Chapter 7: Hypothesis Testing

- A statistical test examines a set of sample data and, on the basis of an expected distribution of data, leads to a decision about whether to accept or to reject the hypothesis.

- **Hypothesis:** A prediction about a single population or about the relationship between two or more populations.

Examples:

- Average length of stay of patients admitted to the hospitals is five days.
- A certain drug will be effective in 90 percent of the cases for which it used.
- A particular educational program will result in improved communication between nurses and patients.

Note: Hypothesis testing is a procedure in which sample data are employed to evaluate a hypothesis

Types of Hypothesis:

- **Research hypothesis:** The conjecture, supposition or general statement of what a researcher predicts.

  Examples:

  (a) The average IQ of all males is some value other than 100.

  (b) Clinically depressed patients who take antidepressant for six months will be less depressed than clinically depressed patients who take a placebo for six months.

Statistical Hypothesis:

- The statistical hypothesis summarizes the research hypothesis with reference to the population parameter or parameters.

  Types of Statistical hypothesis:

  Null hypothesis (H_0): A statement of no effect or no difference. The null hypothesis will be generally be the hypothesis that the researcher expects to be rejected. Therefore the statement of the research hypothesis generally predicts the presence of an effect or a difference with respect to whatever it is that is being studied.

Continues…

- **Alternative hypothesis** (H_A)**:**

  Represents a statistical statement indicating the presence of an effect or a difference. In general, a statement which disagree with null hypothesis. If the null hypothesis is rejected as a result of sample evidence, then the alternative hypothesis is the conclusion.

  Note:

  The alternative hypothesis determines the type of test (one tailed or two tailed) needed in hypothesis testing.

Previous examples…

- Assume a study is conducted in which an IQ score is obtained for each of *n* males who have been randomly selected from a population comprised of *N*.

  Null hypothesis: $H_0 : \mu = 100$

  i.e., the mean of the population is equal to 100.

  Alternative hypothesis: $H_A : \mu \neq 100$

  i.e., mean of the population is not equal to 100.

Remarks:

- The alternative hypothesis is non directional (also referred to as two-tailed).
- The alternative hypothesis can be stated directionally. In above example, the alternative hypothesis can be $H_A : \mu > 100$ or $H_A : \mu < 100$.
- If we want to know that the population mean is greater than 100, $H_0 : \mu \leq 100$ and $H_A : \mu > 100$
- If we want to know that the population mean is less than 100, our hypotheses are

  $H_0 : \mu \geq 100$ and $H_A : \mu < 100$

Types of errors:

- **Type I error:**

  When a true null hypothesis is rejected, the probability of committing a Type I error is represented by the alpha level (α). Thus, the probability of committing a Type I error if α = 0.01, is 1%, as compared with a 5% probability if α =0.05

Continues…

- **Type II error:**
  When a false null hypothesis is not rejected (i.e., a true alternative hypothesis is false). The probability of committing a Type II error is represented by beta (β).

- **Significance Level (α):** The probability of rejecting a true hypothesis, i.e., probability of making a Type I error.

Table: Types of Errors

		Condition of Null Hypothesis	
		True	False
Possible Action	Fails to reject	Correct Action	Type II error
	reject	Type I error	Correct Action

Test Statistics

$$\text{Test Statistics} = \frac{\text{Statistics} - \text{Parameter}}{\text{Standard Error for Statistics}}$$

7.2: Hypothesis Testing: A Single Population Mean

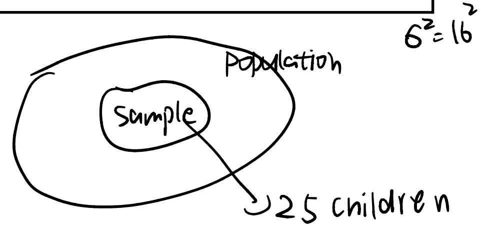
- Testing hypotheses about the single mean are explored under three different situations: When sampling from a population that is
  1. distributed normally with known variance,
  2. distributed normally with unknown variance, and
  3. distributed non-normally.

Normally Distributed Population With Known Variance

- **Example:** Twenty-five randomly sampled children who have lived from birth to age 12 years within 2 miles of a lead smelter are studied to determine if chronic exposure to a lead smelting operation adversely affects performance on a standardized intelligence test. Children in the study group are compared to 12-year-old children in the general population. The mean IQ of children in the study group is 90, with a standard deviation of 14; the mean IQ of 12-year-old children in the general population is 100 and standard deviation of 16.

$\mu = 100$ $\sigma = 16$

$\sigma^2 = 16^2$



In this example our goal is to determine if the performance of children in the study group is comparable to that of the children in the general population or if the performance of the study group has been affected adversely.

- **State the Scientific Hypothesis:**

  The performance on a standardized intelligence test of 12-year-old children who lived from birth to 12 years within 2 miles of a lead smelting operation will be less than that of 12-year-old children in the general population. A directional hypothesis is chosen because substantial published evidence suggests that lead exposure adversely affects children.

Example continues…

- **Formulate the Statistical Hypothesis:**

$$H_0 : \mu = 100$$
$$H_A : \mu < 100$$

- **Test Statistic:** When sampling is from a normally distributed population and the population variance is known, the test statistic for testing $H_0 : \mu = \mu_0$ is

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Example continues…

- When the null hypothesis is true then the distribution of the test statistic is the standard normal curve.
- **Decision Rule:** Identify the level of significance (α) and determine the critical region or rejection area.

  This is a one-tailed test, i.e., all of α will go in the one tail of the distribution. e.g., In the present example, $\alpha = 0.05$.

  Critical value of test statistic (z) = -1.645.

  Our **decision rule** tells us to reject H_0 if the computed value of the test statistic is less than or equal to -1.645.

Example continues…

- **Calculation of the Test Statistic:**

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{90 - 100}{16 / \sqrt{25}} = \frac{-10}{3.2} = -3.1250$$

- **Statistical Decision:**

  Reject the null hypothesis since -3.125 < -1.645.

- **Conclusion:**

  The population mean is smaller than 100.

Example continues…

- **p-value:**

  A decision about the null hypothesis can also be made by comparing the probability of sample results to α. The obtained probability, known as a *p-value*, is the probability of obtaining the result as extreme or more extreme than the sample value, assuming that the null hypothesis is true. Decisions are made about the null hypothesis as follows:

$$p - \text{value} \leq \alpha \Rightarrow \text{Reject } H_0$$
$$p - \text{value} > \alpha \Rightarrow \text{Retain } H_0$$

Example continues…

- The *p*-value for this test is ≈ 0.0009, since P(z ≤ -3.125) = 0.0009 by using table D.

- **Interpretation of the Results:**

  The mean score on standardized intelligence tests of 12-year-old children who have lived from birth to age 12 years within 2 miles of a lead smelter is significantly less than that of 12-year-olds in the general population.

Remarks:

- In the previous example, if one wishes to evaluate the lead smelter results with α=0.01, may do so by comparing the *p*-value (0.0009) to α (0.01).

- Observe that the results are still statistically significant, even if using a decision rule that is more conservative than α=0.05.

Normally Distributed Population With Unknown Variance

- **Example:** The relationship between fetal alcohol exposure and infant development was examined in 31 randomly selected female infants born to women who drank the equivalent of 2 drinks containing 1.5 ounces of alcohol 5 or more times per week during at least 7 $\bar{x}=2.5$; $\sigma_x=0.54$ months of their pregnancy. The mean birth weight of these infants was 2.5 kg, with a standard deviation of 0.54 $\mu=3.3$ kg as compared to 3.3 kg for female newborns in the ~~general population~~. The fetal-alcohol-exposed infants tripled their birth weight at a mean of 15 months, with a standard deviation of 2.7 months as compared to 12 $\mu_2=12$ months for female infants in the general population.

Example continues…

- **State the Scientific Statement:**

  Compared to the female infants in the general population, female infants born to mothers who consumed 2 drinks containing 1.5 ounces of alcohol (or equivalent) 5 or more times per week during at least 7 months of their pregnancy, will differ in mean birth weight and the mean age at which birth weight triples.

Example continues…

- **Formulate the Statistical Hypothesis:**

Birth weight	Age at which birth weight triples
$H_0: \mu = 3.3\,\text{kg}$	$H_0: \mu = 12$ months
$H_A: \mu \neq 3.3\,\text{kg}$	$H_A: \mu \neq 12$ months

- **Test Statistic:** When sampling is from a normally distributed population and the population variance is unknown, the test statistic for testing $H_0: \mu = \mu_0$ is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

which, when null hypothesis is true, is t-distributed with n-1 degree of freedom.

Example continues…

- **Calculation of Test Statistic:**

  Birth weight:

$$\bar{x} = 2.50 \text{ kg}$$
$$s = 0.54, \quad n = 31$$
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.5 - 3.3}{0.0970} = -8.2484$$

  Age at which birth weight triples:

$$\bar{x} = 15 \text{ months}$$
$$s = 2.7, \quad n = 31$$
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{15 - 12}{0.4849} = 6.1868$$

Example continues…

- **Statistical Decision:**

  Compare the calculated and tabulated t-values.

  Birth weight:

  $$t_{\text{cal}} = -8.2484 \text{ and } t_{\text{tab}} = \pm 2.0423$$

  Age at which birth weight triples:

  $$t_{\text{cal}} = 6.1868 \text{ and } t_{\text{tab}} = \pm 2.0423$$

  we reject both null hypotheses.

Example continues…

- ***P*-value:** The exact values for our example cannot be determined from Table E because no t-values greater than 2.7500 are given for two-tailed test with df =30. Consequently to estimate the p-value select the two-tailed area corresponding to the largest t-value in Table E (2.7500) for df=30; this two-tailed area is 0.01. Therefore, We say, then , that the p-value is less than 0.01. The null hypotheses are rejected because the p-value $\leq \alpha$.

Example continues…

- **Conclusion:** The mean birth weight of female infants born to women who drink the equivalent of 2 drinks containing 1.5 ounces of alcohol 5 or more times per week, during at least 7 months of their pregnancy, differs significantly from the mean birth weight of female infants in the general population (p-value < 0.01). Similarly, the mean age at which the fetal-alcohol-exposed infants triple their birth weight differs significantly from female infants in the general population (p-value < 0.01).

When Population Is Not Normally Distributed and Sample Size is ≥ 30

- **Example:** Despite medical advances since 1948, the mean survival time following diagnosis for patients with congestive heart failure remains 2.5 years. An experimental drug undergoing Food and Drug Administration (FDA) review is administered to 50 randomly selected patients with congestive heart failure; all are followed until death. Their mean survival is 3.3 years, with standard deviation of 1.8 years. Does the use of this drug effect survival in patients with congestive heart failure?

Example continues…

- **State the Scientific Statement:**
  Although investigators believe the experimental drug will prolong survival, the FDA is concerned that unexpected side effects may decrease survival after patients take the drug for several weeks or months.
- **Statistical Hypothesis:**

$$H_0 : \mu = 2.5 \text{ years}$$
$$H_A : \mu \neq 2.5 \text{ years}$$

Example continues…

- **Test Statistic:** When sampling is from a non-normally distributed population and the sample size is greater than or equal to 30, the test statistic for testing $H_0 : \mu = \mu_0$ is

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}, \text{ if population variance is known.}$$
$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}, \text{ if population variance is not known.}$$

Example continues…

- **Decision Rule:** Identify the level of significance (α) and determine the critical region or rejection area. This is a two-tailed test.

  e.g., In given example, $\alpha = 0.05$.

  Critical value of test statistic

  $z = \pm 1.96$.

  Our decision rule tells us to reject H_0 if the computed value of the test statistic is either less than or equal to -1.96 or ≥1.96.

Example continues…

- **Calculation of Test Statistic:**

$$\bar{x} = 3.3$$
$$s = 1.8, \quad n = 50$$
$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{3.3 - 2.5}{0.2546} = 3.1422$$

- **Statistical Decision:** Reject the null hypothesis since 3.1422 > 1.96.
- ***P*-value:** Since the test is two tailed, so the p-value for this test is 0.0016.

Example continues…

- **Conclusion:**

  The mean survival time of patients who received the experimental drug differs significantly from 2.5 years (p-value= 0.0016). The observed mean survival of 3.3 years is consistent with anticipated results.

Suggested Homework:

- Examples: 7.2.1- 7.2.4,
- Exercise: 7.2.2, 7.2.5, 7.2.6, 7.2.8, 7.2.14, 7.2.15, 7.2.18, 7.2.19.
- Review Questions and Exercises: 18, 20, 22

7.3: Hypothesis Testing: The Difference Between Two Population Means

- Testing hypotheses about the difference between two means are explored under three different situations: When sampling from a populations that is
  1. distributed normally with known variances,
  2. distributed normally with unknown variances, and
  3. distributed non normally.

Formulation of Hypothesis

1. $H_0: \mu_1 - \mu_2 = 0,$
   $H_A: \mu_1 - \mu_2 \neq 0$
2. $H_0: \mu_1 - \mu_2 \geq 0,$
   $H_A: \mu_1 - \mu_2 < 0$
3. $H_0: \mu_1 - \mu_2 \leq 0,$
   $H_A: \mu_1 - \mu_2 > 0$

- Whether or not the two population means are equal.

Normally Distributed Populations With Known Variances

- **Example:** A researcher conducted a study to determine whether collected data provide sufficient evidence to indicate a difference in mean serum uric acid levels between normal individuals and individuals with Down's syndrome. The data consists of serum uric acid readings on 12 individuals with Down's syndrome and 15 normal individuals. The means are $\overline{x}_1 = 4.5$ mg/100ml and $\overline{x}_2 = 3.4$ mg/100ml. The population variances for Down's syndrome and normal population are 1 and 1.5, respectively.

Example continues…

- **State the Scientific Hypothesis:**

  To find the difference in mean serum uric acid levels between normal individuals and individuals with Down's syndrome.
- **Formulate the Statistical Hypothesis:**

$$H_0: \mu_1 - \mu_2 = 0,$$
$$H_A: \mu_1 - \mu_2 \neq 0$$

or

$$H_0: \mu_1 = \mu_2,$$
$$H_A: \mu_1 \neq \mu_2$$

Example continues…

- **Test Statistic:** When each of two random samples has been drawn from a normally distributed populations and the population variances are known, the test statistic for testing is H_0

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

  where the subscript 0 indicates that the difference is a hypothesized parameter.

Example continues…

- When the null hypothesis is true then the distribution of the test statistic is the standard normal curve.
- **Decision Rule:** This is a two-tailed test.

  $\alpha = 0.05$. Critical value of test statistic (z) = ±1.96. Our decision rule tells us to reject H_0 if the computed value of the test statistic is either less than or equal to -1.96 or greater or equal to 1.96, i.e., $z_{computed} \geq 1.96$ or $z_{computed} \leq -1.96$

Example continues…

- **Calculation of Test Statistic:**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(4.5 - 3.4) - 0}{\sqrt{\frac{1}{12} + \frac{1.5}{15}}} = \frac{1.1}{0.4282} = 2.57$$

- **Statistical Decision:**

  Reject the null hypothesis since 2.57 > 1.96

- **Conclusion:**

  Two population means are different.

- **p-value:** For this test, *p*-value =0.0102.

Normally Distributed Population With Unknown Variance

- **Two possibilities:**
  - The two population variances are equal.
  - The two population variances are not equal.

Two Populations With Equal Variances

- **Example:**

  The effect of environmental exposure to lead on intellectual development is investigated in two randomly selected samples of 7-year-old youngster from similar demographic, social, and economic backgrounds. Serum lead levels in group 1 youngster $(n_2 = 41)$ are > 30 μg/dL; serum lead levels in group 2 youngster $(n_1 = 61)$ are < 10 μg/dL. Because of conflicting published results, the investigators hypothesize that a difference in intelligence test scores exists between these two groups. A standardized and widely used intelligence test is administered to all youngster. The mean intelligence test score for group 1 youngster is 94, with a standard deviation of 17; the mean for group 2 group youngster is 101, with a standard deviation of 8.

Example continues…

- **State the Scientific Statement:**

  Mean intelligence test scores differ in two groups of 7-year-old children who have different serum lead levels.

- **Formulate the Statistical Hypothesis:**

$$H_0 : \mu_1 - \mu_2 = 0,$$
$$H_A : \mu_1 - \mu_2 \neq 0$$

or

$$H_0 : \mu_1 = \mu_2,$$
$$H_A : \mu_1 \neq \mu_2$$

Example continues…

- **Test Statistic:** When sampling is from a normally distributed population, and the population variances are unknown and equal, the test statistic for testing H_0 is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

- where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, and the test statistic is *t*-distributed with $n_1 + n_2 - 2$ degrees of freedom.

Example continues…

- **Decision Rule:** This is a two-tailed test. α = 0.05. Critical value of test statistic, $t_{0.975,\,100}$ = ±1.9840. Our decision rule tells us to reject H_0 if the computed value of the test statistic is either less than or equal to - 1.9840 or greater or equal to 1.9840, i.e.,

$$t_{\text{computed}} \geq 1.9840 \quad \text{or} \quad t_{\text{computed}} \leq -1.9840$$

Example continues…

- **Calculation of Test Statistic:**

Group I	Group II
$\bar{X}_1 = 94$	$\bar{X}_2 = 101$
$s_1 = 17$	$s_2 = 8$
$n_1 = 61$	$n_2 = 41$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(61 - 1)(17)^2 + (41 - 1)8^2}{61 + 41 - 2}$$

$$= 199.0$$

$$s_p = 14.1067$$

Example continues…

- $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(94 - 101) - 0}{\sqrt{\frac{199}{61} + \frac{199}{41}}} = \frac{-7}{2.8488} = -2.4752$$

- **Statistical Decision:**

  Reject the null hypothesis since

  -2.4752 < -1.9840

- ***p*-value:** For this test, *p*-value 0.01<*p*< 0.02.

Example Continues…

- **Interpretation:**

  A significance difference exists in mean intelligence test scores in two groups of 7-year-olds who have different mean serum lead levels. The mean score is 94 in youngster with a mean serum lead level > 30, compared to 101 in youngster with a mean serum lead level < 10.

Two Populations With Unequal Variances

- **Example:**

  Satisfaction with medical care is compared in two groups of randomly selected patients. In one group, patients receive care from clinicians who participate in small multi-specialty group practices (MSGs); Patients in the other group receive care from clinicians who participate in health maintenance organizations (HMOs). Several indices of patient satisfaction are obtained, including the patient's overall satisfaction with the care received during the last 12 months. Satisfaction is rated on a 100-point scale, with the higher rating indicating greater satisfaction. In MSG patients, the mean level of satisfaction is 80, with a standard deviation of 6 ($n_{MSG}=20$). In HMO patients, the mean is 63, with SD of 15 ($n_{HMO}=51$).

	MSG	HMO
n	20	51
$\bar{x}$	80	63
SD	6	15

Example Continues…

- **State the Scientific Statement:**

  The mean level of overall patient satisfaction with care received by MSGs in the 12 months prior to the study differs from that of HMOs.

- **Statistical Hypothesis:**

$$H_0: \mu_{MSG} - \mu_{HMO} = 0,$$
$$H_A: \mu_{MSG} - \mu_{HMO} \neq 0$$

or

$$H_0: \mu_{MSG} = \mu_{HMO},$$
$$H_A: \mu_{MSG} \neq \mu_{HMO}$$

This test is a two-sided test.

Example continues…

- **Test Statistic:**

  When sampling is from a normally distributed population, and the population variances are unknown and unequal, the test statistic for testing H_0 is

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Example continues…

- The critical value of t' for an α level of significance and two-sided test is approximately given by

$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$
$$\text{where } w_1 = \frac{s_1^2}{n_1},\ w_2 = \frac{s_2^2}{n_2},$$
$$t_1 = t_{(1-\alpha/2)} \text{ for } n_1 - 1 \text{ degree of freedom}$$
$$t_2 = t_{(1-\alpha/2)} \text{ for } n_2 - 1 \text{ degree of freedom.}$$

Example continues…

- The critical value of t' for an α level of significance and one-sided test is approximately

$$t'_{1-\alpha} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$
$$\text{where } w_1 = \frac{s_1^2}{n_1},\ w_2 = \frac{s_2^2}{n_2},$$
$$t_1 = t_{(1-\alpha)} \text{ for } n_1 - 1 \text{ degree of freedom}$$
$$t_2 = t_{(1-\alpha)} \text{ for } n_2 - 1 \text{ degree of freedom.}$$

Example Continues…

- **Calculation of the Test Statistic:**

Group I: MSG Patients	Group II: HMO Patients
$\overline{X}_{MSG} = 80$	$\overline{X}_{HMO} = 63$
$s_{MSG} = 6$	$s_{HMO} = 15$
$n_{MSG} = 20$	$n_{HMO} = 51$

$$t' = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{80 - 63}{\sqrt{\frac{6^2}{20} + \frac{15^2}{51}}} = \frac{17}{2.4923} = 6.8210$$

Example Continues…

- **Decision Rule:** This is a two-tailed test with α = 0.05. Critical value of test statistic is

$$w_1 = \frac{s_1^2}{n_1} = \frac{6^2}{20} = 1.8,\ w_2 = \frac{s_2^2}{n_2} = \frac{15^2}{51} = 4.4118$$
$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} = \frac{1.8(2.093) + 4.5(2.0086)}{1.8 + 4.4118}$$
$$= \frac{3.7674 + 9.0387}{6.2118} = 2.0616$$
where
$$t_1 = t_{(1-\alpha/2)} = t_{0.975} = 2.0930 \text{ for } n_1 - 1 = 19 \text{ degree of freedom}$$
$$t_2 = t_{(1-\alpha/2)} = t_{0.975} = 2.0086 \text{ for } n_2 - 1 = 50 \text{ degree of freedom.}$$

Example Continues…

Our decision rule tells us to reject H_0 if the computed value of the test statistic is either ≥ 2.0616 or ≤ -2.0616.

- **Statistical Decision:** Reject the null hypothesis since 6.8122 > 2.0616.
- ***p*-value:** *p*- value < 0.001 (approximately).

Example Continues…

- **Interpretation:**

  In two groups of patients, a significant difference exists in mean patient satisfaction with care received during the 12 months prior to the study; MSG patients express a higher level of overall satisfaction than HMO patients.

When Population Is Not Normally Distributed and Sample Size is ≥ 30

- **Example:** Mean time to coronary artery re-narrowing is compared in two groups of randomly selected patients who undergo surgery for a blocked coronary artery. One group $(n_{\text{Ath}} = 54)$ has atherectomy to open the clogged artery; the other $(n_{\text{Angio}} = 48)$ has balloon angioplasty. Mean time to re-narrowing of the affected artery is 6 months in patients who had atherectomy $(s_{\text{Ath}} = 2.3)$ and 5.3 months $(s_{\text{Angio}} = 2.1)$ in patients who had angioplasty. Assume time to re-narrowing is distributed non-normally in both treatment populations.

Example continues…

- **State the Scientific Statement:**

  Mean time to coronary artery re-narrowing (months) differs in patients having atherectomy to open a clogged coronary artery, versus those undergoing balloon angioplasty.
- **Statistical Hypothesis:**

  $$H_0: \mu_{\text{Ath}} - \mu_{\text{Angio}} \le 0, \quad \text{or} \quad H_0: \mu_{\text{Ath}} \le \mu_{\text{Angio}},$$
  $$H_A: \mu_{\text{Ath}} - \mu_{\text{Angio}} > 0 \qquad\quad H_A: \mu_{\text{Ath}} > \mu_{\text{Angio}}$$

  This test is a one-sided test.

Example continues...

- **Test Statistic:**

  When sampling is from a non-normally distributed population, and the population variances are unknown, the test statistic for testing H_0 is

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- When the null hypothesis is true then the distribution of the test statistic is the standard normal curve.

Example continues...

- **Decision Rule:** This is a one-tailed test.

  $\alpha = 0.01$.

  Critical value of test statistic (z) = 2.33.

  Our decision rule tells us to reject H_0 if the computed value of the test statistic is greater or equal to 2.33, i.e.,

$$z_{\text{computed}} \geq 2.33 .$$

Example continues...

- **Calculation of Test Statistic:**

Group I: Atherectomy	Group II: Angioplasty
$\overline{X}_{\text{Ath}} = 6$ months	$\overline{X}_{\text{Angio}} = 5.3$ months
$s_{\text{Ath}} = 2.3$ months	$s_{\text{Angio}} = 2.1$ months
$n_{\text{Ath}} = 54$	$n_{\text{Angio}} = 48$

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(6 - 5.3) - 0}{\sqrt{\frac{2.3^2}{54} + \frac{2.1^2}{48}}} = \frac{0.7}{0.4357} = 1.6066$$

Example continues...

- **Statistical Decision:**

  Do not reject the null hypothesis since

  1.6066 < 2.33

- ***p*-value:** For this test *p*-value = 0.0537
- **Interpretation:** No significance difference exists in mean time to artery re-narrowing in patients undergoing athrectomy to open a clogged artery and in those having balloon angioplasty.

Suggested Homework:

- Examples: 7.3.1- 7.3.4,
- Exercise: 7.3.1, 7.3.2, 7.3.5, 7.3.6, 7.3.9, 7.3.10.

7.4: Hypothesis Testing: Paired Comparisons Test

- In earlier discussions of difference between two population means, it was assumed that the samples were independent.
- Its frequently happens that true differences do not exist between two populations with respect to the variable of interest, but the presence of extraneous sources of variation may cause rejection of the null hypothesis of no difference.

Paired Comparisons Test:

Objective: To eliminate a maximum number of sources of extraneous variation by making the pairs similar with respect to as many variables as possible.

- In paired comparison tests, samples are dependent, i.e., we focus on the difference between pairs of observations, rather than individual observations.

How to obtain paired observations:

- Same subject may be measured before and after receiving some treatment.
- Litter mates of the same sex may be assigned randomly to receive either a treatment or placebo.

Continues…

- Pairs of twins or siblings may be assigned randomly to two treatments in such a way that members of a single pair receive different treatments.
- Dependent group designs or paired designs are created in one of the three ways: within subject, naturally occurring pairs, and investigator-created pairs.

Example:

- The Food and Drug Administration evaluated an over-the-counter home cholesterol test in 10 randomly selected individuals. Patients fast for 12 hours then measure their total cholesterol level using the home test. Total cholesterol level is then assessed in a certified clinical laboratory. Mean total cholesterol determined from the home test is 201.50 mg/dL. The mean level as measured in the laboratory is 200.6 mg/dL. Is the home test a viable tool to monitor total cholesterol? Assume the population distribution of difference is normal.

Example continues...

Within subject data examines comparability of an over-the-counter home and a laboratory test to measure total cholesterol.

Study Participant	Laboratory Test Result	Home Test Result
1	200	208
2	153	147
3	147	156
4	175	169
5	182	191
6	246	252
7	194	187
8	355	357
9	161	155
10	193	193

Example continues...

- **State the Scientific Statement:**

  A difference exists between the mean total cholesterol levels determined in a certified clinical laboratory and from an over-the counter home test.

- **Formulate the Statistical Hypothesis:**

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d \neq 0$$

Example continues...

- **Test Statistic:** When the n sample differences computed from the n pairs of measurements constitute a simple random sample from a normally distributed population of difference, the test statistic for testing H_0 is

$$t = \frac{\bar{d} - \mu_{d_0}}{s_{\bar{d}}}$$

where $\bar{d}$ is the sample mean difference, μ_{d_0} is the hypothesized population mean difference, $s_{\bar{d}} = s_d / \sqrt{n}$, is the number of sample differences, and s_d is the standard deviation of the sample differences.

Example continues…

- When the null hypothesis is true, then the distribution of the test statistic is distributed as t with n-1 degree of freedom.
- The sampling distribution is the random sampling distribution of paired differences.
- The analysis of paired data is similar conceptually to problems involving one mean.

Example continues…

- **Decision Rule:** This is a two-tailed test.
  $\alpha = 0.05$ and degree of freedom: n-1 = 9.
  Critical value of test statistic (t) = ±2.2622.
- Our decision rule tells us to reject H_0 if the computed value of the test statistic is either less than or equal to -2.2622 or greater or equal to 2.2622, i.e.,

$$2.2622 \le t_{\text{computed}} \ \text{or}\ t_{\text{computed}} \le -2.2622$$

Example continues…

- **Calculation of Test Statistic:**

Determine the difference for each pair of observations:

Pair	(Lab-Home)	Difference
1	(200-208)	-8
2	(153-147)	6
3	(147-156)	-9
4	(175-169)	6
5	(182-191)	-9
6	(246-252)	-6
7	(194-187)	7
8	(355-357)	-2
9	(161-155)	6
10	(193-193)	0

Example continues…

- **Calculations:**

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{(-8+6+\dots+0)}{10} = \frac{-9.0}{10} = -0.90$$

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{n\sum d_i^2 - (\sum d_i)^2}{n(n-1)}$$

$$s_d^2 = \frac{10(423) - (-9)^2}{10(9)} = \frac{4149}{90} = 46.10$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{6.7897}{\sqrt{10}} = 2.1471$$

$$t = \frac{\bar{d} - \mu_{d_0}}{s_{\bar{d}}} = \frac{-0.9 - 0}{2.1471} = -0.4192$$

$$\bar d = \frac{\sum d_i}{n}$$

$$\sum(d_i^2 - 2d_i\bar d + (\bar d)^2)$$
$$= \sum d_i^2 - 2n(\bar d)^2 + \sum(\bar d)^2$$
$$= \sum d_i^2 - n(\bar d)^2 = \sum d_i^2 - \frac{(\sum d_i)^2}{n^2}\cdot n$$
$$= \sum d_i^2 - \frac{(\sum d_i)^2}{n}$$

$$S_d^2 = \frac{\sum(d_i - \bar d)^2}{n-1} = \frac{n\cdot\sum d_i^2 - (\sum d_i)^2}{n(n-1)}$$

Example continues…

- **Statistical Decision:**

  Retain the null hypothesis since -0.4192 > -2.2622
- **Conclusion:**

  No significant difference exists between the mean total cholesterol levels determined in a certified clinical laboratory and from an over-the counter home cholesterol test.
- ***p*-value:** For this test, *p*-value > 0.20.

Confidence Interval For $\mu_{\bar{d}}$

- A 100(1-α)% confidence interval for $\mu_{\bar{d}}$ is given by

$$\bar{d} \pm t_{\frac{\alpha}{2}} s_{\bar{d}}.$$

- In our example, the 95% confidence interval is -0.90 ± 2.2622 (2.1471) = -0.90 ± 4.8567

  = (-5.7567, 3.9567).

Suggested Homework:

- Examples: 7.4.1
- Exercises: 7.4.1, 7.4.2, 7.4.3,

7.5: Hypothesis Testing: A Single Population Proportion

- **Example:** Major systemic embolism is an important complication in patients who undergo heart valve replacement. Consider that 12% of patients suffer a major systemic embolism within 4 years after valve replacement surgery. Three clinicians want to determine if aspirin use affects the percentage of patients experiencing embolic complications. Following heart valve replacement (HVR) surgery, 186 randomly chosen patients are maintained on 100 mg aspirin daily throughout the 4-year follow-up period. Thirteen patients experience a major systemic embolism during the follow-up period.

Example continues…

- **State the Scientific Statement:**
  Among patients who have surgery alone and those have surgery and take 100 mg of aspirin daily, a difference exists in the proportion experiencing major embolic complications within 4 years following HVR surgery.
- **Formulate the Statistical Hypothesis:**

$$H_0: p = 0.12,$$
$$H_A: p \neq 0.12$$

Example continues…

- **Test Statistic:** When the sample size is sufficiently large, the test statistics for testing H_0 is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

- when null hypothesis is true, z is distributed approximately as the standard normal.

Example continues…

- **Decision Rule:** This is a two-tailed test.
  $\alpha = 0.05$. Critical value of test statistic
  $z_{0.975} = \pm 1.96$.
  Our decision rule tells us to reject H_0 if the computed value of the test statistic is either less than or equal to - 1.96 or greater or equal to 1.96 i.e.,

$$z_{\text{computed}} \geq 1.96 \quad \text{or} \quad z_{\text{computed}} \leq -1.96$$

Example continues…

- **Calculation of Test Statistic:**

$$p = 0.12, \quad n = 186, \quad \hat{p} = \frac{13}{186} = 0.0699$$
$$np = 186(0.12) = 22.32 > 5,$$
$$nq = 186(1 - 0.12) = 163.68 > 5.$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.0699 - 0.12}{\sqrt{\frac{0.12(1 - 0.12)}{186}}} = \frac{-0.0501}{0.0238} = -2.1050$$

Example continues…

- **Statistical Decision:**

  Reject the null hypothesis since

  $-2.1050 < -1.96$.
- ***p*-value:** For this test, p-value = 0.0348.
- **Interpretation:** With respect to the proportion who experience embolic complications within 4 years of surgery, a significant difference exists between HRV patients who have surgery alone and those who have surgery and take 100 mg of aspirin daily.

7.6 Hypothesis Testing: The Difference Between Two Population Proportions

- **Example:** In a study of nutrition care in nursing homes, researchers found that among 55 patients with hypertension, 24 were on sodium-restricted diets. Of 149 patients without hypertension, 36 were on sodium-restricted diets. May we conclude that in the sampled populations the proportion of patients on sodium-restricted diets is higher among patients with hypertension than among patients without hypertension?

Example Continues…

- **State the Scientific Statement:**

  In the sampled populations, the proportion of patients on sodium-restricted diets is higher among patients with hypertension than among patients without hypertension.
- **Statistical Hypothesis:**

$$H_0: p_H - p_{\bar{H}} \le 0, \quad \text{or} \quad H_0: p_H \le p_{\bar{H}},$$
$$H_A: p_H - p_{\bar{H}} > 0, \qquad\quad H_A: p_H > p_{\bar{H}},$$

p_H = the proportion on sodium-restricted diets in the population of hypertensive patients.

$p_{\bar{H}}$ = the proportion on sodium-restricted diets in the population of patients without hypertensive patients

Example continues…

- **Test Statistic:**

  The test statistics for testing H_0 is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

where

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Example continues...

- $\bar{p}$ is a pooled estimate of the hypothesized common proportion.
- x_1 and x_2 are the numbers in the first and second samples, respectively, possessing the characteristic of interest.
- *z*-statistic is distributed approximately as the standard normal if the null hypothesis is true.

Example continues...

- **Decision Rule:** This is a one-tailed test.

  α = 0.05. Critical value of test statistic $z_{0.95} = 1.645$

  Our decision rule tells us to reject H_0 if the computed value of the test statistic is greater or equal to 1.645, i.e.,

  $$z_{\text{computed}} \geq 1.645$$

Example Continues...

- **Calculation of Test Statistic:**

Group I

$$x_H = 24$$
$$n_H = 55$$
$$\hat{p}_H = \frac{24}{55} = 0.4364$$

Group II

$$x_{\overline{H}} = 36$$
$$n_{\overline{H}} = 149$$
$$\hat{p}_{\overline{H}} = \frac{36}{149} = 0.2416$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{24 + 36}{55 + 149} = 0.2941$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}$$

$$= \sqrt{\frac{(0.2941)(0.7059)}{55} + \frac{(0.2941)(0.7059)}{149}} = 0.0719$$

Example Continues...

- $$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{0.4364 - 0.2416}{0.0719} = 2.71$$

- **Statistical Decision:**

  Reject the null hypothesis since

  2.71 > 1.645.

- ***p*-value:** For this test, *p*-value = 0.0034.

Example Continues…

- **Interpretation:**
  The proportion of patients on sodium-restricted diets is higher among hypertensive patients than among patients without hypertension.

Suggested Homework:

- Examples: 7.5.1, 7.6.1
- Exercises: 7.5.2, 7.5.3, 7.5.5, 7.6.1, 7.6.3, 7.6.4.

7.9: Type II Error and Power of a Test

- Recall that if a null hypothesis is false and we fail to reject it, we commit a type II error.
- For a given hypothesis test, it is of interest to know how well the test controls type II errors.
- An important concept in connection with hypothesis testing is the **power** of a test.

Observations:

- Note that only one value of α is associated with a given hypothesis test, but there are many values of β, one for each value of μ if null hypothesis value is not the true value of μ as hypothesized.
- β is relatively large compared with α, unless alternative values of μ are much larger or smaller than null hypothesis value.

Purpose for hypothesis-testing:

- Most often in those cases in which, when null hypothesis is false and the true value of the parameter is fairly close to the hypothesized value.
- In most cases, β, the computed probability of failing to reject a false null hypothesis, is larger than α, the probability of rejecting a true null hypothesis.
- These facts are true in a sense that a decision based on a rejected null hypothesis is more conclusive than a decision based on a null hypothesis that is not rejected.

Continues…

- The power is $1-\beta$, the complement of β (the type II error), and is the probability of rejecting the null hypothesis when it is false and the alternative hypothesis is correct.

Objective:

- To have the quantity $1-\beta$ be as large as possible, and the quantity β as small as possible.

Continues…

- If the hypothesized value of the parameter is not a true value then the value of β depends on:
  1. the true value of the parameter of interest,
  2. the hypothesized value of the parameter,
  3. the value of α, and
  4. the sample size.
- For a given test, we may specify any number of possible values of the parameter of interest and, for each, compute the value of $1-\beta$. The result is called the power function.

Continues…

- The graph of a power function is known as a **power curve**.
- The appearance of a power curve for the two-sided test is V-shaped.
- A two-sided test that discriminates well between the value of the parameter in null hypothesis and values in alternative hypothesis results in a narrow V-shaped power curve.
- A wide V-shaped curve indicates that the test discriminates poorly over a relatively wide interval of alternative values of the parameter.

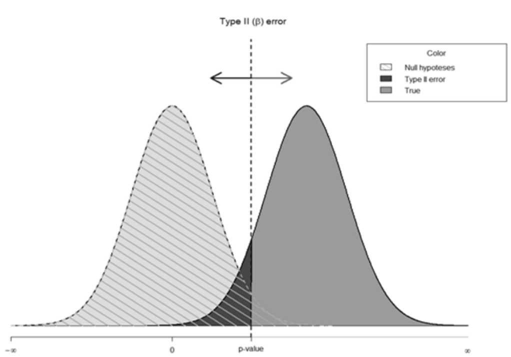
Continues…

- The shape of a power curve for a one-sided test with the rejection region in the upper tail is an elongated S.
- If the rejection region of a one-sided test is located in the lower tail of the distribution, the power curve takes the form of a reverse elongated S.

Importance of Power Analysis:

- A study may have low power because the sample size was too small to detect the presence of reasonable differences; it is a "negative" study.
- Power calculations are also important when an investigator wants to conclude that two drugs or two procedures are not significantly different. In principle, the conclusion of "no significant difference" can be guaranteed by using a very small sample size.

Power



Continues…

- For a given sample size, *n*, the value of α is related inversely to the value of β, i.e., lower probabilities of committing a type I error are associated with the higher probabilities of committing a type II error.
- Both types of error may be reduced simultaneously by increasing *n*. Thus, for a given α, larger samples will result in statistical testing with greater power $1- \beta$.

7.10: Determining Sample Size to Control Type II Errors

- The method of determining the sample size as presented earlier takes into account the probability of type I error, but not a type II error since the level of confidence is determined by the confidence coefficient, 1-α.
- **Objective:**

  To accommodate the type II error as well as type I error when determining the sample size.

Example:

- In a study, the sample size = 16, and the hypotheses are $H_0: \mu \le 4.50,\ H_A: \mu > 4.50.$

  The population standard deviation is 0.020, and the probability of type I error is set at 0.01. Suppose that we want the probability of fail to reject null hypothesis to be 0.05 (β) if H_0 is false, because the true mean is 4.52 rather than the hypothesized 4.50. Find n and C i.e., how large a sample we need to realize, simultaneously, the desired levels of α and β.

Example continues…

- **Critical value (C):** It is a function of z_0 and μ_0; it is also a function of z_1 and μ_1, where μ_0 is the hypothesized mean and μ_1 the mean under the alternative hypothesis.
- When we transform the sampling distribution of $\bar{x}$ that has a mean of μ_0 (μ_1) to the standard normal distribution, we call the z that results $z_0(z_1)$.

Example Continues…

- **Formula for C's:**

$$\left.\begin{array}{l} C = \mu_0 + z_0 \dfrac{\sigma}{\sqrt{n}} \\ C = \mu_1 - z_1 \dfrac{\sigma}{\sqrt{n}} \end{array}\right\rangle \text{if } H_0: \mu \le \mu_0 \text{ and } H_A: \mu > \mu_0$$

$$\left.\begin{array}{l} C = \mu_0 - z_0 \dfrac{\sigma}{\sqrt{n}} \\ C = \mu_1 + z_1 \dfrac{\sigma}{\sqrt{n}} \end{array}\right\rangle \text{if } H_0: \mu \ge \mu_0 \text{ and } H_A: \mu < \mu_0$$

- **Formula for n:**

$$n = \left[\frac{(z_0 + z_1)\sigma}{(\mu_0 - \mu_1)}\right]^2$$

Example continues…

- Here, z_0 = the value of z that has 0.01 of the area to its right is 2.33.
  z_1 = the value of z that has 0.05 of the area to its left is -1.645.
  Both z_0 and z_1 are taken as positive.
  $\mu_0 = 4.500,\ \mu_1 = 4.520,\ \sigma = 0.020$
  $$n = \left[\frac{(z_0 + z_1)\sigma}{(\mu_0 - \mu_1)}\right]^2 = \left[\frac{(2.33 + 1.645)(0.020)}{4.500 - 4.520}\right]^2 = 15.80 \approx 16$$
- $$C = \mu_0 + z_0 \frac{\sigma}{\sqrt{n}} = 4.5 + 2.33\left[\frac{0.020}{\sqrt{16}}\right] = 4.512$$
  $$C = \mu_1 - z_1 \frac{\sigma}{\sqrt{n}} = 4.52 - 1.645\left[\frac{0.020}{\sqrt{16}}\right] = 4.512$$
- The difference between the two results is due to a rounding error.

Continues…

- **Decision Rule:**

  Use the value of C.

  Select a sample of size 16 and compute $\bar{x}$. If $\bar{x} \geq 4.512$, reject the null hypothesis. If $\bar{x} < 4.512$, do not reject the null hypothesis.

Suggested Homework:

- Examples:
  7.9.1, 7.9.2, 7.10.1
- Exercise:
  7.10.1, 7.10.2, 7.10.3