Chapter 5 Estimation and Confidence Interval

Sampling Theory

- ⇨ A study of relationships between a population and samples that drawn from the population
- ⇨ It is useful in the following aspect of statistics
  - ♦ Estimation of unknown population parameters (e.g. population mean, variance) from a knowledge of corresponding sample statistics
  - ♦ Significance and hypothesis testing which are important in the theory of decisions

Sampling Distribution

- ⇨ Consider ALL possible samples of size *n* which can be drawn from a given population
- ⇨ For each sample, we can compute a statistic (mean, variance, etc.) which will vary from sample to sample
- ⇨ We obtain a distribution of statistic which is called sampling distribution

Examples of sampling distribution:

1. sampling distribution of sample mean
2. sampling distribution of sample proportion

Sampling distribution of sample means from a Normal population

If $X_1, X_2, \cdots, X_n$ is a random sample of size *n* taken from a Normal distribution with mean μ and variance σ^2 such that $X \sim N(\mu, \sigma^2)$, then the distribution of $\bar{X}$ is also Normal and $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ where $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$.

Note: $\frac{\sigma}{\sqrt{n}}$ is known as the standard error of the mean

Example 1:
A random sample of size 15 is taken from a Normal distribution with mean 60 and standard deviation 4. Find the probability that the mean of the sample is less than 58.

Solution:

Given X ~ N(   ). Then, for samples of size 15, $\overline{X} \sim N\left(\quad , \frac{}{} \right)$

$P(\overline{X} < 58) =$

Example 2:
Long-distance telephone calls are normally distributed with μ = 8 minutes and σ = 2 minutes. If random samples of 25 calls were selected, what proportion of the sample means would be between

a) 7.8 and 8.2 minutes?

b) 7.5 and 8 minutes?

Solution:
Let X $\equiv$ long-distance telephone calls
Given $X \sim N(8, 2^2)$

a) $P(7.8 < \overline{X} < 8.2) =$

b) $P(7.5 < \overline{X} < 8) =$

<u>Sampling distribution of sample means from any population</u>

<u>The Central Limit Theorem</u>

If $X_1, X_2, \cdots, X_n$ is a random sample of size *n* from ANY distribution with mean μ and variance σ^2, then for LARGE *n*, the distribution of sample mean $\overline{X}$ is approximately Normal and $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ where $\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$.

Note:
The sample size is usually considered to be large if $n \geq 30$

Example 3:
If a random sample of size 30 is taken from each of the following distribution, find, for each case, the probability that the sample mean exceeds 5.

a) $X \sim P_o(4.5)$

b) X ~ B(9, 0.5)

Solution:

a) $X \sim P_o(4.5)$

By the Central Limit Theorem, $\overline{X} \sim N(\quad , \quad)$ approximately

$P(\overline{X} > 5) =$

b) X ~ B(9, 0.5)

Mean = *np* =

Variance = *np(1 – p)* =

By the Central Limit Theorem, $\overline{X} \sim N(\quad)$ approximately

$P(\overline{X} > 5) =$

<u>Sampling distribution of sample proportion</u>

Consider a population in which *p* is the proportion of 'successes'. If a random sample of size *n* is taken from this population and X is the random variable 'the number of successes in the sample', then X ~ B(*n*,*p*) and for large *n*, $X \sim N(np, np(1-p))$ approximately.

Denote P_s is the random variable 'the proportion of successes in a sample'.

⇨ $P_s = \dfrac{X}{n}$

⇨ Mean of P_s is *p* and variance of P_s is $\dfrac{p(1-p)}{n}$.

⇨ $P_s \sim N\left(p, \dfrac{p(1-p)}{n}\right)$

Note:

1. The larger the sample size, the better the approximation
2. The distribution of P_s is known as sampling distribution of proportion
3. $\sqrt{\dfrac{p(1-p)}{n}}$ is known as the standard error of proportion

Estimation of parameters

⇨ The statistical technique of estimating unknown population parameters based on a value of the corresponding sample statistic

The estimation procedure involves the following steps.

1. Select a sample
2. Collect the required information from the members of the sample
3. Calculate the value of the sample statistic
4. Assign value(s) to the corresponding population parameter

Estimate

⇨ The value(s) assigned to a population parameter based on the value of a sample statistic

Estimator

⇨ The sample statistic that is used to estimate a population parameter

Two types of estimates

1. Point estimate
   - ⇨ The value (single number) of a sample statistic that is used to estimate a population parameter
   - ⇨ Example: $\hat{\mu} = \bar{x} = 77$ and $\hat{\sigma}^2 = s^2 = 6$

2. Interval estimate (Confidence Intervals)
   - ⇨ An estimate of a population parameter given by two numbers between which the parameter may be considered to lie on
   - ⇨ Example: $66 \le \mu \le 88$

Consider a population with unknown parameter θ. If we can find an interval (a, b) such that P(a < θ < b) = 0.95, we say that (a, b) is a 95% confidence interval for θ. In this case, 0.95 is the probability that the interval includes θ.

- The 100(1-α)% confidence interval for the population mean, μ when the population variance σ^2 is known is given by

$$\boxed{\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}} \quad \text{or} \quad \boxed{\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

Example 4:

To determine the mean waiting time for his customers, a bank manager took a random sample of 50 customers and found that the mean waiting time was 7.2 minutes. Assuming that the population standard deviation is known to be 5 minutes, find the 90% confidence interval of the mean waiting time for all of the bank's customers.

Solution:

$n =$; $\bar{x} =$; $\sigma =$; $\alpha =$; $Z_{\frac{\alpha}{2}} = Z_{0.05} =$

The 90% confidence interval of the mean waiting time for all of the bank's customers is

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} =$$

Example 5:

In a random sample of 70 students in a large university, a dean found that the mean weekly time spent doing homework was 14.3 hours. If we assume that homework time is normally distributed with a standard deviation of 4.0 hours, find the 99% confidence interval estimate of the weekly time spent doing homework for all the university's students.

Solution:

$n =$; $\bar{x} =$; $\sigma =$; $\alpha =$; $Z_{\frac{\alpha}{2}} = Z_{0.005} =$

The 99% confidence interval estimate of the weekly time spent doing homework for all the university's students is given by

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} =$$

- The 100(1-α)% confidence interval for the population mean, μ when the population variance σ^2 is unknown and the sample size n is large $(n \geq 30)$ is given by

$$\boxed{\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}} \text{ or } \boxed{\bar{x} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}}$$

Example 6:

Measurements of the diameters of a random sample of 200 ball bearings made by a certain machine during 1 week showed a mean of 8.24 mm and a standard deviation of 0.42 mm. Find the 95% and 99% confident interval for the mean diameter of all the ball bearings.

Solution:

$n =$; $\overline{x} =$; $s =$

$\alpha =$; $Z_{\frac{\alpha}{2}} = Z_{0.025} =$ (from Table 4)

The 95% confident interval for the mean diameter of all the ball bearings is

$$\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} =$$

$\alpha =$; $Z_{\frac{\alpha}{2}} = Z_{0.005} =$

The 99% confident interval for the mean diameter of all the ball bearings is

$$\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} =$$

Example 7:

A random sample of 35 drums of a wax-base floor cleaner has a standard deviation of 12 pounds and a mean weight of 240 pounds. Construct a 95% confidence interval for the actual mean weight of all these drums.

Solution:

$n =$; $\overline{x} =$; s = $\alpha =$; $Z_{\frac{\alpha}{2}} = Z_{0.025} =$ (from Table 4)

The 95% confidence interval for the actual mean weight of all these drums is

$$\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} =$$

- The 100(1-α)% confidence interval for the true population proportion, π is given by

$$\boxed{p_s \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p_s(1-p_s)}{n}}} \quad \text{or} \quad \boxed{p_s - Z_{\frac{\alpha}{2}}\sqrt{\frac{p_s(1-p_s)}{n}} < \pi < p_s + Z_{\frac{\alpha}{2}}\sqrt{\frac{p_s(1-p_s)}{n}}}$$

Example 8:

A manufacturer wants to assess the proportion of defective items in a large batch produced by a particular machine. He tests a random sample of 300 items and finds that 45 are defective. Calculate a 98% confidence interval for the proportion of defective items in the complete batch.

Solution:

$$p_s = \frac{x}{n} = \qquad ; \quad \alpha = \qquad ; \qquad Z_{\frac{\alpha}{2}} = Z_{0.01} =$$

The 98% confidence interval for the proportion of defective items in the complete batch is given by

$$p_s \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p_s(1-p_s)}{n}} =$$

Example 9:

A sample poll of 100 voters chosen at random from all voters in a given town indicated that 55% of them were support Mr. ABC. Find the 95% confidence interval for the proportion of all the voters that support Mr. ABC.

Solution:

$$p_s = \qquad ; \ \alpha = \qquad ; \ Z_{\frac{\alpha}{2}} = Z_{0.025} =$$

The 95% confidence interval for the proportion of all the voters that support candidate ABC is given by

$$p_s \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p_s(1-p_s)}{n}} =$$

Determination of a proper sample size

Suppose that we require a sample statistic to provide an estimate of the population parameter to a specified degree of accuracy ($\pm r$).

1. Sample size for estimating a population mean, μ

   At 100(1-α)% level of confidence, $\bar{x} \pm Z_{\frac{\alpha}{2}} S.E. = \bar{x} \pm r$

   $$\Rightarrow \quad Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = r \quad \Longrightarrow \quad n = \left(\frac{Z_{\alpha/2} s}{r} \right)^2$$

2. Sample size for estimating a population proportion, π

   At 100(1 - α)% level of confidence, $p_s \pm Z_{\frac{\alpha}{2}} S.E. = p_s \pm r$

   $$\Rightarrow \quad Z_{\frac{\alpha}{2}} \sqrt{\frac{p_s(1-p_s)}{n}} = r \quad \Longrightarrow \quad n = \frac{(Z_{\alpha/2})^2 p_s(1-p_s)}{r^2}$$

   Note:
   1. *s* is replaced by σ if σ is known
   2. If no information about sample proportion, use $n = \frac{(Z_{\alpha/2})^2}{4r^2}$

Example 10:
Suppose that we wished to find a 90% confidence interval for the purchase price of TVs in various retail stores in a given area such that the sample mean will differ by no more than $25. Assume that σ is known and equal to $35. How large should *n* be to satisfy these requirements?

Solution:

$\alpha = 0.10$; $r =$; $\sigma = 35$; $Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.6449$

$$\Rightarrow \quad n = \left(\frac{Z_{\alpha/2} \sigma}{r} \right)^2 =$$

Example 11:
A manufacturer wishes to estimate the proportion of defective components. He would be satisfied if he obtained an estimate within 0.5% of the true proportion, and was 99% confident of his result. An initial (large) sample indicated that $p_s = 0.02$. What is the size of the sample he should examine?

Solution:
$p_s = 0.02$; $r = 0.5\% = 0.005$; $\alpha = 0.01$; $Z_{\frac{\alpha}{2}} = Z_{0.005} =$

$$\Rightarrow \quad n = \frac{(Z_{\alpha/2})^2 p_s(1 - p_s)}{r^2} =$$

Example 12:
Suppose that we want to estimate what proportion of all drivers exceed the maximum speed limit on a certain road. How large a sample will we need so that the error of our estimate is at most 0.04 if the desired level of confidence is 99%.

Solution:

Finite population adjustments

So far, we have been assuming that we sample from an infinite population. If the population is finite, then an adjustment has to make.

The correct value for the standard error of the mean is

$$S.E. = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

where N is the population size

The correct value for the standard error of the proportion is

$$S.E. = \sqrt{\frac{p_s(1-p_s)}{n}} \times \sqrt{\frac{N-n}{N-1}}$$

The quantity $\sqrt{\frac{N-n}{N-1}}$ is called the finite population adjustment factor

Example 14:
A random sample of 100 items taken from a large batch of articles of 2696 items contains 5 defective items. Set up 95% confidence interval for the proportion of defective items.

Solution:

n =; N =; x = 5; $\alpha = 0.05$; $Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$; $p_s = \frac{x}{n} =$

The 95% confidence interval for the proportion of defective items is given by

$$p_s \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p_s(1-p_s)}{n}} \times \sqrt{\frac{N-n}{N-1}} =$$

Example 15:

A sample survey conducted in a certain town in 2012 showed that 200 families spent on the average RM85.44 per week on food with a standard deviation of RM9.12. Construct a 99% confidence interval for the actual average weekly food expenditures of the 1,000 families in this town.

Solution:

$n =$; $\overline{x} =$; s = ; $N =$; $\alpha = 0.01$; $Z_{\frac{\alpha}{2}} = Z_{0.005} =$

The 99% confidence interval for the actual average weekly food expenditures of the 1,000 families in this town is given by

$$\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} =$$