Tutorial Letter 203/2/2016

Applied Statistics II

STA2601

Semester 2

Department of Statistics

Solutions to Assignment 3

BAR CODE

Learn without limits.

UNISA university of south africa

QUESTION 1

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

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

Are they met? If we assume that the weight of one packet cannot influence the following packet, **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 and there also seems to be a systematic deviation around the line in the Normal Quantile Plot. Luckily the test is not too sensitive and we may proceed. (4)

(b) We have to test $H_0 : \mu = 0.500$ against $H_0 : \mu < 0.500$.

$n = \qquad \sum X_i = 12.136 \qquad \sum \left(X_i - \overline{X}\right)^2 = 0.01105216$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{12.136}{25} = 0.48544$$

(= "Mean" if we use output from "Moments")

$$S_X^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n-1} = \frac{0.01105216}{24} = 0.00046051$$

$$\therefore S_X = \sqrt{0.00046051} = 0.021459$$

(= "Std Dev" if we use output from "Moments")

$$\therefore T = \frac{\sqrt{25}\,(0.48544 - 0.500)}{0.021459} = \frac{-0.0728}{0.021459} \approx -3.3925$$

Since this is left-sided testing, we will reject H_0 if $T \leq -t_{0,05;25-1} = T \leq -t_{0,05;24} = -1.711$ (Stoker, Table III).

Since $-3.392\,5 < -1.711 \Longrightarrow$ we reject H_0. The packets weigh on average significantly less than 0.500 kg and it looks like cheating!

(8)

(c) If we know that $\sigma = 0.02$ we will use the test statistic

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

For this specific sample, it becomes

$$\begin{aligned} Z &= \frac{\sqrt{25}\,(0.48544 - 0.500)}{0.02} \\ &= -3.64. \end{aligned}$$

We will reject H_0 if $Z \leq -z_{0.05} = -1.645$.

Since $-3.64 < -1.645 \Longrightarrow$ we reject H_0 at the 5% level of significance. The packets weigh on average significantly less than 0.500 kg and it looks like cheating!

(6)

(d) A 90% two-sided confidence interval is computed as $\overline{X} - \left(\frac{S}{\sqrt{n}}\right)(t_{0.05;24}) < \mu < \overline{X} + \left(\frac{S}{\sqrt{n}}\right)(t_{0.05;24})$

where $0.48544 \pm \left(\frac{0.021459}{\sqrt{25}}\right)(1.711) = 0.48544 \pm (0.0042918)\,(1.711) = 0.4854 \pm 0.0073 = (0.4781;\ 0.4927)$

$\Longrightarrow 0.4781 \leq \mu \leq 0.4927.$

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 $\mu \leq 0.5$. (This means we reject $H_0 : \mu = 0.500$ which confirms our conclusion.)

(e) For the test in (b) we see the p-value is $0.0012 = P(t < -3.3924)$:

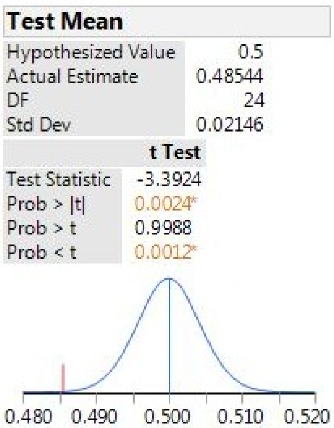


Figure 1: t-test

For the test in (c) we see the p-value is $0.0001 = P(z \leq -3.64)$:

Test Mean	
Hypothesized Value	0.5
Actual Estimate	0.48544
DF	24
Std Dev	0.02146
Sigma given	0.02

	z Test
Test Statistic	-3.6400
Prob > |z|	0.0003*
Prob > z	0.9999
Prob < z	0.0001*

Figure 2: z-test

We reject H_0 in both cases.

(12)

[15]

QUESTION 2

We are testing $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$ and we assume that $\mu_0 = 50 + 0.75\sigma$

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

$$\begin{aligned}
\delta &= \frac{\sqrt{n}(\mu - \mu_0)}{\sigma} \\
&= \frac{\sqrt{n}(50 + 0.75\sigma - 50)}{\sigma} \\
&= \sqrt{13}(0.75) = 2.7042 \\
\Longrightarrow \Phi = \frac{\delta}{\sqrt{2}} &= \frac{2.7042}{\sqrt{2}} = 1.912\,2
\end{aligned}$$

From table F:

For $n = 13,\quad \upsilon = 12,\quad \Phi = 1.912\,2 = 1.9$ the power is 0.69 at the 5% level of significance. (4)

(b) Let the probability of a Type II error $= \beta$.

$\beta = 1-$ power $= 1 - 0.69 = 0.31$ (for a Type I $= \alpha = 0,05$).

(1)

[5]

QUESTION 3

(a) A 99% confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{\frac{\alpha}{2};\ n_1+n_2} S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$S^2 = \frac{\left[(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2\right]}{(n_1 + n_2 - 2)}$$

$$= \frac{\left[399\,(257)^2 + 449\,(251)^2\right]}{400 + 450 - 2}$$

$$= \frac{54\,641\,000}{848} = 64\,435.14151$$

$$\therefore s = 253.8408;$$

$$t_{0.005;\ 848} = 2.576;$$

$$\left(\overline{X}_1 - \overline{X}_2\right) = 1252 - 1330 = -78;$$

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{400} + \frac{1}{450}} = \sqrt{0.00472222} = 0.0687.$$

So, the 99% confidence interval is:

$$-78 \pm 2.576\,(253.8408)\,(0.0687)$$
$$-78 \pm 44.9225$$
$$(-122.9225;\ -33.0775)\,.$$

(10)

(b) Since the confidence interval does not include the value 0, we can reject $H_0 : \mu_1 - \mu_2 = 0$ in favour of $H_1 : \mu_1 - \mu_2 \neq 0$ (two-sided!) and conclude that the mean usage per household has changed between the two years. (2)

(c) **Assumptions**:

(i) That the two samples are mutually independent (this was given) and that the observations in each sample are independent (also given by the fact that the sample was **random**);

(ii) The observations are **normally distributed.** Even without this restriction on the population the very large samples would ensure (from the central limit theorem) that $\overline{X}_1$ and $\overline{X}_2$ are normally distributed;

(iii) The two population variances are equal. (This can be verified by comparing)

We have to test:

$H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$

$$\begin{aligned} F &= \frac{\sigma_Y^2}{\sigma_X^2} \times \frac{S_X^2}{S_Y^2} \\ &= 1 \times \frac{257^2}{251^2} \\ &\approx 1.0484 \end{aligned}$$

The critical value are $F_{\alpha;n_x-1;n_y-1} = F_{0.005;399;449} = 1.28$ and

$F_{1-\alpha/2;n_1-1;n_2-1} = \frac{1}{F_{\alpha/2;n_2-1;n_1-1}} = \frac{1}{F_{0.005;449;399}} = \frac{1}{1.29} \approx 0.78$.(internet calculator).

Reject H_0 if $F < 0.78$ or $F > 1.29$.

Since $0.78 < 1.0484 < 1.29$, $\Longrightarrow$ we can not reject H_0 at the 5% level of significance. The assumption of equal variances is not violated.

So, yes, all the assumptions are satisfied.

(10)

(d) Because the assumption of independent samples is violated, we cannot apply the procedure used in (a) to these samples. (3)

[25]

QUESTION 4

OPTION A: Manual ANOVA Test:

(a)

Group	Control	A	B	C
n	10	10	10	10
$\sum X_{ij}$	11.8	9.3	9	12.4
$\overline{X}_i$	1.18	0.93	0.9	1.24
$\sum \left(X_{ij} - \overline{X}_i\right)^2$	0.156	0.091	0.065	0.084

$$\begin{aligned} S_1^2 &= \frac{1}{n_1 - 1} \sum \left(X_{1j} - \overline{X}_1\right)^2 \\ &= \frac{1}{10-1}(0.156) \\ &= \frac{1}{9}(0.156) \\ &= 0.017333 \end{aligned}$$

$$\begin{aligned} S_2^2 &= \frac{1}{n_2 - 1} \sum \left(X_{2j} - \overline{X}_2\right)^2 \\ &= \frac{1}{10-1}(0.091) \\ &= \frac{1}{9}(0.091) \\ &= 0.010111 \end{aligned}$$

$$\begin{aligned} S_3^2 &= \frac{1}{n_3 - 1} \sum \left(X_{3j} - \overline{X}_3\right)^2 \\ &= \frac{1}{10-1}(0.065) \\ &= \frac{1}{9}(0.065) \\ &= 0.007222 \end{aligned}$$

$$\begin{aligned} S_4^2 &= \frac{1}{n_4 - 1} \sum \left(X_{4j} - \overline{X}_4\right)^2 \\ &= \frac{1}{10-1}(0.084) \\ &= \frac{1}{9}(0.084) \\ &= 0.009333 \end{aligned}$$

We have to test

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

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

$$\begin{aligned} U &= \frac{\max S_i^2}{\min S_i^2} \\ &= \frac{0.017333}{0.007222} \\ &\approx 2.4000 \end{aligned}$$

From **Table E** with $k = 4$ and $\nu = n - 1 = 10 - 1 = 9$, we find that the critical value is 6.31. Reject H_0 if $U > 6.31$.

Since $2.4 < 6.31$, we cannot reject H_0 at the 5% level of significance and we may assume that the variances are equal.

(10)

(b) Assuming the populations to be independent and normally distributed and the population variances to be equal, 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 pair $p \neq q$.

We use the test statistic F by computing $F = \frac{MSTr}{MSE} = \frac{n \sum_{i=1}^{k} \left(\overline{X}_i - \overline{X}\right)^2 / (k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_i\right)^2 / (kn-k)}$

(which is defined for **sub-samples of equal sizes** $n_1 = n_2 = ... = n_4 = 10$).

We have

$$k = 4; \qquad n = 10; \qquad kn - k = 36; \qquad k - 1 = 3;$$

$$\overline{X}_1 = \frac{11.8}{10} = 1.18; \qquad SS_1 = \Sigma \left(X_{1i} - \overline{X}_1\right)^2 = 0.156$$

$$\overline{X}_2 = 0.93 \qquad SS_2 = 0.091$$

$$\overline{X}_3 = 0.90 \qquad SS_3 = 0.065$$

$$\overline{X}_4 = 1.24 \qquad SS_4 = 0.084$$

$$\overline{X} = \frac{42.5}{40} = 1.0625 \qquad SSE = SS_1 + ... + SS_4 = 0.396$$

$$MSE = S^2 = \frac{0.396}{36} = 0.011$$

Furthermore

$$\sum_{i=1}^{4} \left(\overline{X}_i - \overline{X}\right)^2 = (1.18 - 1.0625)^2 + .. + (1.24 - 1.0625)^2 = 0.089275$$

$$SSTr = n\Sigma \left(\overline{X}_i - \overline{X}\right)^2 = 10(0.089275) = 0.89275$$

$$MSTr = \frac{n\Sigma \left(\overline{X}_i - \overline{X}\right)^2}{(k-1)} = \frac{0.89275}{3} \approx 0.297583$$

$$F = \frac{MSTr}{MSE} = \frac{0.297583}{0.011} \approx 27.0530$$

The analysis is often summarized in tabular form called an ANOVA table.

ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Treatments	0.892750	3	0.297583	27.053
Error	0.396	36	0.011	
Toal	1.28875	39		

From table V (Stoker) we find $F_{0.05;3;36} = 2.88$. Reject H_0 if $F > F_{0.05;3;36} = 2.88$.

Since $27.053 > 2.88$, we reject H_0 at the 5% level of significance. This implies that $\mu_p \neq \mu_q$ for at least one pair $p \neq q$.

(10)

(c) Multiple comparisons on all pairs of means:

We may compute, for each pair of means $\overline{X}_p$ and $\overline{X}_q$, a t-statistic

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

and reject $H_0\left(p;q\right) : \mu_p = \mu_q$ in favour of
$H_1\left(p;q\right) : \mu_p \neq \mu_q$ if $\left|T_{pq}\right| > \sqrt{(k-1)\,F_{\alpha;k-1;kn-k}}$.

$$\begin{aligned} T_{pq} &= \frac{\sqrt{n}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}S} \\ &= \frac{\sqrt{10}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}\sqrt{0.011}} \\ &= (21.3201)\left(\overline{X}_p - \overline{X}_q\right) \end{aligned}$$

$$\left(\overline{X}_p - \overline{X}_q\right) = \frac{1}{21.3201}T_{pq}$$

We reject $H_0\left(p;q\right) : \mu_p = \mu_q$ if

$$\left|T_{pq}\right| > \sqrt{(k-1)\,F_{\alpha;k-1;kn-k}} = \sqrt{3(2.88)} = 2.9394$$

$$\therefore \left|\overline{X}_p - \overline{X}_q\right| > \frac{2.9394}{21.3201} = 0.1379$$

$$\begin{array}{llllll}
\left|\overline{X}_1 - \overline{X}_2\right| & = & |1.18 - 0.93| & = & 0.25 > 0.1379 & \Longrightarrow \mu_1 \neq \mu_2 \\
\left|\overline{X}_1 - \overline{X}_3\right| & = & |1.18 - 0.9| & & 0.28 > 0.1379 & \Longrightarrow \mu_1 \neq \mu_3 \\
\left|\overline{X}_1 - \overline{X}_4\right| & = & |1.18 - 1.24| & & 0.06 < 0.1379 & \Longrightarrow \mu_1 = \mu_4 \\
\left|\overline{X}_2 - \overline{X}_3\right| & = & |0.93 - 0.9| & & 0.03 < 0.1379 & \Longrightarrow \mu_2 = \mu_3 \\
\left|\overline{X}_2 - \overline{X}_4\right| & = & |0.93 - 1.24| & & 0.31 > 0.1379 & \Longrightarrow \mu_2 \neq \mu_4 \\
\left|\overline{X}_3 - \overline{X}_4\right| & = & |0.9 - 1.24| & & 0.34 > 0.1379 & \Longrightarrow \mu_3 \neq \mu_4
\end{array}$$

Now $\overline{X}_4 - \overline{X}_3 = 1.24 - 0.90 = 0.34$ (the largest observed difference); $\overline{X}_4 - \overline{X}_2 = 1.24 - 0.93 = 0.31$; $\overline{X}_1 - \overline{X}_3$ and $\overline{X}_4 - \overline{X}_2$ are all significant. We note, however, that $\overline{X}_1 = 1.18$ and $\overline{X}_4 = 1.24$ are rather close together, that $\overline{X}_2 = 0.93$ and $\overline{X}_3 = 0.90$ are close together, but that the two pairs are comparatively more different. We cannot reject $\mu_{\text{Control}} = \mu_{\text{Additive C}}$ and we cannot reject $\mu_{\text{Additive A}} = \mu_{\text{Additive B}}$.

(10)

[25]

Alternative for **QUESTION 4**

OPTION B: (JMP solution)

(a) Yes, it is reasonable to assume that the four groups may be considered as *independent groups* because cars in one group cannot influence cars in the other groups. (2)

(b) No formal tests for normality are included in the output and the graphical output shows only the "Means Diamonds" which is not a graphical test for normality. To perform the ANOVA we simply have to assume that the four groups may be considered as *coming from normal populations.* (4)

(c) 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 $p \neq q$.

We have to test from **Figure 5** we conclude that not any of the tests for the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ are significant at the 5% level of significance. (smallest p-value $= 0.3499 >> \alpha$). It looks as if the assumption of *equal population variances* may be assumed. (4)

(d) The ANOVA test:

(i) 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 pair $p \neq q$. (2)

(ii) From the output in **Figure 4** we see that $F = 27.0530$ with a **p-value** $< \mathbf{0.0001}$ **which is highly significant.** The **mean running times** for the different types of additives differ significantly (at any level of significance). This implies that $\mu_p \neq \mu_q$ for at least one pair $p \neq q$. (5)

(e) Most pairs of means differ significantly except for the pair *"Control"* and *"Additive C"* and the pair *"Additive A"* and *"Additive B"*. This is graphically confirmed by the "Means Diamonds" where we can see that *"Control"* and *"Additive C"* have almost identical pictures and their two circles overlap to a large extent on the "All Pairs Tukey-Kramer" display. The share the letter A and the Abs(Dif)-HSD is -0.06632. The same is true for the pair *"Additive A"* and *"Additive B"* (their two circles overlap almost completely). The share the letter B and the Abs(Dif)-HSD is -0.09632.

From the output of the formal statistical test (**Figure 6: Multiple Comparisons)** we see that the confidence interval for the difference of the mean running time (*"Additive C"* - *"Control"*) $= (-0.066324 : 0.1863236)$. We also see that the confidence interval for the difference of the mean running time (*"Additive A"* - *"Additive B"*) $= (-0.096324 : 0.1563236)$. These are the only intervals **which includes zero** and implies we cannot reject $\mu_{\text{Additive C}} = \mu_{\text{Control}}$ and we cannot reject $\mu_{\text{Additive A}} = \mu_{\text{Additive B}}$.

All the other intervals for the difference of the means are (positive value; positive value) **which excludes zero** and means **we reject** $\mu_p = \mu_q \implies \mu_p \neq \mu_q$. (8)

[25]

QUESTION 5

(a)

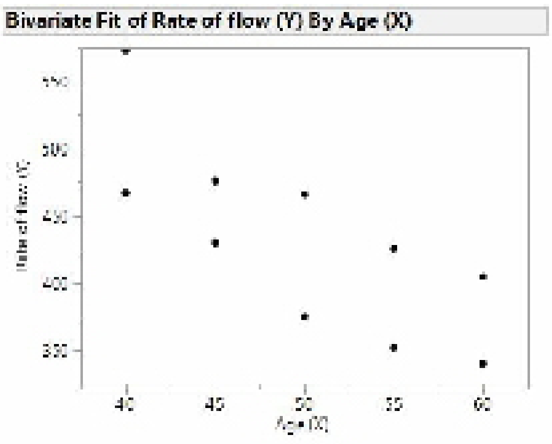


Figure 3: The Scatterplot

Linear regression seems to be applicable since there seem to be a strong negative relationship. (4)

(b)

x_i	y_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$y_i\ (x_i - \overline{x})$	$\hat{y}_i$
40	467	−10	100	−4 670	502.8
40	573	−10	100	−5 730	502.8
45	430	−5	25	−2 150	466.9
45	476	−5	25	−2 380	466.9
50	466	0	0	0	431
50	375	0	0	0	431
55	352	5	25	1 760	395.1
55	426	5	25	2 130	395.1
60	340	10	100	3 400	359.2
60	405	10	100	4 050	359.2
500	4310	0	500	−3 590	
$\overline{x} = 50$	$\overline{y} = 431$				

This additional column is needed for question (c). (10)

(c)

$$\hat{\beta}_1 = \frac{\Sigma y_i\ (x_i - \overline{x})}{\Sigma\ (x_i - \overline{x})^2} = \frac{-3590}{500} = -7.18$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 431 - (-7.18)\ (50)$$

$$= 790$$

So the least squares regression equation for Y on X is:

$$Y = 790 - 7.18X.$$

Using this equation we can compute $\hat{y}_i$ for each observed x_i-value (which is shown in the last column of the first computational table). (6)

(d)

x_i	y_i	$(y_i - \overline{y})^2$	$(x_i - \overline{x})^2$	$(y_i - \widehat{y}_i)^2$
40	467	1 296	100	1 281.64
40	573	20 164	100	4 928.04
45	430	1	25	1 361.61
45	476	2 025	25	82.81
50	466	1 225	0	1 225.00
50	375	3 136	0	3 136.00
55	352	6 241	25	1 857.61
55	426	25	25	954.81
60	340	8 281	100	368.64
60	405	676	100	2 097.64
		43 070	500	17 293.80

Thus $\sum_{i=1}^{10} \left(y_i - \hat{y}_i\right)^2 = 17\,293.80.$ (10)

(e)

$$\sum_{i=1}^{10} (y_i - \overline{y})^2 - b_1^2 \sum_{i=1}^{10} (x_i - \overline{x})^2$$
$$= 43\,070 - (-7.18)^2\, 500$$
$$= 17\,293.80$$

(3)

(f) Replace $X = 70$ in the regression equation in (b) then

$$Y = 790 - 7.18(70)$$
$$= 287.4$$

$\therefore$ The predicted rate of flow of blood through the kidney for a 70 year old person would be 287.4

Now $Var\widehat{Y}(x) = \sigma^2 \left[1 + \frac{1}{n} + (x - \overline{x})^2 / d^2\right]$ since $x = 70$ is a future observation (See p. 295.)

Please note that we need this sum in (d) to compute s^2.

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{10} \left(y_i - \hat{y}_i\right)^2 .$$

It is usually quite laborious to compute $\Sigma \left(y_i - \hat{y}_i\right)^2$. This computation is shown in the last column of the computational table for question (d). **Please note that** $(y_i - \overline{y})^2$ **and** $\left(y_i - \hat{y}_i\right)^2$ **are totally different values.**

$$\begin{aligned} \therefore s^2 &= \frac{17\,293.8}{8} \\ &= 2\,161.725 \\ \Longrightarrow s &= 46.4944 \end{aligned}$$

$\therefore s = 46.4944$ which is an estimate of σ.

So the approximate (or estimated) value for this variance when $x = 70$ is:

$$\begin{aligned} \widehat{Var\widehat{Y}}(x) &= s^2\left[1 + \frac{1}{10} + \frac{(70-50)^2}{500}\right] \\ &= 2\,161.725\,[1 + 0.1 + 0.8] \\ &= 2\,161.725\,(1.9) \\ &= 4\,107.2775 \end{aligned}$$

$\therefore$ Standard error of estimate is 64.088. (6)

(g) $t_{0.005;\,8} = 3.355$ and $s = 46.4944$

$$\begin{aligned} &\left(\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{0.005;\,8} S\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{d^2}}\right) \\ = &\left((790 - 7.18\,(60)) \pm 3.355\,(46.4944)\sqrt{\frac{1}{10} + \frac{(60-50)^2}{500}}\right) \\ = &\left((790 - 430.8) \pm 155.988712\sqrt{0.1 + 0.2}\right) \\ = &\ (359.2 \pm 85.4385) \\ = &\ (273.7615;\ 444.6385) \end{aligned}$$

(6)

(h) We have to test $H_0: \ \beta_1 = 0$ against

$H_1: \ \beta_1 \neq 0.$

The test statistic is $T = \dfrac{\hat{\beta}_1 - \beta_1}{s/d} \sim t_{n-2}$.

So

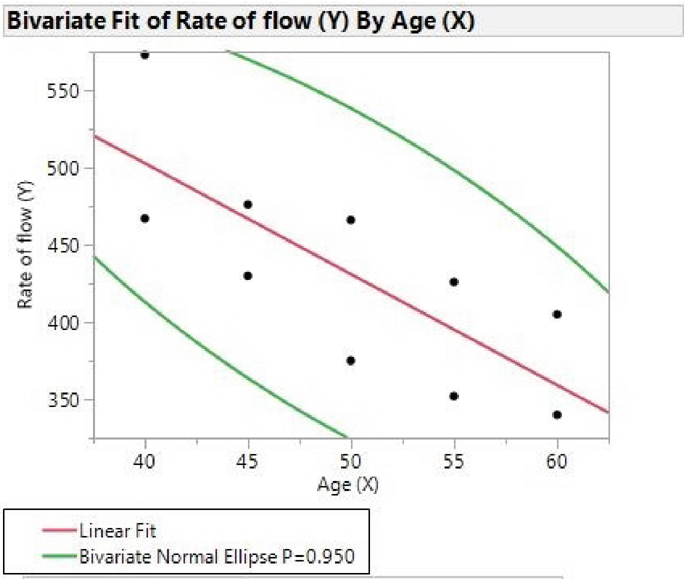
$$T = \frac{-7.18 - 0}{\sqrt{\frac{2\,161.725}{500}}} = \frac{-7.18}{\sqrt{4.32345}} = \frac{-7.18}{2.079290744} \approx -3.4531$$

We will reject H_0 at the 5% level of significance if $T \leq -t_{\frac{\alpha}{2};\,8}$ or if $T \geq t_{\frac{\alpha}{2};\,8}$ where $t_{0.025;\,8} = 2.306$.

Since $-3.4531 < -2.306$ we reject H_0 at the 5% level of significance and conclude that $\beta_1 \neq 0$.This means that the regression line is significant to explain the variability in y. (Only when $\beta_1 = 0$, does it imply that regression is meaningless.)

(7)

(i) The SAS JMP output is



Linear Fit

Rate of flow (Y) = 790 - 7.18*Age (X)

Summary of Fit

RSquare	0.598472
RSquare Adj	0.548281
Root Mean Square Error	46.49435
Mean of Response	431
Observations (or Sum Wgts)	10

Figure 4a: The Simple Linear Regression Model

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	3	1626.800	542.27	0.1731
Pure Error	5	15667.000	3133.40	Prob > F
Total Error	8	17293.800		0.9102
				Max RSq
				0.6362

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	25776.200	25776.2	11.9239
Error	8	17293.800	2161.7	Prob > F
C. Total	9	43070.000		0.0087*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	790	104.999	7.52	<.0001*
Age (X)	-7.18	2.079291	-3.45	0.0087*

Bivariate Normal Ellipse P=0.950

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Age (X)	50	7.45356	-0.77361	0.0087*	10
Rate of flow (Y)	431	69.17771			

Figure 4b: The Simple Linear Regression Model

(8)

[60]

[150]