STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

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

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

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

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

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

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

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

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

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?

10. Answer: Bayes' Theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probability. Conditional probability is the likelihood of an outcome occurring, based on a previous outcome having occurred in similar circumstances. Bayes' theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence. In finance, Bayes' Theorem can be used to rate the risk of lending money to potential borrowers. The theorem is also called Bayes' Rule or Bayes' Law and is the foundation of the field of Bayesian statistics.

11. Answer: Simply put, a z-score (also called a standard score) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

Simply put, a z-score (also called a standard score) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

A z-score can be placed on a normal distribution curve. Z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). In order to use a z-score, you need to know the mean μ and also the population standard deviation σ.

12. Answer: The t test tells you how significant the differences between group means are. It lets you know if those differences in means could have happened by chance. The t test is usually used when data sets follow a normal distribution but you don't know the population variance.

For example, you might flip a coin 1,000 times and find the number of heads follows a normal distribution for all trials. So you can calculate the sample variance from this data, but the population variance is unknown. Or, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). So while the control group may show an average life expectancy of +5 years, the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population.

The t test tells you how significant the differences between group means are. It lets you know if those differences in means could have happened by chance. The t test is usually used when data sets follow a normal distribution but you don't know the population variance.

For example, you might flip a coin 1,000 times and find the number of heads follows a normal distribution for all trials. So you can calculate the sample variance from this data, but the population variance is unknown. Or, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). So while the control group may show an average life expectancy of +5 years, the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population.

13. Answer: “Percentile” is in everyday use, but there is no universal definition for it. The most common definition of a percentile is a number where a certain percentage of scores fall below that number. You might know that you scored 67 out of 90 on a test. But that figure has no real meaning unless you know what percentile you fall into. If you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test.

Percentiles are commonly used to report scores in tests, like the SAT, GRE and LSAT. for example, the 70th percentile on the 2013 GRE was 156. That means if you scored 156 on the exam, your score was better than 70 percent of test takers.

The 25th percentile is also called the first quartile.

The 50th percentile is generally the median

The 75th percentile is also called the third quartile.

The difference between the third and first quartiles is the interquartile range.

14. Answer: An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

Basically, you’re testing groups to see if there’s a difference between them. Examples of when you might want to test different groups:

A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.

A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.

Students from different colleges take the same exam. You want to see if one college outperforms the other.

15. Answer: ANOVA help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.