

- ابزارهای آمار

ابزارهای آمار در SPSS دو گونه هستند: الف- ابزارهای آمار توصیفی، ب- ابزارهای آمار استنباطی.

ابزارهای آمار توصیفی، تصویری از وضعیت موجود را نشان می دهند و هیچ نقشی در پیش بینی و تخمین ندارند. بعنوان مثال میانگین قیمت دلار در ۲۰ روز گذشته یا متوسط نمرات دانشجویان کلاس ۲۵ نفره در درس روش تحقیق.

- ابزارهای آمار توصیفی

ابزارهای آمار توصیفی به دو صورت هستند.

الف- عددی (شاخص های توصیف)، ب- گرافیکی (رسم نمودارها).

در حالیکه ابزارهای آمار استنباطی، ابزارهایی هستند که بوسیله آنها تخمین زده و پیش بینی های بیشتری انجام می شود.

- ابزارهای آمار توصیفی

- عددی (شاخص های توصیف)

۱- توزیع فراوانی

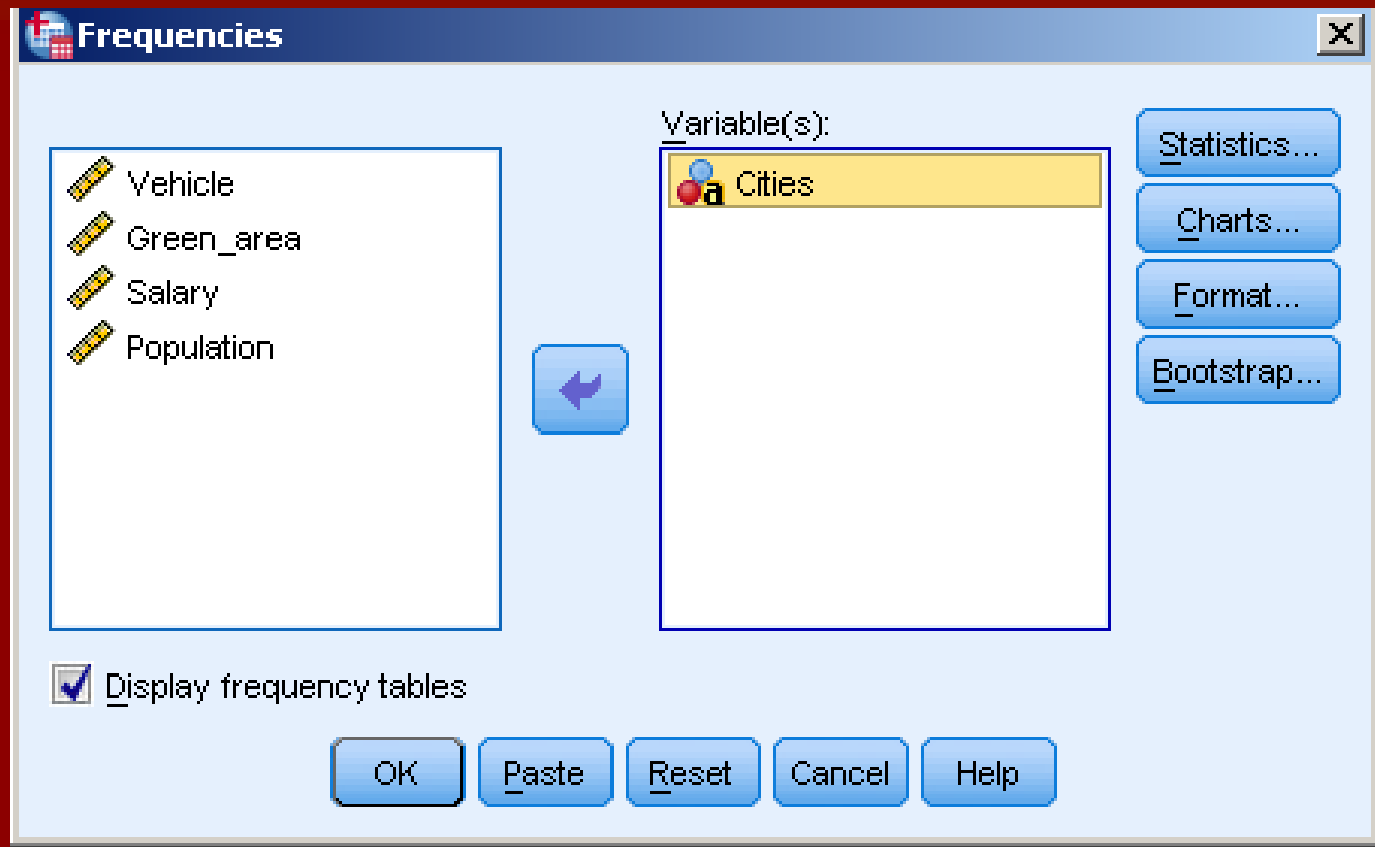
توزیع فراوانی روشی مناسب جهت توصیف داده ها بشمار می آید. توزیع فراوانی برای متغیرهای طبقه ای و توزیع فراوانی برای متغیرهای کمی.

-توزیع فراوانی برای متغیرهای طبقه ای

برای تعیین توزیع فراوانی به روش زیر عمل می کنیم:

Analyze → Descriptive Statistics → Frequencies

متغیر Cities را با کلیک روی دکمه → به کادر Variable(s) منتقل کنید. روی Ok کلیک نمایید.



نتیجه این دستور شامل دو خروجی است. خروجی اول، به ترتیب تعداد داده های معتبر و از دست رفته را نشان می دهد. و خروجی دوم ستون Frequency فراوانی هر طبقه، ستون Percent درصد داده های هر طبقه، ستون Valid Percent درصد داده های معتبر و ستون Cumulative Percent درصد تجمعی را نشان می دهد. تفاوت ستون Percent و ستون Valid Percent آن است که در ستون Valid Percent درصد داده ها بدون احتساب داده های از دست رفته (Missing) را محاسبه می کند.

- توزیع فراوانی برای متغیرهای کمی

برای متغیرهای کمی با سطوح محدود، همانند (جمعیت، حقوق و فضای سبز در مثال قبلی)، همانند روش توضیح داده شده برای متغیرهای طبقه ای عمل می شود.

ولی برای متغیرهای کمی با سطوح زیاد، باید داده ها را دسته بندی کنیم. بدین منظور از دستور Recode استفاده می شود و متغیر متفاوتی تعریف می شود. سپس همانند روش قبلی، فراوانی متغیر جدید تعریف شده بدست می آید.

۲- شاخص های توصیف داده ها

شاخص های توصیف داده ها به سه گروه شاخص های مرکزی (همانند میانگین، میانه و نما) شاخص های پراکندگی (همانند واریانس و انحراف معیار) و شاخص های شکل توزیع (شاخص های چولگی و شاخص های کشیدگی)

۲-۱- شاخص های مرکزی

هر معیار عددی که معرف مرکز داده ها باشد، شاخص مرکزی نامیده می شود.

■ ۲-۱-۱- میانگین (Mean)

- اصلی ترین و مورد استفاده ترین شاخص مرکزی میانگین است. اگر داده ها روی یک محور به صورت منظم ردیف شوند، مقدار میانگین دقیقا در نقطه تعادل یا مرکز ثقل توزیع قرار می گیرد.
- میانگین حسابی، معدل مجموعه ای از مشاهدات بشمار می آید و به کمک رابطه ریاضی محاسبه می شود.

$$\bar{X} = \frac{\sum x_i}{n}$$

■ متداولترین معیار مرکزی که عموماً برای داده های حاصل از مقیاس های فاصله ای و نسبتی بکار برده می شوند، انواع میانگین ها هستند که بسته به داده های مورد مطالعه، نوع مناسب آنها بکار گرفته می شوند. بنابراین اگر اطلاعات بصورت کمی باشند، میانگین یکی از دقیق ترین شاخصهای آماری می باشد. ضمن اینکه در متغیرهای کیفی، شاخص میانگین نداریم.

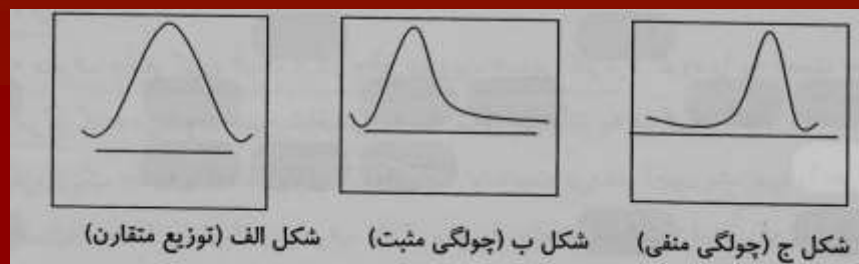
۲-۱-۲- نما (Mode)

مد در آمار مقداری است که بیشترین تکرار را در میان مشاهدات داشته باشد. از این شاخص برای نمایش داده‌هایی استفاده می‌شود که نسبت به سایر مقادیر تکرار بیشتری دارند.

چون نما همیشه در مرکز توزیع فراوانی قرار ندارد، یک شاخص بی ثبات گرایش مرکزی است و کاربرد کمتری دارد. این شاخص به تکرار تاکید دارد. مناسب ترین استفاده آن زمانی است که مقیاس اندازه گیری، طبقه ای باشد. بعنوان مثال در مواردی که برای یک پزشک میانگین سن افرادی که آپاندیس خود را عمل کرده اند، اهمیت آنچنانی نداشته باشد و در عوض تعداد این افراد برایش مهم باشد. در مثال دیگر، اهمیت دارد که ببینیم بیشترین مصرف اراضی مربوط به چه نوع کاربردی است.

چنانچه فراوانی رده هایی که بیشترین فراوانی را دارند، با هم تقریبا یکسان باشند، نما یک شاخص ضعیف برای توزیع بشمار می آید. همچنین اگر همه داده ها در یک توزیع، دارای فراوانی یکسانی باشند، توزیع فاقد نما است. تنها شاخص مرکزی که ممکن است بیش از یکی باشد، نما است. زمانی که توزیع فقط یک نما داشته باشد، در منحنی یک نقطه یا برآمدگی وجود دارد که دارای حداکثر ارتفاع است و آن را تک نمایی گویند. در صورتیکه دارای دو برآمدگی باشد، آن را دو نمایی می گویند، و زمانی که دارای بیش از دو برآمدگی باشد، آنرا چند نمایی گویند.

توزیع های داده ها بر حسب فراوانی می تواند به سه صورت شکل الف (توزیع متقارن)، شکل ب (چولگی مثبت) و شکل ج (چولگی منفی) انجام شود.



همچنین انواع توزیع ها داده ها براساس تعداد نما، در شکل زیر نشان داده شده است.



■ ۲-۱-۳- چندک ها

چندک ها مقادیری از مشاهدات هستند که دامنه تغییرات را به فاصله های چندکی مورد نیاز تقسیم می کنند. بنابراین اگر دامنه تغییرات را به چهار قسمت تقسیم کنیم به چارک ها (Quartile)، اگر به ده قسمت تقسیم کنیم به دهک ها (Decile) و اگر به صد قسمت تقسیم کنیم به صدک ها (Percentile) می رسیم.

۲-۱-۴- میانه (Median)

میانه مقداری است، که در صورت مرتب کردن داده ها، در وسط قرار می گیرد.

میانه مقداری است، که در صورت مرتب کردن داده ها، در وسط قرار می گیرد. میانه مقداری است که ۵۰ درصد داده های جامعه پایین تر از آن و ۵۰ درصد بالاتر از آن قرار می گیرند.

مثال- فرض کنید پول های موجودی ۵ نفر مبالغ زیر بر حسب تومان باشد، شاخص مرکزی میانه را برای هر سری از داده های زیر بدست آورید.

$$۱,۰۰۰ - ۲,۰۰۰ - \underline{۳,۰۰۰} - ۱۰۵,۰۰۰ - ۲۵۰,۰۰۰$$

$$\text{میانه } Md = ۳,۰۰۰$$

$$۱,۰۰۰ - ۲,۰۰۰ - \underline{۳,۰۰۰ - ۱۰۵,۰۰۰} - ۲۵۰,۰۰۰ - ۵۰۰,۰۰۰$$

$$\text{میانه } Md = ۵۴,۰۰۰$$

در محاسبه میانه، مقدار بزرگترین عددی که در بالاترین وضعیت قرار دارد، در میزان میانه تاثیری ندارد. بدین ترتیب در مواقعی که توزیع دارای چولگی قابل ملاحظه باشد، از میانه استفاده می کنیم. بنابراین یکی از مزایای میانه، رفع عیب مربوط به شاخص میانگین می باشد. میانگین زمانی که توزیع نرمال باشد، مورد استفاده قرار می گیرد و بنابراین در مطالعات امروزی، شاخص مرکزی میانه جایگاه مناسبی را بدست آورده است. همچنین در مواقعی که وقت کافی برای محاسبه میانگین نداشته باشیم یا قسمتی از یک توزیع داده ها در دسترس ما باشد، میانه را محاسبه می کنیم.

۲-۲- شاخص های پراکندگی

برای اینکه داده ها بطور واقعی تری توصیف شوند و مجموعه مشاهدات با هم قابل مقایسه باشند، باید میزان تفاوت های آنها نیز از طریق معیارهای عددی سنجیده شود. برای این منظور نیاز به آشنایی با شاخص های پراکندگی می باشد.

۲-۲-۱- دامنه تغییرات (Range)

یکی از ساده ترین شاخص های پراکندگی، دامنه تغییرات است. این شاخص با تفاضل کوچکترین مشاهده از بزرگترین مشاهده محاسبه می شود.

تعریف ریاضی دامنه تغییرات:

$$R = \text{Max} (x_1) - \text{Min} (x_2)$$

۲-۲-۲- واریانس (Variance)

یکی دیگر از شاخص های اندازه گیری پراکندگی داده ها نسبت به میانگین، واریانس است. این شاخص به بررسی داده ها نسبت به میانگین می پردازد.

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

۲-۲-۳- انحراف معیار (Standard Deviation)

اندازه گیری واریانس، مجذور واحد اصلی متغیر است و به همین دلیل تفسیر آن به سادگی صورت نمی گیرد. این مسئله با جذر گرفتن از واریانس حل می شود. شاخص حاصل از جذر واریانس را انحراف معیار می نامند. بدین ترتیب معیاری از پراکندگی بدست می آید که با همان واحد مشاهدات اصلی بیان می شود.

$$s = \sqrt{s^2}$$

از انحراف استاندارد هنگامی استفاده می شود که مقیاس، حداقل فاصله ای باشد و بررسی های آماری لازم باشد. داده های خیلی بزرگ یا کوچک در معیار پراکندگی انحراف معیار تاثیر گذار هستند و زمانی از انحراف معیار استفاده می شود که میانگین بعنوان شاخص مرکزی مورد استفاده قرار گرفته باشد.

بنابراین در حالیکه، دامنه تغییر، فقط به دو نمره انتهایی توجه دارد، انحراف معیار تحت تاثیر تمام داده های داخل توزیع قرار می گیرد و مشخص می کند که هر داده بطور متوسط، چه مقدار از میانگین فاصله دارد. انحراف معیار همانند واریانس با ثبات است و ثبات آن از دامنه تغییر بیشتر است و برخلاف واریانس، می تواند مستقیما بر حسب واحدهای داده تفسیر شود.

■ ۲-۲-۴- خطای استاندارد میانگین (Standard Error of Mean)

■ شاخص خطای معیار میانگین، انحراف معیار میانگین نمونه است.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

■ وقتی که فعالیت آماری روی نمونه انجام می شود. نمی توان در تحلیل ها تمام جامعه را پوشش داد، بنابراین در بررسی نمونه، همیشه خطای استاندارد میانگین (Std Error of Mean) را داریم.

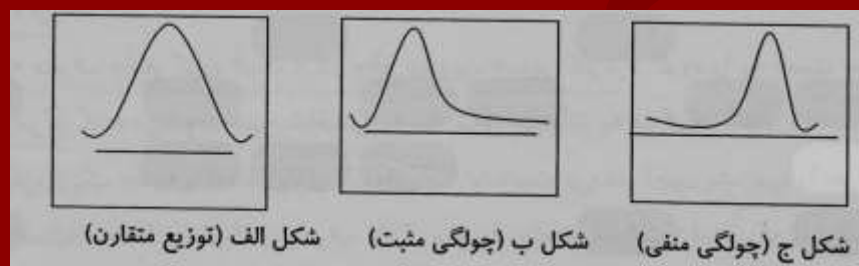
۲-۳- شاخص های شکل توزیع

این شاخص ها به بررسی شکل توزیع جوامع آماری می پردازد و شامل چولگی (Skewness) و کشیدگی (Kurtosis) می باشد.

۲-۳-۱- چولگی (انحراف از قرینگی)

چولگی، عدم تقارن توزیع نسبت به شاخص معین (معمولا نسبت به میانگین) را نشان می دهد. توزیع داده ها زمانی چوله به راست است یعنی دنباله توزیع، سمت راست آن قرار گیرد و زمانی چوله به چپ است که دنباله جامعه سمت چپ آن واقع شود.

■ مهمترین شاخص چولگی، ضریب چولگی است. اگر ضریب چولگی منفی باشد، توزیع دارای چوله به سمت چپ است و در صورتیکه ضریب چولگی مثبت باشد، توزیع دارای چوله به سمت راست است. اگر توزیع متقارن باشد، ضریب چولگی مساوی صفر خواهد بود. بدیهی است هر چقدر ضریب چولگی بزرگ تر باشد، تفاوت جامعه از نظر قرینگی با توزیع متقارن بیشتر است.



■ اگر قدر مطلق ضریب چولگی از ۰/۵ بزرگتر باشد، چولگی توزیع زیاد است. ضریب چولگی (SK) بر پایه رابطه زیر محاسبه می شود.

$$SK = \frac{r_3}{s_x^3}, \quad r_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$

■ خطای استاندارد ضریب چولگی

(Standard Error of Skewness)، می تواند بعنوان آزمون نرمال بودن استفاده شود (اگر کوچک تر از ۲- یا بزرگتر از ۲+ باشد، نرمال بودن رد می شود).

■ ۲-۳-۲- کشیدگی (Kurtosis)

شاخص سنجش پراکندگی جامعه نسبت به توزیع نرمال، ضریب کشیدگی نام دارد. توزیع ها با توجه به مقدار ضریب کشیدگی به سه گروه تقسیم می شوند:

الف- آن دسته از توزیع ها که نسبت به توزیع نرمال از پراکندگی بیشتری برخوردارند، یعنی منحنی توزیع نسبت به توزیع نرمال کوتاه تر است. این دسته از توزیع ها دارای ضریب کشیدگی منفی خواهد بود.

■ ب- توزیع هایی که از توزیع نرمال بلندترند. علت اوج گرفتن این توزیع ها آن است که داده ها حول میانگین متمرکزترند. عبارت دیگر از پراکندگی داده ها کاسته شده است. ضریب کشیدگی این دسته از توزیع ها مثبت خواهد بود.

■ ج- توزیع هایی که کشیدگی آنها با کشیدگی توزیع نرمال یکسان است. این توزیع ها، توزیع هایی با کشیدگی متوسط هستند. ضریب کشیدگی در این توزیع ها مساوی صفر است. اگر قدرمطلق ضریب کشیدگی بزرگتر از $0/5$ باشد، کشیدگی توزیع تفاوت زیادی با توزیع نرمال دارد.

کشیدگی (E) با توجه به رابطه زیر بدست می آید.

$$E = \frac{r_4}{s_x^4} - 3, \quad r_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

خطای استاندارد ضریب کشیدگی
(Standard Error of Kurtosis)، می تواند بعنوان آزمون
نرمال بودن استفاده شود (اگر کوچک تر از ۲- یا بزرگتر از ۲+
باشد، نرمال بودن رد می شود).

■ -تعیین شاخص های توصیف داده ها با استفاده از SPSS
نرم افزار SPSS از طریق دو دستور Frequencies و
Descriptive قابلیت ارائه شاخص های توصیفی را دارا
می باشد.

- اجرای دستور Frequencies

Analyze → Descriptive Statistics → Frequencies

متغیرهای کمی، همانند Salary، Population را انتخاب کرده و
به کادر Variable(s) منتقل کنید.

روی دکمه Statistics کلیک نمایید تا کادر
Frequencies : Statistics باز شود.

- گزینه Mean (میانگین داده ها)، Median (میانه)، Mode
(مد) و Sum (جمع داده ها) را در کادر Central Tendency
انتخاب کنید.

- گزینه Std.deviation (انحراف معیار)، Variance (واریانس)،
Range (دامنه تغییرات)، Maximum (بیشترین داده)،
Minimum (کمترین داده)، S.E.mean (خطای معیار
میانگین) را در کادر Dispersion انتخاب کنید.

■ گزینه Skewness (چولگی)، Kurtosis (کشیدگی) را در کادر Distribution انتخاب کنید.

■ در کادر Percent Value روی گزینه Quartiles کلیک کنید. با این عمل چارک های ۱، ۲ و ۳ محاسبه خواهند شد. به منظور ارائه صدک های روی گزینه Percentile(s) کلیک کرده و مقدار مورد نظر (برای مثال ۹۰) را در کادر مربوطه تایپ کنید. با کلیک روی دکمه Add این صدک انتخاب می شود و می توان برای دیگر صدک ها این عمل را تکرار کرد. روی Continue کلیک کنید تا به کادر Frequencies برگردد.

■ مثال- شاخص های پراکندگی (دامنه تغییرات، مینیمم، ماکزیمم، واریانس، انحراف استاندارد، چارک ها) و شاخص های شکل توزیع (چولگی و کشیدگی) را برای متغیر حقوق اولیه و همچنین حقوق جاری بصورت مجزا با استفاده از SPSS اجرا کنید.



Frequencies: Statistics



Percentile Values

- ☒ Quartiles
- ☐ Cut points for: equal groups
- ☒ Percentile(s):

Add

90.0

Change

Remove

Central Tendency

- ☒ Mean
- ☒ Median
- ☒ Mode
- ☒ Sum

☐ Values are group midpoints

Dispersion

- | | |
|--|---|
| ☒ Std. deviation | ☒ Minimum |
| ☒ <u>V</u> ariance | ☒ <u>M</u> aximum |
| ☒ <u>R</u> ange | ☒ S. <u>E.</u> mean |

Distribution

- ☒ Skewness
- ☒ Kurtosis

Continue

Cancel

Help

Statistics

		Salary	Population
N	Valid	3	3
	Missing	0	0
Mean		1216666.666	2387083.000
Std. Error of Mean		252212.4325	1016241.116
Median		1100000.000	1852305.000
Mode		850000.00 ^a	956312.00 ^a
Std. Deviation		436844.7474	1760181.245
Variance		1.908E11	3.098E12
Skewness		1.116	1.241
Std. Error of Skewness		1.225	1.225
Range		850000.00	3396320.00
Minimum		850000.00	956312.00
Maximum		1700000.00	4352632.00
Sum		3650000.00	7161249.00
Percentiles	25	850000.0000	956312.0000
	50	1100000.000	1852305.000
	75	.	.
	90	.	.