

CHAPTER 4 SAMPLING DISTRIBUTIONS AND ESTIMATION

		<u>Page</u>
Contents	4.1 Definition	72
	4.2 Sampling Distribution of Means	72
	4.3 Sampling Distribution of Difference of Means	75
	4.4 Sampling Distribution of Proportions	75
	4.5 Sampling Distribution of Difference of Proportions	76
	4.6 Estimation	76
	4.7 Confidence Intervals	77
	Exercise	81

Objectives: After working through this chapter, you should be able to:

- (i) define and construct sampling distributions of means and proportions;
- (ii) explain the central limit theorem and its importance in statistical inference;
- (iii) calculate confidence intervals for means and proportions.

4.1 Definition

1. A sample statistic is a characteristic of a sample.
A population parameter is a characteristic of a population.
2. A statistic is a random variable that depends only on the observed random sample.
3. A sampling distribution is a probability distribution for a sample statistic. It indicates the extent to which a sample statistic will tend to vary because of chance variation in random sampling.
4. The standard deviation of the distribution of a sample statistic is known as the standard error of the statistic.

4.2 Sampling Distribution of Means

With replacement OR from an infinite population

$$\mu_{\bar{X}} = \mu, \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

4.2.1 Distribution

- (i) The parent population is normally distributed with mean μ and variance σ^2 .
If $\bar{X} = \sum_{i=1}^n X_i / n$, then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ with replacement or infinite population.
- (ii) The distribution of parent population is unknown.

Central Limit Theorem

If repeated samples of size n are drawn from any infinite population with mean μ and variance σ^2 , then for n large ($n \geq 30$), the distribution of \bar{X} , the sample mean, is approximately normal, with mean μ and variance σ^2/n , and this approximation becomes better as n becomes larger.

Example 1

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Example 2

The mean IQ scores of all students attending a college is 110 with a standard deviation of 10.

- (a) If the IQ scores are normally distributed, what is the probability that the score of any one student is greater than 112?
- (b) What is the probability that the mean score in a random sample of 36 students is greater than 112?
- (c) What is the probability that the mean score in a random sample of 100 students is greater than 112?

4.2.2 Student's t Distribution

The distribution of a statistic T is defined by

$$T = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Where a sample of size n is drawn from a normal distribution with mean μ .

Properties of t-distribution

1. symmetric about a mean of zero
2. bell-shaped
3. the shape of a t-distribution depends on a parameter ν (degrees of freedom). A t-distribution has $n-1$ degrees of freedom when n is the size of the sample.
4. When $n \rightarrow \infty$, t – distribution $\rightarrow Z$ distribution.

Example 3

A random sample of 15 is picked from a normal population.

Use the table to find

(a) $P(T > 2.145)$

(b) $P(-1.345 < T < 2.145)$

(c) a number c for which $P(T > c) = 0.01$ $c = 2.624$

4.3 Sampling Distribution of Difference of Means

With replacement OR from infinite population

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

4.4 Sampling Distribution of Proportions

With replacement OR from infinite population

$$\mu_{\hat{p}} = p, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

4.5 Sampling Distribution of Difference of Proportions

With replacement OR from infinite population

$$\begin{aligned}\mu_{\hat{p}_1 - \hat{p}_2} &= \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2 \\ \sigma_{\hat{p}_1 - \hat{p}_2} &= \sqrt{\sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2} \\ &= \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}.\end{aligned}$$

4.6 Estimation

Estimation is the process of using sample data to estimate the values of the unknown parameters.

$$\begin{array}{ccc} \text{Sample statistic} & \longrightarrow & \text{Population parameter} \\ \hat{\theta} & \longrightarrow & \theta \end{array}$$

Definition:

1. An estimator of a parameter is a statistic relevant for estimating the parameter. An estimator is a random variable.
2. The value of the estimator is the estimate of the parameter.

Example

	Estimator	Population parameter
1.	\bar{X}	μ
2.	s^2	σ^2

Definition:

1. A point estimate is a single number that is used to estimate a population parameter.
2. An interval estimate of a population parameter is an interval of finite width, $\hat{\theta} \pm k$, centered at the point estimate of the parameter, that is expected to contain the true value of the parameter.
3. A confidence interval gives an interval of values, centered on the point estimate, in which the population parameter is thought to lie, with a known risk of error.

A $100(1-\alpha)\%$ confidence interval for θ

$$\text{i.e. } P(\hat{\theta} - k < \theta < \hat{\theta} + k) = 1 - \alpha$$

$0 < \alpha < 1$ and

$(1 - \alpha)$: confidence coefficient

4.7 Confidence Intervals (Sample statistics normally distributed)

4.7.1 Means

- (a) Large samples ($n \geq 30$) OR σ is known

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

where $1 - \alpha$ is the degree of confidence and σ is usually estimated by s .

- (b) Small samples and σ unknown

$$\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}, \quad \nu = n - 1$$

Examples 4

- (a) The mean and standard deviation for the grade point averages of a random sample of 36 college seniors are calculated to be 2.6 and 0.3, respectively. Find the 95% confidence interval for the mean of the entire senior class.
- (b) If we use a point estimate to estimate the population mean of grade point average, how large a sample is required in (a) if we want to be 95% confident that the maximum error is 0.05?

Example 5

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 liters. Find a 95% confidence interval for the mean of all such containers, assuming an approximate normal distribution.

4.7.2 Proportions

$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example 6

In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it was found that $x = 340$ owned color sets. Find a 95% confidence interval for the actual proportion of families in this city with color sets.

4.7.3 Difference of Means

$$(a) \quad (\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (\text{Large samples OR } \sigma_1 \text{ and } \sigma_2 \text{ are known})$$

$$(b) \quad (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, S_P} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (\text{Small samples, } \sigma_1 = \sigma_2 \text{ \& unknown})$$

$$\nu = n_1 + n_2 - 2 = (n_1 - 1) + (n_2 - 1)$$

$$S_P = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Example 7

A standardized chemistry test was given to 50 girls and 75 boys. The girls made an average grade of 76 with a standard deviation of 6, while the boys made an average grade of 82 with a standard deviation of 8. Find a 96% confidence interval for the difference μ_1 and μ_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.

Example 8

In a batch chemical process, two catalysts are being compared for their effect on the output of the process reaction. A sample of 12 batches was prepared using catalyst 1 and a sample of 10 batches was obtained using catalyst 2. The 12 batches for which catalyst 1 was used gave an average yield of 85 with a sample standard deviation of 4, while the average for the second sample gave an average of 81 and a sample standard deviation of 5. Find a 90% confidence interval for the difference between the population means, assuming the populations are approximately normally distributed with equal variances.

4.7.4 Difference of Proportions

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Example 9

A certain change in a manufacturing procedure for component parts is being considered. Samples are taken using both the existing and the new procedure in order to determine if the new procedure results in an improvement. If 75 of 1500 items from the existing procedure were found to be defective and 80 of 2000 items from the new procedure were found to be defective, find a 90% confidence interval for the true difference in the fraction of defectives between the existing and the new process.

EXERCISE: SAMPLING DISTRIBUTIONS AND ESTIMATION

1. In a certain city, the daily food expenditure of families is normally distributed with a mean of \$150 and a standard deviation of \$30. What is the probability that a random sample of 25 families will have an average daily food expenditure between \$141 and \$162?
2. In a food processing plant, machine A and machine B are filling cans of baked beans. The gross mass per can filling by machine A has a mean of 500 grams and a standard deviation of 5.3 grams, and that by machine B has a mean of 504 grams and a standard deviation of 4.8 grams. It can be assumed that both populations are normally distributed. Two samples with twelve cans each are randomly selected from the cans filled by machine A and machine B, respectively. What is the probability that the difference between the means of gross mass per can for these two samples will be less than 2 grams?
3. Let \bar{X} be the mean of a random sample of size n from a distribution which is $N(\mu, 9)$. Find n such that $\Pr(\bar{X} - 1 < \mu < \bar{X} + 1) = 0.90$ approximately.
4. Measurements of the weights of a random sample of 200 ball bearings made by a certain machine during one week showed a mean of 0.824 newton and a standard deviation of 0.042 newton. Find (a) 95% and (b) 99% confidence interval for the mean weight of all the ball bearings.
5. In measuring reaction time, a psychologist estimates that the standard deviation is 0.05 seconds. How large a sample of measurements must be taken in order to be (a) 95% and (b) 99% confident that the error of his estimate will not exceed 0.01 seconds?
6. Ten packets of sugar are randomly selected from a production line; and their weights, in grams, are 23, 26, 28, 30, 31, 24, 29, 27, 22, 21. Find a 95% confidence interval for the population mean of the weights of packets of sugar from that production line. Assume that the population of weights of packets is approximately normal.
7. In a random sample of 500 families owning TV sets in a city, it was found that 160 owned colour sets. Find a 95% confidence interval for the actual proportion of families in the city with colour sets.
8. An investment bank operates two branches in two different towns of a certain country. A sample of salaries was taken from each branch and the results are recorded as follow:

Branch	Sample size	Mean	Standard deviation
A	40	\$53540	\$2630
B	50	\$52480	\$1680

Construct a 96% confidence interval for the difference of mean salaries for all employees between these two branches.

9. A financial manager would like to compare the clients' ratings about the services provided by two financial consultants under his supervision. Independent random samples of service ratings for these two consultants have been selected and the results are summarized below:

	Sample size	Mean	Standard deviation
Consultant I	15	6.82	0.64
Consultant II	10	6.25	0.75

Find a 95% confidence interval for the difference between the population means of service ratings for these two consultants, assuming the populations are approximately normally distributed with equal variances.

10. In a random sample of 400 adults and 600 teenagers who watched a certain TV programme, 100 adults and 300 teenagers indicated that they liked it. Construct (a) 95% and (b) 99% confidence interval for the difference in proportions of all adults and all teenagers who watched the programme and liked it.
11. A company wants to estimate the average number μ of magazine subscriptions per household in a large city. A sample of 120 households is selected and the following data are obtained:

<u>Subscription</u>	<u>Household</u>
0	40
1	37
2	15
3	11
4	9
5	8

- (a) Give a point estimate of μ .
- (b) Give a measure of the degree of precision of your estimate.
- (c) If the company wants, with 99% confidence, that μ be estimated within ± 0.3 , how large the sample should be?