1. Which of the following is the correct formula for total variation?
a) Total Variation = Residual Variation – Regression Variation
b) Total Variation = Residual Variation + Regression Variation
c) Total Variation = Residual Variation * Regression Variation
d) All of the mentioned

Answer: b) Total Variation = Residual Variation + Regression Variation

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

2. Collection of exchangeable binary outcomes for the same covariate data are called _____ outcomes.
a) random
b) direct
c) binomial
d) none of the mentioned

Answer: c) Binomial

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

3. How many outcomes are possible with Bernoulli trial?
a) 2
b) 3
c) 4
d) None of the mentioned

Answer: a) 2

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

4. If Ho is true and we reject it is called
a) Type-I error
b) Type-II error
c) Standard error
d) Sampling error

Answqer: a) Type-I error

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

5. Level of significance is also called:
a) Power of the test
b) Size of the test
c) Level of confidence
d) Confidence coefficient

Answer:

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

6. The chance of rejecting a true hypothesis decreases when sample size is:
a) Decrease
b) Increase
c) Both of them
d) None

Answer: b) Increase

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

7. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

Answer: b) Hypothesis

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

8. What is the purpose of multiple testing in statistical inference?
a) Minimize errors
b) Minimize false positives
c) Minimize false negatives
d) All of the mentioned

Answer: d) All of the mentioned

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

9. Normalized data are centred at _____ and have units equal to standard deviations of the original data

a) 0

b) 5

c) 1

d) 10

Answer: a) 0

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What Is Bayes' Theorem?

11. What is z-score?

12. What is t-test?

13. What is percentile?

14. What is ANOVA?

15. How can ANOVA help?

Answer 10: Bayes' Theorem is a mathematical formula for determining conditional probability which is named after British mathematician Thomas Bayes. Basically, it describes the probability of an event based on prior knowledge of the condition that might relevant to the event.

Beside Statistics, Baye's theorem is also used in various fields like pharmacology, finance.

Formula for Bayes' Theorem

$P(A|B) = P(A \cap B)/ P(B) = P(A) \cdot P(B|A)/ P(B)$

where:

$P(A)$= The probability of A occurring

$P(B)$= The probability of B occurring

$P(A|B)$=The probability of A given B

$P(B|A)$= The probability of B given A

$P(A \cap B)$= The probability of both A and B occurring

Note: events A and B are independent events means (i.e the probability of the outcome of event A doesn't depend on the probability of the outcome of event B)

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

11. What is z-score?

Answer 11 : z-score gives you an idea of how far a data point is away from the mean of data. More technically, it's a measure of how many standard deviation below or above from the mean.

A z-score can be placed on a normal distribution curve where z-score range from -3 standard deviaction (left side from the mean) to +3 standard deviation (right side of the mean). In order to calculate z-score, you need to know mean and standard deviations.

Z-score is a way to compare results to a normal population. Like z-score can tell you where that person weight is compared to the average populations mean weight.

Basic z-score formula is: $z = (x - \mu) / \sigma$

x = data point of data
μ = mean of data
σ = Standard deviation
z = how many standard deviation the data point is from the mean

A z-score of 1 is 1 standard deviation above the mean

A score of -1 is 1 standard deviation below the mean

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

12. What is t-test?

Answer 12: A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process has an effect on the population of interest or whether two groups are different from one another.

You can test the difference between 2 groups using a t-test and null and alternate hypotheses.

. The null hypothesis (H_0) is that the true difference between group means is zero.

. The alternate hypothesis (H_a) is that the true difference is different from zero.

A t-test can only be used when comparing the means of 2 groups otherwise we you have more that 2 groups to compare use an ANOVA test.

The t-test assumes your data:

Are independent

Are normally distributed

Have a similar amount of variance within each group being compared

Few t-test types:

1. If the groups come from a single population (e.g measue before and after treatment), perform a paired t-test
2. If the groups come from two different populations (2 diff species, or peoples from 2 separate cities) perform a two-sample t-test
3. If there is one group being compared against a standard value (comparing the acidity of a liquid to a neutral pH of 7) perform a one-sample t-test

1. If you only care whether the two population are different from one-another, perform a 2-tailed t-test
2. If you want to know whether one population mean is greater that or less than the other, perform one-tailed t-test

The formula for the two-sample t-test is shown below.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s^2(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

When reporting your t-test results, the most important values to include are the *t*-**value**, the *p*-**value**, and the **degrees of freedom** for the test.

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

13. What is percentile?

Answer: A Percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example the 20th percentile means 20% results are below this.

In layman language, lets suppose 100 people sits on exam and you got 95 percentile that means you are on top 5 and 95% population are below you.

The 25th percentile is also known as the first quartile (Q1), the 50th percentile as the median or second quartile (Q2), and the 75th percentile as the third quartile (Q3). The range of values containing the central half of the observations is called the interquartile range: that is, the range between the 25th and 75th percentiles.

Interquartile range is used to handle outliers in dataset.

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

14. What is ANOVA?

Answer: An Anova is a way to find out if survey or experiment results are significant. In other words, it helps to decide if you need to reject the null hypothesis or accept the alternate hypothesis.

The Formula for ANOVA is:

F = MST\ MSE

Where:

F = ANOVA coefficient

MST= Mean sum of squares due to treatment

MSE = Mean sum of squares due to error

The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

15. How can ANOVA help?

Basically, you are testing groups to see if there's a difference between them like the way

1. A group of psychiatric patients are trying 3 different therapies: counselling, medication and biofeedback. You want to see if one therapy is better than the others
2. A manufacturer have 2 different process to make light bulbs. They want to know if one process is better than the other.
3. Students from different colleges take the same exam. You want to see if 1 collegeoutperforms the other.