند از: ۱- ناحیه رد، ۲- P-مقدار، ۳- فاصله اطمینان. که در اکثر خروجی های SPSS نتایج بر اساس این سه روش ارائه می شود. ما نیز با استفاده از خروجی بالا از هر سه راه این فرضیه را در سطح معنی داری ۰/۰۵ آزمون می کنیم.

۱- **روش ناحیه رد:** با توجه به قاعده تصمیم گیری  $|T| > t_{\frac{\alpha}{2}, (n-1)}$  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 معنی داری ۰/۰۵ فرض

$H_0$ ، یعنی فرض برابری، را رد نمی کنیم.

۲- **روش P-مقدار:** همان طور که می دانیم قاعده در این روش به این صورت است که، "اگر P-مقدار

کمتر از  $\alpha$  باشد، فرض  $H_0$  را در سطح معنی داری  $\alpha$  رد می کنیم". در این مثال P-مقدار برابر

۰/۹۴ است و کمتر از ۰/۰۵ نیست، لذا در سطح معنی داری ۰/۰۵ فرض  $H_0$ ، یعنی فرض برابری، را

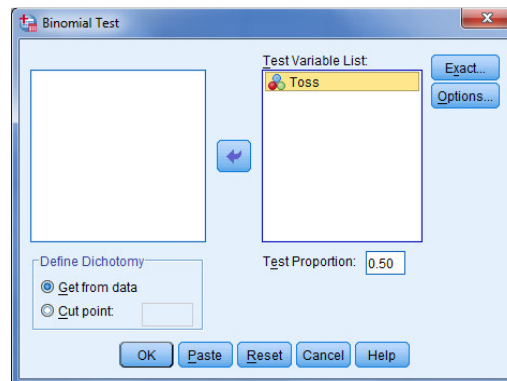
رد نمی کنیم.

WWW.GhadamYar.com

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

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 Dialogs>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	12	.40	.50	.362	
	Group 2	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 داشته باشید 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) در نظر می‌گیرد. در ستون سوم، N، تعداد افراد مشاهده شده 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 ۰/۵ است. در ستون آخر، Exact Sig. (2-tailed)، P-مقدار دقیق را در حالت دوطرفه نشان می‌دهد.

**تفسیر:** با توجه به اینکه P-مقدار دقیق دوطرفه بیشتر از ۰/۵ 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 اطمینان ۹۵ درصد آزمون کنید که:

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

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

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

الف) تمام مراحل مربوط به مثال ۱ را در این جا 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 به صورت زیر خواهیم داشت.

Binomial Test						
	Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)	
X	Group 1	1.00	100	.01	.02	.000 <sup>a</sup>
	Group 2	2.00	9900	.99		
	Total		10000	1.00		

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

ب) به علت وجود تقارن توزیع نرمال و این که در این حالت SPSS آزمون دوطرفه را انجام نمی‌دهد، برای انجام آزمون دوطرفه  $P$ -مقدار آزمون یک طرفه را دو برابر 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 شده بود. 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 مونت کارلو استفاده کنیم. برای این کار 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 به کار 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 واریانس های دو گروه خواهیم داشت. لذا ابتدا بایستی با استفاده از آزمون Levene برابری واریانس های دو گروه را مورد بررسی قرار داد. دو حالت می تواند پیش آید که هر دو حالت در SPSS به طور همزمان اجرا می شود.

۱- **برابری واریانس ها:** چنانچه P-مقدار حاصل از آزمون Levene بیشتر از  $\alpha$  باشد می توان نتیجه

گرفت که واریانس های دو گروه با هم برابر است. در این حالت برای بررسی معنی داری اختلاف بین میانگین های دو گروه، از آزمون "t-دو نمونه مستقل در حالت برابری واریانس ها" استفاده می کنیم.

۲- **نابرابری واریانس ها:** چنانچه P-مقدار حاصل از آزمون Levene کمتر یا مساوی  $\alpha$  باشد می توان

نتیجه گرفت که واریانس های دو گروه با هم برابر نیستند. در این حالت برای بررسی معنی داری

اختلاف بین میانگین های دو گروه از آزمون "t-دو نمونه مستقل در حالت نابرابری واریانس ها"

استفاده می کنیم.

**مثال ۵)** ده دانش آموز پسر و ده دانش آموز دختر به دلخواه انتخاب شده اند و وزن آنها اندازه گیری

شده است. داده ها به صورت زیرند:

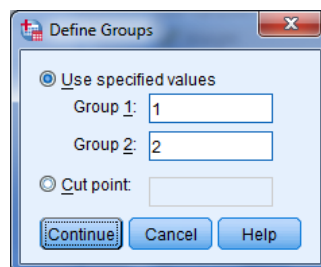
دختر	۵۴	۴۸	۶۵	۶۰	۴۵	۵۷	۴۵	۶۰	۶۳	۵۵
پسر	۷۵	۶۸	۸۰	۷۴	۷۲	۶۸	۷۰	۸۵	۹۰	۷۵

مایلم بدانیم آیا وزن دانش آموزان دختر و پسر از لحاظ آماری با هم برابر است؟

ابتدا داده ها را به صورت زیر در SPSS وارد کنید. مردان با کد ۱ و زنان با کد ۲ مشخص شده اند.

	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 این جا ۱ و ۲) و دکمه Continue و سپس Ok را کلیک کنید تا خروجی به صورت زیر ظاهر شود:

1

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	

2

Independent Samples Test									
Levene's Test for Equality of Variances			t-test for Equality of Means						
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
							Lower	Upper	
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

3

**معرفی جدول:** این جدول شامل ۳ قسمت است:

- ۱- **Group Statistics** : شامل تعداد نمونه و شاخص های آماری به تفکیک هر یک از گروه ها است.
  - ۲- **Levene's Test** : شامل آزمون Levene برای آزمون برابری واریانس ها (آماره F و P-مقدار)
  - ۳- **t-test for Equality of Means** : نتایج آزمون t-دو نمونه ای مستقل است که ستون های آن به ترتیب از چپ به راست عبارتند از: مقدار آماره t، درجه آزادی (df)، P-مقدار آزمون دوطرفه (sig(two-tailed))، اختلاف میانگین ها، یعنی  $\bar{Y} - \bar{X}$  (Difference Mean)، خطای استاندارد اختلاف میانگین ها (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 شده اند. مقادیر خط اول برای حالت برابری واریانس ها و مقادیر خط دوم برای حالت نابرابری واریانس ها هستند.
- تفسیر:** P-مقدار آزمون Levene نشان می دهد که بین واریانس وزن دانش آموزان پسر و دختر اختلاف معنی داری وجود ندارد. لذا باید نتایج آزمون t را از خط دوم قسمت مربوطه گزارش کرد. در این جا نیز می توانیم از سه روش ناحیه رد، P-مقدار، فاصله اطمینان آزمون فرض را انجام دهیم.
- ۱- **روش ناحیه رد:** با توجه به قاعده تصمیم گیری  $|T| \geq t_{\frac{\alpha}{2}, (n_1+n_2-2)}$  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 معنی داری ۰/۰۵ فرض  $H_0$ ، یعنی فرض برابری، را قبول نمی کنیم و معنی داری اختلاف را می پذیریم.

- ۲- **روش P-مقدار:** در این مثال P-مقدار ۰/۰۰۰ گزارش شده است و لذا کمتر از ۰/۰۵ است، بنابراین در سطح معنی داری ۰/۰۵ فرض  $H_0$ ، یعنی فرض برابری، را قبول نمی کنیم و معنی داری اختلاف را می پذیریم.

- ۳- **روش فاصله اطمینان:** در این جا فرضیه  $H_0: \mu_1 - \mu_2 = 0$ ، بررسی می کنیم. بنابراین برای آزمون فرضیه فوق به روش فاصله اطمینان کافیت که چک کنیم آیا فاصله اطمینان  $\mu_1 - \mu_2$  عدد صفر را شامل می شود؟ با توجه به فاصله اطمینان ۹۵٪ در این مثال، در سطح معنی داری ۰/۰۵ فرض  $H_0$ ، یعنی فرض برابری، را قبول نمی کنیم و معنی داری اختلاف را می پذیریم.

توجه:

- ۱- فرض صفر مطرح شده در این آزمون در واقع همان فرض صفر  $\mu_1 - \mu_2 = k$  است که در آن  $k=0$  است.
- ۲- در این آزمون نیز فرض بر این است داده ها، در هریک از گروه ها، از جامعه ای با توزیع نرمال هستند.
- ۳- در آزمون Levene فرض های زیر آزمون می شوند و آماره 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- همان طور که می بینید، این آزمون دوطرفه است. در خروجی این آزمون نیز می توان با نصف کردن P-مقدار، آزمون های یک طرفه را انجام داد.

#### ۴-۱۱ آزمون اختلاف میانگین ها برای دو جامعه وابسته - آزمون t زوجی

همان طور که می دانید، وقتی می خواهیم میانگین یک متغیر کمی را در دو گروه وابسته، مقایسه کنیم، از آزمون t-زوجی استفاده خواهیم کرد. در این حالت فرض های آماری عبارتند از:

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

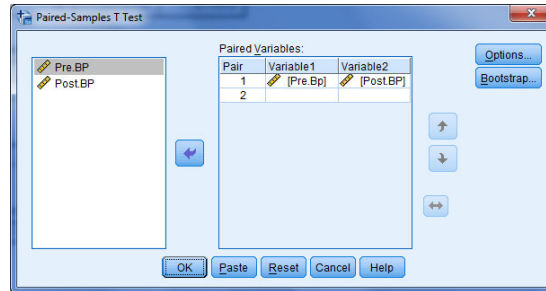
و از آنجایی که در عمل همواره واریانس جامعه نامعلوم است، آزمون مناسب، آزمون t-زوجی می باشد. **مثال ۶)** مثال ۶ (فشار خون ده نفر) از فصل هشتم کتاب را در نظر بگیرید. در این می خواهیم بدانیم آیا رژیم غذایی جدید بر فشارخون بیماران تاثیر داشته است یا خیر؟

ابتدا داده ها را به صورت 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 مکالمه ای Paired-Samples T Test باز شود. در این کادر متغیرهای "Pre.BP" و "Post.BP" را به ترتیب در جعبه Paired Variables، و در قسمت های Variable1 و Variable2 وارد کنید.





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

۱

	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

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

۳

		Paired Differences			95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	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 شاخص های 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 ستون های آن به ترتیب از چپ به راست

عبارتند از: میانگین اختلاف ها یا همان  $\bar{D}$  (Mean) است، انحراف معیار اختلاف ها یا  $\sqrt{S_D^2}$

(Std.Deviation) است، خطای استاندارد میانگین اختلاف هاست (Std. Error Mean)، فاصله

اطمینان ۹۵٪ برای  $\mu_D$  (95% Confidence Interval of the Difference) است، مقدار آماره  $t$ ،

درجه آزادی (df)،  $P$ -مقدار آزمون دوطرفه (sig(two-tailed)).

**تفسیر:** در این جا نیز می توانیم از سه روش آزمون فرض را انجام دهیم.

۱- **روش ناحیه رد:** با توجه به قاعده تصمیم گیری  $|T| \geq t_{\frac{\alpha}{2}, (n-1)}$  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

نمی کنیم، به عبارتی می توانیم 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 نمی کنیم، به عبارتی می توانیم با احتمال ۹۵ درصد قضاوت کنیم که رژیم غذایی جدید بر فشار خون تاثیر ندارد.

۳- **روش فاصله اطمینان:** در این جا فرضیه  $H_0: \mu_1 - \mu_2 = 0$  را بررسی می کنیم. بنابراین برای آزمون فرضیه فوق به روش فاصله اطمینان کافیت که چک کنیم آیا فاصله اطمینان  $\mu_1 - \mu_2$  عدد صفر را شامل می شود؟ با توجه به فاصله اطمینان ۹۵٪ در این مثال، پس فرض  $H_0$ ، یعنی فرض برابری، را رد نمی کنیم، به عبارتی می توانیم 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 دو جامعه به واسطه "جورشدن" نمونه ها نیز وابسته باشند.
- ۲- در این آزمون نیز فرض بر این است که داده ها از جامعه ای با توزیع نرمال هستند.
- ۳- آزمون معادل آزمون t-تک نمونه ای روی جامعه اختلاف ها با  $\mu_0 = 0$  است.
- ۴- همان طور که می بینید، این آزمون در حالت دوطرفه اجرا شده است. در خروجی این آزمون نیز می توان با نصف کردن P-مقدار، آزمون های یک طرفه را انجام داد.

## تمرین:

۱- اداره بهداشت یک شهر می خواهد تعیین کند که آیا میانگین تعداد باکتری ها 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 گردآوری کرده و دیده اند که تعداد باکتری ها عبارتند از:

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

آیا داده ها دلیلی بر نگرانی به دست می دهند؟

۲- از ۲۰ راننده تاکسی در یک شهر در مورد درآمد روزانه آن ها سؤال شده است. داده های حاصل به صورت زیرند (بر حسب هزار تومان):

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

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

۳- دو نوع اسباب بازی مکانیکی و غیر مکانیکی داریم. می خواهیم ببینیم که آیا در کودکان ۵ ساله تمایل به سمت هر کدام از این اسباب بازی ها یکسان است؟ بدین منظور از ۲۰۰ کودک ۵ ساله خواسته ایم که به دلخواه یکی از این اسباب بازی ها را انتخاب کنند و ۱۲۰ کودک اسباب بازی مکانیکی و ۸۰ کودک اسباب بازی غیر مکانیکی را انتخاب کردند. از این داده ها چه استنباطی دارید؟

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

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

آیا می توان گفت که:

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

ب - ۴۰ درصد از بازیکنان فوتبال این شهر بیشتر از 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:

۲۴	۲۷	۱۶	۱۸	۲۱	۱۶	۲۳	۱۱	۲۰	۱۵	روش ۱
۲۸	۲۵	۲۶	۲۸	۱۷	۲۳	۱۹	۱۳	۳۱	۲۳	روش ۲

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  $\alpha = 0.05$  آزمون کنید)  
 ب) فرضیهایی 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 (SBP = steam blanching process) 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 استفاده از دو روش تمیز کردن یک فاصله اطمینان ۹۸٪ بسازید.

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

۸- ادعا شده است که یک برنامه ایمنی صنعتی در کاهش تضييع 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 دانش آموز طی یک ماه، می خواهیم بدانیم آیا این نوع ورزش تفاوتی معنی دار در وزن دانش آموزان یک مدرسه ایجاد کرده است یا نه؟ داده های زیر وزن دانش آموزان را در قبل از انجام دوره و بعد از آن نشان می دهد.

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

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 محقق می‌خواهد 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 می‌کنند فشار خون آنها دوباره ثبت می‌شود. محقق مایل است درباره اثر قرص روی فشار خون به کمک مشاهدات داده شده در جدول زیر استنباطهایی انجام دهد.

جدول اندازه‌های 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 می‌شوند. از بین آنها به طور تصادفی ۱۰ نفر را برای آموزش به وسیله روش ۱ انتخاب می‌کنند، ۱۰ نفر بقیه را با روش ۲ آموزش می‌دهند. بعد از تکمیل دوره آموزشی، همه کارگران در معرض یک آزمون زمان و حرکت قرار می‌گیرند که سرعت انجام یک کار تخصصی را ثبت می‌کند. داده‌های زیر به دست آمده‌اند:

زمان به دقیقه

روش ۱	۱۵	۲۰	۱۱	۲۳	۱۶	۲۱	۱۸	۱۶	۲۷	۲۴
روش ۲	۲۳	۳۱	۱۳	۱۹	۲۳	۱۷	۲۸	۲۶	۲۵	۲۸

الف) آیا از این داده‌ها می‌توان نتیجه گرفت که میانگین زمان کار بعد از آموزش با روش ۱ به طور معنی‌دار از میانگین زمان کار بعد از آموزش با روش ۲ کمتر است؟ (با  $\alpha = 0.05$  آزمون کنید)  
 ب) فرضیهایی را که برای توزیع‌های جامعه در نظر می‌گیرید بیان کنید.  
 ج) برای تفاضل میانگین جامعه‌های زمانهای کار بین دو روش، یک فاصله اطمینان ۹۵٪ بنا کنید.

۳- گمان می‌رود روشی جدید در مقایسه با روش متعارفی که در حال حاضر برای آموزش خواندن به نوآموزان 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) فرضیهایی را که برای تجزیه و تحلیل خود در نظر گرفته اید بیان کنید.

۴-جامعه شناسی مایل به مقایه نرخ باروری زنهای 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 کنند. 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  $\alpha=0.05$  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 (SBP = steam blanching process) 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 لوبیاها 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 استفاده 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 کشف کافی برای بافتن دو رشته طناب است. اندازه های قدرت کشف 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 کشف 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 تضييع 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 تسريع 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

دهند. رشد گیاهان بعد از چهار هفته بر حسب سانتیمتر اندازه گیری 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؟ فرضیهایی 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 تامنسن به عنوان یک نمونه به حجم ۱۴، برای میانگین نسبت جرم به بار، یک فاصله اطمینان ۹۹٪ بنا کنید.