گزیده و نکات بخش دوم، آماره‌های توصیفیِ نمونه‌ها
"Introductory Statistics" by Wonnacott

مهدی صادق‌زاده قمصری

بهمن ۹۹

۱ آماره

هدف اولیه علم آمار، بدست آوردن استنباطی درباره جامعه از طریق نمونه آن است. یک گام اولیه می‌تواند این باشد که نمونه را در چند عدد به طور خلاصه توصیف کرد. به هر کدام از این اعداد توصیف‌کننده‌ی نمونه، آماره[۱]ی نمونه گفته می‌شود.

۲ تکرار[۲]

- فضای گسسته
متغیر تصادفی X متغیر تصادفی گسسته نامیده می‌شود، اگر مقادیر آن محدود و یا نامحدود اما قابل شمارش باشند. خروجی انداختن تاس یک مثال از این متغیر هاست. برای ساده‌سازی درک نمونه، میتوان تکرار حالات مختلف نمونه را به عنوان یک آماره در نظر گرفت. آماره دیگر می‌تواند نسبت این تکرار ها به کل حالات[۳] باشد.
در کنار هم قرار دادن این اعداد برای تمامی حالت یک متغیر تصادفی، به ما توزیع[۴] آنها را خواهد داد که با دیدن نمودار این توزیع‌ها یک تصور از نتیجه نمونه در ذهن شکل می‌گیرد.

---

[۱] statistic
[۲] frequency
[۳] relative frequency
[۴] distribution

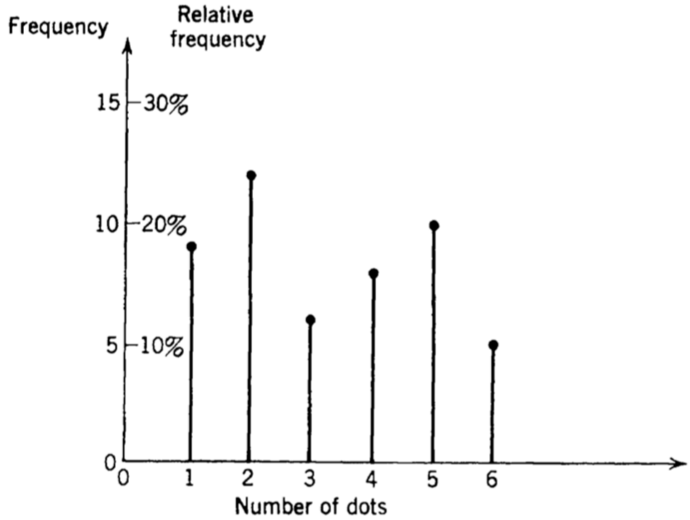

FIG. 2-1 Frequency and relative frequency distribution of the results of a sample of 50 tosses of a die.

- فضای پیوسته

اگر نمونه متعلق به یک فضای پیوسته باشد، متغیر تصادفی X می‌تواند هر مقداری بگیرد و دیگر تکرار یک مقدار مشخص از آن معنی ندارد(چون احتمال تکرار آن به صفر میل میکند)؛ نتیجتا به‌جای تکرار مقادیر مشخص، تکرار مقادیر در بازه های مشخص را می‌شمارند.
تعداد این بازه ها باید با مجموعه حالات تناسب داشته باشد و همچنین وسط بازه ها، نمایانگر تمامی مقادیر بازه باشد.
به نمودار توزیع این متغیر های تصادفی Histogram نیز گفته می‌شود.

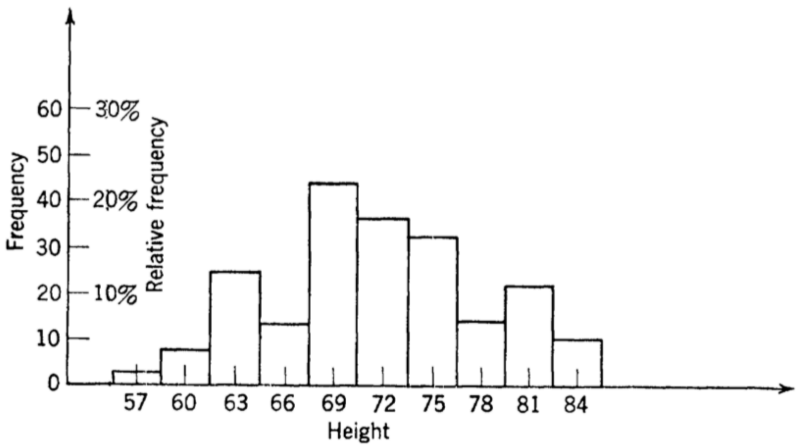

FIG. 2-3 The frequency and relative frequency distribution of a sample of 200 men.

برای توصیف توزیع تکرار یک نمونه، می‌توان از دو آماره توصیفی استفاده کرد: نقطه‌ی

وسط توزیع و میزان گسترش توزیع.

۳ مقیاس های مکانی

تعاریف مختلفی از "وسط" توزیع وجود دارد که ۳ تای از آنها مد، میانه و میانگین هستند.

- مد[۵]: بیشترین مقدار تکرارشده در نمونه
این مقیاس مناسبی برای وسط یک نمونه نیست؛ زیرا به بازه‌بندی داده ها وابسته است؛ و همچنین می‌توان در یک توزیع چند نقطه بیشینه برابر داشت که این تعریف برای آنها گنگ بوده و به آن توزیع ها "bimodal" گفته می‌شود.

- میانه[۶]: پنجاه‌مین صدک داده‌ها
بعضا به آن مقدار میانی نیز گفته می‌شود. اگر تنها داده از نمونه توزیع تکرار آن باشد، باید در انتخاب مقدار از بازه میانی دقت کرد.

- میانگین($\bar{X}$)[۷]: حاصل تقسیم مجموع مقادیر نمونه بر تعداد آنها

$$\bar{X} 1/n \sum_{i=1}^{n} X_i \tag{۱}$$

بیدر زمانی که با بازه ها سر و کار داریم؛ می‌توان مقادیر درون یک بازه را به طور تقریبی برابر مقدار وسط بازه درنظر گرفت:

$$\bar{X} \simeq 1/n(f_1 x_1 + f_2 x_2 + ... + f_k x_k)$$

نتیجتا:

$$\bar{X} \simeq \sum_{i=1}^{m} x_i (f_i/n) \tag{۲}$$

---

[۵] mode
[۶] median
[۷] mean

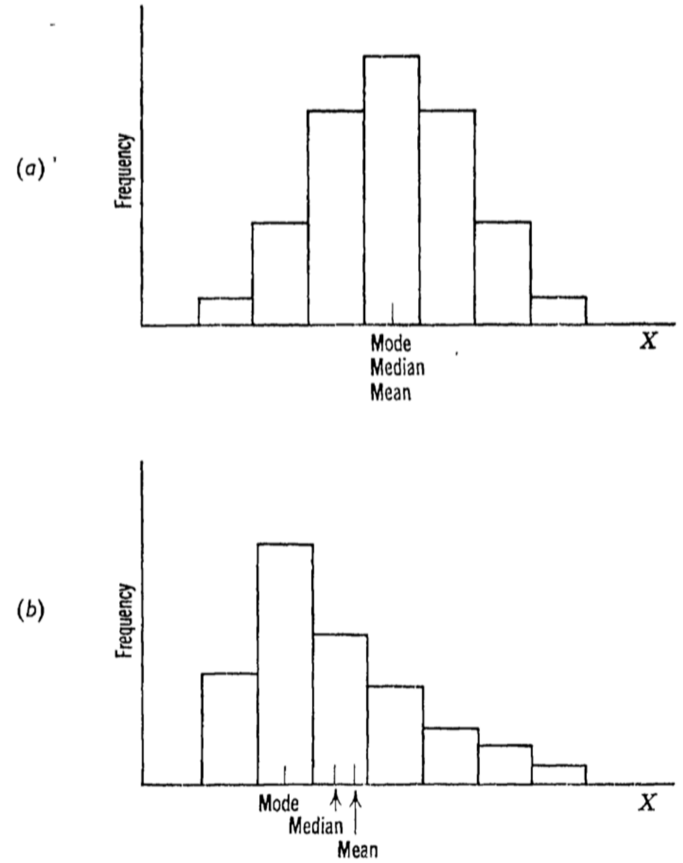

FIG. 2-4(*a*) A symmetric distribution with a single peak. The mode, median, and mean coincide at the point of symmetry. (*b*) A right-skewed distribution, showing mode < median < mean.

نکات:

- برای تساوی ای که مبتنی بر تعریف است از نشانه $\triangleq$ استفاده می‌کنیم.
- برای نشان دادن مقدار تقریبی از نشانه $\simeq$ استفاده می‌کنیم.

۴ مقیاس های پراکندگی

ساده‌ترین مقیاس برای پراکندگی فاصله میان کوچک‌ترین و بزرگ‌ترین مقدار نمونه است که به آن بازه گفته می‌شود. این مقیاس تنها به همین دو مقدار وابسته است و به سایر

مقادیر نمونه کاری ندارد. برای رفع این مشکل مقیاس انحراف[۸] از میانگین تعریف می‌شود؛

$$TheMeanAbsoluteDeviation 1/n \sum_{i=1}^{n} |X_i - \bar{X}| \quad (۳)$$

و برای مشتق‌پذیری بهتر از توان دوی فاصله استفاده می‌شود:

$$TheMeanSquaredDeviation(MSD) 1/n \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad (۴)$$

در بخش ۷ به این می‌پردازیم که برای استنباط آماری بهتر درباره نمونه، می‌توان از n-1 بجای n استفاده کرد. بدین ترتیب آماره واریانس حاصل می‌شود:

$$Variance, s^2 1/(n-1) \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad (۵)$$

$$= 1/(n-1)(\sum_{i=1}^{n} (X_i^2) - n\bar{X}^2)$$

که واریانس با توان دوی فاصله از میانگین رابطه دارد. برای استاندارد کردن این رابطه مقیاس انحراف معیار را تعریف می‌کنیم.

$$StandardDeviation, s \sqrt{1/(n-1) \sum_{i=1}^{n} (X_i - \bar{X})^2} \quad (۶)$$

از میان این مقیاس ها، از میانگین بیشتر به عنوان مقیاس مرکز نمونه استفاده می‌شود؛ و همچنین از انحراف معیار نیز بیشتر به عنوان مقیاس پراکندگی استفاده می‌شود.

۵ تبدیل های خطی

$$\begin{aligned} X_i' &= bX_i + a \\ \Rightarrow \bar{X}' &= b\bar{X} + a \\ \Rightarrow s_{X'} &= |b| s_X \end{aligned}$$

[۸] deviation