$\beta_1 = \frac{\mu_3^2}{\mu_2^3} > 0$

3A, ABM Measures of Central Tendency & Dispersion, Skewness, Kurtosis

Q11. According to Pearson's measure of skewness, in a positively skewed distribution where $\beta_1 > 0$, which of the following orderings of mean (μ), median, and mode holds true?

a. Mode > Median > Mean

b. Mean > Median > Mode

c. Median > Mean > Mode

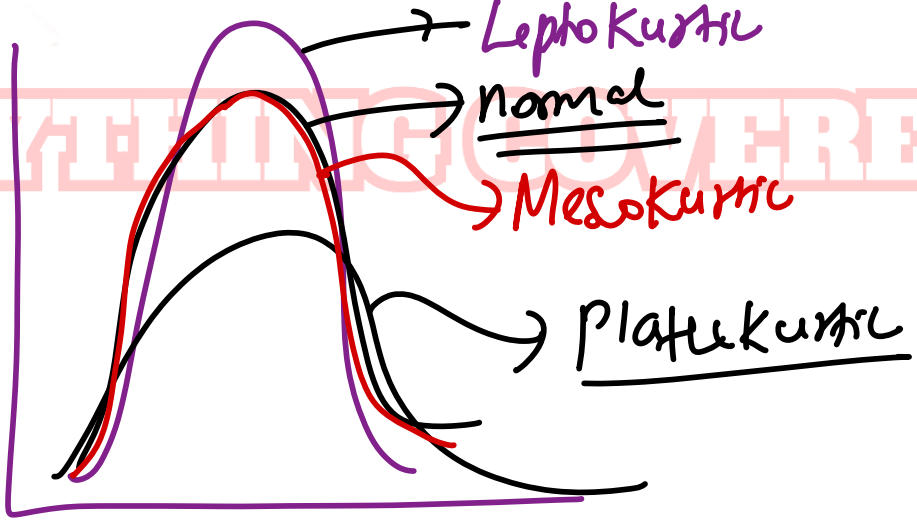
d. Mode > Mean > Median

$\beta_1 = 0$ Symmetrical distribution

$\beta_1 > 0$ +ve skewed

$\beta_1 < 0$ -ve skewed

Kurtosis



$P(A \cap B) = P(A) \times P(B/A)$
$= P(B) \times P(A/B)$

6A ABMTheory of Probability

Q12. Which theorem in probability theory states that the joint probability of two events A and B is given by P(A ∩ B) = P(A) × P(B|A)?

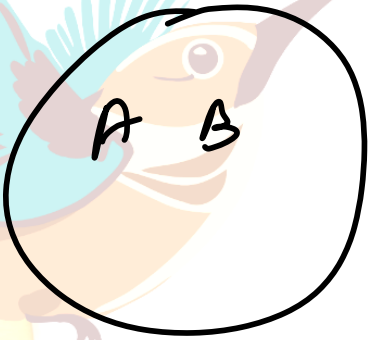
Conditional

a. Addition theorem

b. Bayes' theorem

Ans. c. Multiplication theorem

$P(B/A) \Rightarrow$



d. Central limit theorem

$P(A \cap B)$

$P(A) \times P(B/A)$

4A Correlation and Regression

Q13. Why data used for computing correlation coefficients must be homogeneous before applying the measure?

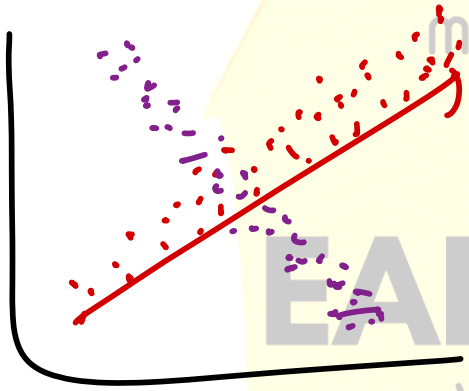
A. To ensure regression residuals are zero

B. Because non-homogeneous data can artificially inflate or deflate r through extra-group variation

Answer

C. To guarantee the relationship is curvilinear

D. Because heterogeneous data leads to singular matrices



Direct Relationship

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$r = \frac{\frac{\sum xy}{N} - \bar{x}\cdot\bar{y}}{\sqrt{\frac{\sum x^2}{N} - (\bar{x})^2}\sqrt{\frac{\sum y^2}{N} - (\bar{y})^2}}$$

Cov

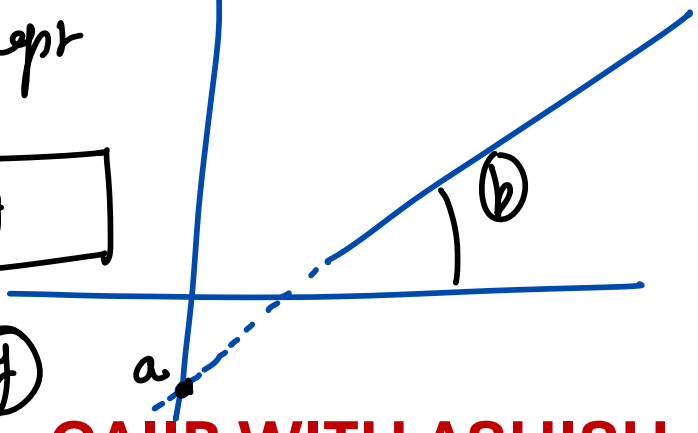
Regression

$y = a + bx$ → slope of line

y-intercept

σ_x σ_y

x	y	x^2	y^2	xy
Σ	Σ	Σ		



$$a = \bar{y} - b\bar{x}$$

$$b = \frac{cov(x,y)}{\sigma_x^2}$$

3A, ABM Measures of Central Tendency & Dispersion, Skewness, Kurtosis

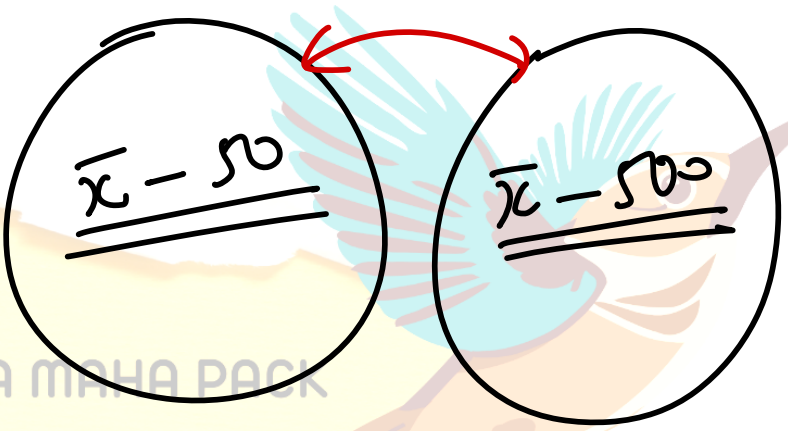
Q14. When comparing variability across two datasets with different means—one set of exam scores around 50 and another around 500—which relative measure best standardizes dispersion for meaningful comparison?

A. Quartile deviation

B. Mean deviation

Ans C. Coefficient of variation

D. Range



$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sigma$$

$$= \frac{\sigma}{\bar{x}}$$

Range

⇒ max − min

Co-efficient of Range

$$= \frac{Max - Min}{Max + Min}$$

6A ABMTheory of Probability

Q15. In formally defining a dice roll experiment, the notation S = {1,2,3,4,5,6} is used. What does this set S represent in probability theory?

A. Event

B. Sample space

C. Random variable

D. Outcome space

Ans.

$4! = 4 \times 3 \times 2 \times 1$

$^5C_2 = \frac{5!}{2!\,3!}$

$^nC_r = \frac{n!}{r!\,(n-r)!}$

$\frac{5 \times 4 \times 3!}{2 \times 1 \times 3!} = 10$

Multiplication

$P(A) = \frac{5}{8}$.

$P(B)$ = Second Ball is Red provided

$P(B/A) = \frac{4}{7}$

R → 5 → 4

T → 8 → 7

$P(A \cap B) = P(A) \times P(B/A) = \frac{5}{8_2} \times \frac{4^1}{7} = \frac{5}{14}$

2A abm sampling distribution-2

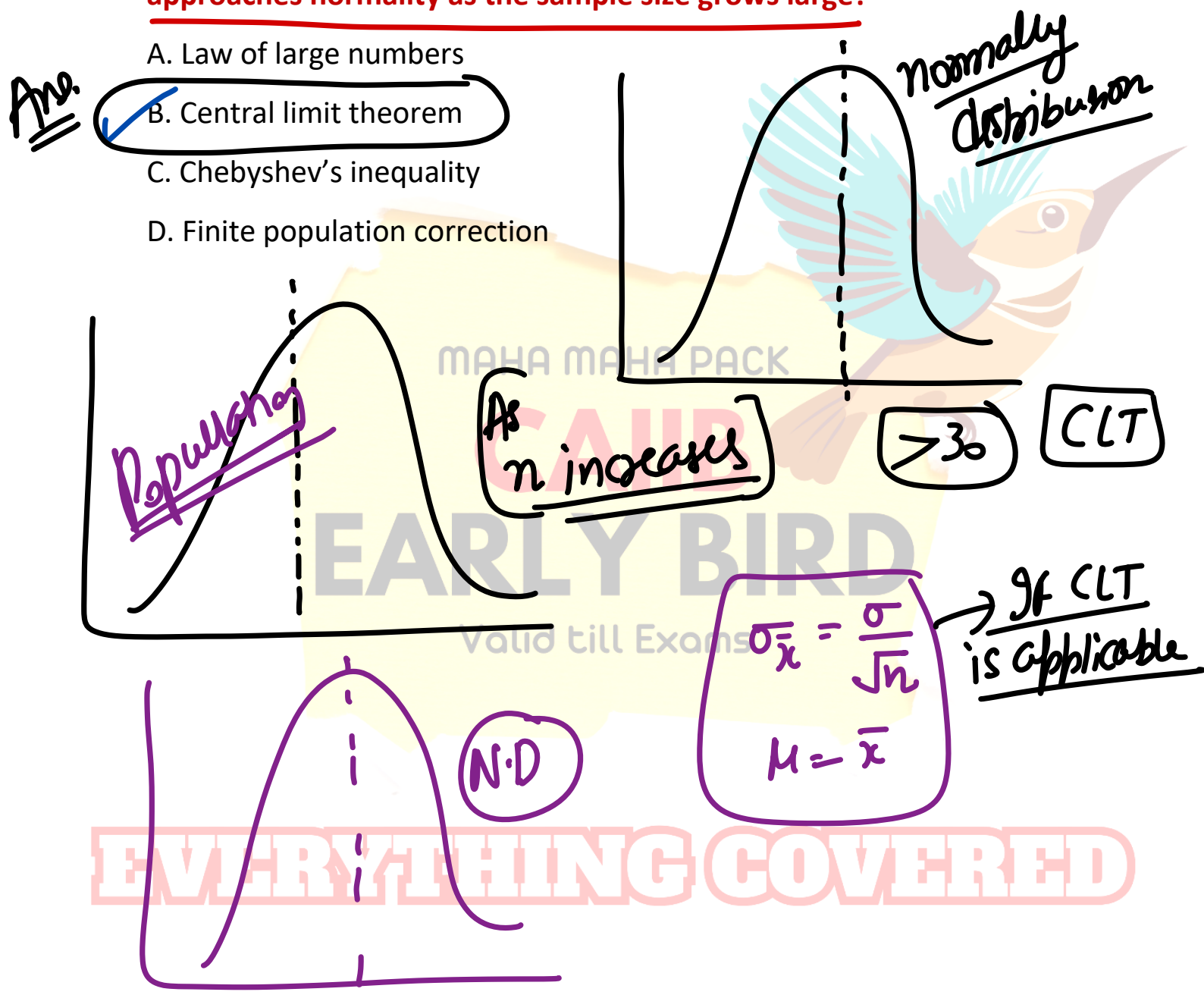
Q16. Which principle ensures that, regardless of a population's original distribution shape, the distribution of sample means approaches normality as the sample size grows large?

A. Law of large numbers

Ans. ✓ B. Central limit theorem

C. Chebyshev's inequality

D. Finite population correction



7A estimation new

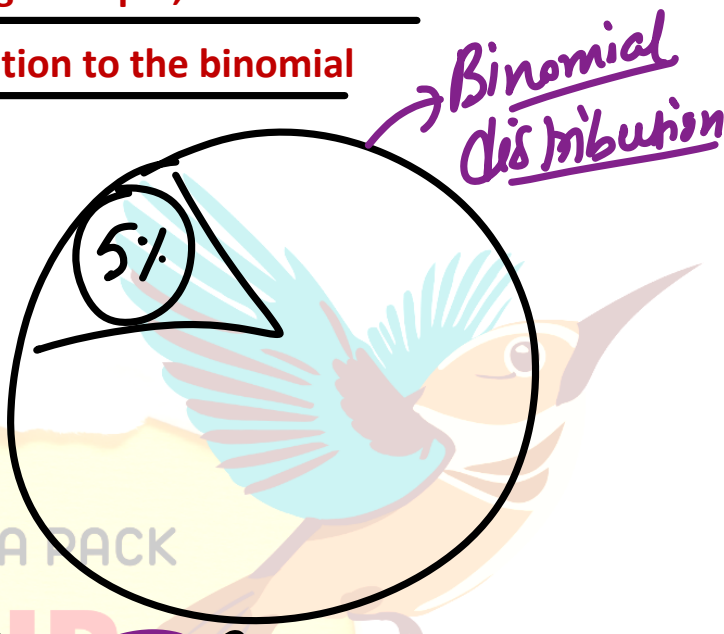
Q17. When constructing an approximate confidence interval for a true population proportion from a large sample, what rule of thumb ensures the normal approximation to the binomial distribution is appropriate?

A. np ≥ 1 and nq ≥ 1

B. np ≥ 5 and nq ≥ 5 Ans.

C. Sample size n > 10

D. p and q both less than 0.5



$q = (1-p)$

$\mu = \frac{np}{n} = p$ proportion

$\mu_{\bar{p}} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$

$\begin{cases} \mu = np \\ s.d = \sqrt{npq} \end{cases}$

n must be large

$np \& nq \geq 5$

1A, ABM DEFINITION OF STATISTICS, IMPORTANCE & LIMITATIONS

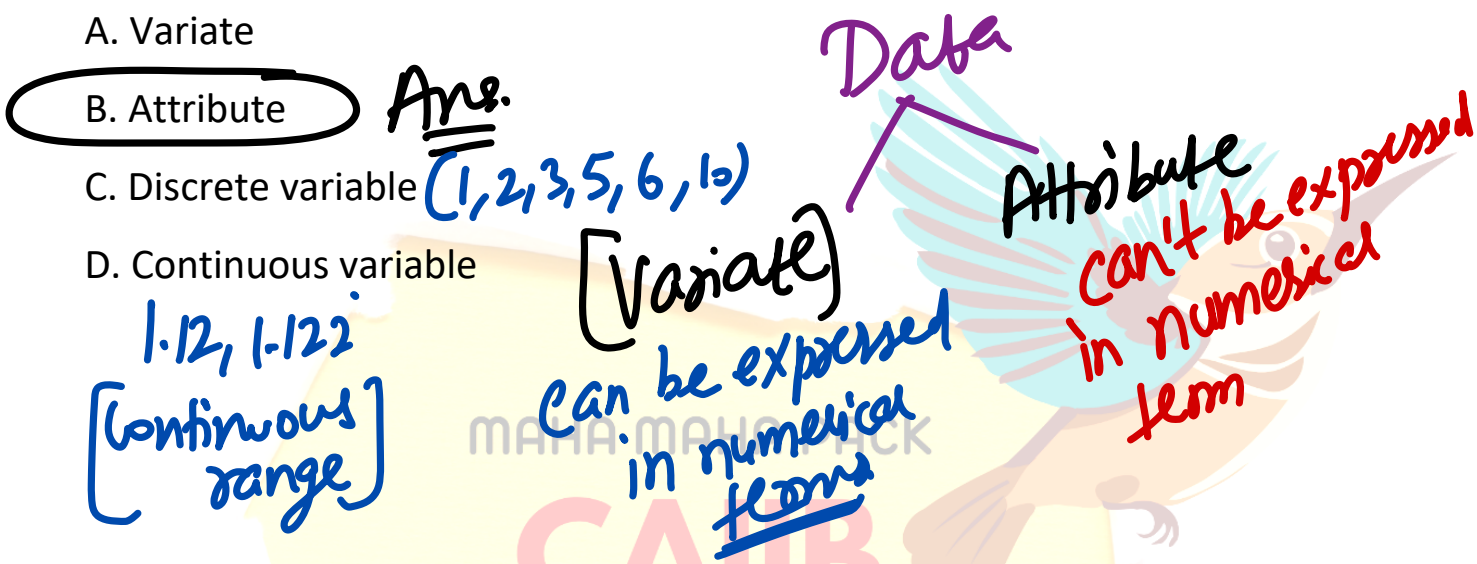
Q18. In census data, when recording responses for "religion" among individuals, the characteristic varies but cannot be quantified numerically. Such a variable is classified as a(n):

A. Variate

B. Attribute

C. Discrete variable

D. Continuous variable



5A Time Series

Q19. If a country's GDP shows ups and downs associated with business cycles that vary in amplitude and duration without a fixed pattern, which type of variation is being described?

A. Secular trend like inflation

Ans. B. Cyclical fluctuation over multiple years

C. Seasonal variation within 12 months.

D. Irregular variation Covid, War etc.

2A abm sampling distribution-2

Q20. When sampling without replacement from a finite population, under what condition can the finite population correction factor be ignored?

a. When n/N > 0.5

b. When n/N < 0.05 Ans

c. When n = N

d. Always

$\sqrt{\left(\frac{N-n}{N-1}\right)}$ → finite population multiplier

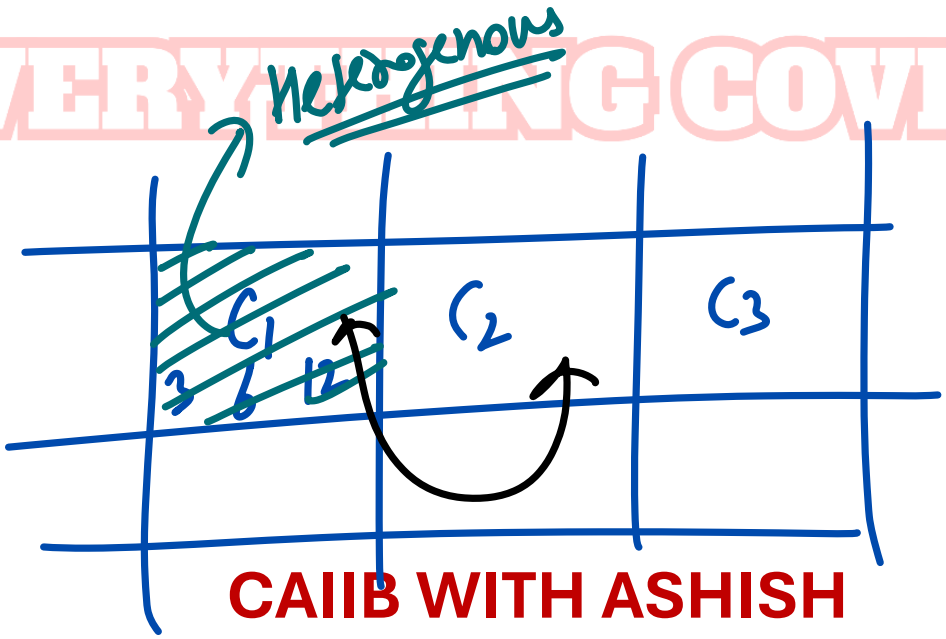
$n/N < 0.05$

5 papers 1999

Mega Marathon 9.30 – 11.30pm

{Numerical case studies imp PYQs}

{Caiib1999 8360944207}



ZA abm sampling distribution-2

21. In designing a survey of customer satisfaction, a researcher considers two approaches: dividing the customer population into income-based strata and then randomly sampling within each stratum (stratified sampling), versus grouping customers by geographic clusters (city zones) and randomly selecting entire clusters for inclusion (cluster sampling). Which option best describes the impact on variance and operational complexity between these two designs?

A. Stratified sampling tends to increase variance but reduce cost, while cluster sampling decreases variance at higher cost.

B. Stratified sampling generally reduces variance by ensuring representation across strata at the expense of higher planning complexity, whereas cluster sampling often increases variance due to intra-cluster homogeneity but can lower field costs by concentrating data collection geographically.

Ans

C. Stratified sampling yields unbiased estimates only if cluster sizes are equal, whereas cluster sampling yields unbiased estimates regardless of cluster size.

D. Both methods yield the same variance properties and cost implications if strata and clusters have equal unit counts.