# STAT 2035 ACE Exam Booklet Solutions (Winter 2026)

## A. Comparing Two Population Means by Using Independent Samples: Variances Known

**Example 1.** a)

$n_1 = 50 \quad n_2 = 65$
$\bar{x}_1 = 75 \quad \bar{x}_2 = 78$
$\sigma_1 = 10 \quad \sigma_1 = 12$

$H_0 \ \mu_1 = \mu_2$
$H_a \ \mu_1 \neq \mu_2 \quad (2\ sided)$
$Z_{crit} = Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$
$DR\ reject\ H_0\ if\ H_0\ if\ Z_{test} > 1.96\ or\ < -1.96$

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

$$\therefore z_{test} = \frac{75-78}{\sqrt{\frac{10^2}{50} + \frac{12^2}{65}}} = \frac{-3}{2.053140184} = -1.46$$

$Z_{test} = -1.46$
$Z_{test} > -1.96$
$\therefore do\ not\ reject\ \ H_0$
$\therefore there\ is\ no\ significant\ difference\ between\ means$

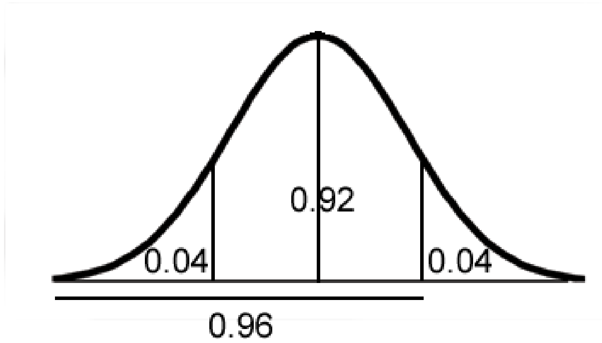
Or use p-values and look up the
p-value= 2(Pr(Z<-1.46)=0.0721(2)=0.1388 >alpha=0.05,
do not reject H0
$\therefore there\ is\ no\ significant\ difference\ between\ means$

**Example 2.**

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}$$

$=(20 - 25) \pm 1.96\sqrt{\frac{2^2}{100} + \frac{3^2}{100}}$
$=-5 \pm 0.71$
$=(-5.71, -4.29)$

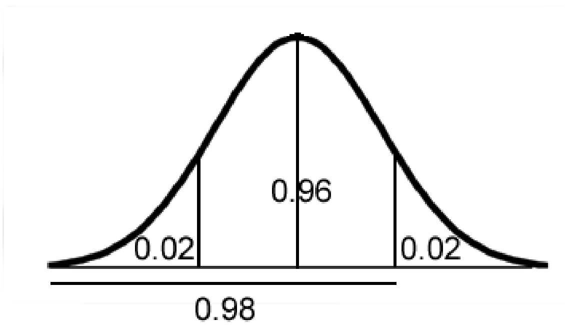
**Example 3**. a)




$\leftarrow\leftarrow\leftarrow\leftarrow\leftarrow |look\ up\ in\ body$

$z = 1.75$

b)




$\leftarrow\leftarrow\leftarrow\leftarrow\leftarrow |look\ up\ in\ body$

$z = 2.05$

A1. $H_0 \quad \mu_1 - \mu_2 = 0$

$H_a \quad \mu_1 - \mu_2 \neq 0\ (2\ sided)$

$\overline{x_1} = 95 \quad \overline{x_2} = 97 \quad n_1 = n_2 = 50$

$\sigma_1 = 5.55 \quad \sigma_2 = 4.55$

$$Z\ test = \frac{\overline{x_1} - \overline{x_2} - do}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{95 - 97 - 0}{\sqrt{\frac{5.55^2}{50} + \frac{4.55^2}{50}}} = -1.97$$

Look up the area and the area below -1.97 is 0.0244, so the p-value for a two-sided test is 0.0244(2) = 0.0488... so the p-value is below 5% and we reject H0 and conclude there is a significant difference in the rates.

A2. $H_0 \quad \mu_1 - \mu_2 = 0$ let alpha = 0.05

$H_a \quad \mu_1 - \mu_2 > 0\ (1\ sided)$

$\overline{x_1} = 250 \qquad \overline{x_2} = 243 \qquad n_1 = 150, n_2 = 200$

$\sigma_1 = 23 \ \ \sigma_2 = 19$

$$Z\ test = \frac{\overline{x_1} - \overline{x_2} - do}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{250 - 243 - 0}{\sqrt{\frac{23^2}{150} + \frac{19^2}{200}}} = 3.03$$

So, it is a one-sided test and we want the area above 3.03. The area below 3.03 is 0.9988 so that means for this greater than test, we want the area 1-0.9988=0.0012. So, our p-value is less than 5% and we reject H0 we conclude that there is statistically significant evidence to support their claim that Florida costs more.

A3. $H_0 \quad \mu_1 - \mu_2 = 0$

$H_a \quad \mu_1 - \mu_2 \neq 0\ (2\ sided)$

$\overline{x_1} = 26 \quad \sigma_1 = 12.1 \qquad n_1 = 75$

$\overline{x_2} = 34 \quad \sigma_2 = 9.6 \quad n_2 = 50$

$$Z\ test = \frac{\overline{x_1} - \overline{x_2} - do}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{26 - 34 - 0}{\sqrt{\frac{12.1^2}{75} + \frac{9.6^2}{50}}} = -4.11$$

It is a two-sided test, so we look up the area below -4.11 on the Z table and we double but the area below -3.49 is almost 0 (0.0002), so the area below -4.11 is almost 0 too and this means it is less than alpha=0.01, so we reject H0 and conclude there is statistical evidence of a difference at the 1% level.

A4.

$n_1 = 20000 \qquad \overline{x_1} = 414 \quad \sigma 1 = 85$

$n_2 = 20000 \qquad \overline{x_2} = 387 \quad \sigma 2 = 75$

$H_0 \quad \mu_1 - \mu_2 = 0$

$H_a \quad \mu_1 - \mu_2 \neq 0$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

$=(414 - 387) \pm 2.576\sqrt{\frac{85^2}{20000} + \frac{75^2}{20000}}$

$=27 \pm 2.064$

$\mu_1 - \mu_2$=(24.936, 29.064) 0 is not in here, so reject H0 and there is a significant difference.

The difference between the mean 1 and 2 is between 24.936 to 29.064.

A5. $H_0 \quad \mu_1 - \mu_2 = 0$
$H_a \quad \mu_1 - \mu_2 \neq 0 \ (2\ sided)$

$$Z\ test = \frac{375 - 362 - 0}{\sqrt{\frac{110^2}{25} + \frac{125^2}{25}}} = 0.39$$

Pr(Z<0.39)=0.6517
Pr(Z>0.39)= 1- 0.6517=0.3413 (area on one side)

p-value=2(0.3483)> alpha=0.05, so we fail to reject H0 and there is no statistical evidence that the machines are different

## B. Comparing Two Population Means: Variances Unknown

**Example 1.**

a) State the appropriate hypotheses to be tested.

**Solution**

$H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 > 0$

we want to know if the treatment 2 takes less time than treatment 1to relief..ie treatment 2 better

$n_1$=10, $n_2$=10, $s_1$=5.2, $s_2$=4.9, $\overline{x_1} = 22.6 \ and \ \overline{x_2} = 19.4$

$$\frac{5.2}{4.9} = 1.06 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{5.2^2(9) + 4.9^2(9)}{10 + 10 - 2} = 25.525$$

$$s_p = 5.05$$

$$df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18 \quad \therefore df = 18$$

b) Compute the test statistic.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{22.6-19.4}{5.05\sqrt{\frac{1}{10}+\frac{1}{10}}} = 1.42$$

c) Find the corresponding *p*-value.

1.330<t=1.42<1.734

0.05<one sided p-value<0.10

p-value>0.05=$\alpha$

d) State the decision and conclusion using a 5% level of significance.

$\therefore$We fail to reject $H_o$. There is not sufficient evidence to suggest the new treatment causes a reduction in throbbing

**Example 2.**

$H_0 \quad \mu_1 - \mu_2 = 0$
$H_a \quad \mu_1 - \mu_2 \neq 0$

$\overline{x1} = 63 \qquad \overline{x2} = 73$
s=23.26 s=8.72

$\frac{23.26}{8.72} = 2.67 > 2 \ldots not\ pooled$

$$t\ test = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{63 - 73}{\sqrt{\frac{23.26^2}{8} + \frac{8.72^2}{8}}} = -1.14$$

$go\ across\ 7\ df$ (smaller of the two df)

$1.14 < t = 1.415$

$2\ sided\ p - value > 2(0.10)$

$\therefore p - value > 0.20$

$p - value > \alpha = 0.10$ and we fail to reject H0 and conclude there is no difference in the performance of the two groups

**Example 3.**

**(a)** Perform a 2-sample *t*-test to test the hypothesis that males and females have the same mean pulse rate.

**Solution:**

$H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$.

n1=30, n2=25, $\overline{x1} = 72.75, \overline{x2} = 73.55, s1 = 5.4, s2 = 7.7$

$$\frac{7.69987}{5.37225} = 1.43 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{7.7^2(24) + 5.4^2(29)}{30 + 25 - 2}$$

$$= 42.8$$

$$s_p = 6.542$$

$$df = n_1 + n_2 - 2 = 30 + 25 - 2 = 53 \quad \therefore use\ df = 50$$

The test statistic is

$$ttest = \frac{\overline{x_1} - \overline{x_2}}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{72.75 - 73.55}{6.542\sqrt{\frac{1}{30} + \frac{1}{25}}} = -0.45$$

The *p*-value is t=0.45<1.299

2 sided p-value >2(0.10)=0.2

From the tables, so since p-value>10%, so we fail to reject $H_0$ at any reasonable significance level because the *p*-value is very large. The largest alpha we use is 10%. (alpha ranges from 1 % to 10 %)

There is no significant difference between male and female pulse rates)

**(b)** Create a 90% confidence interval for the difference in mean pulse rates. Does your interval agree with the conclusion that you drew in the previous question?

$\alpha = 0.10$, df=50 from part a), alpha=0.10, so alpha/2=0.05

t critical=1.676 (50 df, 90% CI)

The 90% confidence interval for $\mu_1 - \mu_2$ is therefore

$$\mu_1 - \mu_2 = (\overline{x_1} - \overline{x_2}) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (72.75 - 73.55) \pm 1.676(6.542)\sqrt{\frac{1}{30} + \frac{1}{25}}$$
$$= -0.8 \pm 2.97 = (-3.77, 2.17)$$

We are 90% confident that the mean pulse rate for men is between 3.77 lower and 2.17 points higher than the mean pulse rate for women. Since 0 is in this interval, there is no evidence of a different in mean pulse rate for men and women. ( same conclusion as part a)

B1.

$H_\circ : \mu_1 - \mu_2 = 0$
$H_a : \mu_1 - \mu_2 \neq 0$ 2-sided p-value

$$\frac{5.8}{4.4} = 1.32 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{5.8^2(14) + 4.4^2(9)}{15 + 10 - 2} = 28.052$$

$$s_p = 5.296$$

$$df = n_1 + n_2 - 2 = 15 + 10 - 2 = 23 \quad \therefore df = 23$$

$$ttest = \frac{\overline{x_1} - \overline{x_2}}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{56 - 59}{5.296\sqrt{\frac{1}{15} + \frac{1}{10}}} = -1.39$$

at 23df we get

1.319<t=1.39<1.714

∴0.05(2)<p-value<0.1(2)

∴0.1<p-value<0.2

∴p-value> $\alpha = 0.10$

So, we fail to reject $H_0$ and conclude there is no difference between the two groups.

B2. a) Construct a 90% confidence interval for the difference between the population mean service-rating scores given by male and female guests at Jamaican 5-star hotels.

$$\frac{6.94}{6.73} = 1.03 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{6.73^2(126) + 6.94^2(113)}{127 + 114 - 2} = 46.65$$

$$s_p = 6.83$$

$$df = n_1 + n_2 - 2 = 127 + 114 - 2 = 239 \ \therefore df = 200$$

$$\mu_1 - \mu_2 =$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (39.08 - 38.79) \pm 1.653(6.83)\sqrt{\frac{1}{127} + \frac{1}{114}}$$

$$= 0.29 \pm 1.46$$

$$= (-1.17, 1.75)$$

0 is in the interval, so we fail to reject H0 and conclude the two hotels are statistically equal.

b) Use the interval to make an inference about whether the perception of service quality at five-star hotels in Jamaica differs by gender.

$$H_\circ : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

Since the 90% confidence interval contains 0, we fail to reject H0. So, we cannot conclude at the 10% level of significance that the perception of service quality differs by gender.

$\mu_1 - \mu_2 =$ -1.17 to 1.75
Mean 1 is between 1.7 lower than Mean 2 to 1.75 higher than Mean 2.

B3.

a) Set up the null and alternative hypothesis for determining whether the average height of Australian boys who repeated a grade is less than the average height of boys who never repeated.

$H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 < 0$

$\mu_1$ is mean of repeated

$\mu_2$ is mean of never repeated

* one-sided p-value

b) Conduct the test in part a) using a significance level of 5%.

$$\frac{1.18}{0.98} = 1.204 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{1.18^2(84) + 0.98^2(1349)}{85 + 1350 - 2} = 0.986$$

$$s_p = 0.993$$

$$df = n_1 + n_2 - 2 = 1350 + 85 - 2 = 1433 \quad \therefore use\ df = 200$$

$$ttest = \frac{\overline{x_1} - \overline{x_2}}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-0.05 - 0.35}{0.993\sqrt{\frac{1}{85} + \frac{1}{1350}}} = -3.60$$

Critical method: reject H0 if t test < -1.653 (df=200, 90% CI)

200df

t= -3.60< -3.131 (less than test)

one-sided p-value<0.001

∴p-value < $\alpha = 0.05$

So, we reject $H_0$

∴There is sufficient evidence to suggest that the mean height for boys who repeated is lower than those who did not repeat

B4. $n_2 = 8 \quad n_1 = 6$
$\bar{x}_2 = 40.5 \quad \bar{x}_1 = 52$
$s_2 = 2.62 \quad s_1 = 8.05$
$= \frac{8.05}{2.62} = 3.07 > 2 \ldots not\ pooled$

$H_0 \quad \mu_1 - \mu_2 = 0$
$H_a \quad \mu_1 - \mu_2 \neq 0$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}} = \frac{52 - 40.5}{\sqrt{\frac{8.05^2}{6} + \frac{2.62^2}{8}}} = 3.37$$

$go\ across\ 5\ df$ (smaller of the two df)

3.365 < t test=3.37 <4.032
0.005(2)< 2 sided p-value < 0.01(2)
0.01<p<0.02

$\therefore pvalue < \alpha = 0.05$
$\therefore\ reject\ H_0 \quad \therefore there\ is\ a\ significant\ difference\ in\ filler\ content$
Critical Method
Go to 0.05/2=0.025 on top and 5df down the side and get 2.571
Reject H0 if t test >2.571 or if t test < -2.571
t test = 3.37 which is in the rejection region on the right, so reject H0

b) 0.005<p<0.001 <0.05 and still the same conclusion
Critical Method
Go to 0.05 on top and 5df down the side and get 2.015, and it is a greater than test, we would reject if t test>2.015. Since t test=3.37>2.015, so we still reject H0.

B5. $\bar{x}_1 = 77.1 \quad \bar{x}_2 = 63$
$n_1 = 8 \quad n_2 = 8$
$s_1 = 5.82 \quad s_2 = 13.7$

$H_0 \quad \mu_1 - \mu_2 = 0$
$H_a \quad \mu_1 - \mu_2 \neq 0$

$= \frac{13.7}{5.82} = 2.35 > 2 \ldots not\ pooled$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}} = \frac{77.1 - 63}{\sqrt{\frac{5.82^2}{8} + \frac{13.7^2}{8}}} = 2.68$$

$go\ across\ 7\ df$ (smaller of the two df)

$2.365 < t = 2.68 < 2.998$

$2(0.01) < 2\ sided\ p - value < 2(0.025)$

$\therefore 0.02 < p - value < 0.05$

$pvalue < \alpha = 0.05$ and we reject H0 and conclude there is a difference in the performance of the two groups

B6. CI always 2 sided (regardless of < or > in question)

$$\frac{100}{90} = 1.11 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{90^2(15) + 100^2(15)}{16 + 16 - 2} = 9050$$

$$s_p = 95.13$$

$$df = n_1 + n_2 - 2 = 16 + 16 - 2 = 30 \quad \therefore df = 30$$

90% CI.... $t_\alpha$ =1.697 (0.05 in each tail)

$$\mu_1 - \mu_2 =$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_\alpha (s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (600 - 550) \pm 1.697(95.13)\sqrt{\frac{1}{16} + \frac{1}{16}}$$

$$= 50 \pm 57.08$$

$$= (-7.08, 107.08)$$

Since the interval contains “0”, there is no statistically significant evidence to suggest an improvement in score and we would not reject H0

B7. a)

$$\frac{3.1}{2.4} = 1.29 < 2 \ldots use\ pooled$$

$$s_p{}^2 = \frac{s_1{}^2(n_1 - 1) + s_2{}^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{2.4^2(40) + 3.1^2(30)}{41 + 31 - 2} = 7.41$$

$$s_p = 2.72$$

$$df = n_1 + n_2 - 2 = 41 + 31 - 2 = 70 \quad \therefore df = 70$$

$$H_\circ : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{34.3 - 43.7}{2.72\sqrt{\frac{1}{41} + \frac{1}{31}}} = -14.52$$

df=70 and t=14.52>3.211

2 sided p-value <2(0.001)
...p-value<0.002<0.01= alpha

∴reject $H_0$
So, there is evidence of a difference between the delivery times of the two stores.

b) Set up a 99% confidence interval estimate of the difference between the population means between Store A and Store B.

df=70 and 99%CI...$t_\alpha$=2.648

$$\mu_1 - \mu_2 =$$
$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (34.3 - 43.7) \pm 2.648(2.72)\sqrt{\frac{1}{41} + \frac{1}{31}}$$
$$= -9.4 \pm 1.714$$
$$= (-11.11, -7.686)$$

0 is not in this interval, so we reject $H_0$
The difference in means $\mu_1 - \mu_2 = (-11.11, -7.686)$

B8. $H_0 \quad \mu_1 - \mu_2 = 0 \qquad H_a \quad \mu_1 - \mu_2 \neq 0$

$\frac{s_1}{s_2} = \frac{8600}{7300} = 1.18 < 2 \quad pooled$

$df = n1 + n2 - 2 = 185 + 187 - 2 = 370 \quad (use\ 200)$

$s_p{}^2 = \frac{s_1{}^2(n1-1)+s_2{}^2(n2-1)}{n1+n2-2} = \frac{8600^2(184)+7300^2(186)}{370}$

$= 63\,569135.14$

$s_p = 7973.03$

$t = \frac{\overline{x_1}-\overline{x_2}-0}{s_p\sqrt{\frac{1}{n1}+\frac{1}{n2}}} = \frac{15000-16000}{7973.03\sqrt{\frac{1}{185}+\frac{1}{187}}} = -1.21$

At 200 $df \quad t\ test = 1.21 < 1.286$

$\therefore 2\ sided\ p - value\ > 2(0.10) = 0.2$

$\therefore p - value > \alpha \quad fail\ to\ reject\ H_0$ and there is no evidence of a difference.

B9.

$H_0 \quad \mu_1 - \mu_2 = 0$

$H_a \quad \mu_1 - \mu_2 \neq 0$ (make sure you take the square root to get s, instead of variance)

$\frac{s_1}{s_2} = \frac{12}{9} = 1.33 < 2 \quad pooled$

95% $CI \quad df = 25 + 25 - 2 = 48\ (use\ 40)$ t crit= 2.021 (95%)

reject H0 if t test >2.021 or <-2.021

$s_p{}^2 = \frac{s_1{}^2(n1-1)+s_2{}^2(n2-1)}{n1+n2-2} = \frac{9^2(24)+12^2(24)}{48}$

=112.5

$s_p$ =10.6

$$\mu_1 - \mu_2 = (\overline{x_1} - \overline{x_2}) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (95 - 100) \pm 2.021(10.6)\sqrt{\frac{1}{25} + \frac{1}{25}}$$

$$= -5 \pm 6.059 = (-11.06, 1.06)$$

The answer is A).

B10.
$H_0 \quad \mu_1 - \mu_2 = 0$
$H_a \quad \mu_1 - \mu_2 < 0$

df=48 (use 40)...one sided test

$$t = \frac{\overline{x_1} - \overline{x_2} - 0}{S_p\sqrt{\frac{1}{n1} + \frac{1}{n2}}} = \frac{95 - 100}{10.6\sqrt{\frac{1}{25} + \frac{1}{25}}} = -1.67$$

go across to 40 df and find 1.67

1.303< 1.67<1.684
find one sided p-values

0.05< one sided p<0.10

So, at the 10% level of significance, p-value would be less than alpha and we would reject and at the 5% level we would fail to reject $H_0$

At the 10% level, there is statistical evidence that $\mu_1 < \mu_2$

At the 5%, 2% or 1%, there is no statistical evidence $\mu_1 < \mu_2$

## C. Comparing Two Population Proportions

**Example 1.**

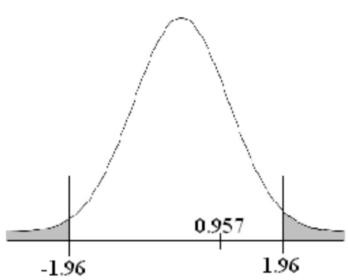
a) At the 5% level of significance, is there a significant difference between English-speaking Catholics and French-speaking Catholics in the proportion that agree that divorcees should be able to remarry in the Church?

$$H_o : p_1 - p_2 = 0$$
$$H_a : p_1 - p_2 \neq 0$$

$$\hat{p}_1 = \frac{169}{225} = 0.751$$
$$\hat{p}_2 = \frac{160}{225} = 0.711$$
$$\hat{p} = \frac{169+160}{225+225} = 0.731$$

$$z = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.751 - 0.711}{\sqrt{0.731(1-0.731)(\frac{1}{225} + \frac{1}{225})}} = 0.957$$

See the diagram below, reject H0 if Z test > 1.96 or Z test < -1.96
If you use critical values, 0.957 is not in the rejection regions




∴ *At the 5% level of significance, there is not sufficient evidence to suggest that there is a difference between English and French speaking Catholics who agree divorces should be able to remarry.*

p-value=2(0.1685)>$\alpha = 0.05$, so we fail to reject $H_0$

b) Set up a 95% confidence interval estimate of the difference between the population proportions of English-speaking Catholics and French-speaking Catholics that agree that divorcees should be able to remarry in the Church.

$$\hat{p}_1 = \frac{169}{225} = 0.751$$

$$\hat{p}_2 = \frac{160}{225} = 0.711$$

$$p1 - p2 = (\widehat{p_1} - \widehat{p_2}) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

$$p1 - p2 = (0.751 - 0.711) \pm 1.96\sqrt{\frac{0.751(0.249)}{225} + \frac{0.711(0.289)}{225}}$$

$$p1 - p2 = 0.04 \pm 0.084 = (-0.044, 0.124)$$

**0 is in this interval, so we fail to reject $H_0$** and there is no evidence of a difference

**<u>Example 2.</u>**

$H_0$ $p_1$-$p_2$=0

$H_a$ $p_1$-$p_2$<0 (one sided p-value)

$$\widehat{p_1} = \frac{56}{2051} = 0.0273$$

$$\widehat{p_2} = \frac{84}{2030} = 0.0414$$

a) standard error= $\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$= $\sqrt{\frac{0.0273(0.9727)}{2051} + \frac{0.0414(0.9586)}{2030}} = 0.0057$

b) 90% confidence interval

$$p1 - p2 = (\widehat{p_1} - \widehat{p_2}) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

=$(0.0273 - 0.0414) \pm 1.645\sqrt{\frac{0.0273(0.9727)}{2051} + \frac{0.0414(0.9586)}{2030}}$ or substitute part a) for the whole square root and get $0.0141 \pm 1.645(0.0057)$

=$-0.0141 \pm 0.00938$

=(-0.0235, -0.00472)

c) Since 0 is not in the interval from part b), we reject $H_0$ and conclude there is evidence that Gemfibrozil lowers the risk of heart attack.

Therefore, the test is statistically significant.

C1. $H_0$ $p_1$-$p_2$=0
$H_a$ $p_1$-$p_2$>0 (one sided p-value) (greater than test, shade above z test)
This means p1>p2 ie. improved proportion for the gastric group

$\widehat{p_1} = \frac{25}{80} = 0.3125$ gastric

$\widehat{p_2} = \frac{30}{79} = 0.3797$ placebo

$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{25+30}{80+79} = 0.3459$

$\hat{q} = 1 - 0.3625 = 0.6541$

$$z = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.3125 - 0.3797}{\sqrt{0.3459(0.6541)(\frac{1}{80} + \frac{1}{79})}} = -0.89$$

p-value=Pr(Z> -0.89)=1-0.1867 > 0.10 so there is no evidence in favour of $H_a$
∴fail to reject $H_0$ and conclude there is no difference with this method.

C2.
$H_0$ $p_1$-$p_2$=0
$H_a$ $p_1$-$p_2$>0 (one sided p-value) higher recall for group 1

2-sided p-value

$$\hat{p}1 = \frac{31}{40} = 0.775$$

$$\hat{p}2 = \frac{22}{40} = 0.55$$

$$\hat{p} = \frac{31+22}{80} = 0.6625$$

$$z = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.775 - 0.55}{\sqrt{0.6625(0.3375)(\frac{1}{40} + \frac{1}{40})}} = 2.13$$

Using critical values, Z critical = 1.645(area 0.05 on the right, or area below the line of 0.95)
And Z test=2.13 is in the rejection region above 1.645, so reject H0.

Or use p-value=1-0.9834=0.0166<0.05=$\alpha$
∴reject $H_0$

∴At the 5% level of significance, there is sufficient evidence to suggest the younger group has a higher recall rate.

C3.

$\widehat{p_1} = \frac{50}{2000} = 0.025$

$\widehat{p_2} = \frac{90}{2050} = 0.0439$

$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{50+90}{2000+2050} = 0.03457$

$\hat{q} = 1 - 0.03457 = 0.96543$

At the 5% level of significance, is there sufficient evidence of a difference in the incidence rates of cardiac events for the two groups? Give supporting details.

$H_o : p_1 - p_2 = 0$

$H_a : p_1 - p_2 \neq 0$

$$z = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.025 - 0.0439}{\sqrt{0.03457(0.96543)(\frac{1}{2000} + \frac{1}{2050})}} = -3.29$$

p-value=2P(Z<-3.29)=2(0.0005)=0.001< alpha=0.05

So, we reject the null hypothesis and conclude there is evidence of a difference

C4. Use a significance level of 5%.

$H_o : p_1 - p_2 \leq 0$

$H_a : p_1 - p_2 > 0$ one sided p-value p1>p2 or p2<p1(decreases)

$$\widehat{p_1} = \frac{575}{1500} = 0.3833$$

$$\widehat{p_2} = \frac{579}{1700} = 0.3406$$

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{575+579}{1500+1700} = 0.3606$$

$$\hat{q} = 1 - 0.3606 = 0.6394$$

$$z = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.3833 - 0.3406}{\sqrt{0.3606(0.6394)(\frac{1}{1500} + \frac{1}{1700})}} = 2.51$$

p-value=Pr(Z>2.51)= 1 – 0.994=0.006<0.05 so we reject $H_0$

There is sufficient evidence that the proportion of smokers has decreased.

C5.

a) Is there evidence of a significant difference between males and females in the proportion who enjoy shopping for clothing at the 1% level of significance?

$H_0$ $p_1 - p_2 = 0$

$H_a$ $p_1 - p_2 \neq 0$ (two sided p-value) alpha=0.01

$$\widehat{p_1} = \frac{130}{250} = 0.52$$

$$\widehat{p_2} = \frac{225}{250} = 0.9$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{130+225}{250+250} = 0.71$$

$$\hat{q} = 1 - 0.71 = 0.29$$

$$Z\ test = \frac{(\widehat{p_1} - \widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.52 - 0.9}{\sqrt{0.71(0.29)(\frac{1}{250} + \frac{1}{250})}} = -9.36$$

b) Find the *p*-value in (a) and interpret its meaning.

*P-value= 2P( z<-9.36) The area below -3.49 is nearly 0*
*= 2(0 )= 0*

*p-value<0.01 so we can reject $H_0$*

*∴ There is very strong evidence to suggest that there is a difference*

*∴ There is sufficient evidence to suggest that there is a difference between the proportion of males and females who enjoy shopping online*

c) Set up a 99% confidence interval estimate of the difference between the proportions of males and females who enjoy shopping online

$$= (\widehat{p_1} - \widehat{p_2}) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

$$=(0.52 - 0.90) \pm 2.576\sqrt{\frac{0.52(0.48)}{250} + \frac{0.9(0.1)}{250}}$$

$$=-0.38 \pm 0.0949$$

=(-0.4749, -0.285)

0 is not in the interval so we can reject $H_0$ and conclude there is statistically significant evidence of a difference.

C6. a) $H_0$ $p_1$-$p_2$=0
$H_a$ $p_1$-$p_2$≠0 (two sided p-value)

$$\widehat{p_1} = \frac{65}{100} = 0.65$$

$$\widehat{p_2} = \frac{82}{100} = 0.82$$

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{65+82}{100+100} = 0.735$$

$$\hat{q} = 1 - 0.735 = 0.265$$

$$= (\widehat{p_1} - \widehat{p_2}) \pm Z_{\alpha/2}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

$$=(0.65 - 0.82) \pm 1.96\sqrt{\frac{0.65(0.35)}{100} + \frac{0.82(0.18)}{100}}$$

$$= -0.17 \pm 0.12$$

b) $z = \frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

$$z = \frac{0.65 - 0.82}{\sqrt{0.735(0.265)(\frac{1}{100} + \frac{1}{100})}} = -2.72$$

p=2Pr(Z<-2.72)=2(0.0033)=0.0066, so D).

c) p-value=0.0066< 1% so at 1% we reject H0
Also, the p-value is less than 5%, 2% and 10%, so the answer is E).

C7.

a)

$H_0$ $p_1$-$p_2$=0
$H_a$ $p_1$-$p_2$≠0 (two sided p-value)

b)

$$\widehat{p_1} = \frac{12}{100} = 0.12$$

$$\widehat{p_2} = \frac{17}{100} = 0.17$$

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{12+17}{200} = 0.145$$

$$\hat{q} = 1 - 0.145 = 0.855$$

$$z = \frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$$

$$z = \frac{0.12 - 0.17}{\sqrt{0.145(0.855)(\frac{1}{100}+\frac{1}{100})}} = -1.00$$

c) p-value=2P(Z<-1)=2(0.1587)=0.3174

d) ∴p-value> 0.10 (the largest alpha value we use) and there is no evidence that $H_a$ is true and we fail to reject $H_0$ and we conclude there is no difference in the infestation rates.

## D. Comparing Two Population Means by Using Paired Differences

**Example 1**.

| Students | Nov. grades | February grades | difference |
|---|---|---|---|
| 1 | 68 | 72 | 4 |
| 2 | 79 | 76 | -3 |
| 3 | 54 | 70 | 16 |
| 4 | 73 | 73 | 0 |
| 5 | 59 | 69 | 10 |
| 6 | 88 | 94 | 6 |
| 7 | 44 | 42 | -2 |
| 8 | 70 | 82 | 12 |
| 9 | 67 | 65 | -2 |

$\overline{d} = \frac{41}{9} = 4.56 \qquad S_d = 6.91$

$H_0 \;\; \mu_d = 0 \qquad H_a \;\; \mu_d > 0 \;\; (1 \; sided)$

$t_{crit} = t_{nD-1,\alpha} = t_{8,0.01} = 2.896$

DR reject $H_0 \; if \; t_{test} > 2.896$

$t_{test} = \frac{\bar{d}-\mu_D}{s_d/\sqrt{n}} = \frac{4.56-0}{6.91/\sqrt{9}} = 1.98$

$\therefore t_{test} < 2.896 \quad \therefore do \; not \; reject \; H_0 \quad \therefore no \; statistical \; evidence$

Or use p-value and look up 1.98 on the t-table with n-1=9-1=8 df

1.895<1.98<2.365

0.025< 1 sided p-value<0.05, so p-value greater than 1%, so we do not reject $H_0$

b) How would your answer differ if a 10% level of significance was used?

$t_{crit} = t_{8,0.10} = 1.397$ or use p-value less than 10%, so reject H0

$t_{t\;test} = 1.98 > 1.397 \quad \therefore reject \; H_0$

$\therefore at \; a \; 10\% \; level, there$

$is \; statistical \; evidence \; that \; the \; mean \; grades \; increased$

c) Form a 98% confidence interval to estimate the true mean difference between the November midterm and February midterm scores.

$$\alpha = \frac{100-98}{2} = 1\% \quad t_{crit} = t_{8,0.01} = 2.896$$

$\mu_d = \overline{d} \pm t_{crit}\left(\frac{s_d}{\sqrt{n}}\right) = 4.56 \pm 2.896\left(\frac{6.91}{\sqrt{9}}\right) = 4.56 \pm 6.67 = (-2.11, 11.23)$

$\mu_D$is from -2.11 to 11.23. the 98% CI is (-2.11, 11.23)

**Example 2.**

| Vehicle | Traditional Tire (1000s of km) | New Tire (1000s of km) | Difference |
|---|---|---|---|
| 1 | 85 | 88 | 3 |
| 2 | 75 | 76 | 1 |
| 3 | 68 | 64 | -4 |
| 4 | 59 | 62 | 3 |
| 5 | 100 | 105 | 5 |
| 6 | 78 | 81 | 3 |
| 7 | 52 | 49 | -3 |
| 8 | 84 | 82 | -2 |
| 9 | 79 | 82 | 3 |
| 10 | 96 | 93 | -3 |

Does the additive significantly increase cars' mileage per gallon (α ≤ .05)?

Comparing same vehicles $\therefore paired\ data$ $H_0\ \ \mu_D = 0$

$H_a\ \ \mu_D > 0$

$\bar{x}_D = 0.6$

$s_D = 3.27$

$n_D = 10$

$t_{crit} = t_{nD-1,\alpha} = t_{9,0.05} = 1.833$

DR reject $H_0\ if\ t_{test} > 1.833$

$t_{test} = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}} = \frac{0.6-0}{3.27/\sqrt{10}} = 0.58$

$\therefore t_{test} < 1.833 \quad \therefore do\ not\ reject\ H_0$

$\therefore no\ statistical\ difference$ and the new tires do not increase life of the tires.

Or use p-value

At df=9, 0.58<1.383

p-value>0.10> alpha=0.05

So, no statistical evidence the new tech tires increase tire life

D1. Alpha/2=0.05/2=0.025

$t_{crit} = t_{nD-1,\alpha} = t_{7,0.025}$ =2.365

$$n = 8$$
$$df = n - 1 = 7$$
$$\sum x = 14$$
$$\sum x^2 = 106$$
$$\bar{d} = \frac{\sum x}{n} = \frac{6 + 5 + (-2) + ... + (-4)}{8} = \frac{14}{8} = 1.75$$
$$s_D = \sqrt{\frac{(6-1.75)^2 + (5-1.75)^2 + ... + (-4-1.75)^2}{8-1}} = 3.41$$

c) Compute a 95% confidence interval estimate for the mean difference.

$$\mu_d = \bar{d} \pm t_{crit}\left(\frac{s_d}{\sqrt{n}}\right) = 1.75 \pm 2.365\left(\frac{3.41}{\sqrt{8}}\right) = 1.75 \pm 2.85 = (-1.1, 4.6)$$

Interpret the interval.
Interval includes 0 (ie. there is no difference). Therefore, there is no evidence that supports the claim that weight loss amongst quitters is different.

d) $H_0$ $\mu_D = 0$
$H_a$ $\mu_D \neq 0$ ( 2-sided)

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{1.75 - 0}{\frac{3.41}{\sqrt{8}}} = 1.45$$

we want to see if there is a difference in weight, so we use a two sided test
df=7...go across at 7df and find t=1.45

1.415< 1.45< 1.895
0.05(2)< pvalue<0.1(2)
0.10< two sided p<0.20

So, p-value > alpha=0.01, so we fail to reject $H_0$ and conclude there is no significant difference in weight

D2.a) State the appropriate hypothesis for a test aimed at determining if the drug lessens anxiety.

*Concept: Two sample mean inference (matched pair)*

If $\mu_1 - \mu_2 = \mu_D > 0$ , then anxiety is lessened.
$H_0 = \mu_D = 0$
$H_a = \mu_D > 0$

b) Calculate the observed test statistic value and give the *p*-value for your test in part (a).

$$n = 9$$
$$df = n - 1 = 8$$
$$\sum x = 17$$
$$\sum x^2 = 185$$
$$\bar{d} = \frac{\sum x}{n} = \frac{3 + 6 + 4 + ... + 7}{9} = 1.89$$
$$s_d = \sqrt{\frac{(3 - 1.89)^2 + ... + (7 - 1.89)^2}{9 - 1}} = 4.37$$
$$t = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{1.89}{4.37 / \sqrt{9}} = 1.29$$

at 8df, t test=1.29 < 1.397, so the one sided p-value > 0.10

c) If you were to use a significance level of 0.10, what would your decision and conclusion be?

p-value > 0.10
Therefore, we cannot reject the Null Hypothesis. There is no evidence supporting the claim that the tranquilizer reduces anxiety.

D3. A) We have paired data here because we have two observations from the same child

b) after-before=5,4,0,8,-4,13,2,4,9,2

mean difference=4.3

So, $\overline{d} = 4.3$
The standard deviation is: $S_d$=4.02
$H_0$: $\mu_d \leq 0$
$H_a$: $\mu_d > 0$

c) $t = \frac{4.3-0}{4.02/\sqrt{10}}$=3.38

df=10-1=9

3.25< t test=3.38< 4.297
0.001<one sided p-value<0.005

d) p-value< alpha=0.01, so reject Ho at 1%
p-value < alpha=0.05, so we reject H0 at the 5% level

D4. n=10, $\overline{d} = 0.09, s_d = 0.7, df = 10 - 1 = 9$
$t_{\alpha/2} = 9df, 95\% = 2.262$

$$\overline{d} \pm t_{\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right)$$
$$= 0.09 \pm 2.262\left(\frac{0.7}{\sqrt{10}}\right)$$
=0.09$\pm 0.50$
=(-0.41, 0.59)

D5.

| Subject | Month 3 Reduction (%) | Month 6 Reduction (%) | Difference |
|---|---|---|---|
| 1 | 5 | 9 | 4 |
| 2 | 8 | 10 | 2 |
| 3 | 12 | 11 | -1 |
| 4 | 10 | 14 | 4 |
| 5 | 7 | 15 | 8 |
| 6 | 9 | 16 | 7 |
| 7 | 11 | 13 | 2 |
| 8 | 6 | 6 | 0 |
| 9 | 1 | 7 | 6 |

Perform the appropriate analysis with $\alpha \leq .05$.

$H_0 \ \ \mu_D = 0$

$H_a \ \ \mu_D > 0$

$\bar{x}_D = 3.56$

$S_D = 3.09$ $\qquad t_{crit} = t_{nD-1,\alpha} = t_{8,0.05} = 1.86$

$n_D = 9$

DR reject $H_0 \ if \ t_{test} > 1.86$

$t_{test} = \frac{3.56-0}{3.09/\sqrt{9}} = 3.46$

$\therefore t_{test} = 3.46 > 1.86 \quad \therefore reject \ H_0$

$\therefore there \ is \ significant \ evidence \ the \ difference \ is \ greater \ than \ 0$

D6.

| Participant | New BB | Old BB | Difference |
|---|---|---|---|
| 1 | 60 | 59 | -1 |
| 2 | 67 | 64 | -3 |
| 3 | 70 | 66 | -4 |
| 4 | 60 | 58 | -2 |
| 5 | 65 | 62 | -3 |
| 6 | 72 | 69 | -3 |
| 7 | 68 | 70 | 2 |

$\bar{d} = -2$ sD=2

Is the manufacturer's claim justified ($\alpha \leq 0.05$)?

$H_0 \ \ \mu_d = 0 \qquad H_a \ \ \mu_d < 0 \quad (1 \ sided)$

if you do old – new<0, it is the same as new-old>0

**<u>Critical Method</u>**

$t_{crit} = t_{n-1,\alpha} = t_{6,0.05} = 1.943$ (look up df=6 and 0.05 along the top)

DR reject $H_0 \ if \ t_{test} < -1.943$ (since it is a less than test)

$t_{test} = \frac{-2-0}{2/\sqrt{7}} = -2.65$

$\therefore t_{test} < -1.943 \quad \therefore \ reject \ H_0 \ \ and \ the \ new \ arrow \ does \ go \ farther$

**P-Value Method**

Go across from 6df and find t test=2.65 and since it is a less than test we use -2.65

-2.447< t test= - 2.65< 3.143

0.01<1 sided p-value<0.025

p-value <0.05=alpha, so we reject H0

D7.

$difference\ \ 5, 2, -1, 3, 0, 4, 1, 3, -2, 1$

$$\bar{d} = \frac{5 + 2 + \cdots + 1}{10} = 1.6$$

$H_0 \quad \mu_d = 0$

$H_a \quad \mu_d > 0 \quad (1\ sided)$ (during – before >0)

Sd=2.22

$n = 10 \quad \alpha = 0.05$

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{1.6 - 0}{\frac{2.22}{\sqrt{10}}} = 2.28$$

$df = 9 \quad go\ across\ 9\ df$

$2.262 < 2.28 < 2.821$

$0.01 < 1\ sided\ p - value < 0.025$

$\therefore p - value\ < \alpha = 0.05 \quad \therefore reject\ H_0$

$\therefore yes, there\ is\ statistically\ significant\ evidence\ that$
$blood\ pressure\ rises$

or use critical method and look up 9df and 0.05 and get t critical = 1.833 and the rejection region for a greater than test would be above 1.833 and then t test = 2.28 is greater than 1.833 and we would reject H0

D8.

**1.** $df = 8 - 1 = 7 \quad the\ answer\ is\ D.$

**2.** $\bar{d} = \frac{140+80+90+10+100+50-50-20}{8} = \frac{400}{8} = 50$

$$s_d = \sqrt{\frac{(140-50)^2 + (80-50)^2 + \cdots + (-20-50)^2}{7}} = 65.03$$

The answer is B.

**3.** $\bar{d} \pm t_{\alpha/2}(\frac{s_d}{\sqrt{n}})$

$standard\ error = \frac{s_d}{\sqrt{n}} = \frac{65.03}{\sqrt{8}} = 22.99$

The answer is C.

**4.** $df = 7 \quad \alpha = 0.05 \qquad t_{\alpha/2} = 1.895$

$H_0 \quad \mu_d = 0$

$H_a \quad \mu_d > 0 \;\; (1\; sided)$

↑

Look up top 0.05 along and 7 df

The answer is D.

**5.** $t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{50 - 0}{\frac{65.03}{\sqrt{8}}} = 2.175$

$go\ across\ 7\ df$

$1.895 < 2.175 < 2.365$

$0.025 < 1\ sided\ p - value < 0.05$

$\therefore p - value\ < \alpha = 0.05$

$\therefore reject\ H_0$

$The\ answeris\ A$

**6.** $2.175$

**7.** $p - value\ is\ between\ 0.025\ and\ 0.05$

**8.** The answer is A.

## E. Comparing Two Population Variances

**Example 1.**

a) $\alpha = 0.05 \quad H_0 \quad {\sigma_1}^2 = {\sigma_2}^2 \quad H_a \quad {\sigma_1}^2 < {\sigma_2}^2$

${s_1}^2 = 58$
${s_2}^2 = 101$

F test= $\frac{{s_1}^2}{{s_2}^2} = \frac{101}{58} = 1.74$

$use\ 199\ df, 199\ df, 0.05 \quad F\ crit = 1.35$ (use 120,120)

$F\ test > F\ crit \quad \therefore reject\ H_0$

So, there is evidence the first instructor's variance is smaller

b) $\alpha = 0.10 \quad H_0 \quad {\sigma_1}^2 = {\sigma_2}^2 \quad H_a \quad {\sigma_1}^2 \neq {\sigma_2}^2$

$same\ as\ a)$ except alpha divided by 2 is 0.10/2 = 0.05, so same table being used

$use\ 199\ df, 199\ df, 0.05 \quad F\ crit = 1.35$ (use 120,120)

$F\ test > F\ crit \quad \therefore reject\ H_0$

So, there is evidence of a difference in variances

E1.

$\alpha = 0.05 \quad H_0 \quad {\sigma_1}^2 = {\sigma_2}^2 \quad H_a \quad {\sigma_1}^2 > {\sigma_2}^2$

${s_1}^2 = 12.96^2 = 167.9616$
${s_2}^2 = 11.85^2 = 140.4225$

F test= $\frac{{s_1}^2}{{s_2}^2} = \frac{167.9616}{140.4225} = 1.196$

$use\ 120\ df, 120\ df \quad F\ crit = 1.3519$

Reject H0 if F test >1.3519

$F\ test < F\ crit \quad \therefore fail\ to\ reject\ H_0$

<u>P-value method: df (120,120)</u>

0.10 F critical = 1.26
0.05 F critical = 1.3519
0.025 F critical= 1.4327

F test = 1.196<1.26, so p-value > 0.10 and we fail to reject H0 at the 5% level of significance.

E2.

$Let\ \alpha = 0.05 \quad (2 - sided) \quad \frac{\alpha}{2} = 0.025$

$n_1 = n_2 = 10$

$df = 9{,}9$ with alpha/2=0.025..use F table in back of booklet

$F\ crit = 4.026$

Reject H0 if F test >4.026

F test = $\frac{{s_2}^2}{{s_1}^2} = \frac{1.328^2}{0.882^2} = 2.27$

$F\ test = 2.27 < F\ crit \quad \therefore fail\ to\ reject\ H_0$

$\therefore no\ statistically\ significant\ evidence\ of\ a\ difference\ in\ variance$

E3. $\alpha = 0.10 \quad H_0 \quad {\sigma_1}^2 = {\sigma_2}^2 \quad H_a \quad {\sigma_1}^2 \neq {\sigma_2}^2$(two sided test)

${s_1}^2 = 109 \quad n_1 = 40$

${s_2}^2 = 65 \quad n_2 = 20$

$df\ num = 39$

$df\ den = 19$ (use 30, 19)

$F\ test = \frac{{s_1}^2}{{s_2}^2} = \frac{109}{65} = 1.68$

$F\ crit = 2.07$ (use 0.10/2=0.05)

$F\ test = 1.68 < F\ crit \quad \therefore fail\ to\ reject\ H_0$

<u>P-value method</u>

0.10 F critical = 1.76
0.05 F critical=2.07
0.025 F critical=2.3937

F test = 1.68 < 1.76, so the p-value>0.10=alpha

E4. $F\ test = \frac{{s_1}^2}{{s_2}^2} = \frac{210.2}{36.5} = 5.76 \quad one - tail$

$H_0 \quad {\sigma_1}^2 = {\sigma_2}^2$

$H_a \quad {\sigma_1}^2 > {\sigma_2}^2$

$\alpha = 0.05 \quad F\ crit = 3.0729$

Reject H0 if F test >3.0729

$df\ num = 12 \quad df\ den = 9$

$\therefore F\ test > F\ crit \quad \therefore reject\ H_0$

$\therefore yes, evidence\ of\ more\ variability$

<u>P-value method</u>

0.05 F critical=3.0729
0.025 F critical=3.8682

0.01 F critical = 5.11

F test = 5.76>5.11, so the p-value<0.01 and the p-value <0.05=alpha
So, we reject H0

E5. $F\ test = \frac{{s_1}^2}{{s_2}^2} = \frac{87.5}{53.4} = 1.64$ $H_0 \quad {\sigma_1}^2 = {\sigma_2}^2$

$H_a \quad {\sigma_1}^2 < {\sigma_2}^2$

$\alpha = 0.05$
$df = 29\ num\ (use\ 24) \quad > \quad Fcrit = 1.90$
$df = 29\ den$

$F\ test < F\ crit \quad \therefore do\ not\ reject\ H_0$
$\therefore no\ statistically\ significant\ evidence\ that\ 1st\ instructor's\ variance\ is\ smaller.$

Using p-values:

p-value 0.10 F critical=1.65
0.05 F critical=1.90
F test < 1.65, so the p-value >0.10

## F. Analysis of Variance (ANOVA)

**Example 1**

$H_0 \quad \mu1 = \mu2 = \mu3$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$N = 9 \quad C = 3\ groups \quad \bar{x} = 529.2$

$SSC = \sum ni(\bar{xi} - \bar{x})^2$

$= 3(666.67 - 529.2)^2 + 3(473.67 - 529.2)^2 + 3(447.33 - 529.2)^2 = 86\ 052.84$

$SSE = \sum(ni - 1)si^2 = 2(31.18)^2 + 2(49.17)^2 + 2(41.68)^2$

$= 10\ 254.2$

$MSC = \frac{SSC}{C-1} = \frac{86\ 052.84}{2} = 43\ 026.42$

$MSE = \frac{SSE}{N-C} = \frac{10\ 254.2}{6} = 1709.03$

$F_{test} = \frac{MSC}{MSE} = \frac{43\ 026.42}{1709.03} = 25.2$

$F_{crit} = F_{C-1,N-C,\alpha} = F_{2,6,0.05} = 5.14$

$\therefore F_{test} > 5.14 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ differ\ significantly$

Or use p-values and look up various F values
0.10…gives F critical 3.46
0.05 gives F critical of 5.14
0.025 gives F critical of 7.26
0.01 gives F critical of 10.92

Our F test=25.2 > 10.92, so our p-values< 0.01

$b)\ H_0 \quad \mu i = \mu j$ Tukey test

$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{C,N-C,\alpha} = Q_{3,6,0.05} = 4.34$

HSD=Qcrit$\sqrt{\frac{MSE}{n}} = 4.34\sqrt{\frac{1709.03}{3}} = (4.34)23.87 = 103.59$

1 & 2 $\quad |666.67 - 473.67| = 193 > HSD \quad \therefore\ reject\ H_0$

$\therefore$ 1 & 2 $means\ are\ statistically\ different$

1 & 3 $\quad |666.67 - 447.33| = 219.34 > HSD \quad \therefore\ reject\ H_0$

$\therefore$ 1 & 3 $means\ are\ statistically\ different$

2 &3 $\quad |473.67 - 447.33| = 26.34 < HSD \quad \therefore\ fail\ to\ reject\ H_0$

$\therefore$ 2 & 3 $means\ are\ statistically\ equal$

**Example 2.** Total degrees of freedom N-1=29 df, so N=30

N-C=30-27=3

$MSC = \frac{SSC}{C-1}$

$2.521 = \frac{SSC}{2}$

SSC=5.043 or SSC=SST - SSE

$MSE = \frac{SSE}{N-C} = \frac{11.799}{27} = 0.437$

$F_{test} = \frac{MSC}{MSE} = \frac{2.521}{0.437} = 5.8$

**Example 3.**

$MSC = 0.228 \qquad MSE = 0.292$

$SSC = 0.683 \qquad SSE = 11.961$

$MSC = \frac{SSC}{C-1} \qquad \therefore 0.228 = \frac{0.683}{C-1}$

$0.228C - 0.228 = 0.683$

$0.228C = 0.911$

$C = 3.99956 \quad \therefore C = 4$

$MSE = \frac{SSE}{N-C}$

$0.292 = \frac{11.961}{N-4}$

$0.292N - 1.168 = 11.961$

$N = 44.96 \quad \therefore N = 45$

**Example 4.**

**a)**

$H_0 \quad \mu 1 = \mu 2 = \mu 3 = \mu 4$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$\alpha = 0.05$

$F_{crit} = F_{c-1,N-C,\alpha} = F_{3,26,0.05}$

DR reject $H_0\ if\ F_{test} > F_{crit} = 2.98$

$\bar{x} = 39.6$

SSC=$\sum ni(\bar{xi} - \bar{x})^2 = 6(60.33 - 39.6)^2 + 8(41.63 - 39.6)^2 + \cdots + 8(26.25 - 39.6)^2 = 4179.61$

SSE=$\sum (n_i - 1)si^2 = 5(17.851)^2 + 7(12.212)^2 + 7(13.384)^2 + 7(8.155)^2 = 4356.67$

MSC= $\frac{SSC}{c-1} = \frac{4179.61}{3} = 1393.20$

MSE= $\frac{SSE}{N-C} = \frac{4356.67}{26} = 167.56$

$F_{test} = \frac{MSC}{MSE} = \frac{1393.20}{167.56} = 8.3$

$\therefore F_{test} > F_{crit} = 2.98 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ differ\ significantly$

Or do p-values

0.01…..F critical 4.64 with 3, 26 df
0.025…F critical is 3.6097
0.05….F critical is 2.9752
0.10…F critical 2.31

F test = 8.3 is greater than 4.64, so our p-value is less than 0.01

**b) NOTE: I wrote out differences between all means for you!!**

$Q_{crit} = Q_{c,N-C,\alpha} = Q_{4,26,0.05} \quad (24\ df)$

$= 3.90$

$H_0\ \ \mu i = \mu j$

$H_a\ \ \mu i \neq \mu j$

$HSD = Q_{crit}\sqrt{\frac{MSE}{2}(\frac{1}{ni} + \frac{1}{nj})}$

1 & 2 $\quad HSD = 3.90\sqrt{\frac{167.56}{2}(\frac{1}{6} + \frac{1}{8})} = 19.28$

$|60.33 - 41.63| = 18.7 < 19.28 \quad \therefore do\ not\ reject\ H_0$

$\therefore\ \ 1\ \&\ 2\ means\ equal$

1 & 3 $\quad |35.38 - 60.33| = 24.95 > 19.28 \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

1 & 4 $\quad |6.25 - 60.33| = 34.08 > 19.28 \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

$HSD = 3.90\sqrt{\frac{167.56}{2}(\frac{1}{8} + \frac{1}{8})} = 17.85$

3 & 4 $\quad |35.38 - 26.25| = 9.13 < HSD \quad \therefore means\ 3\ \&\ 4\ statistically\ equal$

2 & 4 $\quad HSD = 17.85$

$|\bar{x}_2 - \overline{x_4}| = |41.63 - 26.25| = 15.38 < 17.85$

$\therefore do\ not\ reject\ H_0 \quad \therefore\ 2\ \&\ 4\ means\ are\ equal$

2 & 3 $\quad |41.63 - 35.38| = 6.25 < HSD$

$\therefore means\ 2\ \&\ 3\ are\ statistically\ equal$

$\therefore groups\ 1\ \&\ 3, 1\ \&\ 4\ are\ not\ statistically\ equal$

**F1.**

$\overline{x_1} = 95.3 \quad \overline{x_2} = 84.8 \quad \overline{x_3} = 75.3 \quad \overline{x_4} = 81.8$

$c = 4 \quad N = 24 \quad \bar{x} = 84.3$

$s_1 = 4.8 \quad s_2 = 4.07 \quad s_3 = 4.18 \quad s_4 = 3.82$

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$F_{crit} = F_{c-1,N-c,\alpha} = F_{3,20,0.05} = 3.10$

DR reject $H_0\ if\ F_{test} > 3.10$

SSC= $\sum ni(\bar{xi} - \bar{x})^2 = 6(95.3 - 84.3)^2 + 6(84.8 - 84.3)^2 +$
$6(75.3 - 84.3)^2 + 6(81.8 - 84.3)^2 = 1251$

SSE= $\sum(ni - 1)si^2 = 5(4.8)^2 + 5(4.07)^2 + 5(4.18)^2 + 5(3.82)^2$
$= 358.3$

MSC= $\frac{SSC}{c-1} = \frac{1251}{3} = 417$

$MSE = \frac{SSE}{N-c} = \frac{358.3}{20} = 17.915$

$F_{test} = \frac{MSC}{MSE} = \frac{417}{17.915} = 23.3 > 3.10 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ are\ statistically\ different$

**b)** $Q_{crit} = Q_{c,N-c,\alpha} = Q_{4,20,0.05} = 3.96$

$n = 6\ in\ each\ group$

HSD= $Q_{crit}\sqrt{\frac{MSE}{n}} = 3.96\sqrt{\frac{17.915}{6}} = 6.84$

1 & 2 $|95.3 - 84.8| = 10.49 > HSD \quad reject\ H_0$

1 & 3 $|95.3 - 75.3| = 20 > HSD \quad reject\ H_0$

1 & 4 $|95.3 - 81.8| = 13.5 > HSD \quad reject\ H_0$

2 & 3 $|84.8 - 75.3| = 9.5 > HSD \quad reject\ H_0$

3 & 4 $|75.3 - 81.8| = 6.5 < HSD \quad do\ not\ reject\ H_0$

2 & 4 $|84.8 - 81.8| = 3 < HSD \quad do\ not\ reject\ H_0$

$\therefore$ 1&2, 1&3, 1&4, 2&3 $means\ differ\ significantly$

**F2.** SSC= 6536 SSE= 12 407

Wendy's $\overline{x_1} = 150$

McD $\overline{x_2} = 167$

Check $\overline{x_3} = 169$

B.K. $\overline{x_4} = 171$

L.J.S. $\overline{x_5} = 172$

$N = 20 \times 5 = 100 \qquad c = 5\ groups$

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4 = \mu5$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$MSC = \frac{SSC}{c-1} = \frac{6536}{4} = 1634$

$MSE = \frac{SSE}{N-c} = \frac{12\ 407}{95} = 130.6$

$F_{test} = \frac{MSC}{MSE} = \frac{1634}{130.6} = 12.51$

$F_{crit} = F_{c-1,N-c,\alpha} = F_{4,95,0.05} = 2.53$ (use 60)

DR reject $H_0\ if\ F_{test} > 2.53$

$\therefore F_{test} = 12.51 > 2.53 \quad reject\ H_0$

**b)** $H_0 \quad \mu i = \mu j$

$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{c,N-c,\alpha} = Q_{5,95,0.05} = 3.98 \quad (use\ 60)$

$HSD = Q_{crit}\sqrt{\frac{MSE}{n}} \qquad n = 20\ in\ each\ group$

$= 3.98\sqrt{\frac{130.6}{20}} = 10.17$

1&2 $|150 - 167| = 17 > HSD \quad reject\ H_0$

1&3 $|150 - 169| = 19 > HSD \quad reject\ H_0$

1&4 $|150 - 171| = 21 > HSD \quad reject\ H_0$

1&5 $|150 - 172| = 22 > HSD \quad reject\ H_0$

2&3 $|167 - 169| = 2 < HSD \quad do\ not\ reject\ H_0$

2&4 $|167 - 171| = 4 < HSD \quad do\ not\ reject\ H_0$

2&5 $|167 - 172| = 5 < HSD \quad do\ not\ reject\ H_0$

3&4 $|169 - 171| = 2 < HSD \quad do\ not\ reject\ H_0$

3&5 $|169 - 172| = 3 < HSD \quad do\ not\ reject\ H_0$

4&5 $|171 - 172| = 1 < HSD \quad do\ not\ reject\ H_0$

$\therefore groups\ 1\&2, 1\&3, 1\&4, 1\&5\ are\ all\ statistically\ different$

**F3. a)**

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4$

$H_a$ $at\ least\ 2\ means\ differ\ significantly$

$C = 4\ groups \quad N = 24$

$SSC = 5146.34 - 919.58 =$4226.76

$MSE = \frac{SSE}{N-C} = \frac{919.58}{24-4} = 45.98$

MSC=$\frac{SSC}{c-1} = \frac{4226.76}{3} = 1408.92$

$F_{test} = \frac{MSC}{MSE} = \frac{1408.92}{45.98} = 30.64$

$F_{crit} = F_{c-1,N-C,\alpha} = F_{3,20,0.05} = 3.10$

**b)**

$\therefore F_{test} > 3.10 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ differ\ significantly$

**c)** $\overline{x_1}$=47.2 $\quad \overline{x_2} = 15.7 \quad \overline{x_3} = 31.5 \quad \overline{x_4} = 14.8$

HSD= $Q_{crit}\sqrt{\frac{MSE}{n}} \qquad n = \#\ in\ each\ group$

$Q_{crit} = Q_{c,N-C,\alpha} = Q_{4,24-4,0.05} = Q_{4,20,0.05} = 3.96$

$\therefore HSD = 3.96\sqrt{\frac{45.96}{6}} = 10.96$

1 & 2 $\quad |15.7 - 47.2| = 31.5 > HSD \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

2 & 3 $\quad |31.5 - 15.7| = 15.8 > HSD \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

3 & 4 $\quad |14.8 - 31.5| = 16.7 > HSD \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

1 & 3 $\quad |31.5 - 47.2| = 15.7 > HSD \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

1 & 4 $\quad |14.8 - 47.2| = 32.4 > HSD \quad \therefore reject\ H_0$

$\therefore means\ not\ equal$

2 & 4 $\quad |14.8 - 15.7| = 0.9 < HSD \quad \therefore do\ not\ reject\ H_0$

$\therefore means\ are\ statistically\ equal$

$\therefore ONLY\ 2\ \&\ 4\ are\ equal$

**F4.**

$$N = 15 \quad c = 3\ groups \quad \bar{x} = \frac{5(320)+4(298)+6(461.17)}{15} = 370.6$$

$\overline{x_1} = 320 \quad \overline{x_2} = 298 \quad \overline{x_3} = 461.17$

$s_1 = 60.05 \quad s_2 = 77.91 \quad s_3 = 55.73$

$$F_{crit} = F_{c-1,N-c,\alpha} = F_{2,12,0.05} = 3.89$$

DR reject $H_0\ if\ F_{test} > 3.89$

SSC=$\sum ni(\bar{xi} - \bar{x})^2 = 5(320 - 370.6)^2 + 4(298 - 370.6)^2 + 6(461.17 - 370.6)^2 = 83\ 102.4$

SSE= $\sum(ni - 1)si^2 = 4(60.05)^2 + 3(77.91)^2 + 5(55.73)^2 = 48163.08$

$$MSC = \frac{SSC}{c-1} = \frac{83\ 102.4}{2} = 41\ 551.2$$

$$MSE = \frac{SSE}{N-c} = \frac{48163.08}{15-3} = 4013.59$$

$$F_{test} = \frac{MSC}{MSE} = \frac{41551.2}{4013.59} = 10.35 > 3.89 \quad \therefore\ reject\ H_0$$

∴ at least 2 means differ significantly

**Analysis of Variance Results**

F-statistic value = 10.3518

P-value = 0.00244

| Data Summary | | | | |
|---|---|---|---|---|
| Groups | N | Mean | Std. Dev. | Std. Error |
| Group 1 | 5 | 320 | 60.0458 | 26.8533 |
| Group 2 | 4 | 298 | 77.9145 | 38.9572 |
| Group 3 | 6 | 461.1667 | 55.733 | 22.7529 |

| ANOVA Summary | | | | | |
|---|---|---|---|---|---|
| Source | Degrees of Freedom DF | Sum of Squares SS | Mean Square MS | F-Stat | P-Value |
| **Between Groups** | 2 | 83098.8029 | 41549.4014 | 10.3518 | 0.0024 |
| **Within Groups** | 12 | 48164.8368 | 4013.7364 | | |
| **Total:** | 14 | 131263.6397 | | | |

**F5.** $c = 3\ groups \quad N = 79 + 81 + 81 = 241 \quad N - c = 238$

$\bar{x} = \frac{79(31.78)+81(32.88)+81(34.47)}{79+81+81} = 33.05$

SSC= $\sum ni(\bar{xi} - \bar{x})^2 = 79(31.78 - 33.05)^2 + 81(32.88 - 33.05)^2 + 81(34.47 - 33.05)^2 = 293.09$

SSE= $\sum(ni - 1)si^2 = 78(4.45)^2 + 80(4.4)^2 + 80(4.29)^2 = \quad 4565.723$

MSC= $\frac{SSC}{c-1} = \frac{293.09}{2} = 146.545$

MSE= $\frac{SSE}{N-c} = \frac{4565.723}{238} = 19.18$

$F_{test} = \frac{MSC}{MSE} = \frac{146.545}{19.18} = 7.64$

**b)** $F_{crit} = F_{c-1,N-c,\alpha} = F_{2,238,0.05} = 3.07 \quad (use\ 120)$

DR reject $H_0\ if\ F_{test} > 3.07$

$\therefore F_{test} = 7.64 > 3.07 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ differ\ significantly$

**c)** $H_0 \quad \mu i = \mu j$

$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{c,N-c,\alpha} = Q_{3,238,0.05} = 3.36 \quad (120)$

n's are all different

$\therefore HSD = Q_{crit}\sqrt{\frac{MSE}{2}(\frac{1}{ni} + \frac{1}{nj})}$

1&2 HSD= $3.36\sqrt{\frac{19.18}{2}(\frac{1}{79} + \frac{1}{81})} = 1.645$

$|31.78 - 32.88| = 1.1 < HSD \quad do\ not\ reject\ H_0$

$\therefore means\ are\ statistically\ equal$

1&3 $|31.78 - 34.47| = 2.69 > HSD \quad reject\ H_0$

$\therefore means\ are\ statistically\ different$

2&3 HSD= $3.36\sqrt{\frac{19.18}{2}\left(\frac{1}{81} + \frac{1}{81}\right)} = 1.64$

$|\overline{x_2} - \overline{x_3}| = |32.88 - 34.47| = 1.59 < HSD$

$\therefore do\ not\ reject\ H_0 \quad \therefore means\ are\ statistically\ equal$

$\therefore group\ 1\&3\ means\ are\ statistically\ different$

**F6.**

$\bar{x} = \frac{46(3.7)+111(3.1)+52(2.9)}{46+111+52} = 3.18$

$n_1 = 46 \quad \overline{x_1} = 3.7 \quad s_1 = 2.5 \quad c = 3\ groups$
$n_2 = 111 \quad \overline{x_2} = 3.1 \quad s_2 = 1.8 \quad N = 209$
$n_3 = 52 \quad \overline{x_3} = 2.9 \quad s_3 = 1.8 \quad \therefore N - c = 206$

SSC= $\sum ni(\bar{x}i - \bar{x})^2 = 46(3.7 - 3.18)^2 + 111(3.1 - 3.18)^2 + 52(2.9 - 3.18)^2 = 17.2256$

$SSE = \sum(ni - 1)si^2 = 45(2.5)^2 + 110(1.8)^2 + 51(1.8)^2 = 802.89$

$MSC = \frac{SSC}{c-1} = \frac{17.2256}{2} = 8.61$

$MSE = \frac{SSE}{N-c} = \frac{802.89}{20.6} = 3.897$

$F_{test} = \frac{MSC}{MSE} = \frac{8.61}{3.897} = 2.209$

$F_{crit} = F_{c-1,N-c,\alpha} = F_{2,206,0.05} = 3.07 \quad (120)$

$F_{test} = 2.209 < 3.07 \quad do\ not\ reject\ H_0$

$\therefore all\ means\ are\ statistically\ equal$

**F7.** $MSC = \frac{SSC}{c-1}$

$45.733 = \frac{SSC}{3-1} \quad \therefore SSC = 91.47$=A

**F8.** $MSE = \frac{SSE}{N-C} = \frac{276.4}{27} = 10.23$

$F = \frac{MSC}{MSE} = \frac{45.73}{10.23} = 4.47$=B

## G. Repeated-Measures ANOVA (Block Design)

**Example 1.**

C=4 groups n=8 subjects or 8 rows

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$$F_{crit} = F_{c-1,(C-1)(n-1),\alpha} = F_{3,21,0.05} = 3.07$$

DR $reject\ H_0\ if\ F_{test} > 3.07$

$$SSC = \sum ni(\bar{xi} - \bar{x})^2 = 8(9.375 - 9.406)^2 + 8(11.75 - 9.406)^2 + 8(10 - 9.406)^2 + 8(6.5 - 9.406)^2 = 114.34$$

$$SSR = c(\sum row\ mean - \bar{x})^2 = 4(10.25 - 9.406)^2 + 4(10.25 - 9.406)^2 + 4(4.25 - 9.406)^2 + \cdots + 4(11.25 - 9.406)^2 = 184.469$$

$SST = 377.72$

$SSE = SST - SSC - SSR$

$= 377.72 - 184.469 - 114.34$=78.91

$$MSC = \frac{SSC}{C-1} = \frac{114.347}{3} = 38.1$$

$$MSE = \frac{SSE}{(C-1)(n-1)} = \frac{78.91}{3(7)} = 3.76$$

$$\therefore F_{test} = \frac{MSC}{MSE} = \frac{38.1}{3.76} = 10.1$$

$F_{test} = 10.1 > 3.07 \quad reject\ H_0; at\ least\ 2\ means\ differ\ significantly$

**b)** Critical F=F n-1, (n-1)(c-1), alpha
=F 7, 21, 0.05=2.49

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4 \ldots = \mu8$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$$MSR = \frac{SSR}{n-1} = \frac{184.469}{7} = 26.35$$

$$F_R = \frac{MSR}{MSE} = \frac{26.35}{3.76} = 7 > F_{crit} = 2.49$$

$\therefore\ reject\ H_0$

$\therefore there\ is\ a\ statistical\ difference\ among\ subjects$

**c)** $H_0 \quad \mu i = \mu j$
$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{c,(c-1)(n-1),\alpha} = Q_{4,21,0.05} = 3.96$ (use 4, 20)

$HSD = Q_\alpha\sqrt{\frac{MSE}{n}} = 3.96\sqrt{\frac{3.76}{8}} = 2.71$

1&2 $|9.375 - 11.75| = 2.375 < HSD\ do\ not\ reject\ H_0$
$means\ are\ \ statistically\ equal$

2&4 $|11.75 - 6.5| = 5.25 > HSD\ \ reject\ H_0\ ; means\ are$
$statistically\ different$
$\therefore means\ group\ 2\&4\ are\ statistically\ different$

**Example 2.**

C=3 groups n=7 subjects or 7 rows

$H_0 \quad \mu1 = \mu2 = \mu3$
$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$$F_{crit} = F_{c-1,(C-1)(n-1),\alpha} = F_{2,12,0.05} = 3.89$$

DR $reject\ H_0\ if\ F_{test} > 3.89$

$$SSC = \sum ni(\overline{xi} - \bar{x})^2 = 7(10.86 - 9.05)^2 + 7(7.14 - 9.05)^2 + 7(9.14 - 9.05)^2 = 48.5261$$

$$SSR = c(\sum row\ mean - \bar{x})^2 = 3(9 - 9.05)^2 + 3(6 - 9.05)^2 + 3(15 - 9.05)^2 + \cdots + 3(13.33 - 9.05)^2 = 314.343$$

$SST = 382.95$

$$SSE = SST - SSC - SSR = 382.95 - 48.5261 - 314.343 = 20.08$$

$$MSC = \frac{SSC}{C-1} = \frac{48.5261}{2} = 24.263$$

$$MSE = \frac{SSE}{(C-1)(n-1)} = \frac{20.08}{2(6)} = 1.67$$

$$\therefore F_{test} = \frac{MSC}{MSE} = \frac{24.263}{1.67} = 14.5$$

$F_{test} = 14.5 > 3.89 \quad reject\ H_0; at\ least\ 2\ means\ differ\ significantly$

**b)** Critical F=F n-1, (n-1)(c-1), alpha
=F 6, 12, 0.05=2.9961

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4 \ldots = \mu7$
$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$$MSR = \frac{SSR}{n-1} = \frac{314.343}{6} = 52.39$$

$$F_R = \frac{MSR}{MSE} = \frac{52.39}{1.67} = 31.37 > F_{crit} = 2.9961$$

$\therefore\ reject\ H_0$
$\therefore\ there\ is\ a\ statistical\ difference\ among\ subjects$

**c)** $H_0 \quad \mu i = \mu j$

$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{c,(c-1)(n-1),\alpha} = Q_{3,12,0.05} = 3.77$

$HSD = Q_{\alpha}\sqrt{\frac{MSE}{n}} = 3.77\sqrt{\frac{1.67}{7}} = 1.84$

1&3 $|10.86 - 9.14| = 1.72 < HSD\ do\ not\ reject\ H_0$
$means\ are\ statistically\ equal$

d) $Qcrit = Q_{n(c-1)(n-1)} = Q_{7,12,0.05} = 4.95$

$HSD = Qcrit\sqrt{\frac{MSE}{C}} = 4.95\ \sqrt{\frac{1.67}{3}} = 3.69$

$|\overline{x_2} - \overline{x_3}| = |6 - 15| = 9 > HSD = 3.69$

$\therefore\ reject\ H_0$ ∴ subjects 2 & 3 do differ significantly

**G1.** C$= 3$ $columns$ $n = 4$ $rows$ $\alpha = 0.05$

$H_0$ $\mu1 = \mu2 = \mu3$

$H_a$ $at\ least\ 2\ means\ differ\ significantly$

$F_{crit} = F_{C-1,(C-1)(n-1),\alpha} = F_{2,2(3),0.05} = F_{2,6,0.05} = 5.14$

$\bar{x} = 31.42$

$DR\ reject\ H_0\ if\ F_{test} > 5.14$

$SSC = n(\sum \overline{xi} - \bar{x})^2$

$\uparrow column\ mean$

<u>Columns means</u> 41.5, 25.5, 27.75

<u>Row means</u> 34, 28, 31, 32.67

$\therefore SSC = 4(41.5 - 31.58)^2 + 4(25.5 - 31.58)^2 + 4(27.75 - 31.58)^2$
$= 600.17$

$SSR = c(\sum row\ mean - \bar{x})^2$
$= 3(34 - 31.58)^2 + 3(28 - 31.58)^2 + 3(31.67 - 31.58)^2 +$
$3(32.67 - 31.58)^2 = 59.607$

$SST = 796.9177$

$SSE = SST - SSC - SSR$
$= 796.9177 - 600.17 - 59.607 = 137.14$

$MSC = \frac{SSC}{C-1} = \frac{600.17}{2} = 300.09$

$MSE = \frac{SSE}{(C-1)(n-1)} = \frac{137.14}{2(3)} = 22.86$

$\therefore F_{test} = \frac{MSC}{MSE} = \frac{300.09}{22.86} = 13.13$

$F_{test} = 13.13 > 5.14$ $reject\ H_0\ ; at\ least\ 2\ means\ differ$
$significantly$

| *Source of Variation* | *Sum of Squres* | *df* | *MS* | *F* | *p value* |
|---|---|---|---|---|---|
| *Between Subjects* | 59.583336 | 3 | | | |
| *Between treatments* | 600.166664 | 2 | 300.083332 | 13.126368 | 0.006438 |
| *Within* | 137.166662 | 6 | 22.86111 | | |
| *Total* | 796.916662 | 11 | | | |

**b)** Critical F=F n-1, (n-1)(c-1), alpha
=F 3, 6, 0.05=4.76

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$MSR = \frac{SSR}{n-1} = \frac{59.607}{3} = 19.9$

$F_R = \frac{MSR}{MSE} = \frac{19.9}{22.86} = 0.87 < F_{crit} = 4.76$

$\therefore\ do\ not\ reject\ H_0$

$\therefore\ yes, the\ criterion\ is\ satisfied \quad \therefore\ no\ statistical\ difference\ among\ subjects$

**c) which columns differ (tests differ significantly)**

$H_0 \quad \mu i = \mu j$
$H_a \quad \mu i \neq \mu j$

$$Q_{crit} = Q_{c,(c-1)(n-1),\alpha} = Q_{3,6,0.05} = 4.34$$

$$HSD = Q_\alpha\sqrt{\frac{MSE}{n}} = 4.34\sqrt{\frac{22.86}{4}} = 10.4$$

1&2 $|41.5 - 25.5| = 16 > HSD \quad reject\ H_0$

1&3 $|41.5 - 27.75| = 13.75 > HSD \quad reject\ H_0$

2&3 $|25.5 - 27.75| = 2.25 < HSD \quad do\ not\ reject\ H_0\ ; means\ are\ statistically\ equal$

$\therefore means\ group\ 1\&2\ and\ 1\&3\ are\ statistically\ different$

**To test if row means (subjects differ), the critical would be**
**Q n, (n-1)(c-1) and HSD=** $\boldsymbol{Q_\alpha\sqrt{\frac{MSE}{C}}}$

**G2.** $c = 3 \quad n = 4$
Row means $\overline{x_1} = 12.67 \quad \overline{x_2} = 9.67 \quad \overline{x_3} = 10 \quad \overline{x_4} = 15.67$
Column means $\overline{x_1} = 9.5 \quad \overline{x_2} = 10.75 \quad \overline{x_3} = 15.75$

$H_0 \quad \mu_1 = \mu_2 = \mu_3$
$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$\bar{x} = \frac{138}{12} = 11.5$

$F_{crit} = F_{c-1,(c-1)(n-1),\alpha} = F_{2,6,0.05} = 5.14$

DR reject $H_0\ if\ F_{obt}\ or\ F_{test} > 5.14$

$SSC = n(\sum column\ mean - overall\ mean)^2$
$\therefore SSC = 4[(9.5-12)^2 + (10.75-12)^2 + (15.75-12)^2]$
$= 87.5$
$SSR = C[\sum(row\ mean - \bar{x})^2]$
$= 3[(12.67-12)^2 + (9.67-12)^2 + (10-12)^2 + (15.67-12)^2]$
$= 70.04$

$SST = \sum(each\ \# - \bar{x})^2 = 164$

$SSE = SST - SSC - SSR$
$= 164 - 87.5 - 70.04 = 6.46$
$MSE = \frac{SSE}{(n-1)(c-1)} = \frac{6.46}{2(3)} = 1.08$
$MSC = \frac{SSC}{c-1} = \frac{87.5}{2} = 43.75$

$F_c = \frac{MSC}{MSE} = \frac{43.75}{1.08} = 40.5 \quad > 5.14$ $\therefore$at least 2 means differ
Significantly

| *Source of Variation* | *Sum of Squres* | *df* | *MS* | *F* | *p value* |
|---|---|---|---|---|---|
| *Between Subjects* | 70.000005 | 3 | | | |
| *Between treatments* | 87.5 | 2 | 43.75 | 40.384628 | 0.000331 |
| *Within* | 6.499995 | 6 | 1.083333 | | |
| *Total* | 164.0 | 11 | | | |

**b)** Do subject differ at $\alpha = 0.05$

F critical=F 3, 6, 0.05=4.76 use (n-1), (n-1)(c-1) degrees of freedom

$$MSR = \frac{SSR}{n-1} = \frac{70.04}{3} = 23.35$$

$$F_{test} = \frac{MSR}{MSE} = \frac{23.35}{1.08} = 21.62 \quad > 4.76$$

$$\therefore reject\ H_0$$

∴ yes, subjects do differ significantly

**G3.** $c = 3 \quad n = 6$

Row means -3.93, 3.23, 13.93, 21.57, 15.9, 3.03

Column means 8.7, 8.17, 10

$\bar{x} = 8.96$ SST=1454.32

$H_0 \;\; \mu_1 = \mu_2 = \mu_3$

$H_a$ at least 2 means differ significantly

$F_{crit} = F_{c-1,(c-1)(n-1),\alpha} = F_{2,10,0.05} = 4.10$

DR reject $H_0 \; if \; F_{test} > 4.10$

$SSC = n(\sum(column\ mean - overall\ mean)^2)$

$= 6[(8.7 - 8.96)^2 + (8.17 - 8.96)^2 + (10 - 8.96)^2]$

$= 10.64$

$SSR = c[\sum(row\ mean - \bar{x})^2]$

$= 3[(-3.93 - 8.96)^2 + (3.23 - 8.96)^2 + (13.93 - 8.96)^2 + (21.57 - 8.96)^2 + (15.9 - 8.96)^2 + (3.03 - 8.96)^2]$

$= 1398.08$

$SST = \sum(each\ \# - \bar{x})^2$ =1454.32

$SSE = SST - SSC - SSR$

=1454.32 – 10.64 – 1398.08 =45.6

$MSE = \frac{SSE}{(n-1)(c-1)} = \frac{45.6}{(5)(2)} = 4.56$

$MSC = \frac{SSC}{c-1} = \frac{10.64}{2} = 5.32$

$MSR = \frac{SSR}{n-1} = \frac{1398.08}{5} = 279.62$

$\therefore F_{test} = \frac{MSC}{MSE} = \frac{5.32}{4.56} = 1.17 \;\; < 4.10$

$\therefore do\ not\ reject\ H_0$

∴ years don’t differ significantly

$H_0 \;\; \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_a$ at least 2 means differ significantly

F critical=F n-1, (n-1)(c-1)=F 5, 10, 0.05= 3.33

$F_{test} = \frac{MSR}{MSE} = \frac{279.62}{4.56} = 61.3 \; > 3.33$

$\therefore reject\ H_0$

∴ months do differ significantly

| *Source of Variation* | *Sum of Squres* | *df* | *MS* | *F* | *p value* |
|---|---|---|---|---|---|
| *Between Subjects* | 1397.951124 | 5 | | | |
| *Between treatments* | 10.671108 | 2 | 5.335554 | 1.167461 | 0.350201 |
| *Within* | 45.702212 | 10 | 4.570221 | | |
| *Total* | 1454.324444 | 17 | | | |

b) We concluded that the years do not differ significantly, so there is no test to perform.

c) $HSD = q_\alpha\sqrt{\frac{MSE}{C}} = 4.91\sqrt{\frac{4.56}{3}} = 6.05$

$\swarrow$ $\downarrow$

$q_{n,(c-1)(n-1),\alpha} = q_{6,10,0.05} = 4.91$ To see which months differed significantly, you would divide by c in the HSD formula and then calculate, you would find the difference between the means

There are 6 groups, so 6 choose 2 = 15 different pairings to calculate

$|-3.93 - 3.2|$ =7.13> HSD $reject\ H_0$
$|-3.93 - 13.97|$ =17.9>HSD $reject\ H_0$
$|-3.93 - 21.43|$ =25.36> HSD $reject\ H_0$
$|-3.93 - 15.87|$ =19.8>HSD $reject\ H_0$
$|-3.93 - 3.03|$ =6.96>HSD $do\ not\ reject\ H_0$
$|3.2 - 13.97|$ =10.77>HSD $reject\ H_0$
$|3.2 - 21.43|$ =18.23>HSD $reject\ H_0$
$|3.2 - 3.03|$ =0.17 <HSD $do\ not\ reject\ H_0$
$\therefore\ means\ are\ statistically\ equal$

$|13.97 - 15.87|$ =1.9<HSD $do\ not\ reject\ H_0$ $\therefore\ means\ are\ statistically\ equal$

$|13.97 - 3.03|$ =0.94<HSD $do\ not\ reject\ H_0$ $\therefore\ means\ are\ statistically\ equal$

$|21.43 - 15.87| = 5.56 < \text{HSD} \quad do\ not\ reject\ H_0$
$|21.43 - 3.03| = 18.4 > \text{HSD}\ reject\ H_0$
$|15.87 - 3.03| = 12.84 > \text{HSD}\ reject\ H_0$

For all pairs listed as greater than HSD above, we reject H0 and conclude the means are statistically different

**G4.** $n = 5 \quad (blocks) \; c = 4\,(treatments) \quad \bar{x} = 10$

*****note: treatment means are in rows not columns!***

$SSC = n\sum(treatment\ or\ column\ mean - overall\ mean)^2$

$5[(6-10)^2 + (16-10)^2 + (11-10)^2 + (7-10)^2]$

$= 310$

$SSR = c(\sum(row\ or\ block\ mean - overall\ mean)^2)$

$5[(14-10)^2 + (7-10)^2 + (12-10)^2 + (6-10)^2 + (11-10)^2]$

$= 184$

$SST = \sum(each\ \# - \bar{x})^2$

$= 4 + 64 + 4 + 81 + 9 + 100 + 16 + 49 + 4 + 49 + 9 + 9 + 9 + 4 + 16 + 1 + 25 + 0 + 49 + 16$

$= 518$

$SSE = SST - SSC - SSR$

$= 518 - 310 - 184$

$= 24$

$MSC = \frac{SSC}{c-1} = \frac{310}{3} = 103.33$

$MSE = \frac{SSE}{(c-1)(n-1)} = \frac{24}{4(3)} = 2$

$F_c = \frac{MSC}{MSE} = \frac{103.33}{2} = 51.7$ >3.49 so reject H0

G5.

C= 4 groups n=5 rows $\bar{x} = 3$

$H_0 \quad \mu1 = \mu2 = \mu3 = \mu4$

$H_a \quad at\ least\ 2\ means\ differ\ significantly$

$F_{crit} = F_{c-1,(c-1)(n-1),\alpha} \quad (4-1)(5-1)$

$= F_{3,12,0.05} = 3.49$

$DR\ reject\ H_0\ if\ F_{test} > 3.49$

$SSC = n\sum(column\ mean - \bar{x})^2$

$= 5(1-3)^2 + 5(2-3)^2 + 5(4-3)^2 + 5(5-3)^2 = 50$

$SSR = c(\sum row\ block\ mean - \bar{x})^2$

$= 4(5-3)^2 + 4(3-3)^2 + 4(3-3)^2 + 4(2-3)^2$

$+4(2-3)^2 = 24$

$SST = \sum(each\ \# - \bar{x})^2 = 82$

$SSE = SST - SSC - SSR = 82 - 50 - 24 = 8$

MSC= $\frac{SSC}{c-1} = \frac{50}{3} = 16.67$

$MSE = \frac{SSE}{(c-1)(n-1)} = \frac{8}{3(4)} = 0.67$

$F_c = \frac{MSC}{MSE} = \frac{16.67}{0.67} = 24.88 > 3.49 \quad \therefore reject\ H_0$

$\therefore at\ least\ 2\ means\ differ\ significantly$

b) $H_0 \quad \mu1 = \mu2 = \mu3 = \mu4 = \mu5$
$H_a \quad at\ least\ 2\ means\ differ\ significantly$

F critical= F n-1, (n-1)(c-1), alpha=F 4, 12, 0.05=3.26

$MSR = \frac{SSR}{n-1} = \frac{24}{4} = 6$

$F_R = \frac{MSR}{MSE} = \frac{6}{0.67} = 8.96 > 3.26 \quad reject\ H_0$
$\therefore yes, the\ subjects\ differ\ significantly\ from\ one\ another$

**c)** $H_0 \quad \mu i = \mu j$
$H_a \quad \mu i \neq \mu j$

$Q_{crit} = Q_{c,(c-1)(n-1),\alpha} = Q_{4,12,0.05} = 4.20$

$HSD = Q_{crit}\sqrt{\frac{MSE}{n}} = 4.20\sqrt{\frac{0.67}{5}} = 1.54$

1&2 $|\overline{x_1} - \overline{x_2}| = |1 - 2| = 1 < HSD \quad do\ not\ reject\ H_0$

1&3 $|1 - 4| = 3 > HSD \quad reject\ H_0\ ;\ means\ are\ not\ equal$

1&4 $|1 - 5| = 4 > HSD \quad reject\ H_0\ ; means\ are\ not\ equal$

2&3 $|2 - 4| = 2 > HSD \quad reject\ H_0\ ; means\ are\ not\ equal$

2&4 $|2 - 5| = 3 > HSD \quad reject\ H_0\ ; means\ are\ not\ equal$

3&4 $|4 - 5| = 1 < HSD \quad do\ not\ reject\ H_0 \quad \therefore means\ are\ statistically\ equal$

$\therefore groups\ 1\&3, 1\&4, 2\&3\ and\ 2\&4\ are\ statistically\ different$

## H. Simple Linear Regression

**Example 1.**

| $\hat{y} = 10.9 - 0.97x$ | Lawn | Amt. of Chemical (g) $x$ | Surviving Beetles $y$ |
|---|---|---|---|
| 8.96 | A | 2 | 11 |
| 6.05 | B | 5 | 6 |
| 5.08 | C | 6 | 4 |
| 7.99 | D | 3 | 6 |
| 2.17 | E | 9 | 3 |
| | | Sum=25 | Sum=30 |

$\bar{x} = 5 \quad \bar{y} = 6$

a) $SS_{xx} = \sum(x - \bar{x})^2 = 30$ (or use other formula)

$$SS_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2 = 2^2 + 5^2 + 6^2 + 3^2 + 9^2 - \frac{1}{5}(25^2) = 30$$

$SS_{yy} = \sum(y - \bar{y})^2 = 38$ ( or use the other formula)

$$SS_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2 = 11^2 + 6^2 + 4^2 + 6^2 + 3^2 - \frac{1}{5}(30^2)=38$$

$$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y) \quad (sum\ of\ x)\ and\ (sum\ of\ y)$$

$$= 2(11) + 5(6) + 6(4) + 3(6) + 9(3) - \frac{1}{5}(25)(30)$$

$$= 121 - 150 = -29$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-29}{\sqrt{30(38)}} = \frac{-29}{33.7639} = -0.86$$

b) $b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-29}{30} = -0.97$

$$b_0 = \bar{y} - b_1\bar{x} = 6 + 0.97(5) = 10.9$$

$$\hat{y} = b_0 + b_1 x = 10.9 - 0.97x$$

c) $SS_{xx} = \sum(x - \bar{x})^2 = 30$

$t_{\alpha/2} = 3.182 \qquad df = n - 2 = 5 - 2 = 3$ 95% confidence interval

$$SSE = \sum(y - \hat{y})^2 = (11 - 8.96)^2 + (6 - 6.05)^2 + (4 - 5.08)^2 + (6 - 7.99)^2 + (3 - 2.17)^2 = 9.9795$$

**or do** $\boldsymbol{SSE = SS_{yy} - \frac{(SS_{xy})^2}{SS_{xx}}} = 38 - \frac{(-29)^2}{30} = \mathbf{9.97}$

**...this is much faster than calculating all of the predicted y-values**

Find standard error, $s_e$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{9.97}{3}} = 1.82$$

$$b_1 \pm t_{\alpha/2}\frac{Se}{\sqrt{SS_{xx}}} = -0.97 \pm 3.182\left(\frac{1.82}{\sqrt{30}}\right) = -0.97 \pm 1.06$$
$$= (-2.03, 0.09)$$

d) $r^2 = (-0.86)^2 = 0.74 \quad \therefore 74\%$

e) $x_p = 4 \quad \hat{y} = 10.9 - 0.97(4) = 7.02 \quad s_e = 1.82$
$SS_{xx} = 30 \quad n = 5 \quad \bar{x} = 5$

$$\hat{y} \pm t_{\alpha/2}(s_e)\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} = 7.02 \pm 3.182(1.82)\sqrt{\frac{1}{5} + \frac{(4-5)^2}{30}}$$
$$= 7.02 \pm 2.797$$
$$= (4.22, 9.82)$$

f) $\hat{y} \pm t_{\alpha/2}(se)\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$

$$= 7.02 \pm 3.182(1.82)\sqrt{1 + \frac{1}{5} + \frac{(4-5)^2}{30}} = 7.02 \pm 6.4$$

g) $H_0 \quad \beta_1 = 0$

$H_a \quad \beta_1 \neq 0$ $\quad \left.\begin{matrix} \alpha = 0.05 \\ \frac{\alpha}{2} = 0.025 \\ df = n - 2 = 3 \end{matrix}\right\} \quad t_{crit} = +/-3.182$

Rejection regions are below -3.182 and above 3.182

$$t = \frac{b_1 - 0}{\frac{Se}{\sqrt{\sum(x-\bar{x})^2}}} = \frac{-0.97}{\frac{1.82}{\sqrt{30}}} = -2.92$$

$t_{test} > -3.182 \quad \therefore do\ not\ reject\ H_0$

$\therefore no\ statistical\ evidence\ that\ \beta_1 \neq 0$, so there is no significant relationship

or use p-values

go across at 3df 2.353<2.92<3.182

0.025(2) <2 sided p-value <2(0.05)...pvalue> alpha=0.05

so fail to reject Ho and $\beta_1 = 0$.

**Example 2.**

$\bar{x} = 15.93 \quad r^2 = 0.7277 \quad r = 0.85 \quad s_e = 47.9$

$b_0 = 76.535142 \qquad b_1 = 4.3331081 \qquad S_{\widehat{B1}} = 91$

a) $\hat{y} = b_0 + b_1 x = 76.5 + 4.3x$

x=gross in millions

y=# units expected to sell

b) $76.535 \times 1000 = \$76\,535 \qquad (number\ of\ units\ sold\ with\ x = 0\ gross\ sales)$

$\beta_1 = 4.333$(rate of change)

c) $x = 20 \quad (in\ millions)$

$$\hat{y} = 76.535 + 4.3x = 76.535 + 4.3(20)$$

$\hat{y} = 163.2(1000) = 163\,200$ units

d) $\alpha = 0.05 \qquad H_0 \ \ \beta_1 = 0$

$H_a \ \ \beta_1 \neq 0\ (2\ sided)$

$$t\ test = \frac{b_1}{s\ b1}$$

$$t_{test} = \frac{4.33}{0.50084} = 8.65$$

Look up t critical= t (0.05, 28)=1.701 and reject H0 if t test>1.701 or if t test < -1.701 (2 sided test)

n=30

OR use p=value since it's given

$p - value = 2.216 \times 10^{-9} < 0.05$

$\therefore strongly\ reject\ H_0\ and\ strong\ evidence\ that\ video\ sales\ and\ box\ office\ gross\ have\ a\ linear\ relationship$

Without the chart you can say t test = 8.65>3.408 at 28 df

2 sided p-vaue<2(0.001)

So, the p-value is less than alpha=0.05 and we reject H0 and conclude there is statistically significant evidence that $\beta_1 \neq 0$

e) $df = n - 2 = 28$

$b_1 \pm t_{\alpha/2} S_{\widehat{B1}}$

$= 4.333 \pm 2.048(0.500843491)$

$= 4.333 \pm 1.0257$

$= (3.307, 5.3587)$

f) $SS_{xx} = \left(\frac{Se}{S_{\widehat{B1}}}\right)^2 = \left(\frac{47.9}{0.50084}\right)^2 = 9146.8$

$\hat{y} = 76.5 + 4.3(10) = 119.865$

$t_{\alpha/2(0.025,8)} = 2.048$

$s_e = 47.9$

$n = 30$

$$\hat{y} \pm t_{\alpha/2}(s_e)\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

$$= 119.865 \pm (2.048)(47.9)\sqrt{\frac{1}{30} + \frac{(10-15.93)^2}{9146.8}}$$

$$= 119.865 \pm 18.92 = (100.99, 138.7)$$

g) $df = n - 2 = 28$

$$\hat{y} \pm t_{\alpha/2}(s_e)\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

$$= 119.865 \pm 2.048(47.8668)(1.018)$$

$$= 119.865 \pm 99.84 = (20.18, 219.55)$$

h) In (f) we find 95% CI for the ***average*** video sales for a movie grossing $10 million. In (g) we obtained a prediction interval for one video sale.

**Example 3.**

$$\bar{x} = 41.5 \quad \bar{y} = 2.375 \quad \sum x = 332$$

$$\sum y = 19$$

$$SS_{xx} = \sum(x-\bar{x})^2 = (21-41.5)^2 + \cdots + (64-41.5)^2 = 1578$$

$SS_{yy} = \sum(y-\bar{y})^2 = (4-2.375)^2 + \cdots (6-2.375)^2 = 29.875$

$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y)$

$= (21)(4) + (26)(0) + 33(3) + 35(1) + 48(3) + 50(0) + 55(2) \ +64(6) - \frac{1}{8}(332)(19) = 856 - 788.5 = 67.5$

$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{67.5}{\sqrt{(1578)(29.875)}} = 0.3109$

**Find the equation too**

$$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{67.5}{15.78} = 0.0428$$

$$b_0 = \bar{y} - b_1\bar{x} = 2.375 - 0.0428(41.5) = 0.5988$$

$$\hat{y} = b_0 + b_1 x = 0.5988 + 0.0428x$$

**H1.**

$S_x = \sqrt{\frac{1240}{5}} = 15.748 \quad (std\ dev\ of\ x\ values)$

$S_y = \sqrt{\frac{2554}{5}} = 22.6$ (std deviation of y values)

$n = 6$

$\bar{x} = \frac{180}{6} = 30 \quad \bar{y} = \frac{270}{6} = 45$

| | | | | | | |
|---|---|---|---|---|---|---|
| $x\ Rainfall\ (mm)$ | 11 | 24.2 | 25 | 27 | 48 | 59.8 |
| $y\ Demand\ (1000kg)$ | 22.01 | 38 | 28 | 41.4 | 51 | 81 |

a) $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 9592 - \frac{180(270)}{6} = 1492$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} = \frac{9592 - \frac{180(270)}{6}}{\sqrt{(1240)(2554)}} = \frac{1492}{1779.5955} = 0.84$$

$b_1 = r\frac{Sy}{Sx} = 0.84\left(\frac{22.6}{15.748}\right) = 1.2$ OR

use SSxx and SSxy formula! Same answer!!

b) $b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{1492}{1240} = 1.02$

$$b_0 = \bar{y} - b_1\bar{x} = 45 - 1.2(30) = 9$$

$\hat{y} = b_0 + b_1 x = 9 + 1.2x$

c) $sub\ x = 11 \quad \hat{y} = 9 + 1.2(11) = 22.2$

$x = 27$

$\hat{y} = 9 + 1.2(27) = 41.4$

$sub\ y = 38 \quad 38 = 9 + 1.2x \quad \rightarrow x = 24.2$

**H2.**

$r = 0.95 \quad r^2 = 0.95^2 = 0.9025$

The answer is C.

**H3.**

**a) No, since $5200 is outside of the values we are given and our line might not work outside...i.e. no extrapolation**

$\bar{x} = \frac{163}{7} = 23.3 \quad \bar{y} = \frac{46}{7} = 6.6 \quad \sum x = 163$

$\sum y = 46$

$SS_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2$

$= 4295 - \frac{(163)^2}{7} = 499.43$

$SS_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2 = 326 - \frac{(46)^2}{7} = 23.714$

$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y) = 1172 - \frac{(163)(46)}{7}$

$= 100.86$

$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{100.86}{\sqrt{(499.43)(23.714)}} = 0.93$

$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{100.86}{499.43} = 0.2$

$b_0 = \bar{y} - b_1\bar{x} = 6.6 - 0.2(23.3) = 1.94$

$\hat{y} = b_0 + b_1 x = 1.94 + 0.2x$

No, because our data doesn't cover that value and the trend might not continue

b) $\therefore x = 100 \quad 0.2(100) = \$20.00 \quad \textit{increase by } \$20$

c) $x_3 = 16$ Point(16,5)

$\hat{y}p = 1.94 + 0.20(16) = 5.14$

Residual=$y - \hat{y} = 5 - 5.14 = -0.14$

d) False, correlation is a unitless measure. It is just $r = 0.93$.

**H4**.

**a)** $n = 10 \quad r^2 = 0.59 \quad r = 0.7681 \quad se = 0.88$
$b_0 = 8.18 \quad b_1 = 0.58 \quad S_{\widehat{B1}} = 0.17$

$$\hat{y} = b_0 + b_1 x = 8.18 + 0.58x$$

Here it is a positive relationship, since y increases as x increases. Profit increases by 0.58 x1000 as x goes up by $1000

**b)** 95% CI
$df = 10 - 2 = 8 \quad t_{crit} = 2.306$
$b_1 \pm t_{\alpha/2} sb1$ from the table
$= 0.58 \pm 2.306(0.17)$
$= 0.58 \pm 0.392 = (0.188, 0.9720)$

c)$H_0 \quad \beta_1 = 0$
$H_a \quad \beta_1 \neq 0 \quad (2\ sided) \quad t_{crit} = 2.306$

$$t_{test} = \frac{b_1 - 0}{sb1} = \frac{0.58}{0.17} = 3.4118$$

Or do t test =3.41 (from table) p-value<1%. So reject H0

$go\ across\ 8\ df \quad 3.355 < t = 3.4118 < 4.501$
$2(0.001) < p - value < 2(0.005)$
$0.002 < 2\ sided\ p - value < 0.01$
$t_{test} > t_{crit} \quad \therefore reject\ H_0 \quad \therefore evidence\ \beta_1 \neq 0$

**H5.** a)

$r = \frac{SSxy}{\sqrt{SSxxSSyy}} = \frac{-35}{\sqrt{30(90)}} = -0.67$ (Correlation coefficient)
$r^2 = (-0.67)^2 = 0.45$ (Coefficient of determination)
$\therefore 0.45, -0.67$
The answer is C).

b) $b_1 = \frac{SSxy}{SSxx} = \frac{-35}{30} = -1.17$
$b_0 = \bar{y} - b_1\bar{x} = 110 - (-1.17)(18) = 131.06$
$\hat{y} = b_0 + b_1 x$
$= 131.06 - 1.17x$
$= 131.06 - 1.17(15)$
$= 113.51$

**H6.** $\bar{x} = \frac{3398}{20} = 169.9$ $\bar{y} = \frac{972.5}{20} = 48.6$ $\sum x = 3398$
$\sum y = 972.5$

$$SS_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2$$
$$= 701940 - \frac{(3398)^2}{20} = 124619.8$$

$$SS_{yy} = \sum y^2 - \frac{1}{n}\left(\sum y\right)^2 = 49802.31 - \frac{(972.5)^2}{20} = 2514.4975$$
$$SS_{xy} = \sum xy - \frac{1}{n}\left(\sum x\right)\left(\sum y\right) = 182677.8 - \frac{(3398)(972.5)}{20}$$
$= 17450.05$

a) $b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{17450.05}{124619.8} = 0.14$
$b_0 = \bar{y} - b_1\bar{x} = 48.6 - 0.14(169.9) = 24.8$
$\hat{y} = b_0 + b_1 x = 24.8 + 0.14x$

b)
The intercept means that the min. time for any delivery is 24.8 minutes and the slope represents the increase per case, ie. 0.14 minutes per case

c) x=150

$\hat{y} = b_0 + b_1 x = 24.8 + 0.14(150) = 45.8\, min$

d)
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{17450.05}{\sqrt{124619.8(2514.4975)}} = 0.99$$
$r^2 = 0.99(0.99) = 0.98$ So, 98% of the variation in the y-values can be explained by this regression model

e)
$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{17450.05}{\sqrt{124619.8(2514.4975)}} = 0.99$ It is a very strong linear relationship

f) $SSE = SS_{yy} - b_1 SS_{xy}$ n=20
SSE=$2514.4975 - 0.14(17450.05) = 71.4905$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{71.4905}{18}} = 1.993$$

g) df=n-2=18 and 95% confidence, so t crit=2.101

$$\hat{y} \pm t_{\frac{\alpha}{2}}(se)\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SS_{xx}}}$$

$$= 45.8 \pm 2.101(1.993)\sqrt{\frac{1}{20}+\frac{(150-169.9)^2}{124619.8}}$$

$$= 45.8 \pm 2.101(1.993)(0.230603) = 45.8 \pm 0.97$$

h) $\hat{y} \pm t_{\frac{\alpha}{2}}(Se)\sqrt{1+\frac{1}{n}+\frac{(xp-\bar{x})^2}{SS_{xx}}}$

$$= 45.8 \pm 2.101(1.993)\sqrt{1+\frac{1}{20}+\frac{(150-169.9)^2}{124619.8}}$$

$$= 45.8 \pm 2.101(1.993)(1.0262) = 45.8 \pm 4.297$$

$i)\ df = n-2 = 18$

$$= b_1 \pm t_{\alpha/2}\frac{Se}{\sqrt{SS_{xx}}}$$

$$= 0.14 \pm 2.101(1.993/\sqrt{124619.8})$$

$$= 0.14 \pm 0.0119$$

$$= (0.128, 0.152)$$

**H7.**

$$\bar{x} = 137.8 \quad \bar{y} = 129.8$$

a) $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

$SS_{xx} = \sum(x-\bar{x})^2 = (112-137.8)^2 + \cdots + (135-137.8)^2$
$= 2518.8$
$SS_{yy} = \sum(y-\bar{y})^2 = (110-129.8)^2 + \cdots + (129-129.8)^2$
$= 1750.8$

OR

$SS_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2$
$= [112^2 + 123^2 + 178^2 + 141^2 + 135^2]$
$-\frac{1}{5}[112 + 123 + 178 + 141 + 135]^2 = 97\,463 - 94\,944.2$
$= 2518.8$
$SS_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2$
$= [110^2 + 120^2 + 165^2 + 125^2 + 129^2]$
$-\frac{1}{5}[110 + 120 + 165 + 125 + 129]^2 = 1750.8$
$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y)$
$= 112(110) + 123(120) + 178(165) + 141(125) + 135(129)$
$-\frac{1}{5}(689)(649) = 91490 - 89\,432.2 = 2057.8$

$$r = \frac{2057.8}{\sqrt{(2518.8)(1750.8)}} = 0.98$$

b) It would be the same since all values are going down by the same amount.

c) $b_1 = r\frac{Sy}{Sx} = 0.98\left(\frac{20.9}{25.1}\right) = 0.817$

or use $b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{2057.8}{2518.8} = 0.817$

$b_0 = \bar{y} - b_1 \quad \bar{x} = 129.8 - 0.817(137.8) = 17.217$
$\hat{y} = b_0 + b_1x$
$\therefore \hat{y} = 17.217 + 0.817x$

**H8.**

$\bar{x} = 254 \qquad \bar{y} = 110 \sum x = 1270$

$\sum y = 550$

$\sum x^2 = 336700$

$$\sum y^2 = 62500$$

$SS_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2$

$= 336700 - \frac{1}{5}[1270]^2 = 14120$

$SS_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2$

$= [62500] - \frac{1}{5}[550]^2 = 2000$

$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y)$

$= 144600 - \frac{1}{5}(1270)(550) = 4900$

$a)\ b_1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{4900}{14120} = 0.347$

$b_0 = \bar{y} - b_1\bar{x} = 110 - 0.347(254) = 21.86$

$\therefore \hat{y} = 21.86 + 0.35x$

$b)\ r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{4900}{\sqrt{(14120)(2000)}} = \frac{1}{4}\ \frac{(4900)}{(22.4)(59.4)} = 0.92$

$r^2 = 0.92^2 = 0.85$

c) $H_0 \quad \beta_1 = 0$

$H_a \;\; \beta_1 \neq 0 \qquad \alpha = 0.05 \quad t_{0.025}$

$df = 5 - 2 = 3 \qquad t_{crit} = 3.182$

DR Reject H0 if t test>3.182 or < -3.182

$$SSE = SS_{yy} - \frac{(SS_{xy})^2}{SS_{xx}} = 2000 - \frac{(4900)^2}{14120} = 299.58$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{299.58}{3}} = 9.993$$

$t\ test = \frac{b_1 - 0}{\frac{Se}{\sqrt{SS_{xx}}}} = \frac{0.347}{\frac{9.993}{\sqrt{14\,120}}}$ = 4.13>3.182, so we reject H0

or use p-values

$$3.182 < t = 4.13 < 4.541$$

$$0.02 < 2\ sided\ p - value < 0.05$$

$$\therefore p - value\ < \alpha \quad \therefore reject\ H_0 \quad \therefore yes, there\ is\ evidence$$

d) $df = 3 \quad t_{\alpha/2} = 3.182 \quad x_p = 300 \quad \hat{y} = 21.1 + 0.35(300)$

$\hat{y}_p = 126.1$

$$\hat{y} \pm t_{\alpha/2}(s_e)\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

$$= 126.1 \pm 3.182(9.993)\sqrt{1 + \frac{1}{5} + \frac{(300-254)^2}{14\,120}}$$

$$= 126.1 \pm 36.94$$

**H9.**

**a)** $b_1 = r\frac{Sy}{Sx} = 0.9721\left(\frac{139.46}{126.88}\right) = 1.07$

$b_0 = \bar{y} - b_1\bar{x} = 236.56 - 1.07(201.75) = 20.7$

$\hat{y} = 20.7 + 1.07x$

**b)** $r^2 = 0.9721^2 = 0.94$

$\therefore 94\%\ of\ the\ variation\ in\ y\ (selling\ price) can\ be\ explained\ by\ the\ regression\ model\ or\ by\ x(appraised\ value).$

**c)** $x = \frac{50\,000}{1000} = 50$

$1.07(50) = 53.5 \qquad (\times 1000)$

$\therefore\uparrow \ \$53\,500\ in\ selling\ price$

**d)** $\hat{y} = 20.7 + 1.07(475) = 528.95$
$\therefore$ *selling price is* \$ 528 950.

**H10.**

$\bar{x} = 622.3$ $r^2 = 0.797$ $r = 0.89$ $Se = 0.156$ $n = 20$
$b_0 = 0.30032331$ $b_1 = 0.00487023$
$S_{\widehat{B1}} = 0.00057786$

a) $\hat{y} = 0.3 + 0.00487x$

b) $\hat{B}_0 = GPI \text{ when } GMAT \text{ score} = 0$
$\hat{B}_1 = GPI \text{ increase by } 0.00487023 \text{ for every } 1 \text{ pt increase in } GMAT \text{ score}$

c) $x = 600$ $\hat{y} = 0.3 + 0.00487(600) = 3.22$
$\therefore 3.22\ GPI$

d) $s = Se = 0.1559$ (first table)

e) $r^2 = 0.797$ 79.7% *of variation in y can be explained by regression model.* (first table)

f) $r = \sqrt{0.797} = 0.89$

g) $H_0$ $\beta_1 = 0$
$H_a$ $\beta_1 \neq 0$ p-value method
$p - value = 1.158 \times 10^{-9} < 0.05$
$\therefore$ *strongly reject* $H_0$ $\therefore$ strong evidence that GPI and GMAT have a linear relationship

h) $x_p = 600$ $\hat{y} = 0.3 + 0.00487(600) = 3.22$

$$\sum(x - \bar{x})^2 = SS_{xx} = \left(\frac{Se}{S_{\widehat{B1}}}\right)^2 = \left(\frac{0.156}{0.00057786}\right)^2 = 72\,879.3$$

$$\hat{y}_p \pm t_{\alpha/2}(s_e)\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

$$= 3.22 \pm (2.101)(0.156)\sqrt{\frac{1}{20} + \frac{(600 - 622.8)^2}{72\,879.3}}$$

$$= 3.22 \pm 0.078$$

i) $$\hat{y}_p \pm t_{\alpha/2}(s_e)\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

$$= 3.22 \pm (2.101)(0.156)\sqrt{1 + \frac{1}{20} + \frac{(600 - 622.8)^2}{72\,879.3}}$$

$$= 3.22 \pm 0.337$$

j) $df = 18 \quad t_{\alpha/2} = 2.101$

$b_1 \pm t_{\alpha/2} S_{\widehat{B1}}$

$= 0.00487023 \pm 2.101(0.00057786)$

$= 0.00487023 \pm 0.0012 = (0.00366, 0.00608)$

## I. Multiple Regression

**Example 1.**

a) R^2 = 0.59.

Therefore, 59% of the total variation can be explained by the model

b) y = 37724.15 + 34.16x1 – 10848.75x2 + 24668.46x3 + 4588.64x4

c) Test for the overall significance of the model. Clearly state the hypothesis being tested, as well as your conclusion

n-k-1=44-4-1=39

F test = 14.28 (MSR/MSE) p-value=0<alpha=5%
Reject Null hypothesis – model has some predictive power, not all coefficients are zero.

Or do F critical = F k, n-k-1=F 5, 38, 0.05=2.53 and reject if F test>2.53
So, F test=14.28>2.53, so reject H0 and conclude at least one Bi$\neq 0$
and therefore at least one variable is significant

d) Are there any variables that could be dropped from the model? Provide a statistical argument.

Can drop B4 – number of cars in garage, as when testing individually we can see that this variable has no predictive power since the p-value is greater than 10%.

**Example 2.**

n-k-1=63-4-1=58 (from table, error df)
k=4 (from table, regression df)
n-1=62 from the table (total df)
a) $\hat{y} = 3829.5 - 5056b_1 + 1667.7b_2 + 804.12b_3 - 31.49b_4$
b) display should be omitted because the p-value=0.558 > 10% level of significance

c) If the value of b3 is increased by 1 unit while keeping all of the other independent variable unchanged, then y will increase by 804.12 units.

d) $99\%\ CI \quad df = n - k - 1 = 58 \ \ (use\ 50)$

$$t_{\alpha/2} = 2.678$$

$$CI = b_3 \pm t_{\alpha/2}(sb_3)$$

$$= 804.12 \pm (2.678)(86.75)$$

$$= 804.12 \pm 232.32$$

e) $H_0 \quad \beta_3 = 0$
$H_a \quad \beta_3 > 0$
$t\ test = \frac{b_3}{s_{b3}} = \frac{804.12}{86.75} = 9.27$ use p-value in table=0 < alpha, so reject H0
Or use critical values
alpha=5%
df=n-k-1=58 and t critical=1.676
D.R. reject $H_0$ if t test>1.676 and t test=9.27, so reject $H_0$ and conclude that yes, $\beta_3 > 0$ .

f) testing overall significance
$H_0 \quad \beta 1 = \beta 2 = \beta 3 = \beta 4 = 0$
$H_a \quad at\ least\ one\ \beta i \neq 0$

$F_{crit} = F_{k,n-k-1,\alpha} = F_{4,58,0.05}(use\ 4,40) = 2.606$
$DR\ reject\ H_0\ if\ F_{test} > 2.606$

$$F\ test = \frac{MSR}{MSE} = \frac{4593416}{26520} = 173.21 > 2.606$$

Therefore, reject H0 and conclude $at\ least\ one\ \beta i \neq 0$
Or use p-value in table=0< alpha, so we reject H0

If asked for p-value, look up (4,40)df for various alpha levels

0.10 F critical=2.09

0.05 F critical = 2.61

0.01 F critical = 3.83

F test=173.21 so the p-value is less than 0.01

**Example 3.**

1. The answer is d).
2. The answer is b).
k=4 and n=35, so n-k-1=35-4-1=30
F critical=$F_{k,n-k-1,\alpha} = F_{4,30,0.05=3.25}$
Find F test= $\frac{R^2/_k}{1-R^2/_{n-k-1}} = \frac{0.923/_4}{1-0.923/_{30}} = 89.9 > 3.25$
3. $t\ test = \frac{b_1}{s_{b1}} = \frac{1.103130}{0.359573} = 3.07$
2.75<3.07<3.385 at df=30
0.001<1 sided p-value<0.005(best you can do)
The answer is d).
4. The answer is c) since it has the smallest p-value
5. 1-0.95=0.05 and 0.05/2=0.025, so look up t crtical 30, 0.025=2.042
a 95% confidence interval is $b_1 \pm t_{\alpha/2}s_{b1}$=1.103$\pm(2.042)(0.359573) = 1.103 \pm 0.734$
The answer is c).
6. The answer is c), 30 df.
7. The answer is d), the salaries, or the y in this question.

**Example 4.**
n-k-1=30-k-1=29-k

Standard error of estimate:$s_e = \sqrt{\frac{SSE}{n-k-1}}$

$1.645 = \sqrt{\frac{65}{29-k}}$

$2.706025 = \frac{65}{29-k}$

$2.70625(29-k) = 65$

29-k=24

k=5. The answer is B).

**Example 5.**

$SSE = SS_{yy}(1-r^2)$

120=$SS_{yy}$(1-0.98)

$SS_{yy} = 6\ 000 = SST$

Adjusted R² = $1 - \frac{\frac{SSE}{n-k-1}}{\frac{SST}{n-1}} = 1 - \frac{\frac{120}{9}}{\frac{6000}{14}} = 0.969$

So, it is closest to 0.97 answer B).

**Example 6.**

(a) $x_1 \uparrow by\ 1000 \quad \therefore 1000(-0.009) = -9$

$x_2 \uparrow by\ 1000 \quad \therefore 1000(0.0023) = 2.3$

Both $-9 + 2.3 = -6.7$

$\therefore$ go down by 6.7 units

(b) $n - k - 1 = 25 - 3 - 1 = 21$

$H_0\ B_1 = 0 \qquad H_a\ B_1 \neq 0(2\ sided)$

$t_1 test = \frac{b_1}{sb_1} = \frac{-0.009}{0.0024} = -3.75$

$t_3 test = \frac{b_3}{sb_3} = \frac{0.875}{0.234} = 3.74$

$t_{crit} = t_{n-k-1}, \alpha/2 = t_{crit} 21df, 0.025 = 2.08$

Reject if $t\ test > 2.08\ or < -2.08$

$\therefore t_1\ reject\ H_0 \qquad t_3\ reject\ H_0$

$\therefore x_1 and\ x_3$ are significant

I1.a) $x_1 \uparrow by\ 1000 \quad \therefore 1000(-6.2) = -6200$

$x_2\ down\ by\ 1000 \quad \therefore (-1000)(2.3) = -2300$

Both $-6200 - 2300 = -8500$

$\therefore$ go down by 8500 units

$b)\ n - k - 1 = 25 - 3 - 1 = 21$

$H_0\ B_1 = 0 \qquad H_a\ B_1 \neq 0(2\ sided)$

$t_1 test = \frac{b_1}{sb_1} = -\frac{6.2}{1.4} = -4.43$

$t_3 test = \frac{b_3}{sb_3} = \frac{.875}{0.54} = 1.62$

$t_{crit} = t_{n-k-1}, \alpha/2 = t_{crit} 21df, 0.025 = 2.08$

Reject if $t\ test > 2.08\ or < -2.08$

$\therefore t_1\ reject\ H_0 \qquad t_3\ do\ not\ reject\ H_0$

$\therefore x_1\ is$ significant

I2.

a) $\hat{y} = 80 + 16x_1 + 8x_2 + 24x_3 + 5x_4$

b) Calculate $R^2$. Conduct a formal hypothesis test of the validity of the model at a significance level of 5%.

$$R^2 = \frac{SSR}{SST} = \frac{1244}{3303} = 0.377$$

Test for validity So test would be: $H_0: \beta 1 = \beta 2 = \beta 3 = \beta 4 = 0$ vs. $Ha$ : at least one of them is not equal to 0.

$$F = \frac{MSR}{MSE} = \frac{SSR/_k}{SSE/_{n-k-1}} = \frac{1244/_4}{2059/_{21}} = 3.172$$

The rejection region is greater than F critical=F k, n-k-1, alpha
=F (4, 21, 0.05)=2.84
Since F = 3.172>2.84 we reject the null hypothesis. Therefore, the model is valid.

c)

Hypothesis test: $H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

Test statistic: $t = \frac{b_2 - \beta_2}{s_{b_2}} = \frac{8-0}{4} = 2$

Rejection region: $t > t_{\alpha/2, n-k-1} = t_{0.025,21} = 2.08$ or $t < -2.08$

Since t = 2 we do not reject the null hypothesis that $\beta_2 = 0$. The test suggests that in this model the term for advertising is insignificant for determining sales.
Sales do not depend on advertising in a quadratic manner in this model.

d)

Hypothesis test:

$H_0 : \beta_3 = 0$
$H_1 : \beta_3 \neq 0$

Test statistic:

$$t = \frac{b_3 - \beta_3}{s_{b_3}} = \frac{24-0}{10} = 2.4$$

Rejection region: $t > t_{\alpha/2, n-k-1} = t_{0.025,21} = 2.08$, or $t < -2.08$

Since $t$ = 2.4 we reject the null hypothesis that $\beta_3 = 0$. Therefore, the infomercial parameter is significant.

I3. k=4, n=12, n-k-1=12-4-1=7
F critical= $F_{0.05, 4, 7}$ =4.12 and we would reject if F test>4.12
The answer is D).

I4.(a) Briefly interpret the regression coefficients for AGE and EDUC.

As age goes up by one year, the number of weeks a manufacturing worker has been jobless increases on average by 20 holding all else constant.

As number of years of education goes up by one year, the number of weeks a manufacturing worker has been jobless decreases on average by 10 holding all else constant.

b) Hypothesis:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ (i.e., model is not significant)
$H_1$: Not all $\beta_i = 0, \ i = 1,2,\ldots,7$ (i.e., model is significant)
n-k-1=50-7-1=42
MSR $= {SSR}/{k}$ = 7840 / 7 = 1120
MSE $= {SSE}/{n-k-1}$ = 1960 / 42 = 46.667

F = MSR / MSE =1120/46.67= 24

When the significance level is 5% we reject the null hypothesis when F > 2.14 (as found from the table with df (7,42)). Since here 24 > 2.14 we reject the null and conclude that the regression is significant.

c) Write down the estimated multiple regression equation for married men who are heads of households. Let M=1, HEAD=1
Y = 10 + 20*AGE – 10*EDUC + 25*1 + 20*1 +10*TENURE + 8.5*MGT + 6.5*SALES

Y = 55 + 20*AGE – 10*EDUC +10*TENURE + 8.5*MGT + 6.5*SALES

(d) Hypothesis:

$H_0 : \beta_i = 0$, $H_1 : \beta_i \neq 0$ ; $i = 1,2,\ldots,7$

| *Predictor* | *Coefficient* | *Standard Error* | *t-test* |
|---|---|---|---|
| *Constant* | *10* | *2* | *5* |
| *AGE* | *20* | *5* | *4* |
| *EDUC* | *-10* | *4* | *-2.5* |
| *MARRIED* | *25* | *10* | *2.5* |
| *HEAD* | *20* | *10* | *2* |
| *TENURE* | *10* | *8* | *1.25* |
| *MGT* | *8.5* | *2.5* | *3.4* |
| *SALES* | *6.5* | *3.5* | *1.857* |

*Do each t test=bi/sbi*

*For #1, t test=20/5 = 4…all completed in table above*
$-2.021 < t < 2.021$,

*With a significance level of 5%, from the table we will fail to reject the null hypothesis if –2.021 < t < 2.021. For the variables above we reject all hypotheses except for HEAD, TENURE, and SALES.*

**I5.**a) Student with higher average grade in high school math tends to have higher expected GPA. The marginal increase is 0.146.

b) K=5 variables n=224, n-k-1=224-5-1=218 (use 120)

$R^2 = SSR / SST = 28.44 / 135.46 \approx 0.2099$
About 21% of the variation in GPA can be explained by the model.

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
Ha at least one of them is not equal to zero.

$$F = \frac{(SSR)/k}{SSE/(n-k-1)} = \frac{28.44/5}{106.82/(224-5-1)} = \frac{5.688}{0.49} \approx 11.608 >$$
> F 5,120,0.05=2.29

So, $R^2$ is significantly greater than zero

c) $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 > 0$

n-k-1=224-5-1=218 (use 200)

t test=$\frac{b_3}{s_{b_3}} = \frac{0.055}{0.02} = 2.75 > t_{0.05,200} = 1.653$ reject H0

So, $\beta_3$ is significantly greater than zero.

Or use p-value, 2.601<2.75<3.131
0.001<1-sided p-value<0.005
p-value<0.05, so we reject H0

## J. Dummy Variables

**Example 1.**
a) y=20+0.2E, slope is positive
b) 20 is the y-intercept
c) y=19.2+0.3E, slope is positive
d) y intercept of men's equation is 19.2
e) y=20+0.2(20)=24
f) 23.7=19.2 + 0.3 E
4.5=0.3E
E=15

J2. Wins=71.87 +0.101 Payroll - 0.06 League
a) substitute P=1 in for payroll
Wins=71.87 +0.101 Payroll - 0.06 League
Team A would get 0.101(1) more wins than team B.

The answer is A).

b) Team A in National, so league=0 and team B is American, so league=1
Wins = 71.87 - 0.06(1)...so B loses 0.06 more games than team A, so team A wins 0.06 more games than team B.

The answer is C).

c) Slope for American ie. League=1
Wins=71.87 + 0.101P - 0.06(1)...slope = 0.101

The answer is D).

d) Intercept for American
Wins=71.87 + 0.101P - 0.06
71.81 + 0.101P...intercept is 71.81

The answer is E).

e) National League=0
$98 million (in millions) so subst. 98 for payroll
Wins=71.87 + 0.101(98) - 0.06(0)=81.768...x 1 000 000
The answer is E).

f) American League $108
Won 88
Prediction
Wins=71.87 + 0.101P - 0.06(1)
W=71.87+0.101(108)-0.06=82.718

Residual=Observed - Predicted=88-82.718=5.282

The answer is B).

J3.a) Southern, so let S=1
C= -321.9 +4.69 U + 39.3 P – 649.3(1)+12.1(1)-5.84 P(1)
=-971.2 + 16.79 U + 33.46 P
The answer is D).

b) C=-971.2 + 16.79 (55.4) + 33.46(13.7)=417.4
The answer is A).

c) Non-Southern, so let S=0
C= -321.9 + 4.69(65.6) +39.3(8)
=300.2
The answer is A).

d) F= MSR/MSE=412091/19604=21.02
The answer is D).

e) F 5, 45 df
The answer is D).

f) At least one of them (all but poverty south)
The answer is D).

g) Poverty south=0.728 > 10%, so remove it
The answer is E).
Prediction
Wins=71.87 + 0.101P - 0.06(1)
W=71.87+0.101(108)-0.06=82.718
Residual=Observed - Predicted=88-82.718=5.282

## K. Analysis of Categorical Data

**Example 1.** $E = np$

| Motor vehicle | Falls | Drowning | Fire | Poison | Other |
|---|---|---|---|---|---|
| 45% | 15% | 4% | 3% | 16% | 17% |
| 0.45×990=445.5 | 0.15× 990= 148.5 | 0.04× 990=39.6 | 0.03×990= 29.7 | 0.16×990= 158.4 | 0.17×990= 168.3 |

Total of all accidental deaths = 990

$$k = 6\ groups \qquad \alpha = 0.05$$

$H_0\ P_1 = 0.45\ \ P_2 = 0.15\ \ P_3 = 0.04\ \ P_4 = 0.03\ \ P_5 = 0.16\ \ P_6 = 0.17$

$H_a\ \ at\ least\ one\ Pi \neq\ H_o\ value$

$x^2 crit = {x^2}_{5,0.05} = 11.0705$

DR reject $H_o\ if\ x^2 test\ > 11.0705$

$$x^2{}_{test} = \frac{[n_i - E(n_i)]^2}{E(n_i)\ pred} = \frac{(442-445.5)^2}{445.5} + \frac{(161-148.5)^2}{148.5} + \frac{(42-39.6)^2}{39.6} + \frac{(33-29.7)^2}{29.7} + \frac{(162-158.4)^2}{158.4} + \frac{(150-168.3)^2}{168.3}$$
$$= 3.66 < x^2 crit \quad \therefore do\ not\ reject\ H_o$$

If you are only asked for the first term, it would be: $\frac{(442-445.5)^2}{445.5}$=0.0275

**Example 2.**

$$C = 2 \quad R = 2 \quad E = np$$

| | Sidney Crosby | Alexander Ovechkin | |
|---|---|---|---|
| High psychopathy score | 12 | 14 | 26 |
| Low psychopathy score | 18 | 15 | 33 |
| | 30 | 29 | |

N=59

$H_o$ $no\ significant\ association$ $H_a$ $there\ is\ a\ significant\ association$

$x^2{}_{crit} = x^2{}_{(R-1)(C-1),2}x^2{}_{1,0.05} = 3.84146$

$DR\ reject\ H_o\ if\ x^2{}_{obt} > 3.84146$

| | |
|---|---|
| $E_{1,1}$ $\frac{26(30)}{59} = 13.22$ | $E_{1,2}$ $\frac{26(29)}{59} = 12.78$ |
| $E_{2,1}$ $\frac{33(30)}{59} = 16.78$ | $E_{2,2,}$ $\frac{33(29)}{59} = 16.22$ |
| | |

$$x^2{}_{test} = \sum \frac{(obs-pred)^2}{pred} = \frac{(12-13.22)^2}{13.22} + \frac{(14-12.78)^2}{12.78} + \frac{(18-16.78)^2}{16.78} + \frac{(15-16.22)^2}{16.22} = 0.41 < 3.84146$$

$\therefore do\ not\ reject\ H_o$ $\therefore no\ significant\ association$

If you are only asked for the first term, it would be: $\frac{(12-13.22)^2}{13.22}$=0.1126

**K1.** $(R-1)(C-1) = 1 \quad C = 2 \quad R = 2 \quad N = 37$

$x^2crit = x^2{}_{1,0.10} = 2.70554$

$DR\ reject\ H_o\ \ if\ x^2obt > 2.70554$

$H_o\ \ no\ significant\ relationship$

$H_a\ \ there\ is\ a\ significant\ relationship\ between\ gender\ \&\ age$

| | |
|---|---|
| $E1,1$ <br> $\frac{26(20)}{37} = 14.05$ | $E1,2$ <br> $\frac{26(17)}{37} = 11.95$ |
| $E2,1$ <br> $\frac{11(20)}{37} = 5.95$ | $E2,2$ <br> $\frac{11(17)}{37} = 5.05$ |

$x^2test = \frac{(obs-pred)^2}{pred} = \frac{(17-14.05)^2}{14.05} + \frac{(9-11.95)^2}{11.95} + \frac{(3-5.95)^2}{5.95}$

$+\frac{(8-5.05)^2}{5.05} \ = 4.53 \ < 2.70554$

$\therefore reject\ H_o \quad \therefore there\ is\ a\ significant\ relationship\ between\ age\ \&$
$gender\ \ ie.\ they\ are\ dependent$

**K2.** $\alpha = 0.05 \quad C = 2 \quad R = 2$

Leadership Qualities

| | Leaders | Followers | |
|---|---|---|---|
| Small | 14 | 9 | 23 |
| Large | 6 | 15 | 21 |
| | 20 | 24 | N=44 |

$x^2crit = x^2{}_{(R-1)(C-1),\alpha} = x^2{}_{1,0.05} = 3.84146$

$DR\ reject\ H_o\ \ if x^2test > 3.84146$

$H_o\ no\ significant\ relationship\ or\ association$

$H_a\ \ there\ is\ a\ significant\ association$

| | |
|---|---|
| $E1,1$ <br> $\frac{23(20)}{44} = 10.45$ | $E1,2$ <br> $\frac{23(24)}{44} = 12.54$ |
| $E2,1$ <br> $\frac{21(20)}{44} = 9.54$ | $E2,2$ <br> $\frac{21(24)}{44} = 11.45$ |

$x^2test = \frac{(obs-pred)^2}{pred} = \frac{(14-10.45)^2}{10.45} + \frac{(9-12.54)^2}{12.54} + \frac{(6-9.54)^2}{9.54} + \frac{(15-11.45)^2}{11.45}$

$= 4.62 \ > 3.84146$

$\therefore reject\ H_o \quad \therefore there\ is\ a\ significant\ association$

**K3.** $k = 4 \quad \frac{1600}{4} = 400 \quad let\ \alpha = 0.05$

| | Observed | $Ei = expected\ (or\ predicted)$ |
|---|---|---|
| Spades | 405 | 400 |
| Hearts | 419 | 400 |
| Diamonds | 401 | 400 |
| Clubs | 375 | 400 |

$H_o\ \ p_1 = p_2 = p_3 = p_4 = 0.25$

$H_a\ \ at\ least\ one\ p_i \neq 0.25$

$x^2crit = x^2{}_{k-1,\alpha} = x^2{}_{3,0.05} = 7.81473$

$DR\ reject\ H_o\ if\ x^2obt > 7.81473$

$x^2test = \frac{(obs-pred)^2}{pred} = \frac{(405-400)^2}{400} + \frac{(419-400)^2}{400} + \frac{(401-400)^2}{400} + \frac{(375-400)^2}{400} = 2.53$

$x^2test < 7.81473 \quad \therefore do\ not\ reject\ H_o \quad \therefore all\ p_i = 0.25$

**K4.** $let\ \alpha = 0.05 \quad k = 4\ groups \quad N = 100$

$E = np = 100(0.25) = 25$

$H_o\ \ p_1 = p_2 = p_3 = p_4 = 0.25$

$H_a\ \ reject\ H_o\ \ at\ least\ one\ p_i \neq 0.25$

$x^2crit = x^2{}_{k-1,\alpha} = x^2{}_{3,0.05} = 7.81473$

$DR\ reject\ H_o\ if\ x^2obt > 7.81473$

$x^2test = \frac{(obs-pred)^2}{pred} = \frac{(40-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(15-25)^2}{25}$

$= 14 > 7.81473$

$\therefore reject\ H_o \quad \therefore at\ least\ one\ p_i \neq 0.25$

**K5**. $x^2crit = x^2{}_{(R-1)(c-1),\alpha} = x^2{}_{4,0.05} = 9.487$

$DR\ reject\ H_o\ if\ x^2test > 9.487$

$H_o\ \ no\ significant\ association$

$H_a\ \ there\ is\ a\ significant\ association$

R=3, C=3 and N=200

Row totals 60,70,71 column totals 40,60,101 and grand total=201

| | | |
|---|---|---|
| $E11$<br>$\frac{60(40)}{201} = 11.94$ | $E12$<br>$\frac{60(60)}{201} = 17.9$ | $E13$<br>$\frac{60(101)}{201} = 30.15$ |
| $E21$<br>$\frac{70(40)}{201} = 13.93$ | $E22$<br>$\frac{70(60)}{201} = 20.9$ | $E23$<br>$\frac{70(101)}{201} = 35.17$ |
| $E31$<br>$\frac{71(40)}{201} = 14.1$ | $E32$<br>$\frac{71(60)}{201} = 21.19$ | $E33$<br>$\frac{71(101)}{201} = 35.68$ |

$x^2test = \frac{(16-11.94)^2}{12} + \frac{(23-17.9)^2}{18} + \cdots + \frac{(21-30.15)^2 + \cdots}{30} = 26.4$

$x^2test > 9.487 \quad \therefore reject\ H_o$

and therefore, the is a significant association

**K6.** $R-1=1 \quad c-1=2 \quad c=3 \quad R=2$

| | 18 to 35 | 35 to 59 | 60+ | |
|---|---|---|---|---|
| Party A | 85 | 95 | 131 | 311 |
| Party B | 168 | 197 | 173 | 538 |
| | 253 | 292 | 304 | 849=N |

$H_o$ *no significant relationship or association*

$H_a$ *there is a significant association*

| | | |
|---|---|---|
| $E1{,}1$ $\frac{311(253)}{849}=92.68$ | $E1{,}2$ $\frac{311(292)}{849}=106.96$ | $E1{,}3$ $\frac{311(304)}{849}=111.36$ |
| $E2{,}1$ $\frac{538(253)}{849}=160.32$ | $E2{,}2$ $\frac{538(292)}{849}=185.04$ | $E2{,}3$ $\frac{538(304)}{849}=192.64$ |

$x^2{}_{crit}=x^2{}_{(R-1)(c-1),\alpha}=x^2{}_{2,0.05}=5.99147$

$DR\ reject\ H_o\ if\ x^2{}_{test}>5.99147$

$x^2{}_{test}=\frac{(85-92.68)^2}{92.68}+\frac{(95-106.96)^2}{106.96}+\cdots+\frac{(173-192.64)^2}{192.64}$

$=8.6>5.99147$

$\therefore reject\ H_o$

$\therefore$ *age and party are associated (not independent)*

**K7.**

$H_o\ \ p_1=0.40\ \ p_2=0.30\ \ p_3=0.20\ p_4=0.10$

$H_a$ *reject* $H_o$ *at least one* $p_i\neq H_o$ *value*

| | obs | pred |
|---|---|---|
| Type | Number of People | |
| Connor | 190 | $0.4\times450=180$ |
| Nathan | 110 | $0.3\times450=135$ |
| Leon | 90 | $0.2\times450=90$ |
| Austin | 60 | $0.10\times450=45$ |
| | 450 | 450 |

$k=4$

a) $x^2crit=x^2{}_{k-1,\alpha}=x^2{}_{3,0.01}=11.3449$

$DR\ reject\ H_o\ \ if\ x^2test>11.3449$

$x^2test=\frac{(obs-pred)^2}{pred}=\frac{(190-180)^2}{180}+\frac{(110-135)^2}{135}+\frac{(90-90)^2}{90}+\frac{(60-45)^2}{45}$

$=10.2<11.3449$

$\therefore$ *DO NOT reject* $H_o$ $\quad\therefore$ *All of the pi are equal to its* $H_o$ *value* b)

| Region | McDavid | MacKinnon | Draisaitl | Matthews | |
|---|---|---|---|---|---|
| A | 110 | 25 | 60 | 20 | 215 |
| B | 50 | 45 | 30 | 10 | 135 |
| C | 40 | 30 | 30 | 25 | 125 |
| | 200 | 100 | 120 | 55 | N=475 |

$$C = 4 \qquad R = 3 \qquad (R-1)(C-1) = 2(3) = 6$$
$$x^2crit = x^2{}_{(R-1)(c-1),\alpha} = x^2{}_{6,0.05} = 12.5916$$

$DR\ reject\ H_o\ if\ x^2test > 12.5916$

$H_o$ $no\ significant\ difference$
$H_a$ $there\ is\ a\ significant\ difference$

$x^2test = 39.11$

$x^2test > 12.5916 \quad \therefore reject\ H_o$

$\therefore there\ is\ a\ significant\ difference$

**K8.** k=4

| 4 (or more) Hobbies | 3 Hobbies | 2 Hobbies | 0 or 1 Hobbies |
|---|---|---|---|
| 65 | 57 | 45 | 33 |

$E = np \quad 200(0.45) = 90$

$200(0.3) = 60$

$200(0.15) = 30$

$200(0.10) = 20$

$p_1 = 0.45 \quad p_2 = 0.30 \quad p_3 = 0.15 \quad p_4 = 0.10$

**a)** $H_o \quad p_1 = 0.45\, p_2 = 0.30\, p_3 = 0.15 \;\; p_4 = 0.10$

$H_a \quad H_o \; false \quad \therefore at\ least\ one\ pi \neq it's\ H_o\ value$

$x^2{}_{crit} = x^2{}_{k-1,\alpha} = x^2{}_{3,0.05} = 7.81473$

$DR\ reject\ H_o\ if\ x^2{}_{test} > 7.81473$

$x^2{}_{test} = \frac{(obs-pred)^2}{pred} = \frac{(65-90)^2}{90} + \frac{(57-60)^2}{60} + \frac{(45-30)^2}{30} + \frac{(33-20)^2}{20}$

$x^2 = 23.04 > 7.81473$

$\therefore reject\ H_o \quad \therefore at\ least\ one\ p_i\ different$

**b)** C=4 R=2

| | 4(or more) TVs | 3 TVs | 2 TVs | 0 or 1 TV | |
|---|---|---|---|---|---|
| Over 45 | 35 | 29 | 9 | 6 | 79 |
| Under 50 | 44 | 23 | 14 | 31 | 112 |
| | 79 | 52 | 23 | 37 | 191 |

$H_o$ $no\ significant\ relationship$
$H_a$ $there\ is\ a\ significant\ relationship$

$x^2{}_{crit} = x^2{}_{(R-1)(C-1),\alpha} = x^2{}_{(1)(3),0.05} = 7.81473$
$DR\ \ reject\ H_o\ \ if\ x^2{}_{test} > 7.81473$

| | | | |
|---|---|---|---|
| $E11$ $\frac{79(79))}{191} = 32.68$ | $E12$ $\frac{79(52)}{191} = 21.5$ | $E13$ $\frac{79(23)}{191} = 9.51$ | E1 $\frac{79(37)}{191} = 15.3$ |
| $E21$ $\frac{112(79)}{191} = 46.32$ | $E22$ $\frac{112(52)}{191} = 30.49$ | $E23$ $\frac{112(23)}{191} = 13.49$ | E24 $\frac{112(37)}{191} = 21.7$ |

$x^2 test = \frac{(35-32.68)^2}{32.68} + \frac{(29-21.5)^2}{21.5} + \cdots + \frac{(31-21.7)^2 + \cdots}{21.7} = 14.42$
$x^2 test > 9.487 \quad \therefore reject\ H_o$

Over 45, 0 or 1 = $\frac{(6-15.3)^2}{15.3}$=5.65

**Best of luck on the exam!!!!**