**CSUSM**

**BUS 204**

**Spring 2023**

**Quiz 2 Class Notes**

as of 3/25/23

**Target Population:** The population for which statistical inferences such as point estimates are made. It is important for the target population to correspond as closely as possible to the sampled population. There can be unlimited number of target populations. There is just one point estimate for each population.

**Simple Random Sample:** A sample selected so that each item or person in the population has the same chance of being included.

**Systematic Random Sample:** A random starting point is selected, and then every kth member of the population is selected.

**Stratified Random Sample:** A population is divided into subgroups, called strata, and a sample is randomly selected from each stratum.

**Cluster Sample:** A population is divided into clusters using natural occurring geographic or other boundaries. Then, clusters are randomly selected, and a sample is collected by randomly selecting from each cluster.

**Convenience Sampling:** A non-probability method of sampling whereby elements are selected for the sample on the basis of convenience.

**Judgment Sampling:** A non-probability method of sampling whereby elements are selected for the sample based on the judgement of the person doing the study.

**Census:** Collection of data from every parameter in the population of interest.

**Statistical Inference:** The process of making estimates and drawing conclusions about one or more characteristics of a population (the value of one or more parameters.) through the analysis of sample data drawn from the population. It is a synonym of Inferential Statistics. Selection and use of sample data to produce information about the larger population from which the sample was selected.

**Sampling Error:** The difference between a sample statistic and its corresponding population parameter. It occurs because a random sample is used to estimate the population parameter.

**Sampling Distribution of the Sample Mean:** a probability distribution of all possible sample means of a given sample size. For a large enough sample size, the shape of the sampling distribution will be approximately normal. This distribution is centered at the mean of the population. The standard deviation of the sampling distribution can be computed as the population standard deviation divided by the square root of the sample size.

**Central Limit Theorem:** A theorem stating that when enough independent random variables are added, the resulting sum is normally distributed random variable. This result allows one to use the normal probability distribution to approximate sampling distributions of the sample mean and sample proportion for sufficiently large sample size. It assumes that all samples of a particular size can be selected from any population.

**Parameter:** A numerical characteristic of a population, such as a population mean, population standard deviation, a population proportion. A measurable factor that defines a characteristic of a population, process, or system.

**Sampled Population:** The population from which the sample is taken.

**Frame:** A listing of the elements that the sample will be selected from.

**Target Population:** The population for which statistical inferences such as point estimates. It is important for the target population to correspond as closely as possible to the sampled population. There can be an unlimited number of target populations, but each will have just one point estimate.

**Sampling Distribution of the Sample Mean:** A probability distribution consisting of all possible values of a sample statistic. The sample mean is unbiased because the mean of all possible sample means (of a given sample size) is equal to the population mean.

**Unbiased:** A property of a point estimate that is present when the expected value of the point estimator is equal to the population parameter it estimates.

**Point Estimate:** The statistic, computed from sample information, which is used to estimate the population parameter. It is a parameter computed from sample information.

**The Point Estimator:** is the sample statistic, such as $\overbar{x},\overbar{p}.s $, that provides the point estimate of the population parameter.

**Standard Error:** The standard deviation of a point estimator. It is a range of point estimators.

**Confidence Interval:** A range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called the level of confidence. Another name for an interval estimate.

**Confidence Level:** The confidence associated with an interval estimate. For example, if an interval estimation procedure provides intervals such that 95% of the intervals formed using the procedure will include the population parameter, the interval estimate is said to be constructed at 95% confidence level. It is the percentage of all possible confidence intervals that will contain the true population parameter.

**Confidence Coefficient:** The confidence level expressed as a percentage decimal value. For example, 0.95 is the confidence coefficient for a 95% confidence level. It is a range of values within which the population parameter is expected to occur, it will determine the width of a confidence interval, the variability in population, is usually estimated by the ‘s”.

**Proportion:** The fraction, ratio, or percent indicating the part of the sample or the population having a particular trait of interest.

**Margin of Error:** The +/- value added to and subtracted from a point estimate in order to develop an interval estimate of a population parameter. It is a measure of how close we expect the point estimate to be to the population parameter with the specific level of confidence.

**Level of Significance α:** The probability that the interval estimation procedure will generate an interval that does not contain the value of the parameter of interest. This is also the maximum allowable probability of committing a Type I error it is the level of acceptable risk.

**Interval Estimate:** An estimate of a population parameter that provides an interval believed to contain the value of the parameter. It has the form of point estimate +/- margin of error. It is an element of the process of using sample data to calculate the range of values that is believed to include the unknown value of a population parameter. It has the values of the Upper and the Lower limits.

**t-distribution:** A family of probability distributions that can be used to develop an interval estimate of a population mean whenever the population standard deviation is unknown and is estimated by the sample standard deviation (s.) It is generally bell-shaped and symmetrical and tends to be flatter and broader than the standard normal distribution. It is also called the Students’ distribution.

**Degrees of Freedom:** A parameter of the t-distribution. Calculated as n-1 where the n is the sample size. As the number of “degrees of freedom” increases, the shape of the t-distribution approaches the standard normal distribution. Each t-distribution is defined by its degrees of freedom

**Hypothesis:** A statement about a population parameter subject to verification.

**Hypothesis Testing:** A procedure based on sample evidence and probability theory to determine whether the hypothesis is a reasonable statement. The process of making a conjecture about the value of the population parameter, collecting sample data that can be used to assess this conjecture, measuring the strength of evidence against this conjecture that is provided by the sample, and using these results to draw a conclusion about the conjecture. It is a procedure based on sample evidence and probability theory. It leads to either acceptance or rejection of Ha.

**Null Hypothesis (Ho):** A statement about the value of a population parameter developed for the purpose of testing numerical evidence. The hypothesis that is assumed to be tentatively true in the hypothesis testing procedure. It is the statement that is accepted if the sample data provide sufficient evidence that Ho is true. This is a proposition that is directly challenged by the evidence.

**Alternate Hypothesis (Ha):** A statement that is accepted if the sample data provide sufficient evidence that the null hypothesis is false. It tests the numerical evidence, it helps to disprove the Ho based on the sample data, Ho is either rejected or not rejected based on the value of Ha. If Ho is not rejected based on the sample data, we still cannot say that it true.

**Type I Error:** Rejecting the Null Hypothesis when it is true.

**Type II Error:** Accepting the Null Hypothesis when it is false.

**Test Statistic:** A value, calculated from sample information necessary for determining whether there is enough evidence to reject the Ho. This is done by examining the location of the test statistic (in regions of rejection or acceptance of the Ha.) This will always be one number.

**Critical Value:** The dividing point between the region where the null hypothesis is rejected and the region where it is not rejected.

**Summary of the Steps in Hypothesis Testing:**

1. Define the Null hypothesis (Ho)

2. Define the Alternate Hypothesis (Ha)

3. Select the level of significance

4. Formulate the decision rule

5. Calculate critical values

6. Calculate test statistic

7. Formulate findings

8. Decide re: Ho

**One-tailed Test:** A hypothesis test in which rejection of Ho occurs for values of the test statistic in one tail of its sampling distribution.

**Two-tailed Test:** A hypothesis test in which rejection of Ho occurs for values of the test statistic in either tail of its sampling distribution. It can be used with the z distribution as well as with the t-distribution.

**Students’ Distribution:** It is a family distribution that is bell-shaped and symmetrical like standard normal distribution. It is another name for the t-distribution renamed by Guinness Brewing Company.”

**Sample Proportion:** is the fraction of items in a sample that have the attribute of interest.

**Law of Large Numbers:** a sample mean will approach the mean of the parent population as sample size increases.

**Sample size:** the value of α is directly related to the sample size.

**Sample size of ≥ 30:** is considered to be a large sample. In sch a case the “s” value can be substituted for $σ$

**When** $σ$**:** is unknown and the sample size is small we need to use the t-distribution.

**α:** represents the level of significance

**A Decision Rule:** indicates the condition or conditions when the Ho is rejected, in two-tailed test, the rejection region is evenly split between the upper and lower tail, in one-tailed test, all rejection region is either the upper or the lower tail