CHAPTER 5 TEST OF HYPOTHESIS

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

(i) understand the meaning of a statistical hypothesis and significance test;

(ii) understand the reasoning relating to the procedure in carrying out a significance test;

(iii) differentiate small sample test from large sample test;

(iv) understand the tests concerning means and proportions;

(v) understand the applications of chi-square test to perform test of goodness-of-fit and test of independence.

5.1 Introduction to Test of Hypothesis

5.1.1 Statistical Hypothesis

When a random sample is drawn from a population, the information obtained can be used to make inferential statements about the characteristics of the population. One possibility is to estimate the unknown population parameters through the calculation of point estimates or confidence intervals. Alternatively, the sample information can be used to assess the validity of some conjecture, or **hypothesis**.

A statistical hypothesis is an assertion or conjecture concerning one or more populations.

Significance test of a population parameter is a procedure, using the sample information, to infer whether the true population parameter differs significantly from the hypothesized value.

An Illustrative Example

To test the hypothesis that a coin is fair, the following rule of decision is adopted: (1) accept the hypothesis if the number of heads in a single sample of 100 tosses is between 40 and 60 inclusive, (2) reject the hypothesis otherwise.

(a) Find the probability of rejecting the hypothesis when it is actually correct.

(b) Interpret graphically the decision rule and the results of part (a).

Solution:

a) X: number of heads in the 100 tosses

$X \sim b(100, 0.5)$

$\mu = np = 100(0.5) = 50$, $\sigma = \sqrt{npq} = \sqrt{100(0.5)(0.5)} = 5$

$\because$ $0.1 < p < 0.9$, and both *np* and *nq* are greater than 5.

$\therefore$ Normal approximation can be employed.

Required probability = $\Pr(X < 40) + \Pr(X > 60)$

By normal approximation, required probability

= $\Pr(X < 39.5) + \Pr(X > 60.5)$

$$= \Pr\left(Z < \frac{39.5 - 50}{5}\right) + \Pr\left(Z > \frac{60.5 - 50}{5}\right)$$

$$= \Pr(Z < -2.1) + \Pr(Z > 2.1)$$

$$= 2 \times 0.0179$$

$$= 0.0358$$

b) **Graphical Presentation**:

If a single sample of 100 tosses yields a number of heads whose *Z* score lies in the rejection region (**critical region**), we would say that this score differed "**significantly**" from what would be expected if the hypothesis were true.

From this reason the area of rejection region is called the "**level of significance**" of the decision rule and equals 0.0358 in this case. Thus we speak of rejecting the hypothesis at 3.58% level of significance.

5.1.2 Some Hypothesis Testing Terminology

1. Null hypothesis, H_0
   A maintained hypothesis that is held to be true until sufficient evidence to the contrary is obtained.
   $$H_0 : \theta = \theta_0$$

2. Alternative hypothesis, H_1 or H_a

   A hypothesis against which the null hypothesis is tested, and which will be held to be true if the null is held false.

3. Test statistics
   The test statistic is the value, based on the sample, used to determine whether the null hypothesis should be rejected or accepted.

4. Acceptance or rejection (critical) region
   Acceptance region: These values support H_0.
   Rejection (Critical) region: These values support H_a.

5. The significance level, α
   The probability of rejecting a null hypothesis when in fact it is true.

6. Types of error

   (a) Type I error: Reject H_0 when H_0 is true
   $$P(\text{Type I error}) = \alpha$$

   (b) Type II error: Accept H_0 when H_a is true
   $$P(\text{Type II error}) = \beta$$

7. Two-tailed test
   A two-tailed test is used when we are concerned about a possible deviation in either direction of the population parameter from its hypothesized value.

8. One-tailed test
   A one-tailed test is used when we are concerned about the possible deviation in only one direction of the population parameter from its hypothesized value.

5.1.3 Basic Steps in Testing Hypothesis

1. Formulate the null hypothesis.
2. Formulate the alternative hypothesis.
3. Specify the level of significance to be used.
4. Select the appropriate test statistic and establish the critical region.
5. Compute the value of the test statistic.

6. Decision: Reject H_0 if the statistic has a value in the critical region, otherwise accept H_0.
7. Draw conclusion (i.e. to answer the question asked in the problem) according to the decision made in step 6.

5.2 Tests Concerning Means

5.2.1 Testing the Mean of a Population

Example 1a

A manufacturer of sports equipment has developed a new synthetic fishing line that he claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

Example 1b

The average length of time for students to register for fall classes at a certain college has been 50 minutes with a standard deviation of 10 minutes. A new registration procedure using modern computing machines is being tried. If a random sample of 12 students had an average registration time of 42 minutes with a standard deviation of 11.9 minutes under the new system, test the hypothesis that the population mean is now less than 50, using a level of significance of (1) 0.05, and (2) 0.01. Assume the population of times to be normal.

5.2.2 A Summary of Test Statistics of Various Tests Concerning Means

H_0	Test statistic	H_1	Critical region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ if σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_o$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\frac{\alpha}{2}}$ & $z > z_{\frac{\alpha}{2}}$
	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $\upsilon = n-1$ if σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_o$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\frac{\alpha}{2}}$ & $t > t_{\frac{\alpha}{2}}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ if σ_1, σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\frac{\alpha}{2}}$ & $z > z_{\frac{\alpha}{2}}$
	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ with $\upsilon = n_1 + n_2 - 2$ and $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$ if $\sigma_1 = \sigma_2$ but unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\frac{\alpha}{2}}$ & $t > t_{\frac{\alpha}{2}}$
	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ with $\upsilon = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$ if $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\frac{\alpha}{2}}$ & $t > t_{\frac{\alpha}{2}}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$ with $\upsilon = n-1$	$\mu_d < d_0$ $\mu_d > d_0$ $\mu_d \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\frac{\alpha}{2}}$ & $t > t_{\frac{\alpha}{2}}$

*We can use z-test instead of t-test if n is large, say $n \geq 30$.

5.2.3 Testing the Difference between two Means

Example 2

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested, by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a standard deviation of 4, while the samples of material 2 gave an average of 81 and a standard deviation of 5. Test the hypothesis that the two types of material exhibit the same mean abrasive wear at the 0.10 level of significance. Assume the populations to be approximately normal with equal variances.

Example 3

Five samples of a ferrous-type substance are to be used to determine if there is a difference between a laboratory chemical analysis and an X-ray fluorescence analysis of the iron content. Each sample was split into two sub-samples and the two types of analysis were applied. Following are the coded data showing the iron content analysis:

	Sample				
Analysis	1	2	3	4	5
x-ray	2.0	2.0	2.3	2.1	2.4
Chemical	2.2	1.9	2.5	2.3	2.4

Assuming the populations normal, test at the 0.05 level of significance whether the two methods of analysis give, on the average, the same result.

5.3 Tests Concerning Proportions

5.3.1 A Summary of Test Statistics of Various Tests Concerning Proportions

H_0	test statistic	H_1	critical region
$p = p_0$	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ if $n \geq 30$	$p < p_0$	$z < -z_{\alpha}$
		$p > p_0$	$z > z_{\alpha}$
		$p \neq p_0$	$z < -z_{\frac{\alpha}{2}}$ & $z > z_{\frac{\alpha}{2}}$
$p_1 - p_2 = 0$	$z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$ if $n_1, n_2 \geq 30$	$p_1 - p_2 < 0$	$z < -z_{\alpha}$
		$p_1 - p_2 > 0$	$z > z_{\alpha}$
		$p_1 - p_2 \neq 0$	$z < -z_{\frac{\alpha}{2}}$ & $z > z_{\frac{\alpha}{2}}$

5.3.2 Testing a Proportion

Example 4

A manufacturing company has submitted a claim that 90% of items produced by a certain process are non-defective. An improvement in the process is being considered that they feel will lower the proportion of defective below the current 10%. In an experiment 100 items are produced with the new process and 5 are defective. Is this evidence sufficient to conclude that the method has been improved? Use a 0.05 level of significance.

5.3.3 Testing the Difference between two Proportions

Example 5

A vote is to be taken among the residents of a town and the surrounding country to determine whether a proposed chemical plant should be constructed. The construction site is within the town limits and for this reason many voters in the country feel that the proposal will pass because of the large proportion of town voters who favor the construction. To determine if there is a significant difference in the proportion of town voters and county voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal and 240 of 500 county residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use a 0.025 level of significance.

5.4 Other Chi-square Tests

5.4.1 Goodness-of-fit Test

A test to determine if a population has a specified theoretical distribution. The test is based on how good a fit we have between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution.

Theorem: A goodness-of-fit test between observed and expected frequencies is based on the quantity

$$\chi^2_{\text{test}} = \sum \frac{(O_i - E_i)^2}{E_i}$$

where χ^2_{test} is a value of the random variable whose sampling distribution is approximated very closely by the Chi-square distribution,

O_i is the observed frequency of cell i, and E_i is the expected frequency of cell i.

For a level of significance equal to $\alpha, \chi^2_{\text{test}} > \chi^2_{\alpha}$ constitutes the critical region. The decision criterion described here should not be used unless each of the expected frequencies is at least equal to 5. Combine adjacent cells if $E_i < 5$.

The number of degrees of freedom (γ) in a Chi-square goodness-of-fit test = number of cells after combination – 1 – number of parameters estimated for calculating the expected frequencies.

Example 6

Consider the tossing of a die

	Faces					
	1	2	3	4	5	6
Observed	20	22	17	18	19	24
Expected						

Example 7

The following distribution of battery lives may be approximated by the normal distribution.

Class boundaries	O_i	z-value	p-value	E_i
1.45 - 1.95	2			
1.95 - 2.45	1			
2.45 - 2.95	4			
2.95 - 3.45	15			
3.45 - 3.95	10			
3.95 - 4.45	5			
4.45 - 4.95	3			

Solution:

H_0 : the battery lives are normally distributed
H_1 : the battery lives are not normally distributed

$\alpha = 0.05$

Class boundaries	**O_i (Observed frequency)**	**Z-value**	**Probability**	**E_i (Expected frequency)**
1.45 - 1.95	2	< -2.098	0.0179	0.716
1.95 - 2.45	1	-2.098 — -1.381	0.0659	2.636
2.45 - 2.95	4	-1.381 — -0.664	0.1708	6.832
2.95 - 3.45	15	-0.664 — 0.054	0.2653	10.612
3.45 - 3.95	10	0.054 — 0.771	0.2595	10.38
3.95 - 4.45	5	0.771 — 1.489	0.1525	6.1
4.45 - 4.95	3	>1.489	0.0681	2.724

Sample mean, $\bar{x} = \dfrac{\sum xf}{\sum f} = \dfrac{(1.7)(2)+(2.2)(1)+...+(4.7)(3)}{2+1+...+3} = 3.4125$

Sample standard deviation, $s = \sqrt{\dfrac{\sum x^2 f - \dfrac{\left(\sum xf\right)^2}{\sum f}}{\sum f - 1}} = 0.6969$

Hence, Z-value = $\dfrac{L-3.4125}{0.6969} < Z < \dfrac{U-3.4125}{0.6969}$

Where L = lower class boundary of a class
U = upper class boundary of the same class

p-value = $\Pr\left(\dfrac{L-3.4125}{0.6969} < Z < \dfrac{U-3.4125}{0.6969}\right)$

Expected frequency, E_i = (Total observed frequency) (probability)

$$= (40)\ \Pr\left(\frac{L-3.4125}{0.6969} < Z < \frac{U-3.4125}{0.6969}\right)$$

Critical Region, $\chi^2 > 3.841$ $\quad \nu = 4 - 1 - 2$
Test Statistics,

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(7-10.184)^2}{10.184} + \frac{(15-10.612)^2}{10.612} + \frac{(10-10.38)^2}{10.38} + \frac{(8-8.824)^2}{8.824} = 2.9007$$

Conclusion: Do not reject H_0

5.4.2 Test for Independence

The Chi-square test procedure can also be used to test the hypothesis of independence of two variables/attributes. The observed frequencies of two variables are entered in a two-way classification table, or contingency table.

Remark: The expected frequency of the cell in the i[th] row and j[th] column in the contingency table

$$E_{ij} = \frac{\text{(total of row i)} * \text{(total of column j)}}{\text{grand total}}$$

The degrees of freedom for the contingency table is equal to (r – 1) (c –1) where r is the number of rows and c is the number of columns in the table.

Example 8

Suppose that we wish to study the relationship between grade point average and appearance.

	Grade Point Average				
Appearance	1	2	3	4	Totals
attractive	14 ( )	11 ( )	10 ( )	5 ( )	40
ordinary	10 ( )	16 ( )	16 ( )	14 ( )	56
unattractive	3 ( )	4 ( )	7 ( )	10 ( )	24
Totals	27	31	33	29	120

EXERCISE: HYPOTHESIS TESTING

1. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean lifetime of all the bulbs produced by the company, test the hypothesis $\mu = 1600$ hours against the alternative hypothesis $\mu \neq 1600$ hours, using a level of significance (a) 0.05, (b) 0.01.

2. A random sample of 36 drinks from a soft-drink machine has an average content of 7.4 units with a standard deviation of 0.48 unit. Test the hypothesis that $\mu = 7.5$ units against the alternative hypothesis $\mu < 7.5$ at the 0.05 level of significance.

3. The average height of males in the freshman class of a certain college has been 174.0 cm, with a standard deviation of 6.9 cm. Is there reason to believe that there has been a change in the average height if a random sample of 50 males in the present freshman class has an average height of 177.0 cm? Use a 0.02 level of significance.

4. It is claimed that an automobile is driven on the average not more than 12,000 km per year. To test this claim, a random sample of 100 automobile owners are asked to keep a record of the km they travel. Would you agree with this claim if the random sample showed an average of 14,500 km and a standard deviation of 2400 km? Use a 0.01 level of significance.

5. The mean mass of 50 male students who showed above average participation in college athletics was 68.2 kg with a standard deviation of 2.5 kg, while 50 male students who showed no interest in such participation had a mean mass of 67.5 kg with a standard deviation of 2.8 kg. Test whether male students who participate in college athletics are more massive than other male students. Use a 0.05 level of significance.

6. Test the hypothesis that the average weight of containers of a particular lubricant is 10 grams if the weights of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3 and 9.8 grams. Use a 0.01 level of significance and assume that the distribution of weights is normal.

7. It is claimed that the average nicotine content of a cigarette does not exceed 17.5 milligrams. To test the claim, the nicotine contents in milligrams of 8 randomly selected cigarettes were examined.

21.0	16.2	21.5	20.9
15.7	16.3	17.8	19.4

   Is it in line with the manufacturer's claim? Use a 10% significance level.

8. A male student will spend, on the average, \$8 for a Saturday evening fraternity party. Test the hypothesis at the 0.1 level of significance that $\mu = \$8$ against the alternative $\mu \neq \$8$ if a random sample of 12 male students attending a homecoming party showed an average expenditure of \$8.90 with a standard deviation of \$1.75. Assume that the expenses are approximately normally distributed.

9. A taxi company is trying to decide whether to purchase brand A or brand B tyres for its fleet of taxis. To help arrive at a decision an experiment is conducted using 12 of each brand. The tyres are run until they wear out. The results are:

$$\text{Brand A} : \bar{x} = 23{,}600 \text{ miles}, \quad S_x = 3{,}200 \text{ miles}$$
$$\text{Brand B} : \bar{y} = 24{,}800 \text{ miles}, \quad S_y = 3{,}700 \text{ miles}$$

Test the hypothesis at the 0.05 level of significance that there is no difference in the two brands of tyres. Assume the two populations to be approximately normal with equal variances.

10. A manufacturer of cigarettes claims that 20% of the cigarette smokers prefer brand X. To test this claim a random sample of 100 cigarette smokers are selected and asked what brand they prefer. If 30 of the 100 smokers prefer brand X, what conclusion do we draw? Use a 0.01 level of significance.

11. A random sample of 100 men and 100 women at a college are asked if they have an automobile on campus. If 31 of the men and 24 of the women have cars, can we conclude that more men than women have cars on campus? Use a 0.01 level of significance.

12. In a study on the comparison of sorbic acid in ham before and after storage, the following data on sorbic acid residuals in parts per million of 8 slices of ham immediately after dipping in a sorbate solution and after 60 days of storage were recorded.

	Sorbic Acid Residuals	
Slice	Before Storage	After Storage
1	224	116
2	270	96
3	400	239
4	444	329
5	590	437
6	660	597
7	1400	689
8	680	576

Assuming the populations to be normally distributed, is there sufficient evidence, at the 0.05 level of significance, to say that the length of storage influences sorbic acid residual concentrations?

EXERCISE: CHI-SQUARE TESTS

1. The grades in a statistics course for a particular class were as follows:

GRADE	A	B	C	D	E
f	14	18	32	20	16

Test the hypothesis, at the 0.05 level of significance, that the distribution of grades is uniform.

2. A manufacturer of fashion garment for the younger age group suspects that the market for his product has changed recently. Sales records for previous years show that 15% of buyer were below 17 years of age, 36% were 18-21 of age, 28% were 22 to 25 years and 21% were over 25. A random sample of 220 recent buyers, however, showed the following results:

Age	Under 17	18-21	22-25	Over 25
f	35	80	65	40

Carry out a χ^2 test at 0.05 level of significance to determine whether the above observations are consistent with the sales records.

3. It is often not clear whether all properties of a binomial experiment are actually met in a given application. A goodness-of-fit test is desired in such cases. Suppose an experiment consisting 4 trials was repeated 120 times with number of successes as follows. Let $\alpha = 0.05$ and test the hypothesis that the distribution is binomial.

Number of successes	0	1	2	3	4
Observed Frequency	19	53	33	11	4

4. The time to train new employees to use the existing semi-automatic equipment of a large manufacturing company is thought to be approximately normally distributed. A sample of 100 new employees in the past 6 months revealed the following information:

Training time (hours)	Frequency
4.4 – less than 4.6	9
4.6 – less than 4.8	16
4.8 – less than 5.0	27
5.0 – less than 5.2	21
5.2 – less than 5.4	18
5.4 – less than 5.6	9

Test the hypothesis that the time to train new employees is normally distributed with $\alpha = 0.05$.

5. A random sample of 200 retired married men was classified according to education and number of children.

	Number of Children		
Education	0-1	2-3	Over 3
Elementary	14	37	32
Secondary	19	42	17
College	12	17	10

Test the hypothesis, at the 0.05 level of significance, that the size of a family is independent of the level of education attained by the father.

6. A marketing researcher wishes to investigate whether there is any relationship between different groups of people (i.e. Men, Women and Children) and their reaction to the flavour of a new soft-drink. An opinion survey was conducted among a random sample of 500 people. The interviewees were given the new brand of soft-drink to taste and their feedbacks were tabulated as follows:

		People		
		Men	Women	Children
The interviewee's taste	Like flavour	63	66	81
	Dislike flavour	88	54	74
	No opinion	20	32	22

Use the 0.05 level of significance to test the null hypothesis that there is no association between different groups of people and their reaction to the flavour of the new soft-drink.