Tutorial Letter 203/1/2015

Applied Statistics II

STA2601

Semester 1

Department of Statistics

Solutions to Assignment 3

BAR CODE

Learn without limits.

UNISA university of south africa

QUESTION 1

(a) $n = 20 \qquad \overline{X} = 6.003 \qquad s = 0.017$

The $100(1-\alpha)\%$ confidence interval for the unknown variance σ^2 is

$$\left(\frac{\sum \left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;\ n-1}}; \ \frac{\sum \left(X_i - \overline{X}\right)^2}{X^2_{1-\frac{1}{2}\alpha;\ n-1}} \right)$$

$$\begin{aligned}
(n-1)\, s^2 &= \sum \left(X_i - \overline{X}\right)^2 \\
19 \times 0.017^2 &= \sum \left(X_i - \overline{X}\right)^2 \\
0.005491 &= \sum \left(X_i - \overline{X}\right)^2
\end{aligned}$$

$$\alpha = 0.01 \qquad \alpha/2 = 0.005$$

$$\begin{aligned}
\chi^2_{\frac{\alpha}{2}}; n-1 &= \chi^2_{0.005;19} = 38.5822 \\
\chi^2_{1-\frac{\alpha}{2};\ n-1} &= \chi^2_{0.995;19} = 6.84398
\end{aligned}$$

Now the 99% confidence interval for the variance of the diameter of bearings is

$$\begin{aligned}
\frac{\sum \left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;\ n-1}} &< \sigma^2 < \frac{\sum \left(X_i - \overline{X}\right)^2}{X^2_{1-\frac{1}{2};\ n-1}} \\
\frac{0.005491}{38.5822} &< \sigma^2 < \frac{0.005491}{6.84398} \\
0.000142319 &< \sigma^2 < 0.000802331
\end{aligned}$$

Thus the 99% confidence interval for σ is

$$\begin{aligned}
\sqrt{0.000142319} &< \sigma < \sqrt{0.00080231} \\
0.0119 &< \sigma < 0.0283
\end{aligned}$$

(5)

(b) Let first sample be males and second sample be females.

$n_m = 11 \qquad \sigma_m^2 = 12$

$n_f = 13 \qquad \sigma_f^2 = 4$

$H_0: \ \sigma_m^2 = \sigma_f^2 \qquad\qquad H_1: \sigma_m^2 > \sigma_f^2$

The test statistic is

$$\begin{aligned} F &= \frac{\sigma_f^2}{\sigma_m^2} \times \frac{S_m^2}{S_f^2} \\ &= 1 \times \frac{12}{3} \\ &= 3 \end{aligned}$$

Test is one-tailed. The critical value is $F_{\alpha;n_m-1;n_f-1} = F_{0.05;10;12} \approx 2.75$ Reject H_0 if $F > 2.75$.

Since $3 > 2.75$, we reject H_0 at the 5% level of significance and conclude that $\sigma_m^2 > \sigma_f^2$, that is, the reaction times of males are more variable than the reaction times of females.

(9)

[14]

QUESTION 2

(a) Based on the assumption of **independent observations** and the assumption that the weights have a **normal distribution** (i.e. that the sample comes from a normal population) we may assume that

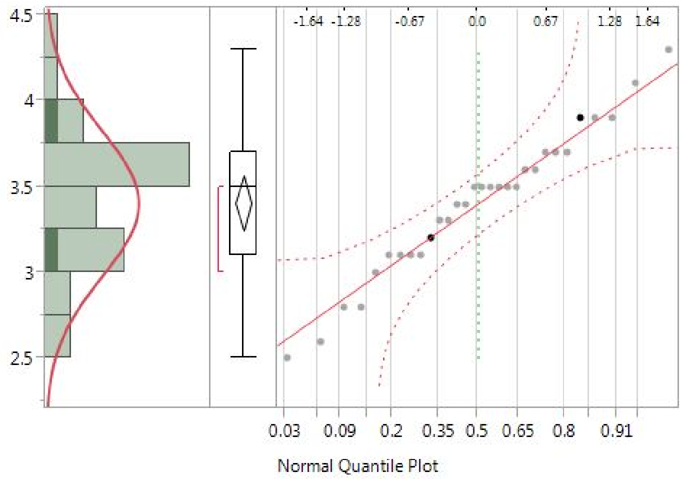
$$T = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{S} \sim t_{n-1}.$$

(2)

(b)

Distributions

Weight



Normal(3.40333,0.42384)

Quantiles

100.0%	maximum	4.3
99.5%		4.3
97.5%		4.3
90.0%		3.9
75.0%	quartile	3.7
50.0%	median	3.5
25.0%	quartile	3.1
10.0%		2.8
2.5%		2.5
0.5%		2.5
0.0%	minimum	2.5

Summary Statistics

Mean	3.4033333
Std Dev	0.4238439
Std Err Mean	0.077383
Upper 95% Mea	3.5615993
Lower 95% Mean	3.2450674
N	30

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	3.4033333	3.2450674	3.5615993
Dispersion	σ	0.4238439	0.3375522	0.5697798

-2log(Likelihood) = 32.6329133184445

Figure 1: Testing for Normality

Are they met? If we assume that the weight of one baby (no multiple births) does not influence the other baby, **independent observations are OK.**

Maybe, the normality assumption is slightly violated because from the JMP graphical output we see that the normal curve does not fit the histogram very well, the box plot has a long tail to the left suggesting negative skewness and there also seems to be a slight systematic deviation around the line in the Normal Quantile Plot especially on the lower end (It is subjective). Luckily the test is not too sensitive and we may proceed. (7)

(c) We have to test $H_0 : \mu = 3.6$ against $H_0 : \mu < 3.6$.

<u>**Method I: Using the critical value approach**</u>:

$n = 30 \qquad \sum X = 102.1 \qquad \sum X^2 = 352.69$

$\overline{X} = \dfrac{\sum X}{n} = \dfrac{102.1}{30} \approx 3.4033.$

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right) \\
&= \frac{1}{30-1}\left(352.69 - \frac{(102.1)^2}{30}\right) \\
&= \frac{1}{29}\,(352.69 - 347.4803333) \\
&= \frac{1}{29}\,(5.209666667) \\
\therefore S &= \sqrt{0.179643678} \\
&\approx 0.4238
\end{aligned}$$

$$\begin{aligned}
t_{calc} &= \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{S} \\
&= \frac{\sqrt{30}\,(3.4033 - 3.6)}{0.4238} \\
&= \frac{\sqrt{30}\,(-0.1967)}{0.4238} \\
&= \frac{-1.077370271}{0.4238} \\
&\approx -2.5422
\end{aligned}$$

Test is one-tailed. $\alpha = 0.05$. The critical value is $t_{\alpha;n-1} = t_{0.05;29} = -1.699$. Reject H_0 if t_{calc} is less than -1.699.

Since $-2.5422 < -1.699$, we reject H_0 at the 5% level of significance and conclude that the mean weight of a full time baby is significantly less than 3.6kg.

(8)

(d) If we know that $\sigma = 0.5$ we will use the test statistic

$$Z = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} \sim n\left(0; 1\right).$$

For this specific sample, it becomes

$$\begin{aligned} Z &= \frac{\sqrt{30}\left(3.4033 - 3.6\right)}{0.5} \\ &= \frac{-1.077370271}{0.5} \\ &\approx 2.1547. \end{aligned}$$

The critical value is $z_\alpha = z_{0.05} = 1.645$. Reject H_0 if z_{calc} is less than -1.645.

Since $-2.1547 < -1.645$, we reject H_0 at the 5% level of significance and conclude that the mean weight of full time baby is significantly less than 3.6kg.

(6)

(e) A 90% two-sided confidence interval is computed as $\overline{X} - \left(\frac{S}{\sqrt{n}}\right)\left(t_{0.05;29}\right) < \mu < \overline{X} + \left(\frac{S}{\sqrt{n}}\right)\left(t_{0.05;29}\right)$

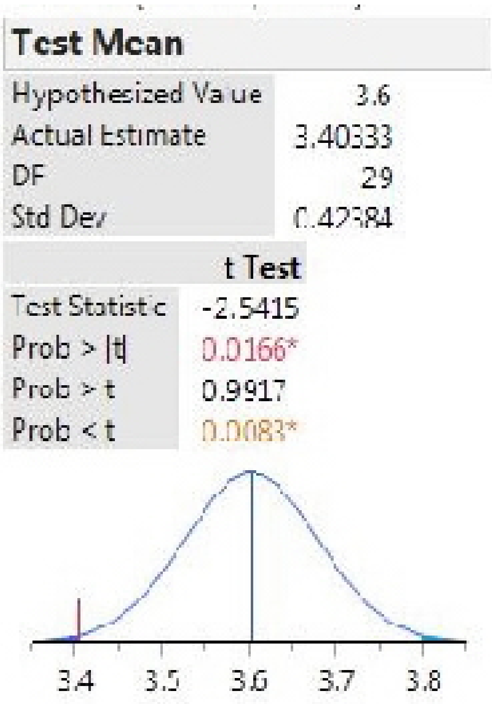
where

$$\begin{aligned} 3.4033 - \left(\frac{0.4238}{\sqrt{30}}\right)(1.699) &< \mu < 3.4033 + \left(\frac{0.4238}{\sqrt{30}}\right)(1.699) \\ 3.4033 - (0.077374939)(1.699) &< \mu < 3.4033 + (0.077374939)(1.699) \\ 3.4033 - 0.1315 &< \mu < 3.4033 + 0.1315 \\ 3.2718 &\leq \mu \leq 3.5348 \end{aligned}$$

Since this lower bound (at the 90% level) will be the same as the 95% one-sided interval we may say we are 95% confident that $\mu \leq 3.5348$ (This means we reject $H_0 : \mu = 3.6$ which confirms our conclusion.)

(5)

(f) For the test in (b) we see the p-value is $0.0083 = P(t \leq -2.5415)$:



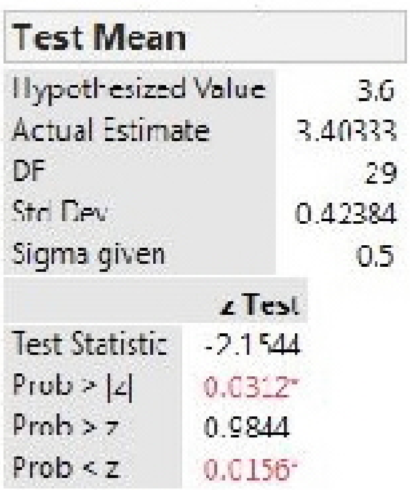
Test Mean

Hypothesized Value	3.6
Actual Estimate	3.40333
DF	29
Std Dev	0.42384

	t Test
Test Statistic	-2.5415
Prob > |t|	0.0166*
Prob > t	0.9917
Prob < t	0.0083*

Figure 2: The t-test

For the test in (c) we see the p-value is $0.0156 = P(z \leq -2.1544)$:



Test Mean

Hypothesized Value	3.6
Actual Estimate	3.40333
DF	29
Std Dev	0.42384
Sigma given	0.5

	z Test
Test Statistic	-2.1544
Prob > |z|	0.0312*
Prob > z	0.9844
Prob < z	0.0156*

Figure 3: The z-test

We reject H_0 in both cases. (15)

(g) We want to test:

$H_0 : \sigma^2 = 0.25$ against $H_1 : \sigma^2 \neq 0.25$

Test Standard Deviation

Hypothesized Value	0.5
Actual Estimate	0.42384
DF	29

Test	ChiSquare
Test Statistic	20.83867
Min PValue	0.2701
Prob < ChiSq	0.1350
Prob > ChiSq	0.8650

Figure 3: Testing for Standard Deviation

<u>Method I: Using the critical value approach</u>:

Assuming μ is unknown, i.e., $\widehat{\mu} = \overline{X}$, then the test statistic is

$$U = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{\sigma^2} = \frac{(n-1)\,s^2}{\sigma^2} \approx 20.83867$$

$$\alpha = 0.05 \qquad \alpha/2 = 0.025$$

$$\begin{aligned} \chi^2_{\frac{\alpha}{2};n-1} &= \chi^2_{0.025;29} = 45.7222 \\ \chi^2_{1-\frac{\alpha}{2};\ n-1} &= \chi^2_{0.975;219} = 16.0471 \end{aligned}$$

Reject H_0 if $U < 16.0471$ or $U > 45.7222$.

Since $16.0471 < 20.83867 < 45.7222$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 0.5$.

<u>Method II: Using the p-value approach</u>

p-value $= 0.2701$. Since $0.2701 > 0.05$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 0.5$. (4)

(h) $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$

$n_x = 30 \quad \overline{X} = 3.4033 \quad S_X^2 = 0.1796$

$n_y = 24 \quad \overline{Y} = 3.5125 \quad S_Y^2 = 0.1803$

The test statistic is

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{nx} + \frac{1}{ny}}}$$

Now

$$\begin{aligned}
S_p^2 &= \frac{(n_x - 1)S_X^2 + (n_y - 1)S_Y^2}{n_1 + n_2 - 2} \\
&= \frac{(30 - 1)0.1796 + (24 - 1)0.1803}{30 + 24 - 2} \\
&= \frac{(29)0.1796 + (23)0.1803}{30 + 24 - 2} \\
&= \frac{5.2084 + 4.1469}{52} \\
&= \frac{9.3553}{52} \\
&\approx 0.179909615 \\
\Longrightarrow & \quad S_{pooled} = \sqrt{0.179909615} \approx 0.4242
\end{aligned}$$

Then

$$\begin{aligned}
T &= \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \\
&= \frac{(3.4033 - 3.5125) - (0)}{0.4242\sqrt{\frac{1}{30} + \frac{1}{24}}} \\
&= \frac{-0.1092}{0.4242\sqrt{0.075}} \\
&= \frac{-0.1092}{0.116171954} \\
&\approx -0.9400
\end{aligned}$$

The critical value is

$$\begin{aligned} t_{\alpha;n_1+n_2-2} &= t_{0.05;52} \\ &= 1.684 + \frac{12}{20}(1.671 - 1.684) \\ &= 1.684 + 0.6(-0.013) \\ &= 1.684 - 0.0078 \\ &\approx 1.676 \end{aligned}$$

Reject H_0 if $T < -1.676$.

Since $-0.94 > -1.676$, we do not reject H_0 at the 5% level and conclude that the means are not significantly different from each other, i.e., $\mu_X = \mu_Y$. (8)

[55]

QUESTION 3

(a) We are testing $H_0 : \mu = 30$ against $H_1 : \mu \neq 30$.

The power of the test is a function of Φ which is defined as $\Phi = \dfrac{\delta}{\sqrt{2}}$

$$\begin{aligned} \delta &= \frac{\sqrt{n}(\mu - \mu_0)}{\sigma} \\ &= \frac{\sqrt{10}(30 + \sqrt{2}\,\sigma - 30)}{\sigma} \\ &= \sqrt{10}\sqrt{2} \\ \Longrightarrow \Phi = \frac{\delta}{\sqrt{2}} &= \frac{\sqrt{10}\sqrt{2}}{\sqrt{2}} \\ &\approx 3.1623 \end{aligned}$$

F**rom table F** we read of the power as 98% (i.e., $1 - \beta = 0.98$) (4)

(b) We have to test $H_0 : \mu = 0$ against $H_0 : \mu < 0$.

Method 1: Using the critical value approach

From the output the test statistics is

$$t = \frac{(\overline{x} - \mu_0)}{\frac{s}{\sqrt{n}}} = \frac{(-1.4875 - 0)}{0.52727} \approx -2.82114$$

The critical value is $t_{\alpha;n-1} = t_{0.05;7} = 1.895$. Reject H_0 if $T \leq -1.895$

Since $-2.82114 < -1.895$, we reject H_0 in favour of H_1 at the 5% level of significance and conclude that $\mu < 0$.

Method II: Using the p-value approach

p-value $= 0.0129$. Since $0.0129 < 0.05$, we can reject H_0 in favour of H_1 at the 5% level of significance and conclude that $\mu < 0$. (6)

[10]

QUESTION 4

Group	A	B	C	D
n	5	5	5	5
$\sum X_{ij}$	1 311	1 174	1 258	1 343
$\overline{X}_i$	262.2	234.8	251.6	268.6
$\sum \left(X_{ij} - \overline{X}_i\right)^2$	154.8	154.8	141.2	209.2

(a)

$$\begin{aligned} S_1^2 &= \frac{1}{n_1 - 1} \sum \left(X_{1j} - \overline{X}_1\right)^2 \\ &= \frac{1}{5-1}(154.8) \\ &= \frac{1}{4}(154.8) \\ &= 38.7 \end{aligned} \qquad \begin{aligned} S_2^2 &= \frac{1}{n_2 - 1} \sum \left(X_{2j} - \overline{X}_2\right)^2 \\ &= \frac{1}{5-1}(154.8) \\ &= \frac{1}{4}(154.8) \\ &= 38.7 \end{aligned}$$

$$\begin{aligned} S_3^2 &= \frac{1}{n_3 - 1} \sum \left(X_{3j} - \overline{X}_3\right)^2 \\ &= \frac{1}{5-1}(141.2) \\ &= \frac{1}{4}(141.2) \\ &= 35.3 \end{aligned} \qquad \begin{aligned} S_4^2 &= \frac{1}{n_4 - 1} \sum \left(X_{4j} - \overline{X}_4\right)^2 \\ &= \frac{1}{5-1}(209.2) \\ &= \frac{1}{4}(209.2) \\ &= 52.3 \end{aligned}$$

From the computations above it, follows that $S_1^2 = 38.7$; $S_2^2 = 38.7$; $S_3^2 = 35.3$ and $S_4^2 = 52.3$.

(4)

(b) (i) Ordinary average = $\dfrac{38.7+38.7+35.3+52.3}{4} = \dfrac{165}{4} = 41.25.$ (2)

(ii) $MSE = \dfrac{SSE}{kn-k}.$

For this $ANOVA$ problem, we have $k = 4$ (there are four groups) and $n = 5$ (the number of observations in each sample).

$$\begin{aligned} SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_i\right)^2 \\ &= 154.8+154.8+141.2+209.2 \\ &= 660 \end{aligned}$$

$$\begin{aligned} \therefore MSE &= \frac{660}{4\,(5)-4} \\ &= \frac{660}{16} \\ &= 41.25. \end{aligned}$$

The result in (i) = result in (ii).

This makes perfect sense! MSE is like a pooled variance or an average variance, because the assumption of $ANOVA$ is that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ and if these variances are unknown, we estimate it by pooling. (2)

(c) It is reasonable to assume that the four samples are **independent**. The outcome of one state can not influence the outcome of the other state.

The other assumption (of equal variances) can be formally tested!

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

$H_1 : \sigma_p^2 \neq \sigma_q^2$ for at least one $p \neq q$.

$$\begin{aligned} U &= \frac{\max S_i^2}{\min S_i^2} \\ &= \frac{52.3}{35.3} \\ &\approx 1.4816 \end{aligned}$$

From **Table E** with $k = 4$ and $\nu = n-1 = 5-1 = 4$, we find that the critical value is 20.6. Reject H_0 if $U > 20.6$.

Since $1.4816 < 20.6$, we cannot reject H_0 at the 5% level of significance and we may assume that the variances are equal.

(If we use the JMP computer output we may also assume that the variances are equal because "Prob $>$ F" is not significant for all the four tests under the heading: "**Tests that the Variances are Equal**".)

(6)

(d) We have to test:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against

$H_1 : \mu_p \neq \mu_q$ for at least one $p \neq q$.

The test statistic is $F = \dfrac{MST_r}{MSE} \sim F_{k-1;kn-k}$

$$\begin{aligned}
MST_r &= \frac{n\sum_{i=1}^{k}\left(\overline{X}_i - \overline{X}\right)^2}{k-1} \\
\text{where } \overline{X} &= \frac{\sum\sum X_{ij}}{N} = \frac{5086}{20} = 254.3 \qquad \text{(overall mean);} \\
\text{and } \sum\left(\overline{X}_i - \overline{X}\right)^2 &= (262.2-254.3)^2 + \cdots + (268.6-254.3)^2 \\
&= (7.9)^2 + (-19.5)^2 + (-2.7)^2 + (14.3)^2 \\
&= 62.41 + 380.25 + 7.29 + 204.49 \\
&= 654.44 \\
\therefore MST_r &= \frac{5\,(654.44)}{4-1} = \frac{3\,272.2}{3} = 1\,090.7333
\end{aligned}$$

We already know that $MSE = 41.2500$ (see question (b)(ii)).

$$\begin{aligned}
\therefore F &= \frac{MST_r}{MSE} \\
&= \frac{1\,090.7333}{41.25} \\
&\approx 26.442.
\end{aligned}$$

(Note that these computations are the same with the JMP output under the heading: "**Analysis of Variance**".)

The critical value is $F_{0,05;3;16} = 3.24$. Reject H_0 if $F > 3.24$.

Since $26.442 > 3.24$, we reject H_0 at the 5% level of significance and conclude that the population mean prices of the four states differ, that is, $\mu_p \neq \mu_q$ for at least one $p \neq q$.

(Note that we reach the same conclusion with the JMP output under the heading: "**Analysis of Variance**" if we consider "Prob $>$ F"< 0.0001)

(8)

(e) For each pair of means, we compute a test statistic

$$T_{pq} = \frac{\overline{X}_p - \overline{X}_q}{S_{pooled}\sqrt{1/n + 1/n}} = \frac{\sqrt{n}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}S} = \frac{\sqrt{5}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}\sqrt{MSE}}.$$

We reject $H_0(p;q)$ if

$$\left|T_{pq}\right| > \sqrt{(k-1)\,F_{\alpha;k-1;kn-k}} = \sqrt{3\,(3.24)} \approx 3.1177$$

This implies that we reject H_0 if

$$\frac{\sqrt{5}\left|\overline{X}_p - \overline{X}_q\right|}{\sqrt{2}\sqrt{41.25}} \geq 3.1177$$

i.e. if $\left|\overline{X}_p - \overline{X}_q\right| \geq \dfrac{(3.1177)\sqrt{2}\sqrt{41.25}}{\sqrt{5}} = \dfrac{28.31791653}{2.236067977} \approx 12.6642$

$$\begin{array}{lclclcl}
\left|\overline{X}_1 - \overline{X}_2\right| & = & |262.2 - 234.8| & = & 27.4 > 12.6642 & \Longrightarrow & \mu_1 \neq \mu_2 \\
\left|\overline{X}_1 - \overline{X}_3\right| & = & |262.2 - 251.6| & & 10.6 < 12.6642 & \Longrightarrow & \mu_1 = \mu_3 \\
\left|\overline{X}_1 - \overline{X}_4\right| & = & |262.2 - 268.6| & & 6.4 < 12.6642 & \Longrightarrow & \mu_1 = \mu_4 \\
\left|\overline{X}_2 - \overline{X}_3\right| & = & |234.8 - 251.6| & & 16.8 > 12.6642 & \Longrightarrow & \mu_2 \neq \mu_3 \\
\left|\overline{X}_2 - \overline{X}_4\right| & = & |234.8 - 268.6| & & 33.8 > 12.6642 & \Longrightarrow & \mu_2 \neq \mu_4 \\
\left|\overline{X}_3 - \overline{X}_4\right| & = & |251.6 - 268.6| & & 17 > 12.6642 & \Longrightarrow & \mu_3 \neq \mu_4
\end{array}$$

All pairs of means are significantly different from each other except the pairs $\overline{X}_1$ and $\overline{X}_3$; and $\overline{X}_1$ and $\overline{X}_4$, that is, $\mu_1 = \mu_3$ and $\mu_1 = \mu_4$. (8)

[30]

QUESTION 5

(a) Yes, it is reasonable to assume that the four groups may be considered as independent groups because suppliers in one state may not influence suppliers in the other states. (2)

(b) Start the *JMP* program

$>$ Enter *State* in the first column and label it *State*.

(make sure to change the scale to nominal)

$>$ Enter *Price* in the second column and label it *Price*.

This is a one-way ANOVA. To fit the model

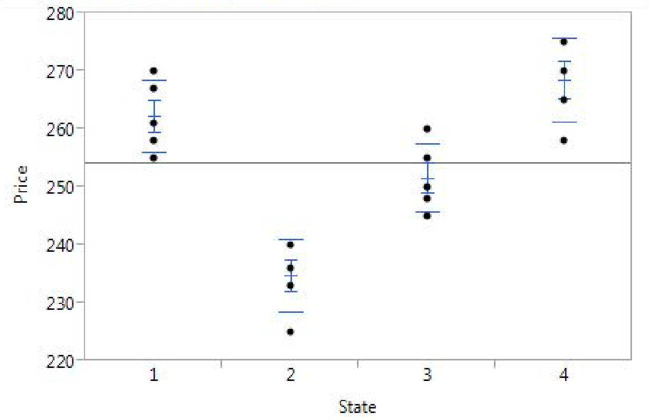
$>$ Choose Analyze$>$Fit Y by X with *State* as X factor and *Price* as Y response.

$>$ Click Ok.

$\Longrightarrow$ Then on the Oneway Analysis of *Price* By *State* click on the **Red** triangle

$>$ Choose Unequal Variances

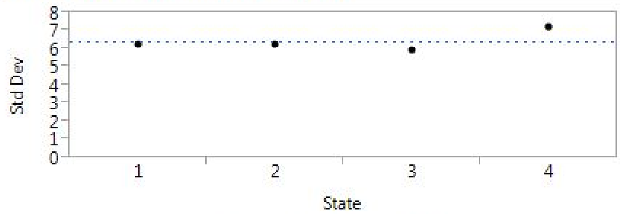
Oneway Analysis of Price By State



Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
1	5	262.200	6.22093	2.7821	254.48	269.92
2	5	234.800	6.22093	2.7821	227.08	242.52
3	5	251.600	5.94138	2.6571	244.22	258.98
4	5	268.600	7.23187	3.2342	259.62	277.58

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
1	5	6.220932	5.040000	5.400000
2	5	6.220932	4.640000	5.000000
3	5	5.941380	4.720000	4.800000
4	5	7.231874	5.680000	6.400000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.1129	3	16	0.9513
Brown-Forsythe	0.2646	3	16	0.8499
Levene	0.1171	3	16	0.9487
Bartlett	0.0556	3	.	0.9828

Warning: Small sample sizes. Use Caution.

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
22.3823	3	8.8701	0.0002*

Figure 5: Testing of Equality of Variances

For your own information:

The standard deviation column shows the estimates you are testing. The p-values are listed under the column called $Prob > F$ and are testing the assumption that the variances are equal. Small p-values suggest that the variance are not equal.

Interpretation:

We have to test:

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, against $H_1 : \sigma_p^2 \neq \sigma_q^2$ for at least one $p \neq q$

Using the Levene's test, p-value $= 0.9487$. Since $0.9487 > 0.05 \implies$ we can not reject H_0 at the 5% level of significance. The assumption of equal variances is not violated.

(8)

(c) $\Longrightarrow$ Click on the triangle "Tests that the variances are equal" to hide the output.

$\Longrightarrow$ Then click on the **Red** triangle on Oneway Analysis of *Price* by *State*.

$>$ Choose Means/ANOVA

$\Longrightarrow$ Click again on the **Red** triangle and choose Means and Std dev.

Oneway Anova

Summary of Fit

Rsquare	0.832155
Adj Rsquare	0.800684
Root Mean Square Error	6.422616
Mean of Response	254.3
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
State	3	3272.2000	1090.73	26.4420	<.0001*
Error	16	660.0000	41.25		
C. Total	19	3932.2000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	5	262.200	2.8723	256.11	268.29
2	5	234.800	2.8723	228.71	240.89
3	5	251.600	2.8723	245.51	257.69
4	5	268.600	2.8723	262.51	274.69

Std Error uses a pooled estimate of error variance

Figure 6: Oneway ANOVA

For your information:

On the plot, the dots shows the response for each *State*. The line across the middle is the grand mean. The diamonds give a 95% confidence interval for each *State* with the middle line of each diamond showing the group mean. If the groups are significantly different, then the diamonds do not overlap.

Interpretation:

(i) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against
$H_1 : \mu_p \neq \mu_q$ for at least one $p \neq q$.

(ii) The test statistic is $F = \frac{MSTr}{MSE} \sim F_{k-1;n-k}$

(iii) From the output: Computations for ANOVA we see that $F = 26.4420$ which is significant with a p–value of $< 0.0001 << 0.05$. We reject H_0 in favour of H_1 at the 5% level of significance and conclude that $u_p \neq \mu_q$ for at least one pair $p \neq q$, that is, the mean price per 1 000 board feet of standard or better grade green Douglas fir framing lumber of the *States* are not the same.

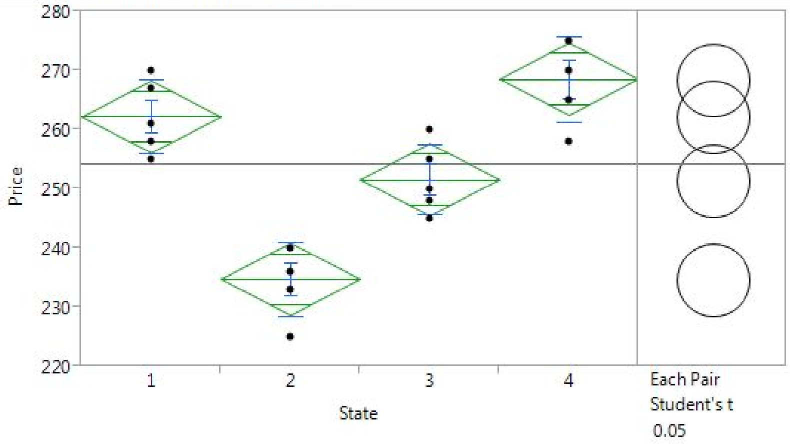
(9)

(d) $\Longrightarrow$ Hide the output "Oneway ANOVA" and "Means and Std deviations" by clicking the triangles.

$\Longrightarrow$ Click on the **Red** triangle on Oneway Analysis of *Price* by *State*.

$\Longrightarrow$ Choose Compare Means $>$ Each Pair, Student's t.

Oneway Analysis of Price By State



Means Comparisons

Comparisons for each pair using Student's t

Confidence Quantile

t	Alpha
2.11991	0.05

LSD Threshold Matrix

Abs(Dif)-LSD

	4	1	3	2
4	-8.611	-2.211	8.389	25.189
1	-2.211	-8.611	1.989	18.789
3	8.389	1.989	-8.611	8.189
2	25.189	18.789	8.189	-8.611

Positive values show pairs of means that are significantly different.

Connecting Letters Report

Level				Mean
4	A			268.60000
1	A			262.20000
3		B		251.60000
2			C	234.80000

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
4	2	33.80000	4.062019	25.1889	42.41110	<.0001*
1	2	27.40000	4.062019	18.7889	36.01110	<.0001*
4	3	17.00000	4.062019	8.3889	25.61110	0.0007*
3	2	16.80000	4.062019	8.1889	25.41110	0.0008*
1	3	10.60000	4.062019	1.9889	19.21110	0.0190*
4	1	6.40000	4.062019	-2.2111	15.01110	0.1347

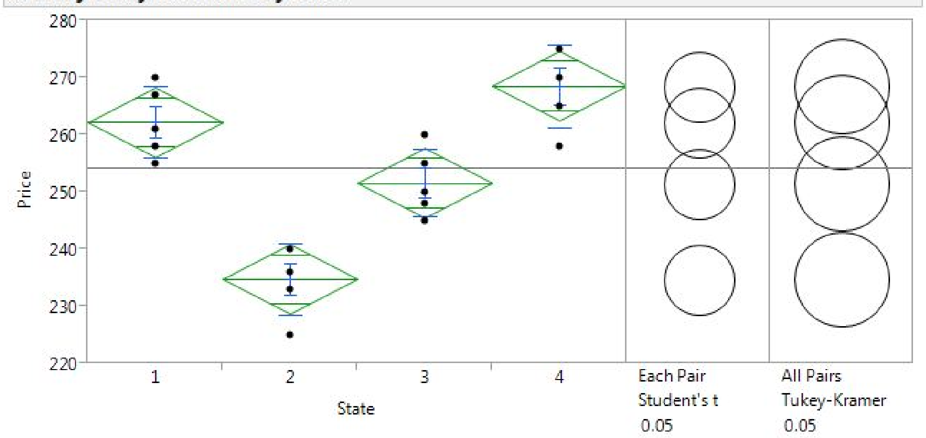
Figure 7: Pairwise t-tests using Student's t

Interpretation:

Yes The $Abs(Dif) - LSD$ for the pair 14 is negative. They all share the letter A and the confidence interval is $(-2.211; 15.0111)$ and it includes zero. We conclude that the means are not significantly different from each other. All the other pairs have $Abs(Dif) - LSDs$ that are positive. Thus we conclude that $\mu_2 \neq \mu_1 = \mu_4$. (9)

(e)

Oneway Analysis of Price By State



Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
2.86102	0.05

LSD Threshold Matrix

Abs(Dif)-HSD

	4	1	3	2
4	-11.622	-5.222	5.378	22.178
1	-5.222	-11.622	-1.022	15.778
3	5.378	-1.022	-11.622	5.178
2	22.178	15.778	5.178	-11.622

Positive values show pairs of means that are significantly different.

Connecting Letters Report

Level				Mean
4	A			268.60000
1	A	B		262.20000
3		B		251.60000
2			C	234.80000

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
4	2	33.80000	4.062019	22.1785	45.42152	<.0001*
1	2	27.40000	4.062019	15.7785	39.02152	<.0001*
4	3	17.00000	4.062019	5.3785	28.62152	0.0035*
3	2	16.80000	4.062019	5.1785	28.42152	0.0039*
1	3	10.60000	4.062019	-1.0215	22.22152	0.0803
4	1	6.40000	4.062019	-5.2215	18.02152	0.4191

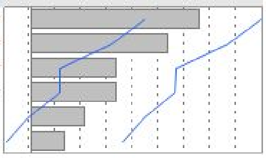


Figure 8: Comparisons of Means using Tukey-HSD

Manually, we should have computed for each pair of means, a test statistic

$$T_{pq} = \frac{\overline{X}_p - \overline{X}_q}{\mathrm{S}_{\text{pooled}}\sqrt{\frac{1}{n} + \frac{1}{n}}}$$

where we have samples of equal sizes if we want to incorporate the principle of the Bonferroni equality.

The Turkey–Kramer HSD that are shown in the JMP out perform individual comparisons that make adjustments for multiple test.

Confidence intervals that do not include zero imply that the pairs of means differ significantly. The pairs that do not include zero are 42, 12, 43 and 32. The confidence interval for the pairs are $(22.1785 : 45.42152)$; $(15.7785; 39.02152)$; $(5.3785; 28.62152)$; and $(5.1785; 28.42152)$; respectively. These are the only intervals that do not include zero and it means we reject the null hypothesis and conclude that $\mu_2 \neq \mu_4$, $\mu_1 \neq \mu_2$, $\mu_3 \neq \mu_4$ and $\mu_2 \neq \mu_3$. This is also supported by the fact that the p-values for the differences between the means are < 0.0001, < 0.0001 , 0.0035 and 0.0039 respectively. All p-values are $<< 0.05$ (highly significant), leading to the rejection of the null hypothesis of equal means.

The pairs that do not have the same letter connecting them means that the pairs are significantly different from each other.

Confirming this is the **Abs(Dif)-LSD** which are 22.178; 15.778; 5.378 and 5.178 for the pairs 42, 12, 43 and 32 respectively. Since all of them are positive thus, the means are significantly different. (Recall a negative value of **Abs(Dif)-LSD** means the groups are not significantly different from each other.)

(5)

[33]

QUESTION 6

(a)

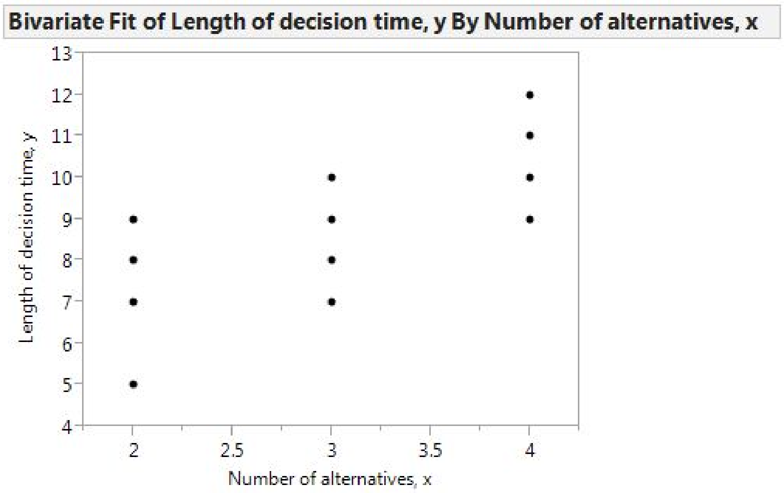


Figure 9: Simple Linear Regression Plot of Y vs X

(3)

(b) $n = 15$ $\quad \Sigma X_i = 45$ $\quad \Sigma X_i^2 = 145$

$\Sigma X_i Y_i = 411$ $\quad \Sigma Y_i = 132$ $\quad \Sigma Y_i^2 = 1\,204$

$$\begin{aligned} b &= \frac{n \Sigma X_i Y_i - (\Sigma X_i)(\Sigma Y_i)}{n \Sigma X_i^2 - (\Sigma X_i)^2} \\ &= \frac{15\,(411) - (45)\,(132)}{15\,(145) - (45)^2} \\ &= \frac{6\,165 - 5\,940}{2\,175 - 2\,025} \\ &= \frac{225}{150} \\ &= 1.5 \end{aligned}$$

$$\begin{aligned} a &= \frac{\Sigma Y_i - b\,(\Sigma X_i)}{n} \\ &= \frac{132 - 1.5\,(45)}{105} \\ &= \frac{132 - 67.5}{15} \\ &= \frac{64.5}{15} \\ &= 4.3 \end{aligned}$$

The estimated regression line is $\widehat{Length\ of\ decision\ time} = 4.3 + 1.5 No.\ of\ alternatives.$ (6)

(c)

X_i	Y_i	$\widehat{Y_i} = \widehat{\beta}_0 + \widehat{\beta}_1 X$	$e_i = Y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X\right)$	$e_i^2 = (Y_i - Y_i)^2$
2	5	7.3	−2.3	5.29
2	8	7.3	0.7	0.49
2	8	7.3	0.7	0.49
2	7	7.3	−0.3	0.09
2	9	7.3	1.7	2.89
3	7	8.8	−1.8	3.24
3	9	8.8	0.2	0.04
3	8	8.8	−0.8	0.64
3	9	8.8	0.2	0.04
3	10	8.8	1.2	1.44
4	10	10.3	−0.3	0.09
4	11	10.3	0.7	0.49
4	10	10.3	−0.3	0.09
4	12	10.3	1.7	2.89
4	9	10.3	−1.3	1.69
$\sum$				19.9

(5)

(d) The confidence interval is $\left(\widehat{\beta}_0 + \widehat{\beta}_1 X\right) \pm t_{\alpha/2;n-2} \times S\sqrt{\frac{1}{n} + \frac{\left(X-\overline{X}\right)^2}{d^2}}$

$$\begin{aligned} s^2 &= \frac{\sum (y_i - \widehat{y}_i)^2}{n-2} \\ &= \frac{19.9}{13} \\ &\approx 1.530769231 \\ \implies & s = \sqrt{1.530769231} \approx 1.2372 \end{aligned}$$

$$\begin{array}{lll} n = 15 & \widehat{\beta}_0 + \widehat{\beta}_1 X = 8.8 & X = 3 \\ \overline{X} = 3 & d^2 = \sum \left(X - \overline{X}\right)^2 = 10 & s \approx 1.2372 \\ \alpha = 0.05 & \alpha/2 = 0.025 & t_{\alpha/2;n-2} = t_{0.025;13} = 2.160 \end{array}$$

The 95% confidence interval for the average length of time necessary to make a decision when three alternatives are presented is

$$\begin{array}{lll} \widehat{\beta}_0 + \widehat{\beta}_1 X & \pm & t_{\alpha/2;n-2} \times S\sqrt{\frac{1}{n} + \frac{\left(X - \overline{X}\right)^2}{d^2}} \\ 8.8 & \pm & 2.160 \times 1.2372\sqrt{\frac{1}{15} + \frac{(3-3)^2}{10}} \\ 8.8 & \pm & 2.672352\sqrt{0.066666667 + 0} \\ 8.8 & \pm & 2.672352\sqrt{0.066666667} \\ 8.8 & \pm & 0.69 \\ (8.8 - 0.69) & ; & 8.8 + 0.69) \\ (8.11 & ; & 9.49) \end{array}$$

Then the interval is (8.11 to 9.49)

(6)

(e) $H_0 : \beta_1 = 0$ $\quad H_1 : \beta_1 > 0$

$\alpha = 0.01$ $\quad t_{\alpha;n-2} = t_{0.01;13} = 2.650$. Reject H_0 if T is greater than 2.65. Now

$$\begin{aligned} T &= \frac{\hat{\beta}_1 - B_1}{s/d} \\ &= \frac{1.5 - 0}{1.2372/\sqrt{10}} \\ &= \frac{1.5}{0.391236992} \\ &= 3.8340 \end{aligned}$$

Since $3.834 > 2.65$, we reject H_0 at the 1% level significance and conclude that $\beta_1 > 0$. (6)

(f) Model fitted is $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$

Commands for the Output:

Start the JMP program

$>$ *Enter number of alternatives in the first column and label it* $\underline{Number\ of\ alternatives\ (x)}$.

$>$ *Enter length of decision time in the second column and label it* $\underline{Length\ of\ decision\ time\ (y)}$

To plot:

$>$ *Choose* Analyze>Fit Y by X *with* $\underline{Number\ of\ alternatives\ (x)}$ *as* X *factor and* $\underline{Length\ of\ decision\ time\ (y)}$ *as* Y *response.*

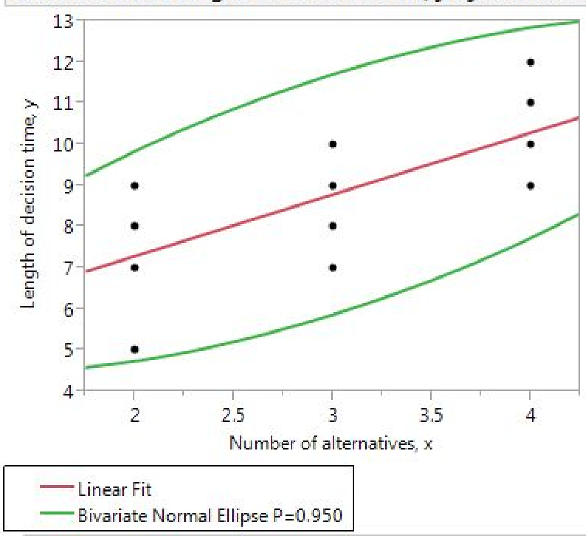
$>$ Click Ok.

Click on the ***Red*** *triangle on Bivariate Fit of* $Length\ of\ decision\ time\ (y)$ *by* $Number\ of\ alternatives\ (x)$.

$>$ *Choose* Fit Line

The JMP output is

Bivariate Fit of Length of decision time, y By Number of alternatives, x



Linear Fit

Length of decision time, y = 4.3 + 1.5*Number of alternatives, x

Summary of Fit

RSquare	0.53066
RSquare Adj	0.494557
Root Mean Square Error	1.237243
Mean of Response	8.8
Observations (or Sum Wgts)	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	22.500000	22.5000	14.6985
Error	13	19.900000	1.5308	Prob > F
C. Total	14	42.400000		0.0021*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.3	1.216447	3.53	0.0037*
Number of alternatives, x	1.5	0.39125	3.83	0.0021*

Correlation

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Number of alternatives, x	3	0.845154	0.728464	0.0021*	15
Length of decision time, y	8.8	1.740279			

Figure 10: The Simple Linear Regression Model

(7)

[33]

[Total Marks: 175]