Chapter 6
Using sample data to make estimates about population parameters (P162-172)

- **Key words:**

- Point estimate, interval estimate, estimator, Confident level ,α , Confident interval for mean μ, Confident interval for two means, Confident interval for population proportion P, Confident interval for two proportions

- **6.1 Introduction:**
- Statistical inference is the procedure by which we reach to a conclusion about a population on the basis of the information contained in a sample drawn from that population.
- **Suppose that:**
- an administrator of a large hospital is interested in the mean age of patients admitted to his hospital during a given year.

1. It will be too expensive to go through the records of all patients admitted during that particular year.
2. He consequently elects to examine a sample of the records from which he can compute an estimate of the mean age of patients admitted to his that year.

- To any parameter, we can compute two types of estimate: a point estimate and an interval estimate.
- A ***point estimate*** is a single numerical value used to estimate the corresponding population parameter.
- An ***interval estimate*** consists of two numerical values defining a range of values that, with a specified degree of confidence, we feel includes the parameter being estimated.
- ***The Estimate and The Estimator:***
- The estimate is a single computed value, but the estimator is the rule that tell us how to compute this value, or estimate.
- **For example,** $$\bar{x} = \sum_i x_i$$
- is an estimator of the population mean, μ. The single numerical value that results from evaluating this formula is called an estimate of the parameter μ.

6.2 Confidence Interval for a Population Mean: (C.I)

Suppose researchers wish to estimate the mean of some normally distributed population.

- They draw a random sample of size n from the population and compute , which they use as a point estimate of μ.
- Because random sampling involves chance, then $\bar{x}$ can't be expected to be equal to μ.
- The value of $\bar{x}$ may be greater than or less than μ.
- It would be much more meaningful to estimate μ by an interval.

The 1-α percent confidence interval (C.I.) for μ:

- We want to find two values L and U between which μ lies with high probability, i.e.

$$P(L \leq \mu \leq U) = 1-\alpha$$

For example:

- **When,**
- $\alpha = 0.01$,

  then $1-\alpha =$
- $\alpha = 0.05$,

  then $1-\alpha =$
- $\alpha = 0.05$,

  then $1-\alpha =$

We have the following cases

a) When the population is normal

1) When the **variance is known** and the **sample size is large or small,** the C.I. has the form:

- $P(\bar{X} - Z_{(1-\alpha/2)}\, \sigma/\sqrt{n} < \mu < \bar{X} + Z_{(1-\alpha/2)}\, \sigma/\sqrt{n}) = 1-\alpha$

2) When **variance is unknown,** and the **sample size is small,** the C.I. has the form:

$$\mathbf{P(\bar{X} - t_{(1-\alpha/2),n-1}\, s/\sqrt{n} < \mu < \bar{X} + t_{(1-\alpha/2),n-1}\, s/\sqrt{n}) = 1-\alpha}$$

b) When the population is not normal and n large (n>30)

1) When the **variance is known** the C.I. has the form:

$$P(\bar{x} - Z_{(1-\alpha/2)}\, \sigma / \sqrt{n} < \mu < \bar{x} + Z_{(1-\alpha/2)}\, \sigma / \sqrt{n}) = 1-\alpha$$

2) When **variance is unknown,** the C.I. has the form:

$$P(\bar{x} - Z_{(1-\alpha/2)}\, s / \sqrt{n} < \mu < \bar{x} + Z_{(1-\alpha/2)}\, s / \sqrt{n}) = 1-\alpha$$

Example 6.2.1 Page 167:

- Suppose a researcher , interested in obtaining an estimate of the average level of some enzyme in a certain human population, takes a sample of 10 individuals, determines the level of the enzyme in each, and computes a sample mean of approximately $\bar{x} = 22$ Suppose further it is known that the variable of interest is approximately normally distributed with a variance of 45. We wish to estimate μ. **(α=0.05)**

Solution:

- $1-\alpha=0.95 \rightarrow \alpha=0.05 \rightarrow \alpha/2=0.025$, $\bar{x}=22$
- variance $= \sigma^2 = 45 \rightarrow \sigma=\sqrt{45}, n=10$
- 95%confidence interval for μ is given by:

$$P(\bar{X} - Z_{(1-\alpha/2)}\,\sigma/\sqrt{n} < \mu < \bar{X} + Z_{(1-\alpha/2)}\,\sigma/\sqrt{n}) = 1-\alpha$$

- $Z_{(1-\alpha/2)} = Z_{0.975} = 1.96$ (refer to table D)
- $Z_{0.975}(\sigma/\sqrt{n}) = 1.96\,(\sqrt{45}/\sqrt{10}) = 4.1578$
- $22 \pm 1.96\,(\sqrt{45}/\sqrt{10}) \rightarrow$
- $(22-4.1578,\ 22+4.1578) \rightarrow (17.84,\ 26.16)$
- **Exercise example 6.2.2 page 169**

Example

The activity values of a certain enzyme measured in normal gastric tissue of 35 patients with gastric carcinoma has a mean of 0.718 and a standard deviation of 0.511.We want to construct a 90 % confidence interval for the population mean.

- **Solution:**

- Note that the population is not normal,
- n=35 (n>30) n is large and σ is unknown ,s=0.511
- 1- α=0.90$\rightarrow$ α=0.1
- $\rightarrow$ $\alpha/2$=0.05$\rightarrow$ 1-$\alpha/2$=0.95,

Then 90% confident interval for μ is given by :

$$P(\bar{X} - Z_{(1-\alpha/2)}\, s/\sqrt{n} < \mu < \bar{x} + Z_{(1-\alpha/2)}\, s/\sqrt{n}) = 1-\alpha$$

- $Z_{(1-\alpha/2)} = Z_{0.95} = 1.645$ (refer to table D)
- $Z_{0.95}(s/\sqrt{n}) = 1.645\,(0.511/\sqrt{35}) = 0.1421$

  $0.718 \pm 1.645\,(0.511)/\sqrt{35} \rightarrow$

  $(0.718-0.1421,\ 0.718+0.1421) \rightarrow$

  $(0.576, 0.860)$.
- **<u>Exercise example 6.2.3 page 164:</u>**

Example6.3.1 Page 174:

- Suppose a researcher , studied the effectiveness of early weight bearing and ankle therapies following acute repair of a ruptured Achilles tendon. One of the variables they measured following treatment the muscle strength. In 19 subjects, the mean of the strength was 250.8 with standard deviation of 130.9

  we assume that the sample was taken from is approximately normally distributed population. Calculate 95% confident interval for the mean of the strength ?

Solution:

- $1-\alpha=0.95 \rightarrow \alpha=0.05 \rightarrow \alpha/2=0.025$, $\bar{x} = 250.8$
- Standard deviation= S = 130.9 ,n=19
- 95%confidence interval for μ is given by:

$$P(\bar{X} - t_{(1-\alpha/2),n-1}\, s/\sqrt{n} < \mu < \bar{X} + t_{(1-\alpha/2),n-1}\, s/\sqrt{n}) = 1-\alpha$$

- $t_{(1-\alpha/2),n-1} = t_{0.975,18} = 2.1009$ (refer to table E)
- $t_{0.975,18}(s/\sqrt{n}) = 2.1009\ (130.9 / \sqrt{19}) = 63.1$
- $250.8 \pm 2.1009\ (130.9 / \sqrt{19}) \rightarrow$
- $(250.8- 63.1\ ,\ 22+63.1) \rightarrow (187.7, 313.9)$
- **Exercise 6.2.1 ,6.2.2**
- **6.3.2 page 171**

6.3 Confidence Interval for the difference between two Population Means: (C.I)

If we draw two samples from two independent population and we want to get the confident interval for the difference between two population means , then we have the following cases :

a) When the population is normal

1) When the **variance is known** and the **sample sizes is large or small,** the C.I. has the form:

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

2) When **<u>variances are unknown but equal,</u>** and the **<u>sample size is small</u>,** the C.I. has the form:

$$(\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2},(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2},(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

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

a) When the population is normal

1) When the **variance is known** and the **sample sizes is large or small,** the C.I. has the form:

$$(\bar{x}_1 - \bar{x}_2) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example 6.4.1 P174:

The researcher team interested in the difference between serum uric and acid level in a patient with and without Down's syndrome .In a large hospital for the treatment of the mentally retarded, a sample of 12 individual with Down's Syndrome yielded a mean of $\bar{x}_1 = 4.5$ mg/100 ml. In a general hospital a sample of 15 normal individual of the same age and sex were found to have a mean value of $\bar{x}_2 = 3.4$ If it is reasonable to assume that the two population of values are normally distributed with variances equal to 1 and 1.5,find the 95% C.I for $\mu_1 - \mu_2$

Solution:

$1-\alpha=0.95 \rightarrow \alpha=0.05 \rightarrow \alpha/2=0.025 \rightarrow Z_{(1-\alpha/2)} = Z_{0.975} = 1.96$

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (4.5 - 3.4) \pm 1.96 \sqrt{\frac{1}{12} + \frac{1.5}{15}}$$

$1.1 \pm 1.96(0.4282) = 1.1 \pm 0.84 = (0.26 , 1.94)$ ■

Example 6.4.1 P178:

The purpose of the study was to determine the effectiveness of an integrated outpatient dual-diagnosis treatment program for mentally ill subject. The authors were addressing the problem of substance abuse issues among people with sever mental disorder. A retrospective chart review was carried out on 50 patient ,the recherché was interested in the number of inpatient treatment days for physics disorder during a year following the end of the program. Among 18 patient with schizophrenia, The mean number of treatment days was 4.7 with standard deviation of 9.3. For 10 subject with bipolar disorder, the mean number of treatment days was 8.8 with standard deviation of 11.5. We wish to construct 99% C.I for the difference between the means of the populations Represented by the two samples

Solution :

- $1-\alpha = 0.99 \rightarrow \alpha = 0.01 \rightarrow \alpha/2 = 0.005 \rightarrow 1-\alpha/2 = 0.995$
- $n_1+n_2-2 = 18 + 10 - 2 = 26$
- $t_{(1-\alpha/2),(n_1+n_2-2)} = t_{0.995,26} = 2.7787$, then 99% C.I for $\mu_1 - \mu_2$

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

where

- $$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(17x9.3^2)+(9x11.5^2)}{18+10-2} = 102.33$$

then

$(4.7-8.8) \pm 2.7787 \sqrt{102.33} \sqrt{(1/18)+(1/10)}$

- $4.1 \pm 11.086 = (-15.186, 6.986)$

Exercises: 6.4.2 , 6.4.6, 6.4.7, 6.4.8 Page 180