Tutorial letter 203/2/2017

Applied Statistics II

STA2601

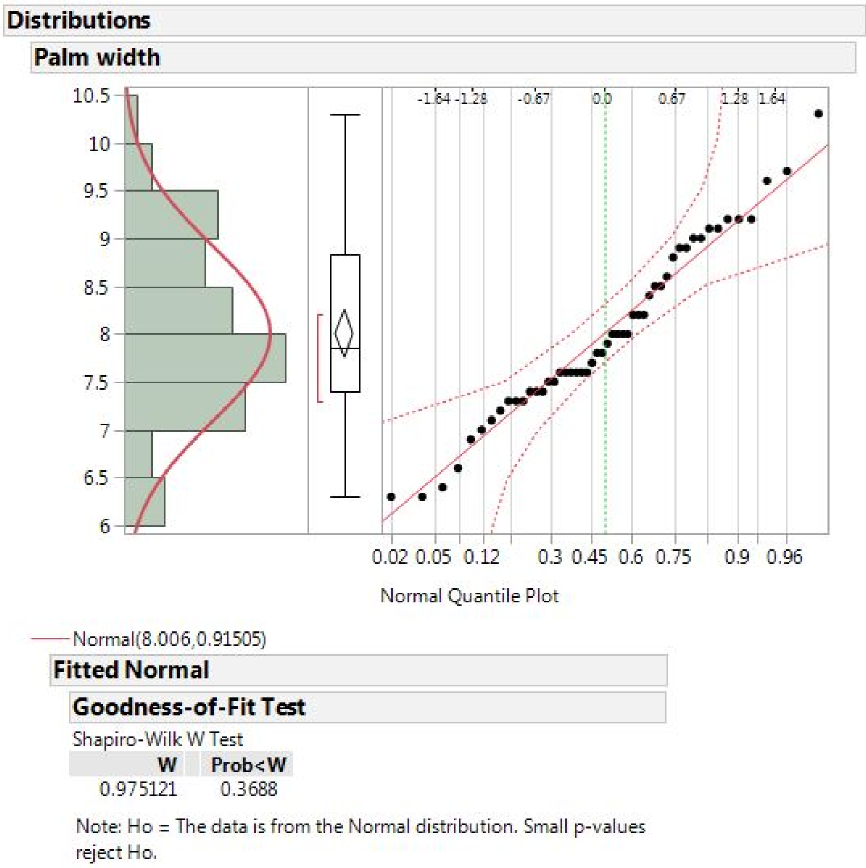
Semester 2

Department of Statistics

Solutions to Assignment 03

QUESTION 1

(a) (i)

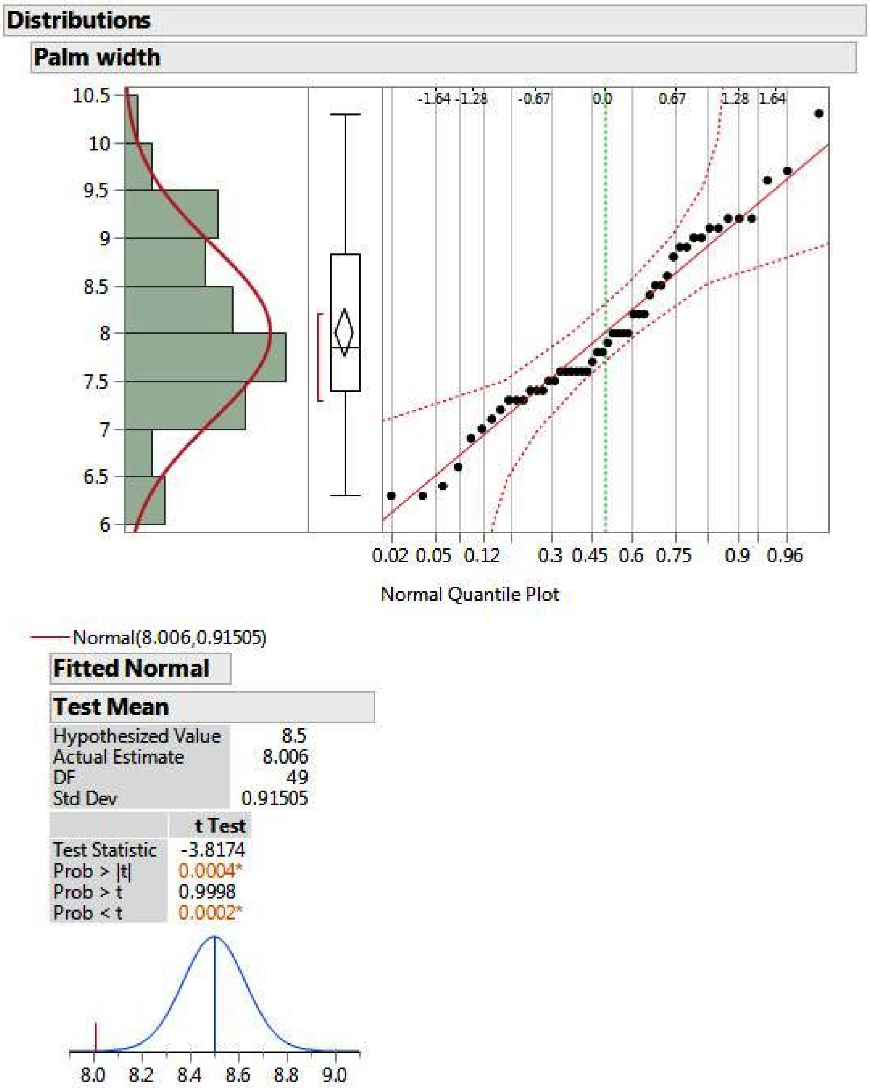


The normal quantile plot shows that the points at both ends are not following a diagonal. They seem to slightly deviate from the line. Secondly the histogram and box plot shows that data is positively skewed (Its subjective).

We need a proper test. The Shapiro-Wilk test for normality shows that the null hypothesis (H_0: Data comes from a normal distribution) would not be rejected (p-value $= 0.3688$), indicating that we may assume that the data does come from a normal distribution.

(7)

(ii)



We have to test $H_0 : \mu = 8.5$ against $H_1 : \mu \neq 8.5$.

From the output $\overline{X} = 8.006$ and $s = 0.91505$.

Method 1: Using the critical value approach

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{s} = \frac{\sqrt{50}\,(8.006 - 8.5)}{0.91505} \approx -3.8174$$

The critical value is

$$
\begin{aligned}
t_{\alpha/2;n-1} &= t_{0.025;49} \\
&= 2.021 + \frac{9}{20}(2.000 - 2.021) \\
&= 2.021 + \frac{9}{20}(-0.021) \\
&= 2.021 - 0.00945 \\
&\approx 2.012
\end{aligned}
$$

We will reject H_0 if $T \leq -2.012$, or if $T > 2.012$ or if $|T| > 2.012$.

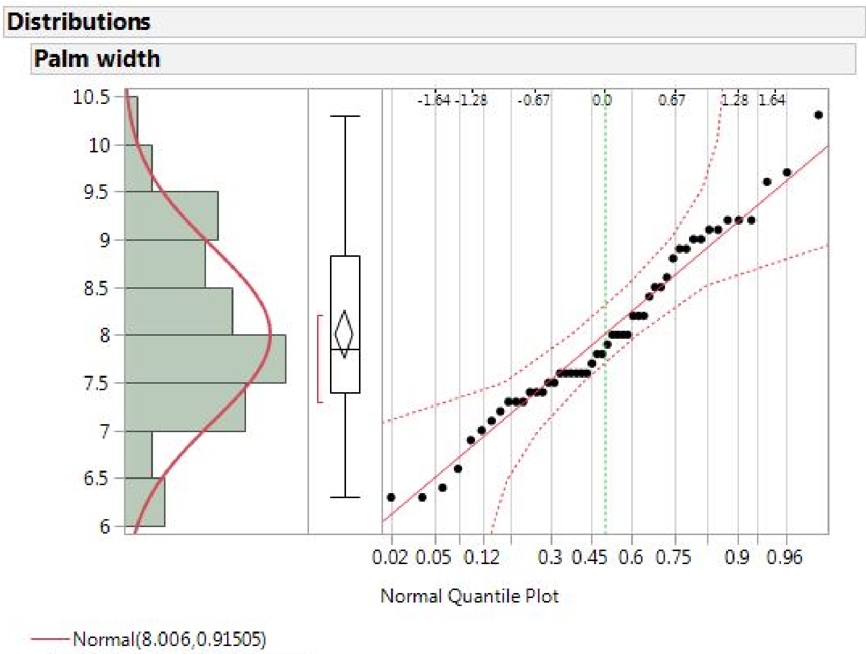
Since $-3.8174 < -2.012$ we reject H_0 at the 5% level of significance and conclude that $\mu \neq 8.5$, i.e., the mean palm width of the right hand is significantly different from 8.5.

Method II: Using the p-value approach

p-value $= 0.0004$. Since $0.0004 < 0.05$, we reject H_0 at the 5% level of significance and conclude that $\mu \neq 8.5$, i.e., the mean palm width of the right hand is significantly different from 8.5.

(10)

(iii)



Summary Statistics

Mean	8.006
Std Dev	0.9150466
Std Err Mean	0.1294071
Upper 95% Mean	8.2660534
Lower 95% Mean	7.7459466
N	50

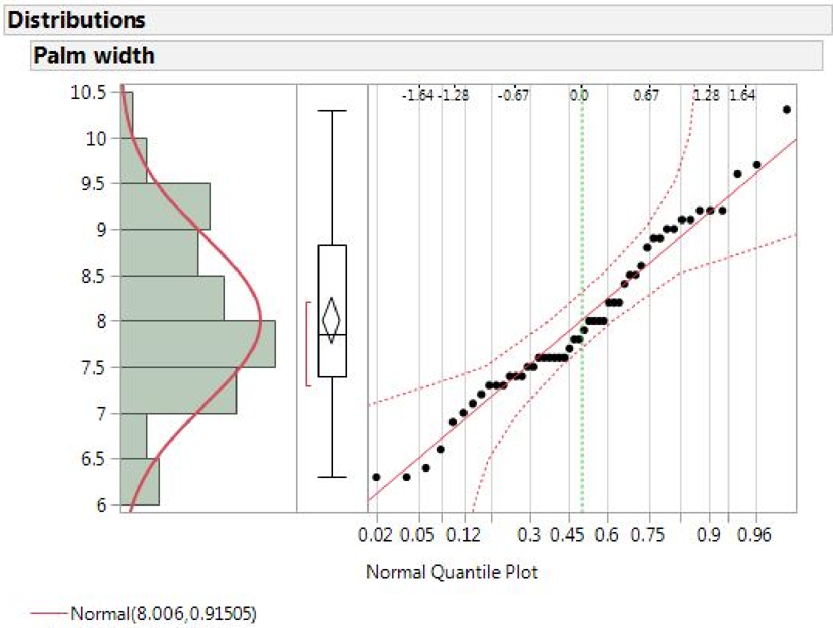
Fitted Normal

From the output, the 95% confidence interval for μ is 8.2661 to 7.7459. The interval supports the conclusion in part (ii). Since the 95% confidence interval is the same as testing a two sided test at the 5% level. Now we are 95% confident that $8.2661 \leq \mu \leq 7.7459$. The two tailed 5% test can be compared to a 95% confidence interval. In this case the value 8.5 does not lie in the interval and thus we reject H_0 at the 5% level of significance and conclude that $\mu \neq 8.5$, i.e., the mean palm width of the right hand is significantly different from 8.5.

(7)

(iv) We have to test $H_0 : \sigma = 1$

against $H_1 : \sigma \neq 1$



Fitted Normal

Test Standard Deviation

Hypothesized Value	1
Actual Estimate	0.91505
DF	49

Test	ChiSquare
Test Statistic	41.0282
Min PValue	0.4323
Prob < ChiSq	0.2162
Prob > ChiSq	0.7838

Method 1: Using the critical value approach

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

$$U = \frac{(n-1)\,s^2}{\sigma^2} = \frac{49\,(0.9150466)^2}{1} \approx 41.0282$$

The critical values are

$$\begin{aligned}\chi^2_{1-\alpha/2;n-1} &= \chi^2_{0.975;49} \\ &= 24.4331 + \tfrac{9}{10}(32.3574 - 24.4331) \\ &= 24.4331 + \tfrac{9}{10}(7.9243) \\ &= 24.4331 + 7.13187 \\ &\approx 31.565\end{aligned}$$

$$\begin{aligned}\chi^2_{\alpha/2;n-1} &= \chi^2_{0.025;29} \\ &= 59.3417 + \tfrac{9}{10}(71.4202 - 59.3417) \\ &= 59.3417 + \tfrac{9}{10}(12.0785) \\ &= 59.3417 + 10.87065 \\ &\approx 70.2124\end{aligned}$$

Reject H_0 if $U < 31.565$ or $U > 70.2124$

Since $31.565 < 41.0282 < 70.2124$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 1$.

Method II: Using the p-value approach

p-value $= 0.4323$. Since $0.4323 > 0.05$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 1$.

The assumption made was that the mean μ is unknown and hence the test statistic $U = \dfrac{\Sigma\left(X_i - \overline{X}\right)^2}{\sigma^2}$ was used at χ^2_{n-1}.

(10)

(b) (i) In order to perform the tests we assumed that:

- the observations in each sample are independent and also the two samples are mutually independent.
- the observations are normally distributed.
- the two population variances are equal.

The samples are independent (stated). We need to test for equal variances.

Men:	$n_1 = 20$	$\overline{X}_1 = 57.4$	$S_1 = 8.124$
Women:	$n_2 = 25$	$\overline{X}_2 = 63.4$	$S_2 = 7.874$

We have to test $H_0 : \sigma_1^2 = \sigma_2^2$

against $H_1 : \sigma_1^2 \neq \sigma_2^2$

The test statistic is

$$\begin{aligned} F &= \frac{\sigma_2^2}{\sigma_1^2} \times \frac{S_1^2}{S_2^2} \\ &= 1 \times \frac{8.124^2}{7.874^2} \\ &\approx 1.0645 \end{aligned}$$

The critical values are:

$F_{\alpha/2;n_1-1;n_2-1} = F_{0.025;19;24} = 2.44 + \frac{4}{5}(2.33 - 2.44) = 2.44 + 0.8\,(-0.11) = 2.352 \approx 2.35$
and $F_{1-\alpha/2;n_1-1;n_2-1} = \dfrac{1}{F_{\alpha/2;n_2-1;n_1-1}} = \dfrac{1}{F_{0.025;24;19}} = \dfrac{1}{2.45} \approx 0.41.$

Reject H_0 if $F > 2.35$ or $F < 0.41$.

[**Note**: if you approximate from your tables without interpolation $F_{0.025;19;24} \approx 2.33$ and $\dfrac{1}{F_{0.025;24;19}} = \dfrac{1}{2.45} \approx 0.41$.]

Since $0.41 < 1.0645 < 2.35$, we do not reject H_0 at the 5% level of significance and conclude that the variances are equal i.e. $\sigma_1^2 = \sigma_2^2$.

(11)

(ii) $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$

$n_1 = 20 \quad \overline{X}_1 = 57.4 \quad S_1 = 8.124$

$n_2 = 25 \quad \overline{X}_2 = 63.4 \quad S_2 = 7.874$

The test statistic is

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Now

$$\begin{aligned} S_p^2 &= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \\ &= \frac{(20-1)8.124^2 + (25-1)7.874^2}{20+25-2} \\ &= \frac{19\,(65.999376) + 24\,(61.999876)}{43} \\ &= \frac{1\,253.988144 + 1\,487.997024}{43} \\ &= \frac{2\,741.985168}{43} \\ &= \approx 63.7671 \\ \Longrightarrow &\quad S_{pooled} = \sqrt{63.7671} \approx 7.9854 \end{aligned}$$

The test statistic is

$$\begin{aligned}
T &= \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\
&= \frac{(57.4 - 63.4) - (0)}{7.9854\sqrt{\frac{1}{20} + \frac{1}{25}}} \\
&= \frac{-6}{7.9854\sqrt{0.09}} \\
&= \frac{-6}{2.39562} \\
&\approx -2.5046
\end{aligned}$$

Test is one tailed. The critical value is $t_{\alpha;(n_1+n_2-2)} = t_{0.05;43}$

Interpolating $t_{0.05;40} = 1.684$ and $t_{0.05;60} = 1.671$.

$$\begin{aligned}
t_{0.05;43}. &= 1.684 + \frac{3}{20}(1.671 - 1.684) \\
&= 1.684 + \frac{3}{20}(-0.013) \\
&= 1.684 - 0.00195 \\
&\approx 1.682
\end{aligned}$$

Reject H_0 if $T < -1.682$.

Since $-2.5046 < -1.682$, we reject H_0 at the 5% level and conclude that $\mu_1 < \mu_2$, that is, women are on average socially more skillful than men.

(10)

[55]

QUESTION 2

(a) We want 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 $p \neq q$

$\overline{Y}_1 = 15 \qquad \sum Y_{1j} = 75 \qquad \sum Y_{1j}^2 = 1\,129$

$\overline{Y}_2 = 17 \qquad \sum Y_{2j} = 85 \qquad \sum Y_{2j}^2 = 1\,459$

$\overline{Y}_3 = 19 \qquad \sum Y_{3j} = 95 \qquad \sum Y_{3j}^2 = 1\,809$

$\overline{Y}_4 = 21 \qquad \sum Y_{3j} = 105 \qquad \sum Y_{3j}^2 = 2\,211$

$$\begin{aligned} S_1^2 &= \frac{1}{n-1}\left(\sum X_{1j}^2 - \frac{\left(\sum X_{1j}\right)^2}{n}\right) \\ &= \frac{1}{5-1}\left(1\,129 - \frac{(75)^2}{5}\right) \\ &= \frac{1}{4}\,(1\,129 - 1\,125) \\ &= \frac{1}{4}\,(4) \\ &= 1 \end{aligned}$$

$$\begin{aligned} S_2^2 &= \frac{1}{n-1}\left(\sum X_{2j}^2 - \frac{\left(\sum X_{2j}\right)^2}{n}\right) \\ &= \frac{1}{5-1}\left(1\,459 - \frac{(85)^2}{5}\right) \\ &= \frac{1}{4}\,(1\,459 - 1\,445) \\ &= \frac{1}{4}\,(14) \\ &= 3.5 \end{aligned}$$

$$\begin{aligned} S_3^2 &= \frac{1}{n-1}\left(\sum X_{3j}^2 - \frac{\left(\sum X_{3j}\right)^2}{n}\right) \\ &= \frac{1}{5-1}\left(1\,809 - \frac{(95)^2}{5}\right) \\ &= \frac{1}{4}\,(1\,809 - 1\,805) \\ &= \frac{1}{4}\,(4) \\ &= 1 \end{aligned}$$

$$\begin{aligned} S_4^2 &= \frac{1}{n-1}\left(\sum X_{4j}^2 - \frac{\left(\sum X_{4j}\right)^2}{n}\right) \\ &= \frac{1}{5-1}\left(2\,211 - \frac{(105)^2}{5}\right) \\ &= \frac{1}{4}\,(2\,211 - 2\,205) \\ &= \frac{1}{4}\,(6) \\ &= 1.5 \end{aligned}$$

From the computations above it, follows that $S_1^2 = 1$; $S_2^2 = 3.5$; $S_3^2 = 1$ and $S_4^2 = 1.5$.

The test statistic is

$$\begin{aligned} U &= \frac{\max_i S_i^2}{\min_i S_i^2} \\ &= \frac{3.5}{1} \\ &= 3.5 \end{aligned}$$

The critical value is 20.6. H_0 is rejected if $U > 20.6$.

Since $3.5 < 20.6$, we do not reject H_0 at the 5% level of significance and conclude that the variances of the four populations are equal.

(15)

(b) $k = 4$ $\quad n = 5$ $\quad kn - k = 16$ $\quad k - 1 = 3$

$$\overline{X}_1 = 15 \qquad SS_1 = \sum_{j=1}^{5} \left(X_{1j} - \overline{X}_1\right)^2 = 4$$

$$\overline{X}_2 = 17 \qquad SS_2 = \sum_{j=1}^{5} \left(X_{2j} - \overline{X}_2\right)^2 = 14$$

$$\overline{X}_3 = 19 \qquad SS_3 = \sum_{j=1}^{5} \left(X_{3j} - \overline{X}_3\right)^2 = 4$$

$$\overline{X}_3 = 21 \qquad SS_4 = \sum_{j=1}^{5} \left(X_{3j} - \overline{X}_3\right)^2 = 6$$

$$\begin{aligned} SSE &= SS_1 + SS_2 + SS_3 + SS_4 \\ &= 4 + 14 + 4 + 6 \\ &= 28 \end{aligned}$$

$$MSE = S^2 = \frac{SSE}{kn - k} = \frac{28}{16} = 1.75$$

$$\overline{X} = \frac{(75 + 85 + 95 + 105)}{20} = \frac{360}{20} = 18$$

$$\begin{aligned}\sum_{i=1}^{4} \left(\overline{X}_i - \overline{X}\right)^2 &= (15-18)^2 + (17-18)^2 + (19-18)^2 + (21-18)^2 \\ &= (-3)^2 + (-1)^2 + (1)^2 + (3)^2 \\ &= 9 + 1 + 1 + 9 \\ &= 20\end{aligned}$$

$$SSTr = n\Sigma\left(\overline{X}_i - \overline{X}\right)^2 = 5(20) = 100$$

$$MSTr = \frac{n\Sigma\left(\overline{X}_i - \overline{X}\right)^2}{(k-1)} = \frac{100}{3} \approx 33.3333$$

$$F = \frac{MSTr}{MSE} = \frac{33.3333}{1.75} \approx 19.0476$$

The ANOVA table is

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Treatments	100	3	33.3333	19.0476
Error	28	16	1.75	
Total	128	19		

Testing $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against $H_1 : \mu_p \neq \mu_p$ for at least one pair $p \neq q$

The critical value is $F_{0.05;3;16} = 3.24$. Reject H_0 if $F > 3.24$

Since $F > F_{0.05;3;16}$, i.e., $19.0476 > 3.24$, we reject H_0 at the 5% level of significance and conclude that at least one pair is significantly different from each other.

(20)

(c) For each pair of means we compute a test statistic:

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

We reject $H_0(p, q)$ if

$$\left|T_{pq}\right| > \sqrt{(k-1)\,F_{\alpha;k-1;kn-k}} = \sqrt{3\,(3.24)} = \sqrt{9.72} \approx 3.1177$$

This implies that we reject H_0 if

$$\frac{\sqrt{5}\left|\overline{X}_p - \overline{X}_q\right|}{\sqrt{2}\sqrt{1.75}} \geq 3.1177$$

i.e., if

$$\left|\overline{X}_p - \overline{X}_q\right| \geq \frac{3.1177\sqrt{2}\sqrt{1.75}}{\sqrt{5}} = \frac{5.832682617}{2.236067977} \approx 2.6085$$

$$\begin{aligned}
\left|\overline{X}_1 - \overline{X}_2\right| &= |15 - 17| = 2 < 2.6085 \Longrightarrow \mu_1 = \mu_2 \\
\left|\overline{X}_1 - \overline{X}_4\right| &= |15 - 21| = 6 > 2.6085 \Longrightarrow \mu_1 \neq \mu_4 \\
\left|\overline{X}_2 - \overline{X}_4\right| &= |17 - 21| = 4 > 2.6085 \Longrightarrow \mu_1 \neq \mu_4
\end{aligned}$$

The pairs of means $\overline{X}_1$and $\overline{X}_2$ do not differ significantly. However $\overline{X}_1$ and $\overline{X}_2$ differ significantly from $\overline{X}_4$. It can be concluded that $\mu_1 = \mu_2 \neq \mu_4$.

(10)

[45]

QUESTION 3

(a)

Day	**Sales during campaign**	**Sales after campaign**	$Y_i =$ **During- After**
Sunday	18.1	16.6	1.5
Monday	10.0	8.8	1.2
Tuesday	9.1	8.6	0.5
Wednesday	8.4	8.3	0.1
Thursday	10.8	10.1	0.7
Friday	13.1	12.3	0.8
Saturday	20.8	18.9	1.9

$n = 7 \quad \sum Y_i = 6.7 \quad \sum \left(Y_i - \overline{Y}\right)^2 = 2.2771$

We have to test:

$H_0 : \mu_d = 0$ against

$H_1 : \mu_d > 0$

$$\begin{aligned} \overline{Y} &= \frac{1}{n}\sum Y_i & S_y^2 &= \frac{1}{n-1}\sum\left(Y_i - \overline{Y}\right)^2 \\ &= \frac{1}{7}(6.7) & &= \frac{1}{6}(2.2771) \\ &\approx 0.9571 & &= 0.379516666 \\ & & \Longrightarrow S_y &= \sqrt{0.379516666} \\ & & &\approx 0.6160 \end{aligned}$$

The test statistic is

$$\begin{aligned} T &= \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{S_y} \\ &= \frac{\sqrt{7}\,(0.9571 - 0)}{0.6160} \\ &= \frac{2.53224858}{0.6160} \\ &\approx 4.1108 \end{aligned}$$

$t_{\alpha;(n-1)} = t_{0.05;6} = 1.943$. We will reject H_0 if $T \geq 1.943$.

Since $4.1108 > 1.943$, we reject H_0 at the 5% level of significance and conclude that sales increased during campaign. (13)

(b) $n = 7 \qquad \alpha = 0.05 \qquad \alpha/2 = 0.025$

$t_{\alpha/2;(n-1)} = t_{0.025;6} = 2.447$

The 95% confidence interval is

$$\begin{array}{lcl}
\overline{Y} & \pm & t_{\alpha/2;(n-1)} \times \dfrac{S_y}{\sqrt{n}} \\
0.9571 & \pm & 2.447 \times \dfrac{0.616}{\sqrt{7}} \\
0.9571 & \pm & 0.5697 \\
(0.9571 - 0.5697) & ; & 0.9571 + 0.5697) \\
(0.3874) & ; & 1.5268)
\end{array}$$

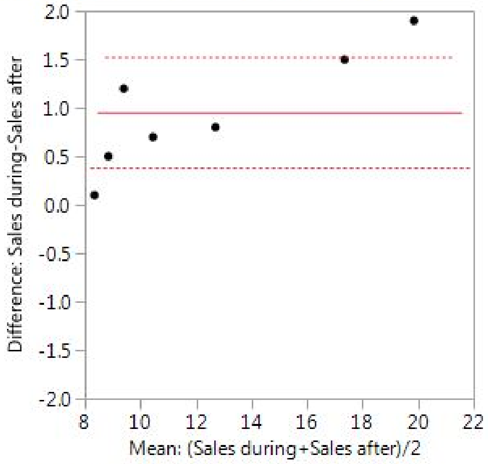
We are 95% confident that the true mean differences, μ_d lies between 0.3874 and 1.5268.

(7)

(c) The output is

Matched Pairs

Difference: Sales during-Sales after



Sales during	12.9	t-Ratio	4.11061
Sales after	11.9429	DF	6
Mean Difference	0.95714	Prob > |t|	0.0063*
Std Error	0.23285	Prob > t	0.0031*
Upper 95%	1.5269	Prob < t	0.9969
Lower 95%	0.38739		
N	7		
Correlation	0.99775		

(5)

[25]

QUESTION 4

(a) Start the *JMP* program

$>$ Enter *Amount of drug* in the first column and label it *Amount of drug*.

(make sure to change the scale to nominal)

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

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

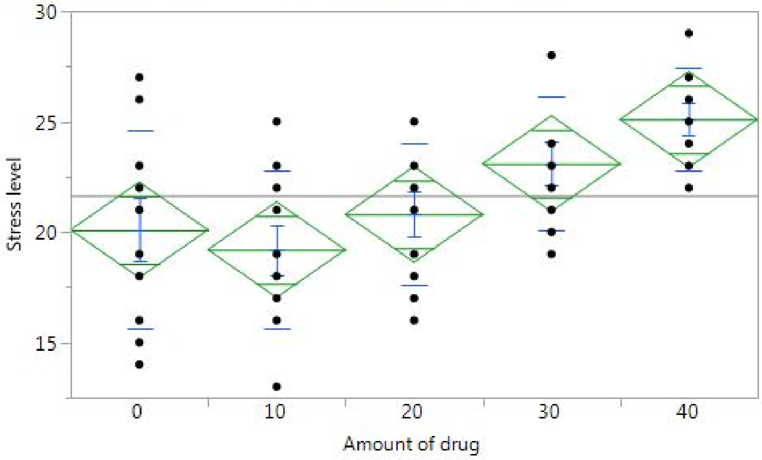
$>$ Choose Analyze$>$Fit Y by X with *Amount of drug* as X factor and *Stress level* as Y response.

$>$ Click Ok.

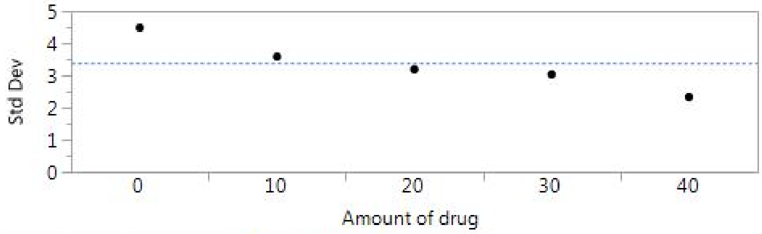
$\Longrightarrow$ Then on the Oneway Analysis of *Stress level* By *Amount of drug* click on the **Red** triangle

$>$ Choose Unequal Variances

Oneway Analysis of Stress level By Amount of drug



Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
0	10	4.483302	3.700000	3.700000
10	10	3.583915	2.840000	2.800000
20	10	3.190263	2.640000	2.600000
30	10	3.034981	2.320000	2.300000
40	10	2.330951	1.900000	1.900000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	1.5943	4	45	0.1923
Brown-Forsythe	1.2498	4	45	0.3037
Levene	1.4232	4	45	0.2417
Bartlett	0.9653	4	.	0.4251

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
6.1967	4	22.239	0.0017*

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 Bartlett's test, p-value $= 0.4251$. Since $0.4251 > 0.05 \Longrightarrow$ we can not reject H_0 at the 5% level of significance. The assumption of equal variances is not violated. (10)

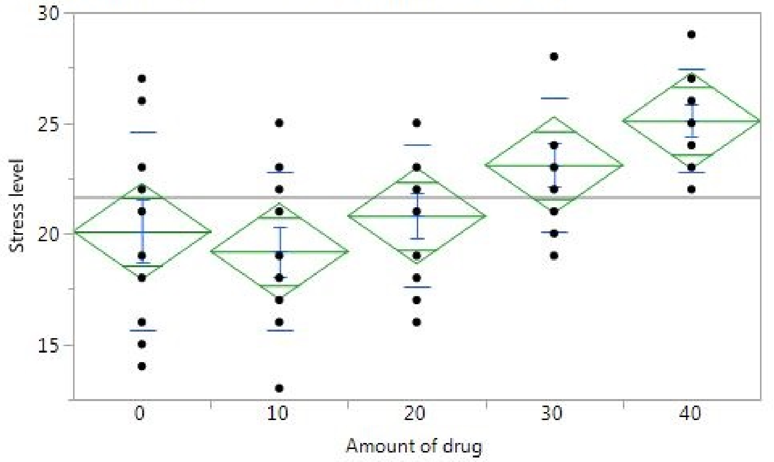
(b) ⟹ Click on the triangle "Tests that the variances are equal" to hide the output.

⟹ Then click on the **Red** triangle on Oneway Analysis of *Stress level* by *Amount of drug*.

> Choose Means/ANOVA

⟹ Click again on the **Red** triangle and choose Means and Std dev.

Oneway Analysis of Stress level By Amount of drug



Oneway Anova

Summary of Fit

Rsquare	0.307926
Adj Rsquare	0.246408
Root Mean Square Error	3.399019
Mean of Response	21.66
Observations (or Sum Wgts)	50

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Amount of drug	4	231.32000	57.8300	5.0055	0.0020*
Error	45	519.90000	11.5533		
C. Total	49	751.22000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
0	10	20.1000	1.0749	17.935	22.265
10	10	19.2000	1.0749	17.035	21.365
20	10	20.8000	1.0749	18.635	22.965
30	10	23.1000	1.0749	20.935	25.265
40	10	25.1000	1.0749	22.935	27.265

Std Error uses a pooled estimate of error variance

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
0	10	20.1000	4.48330	1.4177	16.893	23.307
10	10	19.2000	3.58391	1.1333	16.636	21.764
20	10	20.8000	3.19026	1.0088	18.518	23.082
30	10	23.1000	3.03498	0.9597	20.929	25.271
40	10	25.1000	2.33095	0.7371	23.433	26.767

For your information:

On the plot, the dots shows the response for each *Amount of drug*. The line across the middle is the grand mean. The diamonds give a 95% confidence interval for each *Amount of drug* 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 $F = 5.0055$ which is significant with a p–value of 0.0020. Since $0.0020 < 0.05$ 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 stress level of the companies are not the same.

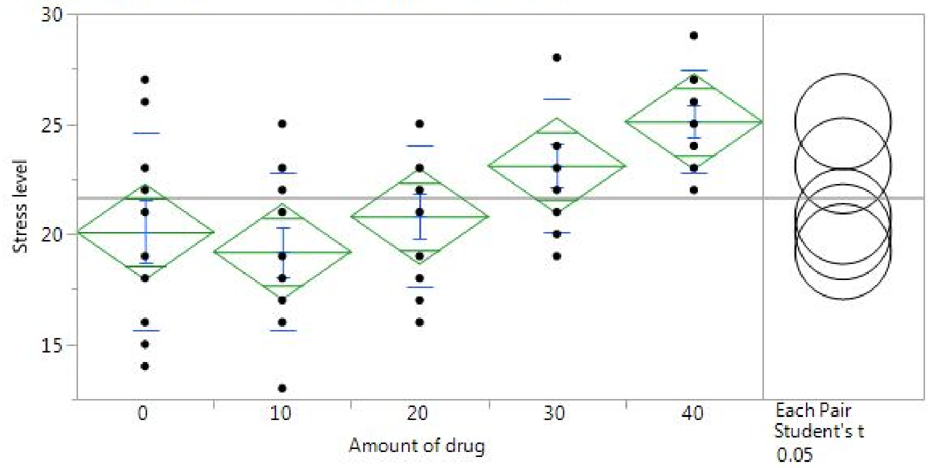
(10)

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

$\Longrightarrow$ Click on the **Red** triangle on Oneway Analysis of *Stress level* by *Amount of drug*.

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

Oneway Analysis of Stress level By Amount of drug



Means Comparisons

Comparisons for each pair using Student's t

Confidence Quantile

t	Alpha
2.01410	0.05

LSD Threshold Matrix

Abs(Dif)-LSD

	40	30	20	0	10
40	-3.0616	-1.0616	1.2384	1.9384	2.8384
30	-1.0616	-3.0616	-0.7616	-0.0616	0.8384
20	1.2384	-0.7616	-3.0616	-2.3616	-1.4616
0	1.9384	-0.0616	-2.3616	-3.0616	-2.1616
10	2.8384	0.8384	-1.4616	-2.1616	-3.0616

Positive values show pairs of means that are significantly different.

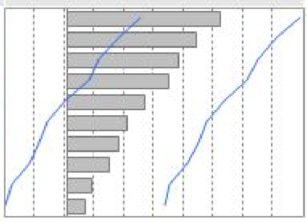
Connecting Letters Report

Level				Mean
40	A			25.100000
30	A	B		23.100000
20		B	C	20.800000
0		B	C	20.100000
10			C	19.200000

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
40	10	5.900000	1.520088	2.83839	8.961614	0.0003*
40	0	5.000000	1.520088	1.93839	8.061614	0.0020*
40	20	4.300000	1.520088	1.23839	7.361614	0.0070*
30	10	3.900000	1.520088	0.83839	6.961614	0.0137*
30	0	3.000000	1.520088	-0.06161	6.061614	0.0546
30	20	2.300000	1.520088	-0.76161	5.361614	0.1373
40	30	2.000000	1.520088	-1.06161	5.061614	0.1949
20	10	1.600000	1.520088	-1.46161	4.661614	0.2982
0	10	0.900000	1.520088	-2.16161	3.961614	0.5568
20	0	0.700000	1.520088	-2.36161	3.761614	0.6474



Amounts of drug injected (*CC*) that share the same letter are not significantly different from each other. *CC30* and *CC40* share the same letter A, *CC0*, *CC20* and *CC30* share the same letter B and *CC0*, *CC10* and *CC20* share the same letter C.

The amounts of drugs which are significantly different from each other have **Abs(Dif)-LSDs** that are positive. The pairs are *CC40-CC20*, *CC40-CC0*, *CC40-CC10* and *CC30-CC10* which are $1.2384, 1.9384, 2.8384$ and 0.8384 respectively. Since they are positive, the means are significantly different. (Recall a negative value of **Abs(Dif)-LSD** means the groups are not significantly different from each other.)

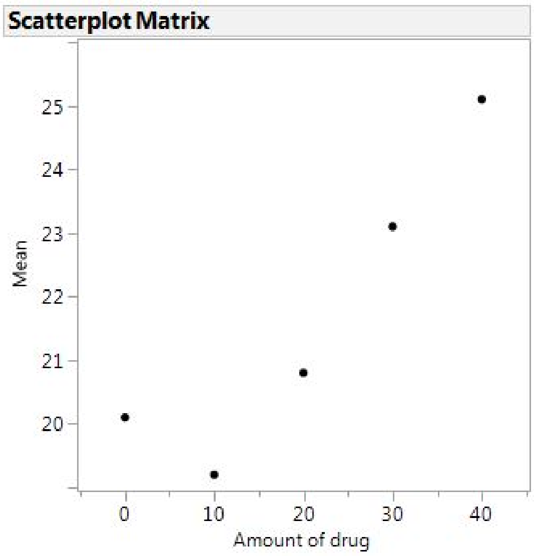
Confidence intervals that do not include zero imply that the pairs of means differ significantly. All pairs include zero except the pair *CC40-CC10, CC40-CC0, CC40-CC20* and *CC30-CC10*. The confidence interval for the pairs are $(2.8384 : 8.9616)$, $(1.9384 : 8.0616)$, $(1.2384 : 7.3616)$ and $(0.8384 : 6.9616)$. These are the only intervals that do not include zero and it means we reject the null hypothesis of equal means and conclude that $\mu_{40} \neq \mu_{10}$, $\mu_{40} \neq \mu_0$, $\mu_{40} \neq \mu_{20}$,and $\mu_{30} \neq \mu_{10}$. The $p-$values are 0.0003, 0.0020, 0.0070 and 0.0137 respectively which are less than 0.05 and thus leading to the rejection of the null hypothesis of equal means.

(14)

(d) Student's t does pairwise comparisons of means. Comparisons of many pairs of means increase the possibility of a *Type I error*. One must remember that using pairwise *t*-tests doesn't control the overall error for all comparisons made (also called the experimental error rate). A Tukey-Kramer tests can be used to control for an overall error rate since it compares all means simultaneously.

(2)

(e) The plot is:



Yes. There is an upward trend.

(4)

[40]

QUESTION 5

(a)

	Height (cm) X_i	Mass (kg) Y_i	$(X_i - \overline{X})$	$Y_i (X_i - \overline{X})$	$(X_i - \overline{X})^2$	$\widehat{Y_i}$	$e_i^2 = (Y_i - \widehat{Y_i})^2$
	160	62	−10	−620	100	66	16
	160	68	−10	−680	100	66	4
	160	68	−10	−680	100	66	4
	165	70	−5	−350	25	71	1
	165	68	−5	−340	25	71	9
	165	74	−5	−370	25	71	9
	170	70	0	0	0	76	36
	170	82	0	0	0	76	36
	170	78	0	0	0	76	4
	175	77	5	385	25	81	16
	175	83	5	415	25	81	4
	175	82	5	410	25	81	1
	180	84	10	840	100	86	4
	180	86	10	860	100	86	0
	180	88	10	880	100	86	4
Total	**2 550**	**1 140**	**0**	**750**	**750**		**148**

Consider the simple linear regression $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X$

Then

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum y_i (x_i - \bar{x})}{d^2} \\ &= \frac{750}{750} \\ &= 1\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x} \\ &= \frac{1\,140}{15} - 1\left(\frac{2\,550}{15}\right) \\ &= 76 - 1\,(170) \\ &= 76 - 170 \\ &= -94\end{aligned}$$

$$
\begin{aligned}
s^2 &= \frac{\sum (y_i - \widehat{y}_i)^2}{n-2} \\
&= \frac{148}{13} \\
&\approx 11.3846
\end{aligned}
$$

(13)

(b) The confidence interval is

$$\widehat{\beta}_1 \pm t_{\alpha/2;n-2} \times \frac{s}{d}$$

$\widehat{\beta}_1 = 1$ $\quad t_{\alpha/2;n-2} = t_{0.05;13} = 1.771$

$d = \sqrt{750}$ $\quad s = \sqrt{11.3846} \approx 3.3741$

The 95% confidence interval for $\widehat{\beta}_1$ is

$$
\begin{array}{lll}
\widehat{\beta}_1 & \pm & t_{\alpha/2;n-2} \times \frac{s}{d} \\
1 & \pm & 1.771 \times \frac{11.3846}{\sqrt{750}} \\
1 & \pm & 0.7362 \\
(1 - 0.7362 & ; & 1 + 0.7362 \\
(0.2638 & ; & 1.7362)
\end{array}
$$

(5)

(c) $x_i = 178$

The expected mass is

$$
\begin{aligned}
\widehat{\text{Mass}} &= -94 + \text{height} \\
&= -94 + 1\,(178) \\
&= -94 + 178 \\
&= 84 \text{ kg}
\end{aligned}
$$

(1)

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

Now

$$\begin{aligned} SE &= S\sqrt{1 + \frac{1}{n} + \frac{\left(X_i - \overline{X}\right)^2}{d^2}} \\ &= 3.3741\sqrt{1 + \frac{1}{15} + \frac{(178 - 170)^2}{750}} \\ &= 3.3741\sqrt{1 + \frac{1}{15} + \frac{32}{375}} \\ &= 3.3741\sqrt{1.152} \\ &\approx 3.6215 \end{aligned}$$

The 95% confidence interval for the expected mass of a man who height is 178 cm tall is

$$\begin{array}{lcl} \widehat{\beta}_0 + \widehat{\beta}_1 X & \pm & t_{\alpha/2;n-2} \times S\sqrt{1 + \frac{1}{n} + \frac{\left(X - \overline{X}\right)^2}{d^2}} \\ 84 & \pm & 1.771 \times 3.6215 \\ 84 & \pm & 6.4137 \\ (84 - 6.4137) & ; & 84 + 6.4137) \\ (77.5863 & ; & 90.4137) \end{array}$$

(4)

(e) The X-values used in the construction of the regression line are 160 to 180. In this case, estimates will be outside the range of X-values used in the construction of the regression line. The limits might become unreliable as the relationship between X and Y outside this range is not known and may be different from the one found in the specified range.

(2)

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

Commands for the Output:

Start the JMP program

> *Enter* height *in the first column and label it* $Height\ (x)$.

> *Enter* mass *in the second column and label it* $Mass\ (y)$

To plot:

> *Choose* Analyze>Fit Y by X *with* $Height\ (x)$ *as* X *factor and* $Mass\ (y)$ *as* Y *response.*

> Click Ok.

Click on the ***Red*** *triangle on Bivariate Fit of* $Height\ (y)$ *by* $Mass\ (x)$.

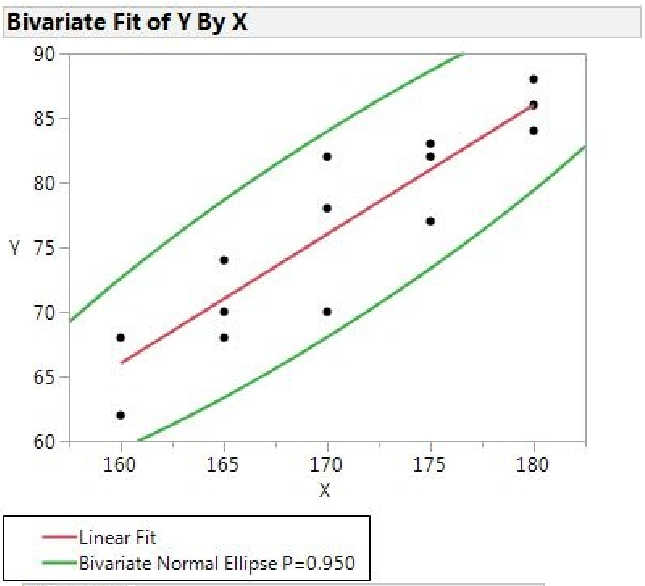
> *Choose* Fit Line

Click on the ***Red*** *triangle on Bivariate Fit of* $Height\ (y)$ *by* $Mass\ (x)$.

> *Choose* Density Ellipse then 0.95

> Click the triangle on *Bivariate Normal Ellipse P=0.95* to display the output.

The JMP output obtained is



Linear Fit

Y = -94 + 1*X

Summary of Fit

RSquare	0.835189
RSquare Adj	0.822512
Root Mean Square Error	3.37411
Mean of Response	76
Observations (or Sum Wgts)	15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	750.00000	750.000	65.8784
Error	13	148.00000	11.385	Prob > F
C. Total	14	898.00000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-94	20.96297	-4.48	0.0006*
X	1	0.123205	8.12	<.0001*

Bivariate Normal Ellipse P=0.950

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
X	170	7.319251	0.913887	<.0001*	15
Y	76	8.008924			

(10)

[35]

[**200]**