

# **Doubt Session - Statistics**

# RECAP

## Statistics

branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data.

### Descriptive

methods for summarizing and describing the features of a data set.

#### Measure of Central Measures

describe the center or typical value within a data set.

Mean  
Median  
Mode  
Quartiles  
Percentile

#### Measure of Variability

provide information about the dispersion or spread of data points within a data set.

Range  
Variance  
Standard Deviation

### Sampling

used to select a subset of individuals, items, or data points from a larger population or dataset.

Random Sampling  
Systematic Sampling  
Stratified Sampling

### Central Limit Theorem

### Inferential

methods are used to make generalizations or predictions about a population based on a sample of data.

#### Error (Bias)

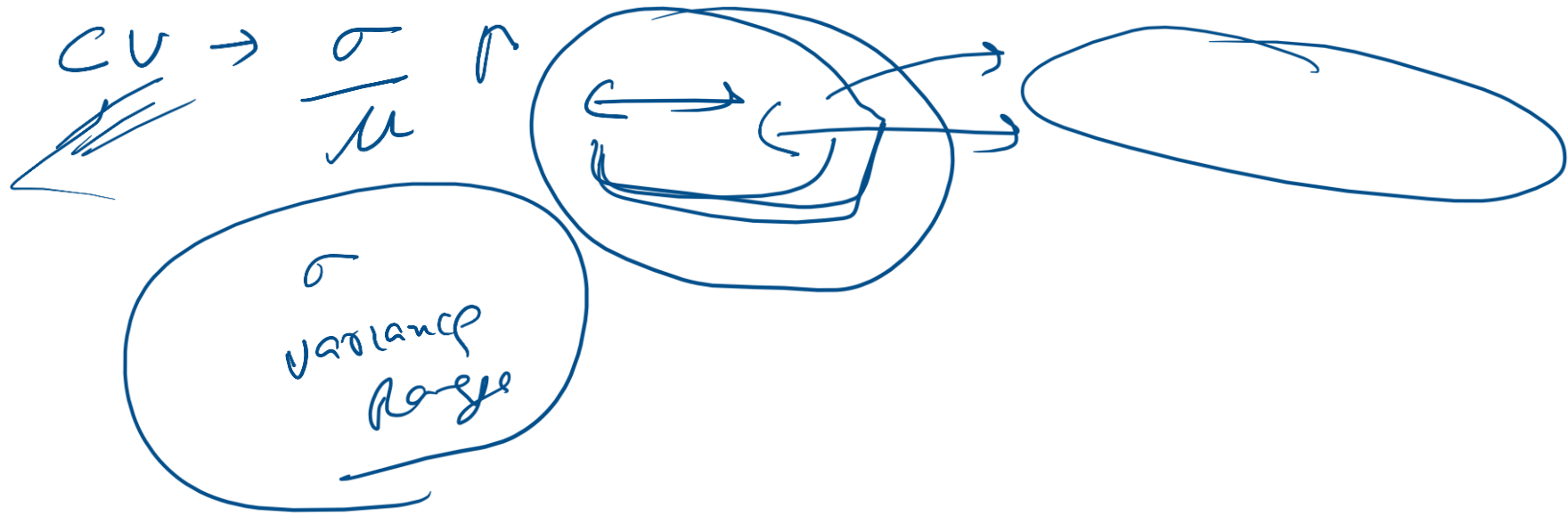
refer to inaccuracies or deviations from the true value or state

Undercover Bias  
Exclusion bias  
Survivorship Bias

#### Point & Range Estimation

Provide the best value (single or range) based on the historical data

describes the characteristics of sampling distributions.



11









What is the mean of the numbers: 10, 20, 30, 40, and 50?

A) 25

✓ B) 30

C) 35

D) 40

$$\mu = \frac{10 + 20 + 30 + 40 + 50}{5} = \underline{\underline{30}}$$



What is the mode of the following dataset: 4, 5, 6, 6, 7, 8, 8, 8, 9?

A) 6

B) 7

C) 8

D) 9

6 - 2 2 #

7 = 1 #

8 - 3 #

1 1 #

3

(6, 8)

**Q. Which of the following is NOT a measure of central tendency?**

a) Mean ✓

b) Mode ✓

✓ c) Range ✓ → Variability

d) Median ✓

**Q. What is the difference between population and sample in statistics?**

A) A sample includes the entire population

B) A population is a subset of a sample

✓ C) A sample is a subset of a population

D) A sample and a population are the same

**Q. What is the main purpose of descriptive statistics?**

A) To infer conclusions from a sample


☒ B) To summarize and describe data

C) To conduct hypothesis testing

☐ D) To predict future outcomes

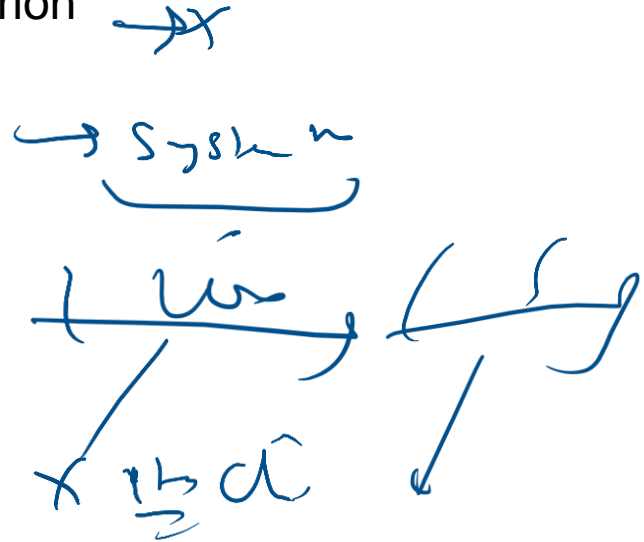
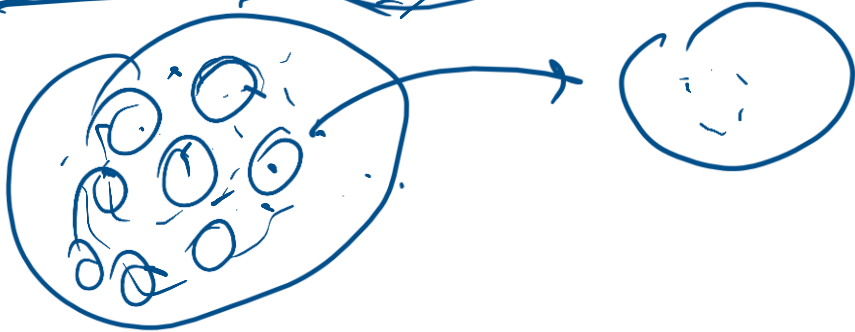
→ inferential

## Q. Why is sampling important in statistics?

- A) It allows researchers to study the entire population
-  B) It is easier and more cost-effective than studying the whole population
- C) It always provides 100% accurate results
- D) It eliminates bias completely

## Q. What is random sampling?

- A) Choosing participants based on convenience
- ☒ B) Every individual has an equal chance of being selected
- C) Selecting only people who meet a certain criterion
- D) Choosing the first 10 people in a population



## What does the range of a dataset represent?

A) The average value of the dataset

☒ B) The difference between the highest and lowest values

C) The middle value of the dataset

D) The most frequently occurring value



2 3 4 50

3.5  $\rightarrow$   $\times 7.5$

mean sensitive

yes

to outlier

yes or NO

$$\text{Range} = X_{\max} - X_{\min}$$

yes  $\rightarrow$  sensitive to outlier?

median sensitive to outlier

NO 2 3 4 5 10

## Which of the following is true about standard deviation?

☒ A) It measures the spread of data

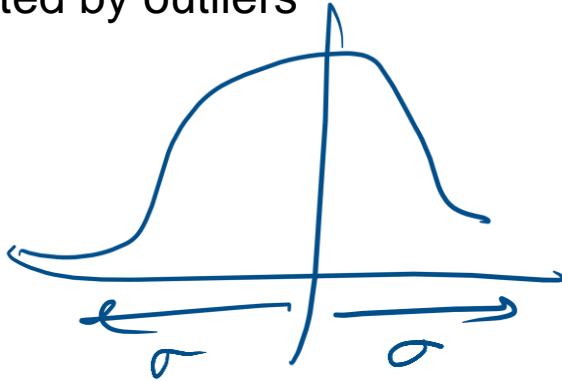
B) It is the same as variance

→  $\text{Variance} = SD^2$

C) It is always greater than the mean

$$SD = \sqrt{\text{Variance}}$$

D) It is unaffected by outliers





## What does the Central Limit Theorem (CLT) state?

A) The sum of data is always normally distributed

B) The distribution of sample means approaches normality as the sample size increases



C) The mean of a population is always equal to the sample mean

D) Every dataset is normally distributed

If the first quartile ( $Q_1$ ) = 20 and third quartile ( $Q_3$ ) = 50, what is the interquartile range (IQR)?

A) 20

☒ B) 30

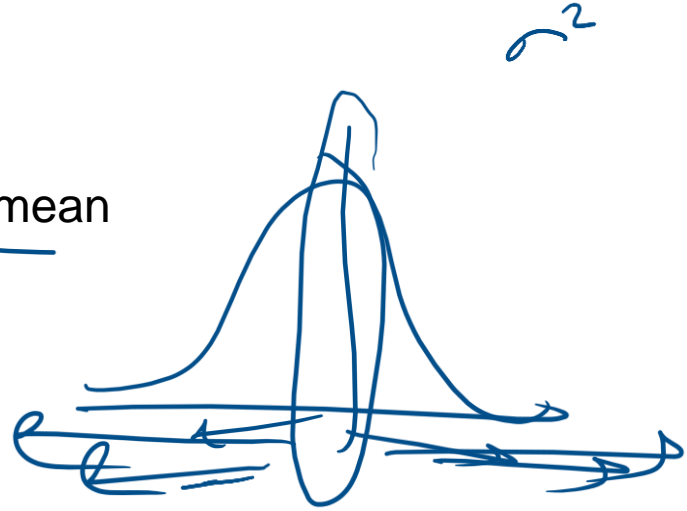
C) 50

D) 70

$$\underline{\underline{IQR = Q_3 - Q_1}}$$

## What does a high variance indicate about the data?

- ☒ A) The data points are close to the mean
- ☒ B) The data points are spread out from the mean
- C) The data set has no outliers
- D) The mean is higher than the median



$V = \text{high}$   
SD  $\rightarrow$  large

## How does systematic sampling differ from random sampling?

- ✓ A) It selects samples at regular intervals → rule
- B) It selects samples using a random number generator
- C) It only selects the first available samples
- D) It is less efficient than convenience sampling

What type of bias occurs when only successful outcomes are considered, ignoring failures?

A) Exclusion bias

B) Survivorship bias

C) Selection bias

D) Recall bias



world war II aircraft

Survivor

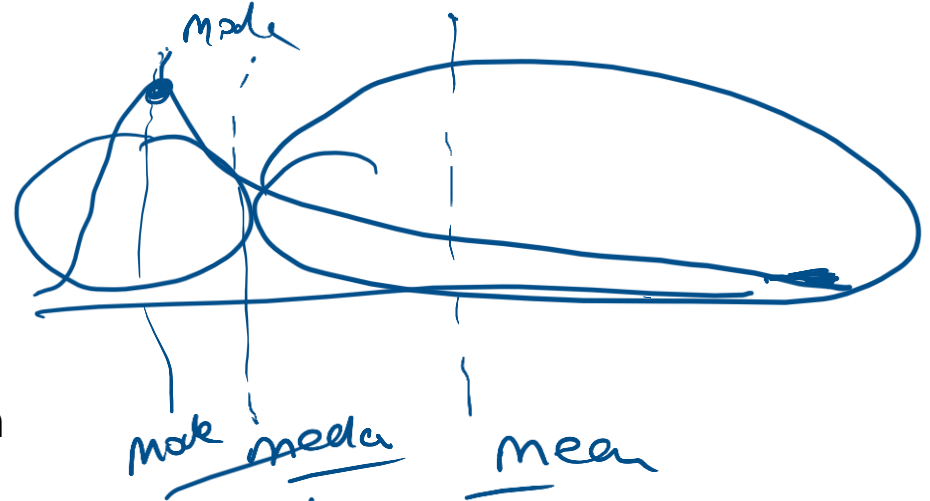
**A dataset has a median of 30 and a mean of 40. What does this suggest?**

☒ A) The data is skewed left

☒ B) The data is skewed right

☒ C) The data is symmetric

☐ D) The mode is greater than the median



Symmetric  $\rightarrow$  mean = median = mode

right skewed  $\rightarrow$  mean > median > mode

left skewed  $\rightarrow$  mean < median < mode

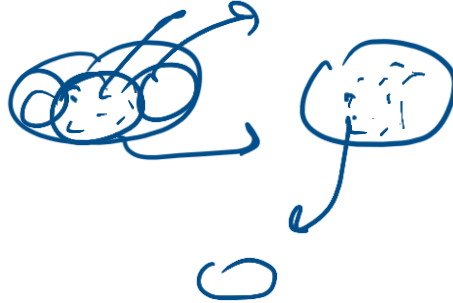
**A survey conducted only among social media users leads to which type of bias?**

A) Selection bias

☒ B) Undercoverage bias

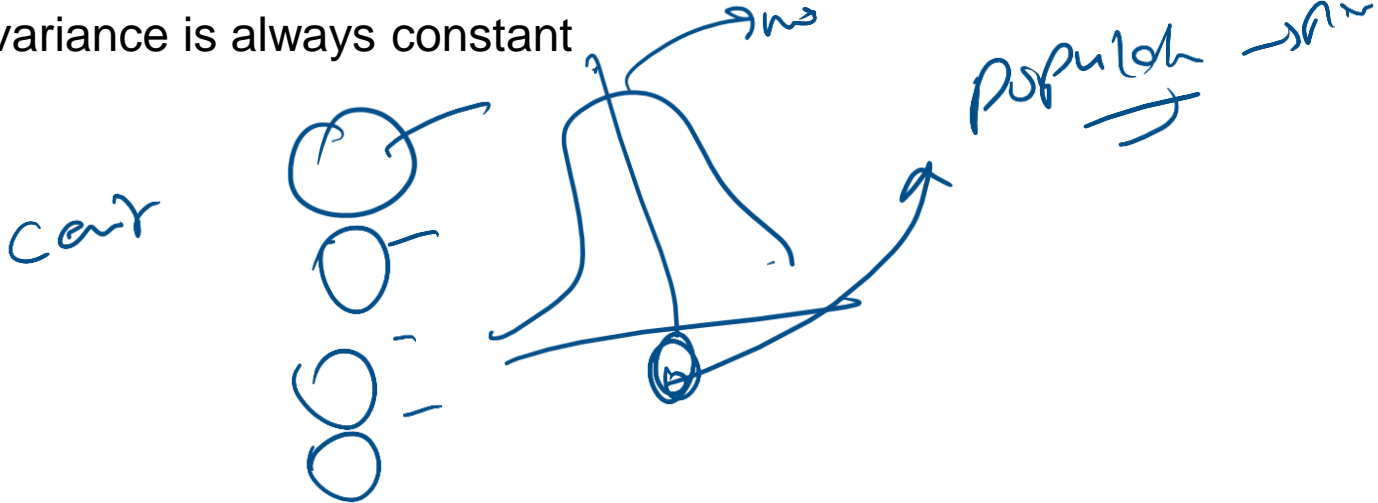
C) Measurement bias

D) Non-response bias



## How does the Central Limit Theorem help in real-world scenarios?

- A) It ensures data is normally distributed
- ✓ B) It allows us to use sample statistics to estimate population parameters
- C) It guarantees equal probability for all sample points
- D) It proves variance is always constant





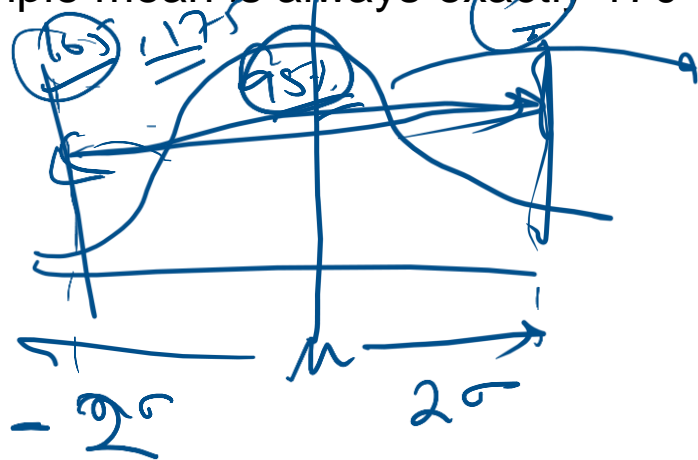
**A 95% confidence interval for the mean height of a group of students is (165 cm, 175 cm). What does this mean?**

A) The true mean height is definitely in this range

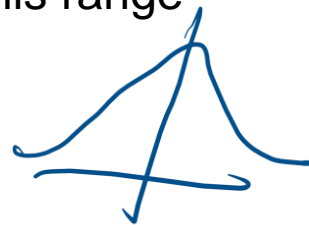
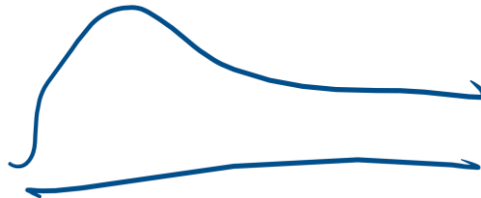
B) There is a 95% chance that the sample mean is between 165 and 175 cm

C) There is a 95% chance that the population mean falls within this range

D) The sample mean is always exactly 170 cm



165, 175



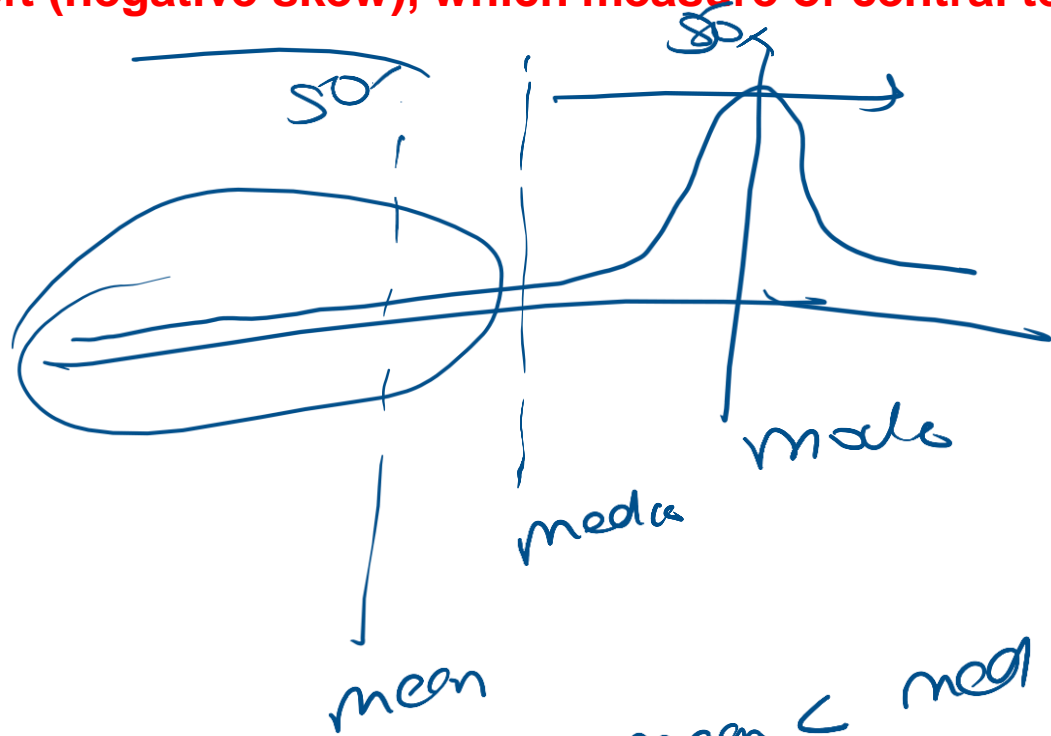
If a dataset is skewed left (negative skew), which measure of central tendency is the largest?

A) Mean ☒

B) Median ☒

☒ C) Mode

D) They are all equal



$mean < median < mode$

**If you take multiple random samples from a population, the means of these samples will form a distribution that follows:**

A) A binomial distribution

B) A uniform distribution

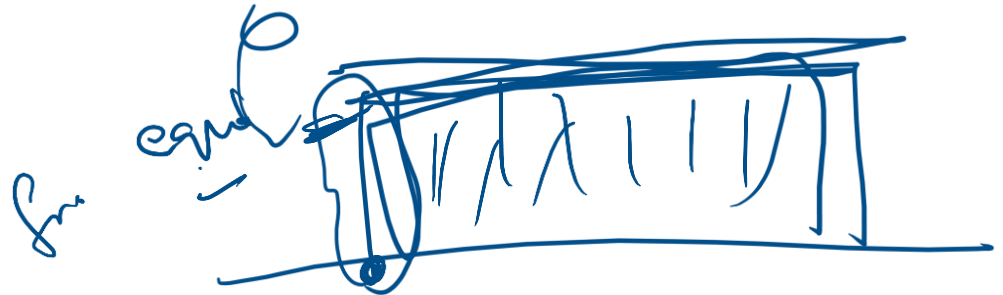
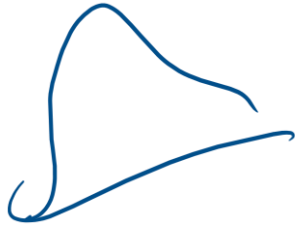
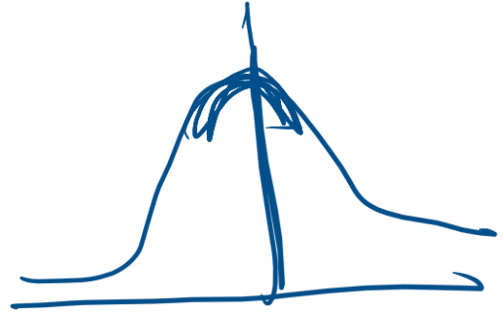
☒ C) A normal distribution

☒ D) A skewed distribution

CNT

MC92

6



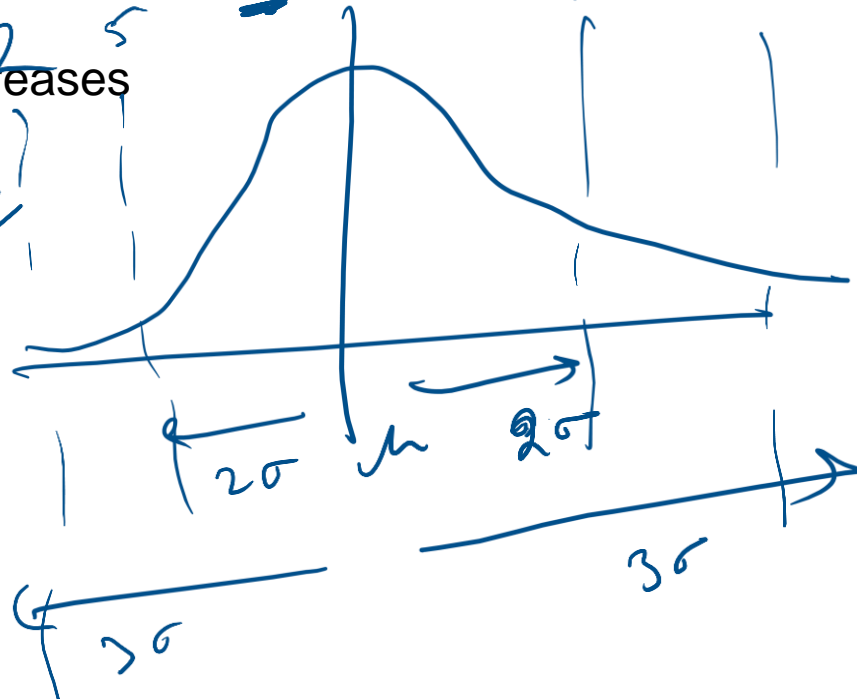
**What is the impact of increasing the confidence level from 95% to 99%?**

A) The confidence interval becomes wider

B) The confidence interval becomes narrower

C) The standard error decreases

D) The mean increases



$F = S_{10}$

$(2, 12)$

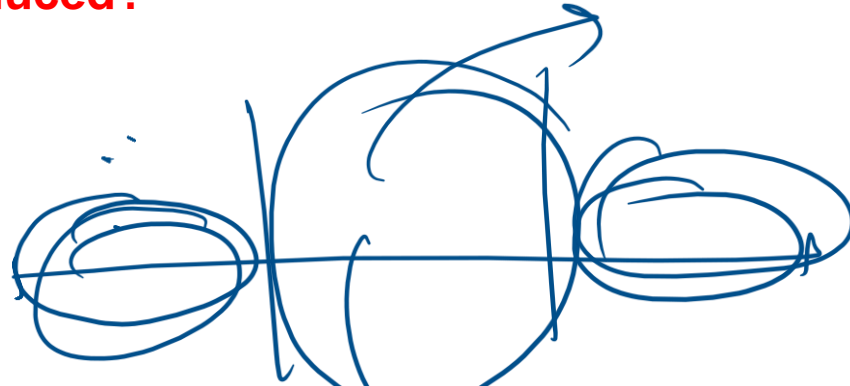
**If a researcher selects only extreme values while ignoring moderate values, what type of bias is introduced?**

A) Exclusion bias

B) Selection bias

C) Outlier bias

D) Measurement bias



left out selection

