324 Stat
Lecture Notes

(7 One- and Two-Sample Estimation Problem)

(Book*: Chapter 8 ,pg265)

Probability& Statistics for Engineers & Scientists
By Walpole, Myers, Myers, Ye

Estimation

- <u>Point estimate:</u>

- Is a single numerical value to estimate parameter
- <u>Example :</u>

$$\bar{X} = \sum \frac{X_i}{n}$$

$$\hat{P} = \frac{X}{n}$$

$$S^2 = \frac{\sum (X_i - \bar{X})}{n-1}$$

- <u>Interval estimate</u>

- Is two numerical values to estimate parameter
- <u>It means to be in an interval s.a</u>

$$(\theta_L < \theta < \theta_U)$$

- Lower bound (L)　　　upper bound (U)

Interval estimation

- An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$ where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$. Since different samples will generally yield different values of $\hat{\Theta}$ and therefore different values of $\hat{\theta}_L$ and $\hat{\theta}_U$. From the sampling distribution of $\hat{\Theta}$ we shall be able to determine $\hat{\theta}_L$ and $\hat{\theta}_U$ such that the $P(\hat{\theta}_L < \theta < \hat{\theta}_U)$ is equal to any positive fractional value we care to specify. If for instance we find $\hat{\theta}_L$ and $\hat{\theta}_U$ such that:

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha \qquad for \quad 0 < \alpha < 1$$

then we have a probability of $(1-\alpha)$ of selecting a random sample that will produce an interval containing θ .

- The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ computed from the selected sample, is then called a $(1-\alpha)100\%$ confidence interval.

- The fraction $(1-\alpha)$ is called confidence coefficient or the degree of confidence.

- The end points $\hat{\theta}_L$ and $\hat{\theta}_U$ are called the lower and upper confidence limits.

- For Example:

when $\alpha = 0.05$ we have a 95% confidence interval and so on, that we are 95% confident that θ is between $\hat{\theta}_L$, $\hat{\theta}_U$

9.4 Single Sample: Estimating the Mean:

1-Confidence Interval on μ (σ^2 Known):

If $\bar{X}$ is the mean of a random sample of size **n** from a population with known variance σ^2 , a $(1-\alpha)100\%$ confidence interval for μ is given by:

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

where $Z_{1-\alpha/2}$ is the -value leaving an area of $\frac{\alpha}{2}$ to the right■

$$\hat{\theta}_L = \bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad , \quad \hat{\theta}_U = \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (2)$$

EX (1):

The mean of the quality point averages of a random sample of **36** college seniors is calculated to be **2.6**.

Find the 95% confidence intervals for the mean of the entire senior class. Assume that the population standard deviation is **0.3**.

Solution:

$$n = 36 \;, \bar{X} = 2.6 \;, \; \sigma = 0.3$$

95% confidence interval for the mean μ :

$$1-\alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \rightarrow Z_{1-\frac{\alpha}{2}} = 1.96$$

See Ex 9.2 pg 271

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

$$2.6 \pm 1.96(\frac{0.3}{\sqrt{36}}) \rightarrow 2.6 \pm 0.098$$

$$2.502 < \mu < 2.698$$

Thus , we have 95% confident that μ lies between 2.502 and 2.698

Theorem 9.1:

If $\bar{X}$ is used as an estimate of μ , we can be $(1-\alpha)100\%$ confident that the error will not be exceed $Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

For example (1): e= (1.96) (0.3/6)=0.098 or

Theorem (2):

If $\bar{X}$ is used as an estimate of μ , we can be $(1-\alpha)100\%$ confident that the error will not exceed a specified amount, e when the sample size is:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} \sigma}{e} \right)^2 \qquad (3)$$

The fraction of **n** is rounded up to next whole number.

EX (2):

How large a sample is required in Ex. (1) if we want to be 95% confident that our estimate of μ is off by less than **0.05?**

Solution

$$\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \rightarrow Z_{1-\frac{\alpha}{2}} = 1.96,$$

$$\sigma = 0.3 \quad , \quad e = 0.05$$

$$n = \left(\frac{(1.96)(0.3)}{0.05} \right)^2 = 138.2976 \approx 138$$

n is rounded up to whole number.

2- Confidence Interval of when σ^2 is Unknown n<30 :

If $\overline{X}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\ \%$ confidence interval for μ is given by:

$$\overline{X} - t_{\frac{\alpha}{2},\, n-1} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2},\, n-1} \frac{S}{\sqrt{n}} \qquad (4)$$

where $t_{\frac{\alpha}{2}}$ is the t-value with **n-1** degrees of freedom leaving an area of $\frac{\alpha}{2}$ to the right.

Ex 9.5 pg 275

The contents of 7 similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10, 10.2, 9.6 liters. Find a 95% confidence interval for the mean of all such containers assuming an approximate normal distribution.

Solution:

confidence interval for the mean :

$$n = 7 \quad , \quad \bar{X} = 10 \quad , \quad S = 0.283 \quad ,$$

$$at: 1-\alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \rightarrow t_{\frac{\alpha}{2}, n-1} = t_{0.025,6} = 2.447$$

$$\bar{X} \pm t_{\frac{\alpha}{2}, n-1}\left(\frac{S}{\sqrt{n}}\right) \Rightarrow 10 \pm (2.447)\left(\frac{0.283}{\sqrt{7}}\right) \Rightarrow 10 \pm 0.262$$

$$9.738 < \mu < 10.262 \rightarrow P(9.738 < \mu < 10.262) = 0.95$$

9:4 Two Samples: Estimating the Difference between Two Means:

1- Confidence Interval for $\mu_1 - \mu_2$ when σ_1^2 and σ_2^2 Known:

If $\bar{X}_1$ *and* $\bar{X}_2$ are the means of independent random samples of size **n_1** and **n_2** from populations with known variances σ_1^2 *and* σ_2^2 respectively, a $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (5)$$

where $Z_{1-\frac{\alpha}{2}}$ is the -value leaving an area of $\frac{\alpha}{2}$ to the right .

EX (4):

A standardized chemistry test was given to **50** girls and **75** boys. The girls made an average grade of **76**, while the boys made an average grade of **82**. Find a 96% confidence interval for the difference $\mu_1 - \mu_2$ where μ_1 is the mean score of all boys and μ_2 is the mean score of all girls who might take this test. Assume that the population standard deviations are **6** and **8** for girls and boys respectively.

Solution:

girls	Boys
$n_1 = 50$	$n_2 = 75$
$\bar{X}_1 = 76$	$\bar{X}_2 = 82$
$\sigma_1 = 6$	$\sigma_2 = 8$

96% confidence interval for the mean $\mu_1 - \mu_2$:

$$1-\alpha = 0.94 \rightarrow \alpha = 0.04 \rightarrow \frac{\alpha}{2} = 0.02 \rightarrow Z_{1-\frac{\alpha}{2}} = 2.05$$

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

$$(82-76) \pm (2.05)\sqrt{\frac{36}{50} + \frac{64}{75}} \Rightarrow 6 \pm 2.571$$

$$3.429 < \mu_1 - \mu_2 < 8.571 \rightarrow P(3.429 < \mu_1 - \mu_2 < 8.571) = 0.96$$

See Ex 9.10 pg 286

2-Confidence Interval for $\mu_1 - \mu_2$ when σ_1^2 and σ_2^2 Unknown but equal variances:

If $\bar{X}_1$ and $\bar{X}_2$ are the means of independent random samples of size $\mathbf{n_1}$ and $\mathbf{n_2}$ respectively from approximate normal populations with unknown but equal variances, a $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2},v} \, S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (6)$$

, where

$$S_P = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \qquad (7)$$

is the pooled estimate of the population standard deviation and $t_{\frac{\alpha}{2},v}$ is the **t** – value with $v = n_1 + n_2 - 2$ degrees of freedom leaving an area of $\frac{\alpha}{2}$ to the right.

EX (9.11- pg 288):

The independent sampling stations were chosen for this study, one located down stream from the acid mine discharge point and the other located upstream. For **12** monthly samples collected at the down stream station the species diversity index had a mean value $\bar{X}_1 = 3.11$ and a standard deviation $S_1 = 0.771$ while **10** monthly samples had a mean index value $\bar{X}_2 = 2.04$ and a standard deviation $S_2 = 0.448$. Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.

Solution:

Station 1	Station 2
$n_1 = 12$	$n_2 = 10$
$\bar{X}_1 = 3.11$	$\bar{X}_2 = 2.04$
$S_1 = 0.771$	$S_2 = 0.448$

90% confidence interval for the mean $\mu_1 - \mu_2$:

$$S_P = \sqrt{\frac{11(0.771)^2 + 9(0.448)^2}{12+10-2}} = 0.646$$

$$at\ \ 1-\alpha = 0.90 \rightarrow \alpha = 0.1 \rightarrow \frac{\alpha}{2} = 0.05 \rightarrow t_{\frac{\alpha}{2}, n_1+n_2-2} \rightarrow t_{0.05,20} = 1.725$$

$$(3.11 - 2.04) \pm (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} \Rightarrow 1.07 \pm 0.477$$

$$0.593 < \mu_1 - \mu_2 < 1.547 \rightarrow P(0.593 < \mu_1 - \mu_2 < 1.547) = 0.90$$

9:10 Single Sample Estimating a Proportion:

Large – Sample Confidence Interval for P:

If $\hat{p}$ is the proportion of successes in a random sample of size n and an approximate $(1-\alpha)100\%$ confidence interval for the binomial parameter is given by:

$$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad (8)$$

$$\hat{q} = 1 - \hat{p}$$

Where $Z_{1-\frac{\alpha}{2}}$ is the Z- value leaving an area of $\frac{\alpha}{2}$ to the right.

EX(6):

A new rocket – launching system is being considered for deployment of small, short – rang rockets. The existing system has $p=0.8$ as the probability of a successful launch. A sample of **40** experimental launches is made with the new system and **34** are successful. Construct a 95% confidence interval for p

Solution:

a 95% confidence interval for *p* .

$$n = 40 \quad , \quad \hat{p} = \frac{34}{40} = 0.85 \quad , \quad \hat{q} = 0.15$$

$$at \quad 1-\alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \rightarrow Z_{1-\frac{\alpha}{2}} = 1.96$$

$$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \rightarrow 0.85 \pm (1.96)\sqrt{\frac{(0.85)(0.15)}{40}} \rightarrow 0.85 \pm (0.111)$$

$$0.739 < p < 0.961 \rightarrow P(0.739 < p < 0.961) = 0.95$$

See Ex 9.14
pg 297

Theorem 3:

If $\hat{p}$ is used as an estimate of *p* we can be $(1-\alpha)100\%$

confident that the error will not exceed $e = Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$.

EX 7:

In Ex. 7, find the error of p .

Solution:

The error will not exceed the following value:

$$e = Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = (1.96)\sqrt{\frac{(0.85)(0.15)}{40}} = 0.111$$

Theorem :

If $\hat{p}$ is used as an estimate of p we can be $(1-\alpha)100\%$ confident that the error will be less than a specified amount **e** when the sample size is approximately:

$$n = \frac{Z^{2}_{1-\alpha/2} \quad \hat{p}\hat{q}}{e^{2}} \qquad (9)$$

Then the fraction of **n** is rounded up.

EX(8):

How large a sample is required in Ex. 7 if we want to be

95% confident that our estimate of *p* is within **0.02?**

Solution:

$$e = 0.02 \quad , \quad Z_{1-\frac{\alpha}{2}} = 1.96 \quad , \quad \hat{p} = 0.85 \quad , \quad \hat{q} = 0.15$$

$$n = \frac{(1.96)^2(0.85)(0.15)}{(0.02)^2} = 1224.51 \approx 1225$$

Two Samples: Estimating the difference between two proportions

Large- Sample confidence interval for p1-p2

If $\hat{p}_1$ and $\hat{p}_2$ are the two proportions of successes in random samples of sizes n1 and n2 respectively, $\hat{q}_1$=1-$\hat{p}_1$ and $\hat{q}_2$=1-$\hat{p}_2$, an approximate $(1-\alpha)100\%$ confidence interval for the difference of two binomial parameters , *p1-p2* is given by

$$(\hat{p}_1\text{-}\hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

where $Z_{1-\frac{\alpha}{2}}$ is the -value leaving an area of $\alpha/2$ to the right .

- Ex 9.17 pg 301

A certain change n the process for manufacturing component parts is being considered. Samples are taken under both the existing and the new process results in an improvement.

If **75 of 1500** items from the **existing process** are found to be defective, and **80 of 2000** items from the **new process** found to be defective, find a 90% confidence interval for the rue difference in the proportions of defectives for the existing and new process respectively.

- Solution

Let p1 and p2 be the true proportion of defectives for the existing and new process respectively.

$\hat{p}_1$=75/1500=0.05

$\hat{p}_2$=80/2000=0.04

$\hat{p}_1$-$\hat{p}_2$=0.05-0.04=0.01

$Z_{1-\frac{\alpha}{2}}$ =1.645

The 90% confidence interval is

$$0.01 \pm (1.645)\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}}$$

(-0.0017, 0.0217)