CHAPTER 7: Using Sample Statistics To Test Hypotheses About Population Parameters:

In this chapter, we are interested in testing some hypotheses about the unknown population parameters.

7.1 Introduction:

Consider a population with some unknown parameter θ. We are interested in testing (confirming or denying) some conjectures about θ. For example, we might be interested in testing the conjecture that $\theta > \theta_o$, where θ_o is a given value.

- A hypothesis is a statement about one or more populations.
- A research hypothesis is the conjecture or supposition that motivates the research.
- A statistical hypothesis is a conjecture (or a statement) concerning the population which can be evaluated by appropriate statistical technique.
- For example, if θ is an unknown parameter of the population, we might be interested in testing the conjecture sating that $\theta \geq \theta_o$ against $\theta < \theta_o$ (for some specific value θ_o).
- We usually test the null hypothesis (H_o) against the alternative (or the research) hypothesis (H_1 or H_A) by choosing one of the following situations:
    - (i) $H_o: \theta = \theta_o$ against $H_A: \theta \neq \theta_o$
    - (ii) $H_o: \theta \geq \theta_o$ against $H_A: \theta < \theta_o$
    - (iii) $H_o: \theta \leq \theta_o$ against $H_A: \theta > \theta_o$
- Equality sign must appear in the null hypothesis.
- H_o is the null hypothesis and H_A is the alternative hypothesis. (H_o and H_A are complement of each other)
- The null hypothesis (H_o) is also called "the hypothesis of no difference".
- The alternative hypothesis (H_A) is also called the research hypothesis.

- There are 4 possible situations in testing a statistical hypothesis:

		Condition of Null Hypothesis H_o (Nature/reality)	
		H_o is true	H_o is false
Possible Action (Decision)	Accepting H_o	Correct Decision	Type II error (β)
	Rejecting H_o	Type I error (α)	Correct Decision

- There are two types of Errors:
    - Type I error = Rejecting H_o when H_o is true
      P(Type I error) = P(Rejecting Ho | Ho is true) = α
    - Type II error = Accepting Ho when Ho is false
      P(Type II error) = P(Accepting Ho | Ho is false) = β

- The level of significance of the test is the probability of rejecting true H_o:
  α = P(Rejecting H_o | H_o is true) = P(Type I error)

- There are 2 types of alternative hypothesis:
    - One-sided alternative hypothesis:
        - $H_O: \theta \geq \theta_o$ against $H_A: \theta < \theta_o$
        - $H_O: \theta \leq \theta_o$ against $H_A: \theta > \theta_o$
    - Two-sided alternative hypothesis:
        - $H_O: \theta = \theta_o$ against $H_A: \theta \neq \theta_o$

- We will use the terms "accepting" and "not rejecting" interchangeably. Also, we will use the terms "acceptance" and "nonrejection" interchangeably.
- We will use the terms "accept" and "fail to reject" interchangeably

The Procedure of Testing H_o (against H_A):

The test procedure for rejecting H_o (accepting H_A) or accepting H_o (rejecting H_A) involves the following steps:

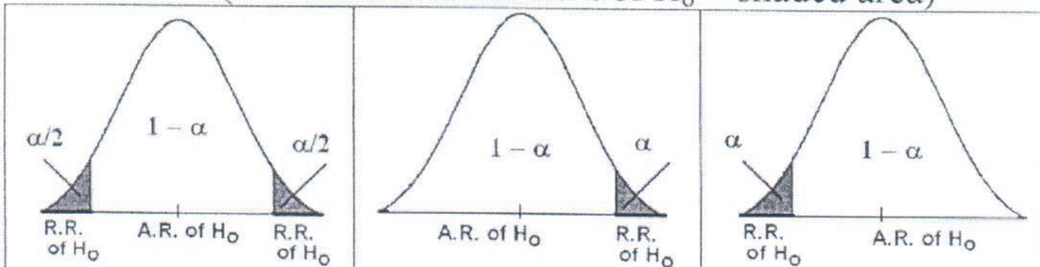
1. Determining a test statistic (T.S.)
We choose the appropriate test statistic based on the point estimator of the parameter.
The test statistic has the following form:

$$\text{Test statistic} = \frac{Estimate - hypothesized\ parameter}{Standard\ Error\ of\ the\ Estimate}$$

2. Determining the level of significance (α):

$$\alpha = 0.01, 0.025, 0.05, 0.10$$

3. Determining the rejection region of H_o (R.R.) and the acceptance region of H_o (A.R.).
The R.R. of H_o depends on H_A and α:
   - H_A determines the direction of the R.R. of H_o
   - α determines the size of the R.R. of H_o
   (α = the size of the R.R. of H_o = shaded area)



$H_A: \theta \neq \theta_o$ Two-sided alternative	$H_A: \theta > \theta_o$ One-sided alternative	$H_A: \theta < \theta_o$ One-sided alternative

4. Decision:
We reject H_o (and accept H_A) if the value of the test statistic (T.S.) belongs to the R.R. of H_o, and vice versa.

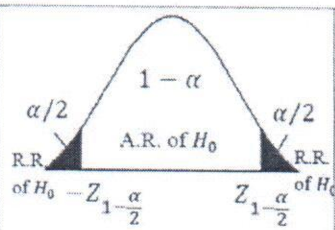
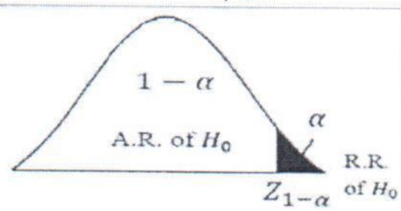
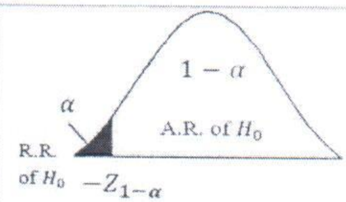
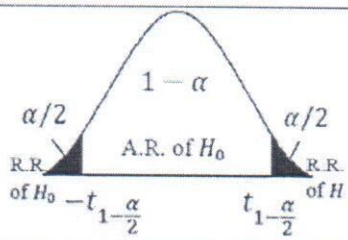
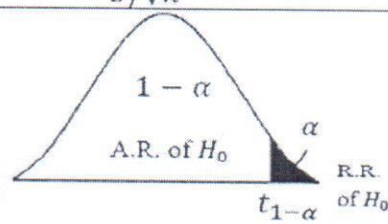
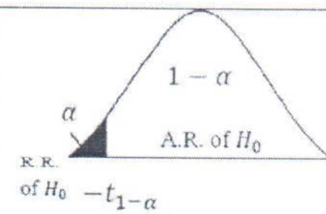
Notes:
1. The rejection region of H_o (R.R.) is sometimes called "the critical region".
2. The values which separate the rejection region (R.R.) and the acceptance region (A.R.) are called "the critical values".

7.2 Hypothesis Testing: A Single Population Mean (μ):

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a distribution (or population) with mean μ and variance σ^2.

We need to test some hypotheses (make some statistical inference) about the mean (μ).

Chapter 7 : Testing Hypothesis about population mean(μ):

Hypothesis	$H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0$	$H_0: \mu \leq \mu_0$ $H_A: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_A: \mu < \mu_0$
First Case	σ is known; Normal or [Non-normal Distribution(n >30)]		
Test Statistic (T.S.)	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	**$-Z_{1-\alpha/2}$ or $Z_{1-\alpha/2}$**	**$Z_{1-\alpha}$**	**$-Z_{1-\alpha}$**
Decision : Reject H_0 if the following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$Z > Z_{1-\alpha/2}$ Or $Z < -Z_{1-\alpha/2}$	$Z > Z_{1-\alpha}$ **(one – Sided Test)**	$Z < -Z_{1-\alpha}$ **(one – Sided Test)**
Second Case	σ is unknown; Normal ,n ≤30 (small)		
Test Statistic (T.S.)	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, \quad df = n-1$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	**$-t_{1-\alpha/2}$ or $t_{1-\alpha/2}$**	**$t_{1-\alpha}$**	**$-t_{1-\alpha}$**
Decision : **Reject H_0 if the Following condition satisfies**	Reject H_0 (Accept H_A) at the significant level α if :		
	$T > t_{1-\alpha/2}$ Or $T < -t_{1-\alpha/2}$ **(Two – Sided Test)**	$T > t_{1-\alpha}$ **(one – Sided Test)**	$T < -t_{1-\alpha}$ **(one – Sided Test)**
Special Case	σ is unknown; Non-Normal ,n >30 (Large)		
Test Statistic (T.S.)	$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$		
Rejection Region	Use the same R.R & A.R as in First Case(Z Case)		

Example: (first case: variance σ^2 is known)

A random sample of 100 recorded deaths in the United States during the past year showed an average of 71.8 years. Assuming a population standard deviation of 8.9 year, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Solution:

.n=100 (large),

σ = 8.9 (σ known)

$\bar{X}$ =71.8, σ =8.9 (σ is known)

μ =average (mean) life span

μ_0= 70

α =0.05

1) Hypotheses:

Ho: $\mu \leq 70$ (μ_0=70)

HA: $\mu > 70$ (research hypothesis)

$H_o: \mu \leq 70 \quad (\mu_o = 70)$

$H_A: \mu > 70$ (research hypothesis)

Test statistics (T.S.) :

$$Z = \frac{\overline{X} - \mu_o}{\sigma / \sqrt{n}} = \frac{71.8 - 70}{8.9 / \sqrt{100}} = 2.02$$

Level of significance:

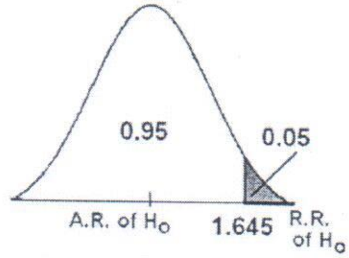
$\alpha = 0.05$

Rejection Region of H_o (R.R.): (critical region)

$-Z_{\alpha} = -Z_{0.05} = 1.645$ (critical value)

We should reject H_o if:

$Z > -Z_{\alpha} = -Z_{0.05} = 1.645$



$- Z_{\alpha} = - Z_{0.05} = 1.645$

Decision:

Since $Z = 2.02 \in$ R.R., i.e., $Z = 2.02 > -Z_{0.05}$, we reject $H_o: \mu \leq 70$ at $\alpha = 0.05$ and accept $H_A: \mu > 70$. Therefore, we conclude that the mean life span today is greater than 70 years.

Note: Using P- Value as a decision tool:

P-value is the smallest value of α for which we can reject the null hypothesis H_o.

Calculating P-value:

* Calculating P-value depends on the alternative hypothesis H_A.
* Suppose that $Z_C = \frac{\overline{X} - \mu_o}{\sigma / \sqrt{n}}$ is the computed value of the test Statistic.
* The following table illustrates how to compute P-value, and how to use P-value for testing the null hypothesis:

Alternative Hypothesis:	HA: $\mu \neq \mu o$	HA: $\mu > \mu o$	HA: $\mu < \mu o$
P - Value	$2 \times P(Z > \|Z_C\|)$	$P(Z > Z_C)$	$P(Z > -Z_C)$
Significance Level =	α		
Decision	Reject Ho if P-value $\leq \alpha$.		

Example:

For the previous example, we have found that:

$$Z_C = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 2.02$$

The alternative hypothesis was HA: $\mu > 70$.

$P-Value = P(Z > Z_C)$

$= P(Z > 2.02) = 1 - P(Z < 2.02) = 1 - 0.9783 = 0.0217$

The level of significance was $\alpha = 0.05$.

Since P-value $\leq \alpha$, we reject H_o.

Example: (second case: variance is unknown)

The manager of a private clinic claims that the mean time of the patient-doctor visit in his clinic is 8 minutes. Test the hypothesis that $\mu = 8$ minutes against the alternative that $\mu \neq 8$ minutes if a random sample of 25 patient-doctor visits yielded a mean time of 7.8 minutes with a standard deviation of 0.5 minutes. It is assumed that the distribution of the time of this type of visits is normal. Use a 0.01 level of significance.

Solution:

The distribution is normal.

.$n = 25$ (small)

$\bar{X} = 7.8$

S=0.5 (sample standard deviation): σ is unknown

μ = mean time of the visit , $\alpha = 0.01$

Hypotheses:

H_0: $\mu = 8$ ($\mu_o = 8$)

H_A: $\mu \neq 8$ (research hypothesis)

Test statistics (T.S.):

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{7.8 - 8}{0.5/\sqrt{25}} = -2$$

df = ν = n-1 = 25 -1= 24

Level of significance:

α = 0.01 , $\alpha/2$=0.005 ,1- $\alpha/2$=0.995

Rejection Region of Ho (R.R.): (critical region)

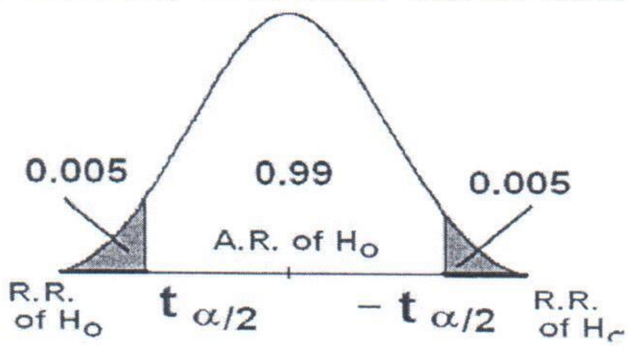
$t_{1-\alpha/2} = t_{0.995} = 2.797$

We should reject Ho if:

$$T < t_{1-\alpha/2} \text{ or } T > - t_{1-\alpha/2}$$

Decision:

Since T = −2 ∈A.R., , we accept H_o: μ =8 at α =0.01 and reject H_A: $\mu \neq 8$. Therefore, we conclude that the claim is correct.



Note:

For the case of non-normal population with unknown variance, and when the sample size is large (n ≥30), we may use the following test statistic:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

That is, we replace the population standard deviation (σ) by the sample standard deviation (S), and we conduct the test as described for the first case.

7.3 Hypothesis Testing: The Difference Between Two Population Means: (Independent Populations)

Suppose that we have two (independent) populations:

- 1-st population with mean μ_1 and variance σ_1^2
- 2-nd population with mean μ_2 and variance σ_2^2
- We are interested in comparing μ_1 and μ_2, or equivalently, making inferences about the difference between the means $(\mu_1-\mu_2)$.
- We independently select a random sample of size n_1 from the 1-st population and another random sample of size n_2 from the 2-nd population:
- Let $\bar{X}_1$ and S_1^2 be the sample mean and the sample variance of the 1-st sample.
- Let $\bar{X}_2$ and S_2^2 be the sample mean and the sample variance of the 2-nd sample.
- The sampling distribution of $\bar{X}_1 - \bar{X}_2$ is used to make inferences about $\mu_1-\mu_2$.

We wish to test some hypotheses comparing the population means.

Hypotheses:

We choose one of the following situations:

(i) $H_o: \mu_1 = \mu_2$ against $H_A: \mu_1 \neq \mu_2$

(ii) $H_o: \mu_1 \geq \mu_2$ against $H_A: \mu_1 < \mu_2$

(iii) $H_o: \mu_1 \leq \mu_2$ against $H_A: \mu_1 > \mu_2$

or equivalently,

(i) $H_o: \mu_1-\mu_2 = 0$ against $H_A: \mu_1 - \mu_2 \neq 0$

(ii) $H_o: \mu_1-\mu_2 \geq 0$ against $H_A: \mu_1 - \mu_2 < 0$

(iii) $H_o: \mu_1-\mu_2 \leq 0$ against $H_A: \mu_1 - \mu_2 > 0$

Test Statistic:

(1) First Case:

For normal populations (or non-normal populations with large sample sizes), and if σ_1^2 and σ_2^2 are known, then the test statistic is:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

(2) Second Case:

For normal populations, and if σ_1^2 and σ_2^2 are unknown but equal ($\sigma_1^2=\sigma_2^2=\sigma^2$), then the test statistic is:

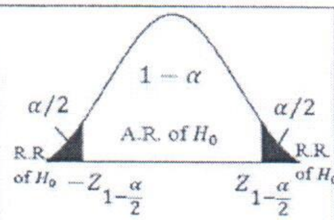
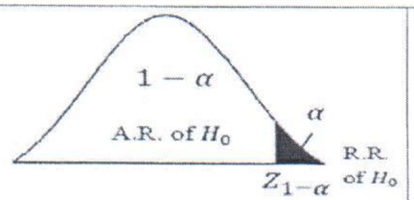
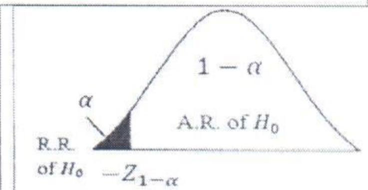
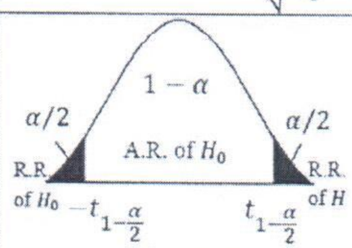
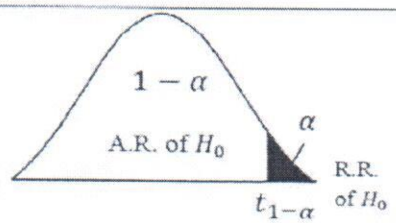
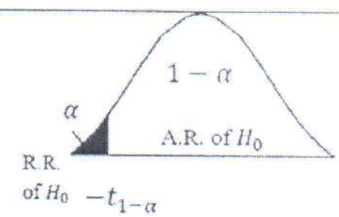
$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim t(n_1+n_2-2)$$

where the pooled estimate of σ^2 is

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

and the degrees of freedom of S_p^2 is df= $\nu=n_1+n_2-2$.

Testing Hypothesis about difference between two population means ($\mu_1 - \mu_2$) : (Independent population)

Hypothesis	$H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_A: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 - \mu_2 \geq 0$ $H_A: \mu_1 - \mu_2 < 0$
First Case	σ_1^2 , σ_2^2 are known		
Test Statistic (T.S.)	$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	$-Z_{1-\alpha/2}$ or $Z_{1-\alpha/2}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision : Reject H_0 if the following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$Z > Z_{1-\alpha/2}$ Or $Z < -Z_{1-\alpha/2}$	$Z > Z_{1-\alpha}$ (one – Sided Test)	$Z < -Z_{1-\alpha}$ (one – Sided Test)
Second Case	σ_1^2 , σ_2^2 are unknown but equal ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)		
Test Statistic (T.S.)	$T = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_P^2}{n_1} + \dfrac{S_P^2}{n_2}}}$, $S_P^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	$-t_{1-\alpha/2}$ or $t_{1-\alpha/2}$	$t_{1-\alpha}$	$-t_{1-\alpha}$
Decision : Reject H_0 if the Following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$T > t_{1-\alpha/2}$ Or $T < -t_{1-\alpha/2}$ (Two –sided test)	$T > t_{1-\alpha}$ (one – Sided Test)	$T < -t_{1-\alpha}$ (one – Sided Test)

Example: (σ_1^2 , σ_2^2 are known)

Researchers wish to know if the data they have collected provide sufficient evidence to indicate the difference in mean serum uric acid levels between individuals with Down's syndrome and normal individuals. The data consist of serum uric acid on 12 individuals with Down's syndrome and 15 normal individuals. The sample means are

$\bar{X}_1 = 4.5$ mg/100ml
$\bar{X}_2 = 3.4$ mg/100ml

Assume the populations are normal with variances

$\sigma_1^2 = 1$

$\sigma_2^2 = 1.5$

. Use significance level $\alpha = 0.05$.

Solution:

μ_1 = mean serum uric acid levels for the individuals with Down's syndrome.

μ_2 = mean serum uric acid levels for the normal individuals.

$n_1 = 12$ $\quad \overline{X}_1 = 4.5$ $\quad \sigma_1^2 = 1$

$n_2 = 15$ $\quad \overline{X}_2 = 3.4$ $\quad \sigma_2^2 = 1.5$.

Hypotheses:

$H_o: \mu_1 = \mu_2$ against $H_A: \mu_1 \neq \mu_2$

or

$H_o: \mu_1 - \mu_2 = 0$ against $H_A: \mu_1 - \mu_2 \neq 0$

Calculation:

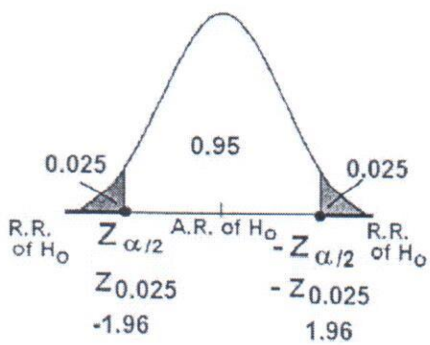
$\alpha = 0.05$

$Z_{0.75} = 1.96$ (1st critical value)

$-Z_{0.75} = -1.96$ (2nd critical value)

Test Statistic (T.S.):

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{4.5 - 3.4}{\sqrt{\frac{1}{12} + \frac{1.5}{15}}} = 2.569$$



Decision:

Since $Z = 2.569 \in$ R.R. we reject $H_o: \mu_1 = \mu_2$ and we accept (do not reject) $H_A: \mu_1 \neq \mu_2$ at $\alpha = 0.05$. Therefore, we conclude that the two population means are not equal.

Notes:

1. We can easily show that a 95% confidence interval for $(\mu_1 - \mu_2)$ is (0.26, 1.94), that is:

$$0.26 < \mu_1 - \mu_2 < 1.94$$

Since this interval does not include 0, we say that 0 is not a candidate for the difference between the population means ($\mu_1-\mu_2$), and we conclude that $\mu_1-\mu_2\neq 0$, i.e., $\mu_1\neq\mu_2$. Thus we arrive at the same conclusion by means of a confidence interval.

2. $P-Value = 2\times P(Z>|Z_C|)$

$$= 2P(Z>2.57) = 2[1-P(Z<2.57)] = 2(1-0.9949) = 0.0102$$

The level of significance was $\alpha = 0.05$.

Since P-value < α, we reject H_o.

Example: ($\sigma_1^2=\sigma_2^2=\sigma^2$ is unknown)

An experiment was performed to compare the abrasive wear of two different materials used in making artificial teeth. 12 pieces of material 1 were tested by exposing each piece to a machine measuring wear. 10 pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average wear of 85 units with a sample standard deviation of 4, while the samples of materials 2 gave an average wear of 81 and a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the mean abrasive wear of material 1 is greater than that of material 2? Assume normal populations with equal variances.

Solution:

Material 1	material 2
n_1=12	n_2=10
$\bar{X}_1$=85	$\bar{X}_2$=81
S_1=4	S_2=5

Hypotheses:

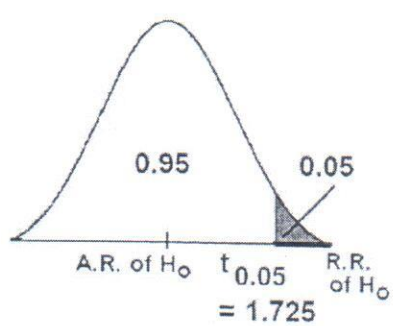
H_o: $\mu_1 \leq \mu_2$

H_A: $\mu_1 > \mu_2$

Or equivalently,

H_o: $\mu_1 - \mu_2 \leq 0$

H_A: $\mu_1 - \mu_2 > 2$



Calculation:

α=0.05

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

$$= \frac{(12 - 1)4^2 + (10 - 1)5^2}{12 + 10 - 2} = 20.05$$

Reliability Coefficient:

$$df = v = 12 + 10 - 2 = 20$$

$\alpha = 0.05$ ------- $1-\alpha = 0.95$ ------- $t_{1-\alpha} = t_{0.95} = 1.725$

Test Statistic (T.S.):

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_P^2}{n_1} + \frac{S_P^2}{n_2}}} = \frac{85 - 81}{\sqrt{\frac{20.05}{12} + \frac{20.05}{10}}} = 2.09$$

Decision:

Since $T = 2.09 \in$ R.R. ($T = 2.09 > t_{0.95} = 1.725$), we reject H_o and we accept **H_A: $\mu_1 - \mu_2 > 0$ (H_A:** $\mu_1 > \mu_2$) at $\alpha = 0.05$. Therefore, we conclude that the mean abrasive wear of material 1 is greater than that of material 2.

7.4 Paired Comparisons:

Paired T-Test :

- In this section, we are interested in comparing the means of two related (non-independent/dependent) normal populations.
- In other words, we wish to make statistical inference for the difference between the means of two related normal populations.
- Paired t-Test concerns about testing the equality of the means of two related normal populations.

Examples of related populations are:

1. Height of the father and height of his son.
2. Mark of the student in MATH and his mark in STAT.
3. Pulse rate of the patient before and after the medical treatment.
4. Hemoglobin level of the patient before and after the medical treatment.

Test procedure:

Let

X: Values of the first population

Y: Values of the Second population

D: Values of X – Values of Y

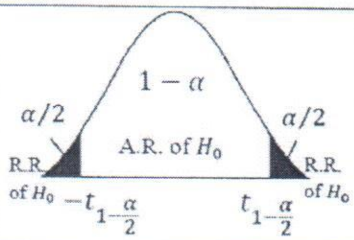
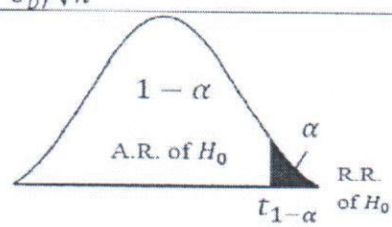
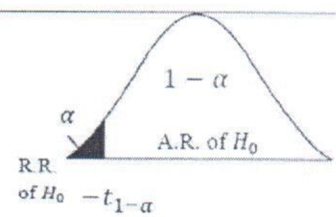
Means :

μ_1= Mean of the first population

μ_2= Mean of the Second population

μ_D =Mean of X – Mean of Y ($\mu_D = \mu_1 - \mu_2$)

Confident Interval and Testing Hypothesis about difference between two population means ($\mu_D = \mu_1 - \mu_2$) : (Dependent/Related population)

Calculate the following Quantities	• The difference (D-observation): $D_i = X_i - Y_i$, $i=1,2,3,4,\ldots\ldots,n$ • Sample mean of the D-Observations : $\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$ • Sample Variance $S_D^2 = \frac{\sum_{i=1}^{n}(D_i-\bar{D})^2}{n-1}$ • Sample Standard Deviation $S_D = \sqrt{S_D^2}$		
Confident Interval for $\mu_D = \mu_1 - \mu_2$			
100(1-α)% Confident Interval for μ_D	$\bar{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$, $df = n-1$		
Testing Hypothesis for $\mu_D = \mu_1 - \mu_2$			
Hypothesis	$H_0:\mu_1 - \mu_2 = 0$ $H_A:\mu_1 - \mu_2 \neq 0$ Or $H_0:\mu_D = 0$ vs $H_A:\mu_D \neq 0$	$H_0:\mu_1 - \mu_2 \leq 0$ $H_A:\mu_1 - \mu_2 > 0$ Or $H_0:\mu_D \leq 0$ vs $H_A:\mu_D > 0$	$H_0:\mu_1 - \mu_2 \geq 0$ $H_A:\mu_1 - \mu_2 < 0$ Or $H_0:\mu_D \geq 0$ vs $H_A:\mu_D < 0$
Test Statistic (T.S.)	$T = \frac{\bar{D}}{S_D/\sqrt{n}}$, $df = v = n-1$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	$-t_{1-\alpha/2}$ or $t_{1-\alpha/2}$	$t_{1-\alpha}$	$-t_{1-\alpha}$
Decision : Reject H_0 if the Following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$T > t_{1-\alpha/2}$ Or $T < -t_{1-\alpha/2}$ (Two –sided test)	$T > t_{1-\alpha}$ (one – Sided Test)	$T < -t_{1-\alpha}$ (one – Sided Test)

Example:

Suppose that we are interested in studying the effectiveness of a certain diet program on ten individual . Let the random variables X and Y given as following table :

Individual(i)	1	2	3	4	5	6	7	8	9	10
Weight before (X_i)	86.6	80.2	91.5	80.6	82.3	81.9	88.4	85.3	83.1	82.1
Weight After (Y_i)	79.7	85.9	81.7	82.5	77.9	85.8	81.3	74.7	68.3	69.7

Find :

1) A 95% Confident Interval for the difference between the mean of weights before the diet program (μ_1) and the mean of weights after the diet program (μ_2).
   [$\mu_D = \mu_1 - \mu_2$]
2) Does the data provide sufficient evidence to allow us to conclude that the diet is good? Use α =0.05 and assume population is normal .

Solution :

1-st population (X) = the weight of the individual before the diet program.
2-nd population (Y)= the weight of the same individual after the diet program.

We assume that the distributions of these random variables are normal with means μ_1 and μ_2 , respectively.

These two variables are related **(dependent/non-independent)**because they are measured on the same individual.

i	Xi	Yi	$D_i = Xi - Y_i$
1	86.6	79.7	6.9
2	80.2	85.9	-5.7
3	91.5	81.7	9.8
4	80.6	82.5	- 1.9
5	82.3	77.9	4.4
6	81.9	85.8	-3.9
7	88.4	81.3	7.1
8	85.3	74.7	10.6
9	83.1	68.3	14.8
10	82.1	69.7	12.4
sum	Σ_X =842	Σ_Y =787.5	Σ_D =54.5

First, we need to calculate :

Sample Mean:

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{54.5}{10} = 5.45$$

Sample Variance :

$$S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1} = \frac{(6.9-5.45)^2 + (-5.7-5.45)^2 + \cdots \ldots\ldots + (12.4-5.45)^2}{10-1} = 50.33$$

Sample Standard Deviation : $S_D = \sqrt{S_D^2} = \sqrt{50.33} = 7.09$

Reliability Coefficient : $t_{1-\alpha/2}$:

α=0.05 ------ 1 – 0.05/2 =1-0.025= 0.975 (df =10 -1=9)

$t_{1-\alpha/2} = t_{0.975} = 2.262$

Then 95% Confident Interval for $\mu_D = \mu_1 - \mu_2$

$$\bar{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

$$5.45 \pm 2.262 \frac{7.09}{\sqrt{10}}$$

$$5.45 \pm 5.0715$$

$$(5.45 - 5.0715 ,\ 5.45 + 5.0715)$$

$$(0.38 , 10.52)$$

$$0.38 < \mu_D < 10.52$$

2)Does the data provide sufficient evidence to allow us to conclude that the diet is good? Use α =0.05 and assume population is normal .

Diet is good means --- weight after will be less than weight befor.

<u>Solution:</u>

μ_1= Mean of the first population

μ_2= Mean of the second population

μ_D =Mean of X – Mean of Y ($\mu_D = \mu_1 - \mu_2$)

Hypothesis :

$H_0: \mu_1 \leq \mu_2$ vs $H_A: \mu_1 > \mu_2$

or $H_0: \mu_1 - \mu_2 \leq 0$ vs $H_A: \mu_1 - \mu_2 > 0$

or $H_0: \mu_D \leq 0$ vs $H_A: \mu_D > 0$

<u>**Test Statistic:**</u>

$$\bar{D} = 5.45,\ S_D = 7.09\ , n = 10$$

$$T = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}} = \frac{5.45}{\frac{7.09}{\sqrt{10}}} = 2.43$$

<u>**Rejection Region(R.R):**</u>

α=0.05 ------ 1-α = 0.95 ----- $t_{1-\alpha} = t_{0.95}$ =1.833 (df =n-1 =9)

Reject H_0 if $T > t_{1-\alpha}$

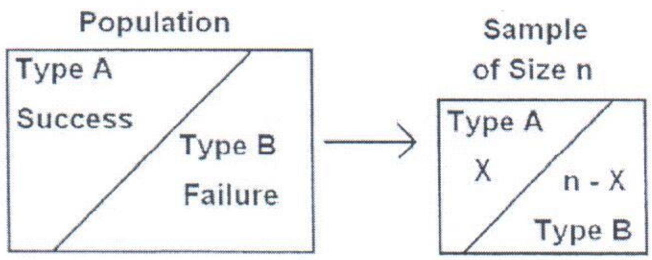
2.45 > 1.833 (condition satisfied)

Then reject H_0 and accept $H_A: \mu_1 > \mu_2$

So, we have a good diet program .

7.5 Hypothesis Testing: A Single Population Proportion (p):

In this section, we are interested in testing some hypotheses about the population proportion (p).



Recall:

- p = Population proportion of elements of Type A in the population

$$p = \frac{\text{no. of elements of type } A \text{ in the population}}{\text{Total no. of elements in the population}}$$

$$p = \frac{A}{N} \qquad (N = \text{population size})$$

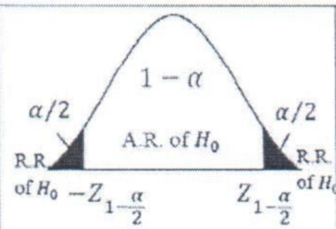
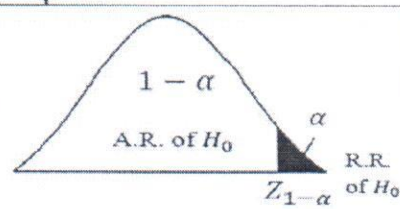
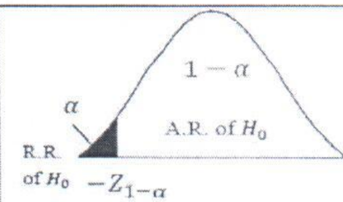
- n = sample size
- X = no. of elements of type A in the sample of size n.
- $\hat{p}$ = Sample proportion elements of Type A in the sample

$$\hat{p} = \frac{\text{no. of elements of type } A \text{ in the sample}}{\text{no. of elements in the sample}}$$

$$\hat{p} = \frac{X}{n} \qquad \text{(n=sample size=no. of elements in the sample)}$$

- $\hat{p}$ is a "good" point estimate for p.
- For large n, ($n \geq 30$, $np > 5$), we have

Test Procedure:(P_0 is known number)

Hypothesis	$H_0: P = P_0$ $H_A: P \neq P_0$	$H_0: P \leq P_0$ $H_A: P > P_0$	$H_0: P \geq P_0$ $H_A: P < P_0$
Test Statistic (T.S.)	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} \quad , \quad q_0 = 1 - p_0$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	$-Z_{1-\alpha/2}$ or $Z_{1-\alpha/2}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision : Reject H_0 if the following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$Z > Z_{1-\alpha/2}$ Or $Z < -Z_{1-\alpha/2}$	$Z > Z_{1-\alpha}$ (one – Sided Test)	$Z < -Z_{1-\alpha}$ (one – Sided Test)

Example:

A researcher was interested in the proportion of females in the population of all patients visiting a certain clinic. The researcher claims that 70% of all patients in this population are females. Would you agree with this claim if a random survey shows that 24 out of 45 patients are females? Use a 0.10 level of significance.

Solution:

p = Proportion of female in the population.

.n=45 (large)

X= no. of female in the sample = 24

$\hat{P}$ = proportion of females in the sample

$$\hat{p} = \frac{X}{n} = \frac{24}{45} = 0.5333$$

$$p_o = \frac{70}{100} = 0.7$$

$\alpha = 0.10$

Hypotheses:

$H_o: p = 0.7 \quad (p_o = 0.7)$

$H_A: p \neq 0.7$

Level of significance:

$\alpha = 0.10$

Test Statistic (T.S.):

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{0.5333 - 0.70}{\sqrt{\frac{(0.7)(0.3)}{45}}} = -2.44$$

Rejection Region of H_o (R.R.):

Critical values:

$Z_{\alpha/2} = Z_{0.05} = -1.645$

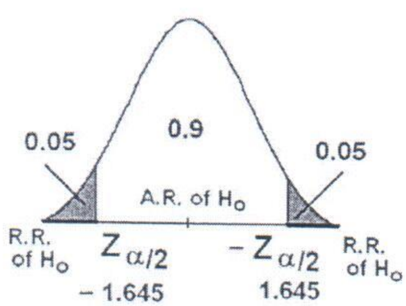
$-Z_{\alpha/2} = -Z_{0.05} = 1.645$

We reject H_o if:

$Z < Z_{\alpha/2} = Z_{0.05} = -1.645$

or

$Z > -Z_{\alpha/2} = -Z_{0.05} = 1.645$



$Z_{\alpha/2} = Z_{0.05} = -1.645$

Decision:

Since $Z = -2.44 \in$ Rejection Region of H_o (R.R), we reject

H_o:p=0.7 and accept H_A:$p \neq 0.7$ at α=0.1. Therefore, we do not agree with the claim stating that 70% of the patients in this population are females.

Example:

In a study on the fear of dental care in a certain city, a survey showed that 60 out of 200 adults said that they would hesitate to take a dental appointment due to fear. Test whether the proportion of adults in this city who hesitate to take dental appointment is less than 0.25. Use a level of significance of 0.025.

Solution:

p = Proportion of adults in the city who hesitate to take a dental appointment.

n= 200 (large)

X= no. of adults who hesitate in the sample = 60

$\hat{p}$ = proportion of adults who hesitate in the sample

$$\hat{p} = \frac{X}{n} = \frac{60}{200} = 0.3$$

p_0=0.25

α=0.025

Hypotheses:

H_o: $p \geq 0.25$ (p_o=0.25)

H_A: $p < 0.25$ (research hypothesis)

Level of significance:

α=0.025

Test Statistic (T.S.):

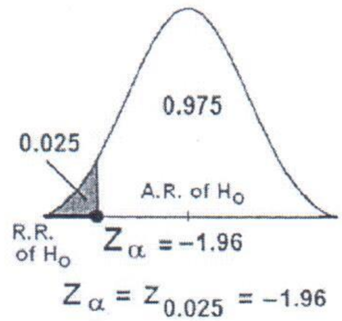
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.3 - 0.25}{\sqrt{\frac{(0.25)(0.75)}{200}}} = 1.633$$

Rejection Region of H_o (R.R.):

Critical value: $Z_\alpha = Z_{0.025} = -1.96$

Critical Region:

We reject H_o if: $Z < Z_\alpha = Z_{0.025} = -1.96$

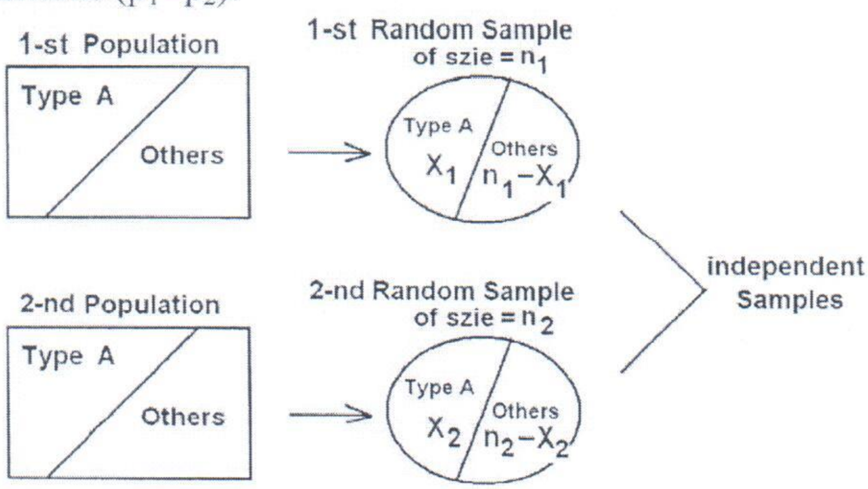


Decision:

Since $Z=1.633 \in$ Acceptance Region of H_o (A.R.), we accept (do not reject) H_o: $p \geq 0.25$ and we reject H_A: $p < 0.25$ at $\alpha=0.025$. Therefore, we do not agree with claim stating that the proportion of adults in this city who hesitate to take dental appointment is less than 0.25.

7.6 Hypothesis Testing: The Difference Between Two Population Proportions (p_1-p_2):

In this section, we are interested in testing some hypotheses about the difference between two population proportions (p_1-p_2).



Suppose that we have two populations:

- p_1 = population proportion of the 1-st population.
- p_2 = population proportion of the 2-nd population.
- We are interested in comparing p_1 and p_2, or equivalently, making inferences about p_1-p_2.
- We <u>independently</u> select a random sample of size n_1 from

the 1-st population and another random sample of size n_2 from the 2-nd population:

- Let X_1 = no. of elements of type A in the 1-st sample.
- Let X_2 = no. of elements of type A in the 2-nd sample.
- $\hat{p}_1 = \frac{X_1}{n_1}$ = the sample proportion of the 1-st sample
- $\hat{p}_2 = \frac{X_2}{n_2}$ = the sample proportion of the 2-nd sample
- The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is used to make inferences about $p_1 - p_2$.
- For large n_1 and n_2, we have

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1) \quad \text{(Approximately)}$$

- $q = 1 - p$

Hypotheses:

We choose one of the following situations:

(i) H_o: $p_1 = p_2$ against H_A: $p_1 \neq p_2$
(ii) H_o: $p_1 \geq p_2$ against H_A: $p_1 < p_2$
(iii) H_o: $p_1 \leq p_2$ against H_A: $p_1 > p_2$

or equivalently,

(i) H_o: $p_1 - p_2 = 0$ against H_A: $p_1 - p_2 \neq 0$
(ii) H_o: $p_1 - p_2 \geq 0$ against H_A: $p_1 - p_2 < 0$
(iii) H_o: $p_1 - p_2 \leq 0$ against H_A: $p_1 - p_2 > 0$

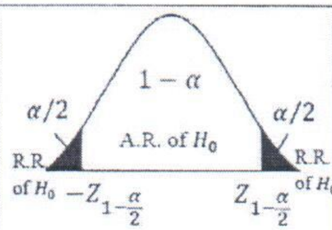
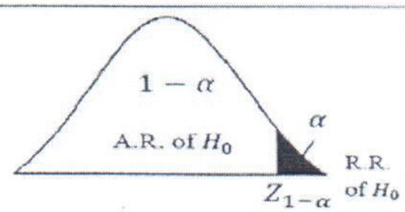
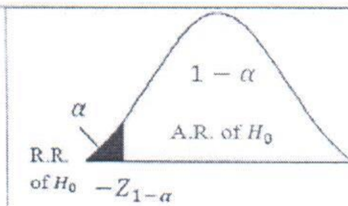
Note, under the assumption of the equality of the two population proportions (H_o: $p_1 = p_2 = p$), the pooled estimate of the common proportion p is:

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} \qquad (\bar{q} = 1 - \bar{p})$$

The test statistic (T.S.) is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}} \sim N(0,1)$$

Test Procedure:

Hypothesis	$H_0: P_1 - P_2 = 0$ $H_A: P_1 - P_2 \neq 0$	$H_0: P_1 - P_2 \leq 0$ $H_A: P_1 - P_2 > 0$	$H_0: P_1 - P_2 \geq 0$ $H_A: P_1 - P_2 < 0$
Test Statistic (T.S.)	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}}$, Pooled proportion: $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ $where \quad \bar{q} = 1 - \bar{p}$		
Rejection Region(R.R) & Acceptance Region(A.R)			
Reliability Coefficient	**$-Z_{1-\alpha/2}$ or $Z_{1-\alpha/2}$**	**$Z_{1-\alpha}$**	**$-Z_{1-\alpha}$**
Decision : Reject H_0 if the following condition satisfies	Reject H_0 (Accept H_A) at the significant level α if :		
	$Z > Z_{1-\alpha/2}$ Or $Z < -Z_{1-\alpha/2}$	$Z > Z_{1-\alpha}$ (one – Sided Test)	$Z < -Z_{1-\alpha}$ (one – Sided Test)

Example:

In a study about the obesity (overweight), a researcher was interested in comparing the proportion of obesity between males and females. The researcher has obtained a random sample of 150 males and another independent random sample of 200 females. The following results were obtained from this study.

	n	Number of obese people(X)
Males	150	21
Females	200	48

Can we conclude from these data that there is a difference between the proportion of obese males and proportion of obese females?
Use **α = 0.05.**

Solution

.p_1 = population proportion of obese males
.p_2 = population proportion of obese females
$\hat{p}_1$= sample proportion of obese males
$\hat{p}_2$= sample proportion of obese females

Males	Females
$n_1 = 150$	$n_2 = 200$
X_1=21	X_2=48
$\hat{p}_1 = \frac{X_1}{n_1} = \frac{21}{150} = 0.14$	$\hat{p}_2 = \frac{X_2}{n_2} = \frac{48}{200} = 0.24$

The pooled estimate of the common proportion p is:

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{21+48}{150+200} = 0.197$$

Hypotheses:

H$_o$: $p_1 = p_2$
H$_A$: $p_1 \neq p_2$

or

H$_o$: $p_1 - p_2 = 0$
H$_A$: $p_1 - p_2 \neq 0$

Level of significance: α=0.05

Test Statistic (T.S.):

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} = \frac{(0.14 - 0.24)}{\sqrt{\frac{0.197 \times 0.803}{150} + \frac{0.197 \times 0.803}{200}}} = -2.328$$

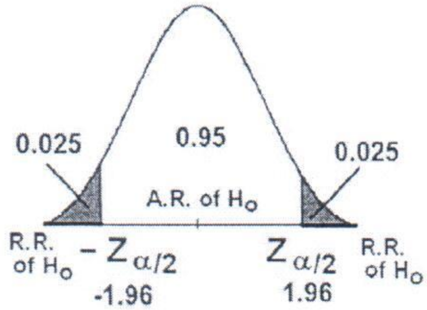
Rejection Region (R.R.) of H$_o$:

Critical values:

$Z_{\alpha/2} = Z_{0.025} = -1.96$
$Z_{1-\alpha/2} = Z_{0975} = 1.96$

Critical region:

Reject H$_o$ if: $Z < -1.96$ or $Z > 1.96$



Decision:

Since Z= −2.328 ∈R.R., we reject H_o: $p_1 = p_2$ and accept H_A: $p_1 \neq p_2$ at α=0.05. Therefore, we conclude that there is a difference between the proportion of obese males and the proportion of obese females. Additionally, since, $\hat{p}_1 = 0.14$ < $\hat{p}_2 = 0.24$, we may conclude that the proportion of obesity for females is larger than that for males.

TABLE OF CONTENTS

Subject	Page
Outline of the course	
Marks Distribution and Schedule of Assessment Tasks	
CHAPTER 1: Organizing and Displaying Data	
Introduction	
Statistics	
Biostatistics	
Populations	
Population Size	
Samples	
Sample Size	
Variables	
Types of Variables	
Types of Quantitative Variables	
Organizing the Data	
Simple frequency distribution or ungrouped frequency distribution	
Grouped Frequency Distributions	
Width of a class interval	
Displaying Grouped Frequency Distributions	
CHAPTER 2: Basic Summary Statistics	
Introduction	
Measures of Central Tendency	
Mean	
Population Mean	
Sample Mean	
Advantages and Disadvantages of the Mean	
Median	
Advantages and Disadvantages of the Median	
Mode	
Advantages and Disadvantages of the Mode	
Measures of Dispersion (Variation)	
Range	
Variance	
Deviations of Sample Values from the Sample Mean	
Population Variance	
Sample Variance	
Calculating Formula for the Sample Variance	
Standard Deviation	
Coefficient of Variation	

TABLE OF CONTENTS

Some Properties of the Mean, Standard Deviation, and Variance	
Calculating Measures from Simple Frequency Table	
Approximating Measures From Grouped Data	
CHPTER 3: Basic Probability Concepts	
General Definitions and Concepts	
Probability	
An Experiment	
Sample Space	
Events	
Equally Likely Outcomes	
Probability of an Event	
Some Operations on Events	
Union of Two events	
Intersection of Two Events	
Complement of an Event	
General Probability Rules	
Applications	
Conditional Probability	
Independent Events	
Bayes' Theorem	
Combinations:	
CHAPTER 4: Probability Distributions	
Introduction	
Probability Distributions of Discrete R.V.'s	
Graphical Presentation	
Population Mean of a Discrete Random Variable	
Cumulative Distributions	
Binomial Distribution	
Poisson Distribution	
Probability Distributions of Continuous R.V.	
Normal Distribution	
Standard Normal Distribution	
Calculating Probabilities of Standard Normal Distribution	
Calculating Probabilities of Normal Distribution	
Sampling Distribution of the Sample Mean	
Results about Sampling Distribution of the Sample Mean	
t-distribution	
CAPTER 5: Statistical Inferences	
Estimation and Hypotheses Testing	

TABLE OF CONTENTS

Estimation	
Estimation of the Population Mean	
Point Estimation of the population mean	
Interval Estimation of the population mean	
Estimation for the Population Proportion	
Point Estimate for the Population Proportion	
Interval Estimation for the Population Proportion	
Tests of Hypotheses	
Single Sample: Tests Concerning a Single Mean	
Single Sample: Tests on a Single Proportion:	
Two-sample: Paired t-test	