**Statistics worksheet 3 Solution**

**Q10.**  What Is Bayes' Theorem?

**Ans1. Bayes’ theorem** describes the probability of occurrence of an event related to any condition. It is also considered for the case of [conditional probability](https://byjus.com/maths/conditional-probability-and-conditional-probability-examples/). Bayes theorem is also known as the formula for the probability of “causes”. For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

**Q11.** What is z-score?

**Ans. A z-score, also called a standard score gives an idea of how far from the**[**mean**](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/)**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](https://www.statisticshowto.com/raw-score/) is.

A z-score can be placed on a [**normal distribution**](https://www.statisticshowto.com/probability-and-statistics/normal-distributions/) 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](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/#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](https://www.statisticshowto.com/arithmetic-mean/)” 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.

**12**. What is t-test?

Ans. A T-test is a statistical method of comparing the means or proportions of two samples gathered from either the same group or different categories. It is aimed at hypothesis testing, which is used to test a hypothesis pertaining to a given population. It is the difference between population means and a hypothesized value.

Its mathematical formula is: T-test = variance between groups / variance within groups

If t-value is large => the two groups belong to different groups.   
If t-value is small => the two groups belong to same group.

**Steps involved in performing a t-test are:**

**Step 1 -** Find the sum of all values in each sample.

**Step 2 -** Square the sum values found in step 1.

**Step 3 -** Find the sum of square of individual values in each sample.

**Step 4 -** Calculate the mean of each sample.

**Step 5 -** Find the degree of freedom **(df)** using **Eq-2**.

**Step 6 -** Insert all the values found in **Steps 1-4** into **Eq-3** and find the calculated t-value.

**Step 7 -** Use the values of df and α (take α = 0.05 if not given) in the two-tails t-table to find the table value of t.

**Step 8 -** Compare values of t found in **Step-6** and **Step-7**.

There are three types of t-tests, and they are categorized as dependent and independent t-tests.

1. **Independent samples t-test:** compares the means for two groups.
2. **Paired sample t-test:** compares means from the same group at different times (say, one year apart).
3. **One sample t-test test:** the mean of a single group against a known mean.

**1. Independent sample t-test**

Independent sample t-test, commonly known as unpaired sample t-test is used to find out if the differences found between two groups is actually significant or just a random occurrence.

**2. Paired sample t-test**

Paired sample t-test, commonly known as dependent sample t-test is used to find out if the difference in the mean of two samples is 0. The test is done on dependent samples, usually focusing on a particular group of people or thing. In this, each entity is measured twice, resulting in a pair of observations.

**3. One sample t-test**

One sample t-test is one of the widely used t-tests for comparison of the sample mean of the data to a particularly given value. Used for comparing the sample mean to the true/population mean.

**Q13. What is percentile?**

Percentile is a term that describes how a score compares to other scores from the same set. It is commonly expressed as the percentage of values in a set of data scores that fall below a given value.

Percentiles show how a given value compares to others. The general rule is that if a value is in the kth percentile, it is greater than K per cent of the total values.

Formula to calculate percentile is: **Px = x(n+1)/100**

Where Px is value at which x percentage of data lie within that value.  
n is total number of observations.

For example: If P90 = 94, then it means that 90% of students got less than 94 and 10% of students got more than 94.

14. What is ANOVA?

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.

Apart from t-test and z-test methods, analysis of variance method is used for statistical analysis .ANOVA i.e. analysis of variance is the extension of the t- and z-tests.

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA. For example, MANOVA (multivariate ANOVA) differs from ANOVA as the former tests for multiple dependent variables simultaneously while the latter assesses only one dependent variable at a time. One-way or two-way refers to the number of independent variables in your analysis of variance test. A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as salary and skill set. It is utilized to observe the interaction between the two factors and tests the effect of two factors at the same time.

In Anova test, a one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

**The Formula for ANOVA is:**

F = MST/MSE

**where:** F=ANOVA coefficient

MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error​

If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

**Q15. How can ANOVA help?**

The ANOVA test helps in analyzing factors that affect a given data set. Once the test is finished, an analyst can perform additional testing on the methodical factors that measurably contribute to the data set's inconsistency. The analyst utilizes the ANOVA test results in an f-test to generate additional data that aligns with the proposed regression models.

The ANOVA test also allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

If no real difference exists between the tested groups, which is called the null hypothesis, the result of the ANOVA's F-ratio statistic will be close to 1. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator degrees of freedom and the denominator degrees of freedom.