 WORKSHEET 4

**STATISTICS WORKSHEET-4**

**Q1to Q15 are descriptive types. Answer in brief.**

1. **What is central limit theorem and why is it important?**

Ans: The central limit theorem says that **the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough**. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

The central limit theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size (N) increases.

1. **What is sampling? How many sampling methods do you know?**

**Ans:** In a statistical study, sampling methods refer to **how we select members from the population to be in the study**. If a sample isn't randomly selected, it will probably be biased in some way and the data may not be representative of the population. There are many ways to select a sample—some good and some bad.

There are **two types of sampling methods**: Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group. Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

1. **What is the difference between type1 and typeII error?**

**Ans:**

|  |  |
| --- | --- |
| **Type 1 Error** | **Type II Error** |
| * **Type I error** is a false positive conclusion | * **Type II error** is a false negative conclusion. |
|  |  |
| * The probability of making a Type I error is the significance level, or alpha (α) | * The probability of making a Type II error is beta (β). |
| * A Type I error means rejecting the null hypothesis when it’s actually true. | * A Type II error means not rejecting the null hypothesis when it’s actually false. |
|  |  |
|  |  |
|  |  |

1. **What do you understand by the term Normal distribution?**

**Ans:** 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 graphical form, the normal distribution appears as a "bell curve".

In a normal distribution, data are symmetrically distributed with no skew. Most values cluster around a central region, with values tapering off as they go further away from the center. The measures of central tendency (mean, mode, and median) are exactly the same in a normal distribution.

1. **What is correlation and covariance in statistics?**

**Ans:** Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related (meaning they change together at a constant rate). It's a common tool for describing simple relationships without making a statement about cause and effect.

1. **Differentiate between univariate ,Biavariate,and multivariate analysis.**

**Ans:**.

|  |  |  |
| --- | --- | --- |
| Univariate Analysis | Bivariate Analysis | Multivariate Analysis |
| This type of data consists of only one variable. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. | This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables. | When the data involves **three or more variables**, it is categorized under multivariate. It is similar to bivariate but contains more than one dependent variable. The ways to perform analysis on this data depends on the goals to be achieved. |
|  |  |  |
| The example of a univariate data can be height. Suppose that the heights of seven students of a class is recorded, there is only one variable that is height and it is not dealing with any cause or relationship. The description of patterns found in this type of data can be made by drawing conclusions using central tendency measures (mean, median and mode), dispersion or spread of data (range, minimum, maximum, quartiles, variance and standard deviation) and by using frequency distribution tables, histograms, pie charts, frequency polygon and bar charts. | Example of bivariate data can be temperature and ice cream sales in summer season. Bivariate data analysis involves comparisons, relationships, causes and explanations. These variables are often plotted on X and Y axis on the graph for better understanding of data and one of these variables is independent while the other is dependent. | Example of this type of data is suppose an advertiser wants to compare the popularity of four advertisements on a website, then their click rates could be measured for both men and women and relationships between variables can then be examined. |
|  |  |  |

1. **What do you understand by sensitivity and how would you calculate it?**

**Ans:** Sensitivity analysis is a financial model that determines how target variables are affected based on changes in other variables known as input variables. It is a way to predict the outcome of a decision given a certain range of variables.

The formula for sensitivity analysis is basically a financial model in excel where the analyst is required to identify the key variables for the output formula and then assess the output based on different combinations of the independent variables.

Mathematically, the dependent output formula is represented as,

**Z = X2 + Y2**

Below are mentioned the steps used to conduct sensitivity analysis:

1. Firstly, the base case output is defined; say the NPV at a particular base case input value (V1) for which the sensitivity is to be measured. All the other inputs of the model  are kept constant.
2. Then the value of the output at a new value of the input (V2) while keeping other inputs constant is calculated.
3. Find the percentage change in the output and the percentage change in the input.
4. The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.
5. **What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?**

**Ans:** Hypothesis Testing is a type of [statistical analysis](https://www.simplilearn.com/what-is-statistical-analysis-article) in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables. There are two types of Hypothesis testing in statistics:

* 1. **Null Hypothesis**: The Null Hypothesis is the assumption that the event will not occur. A null hypothesis has no bearing on the study's outcome unless it is rejected.H0 is the symbol for it, and it is pronounced H-naught.
  2. **Alternative Hypothesis**: The Alternate Hypothesis is the logical opposite of the null hypothesis. The acceptance of the alternative hypothesis follows the rejection of the null hypothesis. H1 is the symbol for it.

**H0 and H1 for two-tail test:-** In two tails, the test sample is checked to be greater or less than a range of values in a Two-Tailed test, implying that the critical distribution area is two-sided.

If the sample falls within this range, the alternate hypothesis will be accepted, and the null hypothesis will be rejected.

1. **What is quantitative data and qualitative data?**

**Ans:** Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code. Start with yourself as an example. To acquire qualitative data, consider identifiers like the color of your clothes, type of hair, and nose shape. For quantitative data, consider measurable values like your height, weight, age, and shoe size.

1. **How to calculate range and interquartile range?**

**Ans:** The range is calculated by subtracting the lowest value from the highest value.

While The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), ​first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

1. **What do you understand by bell curve distribution?**

**Ans:** A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

The bell-shaped curve is a common feature of nature and psychology. The normal distribution is the most important probability distribution in statistics because many continuous data in nature and psychology displays this bell-shaped curve when compiled and graphed.

1. **Mention one method to find outliers.**

**Ans: Outliers** are extreme values that differ from most other data points in a dataset. They can have a big impact on your [statistical analyses](https://www.scribbr.com/category/statistics/) and skew the results of any [hypothesis tests](https://www.scribbr.com/statistics/hypothesis-testing/).

It’s important to carefully identify potential outliers in your dataset and deal with them in an appropriate manner for accurate results.

Statistical method (Z scores) : Statistical outlier detection involves applying **statistical tests** or procedures to identify extreme values.

You can convert extreme data points into [z scores](https://www.scribbr.com/statistics/standard-normal-distribution/) that tell you how many standard deviations away they are from the mean.

If a value has a high enough or low enough z score, it can be considered an outlier. As a rule of thumb, values with a z score greater than 3 or less than –3 are often determined to be outliers.

1. **What is p-value in hypothesis testing?**

**Ans:** The p-value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P-values are used in hypothesis testing to help decide whether to reject the null hypothesis.

A p-value measures the probability of obtaining the observed results, assuming that the null hypothesis is true. The lower the p-value, the greater the statistical significance of the observed difference. A p-value of 0.05 or lower is generally considered statistically significant.

1. **What is the Binomial Probability Formula?**

**Ans:** Binomial probability refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment). If the probability of success on an individual trial is p , then the binomial probability is **nCx⋅px⋅(1−p)n−x** .

1. **Explain ANOVA and it’s applications.**

**Ans:** Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests. 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.

ANOVA is a method to determine if the mean of groups are different. In inferential statistics, we use samples to infer properties of populations. Statistical tests like ANOVA help us justify if sample results are applicable to populations.

ANOVA checks the impact of one or more factors by comparing the means of different samples. We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not. Another measure to compare the samples is called a t-test. When we have only two samples, t-test and ANOVA give the same results.

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.

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