STAT 231

Roadmap

- 5 min recap
- Numerical data summaries

Empirical

Obs | Exp

Obs $\rightarrow$ more conceptual.

$\downarrow$

Sample Survey

Target pop. is finite

Data

- Num
  - Discrete
  - Continuous
- Categorical
  - ORDINAL
  - NON-ORDINAL.

Numerical Data Summaries

(i) MEASURES OF LOCATION

$$\{y_1, \ldots y_n\}$$

S.M : $\bar{y} = \frac{1}{n} \sum y_i$

G.M = $(y_1 \ldots y_n)^{1/n}$.

Median $\hat{m}$ = Middle most observation

$\{y_1, \ldots y_n\}$

$\{y_{(1)}, \ldots y_{(n)}\}$ → Data is ordered

$y_{(1)} \leq y_{(2)} \ldots \leq y_{(n)}$

If n is odd, the median is a member of the data. → $y_{\left(\frac{n+1}{2}\right)}$

1, 3, 7, 13, 25

If n is even, = Average of the two middle ones

$1, 3, \underbrace{7, 13}, 25, 36$

Median: $\frac{7+13}{2} = 10$

$$\frac{y_{(n/2)} + y_{(n/2+1)}}{2}$$

QUARTILES: The data set is divided into four equal parts

Q_1 = Lower quartile : 25% of the obs. lie below it

Q_3 = Upper quartile : 75% . . . .

Algorithm :

Percentile : Data set is divided into 100 equal parts

$y_{(1)}, \ldots\ y_{(n)} \longrightarrow$ Data set.

We want to find the p^{th} percentile. $p \in (0,1)$

Calculate

$m = (n+1) \times p \longrightarrow$ Integer then that obs. is our answer.

$n = 39 \qquad p = 0.75$

$y_{(30)} = Q_3$

If m is not an integer, then the quantile = average of the two nearest integers.

<u>MODE:</u> Observation that occurs with the maximum frequency

1, 1, 3, 3, 3, 5, 13

Mode: 3

Mode need not be unique.

Median is ~~too~~ more robust to extreme observation as compared to the mean.

MEASURES OF DISPERSION

St. Petersburg's paradox

	Prob.	
2	$1/2$	1
4	$1/4$	1
8	$1/8$	1
16	.	.
32	.	.
.	.	.
.	.	.
.	.	.

$$\infty$$

100 — 0.9

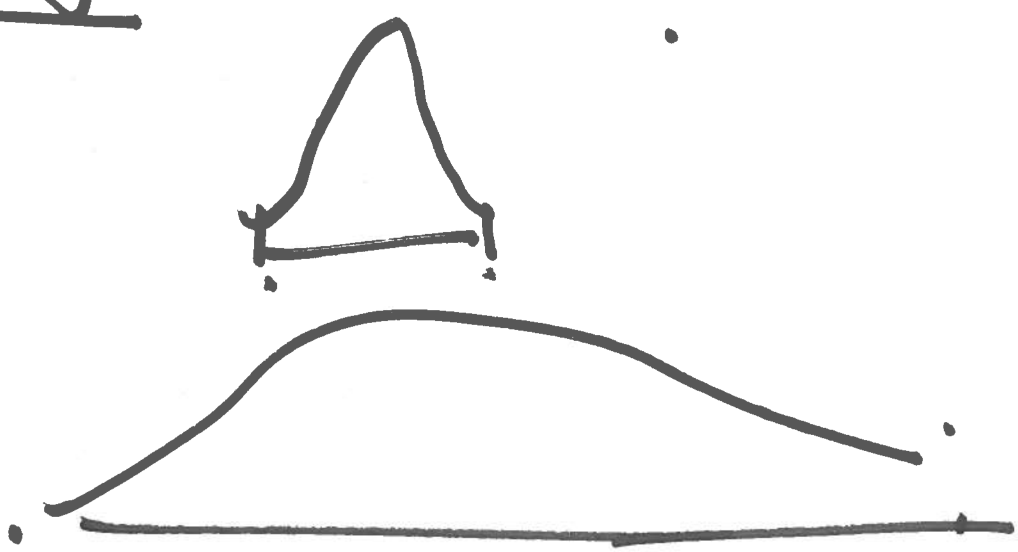
$10 \rightarrow 0.1$

Goalie 1 : 0 6 0 6 0 6

Goalie 2 : 3 3 3 3 3 3

Measures

Range = Max − Min



IQR = Inter-Quartile Range.

$= Q_3 - Q_1$

Sample Variance & Standard deviation

$s^2 =$ "Average" of the squared deviations from the mean.

$y_1, \ldots y_n$.

$$s^2 = \frac{(y_1-\bar{y})^2 + (y_2-\bar{y})^2 + \cdots + (y_n-\bar{y})^2}{n-1}$$

$$\boxed{s^2 = \frac{1}{n-1}\sum (y_i-\bar{y})^2}$$

ST. DEV. $s =$ Positive Square root of the variance.

$y_1, \ldots y_n$

$$\frac{(y_1 - \bar{y}) + \ldots + (y_n - \bar{y})}{n}$$

$= 0$

$\{x_1, \ldots x_n\}$

$y_i = a + b x_i$

$\{y_1, \ldots y_n\}$

Is $\bar{y} = a + b\bar{x}$? Yes.

$$y_i = a + bx_i$$

$$s_y^2 = b^2 s_x^2$$

$$s_y = |b|\, s_x$$

$$s_x^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n-1}$$

$$(a+bx_1) - (a+b\bar{x})$$

New range = $b \times$ Old Range.