CHAPTER 6

Using Sample Data to Make Estimations About Population Parameters

6.1. Introduction :

Statistical Inferences: (Estimation and Hypotheses Testing):

It is the procedure by which we reach a conclusion about a population on the basis of the information contained in a sample drawn from that population.

There are two main purposes of statistics:

- **Descriptive Statistics:** (Chapter 1 & 2): Organization & summarization of the data.

- **Statistical Inference:** (Chapter 6 and 7): Answering research questions about some unknown population parameters.

(1) Estimation: (chapter 6)

Approximating (or estimating) the actual values of the unknown parameters:

- **Point Estimate:** A point estimate is single value used to estimate the corresponding population parameter.
- **Interval Estimate (or Confidence Interval):** An interval estimate consists of two numerical values defining a range of values that most likely includes the parameter being estimated with a specified degree of confidence.

(2) Hypothesis Testing: (chapter 7)

Answering research questions about the unknown parameters of the population (confirming or denying some conjectures or statements about the unknown parameters).

6.2. Confidence Interval

for a Population Mean (μ)

Statistical Inferences: (Estimation and Hypotheses Testing):

Population:

Population Size = N

Population Values: $X_1, X_2, \ldots, X_N$

Population Mean: $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$

Population Variance: $\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$

Sample:

Sample Size = n

Sample values: $x_1, x_2, \ldots, x_n$

Sample Mean: $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

Sample Variance: $S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$

(i) Point Estimation of μ:

A point estimate of the mean is a single number used to estimate (or approximate) the true value of μ .

- Draw a random sample of size n from the population:

  - $x_1, x_2, \ldots, x_n$

- Compute the sample mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Result:

The sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a "good" point estimator of the population mean (μ).

(ii) Confidence Interval (Interval Estimate) of μ:

An interval estimate of μ is an interval (L,U) containing the true value of μ "with a probability of $1-\alpha$".

* $1-\alpha$ = is called the confidence coefficient (level)
* L = lower limit of the confidence interval
* U = upper limit of the confidence interval

Result: (For the case when σ is known)

(a) If $X_1, X_2 \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and known variance σ^2,

or

(b) If $X_1, X_2 \ldots, X_n$ is a random sample of size n from a non-normal distribution with mean μ and known variance σ^2, and if the sample size n is large $(n \geq 30)$,

then:

A $(1-\alpha)100\%$ confidence interval for μ is:

$$\overline{X} \pm Z_{1-\frac{\alpha}{2}} \sigma_{\overline{X}}$$

$$\overline{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\left(\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \, , \, \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Notes:

1. We are $(1-\alpha)100\%$ confident that the true value of μ belongs to the interval $(\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\ ,\ \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$.

2. Upper limit of the confidence interval = $\overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

3. Lower limit of the confidence interval = $\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

4. $Z_{1-\frac{\alpha}{2}}$ = Reliability Coefficient

5. $Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$ = margin of error = precision of the estimate

6. In general the interval estimate (confidence interval) may be expressed as follows:

$$\overline{X} \pm Z_{1-\frac{\alpha}{2}}\ \sigma_{\overline{X}}$$

estimator ± (reliability coefficient) × (standard Error)

estimator ± margin of error

6.3. The t-Distribution:

(Confidence Interval Using t)

Result: (For the case when σ is unknown + normal population)

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and unknown variance σ^2, then:

A $(1-\alpha)100\%$ confidence interval for μ is:

$$\bar{X} \pm t_{1-\frac{\alpha}{2}} \, \hat{\sigma}_{\bar{X}}$$

$$\bar{X} \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$\left(\bar{X} - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \, , \, \bar{X} + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

where the degrees of freedom is:

$$\mathrm{df} = \nu = n - 1.$$

Notes:

1. We are $(1-\alpha)100\%$ confident that the true value of μ belongs to the interval $\left(\overline{X} - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$.

2. $\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$ (estimate of the standard error of $\overline{X}$)

3. $t_{1-\frac{\alpha}{2}}$ = Reliability Coefficient

4. In this case, we replace σ by S and Z by t.

5. In general the interval estimate (confidence interval) may be expressed as follows:

Estimator $\pm$ (Reliability Coefficient) $\times$ (Estimate of the Standard Error)

$$\overline{X} \pm t_{1-\frac{\alpha}{2}} \hat{\sigma}_{\overline{X}}$$

Example:

Suppose that $Z \sim N(0,1)$. Find $Z_{1-\frac{\alpha}{2}}$ for the following cases:

(1) $\alpha = 0.1$ (2) $\alpha = 0.05$ (3) $\alpha = 0.01$

Solution:

(1) For $\alpha = 0.1$:

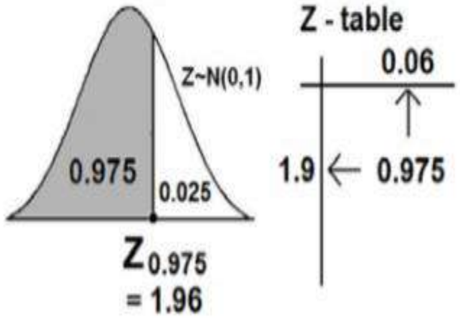
$$1-\frac{\alpha}{2} = 1-\frac{0.1}{2} = 0.95 \quad \Rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.95} = 1.645$$

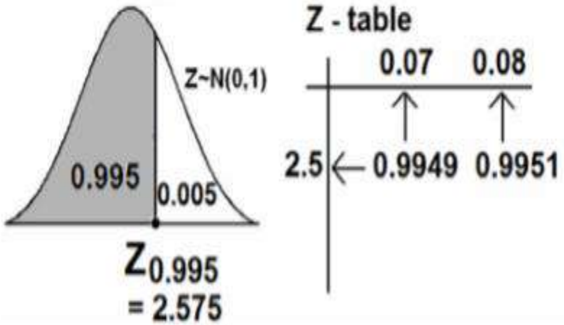
(2) For $\alpha = 0.05$:

$$1-\frac{\alpha}{2} = 1-\frac{0.05}{2} = 0.975 \quad \Rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96.$$

(3) For $\alpha = 0.01$:

$$1-\frac{\alpha}{2} = 1-\frac{0.01}{2} = 0.995 \quad \Rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.995} = 2.575.$$





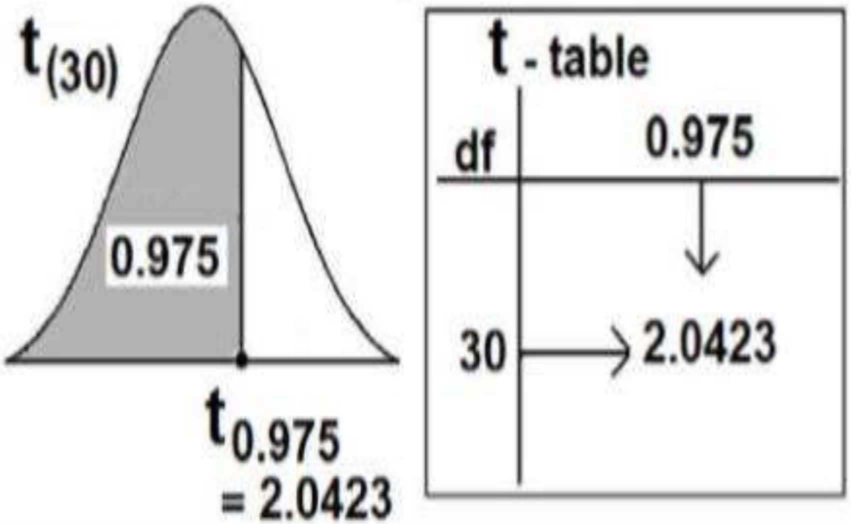
Example:

Suppose that $t \sim t(30)$. Find $t_{1-\frac{\alpha}{2}}$ for $\alpha = 0.05$.

Solution:

$df = \nu = 30$

$$1-\frac{\alpha}{2} = 1-\frac{0.05}{2} = 0.975 \quad \Rightarrow \quad t_{1-\frac{\alpha}{2}} = t_{0.975} = 2.0423$$



Example: (The case where σ^2 is known)

Diabetic ketoacidosis is a potential fatal complication of diabetes mellitus throughout the world and is characterized in part by very high blood glucose levels. In a study on 123 patients living in Saudi Arabia of age 15 or more who were admitted for diabetic ketoacidosis, the mean blood glucose level was 26.2 mmol/l. Suppose that the blood glucose levels for such patients have a normal distribution with a standard deviation of 3.3 mmol/l.

(1) Find a point estimate for the mean blood glucose level of such diabetic ketoacidosis patients.

(2) Find a 90% confidence interval for the mean blood glucose level of such diabetic ketoacidosis patients.

Solution:

σ^2 is known ($\sigma^2 = 10.89$)

X ~ Normal(μ , 10.89)

μ = ?? (unknown- we need to estimate μ)

Sample size: $n = 123$ (large)

Sample mean: $\overline{X} = 26.2$

(1) Point Estimation:

We need to find a point estimate for μ.

$\overline{X} = 26.2$ is a point estimate for μ.

$\mu \approx 26.2$

(2) Interval Estimation (Confidence Interval = C. I.):

We need to find 90% C. I. for μ.

$90\% = (1-\alpha)100\%$

$$1-\alpha = 0.9 \Leftrightarrow \alpha = 0.1 \Leftrightarrow \frac{\alpha}{2} = 0.05 \Leftrightarrow 1-\frac{\alpha}{2} = 0.95$$

The reliability coefficient is: $Z_{1-\frac{\alpha}{2}} = Z_{0.95} = 1.645$

90% confidence interval for μ is:

$$\left(\overline{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\, ,\, \overline{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

$$\left(26.2 - (1.645)\frac{3.3}{\sqrt{123}}\, ,\, 26.2 + (1.645)\frac{3.3}{\sqrt{123}}\right)$$

$$(26.2 - 0.4894714\, ,\; 26.2 + 0.4894714)$$

$$(25.710529\, ,\, 26.689471)$$

We are 90% confident that the true value of the mean μ lies in the interval $(25.71\, ,\, 26.69)$, that is:

$$25.71 < \mu < 26.69$$

Example: (The case where σ^2 is unknown)

A study was conducted to study the age characteristics of Saudi women having breast lump. A sample of 121 Saudi women gave a mean of 37 years with a standard deviation of 10 years. Assume that the ages of Saudi women having breast lumps are normally distributed.

(a) Find a point estimate for the mean age of Saudi women having breast lumps.

(b) Construct a 99% confidence interval for the mean age of Saudi women having breast lumps

Solution:

$X \sim \text{Normal}(\mu, \sigma^2)$

μ = ?? (unknown- we need to estimate μ)

σ^2 = ?? (unknown)

Sample size: $n = 121$

Sample mean: $\overline{X} = 37$

Sample standard deviation: $S = 10$

Degrees of freedom: df =ν =121 – 1= 120

(a) Point Estimation: We need to find a point estimate for μ.

$\overline{X} = 37$ is a "good" point estimate for μ.

$\mu \approx 37$ *years*

(b) Interval Estimation (Confidence Interval = C. I.): We need to find 99% C. I. for μ.

$$99\% = (1-\alpha)100\%$$

$$1-\alpha = 0.99 \Leftrightarrow \alpha = 0.01 \quad \Leftrightarrow \quad \frac{\alpha}{2} = 0.005 \quad \Leftrightarrow \quad 1-\frac{\alpha}{2} = 0.995$$

$\nu = df = 120$

The reliability coefficient is: $t_{1-\frac{\alpha}{2}} = t_{0.995} = 2.617$

99% confidence interval for μ is:

$$\overline{X} \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$37 \pm (2.617)\frac{10}{\sqrt{121}}$$

$$37 \pm 2.38$$

$$(37-2.38,\ 37+2.38)$$

$$(34.62\ ,\ 39.38)$$

6.4. Confidence Interval for the Difference between Two Population Means $(\mu_1 - \mu_2)$

Suppose that we have two populations:

- 1-st population with mean μ_1 and variance $\sigma_1{}^2$
- 2-nd population with mean μ_2 and variance $\sigma_2{}^2$
- We are interested in comparing μ_1 and μ_2, or equivalently, making inferences about the difference between the means $(\mu_1 - \mu_2)$.
- We <u>independently</u> select a random sample of size n_1 from the 1-st population and another random sample of size n_2 from the 2-nd population:
- Let $\bar{X}_1$ and S_1^2 be the sample mean and the sample variance of the 1-st sample.
- Let $\bar{X}_2$ and S_2^2 be the sample mean and the sample variance of the 2-nd sample.
- The sampling distribution of $\bar{X}_1 - \bar{X}_2$ is used to make inferences about $\mu_1 - \mu_2$.

Recall:

1. Mean of $\bar{X}_1 - \bar{X}_2$ is: $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$

2. Variance of $\bar{X}_1 - \bar{X}_2$ is: $\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

3. Standard error of $\bar{X}_1 - \bar{X}_2$ is: $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

4. If the two random samples were selected from normal distributions (or non-normal distributions with large sample sizes) with known variances σ_1^2 and σ_2^2, then the difference between the sample means $(\bar{X}_1 - \bar{X}_2)$ has a normal distribution with mean $(\mu_1 - \mu_2)$ and variance $((\sigma_1^2/n_1) + (\sigma_2^2/n_2))$, that is:

- $$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \, , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

- $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{N}(0,1)$$

Point Estimation of $\mu_1 - \mu_2$:

Result:

$\overline{X}_1 - \overline{X}_2$ is a "good" point estimate for $\mu_1 - \mu_2$.

Interval Estimation (Confidence Interval) of $\mu_1-\mu_2$:

We will consider two cases.

(i) First Case: σ_1^2 and σ_2^2 are known:

If σ_1^2 and σ_2^2 are known, we use the following result to find an interval estimate for $\mu_1-\mu_2$.

Result:

A $(1-\alpha)100\%$ confidence interval for $\mu_1-\mu_2$ is:

$$(\bar{X}_1-\bar{X}_2) \pm Z_{1-\frac{\alpha}{2}}\, \sigma_{\bar{X}_1-\bar{X}_2}$$

$$(\bar{X}_1-\bar{X}_2) \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$\left((\bar{X}_1-\bar{X}_2) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\ ,\ (\bar{X}_1-\bar{X}_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

$$(\bar{X}_1-\bar{X}_2) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Estimator ± (Reliability Coefficient) × (Standard Error)

(ii) Second Case:

Unknown equal Variances: ($\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown):

If σ_1^2 and σ_2^2 are equal but unknown ($\sigma_1^2 = \sigma_2^2 = \sigma^2$), then the pooled estimate of the common variance σ^2 is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

where S_1^2 is the variance of the 1-st sample and S_2^2 is the variance of the 2-nd sample. The degrees of freedom of S_p^2 is

$$df = \nu = n_1 + n_2 - 2.$$

We use the following result to find an interval estimate for $\mu_1 - \mu_2$ when we have normal populations with unknown and equal variances.

Result:

A $(1-\alpha)100\%$ confidence interval for $\mu_1-\mu_2$ is:

$$(\bar{X}_1-\bar{X}_2) \pm t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}$$

$$\left((\bar{X}_1-\bar{X}_2) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}} \;,\; (\bar{X}_1-\bar{X}_2) + t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}} \right)$$

where reliability coefficient $t_{1-\frac{\alpha}{2}}$ is the t-value with

$df=\nu=n_1+n_2-2$ degrees of freedom.

Example: (1[st] Case: σ_1^2 and σ_2^2 are known)

An experiment was conducted to compare time length (duration time) of two types of surgeries (A) and (B). 75 surgeries of type (A) and 50 surgeries of type (B) were performed. The average time length for (A) was 42 minutes and the average for (B) was 36 minutes.

(1) Find a point estimate for $\mu_A - \mu_B$, where μ_A and μ_B are population means of the time length of surgeries of type (A) and (B), respectively.

(2) Find a 96% confidence interval for $\mu_A - \mu_B$. Assume that the population standard deviations are 8 and 6 for type (A) and (B), respectively.

Solution:

Surgery	Type (A)	Type (B)
Sample Size	$n_A = 75$	$n_B = 50$
Sample Mean	$\bar{X}_A = 42$	$\bar{X}_B = 36$
Population Standard Deviation	$\sigma_A = 8$	$\sigma_B = 6$

(1) A point estimate for $\mu_A - \mu_B$ is:

$$\bar{X}_A - \bar{X}_B = 42 - 36 = 6.$$

(2) Finding a 96% confidence interval for $\mu_A - \mu_B$:

$\alpha = ??$

$96\% = (1-\alpha)100\% \Leftrightarrow 0.96 = (1-\alpha) \Leftrightarrow \alpha=0.04 \Leftrightarrow \alpha/2 = 0.02$

Reliability Coefficient: $Z_{1-\frac{\alpha}{2}} = Z_{0.98} = 2.055$

A 96% C.I. for $\mu_A - \mu_B$ is:

$$(\bar{X}_A - \bar{X}_B) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$6 \pm Z_{0.98} \sqrt{\frac{8^2}{75} + \frac{6^2}{50}}$$

$$6 \pm (2.055) \sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$6 \pm 2.578$$

$$3.422 < \mu_A - \mu_B < 8.58$$

Example: (2nd Case: $\sigma_1^2 = \sigma_2^2$ unknown)

To compare the time length (duration time) of two types of surgeries (A) and (B), an experiment shows the following results based on two independent samples:

Type A: 140, 138, 143, 142, 144, 137

Type B: 135, 140, 136, 142, 138, 140

(1) Find a point estimate for $\mu_A - \mu_B$, where μ_A (μ_B) is the mean time length of type A (B).

(2) Assuming normal populations with equal variances, find a 95% confidence interval for $\mu_A - \mu_B$.

Solution:

First we calculate the mean and the variances of the two samples, and we get:

Surgery	Type (A)	Type (B)
Sample Size	$n_A = 6$	$n_B = 6$
Sample Mean	$\bar{X}_A = 140.67$	$\bar{X}_B = 138.50$
Sample Variance	$S^2_A = 7.87$	$S^2_B = 7.10$

(1) A point estimate for $\mu_A - \mu_B$ is:

$$\bar{X}_A - \bar{X}_B = 140.67 - 138.50 = 2.17.$$

(2) Finding 95% Confidence interval for $\mu_A - \mu_B$:

$95\% = (1-\alpha)100\% \Leftrightarrow 0.95 = (1-\alpha) \Leftrightarrow \alpha = 0.05 \Leftrightarrow \alpha/2 = 0.025$

.df = $\nu = n_A + n_B - 2 = 10$

Reliability Coefficient: $t_{1-\frac{\alpha}{2}} = t_{0.975} = 2.228$

$$S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

$$= \frac{(6-1)(7.87) + (6-1)(7.1)}{6+6-2} = 7.485$$

A 95% C.I. for $\mu_A - \mu_B$ is:

$$(\bar{X}_A - \bar{X}_B) \pm t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_p^2}{n_A} + \frac{S_p^2}{n_B}}$$

$$2.17 \pm (2.228) \sqrt{\frac{7.485}{6} + \frac{7.485}{6}}$$

$$2.17 \pm 3.519$$

$$-1.35 < \mu_A - \mu_B < 5.69$$

Note: Since the confidence interval includes zero, we conclude that the two population means may be equal ($\mu_A - \mu_B = 0 \Leftrightarrow \mu_A = \mu_B$). Therefore, we may conclude that the mean time length is the same for both types of surgeries.

6.5. Confidence Interval for a Population Proportion (p)

Recall:

For the population:

$N(A)$ = number of elements in the population with a specified characteristic "A"

N = total number of elements in the population (population size)

The population proportion is:

$$p = \frac{N(A)}{N}$$ (p is a parameter)

Recall:

For the sample:

$n(A)=$ number of elements in the sample with the same characteristic "A"

n = sample size

The sample proportion is:

$$\hat{p}=\frac{n(A)}{n} \qquad (\hat{p} \text{ is a statistic})$$

Recall:

. The sampling distribution of the sample proportion ($\hat{p}$) is used to make inferences about the population proportion (p).

. The mean of ($\hat{p}$) is: $\mu_{\hat{p}} = p$

. The variance of ($\hat{p}$) is: $\sigma^2_{\hat{p}} = \dfrac{p(1-p)}{n}$

. The standard error (standard deviation) of ($\hat{p}$) is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

. For large sample size ($n \geq 30, np > 5, n(1-p) > 5$), the sample proportion ($\hat{p}$) has approximately a normal distribution with mean $\mu_{\hat{p}} = p$ and a variance $\sigma^2_{\hat{p}} = p(1-p)/n$, that is:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \qquad \text{(approximately)}$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1) \qquad \text{(approximately)}$$

(i) Point Estimate for (p):

Result:

A good point estimate for the population proportion (p) is the sample proportion ($\hat{p}$).

(ii) Interval Estimation (Confidence Interval) for (p):

Result:

For large sample size ($n \geq 30, np > 5, n(1-p) > 5$), an approximate $(1-\alpha)100\%$ confidence interval for (p) is:

$$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\left(\hat{p} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \,, \quad \hat{p} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Estimator ± (Reliability Coefficient) × (Standard Error)

Example:

In a study on the obesity of Saudi women, a random sample of 950 Saudi women was taken. It was found that 611 of these women were obese (overweight by a certain percentage).

(1) Find a point estimate for the true proportion of Saudi women who are obese.

(2) Find a 95% confidence interval for the true proportion of Saudi women who are obese.

Solution:

$n = 950$ (950 women in the sample)

$n(A) = 611$ (611 women in the sample who are obese)

The sample proportion (the proportion of women who are obese in the sample.) is:

$$\hat{p} = \frac{n(A)}{n} = \frac{611}{950} = 0.643$$

(1) A point estimate for p is: $\hat{p} = 0.643$.

(2) We need to construct 95% C.I. for the proportion (p).

$$95\% = (1-\alpha)100\% \Leftrightarrow 0.95 = 1-\alpha \Leftrightarrow \alpha = 0.05 \Leftrightarrow \frac{\alpha}{2} = 0.025 \Leftrightarrow 1-\frac{\alpha}{2} = 0.975$$

The reliability coefficient: $Z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$.

A 95% C.I. for the proportion (p) is:

$$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.643 \pm (1.96)\sqrt{\frac{(0.643)(1-0.643)}{950}}$$

$$0.643 \pm (1.96)(0.01554)$$

$$0.643 \pm 0.0305$$

$$(0.6127\ ,\ 0.6735)$$

We are 95% confident that the true value of the population proportion of obese women, p, lies in the interval $(0.61, 0.67)$, that is:

$$0.61 < p < 0.67$$