# Chapter 6 Estimation and Confidence Interval

## ***Estimation and Confidence Interval***

| Single Mean | Two Means |
|---|---|
| $\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ <br> $\sigma$ known | $(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ <br> $\sigma_1$ and $\sigma_2$ known |
| $\bar{X} \pm t_{1-\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}$ <br> $\sigma$ unknown | $(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2},\ (n_1+n_2-2)}\ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ <br> $\sigma_1$ and $\sigma_2$ unknown |
| **Single Proportion** | **Two Proportions** |
| For large sample size $(n \geq 30,\ np > 5,\ nq > 5)$ <br> $\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ | For large sample size $(n_1 \geq 30,\ n_1p_1 > 5,\ n_1q_1 > 5)$ $(n_2 \geq 30,\ n_2p_2 > 5,\ n_2q_2 > 5)$ <br> $(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ |

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

$$\bar{X} \pm \underbrace{\left( \overbrace{Z_{1-\frac{\alpha}{2}}}^{\text{Reliability coefficeint}} \frac{\sigma}{\sqrt{n}} \right)}_{\substack{\text{margin of error} \\ or \\ \text{precision of the estimate}}}$$

**Question 1:**

Suppose we are interested in making some statistical inference about the mean $\mu$, of a normal population with standard deviation $\sigma = 2$. Suppose that a random sample of size $n = 49$ from this population gave a sample mean $\bar{X} = 4.5$.

a. Find the upper limit of 95% of the confident interval for μ

$\sigma = 2 \quad \bar{X} = 4.5 \quad n = 49$

$95\% \rightarrow \alpha = 0.05 \qquad Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$

$$\bar{X} + \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 4.5 + \left(1.96 \times \frac{2}{7}\right) = 5.06$$

b. Find the lower limit of 95% of the confident interval for μ

$$\bar{X} - \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 4.5 - \left(1.96 \times \frac{2}{7}\right) = 3.94$$

**Question 2:**

A researcher wants to estimate the mean of a life span a certain bulb. Suppose that the distribution is normal with standard deviation 5 hours. Suppose that the researcher selected a random sample of 49 bulbs and found that the sample mean is 390 hours.

$\sigma = 5 \quad , \quad \bar{X} = 390 \quad , \quad n = 49$

a. find $Z_{0.975}$ :

$Z_{0.975} = 1.96$

b. find a point estimate for μ

$E(\bar{X}) = \hat{\mu} = \bar{X} = 390$

c. Find the upper limit of 95% of the confident interval for μ

$$\bar{X} + \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 390 + \left(1.96 \times \frac{5}{\sqrt{49}}\right) = 391.4$$

d. Find the lower limit of 95% of the confident interval for μ

$$\bar{X} - \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 390 - \left(1.96 \times \frac{5}{\sqrt{49}}\right) = 388.6$$

**Question 3:**

A sample of 16 college students were asked about time they spent doing their homework. It was found that the average to be 4.5 hours. Assuming normal population with standard deviation 0.5 hours.

$$\sigma = 0.5 \quad \bar{X} = 4.5 \quad n = 16$$

1. The point estimate for μ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0 hours | B | 10 hours | C | 0.5 hours | D | 4.5 hours |

2. The standard error of $\overline{X}$ is:

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{16}}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.125 hours | B | 0.266 hours | C | 0.206 hours | D | 0.245hours |

3. The correct formula for calculating $100\,(1-\alpha)\%$ confidence interval for μ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $\overline{X} \pm t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ | B | $\overline{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ | C | $\overline{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma^2}{n}$ | D | $\overline{X} \pm t_{1-\frac{\alpha}{2}} \frac{\sigma^2}{n}$ |

4. The upper limit of 95% confidence interval for μ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 4.745 | B | 4.531 | C | 4.832 | D | 4.891 |

5. The lower limit of 95% confidence interval for μ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 5.531 | B | 7.469 | C | 3.632 | D | 4.255 |

6. The length of the 95% confidence interval for μ is:

$$\text{Length} = 4.745 - 4.255 = 0.49$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 4.74 | B | 0.49 | C | 0.83 | D | 0.89 |

**Question 4:**

Let us consider a hypothetical study on the height of women in their adulthood. A sample of 24 women is drown from a normal distribution with population mean μ and variance $\sigma^2$. The sample mean and variance of height of the selected women are 151 cm and 18.65 $cm^2$ respectively. Using given data, we want to constract a 99% confidentce interval for the mean height of the adult women in the populatopn from which the sample was drown randomly.

$$\overline{X} = 151 \quad ; \quad n = 24 \quad ; \quad S^2 = 18.65 \Longrightarrow S = 4.32$$

a. Point estimate for μ

$$\hat{\mu} = \overline{X} = 151$$

b. Find the upper limit of 99% of the confident interval for μ

$$\overline{X} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right)$$

$$= 151 + \left(2.807 \times \frac{4.32}{\sqrt{24}}\right) = 153.4753$$

$$99\% \rightarrow \alpha = 0.01$$

$$t_{1-\frac{\alpha}{2},n-1} = t_{1-\frac{0.01}{2},24-1}$$

$$= t_{0.995,23} = 2.807$$

c. Find the lower limit of 99% of the confident interval for μ

$$\overline{X} - \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right)$$

$$= 151 - \left(2.807 \times \frac{4.32}{\sqrt{24}}\right) = 148.5247$$

***Estimation and Confidence Interval: Two Means***

1- $(\bar{X}_1 - \bar{X}_2) \pm \left( Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$

2- $(\bar{X}_1 - \bar{X}_2) \pm \left( t_{1-\frac{\alpha}{2}, n_1+n_2-2}\, Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$

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

## Question 5:

The tensile strength of type I thread is approximately normally distributed with standard deviation of 6.8 kg. A sample of 20 pieces of the thread has an average tensile strength of 72.8 kg. Another type of thread (type II) is approximately followed normal distribution with standard deviation 6.8 kg. A sample of 25 pieces of the thread has an average tensile strength pf 64.4 kg. then for 98% confidence interval of the difference in tensile strength means between type I and type II, we have:

$$\text{Theard 1}: \quad n_1 = 20, \bar{X}_1 = 72.8, \sigma_1 = 6.8$$
$$\text{Thread 2}: \quad n_2 = 25, \bar{X}_2 = 64.4, \sigma_2 = 6.8$$

$$98\% \rightarrow \alpha = 0.02 \rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.99} = 2.325$$

$$(\bar{X}_1 - \bar{X}_2) \pm \left( Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$
$$(72.8 - 64.4) \pm \left( 2.325 \times \sqrt{\frac{6.8^2}{20} + \frac{6.8^2}{25}} \right)$$

$$(3.657, 13.143)$$

(1): The lower limit $= 3.657$
(2): The upper limit $= 13.143$

# Question 6:

| | First sample | Second sample |
|---|---|---|
| Sample size (n) | 12 | 14 |
| Sample mean ($\overline{X}$) | 10.5 | 10 |
| Sample variance ($S^2$) | 4 | 5 |

1. Estimate the difference $\mu_1 - \mu_2$:

$$\hat{\mu}_1 - \hat{\mu}_2 = \overline{X}_1 - \overline{X}_2 = 10.5 - 10 = 0.5$$

2. Find the pooled standard deviation estimator Sp :

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{4(11) + 5(13)}{24} = 4.54 \Longrightarrow \boxed{Sp = 2.13}$$

3. The upper limit of 95% confidence interval for $(\mu_1 - \mu_2)$ is:

$$95\% \to \alpha = 0.05 \to t_{1-\frac{\alpha}{2}, n_1+n_2-2} = t_{0.975, 24} = 2.064,$$

$$(\overline{X}_1 - \overline{X}_2) + \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2} \times Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$(0.5) + \left(2.064 \times 2.13\sqrt{\frac{1}{12} + \frac{1}{14}}\right) = 2.23$$

4. The lower limit of 95% confidence interval for $(\mu_1 - \mu_2)$ is:

$$(\overline{X}_1 - \overline{X}_2) - \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2} \times Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$(0.5) - \left(2.064 \times 2.13\sqrt{\frac{1}{12} + \frac{1}{14}}\right) = -1.23$$

## Question 7:

A researcher was interested in comparing the mean score of female students $\mu_1$, with the mean score of male students $\mu_2$ in a certain test. Assume the populations of score are normal with equal variances. Two independent samples gave the following results:

| | Female | Male |
|---|---|---|
| Sample size | $n_1 = 5$ | $n_2 = 7$ |
| Mean | $\overline{X}_1 = 82.63$ | $\overline{X}_2 = 80.04$ |
| Variance | $S_1^2 = 15.05$ | $S_2^2 = 20.79$ |

1. The point estimate of $\mu_1 - \mu_2$ is:

| A | 2.63 | B | $-2.37$ | C | 2.59 | D | 0.59 |
|---|---|---|---|---|---|---|---|

2. The estimate of the pooled variance $S_p^2$ is:

| A | 17.994 | B | 18.494 | C | 17.794 | D | 18.094 |
|---|---|---|---|---|---|---|---|

3. The upper limit of the 95% confidence interval for $\mu_1 - \mu_2$ is :

| A | 26.717 | B | 7.525 | C | 7.153 | D | 8.2 |
|---|---|---|---|---|---|---|---|

4. The lower limit of the 95% confidence interval for $\mu_1 - \mu_2$ is :

| A | $-21.54$ | B | $-2.345$ | C | $-3.02$ | D | $-1.973$ |
|---|---|---|---|---|---|---|---|

***Estimation and Confidence Interval: Single Proportion***

For large sample size ($n \geq 30,\ np > 5,\ nq > 5$)

$$* \text{ Point estimate for } P \text{ is: } \frac{x}{n}$$

$$* \text{ Interval estimate for } P \text{ is: } \hat{p} \pm \left( Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

## Question 7:

A random sample of 200 students from a certain school showed that 15 student smoke. Let p be the proportion of smokers in the school.

1. Find a point estimate for p.

$$n = 200 \ \ \& \ \ x = 15$$

$$\hat{p} = \frac{x}{n} = \frac{15}{200} = 0.075 \rightarrow \hat{q} = 0.925$$

2. Find 95% confidence interval for p.

$$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

$$\hat{p} \pm \left( Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right) = 0.075 \pm \left( 1.96 \times \sqrt{\frac{0.075 \times 0.925}{200}} \right)$$

The 95% confidence interval is: $(0.038,\ 0.112)$

## Question 8:

A researcher's group has perfected a new treatment of a disease which they claim is very efficient. As evidence, they say that they have used the new treatment on 50 patients with the disease and cured 25 of them. To calculate a 95% confidence interval for the proportion of the cured.

1. The point estimate of p is equal to:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.25 | B | 0.50 | C | 0.01 | D | 0.33 |

2. The reliability coefficient $\left(z_{1-\frac{\alpha}{2}}\right)$ is equal is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 1.96 | B | 1.645 | C | 2.02 | D | 1.35 |

3. The 95% confidence interval is equal to:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | (0.1114,0.3886) | B | (0.3837,0.6163) | C | (0.1614,0.6386) | D | (0.3614,0.6386) |

***Estimation and Confidence Interval: Two Proportions***

For large sample size $(n_1 \geq 30,\ n_1 p_1 > 5,\ n_1 q_1 > 5)$
$(n_2 \geq 30,\ n_2 p_2 > 5,\ n_2 q_2 > 5)$

$$* \text{ Point estimate for } P_1 - P_2 = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$$

$$* \text{ Interval estimate for } P_1 - P_2 \text{ is: } (\hat{p}_1 - \hat{p}_2) \pm \left( Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$$

## Question 9:

A random sample of 100 students from school "A" showed that 15 students smoke. Another independent random sample of 200students from school "B" showed that 20 students smoke. Let $p_1$ be the proportion of smoker in school "A" and let $p_2$ be the proportion of smoker in school "B".

1. Find a point estimate for $P_1 - P_2$.

$$n_1 = 100\,, x_1 = 15 \ \rightarrow \ \hat{p}_1 = \frac{15}{100} = \boxed{0.15} \ \Longrightarrow \hat{q}_1 = 1 - 0.15 = \boxed{0.85}$$

$$n_2 = 200\,, x_2 = 20 \ \rightarrow \ \hat{p}_2 = \frac{20}{200} = \boxed{0.10} \ \Longrightarrow \hat{q}_2 = 1 - 0.10 = \boxed{0.90}$$

$$\boxed{\hat{p}_1 - \hat{p}_2 = 0.15 - 0.1 = 0.05}$$

2. Find 95% confidence interval for $P_1 - P_2$.

$$95\% \rightarrow \alpha = 0.05 \ \rightarrow \ Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

$$(\hat{p}_1 - \hat{p}_2) \pm \left( Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$$

$$= (0.05) \ \pm \left( 1.96 \times \sqrt{\frac{(0.15)(0.85)}{100} + \frac{(0.1)(0.9)}{200}} \right)$$

$$= 0.05 \pm \left( 1.96 \times \sqrt{0.001725} \right)$$

*The 95% confidence interval is:* $(-0.031,\ 0.131)$

**Question 10:**

a first sample of 100 store customers, 43 used a MasterCard. In a second sample of 100 store customers, 58 used a Visa card. To find the 95% confidence interval for difference in the proportion $(P_1 - P_2)$ of people who use each type of credit card?

1. The value of α is:

| A | 0.95 | B | 0.50 | C | 0.05 | D | 0.025 |
|---|---|---|---|---|---|---|---|

2. The upper limit of 95% confidence interval for the proportion difference is:

$$n_1 = 100\ ,\ x_1 = 43 \ \rightarrow\ \hat{p}_1 = \frac{43}{100} = \boxed{0.43} \ \Longrightarrow \hat{q}_1 = 1 - 0.43 = \boxed{0.57}$$
$$n_2 = 100\ ,\ x_2 = 58 \ \rightarrow\ \hat{p}_2 = \frac{58}{100} = \boxed{0.58} \ \Longrightarrow \hat{q}_2 = 1 - 0.58 = \boxed{0.42}$$

$$(\hat{p}_1 - \hat{p}_2) + \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$$

$$= (0.43 - 0.58) + \left(1.96 \times \sqrt{\frac{(0.43)(0.57)}{100} + \frac{(0.58)(0.42)}{100}}\right) = -0.013$$

3. The lower limit of 95% confidence interval for the proportion difference is:

$$(\hat{p}_1 - \hat{p}_2) - \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$$

$$= (0.43 - 0.58) - \left(1.96 \times \sqrt{\frac{(0.43)(0.57)}{100} + \frac{(0.58)(0.42)}{100}}\right) = -0.287$$

## Question from previous midterms and finals:

- In procedure of construction $(1-\alpha)100\%$ confidence interval for the population mean $(\mu)$ of a normal population with a known standard deviation $(\sigma)$ based on a random sample of size n.

1. The width of $(1-\alpha)100\%$ confidence interval for $(\mu)$ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $2\,Z_{1-\alpha}\frac{\sigma^2}{n}$ | B | $2\,Z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$ | C | $2\,Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ | D | $2\,Z_{1-\alpha}\frac{\sigma^2}{\sqrt{n}}$ |

2. For $\mathrm{n}=70$ and $\sigma=4$ the width of a 95% confidence interval for $(\mu)$ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 3.1458 | B | 1.5153 | C | 6.1601 | D | 1.8741 |

3. For $\overline{\mathrm{X}}=60$ and a 95% confidence interval for $\mu$ is $(57\,,\ k)$, then the value of the upper confidence limit k is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 64.5 | B | 66 | C | 61.5 | D | 63 |

4. When comparing the width of the 95% confidence interval (C.I.) for $\mu$ with that of 90% C.I., we found that:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 95% C.I. is shorter | B | 95% C.I. is wider | C | They have the same width | D | We can’t decide |

5. When the sample size n increase, the width of the C.I. will:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Decrease | B | Increase | C | Not be change | D | We can’t decide |

6. The most typical form of a calculated confidence interval is:

| | |
|---|---|
| A | Point estimate $\pm$ standard error |
| B | Population parameter $\pm$ margin of error |
| C | Population parameter $\pm$ standard error |
| D | Point estimate $\pm$ margin of error |

7. Confidence intervals are useful when trying to estimate ............. parameter:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Sample | B | Statistics | C | Population | D | None of these |

8. The following C.I. is obtained for a population proportion $(0.505, 0.545)$, then the margin of error equals (let $\hat{p}=0.525$)

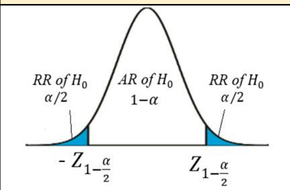
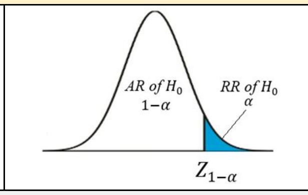
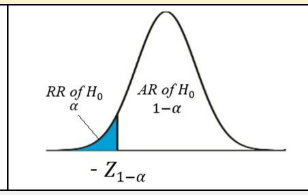
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.01 | B | 0.04 | C | 0.03 | D | 0.02 |

# Chapter 7 Hypotheses Testing

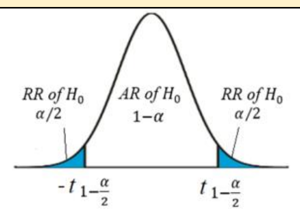
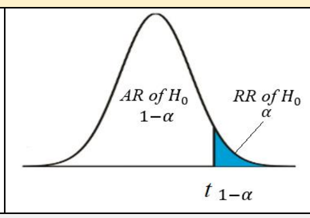
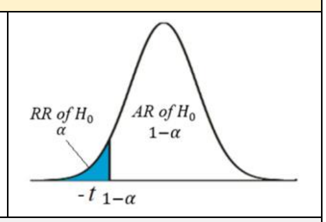
# ***Hypotheses Testing***

## 1-Single Mean

$(\, if\ \sigma\ known\,)$:

<table>
<tr><td>Hypothesis<br>Null $H_0$<br>Alternative (Research) $H_A$</td><td>$H_0: \mu = \mu_o$<br>$H_A: \mu \neq \mu_o$</td><td>$H_0: \mu \leq \mu_o$<br>$H_A: \mu > \mu_o$</td><td>$H_0: \mu \geq \mu_o$<br>$H_A: \mu < \mu_o$</td></tr>
<tr><td>Test Statistics<br>(TS)</td><td colspan="3">$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \ \sim\ N(0,1)$</td></tr>
<tr><td>Rejection Region<br>(RR) of $H_0$<br>Acceptance Region<br>(AR) of $H_0$</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td rowspan="2">Decision</td><td colspan="3">We reject $H_0$ at the significance level $\alpha$ if</td></tr>
<tr><td>$Z < -Z_{1-(\alpha/2)}$<br>or<br>$Z > \ Z_{1-(\alpha/2)}$<br>Two sides test</td><td>$Z > Z_{1-\alpha}$<br><br>One side test</td><td>$Z < -Z_{1-\alpha}$<br><br>One side test</td></tr>
</table>

$(\, if\ \sigma\ unknown\,)$:

<table>
<tr><td>Hypothesis<br>Null $H_0$<br>Alternative (Research) $H_A$</td><td>$H_0: \mu = \mu_o$<br>$H_A: \mu \neq \mu_o$</td><td>$H_0: \mu \leq \mu_o$<br>$H_A: \mu > \mu_o$</td><td>$H_0: \mu \geq \mu_o$<br>$H_A: \mu < \mu_o$</td></tr>
<tr><td>Test Statistics<br>(TS)</td><td colspan="3">$t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \ \sim t_{n-1}$</td></tr>
<tr><td>Rejection Region<br>(RR) of $H_0$<br>Acceptance Region<br>(AR) of $H_0$</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td rowspan="2">Decision</td><td colspan="3">We reject $H_0$ at the significance level $\alpha$ if</td></tr>
<tr><td>$t < -t_{1-(\alpha/2)}$<br>or<br>$t > \ t_{1-(\alpha/2)}$<br>Two sides test</td><td>$t > t_{1-\alpha}$<br><br>One side test</td><td>$t < -t_{1-\alpha}$<br><br>One side test</td></tr>
</table>

## Question 1:

*Suppose that we are interested in estimating the true average time in seconds it takes an adult to open a new type of tamper-resistant aspirin bottle. It is known that the population standard deviation is $\sigma = 5.71$seconds. A random sample of 40 adults gave a mean of 20.6 seconds. Let $\mu$ be the population mean, then, to test if the mean$\mu$ is 21 seconds at level of significant 0.05 ($H_0: \mu = 21$ vs $H_A: \mu \neq 21$ ) then:*

(1) The value of the test statistic is:

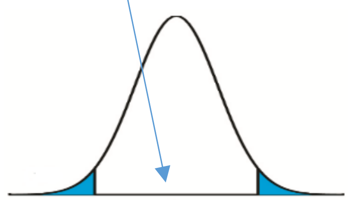
$$\boxed{\sigma = 5.71 \quad n = 40 \quad \bar{X} = 20.6}$$

$$Z = \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}} = \frac{20.6 - 21}{5.71/\sqrt{40}} = \boxed{-0.443}$$

| A | 0.443 | B | −0.012 | C | −0.443 | D | 0.012 |
|---|---|---|---|---|---|---|---|

(2) The acceptance area is:

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$




| A | (−1.96,1.96) | B | (1.96, ∞) | C | (−∞, 1.96) | D | (−∞, 1.645) |
|---|---|---|---|---|---|---|---|

(3) The decision is:

| A | Reject $H_0$ | B | Accept $H_0$ | C | No decision | D | None of these |
|---|---|---|---|---|---|---|---|

$$P - value = 2 \times P(Z < -0.443) = 2 \times 0.32997 = 0.66 > 0.05$$

*or*

$$P - value = 2 \times P(Z > |-0.443|) = 2 \times P(Z > 0.443) = 0.66 > 0.05$$

## Question 2:

If the hemoglobin level of pregnant women (امرأه حامل) is normally distributed, and if the mean and standard deviation of a sample of 25 pregnant women were $\overline{X} = 13$ (g/dl), $s = 2$ (g/dl). Using $\alpha = 0.05$, to test if the average hemoglobin level for the pregnant women is greater than 10 (g/dl) [$H_0$: $\mu \leq 10$, $H_A$: $\mu > 10$].

$$\boxed{s = 2 \ , \ \ n = 25 \ , \ \ \bar{X} = 13}$$

1. The test statistic is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $Z = \frac{\bar{X}-10}{\sigma/\sqrt{n}}$ | B | $Z = \frac{\bar{X}-10}{S/\sqrt{n}}$ | C | $t = \frac{\bar{X}-10}{\sigma/\sqrt{n}}$ | D | $t = \frac{\bar{X}-10}{S/\sqrt{n}}$ |

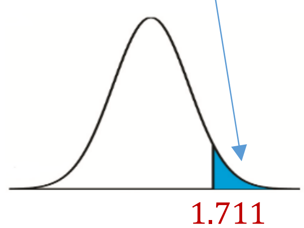
2. The value of the test statistic is:

$$t = \frac{\bar{X} - \mu_o}{S/\sqrt{n}} = \frac{13 - 10}{2/\sqrt{25}} = \boxed{7.5}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 10 | B | 1.5 | C | 7.5 | D | 37.5 |

3. The rejection of $H_0$ is:

$$t_{1-\alpha,n-1} = t_{0.95,24} = 1.711$$



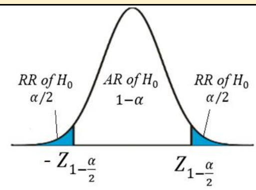
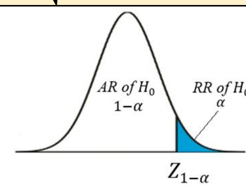
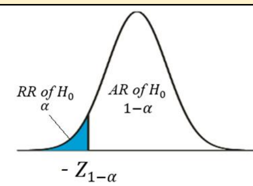

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $Z < -1.645$ | B | $Z > 1.645$ | C | $t < -1.711$ | D | $t > 1.711$ |

4. The decision is:

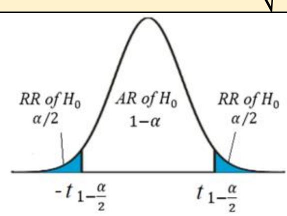
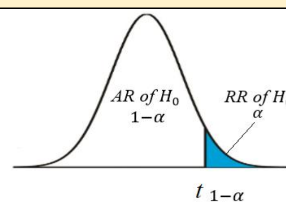
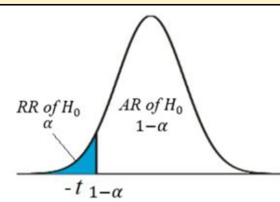
| | |
|---|---|
| A | Reject $H_0$ |
| B | Do not reject (Accept) $H_0$. |
| C | Accept both $H_0$ and $H_A$. |
| D | Reject both $H_0$ and $H_A$. |

**<u>2-Two Means:</u>**

$(\ if\ \sigma_1\ and\ \sigma_2\ known\ ):$

| *Hypothesis*<br>*Null $H_0$*<br>*Alternative (Research) $H_A$* | $H_0: \mu_1 - \mu_2 = d$<br>$H_A: \mu_1 - \mu_2 \neq d$ | $H_0: \mu_1 - \mu_2 \leq d$<br>$H_A: \mu_1 - \mu_2 > d$ | $H_0: \mu_1 - \mu_2 \geq d$<br>$H_A: \mu_1 - \mu_2 < d$ |
|---|---|---|---|
| *Test Statistics*<br>*(TS)* | $Z = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ | | |
| *Rejection Region*<br>*(RR) of $H_0$*<br>*Acceptance Region*<br>*(AR) of $H_0$* |  |  |  |
| *Decision* | We reject $H_0$ at the significance level $\alpha$ if | | |
| | $Z < -Z_{1-(\alpha/2)}$<br>*or*<br>$Z > Z_{1-(\alpha/2)}$<br>*Two sides test* | $Z > Z_{1-\alpha}$<br>*One side test* | $Z < -Z_{1-\alpha}$<br>*One side test* |

$(\ if\ \sigma_1\ and\ \sigma_2\ unknown\ ):$

| *Hypothesis*<br>*Null $H_0$*<br>*Alternative (Research) $H_A$* | $H_0: \mu_1 - \mu_2 = d$<br>$H_A: \mu_1 - \mu_2 \neq d$ | $H_0: \mu_1 - \mu_2 \leq d$<br>$H_A: \mu_1 - \mu_2 > d$ | $H_0: \mu_1 - \mu_2 \geq d$<br>$H_A: \mu_1 - \mu_2 < d$ |
|---|---|---|---|
| *Test Statistics*<br>*(TS)* | $t = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ | | |
| *Rejection Region*<br>*(RR) of $H_0$*<br>*Acceptance Region*<br>*(AR) of $H_0$* |  |  |  |
| *Decision* | We reject $H_0$ at the significance level $\alpha$ if | | |
| | $t < -t_{1-(\alpha/2)}$<br>*or*<br>$t > t_{1-(\alpha/2)}$<br>*Two sides test* | $t > t_{1-\alpha}$<br>*One side test* | $t < -t_{1-\alpha}$<br>*One side test* |

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

## Question 3:

A standardized chemistry test was given to 50 girls and 75 boys. The girls made an average of 84, while the boys made an average grade of 82. Assume the population standard deviations are 6 and 8 for girls and boys respectively. To test the null hypothesis

$$H_0: \mu_1 - \mu_2 \leq 0 \;\; vs \; H_A: \mu_1 - \mu_2 > 0 \quad use\ \alpha = 0.05$$

(1) The standard error of $(\bar{X}_1 - \bar{X}_2)$ is:

$$girls: \quad n_1 = 50,\ \bar{X}_1 = 84,\ \sigma_1 = 6$$
$$boys: \quad n_2 = 75,\ \bar{X}_2 = 82,\ \sigma_2 = 8$$

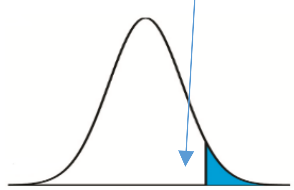
$$S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{6^2}{50} + \frac{8^2}{75}} = 1.2543$$

(2) The value of the test statistic is:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(84-82)}{\sqrt{\frac{6^2}{50} + \frac{8^2}{75}}} = \frac{2}{1.2543} = \boxed{1.5945}$$

(3) The rejection region (RR) of $H_0$ is:

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.645$$




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $(1.645, \infty)$ | B | $(-\infty, -1.645)$ | C | $(1.96, \infty)$ | D | $(-\infty, -1.96)$ |

(4) The decision is:

| | |
|---|---|
| A | Reject $H_0$ |
| B | Do not reject (Accept) $H_0$. |
| C | Accept both $H_0$ and $H_A$. |
| D | Reject both $H_0$ and $H_A$. |

$$P - value = P(Z > 1.59) = 1 - P(Z < 1.59) = 0.056 > 0.05$$

## Question 4:

Cortisol level determinations were made on two samples of women at childbirth. Group 1 subjects underwent emergency cesarean section (عملية قيصرية) following induced labor. Group 2 subjects natural childbirth route following spontaneous labor (الولادة الطبيعية). The random sample sizes, mean cortisol levels, and standard deviations were $(n_1 = 40, \bar{x}_1 = 575, \sigma_1 = 70)$, $(n_2 = 44, \bar{x}_2 = 610, \sigma_2 = 80)$.

If we are interested to test if the mean Cortisol level of group 1 ($\mu_1$) is less than that of group 2 ($\mu_2$) at level 0.05 (or$H_0: \mu_1 \geq \mu_2$ vs $H_1: \mu_1 < \mu_2$), then:

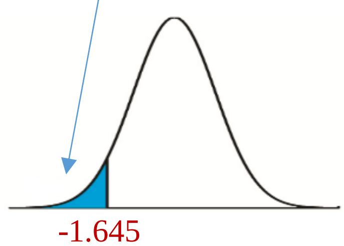
(1) The value of the test statistic is:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(575 - 610)}{\sqrt{\frac{70^2}{40} + \frac{80^2}{44}}} = \boxed{-2.138}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $-1.326$ | B | $-2.138$ | C | $-2.576$ | D | $-1.432$ |

(2) Reject $H_0$ if :

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $Z > 1.645$ | B | $T > 1.98$ | C | $Z < -1.645$ | D | $T < -1.98$ |

(3) The decision is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | No decision | D | None of these |

$$P - value = P(Z < -2.138) = 0.01618 < 0.05$$

## Question 5:

An experiment was conducted to compare time length (duration time in minutes) of two types of surgeries (A) and (B). 10 surgeries of type (A) and 8 surgeries of type (B) were performed. The data for both samples is shown below.

| Surgery type | A | B |
|---|---|---|
| Sample size | 10 | 8 |
| Sample mean | 14.2 | 12.8 |
| Sample standard deviation | 1.6 | 2.5 |

Assume that the two random samples were independently selected from two normal populations with equal variances. If $\mu_A$ and $\mu_B$ are the population means of the time length of surgeries of type (A) and type (B), then, to test if $\mu_A$ is greater than $\mu_B$ at level of significant 0.05 ($H_0: \mu_A \leq \mu_B$ vs $H_A: \mu_A > \mu_B$) then:

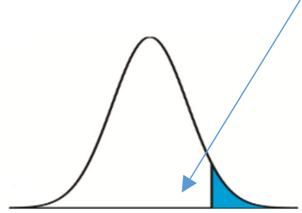
1. The value of the test statistic is:

$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{1.6^2(10-1)+2.5^2(8-1)}{10+8-2} = 4.174$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(14.2-12.8)}{\sqrt{4.174}\sqrt{\frac{1}{10}+\frac{1}{8}}} = \boxed{1.44}$$

2. Reject $H_0$ if:

$$t_{1-\alpha,n_1+n_2-2} = t_{0.95,10+8-2} = t_{0.95,16} = 1.746$$



1.746

| A | $Z > 1.645$ | B | $Z < -1.645$ | C | $T > 1.746$ | D | $T < -1.746$ |
|---|---|---|---|---|---|---|---|

3. The decision is:

| A | Reject $H_0$ | B | Accept $H_0$ | C | No decision | D | None of these |
|---|---|---|---|---|---|---|---|

## Question 6:

A researcher was interested in comparing the mean score of female students $\mu_1$ ,with the mean score of male students $\mu_2$ in a certain test. Assume the populations of score are normal with equal variances. Two independent samples gave the following results:

| | Female | Male |
|---|---|---|
| Sample size | $n_1 = 5$ | $n_2 = 7$ |
| Mean | $\bar{X}_1 = 82.63$ | $\bar{X}_2 = 80.04$ |
| Variance | $S_1^2 = 15.05$ | $S_2^2 = 20.79$ |

Test that is there is a difference between the mean score of female students and the mean score of male students.

1. The hypotheses are:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $H_0: \mu_1 = \mu_2$<br>$H_A: \mu_1 \neq \mu_2$ | B | $H_0: \mu_1 = \mu_2$<br>$H_A: \mu_1 < \mu_2$ | C | $H_0: \mu_1 < \mu_2$<br>$H_A: \mu_1 > \mu_2$ | D | $H_0: \mu_1 \leq \mu_2$<br>$H_A: \mu_1 > \mu_2$ |

2. The value of the test statistic is:

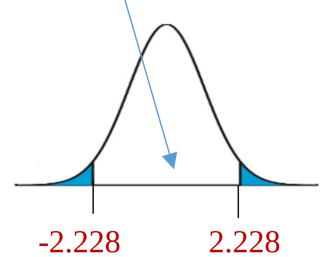
$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{15.05(4)+20.79(6)}{5+7-2} = 18.494$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{82.63-80.04}{\sqrt{18.494}\sqrt{\frac{1}{5}+\frac{1}{7}}} = \boxed{1.029}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 1.3 | B | 1.029 | C | 0.46 | D | 0.93 |

3. The acceptance region (AR) of $H_0$ is:

$$t_{1-\frac{\alpha}{2},n_1+n_2-2}$$
$$= t_{1-\frac{0.05}{2},5+7-2}$$
$$= t_{0.975,10}$$
$$= 2.228$$




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $(2.228, \infty)$ | B | $(-\infty, -2.228)$ | C | (-2.228, 2.228) | D | (-1.96, 1.96) |

## Question 7:

A nurse researcher wished to know if graduates of baccalaureate (بكالوريس) nursing program and graduate of associate degree (الزمالة) nursing program differ with respect to mean scores on personality inventory at α = 0.02. A sample of 50 associate degree graduates (sample A) and a sample of 60 baccalaureate graduates (sample B) yielded the following means and standard deviations:

$$\bar{X}_A = 88.12, \; S_A = 10.5, \; n_A = 50$$
$$\bar{X}_B = 83.25, \; S_B = 11.2, \; n_B = 60$$

1) The hypothesis is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $H_0: \mu_1 = \mu_2$ <br> $H_A: \mu_1 \neq \mu_2$ | B | $H_0: \mu_1 \leq \mu_2$ <br> $H_A: \mu_1 > \mu_2$ | C | $H_0: \mu_1 \geq \mu_2$ <br> $H_A: \mu_1 < \mu_2$ | D | None of these |

2) The test statistic is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Z | B | t | C | F | D | None of these |

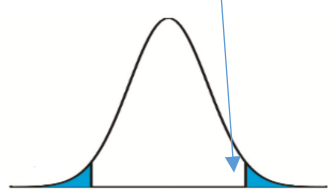
3) The computed value of the test statistic is:

$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{10.5^2(50-1)+11.2^2(60-1)}{50+60-2} = 118.55$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{88.12-83.25}{\sqrt{118.55}\sqrt{\frac{1}{50}+\frac{1}{60}}} = \boxed{2.34}$$

4) The critical region (rejection area) is:

$$t_{1-\frac{\alpha}{2}, n_1+n_2-2}$$
$$= t_{1-\frac{0.02}{2},\; 50+60-2}$$
$$= t_{0.99,108} = 2.36$$

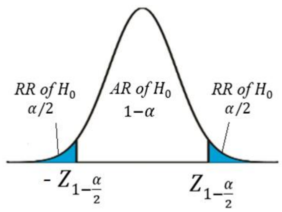
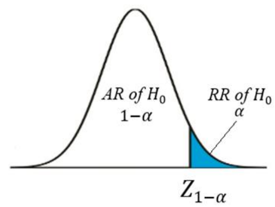
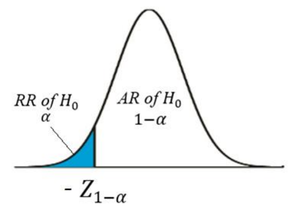



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 2.60 or − 2.60 | B | 2.06 or − 2.06 | C | 2.36 or − 2.36 | D | 2.58 |

5) Your decision is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | Accept $H_A$ | D | No decision |

**3- Single proportion:**

| *Hypothesis*<br>*Null $H_0$*<br>*Alternative (Research) $H_A$* | $H_0: p = p_o$<br>$H_A: p \neq p_o$ | $H_0: p \leq p_o$<br>$H_A: p > p_o$ | $H_0: p \geq p_o$<br>$H_A: p < p_o$ |
|---|---|---|---|
| *Test Statistics (TS)* | $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,1)$ | | |
| *Rejection Region (RR) of $H_0$*<br>*Acceptance Region (AR) of $H_0$* |  |  |  |
| *Decision* | We reject $H_0$ at the significance level $\alpha$ if | | |
| | $Z < -Z_{1-(\alpha/2)}$<br>*or*<br>$Z > Z_{1-(\alpha/2)}$<br>*Two sides test* | $Z > Z_{1-\alpha}$<br>*One side test* | $Z < -Z_{1-\alpha}$<br>*One side test* |

## Question 8:

Toothpaste (معجون الأسنان) company claims that more than 75% of the dentists recommend their product to the patients. Suppose that 161 out of 200 dental patients reported receiving a recommendation for this toothpaste from their dentist. Do you suspect that the proportion is actually morethan75%. If we use 0.05 level of significance to test $H_0: P \leq 0.75$, $H_A: P > 0.75$, then:

(1) The sample proportion $\hat{p}$ is:

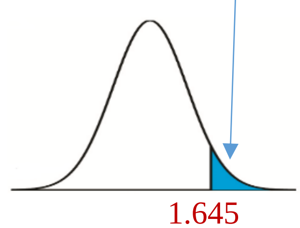
$$n = 200, \quad \hat{p} = \frac{161}{200} = 0.8050$$

(2) The value of the test statistic is:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.805 - 0.75}{\sqrt{\frac{(0.75)(0.25)}{200}}} = \boxed{1.7963}$$

(3) The decision is:

$$\alpha = 0.05$$
$$Z_{1-\alpha} = Z_{0.95} = 1.645$$




| | |
|---|---|
| A | Reject $H_0$ (agree with the claim) |
| B | Do not reject (Accept) $H_0$ |
| C | Accept both $H_0$ and $H_A$ |
| D | Reject both $H_0$ and $H_A$ |

$$P - value = P(Z > 1.7963) = 1 - P(Z < 1.7963) = 1 - 0.96407 = 0.03593 < 0.05$$

**Question 9:**

A researcher was interested in studying the obesity (السمنة) disease in a certain population. A random sample of 400 people was taken from this population. It was found that 152 people in this sample have the obesity disease. If p is the population proportion of people who are obese. Then, to test if p is greater than 0.34 at level 0.05 ($H_0: p \leq 0.34$ vs $H_A: p > 0.34$) then:

(1) The value of the test statistic is:

$$n = 400, \quad \hat{p} = \frac{152}{400} = 0.38$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.38 - 0.34}{\sqrt{\frac{0.34 \times 0.66}{400}}} = \boxed{1.69}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.023 | B | 1.96 | C | 2.50 | D | 1.69 |

(2) The P-value is

$$P - value = P(Z > 1.69) = 1 - P(Z < 1.69) = 1 - 0.9545 = 0.0455$$

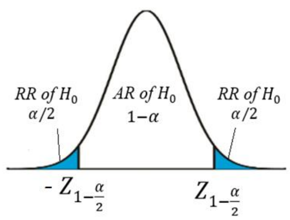
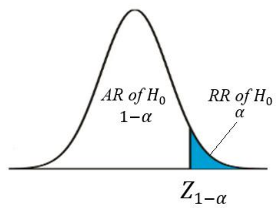
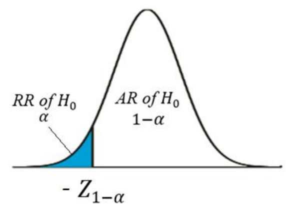
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.9545 | B | 0.0910 | C | 0.0455 | D | 1.909 |

(3) The decision is:

$$P - value = 0.0455 < 0.05$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | No decision | D | None of these |

**4-Two proportions:**

| *Hypothesis*<br>*Null $H_0$*<br>*Alternative (Research) $H_A$* | $H_0: p_1 - p_2 = d$<br>$H_A: p_1 - p_2 \neq d$ | $H_0: p_1 - p_2 \leq d$<br>$H_A: p_1 - p_2 > d$ | $H_0: p_1 - p_2 \geq d$<br>$H_A: p_1 - p_2 < d$ |
|---|---|---|---|
| *Test Statistics (TS)* | $Z = \frac{(\hat{p}_1 - \hat{p}_2) - d}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - d}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$ | | |
| *Rejection Region (RR) of $H_0$*<br>*Acceptance Region (AR) of $H_0$* | <br> | <br> | <br> |
| *Decision* | We reject $H_0$ at the significance level $\alpha$ if | | |
| | $Z < -Z_{1-(\alpha/2)}$<br>*or*<br>$Z > Z_{1-(\alpha/2)}$<br>*Two sides test* | $Z > Z_{1-\alpha}$<br>*One side test* | $Z < -Z_{1-\alpha}$<br>*One side test* |

## Question 10:

In a first sample of 200 men, 130 said they used seat belts and a second sample of 300 women, 150 said they used seat belts. To test the claim that men are more safety-conscious than women ($H_0: p_1 - p_2 \leq 0, H_1: p_1 - p_2 > 0$), at 0.05 level of significant:

(1) The value of the test statistic is:

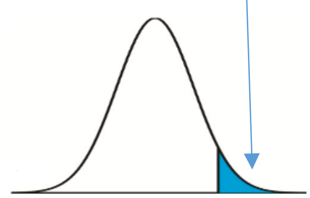
$$n_1 = 200, \ \hat{p}_1 = \frac{130}{200} = 0.65 \qquad n_2 = 300, \ \hat{p}_2 = \frac{150}{300} = 0.5$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{130+150}{200+300} = 0.56$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.65 - 0.5)}{\sqrt{(0.56)(0.44)\left(\frac{1}{200} + \frac{1}{300}\right)}} = \boxed{3.31}$$

(2) The decision is:

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.645$$




| | |
|---|---|
| A | Reject $H_0$ |
| B | Do not reject (Accept) $H_0$ |
| C | Accept both $H_0$ and $H_A$ |
| D | Reject both $H_0$ and $H_A$ |

$$P - value = P(Z > 3.31) = 1 - P(Z < 3.31) = 1 - 0.99953 = 0.00047 < 0.05$$

## Question 11:

In a study of diabetes, the following results were obtained from samples of males and females between the ages of 20 and 75. Male sample size is 300 of whom 129 are diabetes patients, and female sample size is 200 of whom 50 are diabetes patients. If $P_M, P_F$ are the diabetes proportions in both populations and $\hat{p}_M, \hat{p}_F$ are the sample proportions, then:
A researcher claims that the Proportion of diabetes patients is found to be more in males than in female ($H_0: P_M - P_F \leq 0$ vs $H_A: P_M - P_F > 0$) . Do you agree with his claim, take $\alpha = 0.10$

$$n_m = 300 \quad , \quad x_m = 129 \quad \Longrightarrow \hat{p}_1 = \frac{129}{300} = 0.43$$

$$n_f = 200 \quad , \quad x_f = 50 \quad \Longrightarrow \hat{p}_2 = \frac{50}{200} = 0.25$$

(1) The pooled proportion is:

$$\bar{p} = \frac{x_m + x_f}{n_m + n_f} = \frac{129 + 50}{300 + 200} = 0.358$$

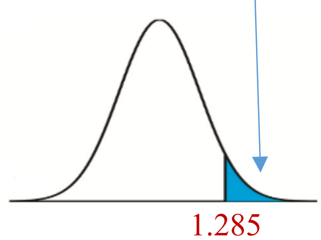
| A | 0.43 | B | 0.18 | C | 0.358 | D | 0.68 |
|---|---|---|---|---|---|---|---|

(2) The value of the test statistic is:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.43 - 0.25)}{\sqrt{(0.358)(1 - 0.358)\left(\frac{1}{300} + \frac{1}{200}\right)}} = \boxed{4.11}$$

(3) The decision is:

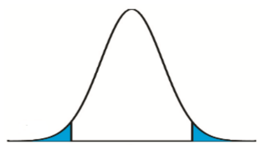
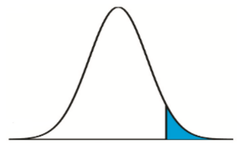
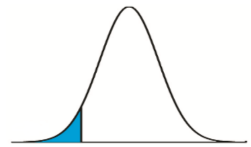
$$Z_{1-\alpha} = Z_{1-0.10} = Z_{0.90} = 1.285$$




| A | Agree with the claim (Reject $H_o$) |
|---|---|
| B | do not agree with the claim |
| C | Can't say |

$$P - value = P(Z > 4.11) = 1 - P(Z < 4.11) = 1 - 1 = 0 < 0.05$$

- *P – value:*

| *Hypothesis* | $H_0: \mu = \mu_o$<br>$H_A: \mu \neq \mu_o$ | $H_0: \mu \leq \mu_o$<br>$H_A: \mu > \mu_o$ | $H_0: \mu \geq \mu_o$<br>$H_A: \mu < \mu_o$ |
|---|---|---|---|
| *RR* |  |  |  |
| *P-value* | $2 \times P(Z > \|TS\|)$ | $P(Z > TS)$ | $P(Z < TS)$ |

| $2 \times P(Z > TS)$<br>*If* $TS > 0$ | $2 \times P(Z < TS)$<br>*If* $TS < 0$ |
|---|---|

| | population normal or not normal<br>n large ($n \geq 30$) | | population normal<br>n small ($n < 30$) | |
|---|---|---|---|---|
| | σ known | σ unknown | σ known | σ unknown |
| Testing | $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ | $Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ | $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ | $T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ |

- **Two Samples Test for Paired Observation**

**<u>Question 1:</u>**

The following contains the calcium levels of eleven test subjects at zero hours and three hours after taking a multi-vitamin containing calcium.

| Pair | 0 hour ( $X_i$ ) | 3 hours ( $Y_i$ ) | Difference $D_i = X_i - Y_i$ |
|---|---|---|---|
| 1 | 17.0 | 17.0 | 0.0 |
| 2 | 13.2 | 12.9 | 0.3 |
| 3 | 35.3 | 35.4 | -0.1 |
| 4 | 13.6 | 13.2 | 0.4 |
| 5 | 32.7 | 32.5 | 0.2 |
| 6 | 18.4 | 18.1 | 0.3 |
| 7 | 22.5 | 22.5 | 0.0 |
| 8 | 26.8 | 26.7 | 0.1 |
| 9 | 15.1 | 15.0 | 0.1 |

The sample mean and sample standard deviation of the differences D are 0.144 and 0.167, respectively. To test whether the data provide sufficient evidence to indicate a difference in mean calcium levels ($H_0: \mu_1 = \mu_2$ against$H_1: \mu_1 \neq \mu_2$) with $\alpha = 0.10$ we have: $\boxed{\bar{D} = 0.144 \;,\; S_d = 0.167 \;,\; n = 9}$

[1]. The reliability coefficient (the tabulated value) is:

$$t_{1-\frac{\alpha}{2},n-1} = t_{1-\frac{0.1}{2},9-1} = t_{0.95,8} = \boxed{1.860}$$

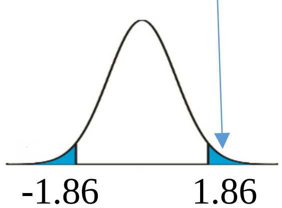
[2]. The value of the test statistic is:

$$\boxed{\begin{matrix} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{matrix}} \Longrightarrow \boxed{\begin{matrix} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{matrix}} \Longrightarrow \boxed{\begin{matrix} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{matrix}}$$

$$T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}} = \frac{0.144 - 0}{0.167/\sqrt{9}} = \boxed{2.5868}$$

[3]. The decision is:

We reject $H_0$



## Question 2:

Scientists and engineers frequently wish to compare two different techniques for measuring or determining the value of a variable. Reports the accompanying data on amount of milk ingested by each of 14 randomly selected infants.

| Pair | DD method ($X_i$) | TW method ($Y_i$) | Difference $D_i = X_i - Y_i$ |
|---|---|---|---|
| 1 | 1509 | 1498 | 11 |
| 2 | 1418 | 1254 | 164 |
| 3 | 1561 | 1336 | 225 |
| 4 | 1556 | 1565 | -9 |
| 5 | 2169 | 2000 | 169 |
| 6 | 1760 | 1318 | 442 |
| 7 | 1098 | 1410 | -312 |
| 8 | 1198 | 1129 | 69 |
| 9 | 1479 | 1342 | 137 |
| 10 | 1281 | 1124 | 157 |
| 11 | 1414 | 1468 | -54 |
| 12 | 1954 | 1604 | 350 |
| 13 | 2174 | 1722 | 452 |
| 14 | 2058 | 1518 | 540 |

1. The sample mean of the differences $\overline{D}$ is:

$$\overline{D} = \frac{11+164+225-9+169+442-312+\cdots+540}{14} = 167.21$$

| A | 167.21 | B | 0.71 | C | 0.61 | D | 0.31 |
|---|---|---|---|---|---|---|---|

2. The sample standard deviation of the differences $S_D$ is:

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}} = 228.21$$

| A | 3.15 | B | −0.71 | C | 71.53 | D | 228.21 |
|---|---|---|---|---|---|---|---|

3. The reliability coefficient to construct 90% confidence interval for the true average difference between intake values measured by the two methods $\mu_D$ is:

$$\text{The reliability coefficient} = t_{1-\frac{\alpha}{2},n-1} = t_{0.95,13} = 1.771$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 1.96 | B | 1.771 | C | 2.58 | D | 1.372 |

4. The 90% lower limit for $\mu_D$ is:

$$= \bar{D} - \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 167.21 - \left(1.771 \times \frac{228.12}{\sqrt{14}}\right) = 59.19$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 24.92 | B | 22.55 | C | 59.19 | D | 44.96 |

5. The 90% upper limit for $\mu_D$ is:

$$= \bar{D} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 167.21 + \left(1.771 \times \frac{228.12}{\sqrt{14}}\right) = 275.23$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 224.92 | B | 322.55 | C | 275.23 | D | 24.96 |

To test $H_0: \mu_D = 0$ versus $H_A: \mu_D \neq 0$ , $\alpha = 0.10$ as a level of significance we have:

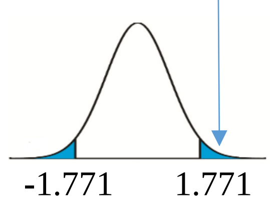
6. The value of the test statistic is:

$$T = \frac{\bar{D}-\mu_D}{S_d/\sqrt{n}} = \frac{167.21-0}{228.12/\sqrt{14}} = \boxed{2.74}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 2.74 | B | −0.7135 | C | −7.153 | D | −0.3157 |

7. The decision is:

Reject $H_0$




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | No decision | D | None of these |

**Question 3:**

In a study of a surgical procedure used to decrease the amount of food that person can eat. A sample of 10 persons measures their weights before and after one year of the surgery, we obtain the following data:

| Before surgery (X) | 148 | 154 | 107 | 119 | 102 | 137 | 122 | 140 | 140 | 117 |
|---|---|---|---|---|---|---|---|---|---|---|
| After surgery (Y) | 78 | 133 | 80 | 70 | 70 | 63 | 81 | 60 | 85 | 120 |
| $D_i = X_i - Y_i$ | 70 | 21 | 27 | 49 | 32 | 74 | 41 | 80 | 55 | -3 |

We assume that the data comes from normal distribution.
For 90% confidence interval for $\mu_D$, where $\mu_D$ is the difference in the average weight before and after surgery.

1. The sample mean of the differences $\bar{D}$ is:

$$\bar{D} = \frac{70+21+27+49+32+74+41+80+55-3}{10} = 44.6$$

2. The sample standard deviation of the differences $S_D$ is:

$$S_D = \sqrt{\frac{\sum(D_i - \bar{D})^2}{n-1}} = 26.2$$

3. The 90% upper limit of the confidence interval for $\mu_D$ is:

$$t_{1-\frac{\alpha}{2},n-1} = t_{0.95,9} = 1.833$$

$$= \bar{D} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 44.6 + \left(1.833 \times \frac{26.2}{\sqrt{10}}\right) = 59.79$$

4. To test $H_0: \mu_D \geq 43$ versus $H_A: \mu_D < 43$, with $\alpha = 0.10$ as a level of significance, the value of the test statistic is:

$$T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}} = \frac{44.6-43}{26.2/\sqrt{10}} = \boxed{0.19}$$

5. The decision is:

$$-t_{1-\alpha,n-1} = -t_{0.90,9} = -1.383 \Longrightarrow \boxed{0.19 \notin RR: (-\infty, -1.383)}$$

| A | Reject $H_0$ | B | Do not reject $H_0$ | C | No decision | D | None of these |
|---|---|---|---|---|---|---|---|

**Questions 4:**

Trace metals in drinking water affect the flavor and an unusually high concentration can pose a health hazard. Ten pairs of data were taken measuring zinc concentration in bottom water and surface water.

The Data is given below:

| | zinc concentration in Bottom water | zinc concentration in Surface water | Difference |
|---|---|---|---|
| 1 | 0.43 | 0.415 | 0.015 |
| 2 | 0.266 | 0.238 | 0.028 |
| 3 | 0.567 | 0.39 | 0.177 |
| 4 | 0.531 | 0.41 | 0.121 |
| 5 | 0.707 | 0.605 | 0.102 |
| 6 | 0.716 | 0.609 | 0.107 |
| 7 | 0.651 | 0.632 | 0.019 |
| 8 | 0.589 | 0.523 | 0.066 |
| 9 | 0.469 | 0.411 | 0.058 |
| 10 | 0.723 | 0.612 | 0.111 |

Note that the mean and the standard deviation of the difference are given respectively by $\overline{D}$ = 0.0804 and $S_D$ = 0.0523 We want to determine the 95 % confidence interval for $\mu_1 - \mu_2$, where $\mu_1$ and $\mu_2$ represent the true mean zinc concentration in Bottom water and surface water respectively. Assume the distribution of the differences to be approximately normal.

1. The 95% lower limit for $\mu_1 - \mu_2$ equals to:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.02628 | B | 0.13452 | C | 0.04299 | D | 0.11781 |

2. The 95% upper limit for $\mu_1 - \mu_2$ equals to:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 0.02628 | B | 0.13452 | C | 0.04299 | D | 0.11781 |

| | Estimation | Testing |
|---|---|---|
| Single mean | $\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ $\sigma$ known | $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ $\sigma$ known |
| | $\bar{X} \pm t_{1-\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}$ $\sigma$ unknown | $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ $\sigma$ unknown |
| Two means | $(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $\sigma_1$ and $\sigma_2$ known | $Z = \frac{(\bar{X}_1-\bar{X}_2)-d}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ $\sigma_1$ and $\sigma_2$ known |
| | $(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2},\ (n_1+n_2-2)}\ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $\sigma_1$ and $\sigma_2$ unknown | $T = \frac{(\bar{X}_1-\bar{X}_2)-d}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\sigma_1$ and $\sigma_2$ unknown |
| Single proportion | $\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ | $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ |
| Two proportions | $(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ | $Z = \frac{(\hat{p}_1-\hat{p}_2)-d}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ |

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

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accepting $H_0$ | Correct decision ✓ | Type II error ($\beta$) |
| Rejecting $H_0$ | Type I error ($\alpha$) | Correct decision ✓ |

| | |
|---|---|
| Type I error = Rejecting $H_0$ when $H_0$ is true<br>P(Type I error) = P(Rejecting $H_0 \mid H_0$ is true) = α | Type II error = Accepting $H_0$ when $H_0$ is false<br>P(Type II error) = P(Accepting $H_0 \mid H_0$ is false) = $\beta$ |

- **Question from previous midterms and finals:**

**Q1.** In the procedure of testing the statistical hypotheses $H_0$ against $H_A$ using a significance level α

1. The type I error occur if we:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Rejecting $H_0$ when $H_0$ is true | B | Rejecting $H_0$ when $H_0$ is false | C | Accepting $H_0$ when $H_0$ is true | D | Accepting $H_0$ when $H_0$ is false |

2. The probability of type I error is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $\beta$ | B | α | C | $1-\beta$ | D | $1-\alpha$ |

3. The level of significance is:

| | | | |
|---|---|---|---|
| A | The probability of rejecting $H_A$ | B | The probability of rejecting $H_0$ |
| C | The probability of making a Type I error | D | The probability of making a Type II error |

4. When we use P-value method, we reject $H_0$ if

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | P- value > α | B | P- value < α | C | P- value < $\beta$ | D | P- value > $\beta$ |

5. If the P-value = 0.0625 and α = 0.05, the decision is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | Reject $H_A$ | D | Accept $H_A$ |

6. To determine the rejection region for $H_0$, it depends on:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | α and $H_A$ | B | $H_0$ | C | α and $H_0$ | D | $\beta$ |

7. Which one is an example of two-tailed test:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $H_A: \mu = 0$ | B | $H_A: \mu \neq 0$ | C | $H_A: \mu < 0$ | D | $H_A: \mu > 0$ |

8. If the P-value = 0.0625 and α = 0.05, the decision is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | Reject $H_0$ | B | Accept $H_0$ | C | Reject $H_A$ | D | Accept $H_A$ |

9. If the distribution of the random sample is normal and standard deviation of the population is known, which type of the interval should be considered?

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | z - interval | B | x - interval | C | t - interval | D | c - interval |

10. An appropriate 95% CI for $\mu$ has been calculated as (-1.5 , 3.5) based on $n_1 = 15$, $n_2 = 17$ observations from two independent population with normal distribution. The hypotheses of interest $H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 \neq \mu_2$. Based on this CI and at $\alpha = 0.05$,

| | | | |
|---|---|---|---|
| A | We should reject $H_0$ | B | We should not reject $H_0$ |

**Q2.** To compare the mean times spent waiting for a heart transplant for two age groups, you randomly select several people in each age group who have had a heart transplant. The result is shown below. Assume both population is are normally distributed with equal variance.

| Sample statistics for heart transplant | | |
|---|---|---|
| Age group | 18-34 | 35-49 |
| Mean | 171 days | 169 days |
| Standard deviation | 8.5 days | 11.5 days |
| Sample size | 20 | 17 |

Do this data provide sufficient evident to indicate a difference among the population means at $\alpha = 0.05$

1. The alternative hypothesis is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $H_A: \mu_1 \neq \mu_2$ | B | $H_A: \mu_1 \leq \mu_2$ | C | $H_A: \mu_1 > \mu_2$ | D | $H_A: \mu_1 = \mu_2$ |

2. The pooled estimator of the common variance $S_p^2$ is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 9935.82 | B | 105.5214 | C | 10.4429 | D | 99.6786 |

3. The appropriate test statistics is:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$ | B | $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | C | $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ | D | $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$ |

4. The 95% confidence interval for the different in mean times spent waiting for heart transplant for the two age groups:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $(-3.548, 7.565)$ | B | $(-0.1306, 4.1306)$ | C | $(-4.6862, 8.6862)$ | D | $(-4.8519, 8.8519)$ |

5. Base on the 95% C.I. in the above question, it can be concluded that:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | $\bar{X}_1 = \bar{X}_2$ | B | $\mu_1 \neq \mu_2$ | C | $\mu_1 = \mu_2$ | D | None of these |