

# STATISTICS

## MCQ

ANSWER 1= a) True

ANSWER 2=a) Central Limit Theorem

ANSWER 3=b) Modelling bounded count data

ANSWER 4=c) The square of a standard normal random variable follows what is called  
chi-squared distribution

ANSWER 5=c) Poisson

ANSWER 6=b) False

ANSWER 7=b) Hypothesis

ANSWER 8=a) 0

ANSWER 9=c) Outliers cannot conform to the regression relationship

## SUBJECTIVE ANSWER

ANSWER 10=A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the *mean* of the distribution. The normal distribution is also known as a *Gaussian distribution*. It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean.

ANSWER 11=Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

Mean imputation: Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.

ANSWER 12= A/B testing is a statistical approach to compare two versions of something, often used in marketing or research. Groups A and B experience different versions, and their outcomes are compared. Statistical analysis helps determine if differences are significant, aiding decisions on which version is more effective.

ANSWER 13. Mean imputation is straightforward but risky. It assumes randomness in missing data and can introduce bias, affecting results. It's acceptable when data is missing at random, but advanced methods like

regression imputation or multiple imputation are more reliable for capturing data relationships.

ANSWER 14. Linear regression is a statistical technique for modelling relationships between a dependent variable and one or more independent variables. It assumes a linear connection, aiming to find the best-fitting line that minimises the difference between observed and predicted values. It's used for prediction, understanding variable impact, and identifying trends.

ANSWER 15. Statistics encompasses several branches. Descriptive statistics summarise data. Inferential statistics draws conclusions from samples. Probability deals with uncertainty. Biostatistics analysis biological data. Econometrics applied statistics to economics. Social statistics studies social trends. Multivariate stats handles multiple variables. Time series analysis examines sequential data. Experimental design plans controlled studies. Statistical machine learning merges statistics with machine learning for predictive modelling, each branch serving distinct analytical goals