Tutorial Letter 203/1/2014

Applied Statistics II

STA2601

Semester 1

Department of Statistics

Solutions to Assignment 3

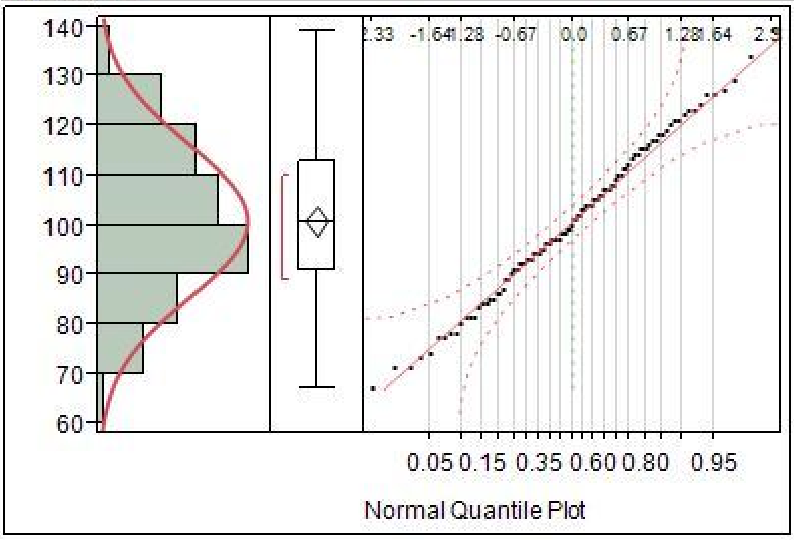
BAR CODE

QUESTION 1

(a)

Distributions

IQ Scores



Normal(100.91,15.4907)

Quantiles

100.0%	maximum	139
99.5%		139
97.5%		131.375
90.0%		121.9
75.0%	quartile	112.75
50.0%	median	100.5
25.0%	quartile	91
10.0%		80.1
2.5%		71
0.5%		67
0.0%	minimum	67

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	100.91	97.836311	103.98369
Dispersion	σ	15.490691	13.600933	17.995153

-2log(Likelihood) = 830.835562335798

(i) The assumptions are:

- observations are independent
- the data follows a normal distribution

Now based on the assumption of **independent observations** and the assumption that the IQ scores have a **normal distribution** (i.e., the sample comes from a normal population) we may assume that

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

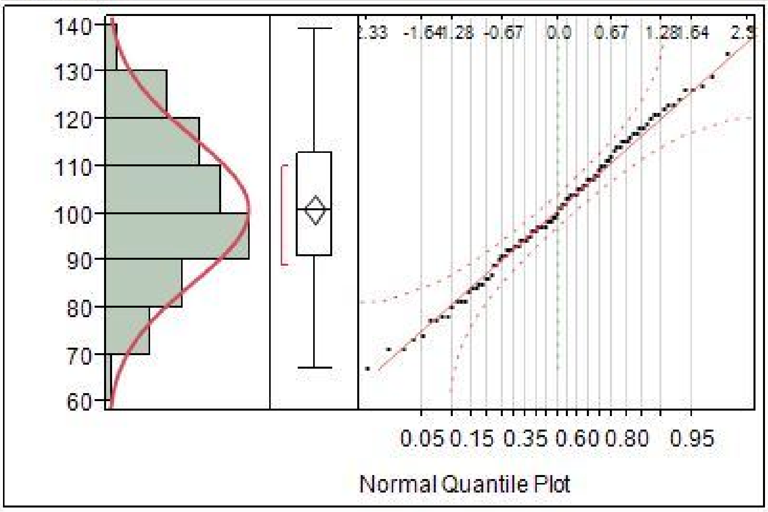
Are they met? The people were drawn randomly, thus the assumption of **independent observations is met.**

The normality assumption is not violated because from the JMP graphical output we see that the normal curve does fit the histogram very well. The box plot depicts a symmetric distribution and the points seem not to be deviating from the diagonal on the Normal Quantile Plot.

(ii) No.
We have to test $H_0 : \mu = 100$ against $H_1 : \mu \neq 100$.

Distributions

IQ Scores



Normal(100.91,15.4907)

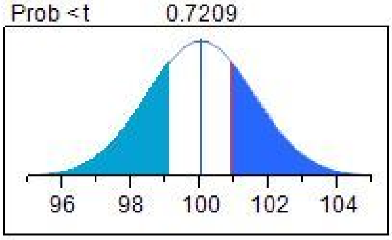
Summary Statistics

Mean	100.91
Std Dev	15.490691
Std Err Mean	1.5490691
Upper 95% Mean	103.98369
Lower 95% Mean	97.836311
N	100

Test Mean

Hypothesized Value	100
Actual Estimate	100.91
DF	99
Std Dev	15.4907

	t Test
Test Statistic	0.5874
Prob > |t|	0.5582
Prob > t	0.2791
Prob < t	0.7209



Method 1: Using the critical value approach

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{s} = \frac{\sqrt{100}\,(100.91 - 100)}{15.4907} \approx 0.5874$$

The critical value is

$$\begin{aligned}
t_{\alpha/2;n-1} &= t_{0.025;99} \\
&= 2.000 + \frac{39}{40}(1.984 - 2.000) \\
&= 2.000 + \frac{39}{40}(-0.016) \\
&= 2.000 - 0.0156 \\
&\approx 1.984
\end{aligned}$$

We will reject H_0 if $T \geq 1.984$ or $T \leq -1.984$ or if $|T| \geq 1.984$.

Since $-1.984 < 0.5874 < 1.984$, we do not reject H_0 at the 5% level of significance and conclude that $\mu = 100$, that is, the mean IQ score is 100.

Method II: Using the p-value approach

p-value $= 0.5582$. Since $0.5582 > 0.05$, we do not reject H_0 at the 5% level of significance and conclude that $\mu = 100$, that is, the mean IQ score is 100.

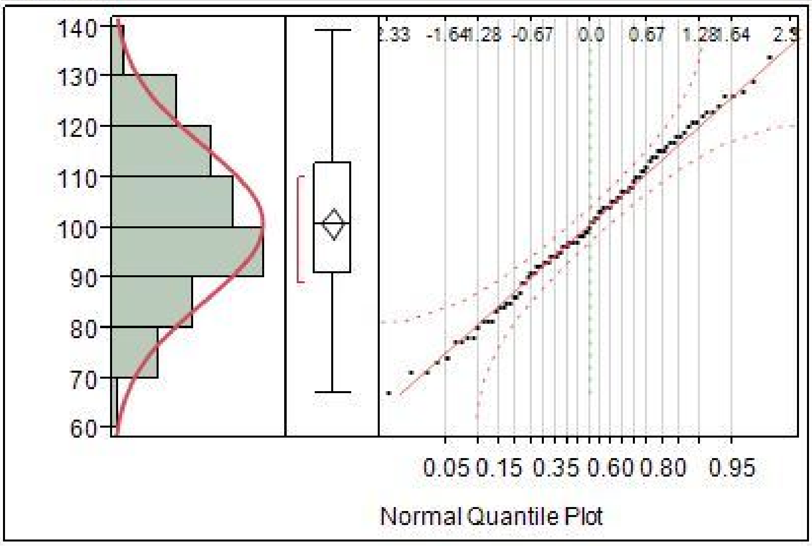
(iii) No.

We have to test $H_0 : \sigma = 15$

against $H_1 : \sigma \neq 15$

Distributions

IQ Scores



Normal(100.91,15.4907)

Summary Statistics

Mean	100.91
Std Dev	15.490691
Std Err Mean	1.5490691
Upper 95% Mean	103.98369
Lower 95% Mean	97.836311
N	100

Test Standard Deviation

Hypothesized Value	15
Actual Estimate	15.4907
DF	99

Test	ChiSquare
Test Statistic	105.5831
Min PValue	0.6136
Prob < ChiSq	0.6932
Prob > ChiSq	0.3068

Method 1: Using the critical value approach

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

$$\begin{aligned} U &= \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{\sigma^2} \\ &= \frac{(n-1)\, S_X^2}{\sigma^2} \\ &= \frac{99 \times 15.4907^2}{15^2} \\ &\approx 105.5831 \end{aligned}$$

The critical values are

$$\begin{aligned} &\chi^2_{1-\alpha/2;n-1} & ; \quad &\chi^2_{\alpha/2;n-1} \\ =\ &\chi^2_{0.975;99} & ; \quad =\ &\chi^2_{0.025;99} \\ =\ &65.6466 + \tfrac{9}{10}(74.2219 - 65.6466) & ; \quad =\ &118.136 + \tfrac{9}{10}(129.561 - 118.136) \\ =\ &65.6466 + 0.9(8.5753) & ; \quad =\ &118.136 + 0.9(11.425) \\ =\ &65.6466 + 7.7177 & ; \quad =\ &118.136 + 10.2825 \\ =\ &73.3644 & ; \quad =\ &128.4185 \end{aligned}$$

Reject H_0 if $U < 73.3644$ or $U > 128.4185$

Since $73.3644 < 105.5831 < 128.4185$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 15$.

Method II: Using the p-value approach

p-value $= 0.6136$. Since $0.6136 > 0.05$, we do not reject H_0 at the 5% level of significance and conclude that $\sigma = 15$.

(b) We have to test: $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$

$n_X = 100 \quad \overline{X} = 100.91 \quad S_X^2 = 239.972$

$n_Y = 60 \quad \overline{Y} = 105.21 \quad S_Y^2 = 240.6537$

The test statistic is

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

Now

$$
\begin{aligned}
S_p^2 &= \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2} \\
&= \frac{(100-1)239.972+(60-1)240.6537}{100+60-2} \\
&= \frac{23\,757.228+14\,198.5683}{158} \\
&= \frac{37\,955.7963}{158} \\
&\approx 240.2265589 \\
\Longrightarrow & \quad S_{pooled}=\sqrt{240.2265589}\approx 15.4992
\end{aligned}
$$

Then

$$
\begin{aligned}
T &= \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}} \\
&= \frac{(100.91-105.21)-(0)}{15.4992\sqrt{\frac{1}{100}+\frac{1}{60}}} \\
&= \frac{-4.3}{15.4992\sqrt{0.026666666}} \\
&= \frac{-4.3}{2.531008761} \\
&\approx -1.6989
\end{aligned}
$$

The critical value is $t_{\alpha;(n_1+n_2-2)}=t_{0.05;158}=1.645$. Reject H_0 if $T\le -1.645$.

Since $-1.6989 < -1.645$, we reject H_0 at the 5% level of significance and conclude that $\mu_X<\mu_Y$, i.e., mean IQ score of population B is higher than the mean IQ score of population A.

QUESTION 2

Group	A	B	C	D
n	3	3	3	3
$\sum X_{ij}$	139	145	153	128
$\overline{X}_i$	46.3333	48.3333	51	42.6667
$\sum\left(X_{ij}-\overline{X}_i\right)^2$	34.6667	20.6667	26	72.6667

(a)

$$\begin{aligned} S_1^2 &= \frac{1}{n_1-1}\sum\left(X_{1j}-\overline{X}_2\right)^2 \\ &= \frac{1}{3-1}(34.6667) \\ &= \frac{1}{2}(34.6667) \\ &\approx 17.3334 \end{aligned} \qquad \begin{aligned} S_2^2 &= \frac{1}{n_2-1}\sum\left(X_{2j}-\overline{X}_2\right)^2 \\ &= \frac{1}{3-1}(20.6667) \\ &= \frac{1}{2}(20.6667) \\ &\approx 10.3334 \end{aligned}$$

$$\begin{aligned} S_3^2 &= \frac{1}{n_3-1}\sum\left(X_{3j}-\overline{X}_3\right)^2 \\ &= \frac{1}{3-1}(26) \\ &= \frac{1}{2}(26) \\ &= 13 \end{aligned} \qquad \begin{aligned} S_4^2 &= \frac{1}{n_4-1}\sum\left(X_{4j}-\overline{X}_4\right)^2 \\ &= \frac{1}{3-1}(72.6667) \\ &= \frac{1}{2}(72.6667) \\ &\approx 36.3334 \end{aligned}$$

From the computations above it, follows that $S_1^2 = 17.3334$; $S_2^2 = 10.3334$; $S_3^2 = 13$ and $S_4^2 = 36.3334$.

(b) (i) Ordinary average = $\dfrac{17.3334+10.3334+13+36.3334}{4} = \dfrac{77.0002}{4} = 19.2501$

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

For this $ANOVA$ problem, we have $k = 4$ (there are four groups) and $n = 3$ (the number of observations in each sample).

$$\begin{aligned} SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_i\right)^2 \\ &= 34.6667+20.6667+26+72.6667 \\ &= 154.0001 \end{aligned}$$

$$\begin{aligned} \therefore MSE &= \frac{154.0001}{4(3)-4} \\ &= \frac{154.0001}{8} \\ &= 19.2500. \end{aligned}$$

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.

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

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 $p \neq q$.

$$\begin{aligned} U &= \frac{\max S_i^2}{\min S_i^2} \\ &= \frac{36.3334}{10.3334} \\ &\approx 3.5161 \end{aligned}$$

From **Table E** with $k = 4$ and $\nu = n - 1 = 3 - 1 = 2$, we find that the critical value is 142. Reject H_0 if $U > 142$.

Since $3.5161 < 142$, 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**".)

(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 $F = \dfrac{MST_r}{MSE} \sim F_{k-1;kn-k}$

$$MST_r = \frac{n\sum_{i=1}^{k}\left(\overline{X}_i - \overline{X}\right)^2}{k-1}$$

$$\text{where } \overline{X} = \frac{\sum\sum X_{ij}}{N} = \frac{565}{12} \approx 47.0833 \qquad \text{(overall mean);}$$

$$\begin{aligned}
\text{and } \sum\left(\overline{X}_i - \overline{X}\right)^2 &= (46.3333 - 47.0833)^2 + \cdots + (42.6667 - 47.0833)^2 \\
&= (-0.75)^2 + (1.25)^2 + (3.9167)^2 + (-4.4166)^2 \\
&= 0.5625 + 1.5625 + 15.34053889 + 19.50635556 \\
&\approx 36.9719
\end{aligned}$$

$$\therefore MST_r = \frac{3\,(36.9719)}{4-1} = \frac{110.9157}{3} = 36.9719$$

We already know that $MSE = 19.2500$ (see question (b)(ii)).

$$\begin{aligned}
\therefore F &= \frac{MST_r}{MSE} \\
&= \frac{36.9719}{19.25} \\
&\approx 1.9206.
\end{aligned}$$

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

The critical value is $F_{0,05;3;8} = 4.07$. Reject H_0 if $F > 4.07$.

Since $1.9206 < 4.07$, we do not reject H_0 at the 5% level of significance and conclude that the population means of the four detergents do not differ, that is, $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

(Note that we reach the same conclusion with the JMP output under the heading: "**Analysis of Variance**" if we consider "Prob $>$ F"$= 0.2048$)

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

$$T_{pq} = \frac{\overline{X}_p - \overline{X}_q}{S_{pooled}\sqrt{1/n + 1/n}} = \frac{\sqrt{n}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}S} = \frac{\sqrt{8}\left(\overline{X}_p - \overline{X}_q\right)}{\sqrt{2}\sqrt{MSE}}.$$

We reject $H_0(p;q)$ if

$$\left|T_{pq}\right| > \sqrt{(k-1)\,F_{\alpha;k-1;kn-k}} = \sqrt{3\,(4.07)} \approx 3.4943$$

This implies that we reject H_0 if

$$\begin{aligned} \frac{\sqrt{3}\left|\overline{X}_p - \overline{X}_q\right|}{\sqrt{2}\sqrt{0,924\,11}} &\geq 3.4943 \\ \text{i.e. if } \left|\overline{X}_p - \overline{X}_q\right| &\geq \frac{(3.4943)\,\sqrt{2}\sqrt{19.25}}{\sqrt{3}} = \frac{21.68156131}{1.732050808} = 12.5179 \end{aligned}$$

$$\begin{aligned} \left|\overline{X}_1 - \overline{X}_2\right| &= 2 < 12.5179 \implies \mu_1 = \mu_2 \\ \left|\overline{X}_1 - \overline{X}_3\right| &= 4.6667 < 12.5179 \implies \mu_1 = \mu_3 \\ \left|\overline{X}_1 - \overline{X}_4\right| &= 3.6666 < 12.5179 \implies \mu_1 = \mu_4 \\ \left|\overline{X}_2 - \overline{X}_3\right| &= 2.6667 < 12.5179 \implies \mu_2 = \mu_3 \\ \left|\overline{X}_2 - \overline{X}_4\right| &= 5.6666 < 12.5179 \implies \mu_2 = \mu_4 \\ \left|\overline{X}_3 - \overline{X}_4\right| &= 8.3333 < 12.5179 \implies \mu_3 = \mu_4 \end{aligned}$$

All pairs of means are not significantly different from each other, i.e., they are equal.

QUESTION 3

(a)

Subject	Before	After	Y_i =After - Before
1	27	29	2
2	21	32	11
3	34	29	-5
4	24	27	3
5	30	31	1
6	27	26	-1
7	33	35	2
8	31	30	-1
9	22	29	7
10	27	28	1
11	33	36	3
12	17	15	-2
13	25	28	3
14	26	26	0
15	23	26	3

$n = 15 \quad \sum Y_i = 27 \quad \sum \left(Y_i - \overline{Y}\right)^2 = 198.4$

We have to test:

$H_0 : \mu_d = 0$ against

$H_1 : \mu_d > 0$

$$\begin{aligned} \overline{Y} &= \frac{1}{n}\sum Y_i & S_y^2 &= \frac{1}{n-1}\sum \left(Y_i - \overline{Y}\right)^2 \\ &= \frac{1}{15}(27) & &= \frac{1}{14}(198.4) \\ &= 1.8 & &= 14.17142857 \end{aligned}$$

$$\begin{aligned} \Longrightarrow S_y &= \sqrt{14.17142857} \\ &\approx 3.7645 \end{aligned}$$

The test statistic is

$$\begin{aligned} T &= \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{S_y} \\ &= \frac{\sqrt{15}\,(1.8 - 0)}{3.7645} \\ &= \frac{6.971370023}{3.7645} \\ &\approx 1.8519. \end{aligned}$$

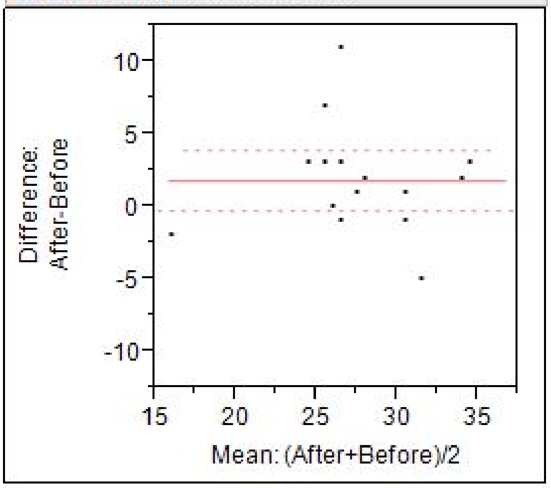
The critical value is $t_{\alpha;(n-1)} = t_{0.05;14} = 1.761$. Reject H_0 if $T > 1.761$.

Since $1.8519 > 1.761$, we reject H_0 at the 5% level of significance and conclude that learning is shown by an increase in score.

(b) The output is

Matched Pairs

Difference: After-Before



After	28.4667	t-Ratio	1.851874
Before	26.6667	DF	14
Mean Difference	1.8	Prob > |t|	0.0852
Std Error	0.97199	Prob > t	0.0426*
Upper 95%	3.88471	Prob < t	0.9574
Lower 95%	-0.2847		
N	15		
Correlation	0.69908		

(c) Paired data since the two observations were taken from the same individual.

QUESTION 4

(a) Start the *JMP* program

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

(make sure to change the scale to nominal)

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

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

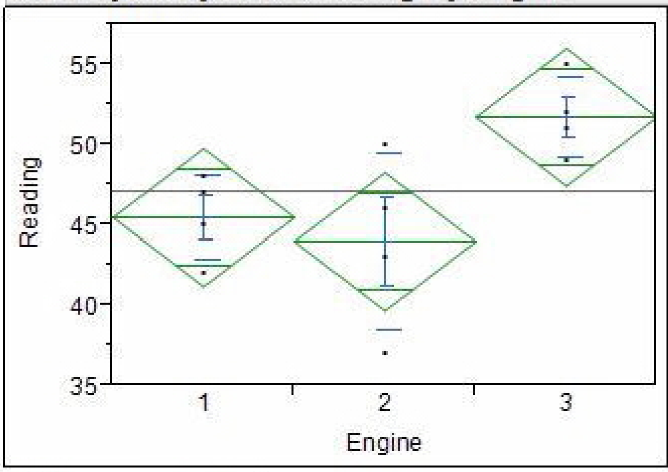
> Choose Analyze>Fit Y by X with *Engine* as X factor and *Reading* as Y response.

> Click Ok.

⟹ Then on the Oneway Analysis of *Reading* By *Engine* click on the **Red** triangle

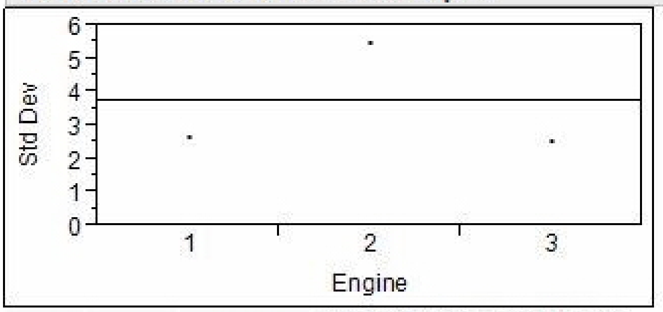
> Choose Unequal Variances

Oneway Analysis of Reading By Engine



Oneway Anova

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
1	4	2.645751	2.000000	2.000000
2	4	5.477226	4.000000	4.000000
3	4	2.500000	1.750000	1.750000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	1.2740	2	9	0.3257
Brown-Forsythe	1.3774	2	9	0.3007
Levene	1.4600	2	9	0.2824
Bartlett	1.0784	2	.	0.3401

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
6.5121	2	5.6461	0.0342*

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:

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

Using the Levene's test, p-value $= 0.2824$. Since $0.2824 > 0.05 \Longrightarrow$ we can not reject H_0 at the 5% level of significance. The assumption of equal variances is not violated.

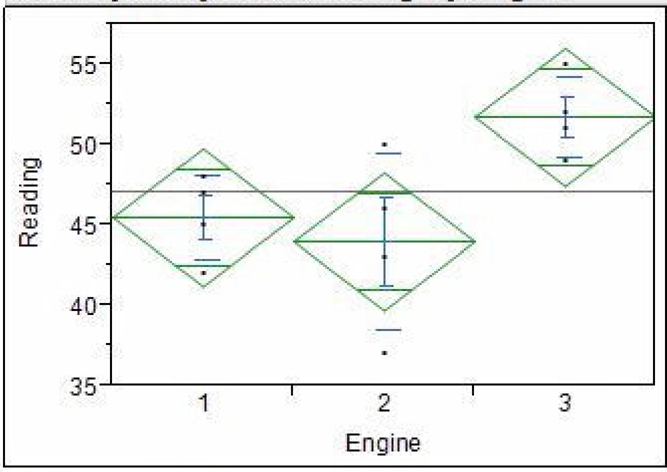
(b) $\Longrightarrow$ Click on the triangle "Tests that the variances are equal" to hide the output.

$\Longrightarrow$ Then click on the **Red** triangle on Oneway Analysis of *Reading* by *Engine*.

$>$ Choose Means/ANOVA

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

Oneway Analysis of Reading By Engine



Oneway Anova

Summary of Fit

Rsquare	0.510223
Adj Rsquare	0.401384
Root Mean Square Error	3.796929
Mean of Response	47.08333
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Engine	2	135.16667	67.5833	4.6879	0.0403*
Error	9	129.75000	14.4167		
C. Total	11	264.91667			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	4	45.5000	1.8985	41.205	49.795
2	4	44.0000	1.8985	39.705	48.295
3	4	51.7500	1.8985	47.455	56.045

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%
1	4	45.5000	2.64575	1.3229	41.290	49.710
2	4	44.0000	5.47723	2.7386	35.285	52.715
3	4	51.7500	2.50000	1.2500	47.772	55.728

For your information:

On the plot, the dots shows the response for each *Engine*. The line across the middle is the grand mean. The diamonds give a 95% confidence interval for each *Engine* 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) $H_0 : \mu_1 = \mu_2 = \mu_3$ against
$H_1 : \mu_p \neq \mu_q$ for at least one $p \neq q$.

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

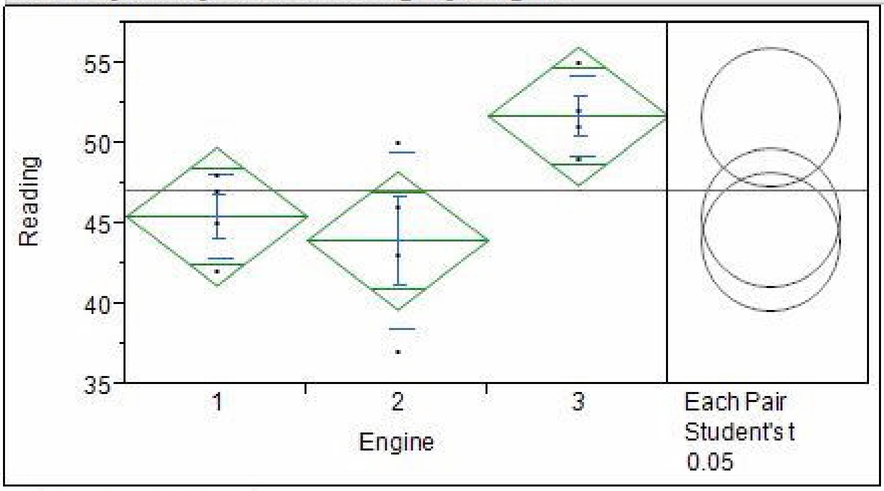
(iii) From the output: Computations for ANOVA we see that $F = 4.6879$ which is significant with a p–value of 0.0403. Since $0.0403 < 0.05$ we reject H_0 in favour of H_1 at the 5% level of significance and conclude that $u_p \neq \mu_q$ for at least one pair $p \neq q$, that is, the mean reading of the engines are not the same.

(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 *Reading* by *Engine*.

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

Oneway Analysis of Reading By Engine



Oneway Anova

Means Comparisons

Comparisons for each pair using Student's t

Confidence Quantile

t	Alpha
2.26216	0.05

LSD Threshold Matrix

Abs(Dif)-LSD

	3	1	2
3	-6.0735	0.1765	1.6765
1	0.1765	-6.0735	-4.5735
2	1.6765	-4.5735	-6.0735

Positive values show pairs of means that are significantly different.

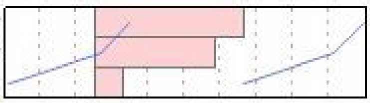
Connecting Letters Report

Level			Mean
3	A		51.750000
1		B	45.500000
2		B	44.000000

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	2	7.750000	2.684834	1.67648	13.82352	0.0180*
3	1	6.250000	2.684834	0.17648	12.32352	0.0449*
1	2	1.500000	2.684834	-4.57352	7.57352	0.5900

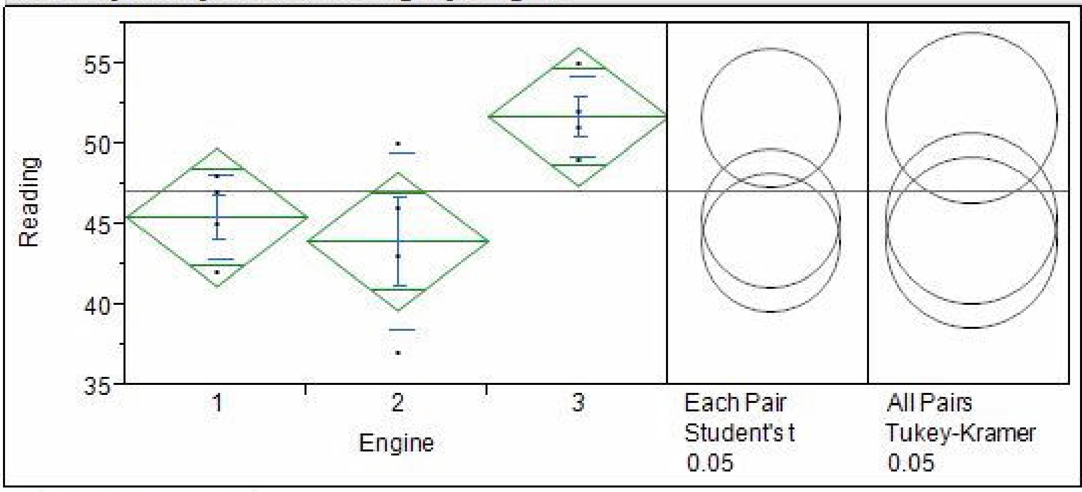


Interpretation:

Yes The $Abs(Dif) - LSD$ for the pair 12 is negative. They all share the letter B and the confidence interval is $(-4.5735; 7.5735)$ and it includes zero. We conclude that the means are not significantly different from each other. Thus we conclude that $\mu_1 = \mu_2 \neq \mu_3$.

(d)

Oneway Analysis of Reading By Engine



Oneway Anova

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
2.79201	0.05

LSD Threshold Matrix

Abs(Dif)-HSD

	3	1	2
3	-7.4961	-1.2461	0.2539
1	-1.2461	-7.4961	-5.9961
2	0.2539	-5.9961	-7.4961

Positive values show pairs of means that are significantly different.

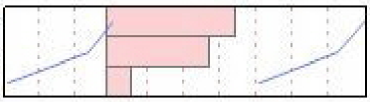
Connecting Letters Report

Level			Mean
3	A		51.750000
1	A	B	45.500000
2		B	44.000000

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
3	2	7.750000	2.684834	0.25392	15.24608	0.0431*
3	1	6.250000	2.684834	-1.24608	13.74608	0.1026
1	2	1.500000	2.684834	-5.99608	8.99608	0.8447



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.

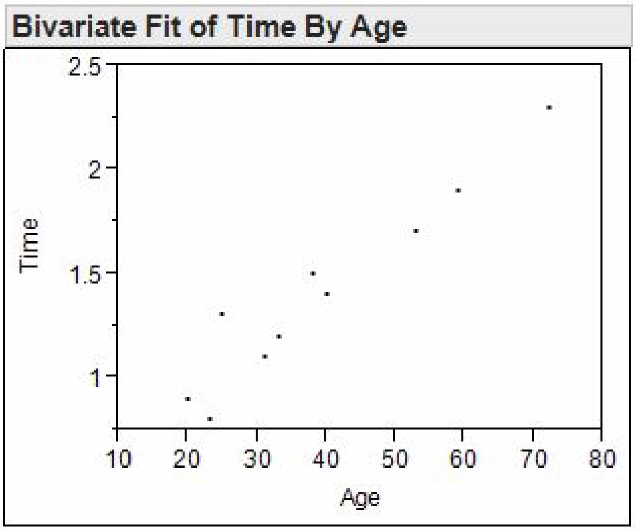
The Turkey–Kramer HSD that are shown in the JMP out perform individual comparisons that make adjustments for multiple test.

Confidence intervals that do not include zero imply that the pairs of means differ significantly. All pairs include zero except the pair 23. The confidence interval for the pair is $(0.2539 : 15.2461)$. This is the only interval that do not include zero and it means we reject the null hypothesis and conclude that $\mu_2 \neq \mu_3$. The $p-$value is 0.0431, thus less than 0.05 and leading to the rejection of the null hypothesis of equal means.

Confirming this is the **Abs(Dif)-LSD** which is 0.2539. Since it is positive, the means are significantly different. (Recall a negative value of **Abs(Dif)-LSD** means the groups are not significantly different from each other.)

QUESTION 5

(a)



(b) $H_0 : \rho = 0$ against $H_1 : \rho > 0$

$n = 10$ $R = 0.964$

The test statistic is

$$\begin{aligned} T &= \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \\ &= \frac{0.964\sqrt{10-2}}{\sqrt{1-0.964^2}} \\ &= \frac{2.726603748}{\sqrt{0.070704}} \\ &= \frac{2.726603748}{0.265902237} \\ &\approx 10.2542 \end{aligned}$$

The critical value is $t_{\alpha;(n-2)} = t_{0.005;8} = 3.355$. Reject H_0 if $T \geq 3.355$.

Since $10.2542 > 3.355$, we reject H_0 at the 5% level of significance and conclude that $\rho > 0$, that is, there is a significant positive correlation.

(c)

$$\begin{aligned}
b &= \frac{n\,\Sigma a_i t_i - (\Sigma t_i)\,(\Sigma a_i)}{n\,\Sigma a_i^2 - (\Sigma a_i)^2} \\
&= \frac{10\,(623.4) - (14.1)\,(394)}{10\,(18\,122) - (394)^2} \\
&= \frac{6\,234 - 5\,555.4}{181\,220 - 155\,236} \\
&= \frac{678.6}{25\,984} \\
&= 0.026116071 \\
&\approx 0.0261
\end{aligned}$$

$$\begin{aligned}
a &= \frac{\Sigma t_i - b\,(\Sigma a_i)}{n} \\
&= \frac{14.1 - 0.026116071\,(394)}{10} \\
&= \frac{14.1 - 10.28973214}{10} \\
&= \frac{3.810267857}{10} \\
&= 0.381026785 \\
&\approx 0.381
\end{aligned}$$

The estimated regression line is $\widehat{Time} = 0.381 + 0.0261\ Age$.

(d) $a_i = 45$

Then

$$\begin{aligned}
\widehat{Time} &= 0.381 + 0.0261 Age \\
&= 0.381 + 0.0261(45) \\
&= 0.381 + 1.1745 \\
&= 1.5555 \\
&\approx 1.56 \text{ seconds}
\end{aligned}$$

$\therefore$ The predicted response time is 1.56 seconds.

(e) $\widehat{\beta}_1 = 0.0261$

For every increase in age of 1 year, the response time increase by 0.0261 seconds.

(f) The standard error of the estimate is given by $SE = S\sqrt{\frac{1}{n} + \frac{(a_i - \overline{a})^2}{d^2}}$.

X	Y	$\widehat{\beta}_0 + \widehat{\beta}_1 X$	$e_i = Y_i$-$(\widehat{\beta}_0 + \widehat{\beta}_1 X)$	e_i^2
20	0.9	0.903348	−0.003348	0.000011
23	0.8	0.981696	−0.181696	0.033013
25	1.3	1.033929	0.266071	0.070794
31	1.1	1.190625	−0.090625	0.008213
33	1.2	1.242857	−0.042857	0.001837
38	1.5	1.373438	0.126562	0.016018
40	1.4	1.425670	−0.02567	0.000659
53	1.7	1.765179	−0.065179	0.004248
59	1.9	1.921875	−0.021875	0.000479
72	2.3	2.261384	0.038616	0.001491
				0.136763

$$\begin{aligned}
MSE &= s^2 \\
&= \frac{1}{n-2}\sum\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X\right)^2 \\
&= \frac{1}{10-2}(0.136763) \\
&= \frac{1}{8}(0.136763) \\
&= 0.017095375 \\
\Longrightarrow s &= \sqrt{0.017095375} \\
&\approx 0.1307
\end{aligned}$$

$$\begin{aligned}
d^2 &= \sum\left(X_i - \overline{X}\right)^2 \\
&= \Sigma a_i^2 - \frac{(\Sigma a_i)^2}{n} \\
&= 18\,122 - \frac{(394)^2}{10} \\
&= 2\,598.4
\end{aligned}$$

$$\begin{aligned}
\overline{a}_i &= \frac{394}{10} \\
&= 39.4
\end{aligned}$$

Then

$$
\begin{aligned}
SE &= S\sqrt{\frac{1}{n} + \frac{(a_i - \overline{a})^2}{d^2}} \\
&= 0.1307\sqrt{\frac{1}{10} + \frac{(45 - 39.4)^2}{2\,598.4}} \\
&= 0.1307\sqrt{0.1 + \frac{31.36}{2\,598.4}} \\
&= 0.1307\sqrt{0.1 + 0.012068965} \\
&= 0.1307\sqrt{0.112068965} \\
&\approx 0.0438
\end{aligned}
$$

(g) The confidence interval for β_1 is

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

$\widehat{\beta}_1 =$ $\qquad$ $t_{\alpha/2;n-2} = t_{0.025;8} = 2.306$

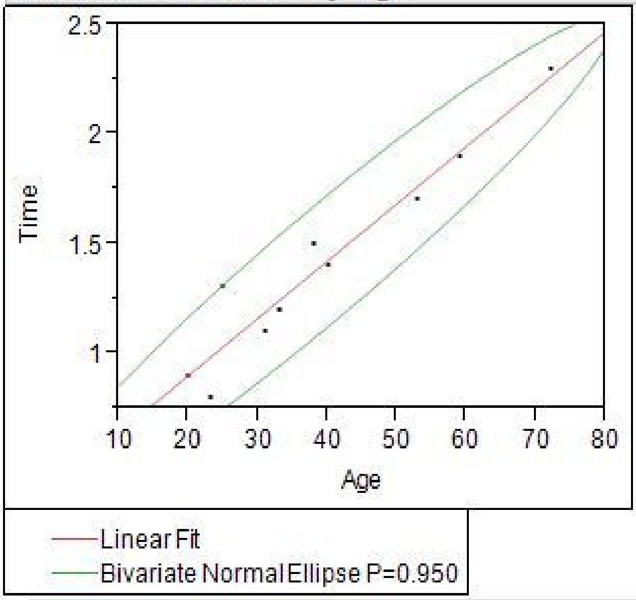
$d = \sqrt{2\,598.4} \approx 50.9745$ $\qquad$ $s = 0.1307$

Thus, the 95% confidence interval for the slope $\widehat{\beta}_1$ is 0.0261

$$
\begin{array}{lcl}
\widehat{\beta}_1 & \pm & t_{\alpha/2;n-2} \times \dfrac{s}{d} \\
0.0261 & \pm & 2.306 \times \dfrac{0.1307}{50.9745} \\
0.0261 & \pm & 2.306 \times 0.002564027 \\
0.0261 & \pm & 0.0059 \\
(0.0261 - 0.0059 & ; & 0.0261 + 0.0059) \\
(0.0202 & ; & 0.032)
\end{array}
$$

(h) The JMP output is

Bivariate Fit of Time By Age



Linear Fit

Time = 0.3810268 + 0.0261161*Age

Summary of Fit

RSquare	0.928359
RSquare Adj	0.919403
Root Mean Square Error	0.130749
Mean of Response	1.41
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.7722366	1.77224	103.6673
Error	8	0.1367634	0.01710	Prob > F
C. Total	9	1.9090000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.3810268	0.109192	3.49	0.0082*
Age	0.0261161	0.002565	10.18	<.0001*

Correlation

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Age	39.4	16.9915	0.963514	<.0001*	10
Time	1.41	0.460555			