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?**
   Ans- > b) Total Variation = Residual Variation + Regression Variation
2. **Collection of exchangeable binary outcomes for the same covariate data are called outcomes.**
   Ans-> c) binomial
3. **How many outcomes are possible with Bernoulli trial?**
   Ans -> a) 2
4. **If Ho is true and we reject it is called.**
   Ans-> a) Type-I error
5. **Level of significance is also called:**
   Ans -> b) Size of the test
6. **The chance of rejecting a true hypothesis decreases when sample size is:**
   Ans -> b) Increase
7. **Which of the following testing is concerned with making decisions using data?**

   Ans -> b) Hypothesis

8. **What is the purpose of multiple testing in statistical inference?**
   Ans -> b) Minimize false positives
9. **Normalized data are centred at and have units equal to standard deviations of the original data.**

   Ans -> a) 0

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

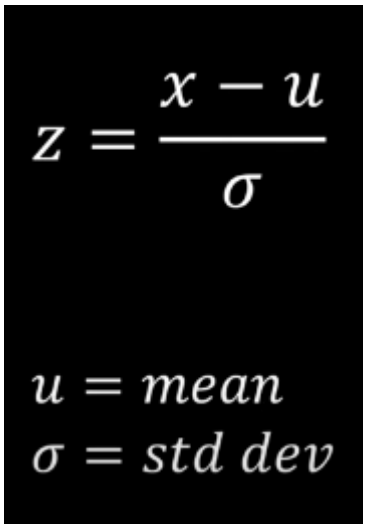
10. What Is Bayes' Theorem?

Ans -> Bayes Theorem provides a principled way for calculating a conditional probability. It is a deceptively simple calculation, although it can be used to easily calculate the conditional probability of events where intuition often fails.

11. What is z-score?

Ans -> How many Standard Deviation away a data points is from mean.

Formula of the Z score :-



Lets Understand -> Suppose we have height columns and we have to find the z score of the height column. So firstly we have to find the mean of the height column and minus with the height and then devide by the height Standard Deviation . So that how we able to find the Z score .

Z score is nothing but it can show How much standard deviation away the data points mean ..

12. What is t-test?

Ans -> T-Test is also called Student Ttest.

- **A t-test is a statistical test that us used to compare the means of two groups.**
- **It is often used in hypothesis is testing to determine wheather a process or treatment actually has an effect on the population of intrest m Or wheather two groups are different from one another.**

When to use t test ?

- **When we have to compare mean of two groups .**

Performing a T test –

- **So if we want to perform t test we can also perform manually by using formula and we also use statistical analysis software .**

T test Formula :

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

t -> T is the t value.

x1 and x2 a-> are the means of the two groups being compared

s2 -> s2 is the pooled standard error of two groups , and

n1 and n2 -> are the numbers of observations in each groups .

Lets Understand with example . Suppose we have a 2 columns data .

X1 =[60,70,80,90,80]

X2 =[10,20,30,40,50]

Step 1 -> Subtract X1 – X2 and check the result.

X1-X2 = [50,50,50,50,30]

Step 2 -> Add up all the values of Step 1 .

50+

Here we got 230.

Step 3 – Square the the difference from step 1 .

[2500,2500,2500,2500,900]

Step 4 -> Add up all of the squared difference from step 3

[10900]

Step 5 -> Use the t test formula to calculate the t score.

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

Put the values in this formula and get the t test score .

After putting the formula . we get some number.s

Step 6 -> Subtract one from the sample size to get the degree of freedom .

Step 7 -> Find the p value in the table , Using the degree of freedom in step 6 if you don't have a specifies alpha level , use 0.05 (5%)

Step 8 -> Compare your t table value from step 7 .

The calculated t value is greater then the table value at an alpha level 0.05 . The p value is less then the alpha level p<0.05.

We can reject the null hypothesis that there is no difference between mean .

13. What is percentile?

Ans -> Before Knowing of Percentile ,Lets Understand What is Percentage .

So , in Scool when we get result . we get our marks in percentage but any competitive exam we get marks in percentile.

So First lest understand How to Calculate Percentage :-

So we have data -> x=[1,2,3,4,5]

And , The Question is How Many percent on the dataset it even number ?

So , Here we can see that two numbers are even 2 and 4 .

So we all know how to calculate percentage ,

Formula :
Percentage = Meeting Characterstics of intrest/Total Observation *100
Lets apply the tha data in formula .

2/5*100=40%.

So here we get , Our 40% data is even data .and here we understand how to calculate percentage .

Lets Understand Percentile :-

A percentile a value below which a certain percentage of observation lie .

Lets understand with Example ->
Suppose we have a dataset :
x=[2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,12]
Lets understand, how to find percentile of 10?
The percentile ranking of any value is equal to the number of values below their value divided by total number values in the set times 100 . So there are 16 numbers below 10 and there are total 20 numbers .

Formula of Percentile ->

Percentile Rank = Value Below X/ Total number of Values * 100

Lets put value in the Formula .

Percentile =16/20*100 = 80 Percentile

The Percentile of the value 10 is 80% Percentile of score lie below it.

14. What is ANOVA?

Ans -> Anova is used to compare difference of means among more then two groups if does this by looking at variation in the data and where that variation is found Specifically , Anova compares the amount of variation between groups with the amount of variation within groups .

- **Null hypothesis , typically is that , all means are equal.**
- **The inpedenent Variables are categorical.**
- **Dependent Variables are continuous .**

F value -> F=Sample means of between groups/sample mean of within groups

RealTime Research on Human Body :-

It different types of drinks (Coffee,Water ,Milk, Soda,Etc). Have the same effect (reaction time in the morning) on human body ?

Null Hypothesis -> All the drinks average time on human body in same .

Alternate Hypothesis -> All the drinks average reaction time on human body is not same .

Hypothesis Construction - The Null Hypothesis for Annova says that , average of dependent varibles are same for all the given groups

alternate Hypothesis says -mean of dependent variable are not same for the given groups..

Ho:muA=muB=muC

Ha:not all are equal

15. How can ANOVA help?

Ans -> The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as the first example: age, sex, income). When you understand how each independent variable's mean is different from the others, you can begin to understand which of them has a connection to your dependent variable (landing page clicks), and begin to learn what is driving that behaviour.