Week 10

Confidence Interval

Statistics (Sample)	Parameter (Population)
$\bar{x}$	μ
s^2	σ^2
s	σ
$\hat{p}$	P
mode	Mode
median	Median

Definition (Point Estimator and Point Estimation): Any single statistical value, for example, arithmetic mean or standard deviation or median that is computed based on a sample size of n is called **a point estimation** of a corresponding parameter. The mathematical formula used to compute this value is called **a point estimator**. For example, sample arithmetic mean ($\bar{x}$) is a point estimator for μ and its numerical value is called point estimation. The mathematical formula is defined by

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Definition (Confidence Interval and Confidence Estimation): Any confidence interval estimator for a parameter is a rule that defines an interval based on utilizing the information available in a sample size of n. Those values of the interval are called the estimations of a confidence interval

Definition: Let θ be a parameter of a population, using the available information in sample, a confidence interval for the parameter θ could be expressed as follows:

$$P(A < \theta < B) = 1 - \alpha$$

Let A and B be two random variables for the parameter called θ. Let a and b be two realizations (numerical values for A and B based on a sample size of n. The interval whose lower and upper bound represented by a and b with a probability 100(1-α) is called a confidence interval for θ. (1-α) is called the confidence level.

1.Confidence Interval for the Mean of a Population

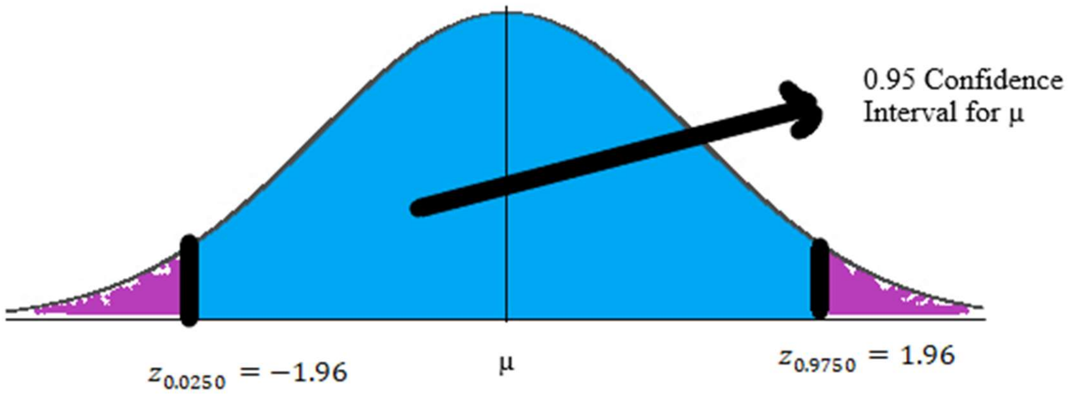
1.1 Confidence Interval for a Mean of the Normal Distribution: Variance of the Normal Distribution is Known

A data set following the normal distribution is used to sample n observations in order to estimate the mean of the population (μ) with any given or chosen probability represented by (1-α). By doing this, the lower bound and upper bound for μ with some probability of 1-α could be obtained. Besides, the assumption of known variance (σ^2) is presumed in advance. Then, the confidence interval for μ is denoted by

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Ex: In a packaging process of spice, the weights of the packages are distributed normally with standard deviation of 1.2 gram. A random sample of 25 packages are picked from the production line, and the sample mean is found 19.8 gr. Find the 95 percent confidence interval of mean value (μ) of the packages of spice.

$z_{0.0250} = -1.96$ and $z_{0.9750} = 1.96$



$$P\left(19.8 - \frac{1.2}{\sqrt{25}}(1.96) \leq \mu \leq 19.8 + \frac{1.2}{\sqrt{25}}(1.96)\right) = 0.95$$

$$P(19.33 \leq \mu \leq 20.27) = 0.95$$

Remark: The population mean value (μ) of the packages of spice is between 19.33 gram and 20.27 gram with a probability of 95.

Note: The distribution is normal and its variance σ^2 is known, which is equal to 1.44.

1.2. Confidence Interval for a Mean of the Normal Distribution: Variance of the Normal Distribution is Unknown

A data set following the normal distribution is used to sample n observations in order to estimate the mean of the population (μ) with any given or chosen probability represented by (1-α). By doing this, the lower bound and upper bound for μ with some probability of 1-α could be obtained. Besides, the assumption of unknow variance (σ^2) is presumed in advance. Then, the confidence interval for μ is denoted by

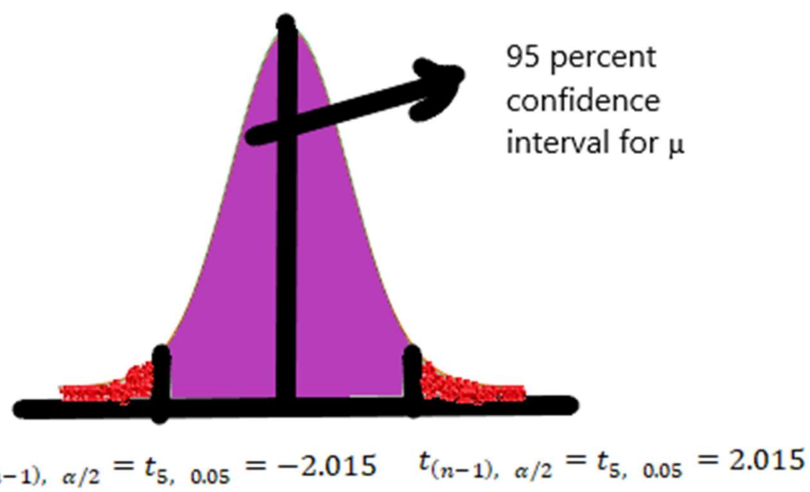
$$P\left(\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}}\right) = 1 - \alpha$$

Ex: Automobiles produced in a particular year are investigated based on fuel consumption. & automobiles are selected randomly and their fuel consumptions are provided below as follows:

18.6/18.4/19.2/20.8/19.4/20.5

Assuming the distribution of the fuel consumption for automobiles distributed normally, find the 90 percent confidence interval for mean consumption value for the population. And Interpret the result.

$$t_{(n-1),\ \alpha/2} = t_{5,\ 0.05} = \pm 2.015$$



$$\bar{x} = 19.48, s^2 = 0.96, s = 0.98$$

$$P\left(\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(19.48 - \frac{0.98}{\sqrt{6}}(2.015) \leq \mu \leq 19.48 + \frac{0.98}{\sqrt{6}}(2.015)\right) = 0.90$$

$$P(18.67 \leq \mu \leq 20.29) = 0.90$$

Remark: The average fuel consumption of the population of the same brand and year automobiles with 90 percent probability could be between 18.67 and 20.29 unit.

1.3. Confidence Interval for the Mean of a Population: Large Sample Size

When the distribution of the data and the variance of the population are not known, sample size is a critical information to determine which distribution would be used to estimate the mean of the population. Central Limit Theorem is used to make this estimation happen. If the sample size is greater than 30, then, any given distribution for the data could converge approximately to Normal Distribution. Then, the confidence interval for the mean of the population could be expressed as follows:

$$P\left(\bar{x} - \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Ex: 172 accounting major students are randomly selected. A question is asked to them to assess "whether job security is a critical issue for any given job". They evaluate this question using a scale between 1 through 5. While 1 corresponds to "not very important" to 5 corresponds to "very important". The sample mean and standard deviation are 4.38 and 0.70, respectively. Find the 99 percent confidence interval and interpret the result.

$$P\left(\bar{x} - \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(4.38-\frac{0.70}{\sqrt{172}}(2.575)\leq \mu \leq 4.38+\frac{0.70}{\sqrt{172}}(2.575)\right)=0.99$$

$$P(3.98\leq \mu \leq 4.39)=0.99$$

Remark: The population mean of the assessments of the accounting students could be between 3.98 and 4.39. hence, it can be said that the population of the accounting students think that job security is important for them.

2.The Confidence Interval for Portion of the Population: Large Sample Size

Definition: Instead of computing the average value for a population, a ratio of some observations which has some certain characteristic should be an interest to study. For example, a certain disease in a population could be dealt with. This parameter is called portion of the population and denoted by P. the sample portion is denoted by $\hat{p}$.

$\hat{p}=\frac{x}{n}$, where x denotes the number of observations related to some certain characteristic. The confidence interval for P is defined by

$$P\left(\hat{p}-\sqrt{\frac{p(1-p)}{n}}z_{\frac{\alpha}{2}}\leq P\leq \hat{p}+\sqrt{\frac{p(1-p)}{n}}z_{\frac{\alpha}{2}}\right)=1-\alpha$$

Ex: Recruitment officers from various companies on campus interview with senior university students. A random sample of 142 are selected to ask “whether the graduation score for the students are important to recruit them”. Out of 87 declared that it is important. Find the 95 percent confidence interval for the population portion of recruitment officers and interpret the result.

$$P\left(\hat{p}-\sqrt{\frac{p(1-p)}{n}}z_{\frac{\alpha}{2}}\leq P\leq \hat{p}+\sqrt{\frac{p(1-p)}{n}}z_{\frac{\alpha}{2}}\right)=1-\alpha$$

$$P\left(0.61-\sqrt{\frac{0.61(0.39)}{142}}1.96\leq P\leq 0.614\sqrt{\frac{0.61(0.39)}{142}}(1.96)\right)=0.95$$

$$P(0.533 \leq \mu \leq 0.693) = 0.95$$

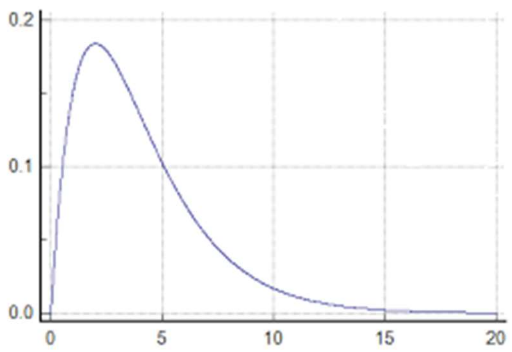
Remark: The population portion of the recruitment officers considering the graduation scores being important could be between 0.53 and 0.69 with 0.95 probability.

3.The Confidence Interval for Variance When the Distribution is Normal

The data is distributed normal and its variance is denoted by σ^2. A sample size of n is selected. The confidence interval for variance population is defined by

$$P\left(\frac{ns^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{n-1,\ 1-\alpha/2}}\right) = 1 - \alpha$$

Ex: A random sample of 15 pain killer whose standard deviation is 0.8 is chosen to determine the confidence interval for the standard deviation for the chemical used in these pain killers. Find the 90 percent confidence interval and interpret the result.



$$P\left(\frac{ns^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{n-1-\alpha/2}}\right) = 1 - \alpha$$

$$P\left(\frac{15(0.64)}{\chi^2_{14,\ 0.05}} \leq \sigma^2 \leq \frac{15(0.64)}{\chi^2_{14,\ 0.95}}\right) = 0.95$$

$$P\left(\frac{15(0.64)}{23.685} \leq \sigma^2 \leq \frac{15(0.64)}{6.571}\right) = 0.95$$

$$P(0.40 \leq \sigma^2 \leq 1.46) = 0.95$$

Remark: The population variance for the chemical used in the pain killer could be between 0.40 and 1.46 with 0.95 probability.