**STATISTICS– WORKSHEET 4**

**Solution:1** The central limit theorem is a result from [probability theory](https://www.thoughtco.com/probability-and-what-were-the-chances-2312523). This theorem shows up in a number of places in the field of statistics. Although the central limit theorem can seem abstract and devoid of any application, this theorem is actually quite important to the practice of statistics.

The Central Limit Theorem (CLT) states that the distribution of a sample mean that approximates the normal distribution, as the sample size becomes larger, assuming that all the samples are similar, and no matter what the shape of the population distribution is. The central limit theorem holds for the sample of size greater than or equal to 30. This theorem is very important for testing hypotheses in statistical analysis.

**Solution:2** When you conduct research about a group of people, it’s rarely possible to collect data from every person in that group. Instead, you select a sample. The sample is the group of individuals who will actually participate in the research. The process of selecting these people is called sampling.

Sampling method:

1. Simple random sampling
2. Systematic sampling
3. Cluster sampling
4. Convenience sampling
5. Snowball sampling
6. Voluntary response sampling

**Solution:3**

|  |  |  |
| --- | --- | --- |
| **Basis for comparison** | **Type I error** | **Type II error** |
| **Definition** | Type 1 error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true. | Type II error is the error that occurs when the null hypothesis is accepted when it is not true. |
| **Also termed** | Type I error is equivalent to false positive. | Type II error is equivalent to a false negative. |
| **Meaning** | It is a false rejection of a true hypothesis. | It is the false acceptance of an incorrect hypothesis. |
| **Symbol** | Type I error is denoted by α. | Type II error is denoted by β. |
| **Probability** | The probability of type I error is equal to the level of significance. | The probability of type II error is equal to one minus the power of the test. |
| **Reduced** | It can be reduced by decreasing the level of significance. | It can be reduced by increasing the level of significance. |
| **Cause** | It is caused by luck or chance. | It is caused by a smaller sample size or a less powerful test. |
| **What is it?** | Type I error is similar to a false hit. | Type II error is similar to a miss. |
| **Hypothesis** | Type I error is associated with rejecting the null hypothesis. | Type II error is associated with rejecting the alternative hypothesis. |
| **When does it happen?** | It happens when the acceptance levels are set too lenient. | It happens when the acceptance levels are set too stringent. |

**Solution:4** Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

* A normal distribution is the proper term for a probability bell curve.
* In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
* Normal distributions are symmetrical, but not all symmetrical distributions are normal.
* In reality, most pricing distributions are not perfectly normal.

**Solution:5** Correlation: Correlation is a step ahead of covariance as it quantifies the relationship between two random variables. In simple terms, it is a unit measure of how these variables change with respect to each other.

Covariance: Covariance measures how the two variables move with respect to each other and is an extension of the concept of variance (which tells about how a single variable varies). It can take any value from -∞ to +∞.

**Solution:6**

1.Univariate statistics summarize only one variable at a time.

2.Bivariate statistics compare two variables.

3.Multivariate statistics compare more than two variables.

**Solution:7** The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as sensitive analysis. It’s usage will depend on one or more input variables within the specific boundaries, such as the effect that changes in interest rates will have on a bond’s price. It is also known as the what – if analysis. Sensitivity analysis can be used for any activity or system. All from planning a family vacation with the variables in mind to the decisions at corporate levels can be done through sensitivity analysis.

The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

**Solution:8** Hypothesis testing in statistics is a way of testing the results of a survey or an experiment in order to get meaningful results. It provides us the way of testing whether the outcomes are valid by figuring out the odd ones that appeared while evaluation. In addition to it, it can be used in statistics by an analyst in order to test an assumption regarding a population parameter.

H0 and H1:

H0 is a null hypothesis while H1 is an alternative hypothesis. Research studies and testing usually formulate two hypotheses. One will describe the prediction while the other will describe all other possible outcomes.

A null hypothesis is a hypothesis that says there is no statistical significance between the two variables. It is usually the hypothesis a researcher or experimenter will try to disprove or discredit. An alternative hypothesis is one that states there is a statistically significant relationship between two variables.

**Solution:9** Qualitative data : Qualitative data represent characteristics. This data should be interpreted using plain language than numbers. Conducting interviews and observations are some methods to obtain qualitative data. The descriptive nature of these data makes them difficult to analyze. Some examples of these type of data is gender, country, city, nationality, etc.

Qualitative data : Qualitative data represents numerical information. These type of data is easier to compute and analyze. Some examples are length, width, height, weight, area, speed, age, lifetime of an electric bulb, etc. Tables, graphs, charts help to represent quantitative data. It is possible to find quantitative data with the help of surveys, observations, interviews, and other experiments.

**Solution:10** Range: Range is the difference between the highest value and lowest value.

Range= highest-lowest

IQR: The difference between the upper and lower quartile is known as the interquartile range. The formula for the interquartile range is given below

Interquartile range = Upper Quartile – Lower Quartile = Q­3 – Q­1

**Solution:11** The term bell curve is used to describe the mathematical concept called normal distribution, sometimes referred to as Gaussian distribution. "Bell curve" refers to the bell shape that is created when a line is plotted using the data points for an item that meets the criteria of normal distribution.

In a bell curve, the center contains the greatest number of a value and, therefore, it is the highest point on the arc of the line. This point is referred to the mean, but in simple terms, it is the highest number of occurrences of an element (in statistical terms, the mode).

**Solution:12** Outliers are data points that are far from other data points. In other words, they’re unusual values in a dataset. Outliers are problematic for many statistical analyses because they can cause tests to either miss significant findings or distort real results**.**

**Z-Score formul:**

To calculate the Z-score for an observation, subtract the mean, and divide by the standard deviation. Mathematically, the formula for that process is the following:

z-score=x-mean / standard deviation

**Solution:13** In statistics, the p-value is the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis.

**Solution:14**

**Binomial Probability Formula**

P (X) = nCx px qn – x

Where,

n = Total number of trials

x = Total number of successful trials

p = probability of success in a single trial

q = probability of failure in a single trial = 1-p

**Solution:15** 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.