

Describing the Central Tendency of Data

Lecture 5

Central Tendency

Reading: Chapter 3: Central Tendency (skip "finding the precise median for continuous variables")
Seeing Statistics 3

Statistics

- Numerical summaries that characterize the shapes of distributions, allowing efficient comparisons between distributions
- Central tendency (typical value)
- Variability (spread or disagreement)
- Skew (long-tail one direction or other)

Central Tendency Statistics

- Mean (M or \bar{X}): Average Value

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

- Median (*median* or \hat{X}): Value of score in the middlemost location (or median location)
 - Value of score in $(n + 1)/2$ location
- Mode (*mode*): Most frequent value

Computing the Mean of 9 arithmetic questions

value	freq.	prop.	cfreq.(a)	cprop.(a)	cfreq.(d)	cprop.(d)
1	0	0.00	0	0.00	114	1.00
2	0	0.00	0	0.00	114	1.00
3	1	0.01	1	0.01	114	1.00
4	2	0.02	3	0.03	113	0.99
5	9	0.08	12	0.11	111	0.97
6	15	0.13	27	0.24	102	0.89
7	15	0.13	42	0.37	87	0.76
8	39	0.34	81	0.71	72	0.63
9	33	0.29	114	1.00	33	0.29

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} = \frac{(3+4+4+5+5+...+9)}{114} = \frac{860}{114} = 7.54 \\ &= \frac{\sum_{j=1}^n n_j X_j}{n} = \frac{(1*3)+(2*4)+(9*5)+...+(33*9)}{114} = \frac{860}{114} = 7.54\end{aligned}$$

Computing the Mean of 9 algebra questions

value	freq.	prop.	cfreq.(a)	cprop.(a)	cfreq.(d)	cprop.(d)
1	0	0.00	0	0.00	114	1.00
2	0	0.00	0	0.00	114	1.00
3	4	0.04	4	0.04	114	1.00
4	23	0.20	27	0.24	110	0.96
5	36	0.32	63	0.55	87	0.76
6	33	0.29	96	0.84	51	0.45
7	9	0.08	105	0.92	18	0.16
8	7	0.06	112	0.98	9	0.08
9	2	0.02	114	1.00	2	0.02

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} = \frac{(3+3+3+3+4+...+9)}{114} = \frac{619}{114} = 5.43 \\ &= \frac{\sum_{j=1}^n n_j X_j}{n} = \frac{(4*3)+(23*4)+...+(2*9)}{114} = \frac{619}{114} = 5.43\end{aligned}$$

Computing the Median of 9 arithmetic questions

value	freq.	prop.	cfreq.(a)	cprop.(a)	cfreq.(d)	cprop.(d)
1	0	0.00	0	0.00	114	1.00
2	0	0.00	0	0.00	114	1.00
3	1	0.01	1	0.01	114	1.00
4	2	0.02	3	0.03	113	0.99
5	9	0.08	12	0.11	111	0.97
6	15	0.13	27	0.24	102	0.89
7	15	0.13	42	0.37	87	0.76
8	39	0.34	81	0.71	72	0.63
9	33	0.29	114	1.00	33	0.29

$$\text{Location of Median} = (114+1)/2 = 57.5$$

$$\text{Median} = 8$$

Computing the Median 9 algebra questions

value	freq.	prop.	cfreq.(a)	cprop.(a)	cfreq.(d)	cprop.(d)
1	0	0.00	0	0.00	114	1.00
2	0	0.00	0	0.00	114	1.00
3	4	0.04	4	0.04	114	1.00
4	23	0.20	27	0.24	110	0.96
5	36	0.32	63	0.55	87	0.76
6	33	0.29	96	0.84	51	0.45
7	9	0.08	105	0.92	18	0.16
8	7	0.06	112	0.98	9	0.08
9	2	0.02	114	1.00	2	0.02

Location of Median = $(114+1)/2 = 57.5$

Median = 5

Finding the median if n is an even number

5,6,6,7,7,9,10,11,11,13

$n = 10$

median location = $\frac{n+1}{2} = 5.5$

median = $\frac{7+9}{2} = 8$

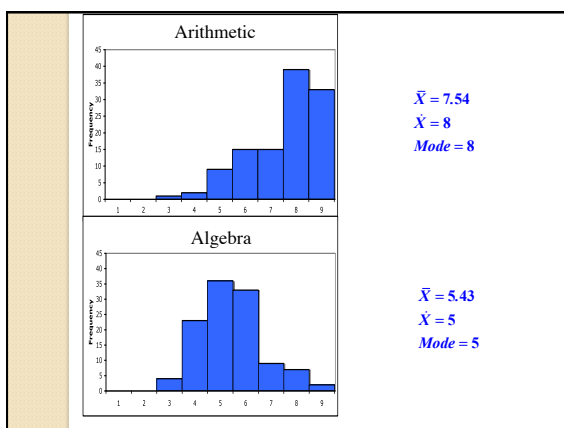
Finding the median if n is an odd number

5,6,6,7,7,9,10,11,11

$n = 9$

median location = $\frac{n+1}{2} = 5$

median = 7



Desirable Properties of Sample Statistics

- **Unbiased:** On average the sample statistic equals the population parameter
 - Both mean and median unbiased
- **Efficient (or Accuracy):** A more efficient statistic is one that varies less from sample to sample; on average it is closer to the population parameter
 - In general, mean is more efficient than the median
 - *Seeing Statistics* 3.4.3
- **Consistent:** A consistent statistic is one where efficiency increases as the sample size increases (We will understand this better later.)
 - Both the mean and the median are consistent

Mean vs. Median

- Both based on a notion of balance
- Mean sensitive to each datum's distance from middle
- Median better for irregular distributions
 - Skew
 - Outliers

Mode

- Most common value
- Peak in the distribution for continuous variables
- Simple and insensitive
- Most useful when mean, median not definable
 - College majors, sex, favorite color