

STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

1. What is central limit theorem and why is it important?

Answer: The Central Limit Theorem is important for statistics because it allows us to safely assume that the sampling distribution of the mean will be normal in most cases. This means that we can take advantage of statistical techniques that assume a normal distribution.

2. What is sampling? How many sampling methods do you know?

Answer: Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

There are two types of sampling methods:

- Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group.
- Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

3. What is the difference between type I and type II error?

Answer: Type I error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true. Type II error is the error that occurs when the null hypothesis is accepted when it is not true. Type I error is equivalent to false positive.

4. What do you understand by the term Normal distribution?

Answer: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

5. What is correlation and covariance in statistics?

Answer: Covariance is a measure of how much two random variables vary together. Correlation is a statistical measure that indicates how strongly two variables are related.

6. Differentiate between Univariate, Bivariate, and multivariate analysis.

Answer: Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.

7. What do you understand by sensitivity and how would you calculate it?

Answer: A sensitivity analysis determines how different values of an independent variable affect a particular dependent variable under a given set of assumptions. In other words, sensitivity analyses study how various sources of uncertainty in a mathematical model contribute to the model's overall uncertainty. The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

8. What is hypothesis testing? What is H_0 and H_1 ? What is H_0 and H_1 for two-tail test?

Answer: A statistical hypothesis is an assertion or conjecture concerning one or more populations. Hypothesis testing is formulated in terms of two hypotheses: H_0 : the null hypothesis; H_1 : the alternate hypothesis. Alternative Hypothesis: H_1 : The hypothesis that we are interested in proving. Null hypothesis: H_0 : The complement of the alternative hypothesis. ... This is the probability of falsely rejecting the null hypothesis. Type II error: do not reject the null hypothesis when it is wrong. Null hypothesis (H_0): The null hypothesis here is what currently stated to be true about the population. ... Alternate hypothesis (H_1): The alternate hypothesis is always what is being claimed. "In our case, Tedd believes (Claims) that the actual value has changed".

9. What is quantitative data and qualitative data?

Answer: Quantitative data is information about quantities, and therefore numbers, and qualitative data is descriptive, and regards phenomenon which can be observed but not measured, such as language.

10. How to calculate range and interquartile range?

Answer: To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

11. What do you understand by bell curve distribution ?

Answer: The term "bell curve" is used to describe a graphical depiction of a normal probability distribution, whose underlying standard deviations from the mean create the curved bell shape. A standard deviation is a measurement used to quantify the variability of data dispersion, in a set of given values around the mean.

12. Mention one method to find outliers.

Answer: The most effective way to find all of your outliers is by using the interquartile range (IQR). The IQR contains the middle bulk of your data, so outliers can be easily found once you know the IQR. Add $1.5 \times (\text{IQR})$ to the third quartile. Any number greater than this is a suspected outlier. Subtract $1.5 \times (\text{IQR})$ from the first quartile. Any number less than this is a suspected outlier.

13. What is p-value in hypothesis testing?

Answer: The p-value, or probability value, tells you how likely it is that your data could have occurred under the null hypothesis. ... The p-value is a proportion: if your p-value is 0.05, that means that 5% of the time you would see a test statistic at least as extreme as the one you found if the null hypothesis was true.

14. What is the Binomial Probability Formula?

Answer: Binomial probability refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment). If the probability of success on an individual trial is p , then the binomial probability is $nCx \cdot p^x \cdot (1-p)^{n-x}$.

15. Explain ANOVA and its applications.

Answer: Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples. Another measure to compare the samples is called a t-test.

ANOVA models can realistically be used in numerous industries and applications:

1. Comparing the gas mileage of different vehicles, or the same vehicle under different fuel types, or road types.
2. Understanding the impact of temperature, pressure or chemical concentration on some chemical reaction (power reactors, chemical plants, etc)
3. Understanding the impact of different catalysts on chemical reaction rates
4. Studying whether advertisements of different kinds solicit different numbers of customer responses
5. Understanding the performance, quality or speed of manufacturing processes based on number of cells or steps they're divided into