آمار برای مدیران   استاد: جناب آقای مجتبی عطائی   تهیه و تنظیم: ناصر ...

مراحل انجام یک پژوهش آماری

۱- تعیین جامعه آماری: جامعه آماری به مجموعه تمام افراد و یا اشیاء پیرامون موضوع مورد تحقیق را گویند.

مثال: پژوهش بررسی رابطه بین میزان درآمد کارمندان دانشگاه و وضعیت تحصیلی فرزندان آنها در استان تهران

جامعه آماری فوق: عبارت است از کلیه کارمندان دانشگاه در استان تهران که حداقل یک فرزند محصل دارند.

۲- تعیین متغیرهای آماری: متغیر آماری کمیت یا ویژگی که می‌تواند از یک واحد جامعه به واحد دیگر تغییر کند.

مثال: تعداد غیبت دانشجویان - جنسیت دانشجو - گروه خونی - اندازه قد - حقوق - میزان بدهی - نمره هوش و ...

متغیر:
- کمی: مقادیر آن بصورت عددی است. مثل: میزان سپرده، درجه حرارت، سن و ...
- کیفی: عددی نبوده و کیفی از سطوح خود را اختیار می‌کند. مثل: گروه خونی (O و A و B و AB)، رشته دانشجو، وضعیت تأهل

متغیر کمی:
- شمارشی: تعداد مراجعه، تعداد دانشجویان
- غیر شمارشی: سن، قد، زمان از دست رفته

۳- اندازه‌گیری متغیر (مقیاس): تخصیص عدد به متغیر را مقیاس گویند.

انواع مقیاس: ۱- اسمی ۲- ترتیبی ۳- فاصله‌ای ۴- نسبی

مقیاس:
- برای اندازه‌گیری متغیرهای کیفی مثل وضعیت جسمی ورزشکاران
  - ۱- اسمی
  - ۲- ترتیبی
- برای اندازه‌گیری متغیرهای کمی غیر شمارشی
  - ۳- فاصله‌ای
  - ۴- نسبی

ترتیبی		اسمی
۱	←	ضعیف
۲	←	متوسط
۳	←	قوی

مقیاس فاصله‌ای: علاوه بر خصوصیات مقیاس ترتیبی فاصله را نیز مشخص می‌کند اما نسبت در این مقیاس بی معنی است.

(چون صفر مطلق وجود ندارد و قراردادی است). مثال: درجه حرارت

تهران ۵°c
اراک ۰
اهواز ۳۰°c

اگر واحد تغییر کند مثلاً فارنهایت گردد دیگر صفر نمی‌باشد و صفر درجه سانتیگراد به ۳۲ درجه فارنهایت تبدیل می‌شود،

کلید نمرات صفرشان قراردادی است. مثال دیگر: نمره هوش

علی ۲۰۰
حسن ۱۰۰

نمی‌گوئیم علی دو برابر باهوشتر است. بلکه می‌گوئیم علی ۱۰۰ نمره نسبت به حسن باهوشتر است.

مقیاس / جهت	جهت تشخیص	معنی داری ترتیب	معنی داری فاصله	معنی داری نسبت	
اسمی	+	−	−	−	متأهل یا مجرد - مرد یا زن (دو سطحی)
ترتیبی	+	+	−	−	وضعیت جسمی ورزشکاران در مثال فوق
فاصله‌ای	+	+	+	−	نمره دانشجویان
نسبی	+	+	+	+	کاملترین مقیاس

مقیاس نسبی: کاملترین مقیاس است که علاوه بر خصوصیات فاصله‌ای، نسبت در این مقیاس با معنی است.

(صفر مطلق نیز وجود دارد). مثال: میزان دارائی، میزان قد، سرعت خودرو، میزان درآمد، وزن

۴- جمع آوری داده های آماری: داده آماری: مقادیر مشاهده شده از یک متغیر آماری که از طریق شمارش یا توسط یکی از مقیاسها اندازه گیری شده است.

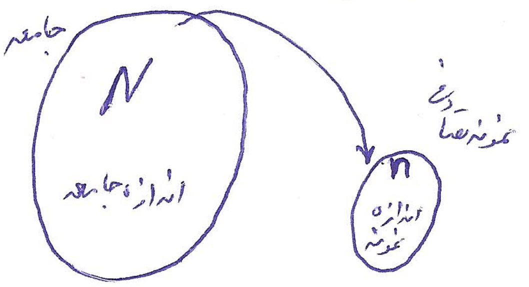
مثال: تعداد غیبت دانشجویان ۰-۱-۴-۰-۰-۳-۱-۱-۲-۳-۴-۴-۴ و...

گروه خونی ۲-۴-۳-۳-۱-۴-۲-۱-۱-۱

۱ ← A
۲ ← B
۳ ← AB
۴ ← O

مقیاس اسمی

جمع آوری داده های آماری:
- سرشماری: تمام واحدهای جامعه بررسی می شود.
- نمونه گیری: انتخاب قسمتی از جامعه (تصادفی) و بررسی آن.



دلایل نمونه گیری:
۱- صرفه جویی در وقت
۲- صرفه جویی در هزینه
۳- صرفه جویی در نیروی انسانی

هدف: استفاده از روشهای آماری برای خلاصه نمودن و تجزیه و تحلیل داده ها و بررسی فرضیه مورد تحقیق در صورت نمونه گیری.

روشهای آماری: ۱- آمار توصیفی ۲- آمار استنباطی.

I) آمار توصیفی: خلاصه نمودن و توصیف داده ها در قالب
۱- جدول توزیع فراوانی
۲- نمودارها
۳- معیارهای مرکزی، معیارهای پراکندگی، چندکها

داده های آماری:
- جدا (گسسته): مربوط به کمیت شمارشی و حاصل از مقیاس اسمی و ترتیبی.
- پیوسته: حاصل از مقیاس فاصله ای و نسبی

II) آمار استنباطی: پارامتری از جامعه مجهول است، مانند میانگین جامعه، نسبت جامعه، واریانس جامعه.

هدف از آمار استنباطی: تصمیم گیری در مورد پارامتر مجهول با استفاده از یک نمونه تصادفی n تایی است، که شامل انواع زیر است:

۱- برآورد نقطه ای.

۲- برآورد فاصله ای (فاصله اطمینان)

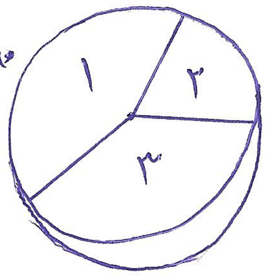
۳- آزمون فرض.

مقدمه ای بر آمار توصیفی

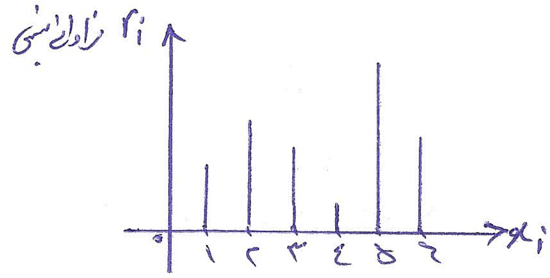
خلاصه کردن داده ها با استفاده از:
- جدول توزیع فراوانی
- نمودارها
- معیارهای مرکزی، پراکندگی و ...

نمودار داده های گسسته: ۱- نمودار میله ای یا ستونی (column chart) - ۲- نمودار دایره ای یا کلوچه ای (pie chart)

$\theta_i = r_i \times 360$

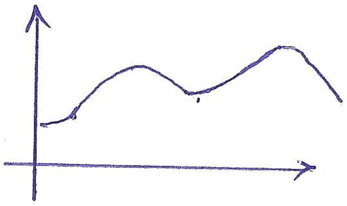


(نمودار دایره ای)

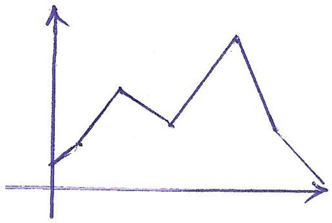


(نمودار میله ای)

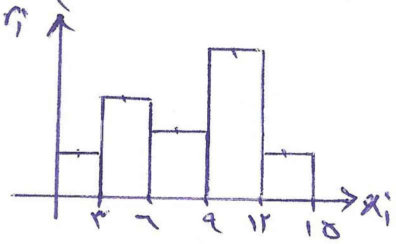
نمودار داده های پیوسته: ۱- هیستوگرام - ۲- چندبر فراوانی (چندضلعی فراوانی) - ۳- منحنی فراوانی



(۳)



(۲)



(۱)

معیارهای مرکزی: این معیارها مرکزیت داده ها را نمایش داده و بعنوان شاخص کل داده ها بکار می روند.

معیارهای مرکزی عبارتند از: ۱- میانگین حسابی - ۲- میانگین هندسی - ۳- میانگین وزنی - ۴- میانه - ۵- مُد

۱- میانگین حسابی (Avrage mean): مجموع داده ها تقسیم بر تعداد داده ها.

میانگین حسابی را با $\bar{x}$ و μ نمایش می دهند.

$$\text{میانگین حسابی} = \frac{\text{مجموع داده ها}}{\text{تعداد داده ها}}$$

x_i: داده

f_i: فراوانی

n: تعداد داده

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \bar{x} = \frac{\sum^{n} x_i}{n}$$

$$n = \sum f_i$$

مثال: محاسبه میانگین از روی جدول مقابل

میانگین تعداد تحصیل کرده در هر خانوار را محاسبه و تفسیر کنید:

x_i: تعداد تحصیل کرده در خانوارها

f_i: فراوانی

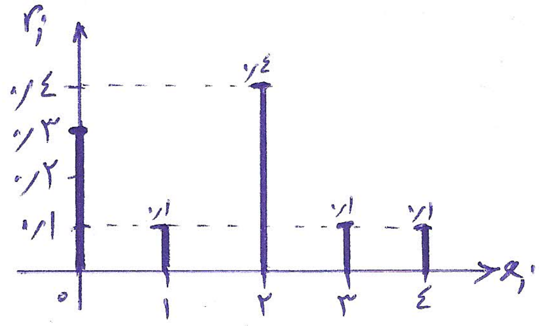
$$n = \sum f_i = 22$$

$$\bar{x} = \frac{\sum^{n} x_i}{n}$$

x_i	f_i	$f_i x_i$
۰	۸	۰
۱	۶	۶
۲	۳	۶
۳	۵	۱۵
جمع	۲۲	۲۷

$$\Rightarrow \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{27}{22} = 1.23$$

تفسیر: هر خانوار به طور متوسط ۱٫۲۳ را تحصیل کرده دارد.



مثال: نمودار میله‌ای تعداد روزهای مرخصی کارمندان انتخاب در ماه به صورت مقابل می‌باشد. میانگین تعداد روزهای مرخصی کارمندان انتخاب را محاسبه و تفسیر کنید.

همواره $\sum r_i = 1$ و $r_i = \frac{f_i}{n}$ فراوانی نسبی

x_i	r_i	$r_i x_i$
۰	/۳	۰
۱	/۱	/۱
۲	/۴	/۸
۳	/۱	/۳
۴	/۱	/۴
جمع	۱	۱٫۶

$$\bar{x} = \sum r_i x_i$$

$$\bar{x} = (0.3 \times 0) + (0.1 \times 1) + (0.4 \times 2) + (0.1 \times 3) + (0.1 \times 4)$$

$$\Rightarrow \bar{x} = 1.6$$

تفسیر: بطور متوسط هر کارمند ۱٫۶ روز مرخصی داشته است.

۲- میانگین هندسی: برای n داده مثبت تعریف می‌کنیم:

$$\bar{x}_g = \sqrt[n]{x_1 x_2 x_3 \cdots x_n}$$

در مواقعی که داده‌ها به صورت تصاعدی یا تقریباً تصاعدی تغییر می‌کند. این معیار بهترین معیار مرکزی است:

$$\log \bar{x}_g = \frac{\sum \log x_i}{n}$$

مثال: میانگین هندسی ۴ و ۲ و ۱

$$\bar{x}_g = \sqrt[3]{1 \times 2 \times 4} = \sqrt[3]{8} = 2$$

درصورتیکه میانگین حسابی آن $\bar{x} = \frac{1+2+4}{3} = 2.33$ خواهد بود.

۳ـ میانگین وزنی : اگر تمام مشاهدات هم ارزش نباشند یعنی (x_1 دارای ارزش وزنی w_1) و (x_2 دارای ارزش وزنی w_2) و ... و (x_k دارای ارزش وزنی w_k) باشد، آنگاه

$$\bar{x}_w = \frac{\sum^k w_i x_i}{\sum^k w_i}$$

مثال : دانشجوئی در ترم گذشته با ۱٤ واحد معدل ۱۵ را اخذ کرده است. اگر نمرات وی در دو درس ٤ واحدی به ترتیب (۱۳) و (۱٤/۵) و در درس ۳ واحدی (۱۷) باشد، نمره او در درس دیگر چقدر است.

w_i	x_i	$w_i x_i$
٤	۱۳	۵۲
٤	۱٤/۵	۵۸
۳	۱۷	۵۱
۳	x	۳x
۱٤ واحد		۱۶۱+۳x

$\bar{x}_w = ۱۵$ و $\sum w_i = ۱٤$

$$\Rightarrow \bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} \Rightarrow ۱۵ = \frac{۱۶۱+۳x}{۱٤} \Rightarrow x = ۱۶/۳۳$$

نمره دانشجو در درس دیگر ۱۶/۳۳ می باشد.

٤ـ میانه (median) : معیاری است که نیمی از داده ها کمتر یا مساوی آن و نیمی دیگر بزرگتر یا مساوی آن است. و با m نمایش داده می شود. در ابتدا داده ها را بصورت صعودی یا نزولی مرتب می کنیم. اگر تعداد داده ها (n) فرد باشد، عدد وسطی میانه است و اگر n زوج باشد، میانگین دو عدد وسطی میانه خواهد بود.

مثال : میانه داده های مقابل را بیابید.   ۱۲   ۱۷   ۲۰   ۸   ۱۵

مرتب می کنیم $\Rightarrow$ ۷   ۸   ۱۲   ۱۵   ۲۰ $\Rightarrow m = ۱۲$

x_i		f_i
خیلی کم	۱	۳
کم	۲	۲
متوسط	۳	۹
زیاد	٤	۱۰
خیلی زیاد	۵	٤
		$n=۲۸$

مثال : میانه را از جدول مقابل بیابید. (x میزان علاقه به درس آمار)

تعداد داده ها ۲۸ و نصف آن ۱٤ داده است، پس میانه بین ۳ و ٤ است.

پس $m = \frac{۳+٤}{۲} = ۳/۵$ می باشد.

نکته : تمام معیارهای مرکزی بین min و max داده ها می باشد، یعنی جواب را باید در ستون x جستجو کرد.

تفسیر : نیمی از دانشجویان میزان علاقه مندی شان متوسط یا کمتر است و نیمی دیگر زیاد تر یا برابر. به عبارت دیگر میزان علاقه به درس آمار متوسط متمایل به زیاد است.

۵- مُد (Mode) : داده ای که دارای بیشترین فراوانی یا بیشترین فراوانی نسبی باشد مُد نامیده می‌شود و با M نمایش داده می‌شود.

نکته اول: اگر دو داده ای که پهلوی هم نیستند دارای بیشترین فراوانی یا فراوانی نسبی باشند دو داده را مُد در نظر می‌گیریم. (دو مُدی)

مثال: در جدول مقابل مُد ۱ و ۳ می‌باشد: $M = 1$ و 3

داده x_i	فراوانی نسبی r_i
۰	۰/۱
۱	۰/۳۵
۲	۰/۱
۳	۰/۳۵
۴	۰/۱

نکته دوم: اگر دو داده ای که پهلوی هم هستند دارای بیشترین فراوانی یا فراوانی نسبی باشند، میانگین دو داده را مُد در نظر می‌گیریم.

مثال: نمرات آمار افراد انتخابی ۱۷ و ۲۰ و ۱۰ و ۱۷ و ۱۷ و ۱۸ و ۱۲ و ۱۵ و ۱۸ و ۱۳ و ۱۸ می‌باشد. مُد را تعیین کنید.

۱۷ و ۱۸ بیشترین فراوانی را دارند و چون پهلوی هم هستند، آنگاه

$$M = \frac{17+18}{2} = 17/5$$

نکته سوم: اگر فراوانی تمام داده ها یکسان باشد آنگاه مُد نداریم (بدون مُد). مثال: مُد وجود ندارد ۴ و ۴ و ۳ و ۳ و ۲ و ۲ و ۱ و ۱

معیارهای پراکندگی: این معیارها برای مقایسه پراکندگی دو یا چند گروه از داده ها بکار رفته و اغلب به تنهایی معنا دار نیستند و عبارتند از: ۱- دامنه تغییرات ۲- واریانس ۳- انحراف معیار ۴- ضریب تغییرات.

۱- دامنه تغییرات (Range): تفاضل بزرگترین داده ها با کوچکترین داده می‌باشد. $R = x_{max} - x_{min}$

این معیار برای مقایسه پراکندگی مناسب نیست. زیرا: ۱- به تعداد داده ها بستگی ندارد ۲- تابعی از تمام داده ها نیست (فقط به کوچکترین و بزرگترین داده بستگی دارد).

نمرات A: ۱۰ ۱۷ ۱۷/۵ ۱۷/۷۵ ۱۷/۷۵ ۱۸ $\Rightarrow R_A = 8$

نمرات B: ۱۰ ۱۲ ۱۳/۵ ۱۵ ۱۸ $\Rightarrow R_B = 8$

هرچند $R_A = R_B$ می‌باشد لکن مشاهده می‌شود پراکندگی B بیشتر از A است.

۲ـ واریانس (Variance) این معیار میزان پراکندگی داده ها را از میانگین آنها نمایش می دهد که هم به تمام داده ها و هم به تعداد داده ها بستگی دارد. واریانس نمونه را با s^2 و واریانس جامعه را با σ^2 نمایش می دهند.

$$\bar{x}=\frac{\sum^n x_i}{n}$$

x_i	x_1	x_2	------	x_n	$\sum^n x_i$
$x_i-\bar{x}$	$x_1-\bar{x}$	$x_2-\bar{x}$	......	$x_n-\bar{x}$	$\sum^n (x_i-\bar{x})=0$
$(x_i-\bar{x})^2$	$(x_1-\bar{x})^2$	$(x_2-\bar{x})^2$	......	$(x_n-\bar{x})^2$	$\sum^n (x_i-\bar{x})^2$

همواره مجموع انحراف از میانگین صفر می باشد

$\bar{x}$ میانگین نمونه $\longrightarrow$ واریانس نمونه $s^2=\dfrac{\sum^n (x_i-\bar{x})^2}{n-1}$

μ میانگین جامعه $\longrightarrow$ واریانس جامعه $\sigma^2=\dfrac{\sum^N (x_i-\mu)^2}{N}$ (N اندازه جامعه است).

مثال: واریانس داده های زیر را بیابید:

x_i	۵	۸	۵	۰	۳	۱۰	٤	۳۵
$x_i-\bar{x}$	۰	۳	۰	۵-	۲-	۵	۱-	۰
$(x_i-\bar{x})^2$	۰	۹	۰	۲۵	٤	۲۵	۱	٦٤

$$\bar{x}=\frac{\sum x_i}{n}=\frac{۳۵}{۷}=۵$$

$$s^2=\frac{\sum^n (x_i-\bar{x})^2}{n-1}=\frac{٦٤}{۷-۱}=\frac{٦٤}{٦}=۱۰/۷ \quad \text{واریانس}$$

نکته: فرمول ساده واریانس:

$$s^2=\frac{\sum (f_i x_i)^2-n\bar{x}^2}{n-1} \quad \text{یا} \quad s^2=\frac{\sum^n x_i^2-n\bar{x}^2}{n-1}$$

در مثال فوق: $\sum^n x_i^2=۵^2+۸^2+۵^2+۰+۳^2+۱۰^2+٤^2=۲۳۹$ و $\bar{x}=۵ \Rightarrow \bar{x}^2=۲۵$

$$\Rightarrow s^2=\frac{۲۳۹-۷\times۲۵}{۷-۱}=\frac{٦٤}{٦}=۱۰/۷ \quad \text{واریانس}$$

۳ـ انحراف معیار: جذر واریانس را انحراف معیار گویند

σ جامعه $=\sqrt{\sigma^2}$ و s نمونه $=\sqrt{s^2}$

۴- ضریب تغییرات: میزان پراکندگی نسبت به میانگین را نمایش می‌دهد.

$$C.V=\frac{\text{انحراف معیار}}{\text{میانگین}}\times 100 \Longrightarrow C.V=\frac{S}{\bar{x}}\times 100$$

خاصیت میانگین: اگر تمام داده‌ها را بصورت خطی و یکسان تغییر دهیم یعنی $y=ax+b$، آنگاه میانگین آنها نیز به همان صورت خطی تغییر می‌کند یعنی $\bar{y}=a\bar{x}+b$

خاصیت واریانس و انحراف معیار: اگر به تمام داده‌ها مقدار ثابتی اضافه یا کم کنیم این معیار تغییری نمی‌کند، و اگر تمام داده‌ها را در عدد ثابتی مانند a ضرب کنیم واریانس داده‌ها a^2 برابر واریانس قبلی می‌شود.
در حالت کلی اگر بصورت $y_i=ax_i+b$ تغییر دهیم آنگاه $S_y^2=a^2\times S_x^2$ و همینطور $S_y=|a|.S_x$

مثال: از حقوق کارمندان یک اداره ۱۰٪ بعنوان مالیات کاسته‌ایم، بدین ترتیب میانگین و انحراف معیار دریافتی آنها به ترتیب ۸۵۰ و ۵۰ هزار تومان شده است. میانگین و واریانس حقوق اولیه کارمندان را محاسبه کنید.

$$y_i=x_i-0.1x_i=0.9x_i \Longrightarrow \begin{cases}\bar{y}=0.9\bar{x}\\ S_y^2=(0.9)^2S_x^2\end{cases} \Longrightarrow$$

$$\Longrightarrow \begin{cases}850=0.9\bar{x} \Rightarrow \bar{x}=\frac{850}{0.9} \Longrightarrow \bar{x}=944.44\\ 2500=0.81S_x^2 \Longrightarrow S_x^2=\frac{2500}{0.81} \Longrightarrow S_x^2=3086.42\end{cases}$$

خاصیت ضریب تغییرات: این معیار به واحد اندازه‌گیری بستگی ندارد و برای مقایسه پراکندگی دو گروه از داده‌ها با واحدهای متفاوت بکار می‌رود. یعنی اگر داده‌ها را در عدد مثبتی ضرب کنیم این معیار تغییری نمی‌کند.

$$a>0,\quad y_i=ax_i \longrightarrow \begin{cases}\bar{y}=a\bar{x}\\ S_y=|a|S_x=aS_x\end{cases} \longrightarrow C.V_y=\frac{S_y}{\bar{y}}\times 100=\frac{aS_x}{a\bar{x}}\times 100=\frac{S_x}{\bar{x}}\times 100=C.V_x$$

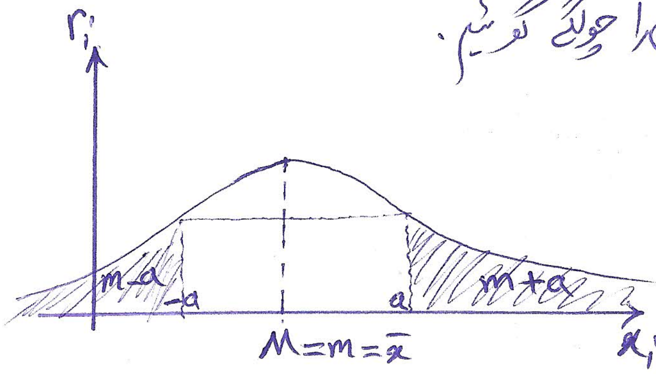
در مثال قبل ضریب تغییرات به چه نسبتی تغییر می‌کند؟

$$y_i=0.9x_i$$

$$\Longrightarrow C.V_y=\frac{50}{850}\times 100=5.88 \quad \text{و} \quad C.V_x=\frac{944.44}{\sqrt{3086.42}}\times 100=5.88$$

تغییر نمی‌کند.

۵- چولگی : عدم تقارن نسبت به منحنی تقارن را چولگی گوییم.

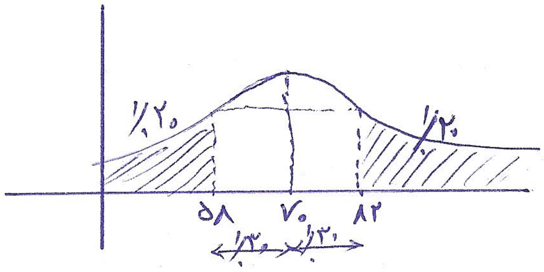


منحنی متقارن :

۱- در منحنی متقارن میانگین، میانه و مُد بر هم منطبق می‌باشند.

۲- در منحنی متقارن درصد داده‌های بزرگتر از m+a با درصد داده‌های کوچکتر از m-a برابر است. (قسمتهای هاشور خورده باهم برابرند).

مثال : اگر وزن ورزشکاران دارای منحنی متقارن بوده و اکثر آنها ۷۰kg باشند و ۲۰٪ بیش از ۸۲kg باشند، وزن چند درصد بین ۵۸ تا ۷۰ کیلوگرم می‌باشد؟

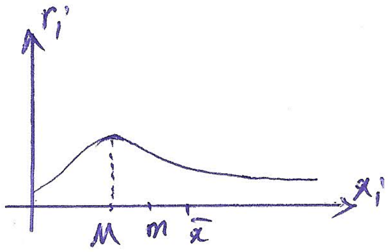


$M=m=\bar{x}=۷۰$

۵۰٪ کمتر از ۷۰kg و ۵۰٪ بیشتر از ۷۰kg

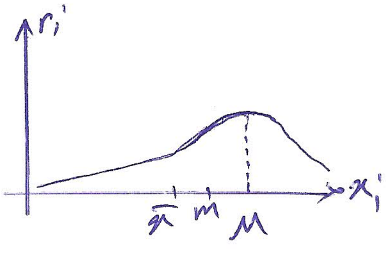
در نتیجه با توجه به شکل مقابل ۵۰٪ - ۲۰٪ = ۳۰٪ بین ۵۸ تا ۷۰ کیلوگرم می‌باشند.

انواع چولگی : ۱- چوله به راست ۲- چوله به چپ ،



۱- چوله به راست : الف) $M < m < \bar{x}$

ب) در منحنی چوله به راست، درصد داده‌های بزرگ از درصد داده‌های کوچک کمتر است



۲- چوله به چپ : الف) $\bar{x} < m < M$

ب) در منحنی چوله به چپ، درصد داده‌های بزرگ از درصد داده‌های کوچک بیشتر است.

* در منحنی‌ها همواره میانه بین میانگین و مُد می‌باشد.

ضریب چولگی پیرسون :

$S.K_1 = \frac{\bar{x}-M}{S}$ نوع اول

$S.K_2 = \frac{3(\bar{x}-m)}{S}$ نوع دوم

با توجه به اینکه $\bar{x}-M \simeq 3(\bar{x}-m)$ آنگاه

نکته :

۱- اگر $S.K > 0$ باشد آنگاه منحنی چوله به راست می باشد.

۲- اگر $S.K < 0$ باشد آنگاه منحنی چوله به چپ می باشد.

۳- هرچه $|S.K|$ افزایش یابد چولگی نیز افزایش می یابد.

۴- اگر $-0.1 \leq S.K \leq 0.1$ باشد آنگاه منحنی تقریباً متقارن می باشد.

۵- ضریب چولگی با واریانس یا انحراف معیار رابطه معکوس دارد.

۶- اگر به همه داده ها مقدار ثابتی اضافه کنیم ضریب چولگی تغییری نمی کند.

۷- در تغییر خطی با شیب مثبت $(a>0)$ ضریب چولگی تغییری نمی کند.

مثال : میانگین و انحراف معیار نمرات آمار به ترتیب ۱۳/۵ و ۱/۵ می باشد. اگر ضریب چولگی نوع اول برابر ۰/۴ باشد، الف) مد را حساب و مقادیر مجاز برای میانه را تعیین کنید. ب) اگر به همه نمرات ۲ نمره بیافزاییم ضریب چولگی چه تغییری می کند؟ ج) اگر نمرات را در ۱/۲ ضرب کنیم ضریب چولگی چه تغییری می کند؟

الف :

$S.K_1 = 0.4$ و $\bar{x} = 13.5$ و $S = 1.5$

$S.K_1 = \frac{\bar{x}-M}{S} \Rightarrow 0.4 = \frac{13.5-M}{1.5} \Rightarrow M = 12.9$ و $S.K_1 > 0 \xrightarrow{\text{چوله به راست}} 12.9 < m < 13.5$

ب : تغییری نمی کند

$S.K_y = \frac{\bar{y}-M_y}{S_y} = \frac{\bar{x}+2-(M_x+2)}{S_x} = \frac{\bar{x}-M}{S} = S.K_x = 0.4$

ج : تغییری نمی کند.

$S.K_y = \frac{\bar{y}-M_y}{S_y} = \frac{1.2\bar{x}-1.2M_x}{1.2S_x} = \frac{\bar{x}-M}{S} = S.K_x = 0.4$

آمار استنباطی

آمار استنباطی پارامتری از جامعه مجهول است. مانند میانگین جامعه (μ) - واریانس جامعه (σ^2) - نسبت جامعه (p)

هدف آمار استنباطی، نتیجه‌گیری و تصمیم‌گیری در مورد پارامتر مجهول جامعه با استفاده از یک نمونه تصادفی n تایی که شامل سه قسمت می‌باشد.

- برآورد نقطه‌ای
- برآورد فاصله‌ای
- آزمون فرض

تئوری نمونه‌گیری

نمونه تصادفی i.i.d : نمونه تصادفی $x_1, \ldots, x_n$ را i.i.d گوییم هرگاه مستقل و هم توزیع باشند.

۳ مرد

۲ زن

احتمال خروج مرد

$\frac{3}{10}$: با جایگزینی

$\frac{2}{9}$: بدون جایگزینی

در نمونه‌گیری با جایگزینی نمونه‌ها مستقلند درحالیکه بدون جایگزینی اینگونه نیست. در نمونه‌گیری بدون جایگزینی اگر جامعه نامتناهی (بزرگ) باشد نمونه‌ها مستقل خواهند بود.

$\frac{n}{N} \leq \%5$ شرط جامعه نامتناهی

$N = 500$ ، $n = 10$ $\longrightarrow \frac{n}{N} = \%2 < \%5$ جامعه نامتناهی

$N = 500$ ، $n = 50$ $\longrightarrow \frac{n}{N} = 0.1 > \%5$ جامعه متناهی

آماره: هر تابعی از نمونه تصادفی $x_1, \ldots, x_n$ که به پارامتر مجهول بستگی ندارد آماره گویند.

میانگین نمونه $\bar{x} = \frac{\sum^n x_i}{n}$ یک آماره است $\longrightarrow$ $\bar{x} - \mu$ آماره نیست

$\frac{\bar{x} - \overset{\text{معلوم}}{\mu_0}}{\frac{s}{\sqrt{n}}}$ آماره است

برآورد نقطه‌ای: هدف تخمین پارامتر مجهول θ توسط یک نمونه تصادفی n تایی است.

برآورد θ را با $\hat{\theta}$ (تتا هت) نمایش می دهند.

نکته: گوییم $\hat{\theta}$ یک برآورد نااریب برای θ است هرگاه $E(\hat{\theta})=\theta$ امید ریاضی

یعنی: تعداد زیادی نمونه n تایی انتخاب نموده و میانگین $\hat{\theta}$ ها را محاسبه کنیم برابر θ شود.

مثال: نشان دهید $\bar{x}$ برآورد نااریب میانگین جامعه (μ) است.

جامعه: $N=3$
سن ۲۵
سن ۲۶
سن ۳۰

$n=2$

$$\mu=\frac{25+26+30}{3}=\frac{81}{3}=27 \quad *$$

تعداد نمونه های n تایی برابر است با: $\binom{N}{n}=\frac{N!}{n!(N-n)!}$

نمونه های ۲ تایی برابر است با: $\binom{3}{2}=\frac{3!}{2!(3-1)!}=\frac{1\times2\times3}{(1\times2)\times(1)}=3$

$$\Rightarrow \left.\begin{array}{l} 25,26 \longrightarrow \bar{x}_1=25.5 \\ 25,30 \longrightarrow \bar{x}_2=27.5 \\ 26,30 \longrightarrow \bar{x}_3=28 \end{array}\right\} \Rightarrow \frac{\bar{x}_1+\bar{x}_2+\bar{x}_3}{3}=\frac{81}{3}=27 \quad **$$

$$\Rightarrow * = **$$

نتایج:

۱- $\bar{x}$ برآورد نااریب μ است.

۲- s^2 برآورد نااریب σ^2 است.

۳- $\hat{p}=\frac{x}{n}$ (نسبت نمونه) برآورد نااریب نسبت جامعه p است.

x تعداد واحد دارای مشخصه در نمونه n تایی است. $\hat{p}=\frac{x}{n}$

مثال: یک نمونه ۴۰۰ تایی از دانشجویان واحد الکترونیک انتخاب و نتایج زیر حاصل شده است.

جنسیت / علاقه به آمار	B_1 کم	B_2 متوسط	B_3 زیاد	جمع
A_1 مرد	۷۰	۳۰	۵۰	۱۵۰
A_2 زن	۴۵	۸۵	۱۲۰	۲۵۰
جمع	۱۱۵	۱۱۵	۱۷۰	۴۰۰

الف) برآورد کنید چه نسبتی از دانشجویان مرد بوده و علاقه زیادی به آمار دارند.

$$\hat{p} = \frac{50}{400} = \frac{1}{8} = 0.125$$

ب) برآورد کنید چه نسبتی از دانشجویان مرد بوده یا علاقه زیادی به آمار دارند.

حل: $n(A_1 \cup B_3) = n(A_1) + n(B_3) - n(A_1 \cap B_3) = 150 + 170 - 50 = 270$

$$\Rightarrow \hat{p} = \frac{270}{400} = 0.675$$

آزمون فرض:

پارامتری از جامعه مجهول است (مانند θ) دو فرض H_0 و H_1 در مورد پارامتر مجهول مطرح می‌شود.

فرض صفر H_0: خلاف ادعا

فرض مقابل H_1: ادعای مطرح شده در مورد θ

بنابراین فرضهای H_0 و H_1 به یکی از سه فرم زیر است: $\theta_0 =$ عدد ثابت

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta \end{cases}$$

یکطرفه راست

$$\begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$

یکطرفه چپ

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

دوطرفه