**Iqra National University**

****

**Name :** Haroon Zazai

**T/name :** Ma’am Wajeeha Amin

**Class :** BBA 4th

**ID :** 15314

**Subject :** Statistical Inferences

**Assignment** : Mid-term

**Date :** 21st August 2020

**Q1:**The average rainfall in an area recorded is 9.22cm for a month. Given the distribution to be normally distributed with a standard deviation of 2.83cm,

* Find the probability that the rainfall in the next month will be less than 1.84cm.
* Rainfall will be between 7 cm and 13.8cm.
* Rainfall is more than 11.05 cm  **(3+4+3=10)**

H= 9.22cm σ = 2.83cm P (x< 1.84)=?

Z = $\frac{1.84-9.22}{2.83}$ Z = -2.607

P (x< 1.84) = P ( Z < -2.607 = 0.0038)

1. P (7<x<13.8) = ?

Z1 = $\frac{x-u }{o}$ = $\frac{7-9022}{2.83}$ **=** 0.784

P (x>7) = ?

P (x>7) = P (Z> -0.784) = 0.2177

P (13.8>x) = ?

Z2 = $\frac{x-u }{o}$ = $\frac{13.8-9.22}{2.83}=1.618$

P ( x<13.8) = P (Z<1.618) = 0.9463

**(2)** P (7<x<13.8)

= P (-0.784) < Z < 1.618)

= 0.9463 – 0.2177

**=** 0.7286

**(3)** P (x > 11.05)

Z=$\frac{x-u }{o}$ Z= $\frac{11.05-9.22}{2.83}$ Z= 0.646

P (x > 11.05)

P ( 2 > 0.646)

**=** 0.7389

**Q2: (a)** Discuss any 3 characteristics of normal distribution and discuss its uses in the business world. **(4)**

 **(b)** Suppose you are going to be conducting a study on students, asking for their opinion on an issue of interest to you (could be related to the university, or a wider societal issue). Describe how you would carry out the sampling of students using the following methods:

(i) simple random sampling

(ii) stratified sampling

(iii) cluster sampling Think about what attributes of the student population make sense to stratify vs. cluster. **(6)**

**2.A**

1. **Characteristics of Normal Distribution**
* The distribution is symmetric; its skewness measure is zero.

 X

Symmetrical distribution occurs when the values of variables occur at regular frequencies and the mean, median and mode occur at the same point. In graph form, symmetrical distribution often appears as a bell curve. If a line were drawn dissecting the middle of the graph, it would show two sides that mirror each other.

Symmetrical distribution is most often used to put price action into context. The further the price action wanders from the value area one standard deviation on each side of the mean, the greater the probability that the underlying asset is being under or overvalued by the market.

* The entire family of normal probability distributions is defined by its mean μ and its standard deviation σ.

 Standard Deviation ở

 X

 Mean µ

Standard deviation is a measure of how spread out a data set is. It's used in a huge number of applications. In finance, standard deviations of price data are frequently used as a measure of volatility. The standard deviation of company A's employees is 1, while the standard deviation of company B's wages is about 5.

The standard deviation is used in conjunction with the mean to summarise continuous data, not categorical data. In addition, the standard deviation, like the mean, is normally only appropriate when the continuous data is not significantly skewed or has outliers.

The standard deviation is a summary measure of the differences of each observation from the mean. If the differences themselves were added up, the positive would exactly balance the negative and so their sum would be zero. Consequently the squares of the differences are added.

* The highest point on the normal curve is at the mean, which is also the median and mode.

 X

The highest point of the Normal curve occurs for the mean of the population. The properties of the Normal distribution ensure that this point is also the median value and the mode. The shape of the Normal curve (relatively narrow or relatively broad) is influenced by the standard deviation (SD) of the data.

The highest point of the Normal curve occurs for the mean of the population. The properties of the Normal distribution ensure that this point is also the median value and the mode.

A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. You can think of it as the tendency of data to cluster around a middle value. In [statistics](https://statisticsbyjim.com/glossary/statistics/), the three most common measures of central tendency are the [mean](https://statisticsbyjim.com/glossary/mean/), [median](https://statisticsbyjim.com/glossary/median/), and [mode](https://statisticsbyjim.com/glossary/mode/). Each of these measures calculates the location of the central point using a different method.

**Question 2.B:**

1. **Simple random sampling**

In this technique, each member of the population has an equal chance of being selected as subject. The entire process of sampling is done in a single step with each subject selected independently of the other members of the population. Another way would be to let a computer do a random selection from your population.

1. Step one: define the population.
2. Step two: choose your sample size.
3. Step three: list the population.
4. Step four: assign numbers to the units.
5. Step five: find random numbers.
6. Step six: Select your sample.

 \* It would be the names of 25 students being chosen out of a hat from a batch of 250 students. In this case the population is all 250 employees and the sample is random because each employee has an equal chance of being chosen. Random sampling is used in science to conduct randomized control test or for blinded experiments.

1. **Stratified sampling**

A sample may be selected from a population through a number of ways, one of which is the stratified random sampling method. A stratified random sampling involves dividing the entire population into homogeneous groups called strata (plural for stratum). Random samples are then selected from each stratum.

1. Define the population. ...
2. Choose the relevant stratification. ...
3. List the population. ...
4. List the population according to the chosen stratification. ...
5. Choose your sample size. ...
6. Calculate a proportionate stratification. ...
7. Use a simple random or systematic sample to select your sample.

\* I will now divide all of the students into two groups then conduct a questioner test then will select the ones with the best results.

This is conducted to save time and money.

The result of a subset of population which represents the entire population.

1. **Cluster sampling Think about what attributes of the student population make sense to stratify vs. cluster.**

1. In stratified sampling, a sample is drawn from each strata (using a random sampling method like simple random sampling or systematic sampling).

2. In cluster sampling, the sampling unit is the whole cluster; Instead of sampling individuals from within each group, a researcher will study whole clusters

Cluster sampling refers to a type of sampling method . With cluster sampling, the researcher divides the population into separate groups, called clusters. Then, a simple random sample of clusters is selected from the population. The researcher conducts his analysis on data from the sampled clusters.

\* I will divide the students in different clusters then select the best group for the test.

I think the cluster formation will be best for the students division.

**Q3: (a) Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience). (4)**

**1. A group of test subjects is divided into twelve groups; then four of the groups are chosen at random.**

**2. A market researcher polls every tenth person who walks into a store.**

**3. The first 50 people who walk into a sporting event are polled on their television preferences.**

**4. A computer generates 100 random numbers, and 100 people whose names correspond with the numbers on the list are chosen.**

**(b) Differentiate between (2+2+2)**

* **Descriptive statistics and inferential statistics**
* **Variance and standard deviation**
* **Cluster and strata**

**3.A**

1. Cluster

 2. Systematic

3. Convenience

4. Simple random

**3.B**

* **Descriptive statistics and inferential statistics**

**Descriptive Statistics**

Descriptive statistics give information that describes the data in some manner. For example, suppose a pet shop sells cats, dogs, birds and fish. If 100 pets are sold, and 40 out of the 100 were dogs, then one description of the data on the pets sold would be that 40% were dogs.

This same pet shop may conduct a study on the number of fish sold each day for one month and determine that an average of 10 fish were sold each day. The average is an example of descriptive statistics.

Some other measurements in descriptive statistics answer questions such as 'How widely dispersed is this data?', 'Are there a lot of different values?' or 'Are many of the values the same?', 'What value is in the middle of this data?', 'Where does a particular data value stand with respect with the other values in the data set?'

A graphical representation of data is another method of descriptive statistics. Examples of this visual representation are histograms, bar graphs and pie graphs, to name a few. Using these methods, the data is described by compiling it into a graph, table or other visual representation.

This provides a quick method to make comparisons between different data sets and to spot the smallest and largest values and trends or changes over a period of time. If the pet shop owner wanted to know what type of pet was purchased most in the summer, a graph might be a good medium to compare the number of each type of pet sold and the months of the year.

**Inferential Statistics**

Now, suppose you need to collect data on a very large population. For example, suppose you want to know the average height of all the men in a city with a population of so many million residents. It isn't very practical to try and get the height of each man.

This is where inferential statistics comes into play. Inferential statistics makes inferences about populations using data drawn from the population. Instead of using the entire population to gather the data, the statistician will collect a sample or samples from the millions of residents and make inferences about the entire population using the sample.

The sample is a set of data taken from the population to represent the population. Probability distributions, hypothesis testing, correlation testing and regression analysis all fall under the category of inferential statistics.

**Variance and standard deviation**

**Standard Deviation**

Standard deviation is a statistic that looks at how far from the mean a group of numbers is, by using the square root of the variance. The calculation of variance uses squares because it weighs outliers more heavily than data closer to the mean. This calculation also prevents differences above the mean from canceling out those below, which would result in a variance of zero.
Standard deviation is calculated as the square root of variance by figuring out the variation between each data point relative to the mean. If the points are further from the mean, there is a higher deviation within the date; if they are closer to the mean, there is a lower deviation. So the more spread out the group of numbers are, the higher the standard deviation.
To [calculate standard deviation](https://www.investopedia.com/ask/answers/021015/how-can-you-calculate-volatility-excel.asp), add up all the data points and divide by the number of data points, calculate the variance for each data point and then find the square root of the variance.

## Variance

The variance is the average of the squared differences from the mean. To figure out the variance, first calculate the difference between each point and the mean; then, square and average the results.
For example, if a group of numbers ranges from 1 to 10, it will have a mean of 5.5. If you square the differences between each number and the mean, and then find their sum, the result is 82.5﻿. To figure out the variance, divide the sum, 82.5, by N-1, which is the sample size (in this case 10) minus 1. The result is a variance of 82.5/9 = 9.17. Standard deviation is the square root of the variance so that the standard deviation would be about 3.03.
Because of this squaring, the variance is no longer in the same unit of measurement as the original data. Taking the root of the variance means the standard deviation is restored to the original unit of measure and therefore much easier to interpret.

* Standard deviation looks at how spread out a group of numbers is from the mean, by looking at the square root of the variance.
* The variance measures the average degree to which each point differs from the mean—the average of all data points.
* The two concepts are useful and significant for traders, who use them to measure market volatility.

**Cluster and strata**

The main difference between stratified sampling and cluster sampling is that with [cluster sampling](https://www.statisticshowto.com/what-is-cluster-sampling/), you have natural groups separating your population. For example, you might be able to divide your data into natural groupings like city blocks, voting districts or school districts. With [stratified random sampling](https://www.statisticshowto.com/stratified-random-sample/), these breaks may not exist\*, so you divide your target population into groups (more formally called "strata").

* In stratified sampling, a sample  is drawn from each strata (using a random sampling method like [simple random sampling](https://www.statisticshowto.com/simple-random-sample/) or [systematic sampling](https://www.statisticshowto.com/systematic-sampling/)). In the image below, let's say you need a sample size of 6. Two members from each group (yellow, red, and blue) are selected randomly. Make sure to sample proportionally: In this simple example, 1/3 of each group (2/6 yellow, 2/6 red and 2/6 blue) has been sampled. If you have one group that's a different size, make sure to adjust your proportions. For example, if you had 9 yellow, 3 red and 3 blue, a 5-item sample would consist of 3/9 yellow (i.e. one third), 1/3 red and 1/3 blue.
* In cluster sampling, the [sampling unit](https://www.statisticshowto.com/sampling-unit/) is the whole cluster; Instead of sampling individuals from within each group, a researcher will study whole clusters. In the image below, the strata are natural groupings by head color (yellow, red, blue). A sample size of 6 is needed, so two of the complete strata are selected randomly (in this example, groups 2 and 4 are chosen).

## https://storage.ning.com/topology/rest/1.0/file/get/3553181755?profile=RESIZE_710x