**STATISTICS– WORKSHEET 6 ANSWERSHEET**

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

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial
3. a) 2
4. (a) Type-I error
5. (b) Size of the test
6. (b) Increases
7. (b) Hypothesis
8. (d) All of the mentioned
9. (a) 0

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

1. **Bayes' Theorem** is a way of finding a probability when we know certain other probabilities.

The formula is:

P(A|B) = P(A) P(B|A) P(B)

Which tells us: how often A happens given that B happens, written P(A|B),

When we know: how often B happens given that A happens, written P(B|A)

and how likely A is on its own, written P(A)

and how likely B is on its own, written P(B)

1. 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.

Z-scores are a way to compare results to a “normal” population. Results from tests or surveys have thousands of possible results and units; those results can often seem meaningless. For example, knowing that someone’s weight is 150 pounds might be good information, but if you want to compare it to the “average” person’s weight, looking at a vast table of data can be overwhelming (especially if some weights are recorded in kilograms). A z-score can tell you where that person’s weight is compared to the average population’s mean weight.

1. A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. It is mostly used when the data sets, like the data set recorded as the outcome from flipping a coin 100 times, would follow a normal distribution and may have unknown variances. A t-test is used as a hypothesis testing tool, which allows testing of an assumption applicable to a population.
2. A percentile is a term used in statistics to express how a score compares to other scores in the same set. While there is technically no standard definition of percentile, it's typically communicated as the percentage of values that fall below a particular value in a set of data scores.

Percentiles are commonly used to report values from norm-referenced tests (in which the average is determined by comparing a set of results in the same group) as the percentages of scores that fall below those of the average of the set. For example, a male child age 12 with a weight of 130 pounds is at the 90th percentile of weight for males of that age, which indicates that he weighs more than 90 percent of other 12-year-old boys.

1. Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

The Formula for ANOVA is:

F= MSE/MST

where:

F=ANOVA coefficient

MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error

​

1. It’s help to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.