FLIP ROBO

Statistics– WORKSHEET 6

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
    ANS B

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
    ANS C

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

4. If Ho is true and we reject it, then it is called:
    (a) Type-I error
    (b) Type-II error
    (c) Standard error
    (d) Sampling error
    ANS A

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

6. The chance of rejecting a true hypothesis decreases when sample size:
    (a) Decreases
    (b) Increases
    (c) Both of them
    (d) None of them
    ANS A

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

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   ANS D

9. Normalized data is centered at_____and has unit equal to standard deviations of the original data.
    (a) 0   (b) 5
    (c) 1   (d) 10
    ANS A

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

10 What Is Bayes' Theorem?

Bayes' theorem is a way to figure out conditional probability. Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events. For example, your probability of getting a parking space is connected to the time of day you park, where you park, and what conventions are going on at any time. Bayes' theorem is slightly more nuanced. In a nutshell, it gives you the actual probability of an event given information about tests.

"Events" Are different from "tests." For example, there is a test for liver disease, but that's separate from the event of actually having liver disease.
Tests are flawed: just because you have a positive test does not mean you actually have the disease. Many tests have a high false positive rate. Rare events tend to have higher false positive rates than more common events. We're not just talking about medical tests here. For example, spam filtering can have high false positive rates. Bayes' theorem takes the test results and calculates your real probability that the test has identified the event.

Bayes' Theorem (also known as Bayes' rule) is a deceptively simple formula used to calculate conditional probability. The Theorem was named after English mathematician Thomas Bayes (1701-1761). The formal definition for the rule is:
bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In most cases, you can't just plug numbers into an equation; You have to figure out what your "tests" and "events" are first. For two events, A and B, Bayes' theorem allows you to figure out p(A|B) (the probability that event A happened, given that test B was positive) from p(B|A) (the probability that test B happened, given that event A happened). It can be a little tricky to wrap your head around as technically you're working backwards; you may have to switch your tests and events around, which can get confusing. An example should clarify what I mean by "switch the tests and events around."

11 What is z-score?

Z scores (Z value) is the number of standard deviations a score or a value (x) away from the mean. In other words, Z-score measures the dispersion of data. Technically, Z-score tells a value (x) is how many standard deviations below or above the population mean (µ). If the Z value is positive, it indicates that the value or score (x) is above the mean. Similarly, if Z value is negative, it means the value (x) is below the mean.

12 **What is t-test?**

The t score is a ratio between the difference between two groups and the difference within the groups. The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different from each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.

A large t-score tells you that the groups are different.
A small t-score tells you that the groups are similar.

T-Values and P-values

How big is "big enough"? Every t-value has a p-value to go with it. A p-value is the probability that the results from your sample data occurred by chance. P-values are from 0% to 100%. They are usually written as a decimal. For example, a p value of 5% is 0.05. Low p-values are good; They indicate your data did not occur by chance. For example, a p-value of .01 means there is only a 1% probability that the results from an experiment happened by chance. In most cases, a p-value of 0.05 (5%) is accepted to mean the data is valid.

Calculating the Statistic / Test Types

There are three main types of t-test:

An Independent Samples t-test compares the means for two groups.
A Paired sample t-test compares means from the same group at different times (say, one year apart).
A One sample t-test tests the mean of a single group against a known mean.
You probably don't want to calculate the test by hand (the math can get very messy, but if you insist you can find the steps for an independent samples t test here.

Use the following tools to calculate the t test:
How to do a T test in Excel.
T test in SPSS.
T distribution on the TI 89.
T distribution on the TI 83.

13 What is a percentile?

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.

14 What is ANOVA?

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.

What Does "One-Way" or "Two-Way Mean?
One-way or two-way refers to the number of independent variables (IVs) in your Analysis of Variance test.

One-way has one independent variable (with 2 levels). For example: brand of cereal,
Two-way has two independent variables (it can have multiple levels). For example: brand of cereal, calories.

15 How can ANOVA help?

ANOVA tests in statistics packages are run on parametric data. If you have rank or ordered data, you'll want to run a non-parametric ANOVA (usually found under a different heading in the software, like "nonparametric tests").
Steps
It is unlikely you'll want to do this test by hand, but if you must, these are the steps you'll want to take:

Find the mean for each of the groups.
Find the overall mean (the mean of the groups combined).
Find the Within Group Variation; the total deviation of each member's score from the Group Mean.
Find the Between Group Variation: the deviation of each Group Mean from the Overall Mean.
Find the F statistic: the ratio of Between Group Variation to Within Group Variation.
ANOVA vs. T Test
A Student's t-test will tell you if there is a significant variation between groups. A t-test compares means, while the ANOVA compares variances between populations.
You could technically perform a series of t-tests on your data. However, as the groups grow in number, you may end up with a lot of pair comparisons that you need to run. ANOVA will give you a single number (the f-statistic) and one p-value to help you support or reject the null hypothesis.