Announcements

- HW 1 Due Thurs Jan 29th 11:59 PM (Do before quiz)
- Quiz 1 Thurs Jan 29th in discussion

- Lecture Review

- Def) Population

The population is the entire group being studied, of size N

- Def) Sample

A sample is a subset of the population which we draw data from, size n

$$\{X_1, X_2, \ldots, X_n\}$$

$$\{X_1, X_2, X_3, X_4\}$$

- Example)

You conduct study, average height UI undergrads by sampling 100 people on Penta Crest

Population : UI undergrads

Sample : 100 people on Pentacrest

- Def) Parameter
  A parameter is a numerical summary of the population
- Def) Statistic
  A statistic is a numerical summary of the sample
- Ex)

Parameter μ = average height of UI undergrads

Statistic $\bar{X}$ = average height of 100 people on Pentacrest

- Def) Sample mean

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

"balance point of the data"

- Example) 5 people pool money together

$\{10, 5, 3, 7, 20\}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{5} \sum_{i=1}^{5} x_i$$

$$= \frac{1}{5} [10 + 5 + 3 + 7 + 20]$$

$$= \frac{1}{5} [45] = 9$$

- Def) Median

"middle" of the data

half the data is below median, half above

- Example

data = $\{\cancel{1}, \cancel{2}, 3, 6, \cancel{7}, \cancel{8}\}$

$\frac{3+6}{2} = \frac{9}{2} = 4.5$ = median

data = $\{\cancel{1}, \cancel{2}, 3, \cancel{7}, \cancel{8}\}$

median = 3

- Def) Mode

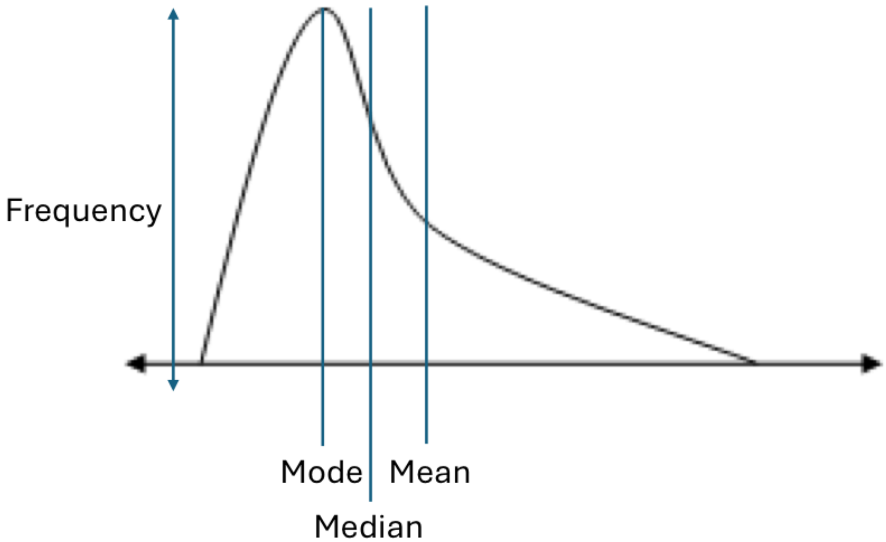
data point that "appears" most often

- Example 3

data = $\{1, 2, 2, 2, 3, 3, 5\}$

mode = 2

This graph is said to have right-skew. Think of it as "right-tailed."
Where is the mean, median, and mode on this graph?



Problem 1: Two samples are drawn and listed below:

List A: {47, 48, 49, 50, 51, 52, 53} $n = 7$

List B: {20, 30, 40, 50, 60, 70, 80}

a. What is the sample mean of each list?

b. What is the sample median of each list?

(a)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{1}{7}\left[47 + 48 + \dots + 53\right]$$

$$= \frac{1}{7}[350] = 50$$

list B, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$= \frac{1}{7}[350] = 50$$

(b) : $\{\cancel{47}, \cancel{48}, \cancel{49}, 50, \cancel{51}, \cancel{52}, \cancel{53}\}$

$$= 50$$

- Def) Population Variance

$$\sigma^2 = \frac{1}{N}\sum (x_i - \mu)^2$$

represents the "spread" of the data

μ = Population mean $= \frac{1}{N}\sum_{i=1}^{N} x_i$

- Def) Sample Variance

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

$\frac{1}{n-1}$ instead of $\frac{1}{n}$

Bessel's correction
error in estimating $\bar{x}$ instead of μ causes sample variance to be too small

- Def) Population & Sample SD

$\sigma = \sqrt{\sigma^2}$, $s = \sqrt{s^2}$
Calculate the "typical distance from the mean"

List A: {47, 48, 49, 50, 51, 52, 53}

List B: {20, 30, 40, 50, 60, 70, 80}

c. Without calculating, which list do you expect to have a lower sample standard deviation?

d. What is the sample variance of each list? (Feel free to use *Excel*)

e. What is the sample standard deviation of each list?

f. What is Σx_i of each list?

(c) Standard deviation is "average spread" of data. List A has lower standard deviation

(f)

$\bar{x} = \frac{1}{n} \sum x_i, \quad \sum x_i = 350$

Problem 2: Answer the following summation problems:

$$x_1 = 9, x_2 = 9, x_3 = 12$$

a. Σx_i

b. $\Sigma(x_i - \bar{x})$

c. $\Sigma(x_i - \bar{x})^2$

(a) Σx_i, $\Sigma_{i=1}^{n} x_i$

$\Sigma x_i = \Sigma_{i=1}^{3} x_i$

$= 9 + 9 + 12 = 30$

(b) $\Sigma(x_i - \bar{x})$

$\bar{x} = \frac{1}{n} \Sigma_{i=1}^{n} x_i$

$= \frac{1}{3} \Sigma_{i=1}^{3} x_i$

$= \frac{1}{3}(30) = 10$

(c) $\Sigma(x_i - \bar{x})^2$

$= [(9-10)^2 + (9-10)^2 + (12-10)^2]$

$= [(-1)^2 + (-1)^2 + (2)^2] = 6$

Problem 3: Mean, Median, and Mode Problems:

a. There are 3 exams in a particular semester where every exam is weighted equally. If you score an 87 and an 85, what do you need on the 3rd exam for your mean to be a 90?

b. Is it possible to have a larger sample size than a population size? Why or why not?

$X_1 = 87,\ X_2 = 85,\ X_3 = ?$

$\bar{X} = 90$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{3}\sum X_i$

$90 = \frac{1}{3}[87 + 85 + X_3]$

$270 = [87 + 85 + X_3]$

$98 = X_3$

(b) No, we cant draw a larger subset than the population

Problem 4: Calculate the sample variance by hand.

$$\{-5, 0, 5\}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$(n=3)$

$$s^2 = \frac{1}{(3)-1} \sum_{i=1}^{3} (x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{3}(-5+0+5)$$

$$= 0$$

$$s^2 = \frac{1}{2} \sum_{i=1}^{3} (x_i - 0)^2$$

$$= \frac{1}{2}\left[(-5)^2 + 0^2 + 5^2\right]$$

$$= \frac{1}{2}[50] = 25$$