Math 343 - Homework 3

Preston Duffield
Western Washington University
April 20, 2023

Question 1

a)

The sample mean of μ_2 can be estimated by $\hat{\mu}_2 = \bar{y}_{2\cdot} = 5753.3$.

b)

The grand sample mean, μ, can be estimated by:

$$\begin{aligned}\hat{\mu} = \bar{y}_{\cdot\cdot} &= \frac{5635.3 + 5753.3 + 4527.3 + 3442.3}{4} \\ &= 4839.55\end{aligned}$$

c)

The treatment effect τ_3 can be estimated by:

$$\begin{aligned}\hat{\tau}_3 &= \hat{\mu}_i - \hat{\mu} \\ &= 4527.3 - 4839.55 \\ &= -312.25\end{aligned}$$

d)

An estimate of the common population variance, σ^2, is:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{(5-1)509.6^2 + (5-1)432.3^2 + (5-1)356.4^2 + (5-1)505.5^2}{(5-1)+(5-1)+(5-1)+(5-1)} \\ &= 207281.665\end{aligned}$$

e)

By noting that $\hat{\sigma}^2$ can be used as MSE, a 90% confidence interval for μ_4 is given by:

$$\begin{aligned}
\bar{y}_{4\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}} &\leq \mu_4 \leq \bar{y}_{4\cdot} + t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}} \\
3442.3 - t_{.05,16}\sqrt{\frac{207281.665}{5}} &\leq \mu_4 \leq 3442.3 + t_{.05,16}\sqrt{\frac{207281.665}{5}} \\
3442.3 - 1.746 \cdot 203.6082832 &\leq \mu_4 \leq 3442.3 + 1.746 \cdot 203.6082832 \\
3086.799938 &\leq \mu_4 \leq 3797.800062
\end{aligned}$$

Therefore, were are 90% confident that the true mean of concrete formula 4 is between 3086.79 and 3797.80.

f)

A 95% confidence interval for $\mu_1 - \mu_3$ is given by:

$$\begin{aligned}
\bar{y}_{1\cdot} - \bar{y}_{3\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{2 \cdot MSE}{n}} &\leq \mu_1 - \mu_3 \leq \bar{y}_{1\cdot} - \bar{y}_{3\cdot} + t_{\alpha/2,N-a}\sqrt{\frac{2 \cdot MSE}{n}} \\
5635.3 - 4527.3 - t_{.025,16}\sqrt{\frac{2 \cdot 207281.665}{5}} &\leq \mu_1 - \mu_3 \leq 5635.3 - 4527.3 + t_{.025,16}\sqrt{\frac{2 \cdot 207281.665}{5}} \\
1108 - 2.120 \cdot 287.9455956 &\leq \mu_1 - \mu_3 \leq 1108 + 2.120 \cdot 287.9455956 \\
497.5553373 &\leq \mu_1 - \mu_3 \leq 1718.444663
\end{aligned}$$

Therefore, were are 95% confident that the true difference in means of concrete formula 1 and concrete formula 3 is between 497.55 and 1718.44.

Question 2

a)

H_0: All means are equal.
H_a: Not all means are equal.
Performing an F test at we can see that the P-value= $0.000 < \alpha = .05$,

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Technique	3	489740	163247	12.73	0.000
Error	12	153908	12826		
Total	15	643648			

Figure 1: The output of the One-way ANOVA from Minitab.

therefore we can conclude the following. There is enough statistical evidence to support the hypothesis that not all means are equal.

b)

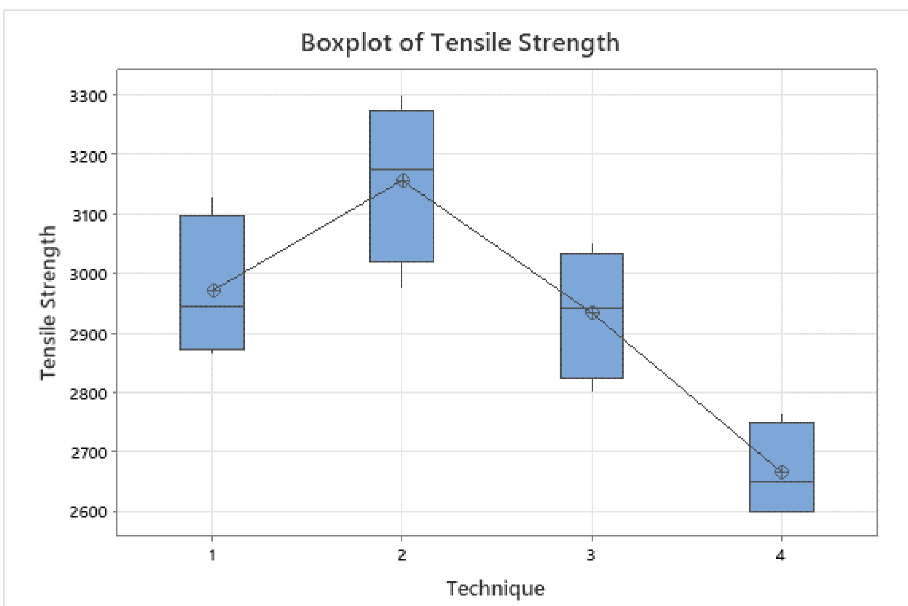


Figure 2: The Boxplot of tensile strength of the One-way ANOVA test from Minitab.

Judging by the boxplot alone, It would seem that the means are significantly different. This is consistent with the above hypothesis test. It looks as though the mean for treatment 1 and treatment 3 could be similar, in a pairwise comparison.

c)

The results of the Fisher's LSD test can be summarized in the graphical results below. Note that a line under the treatments indicates that they are not significantly different.

Mixing Technique	**(4)**	**(3)**	**(1)**	**(2)**

d)

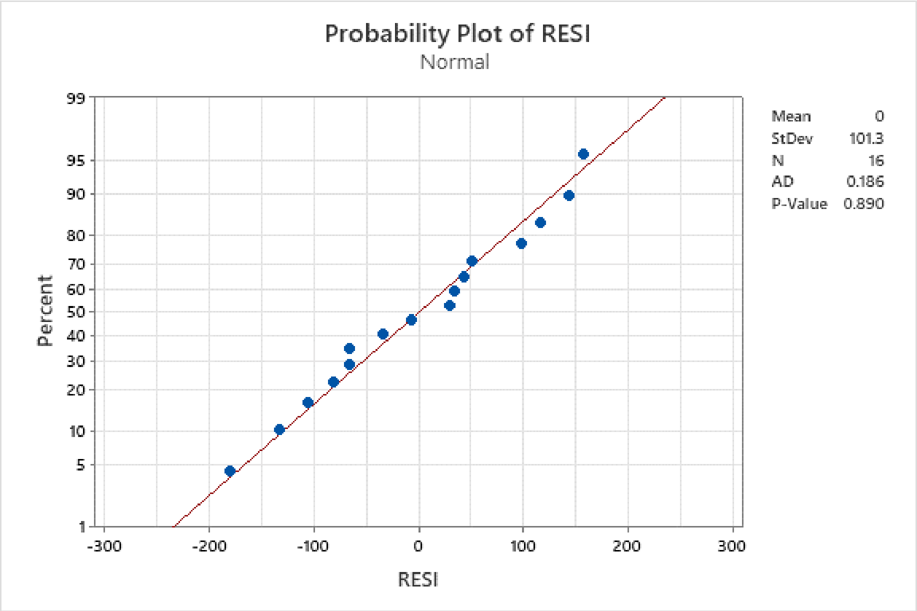


Figure 3: The normal probability plot of the residuals from Minitab.

H_0: The data are drawn from a normal disribution.
H_a: The data are not drawn from a normal disribution.

Since the P-value is very large (0.890), we can conclude the following. The evidence of the data is consistent with the data being drawn from a normal disribution.

e)

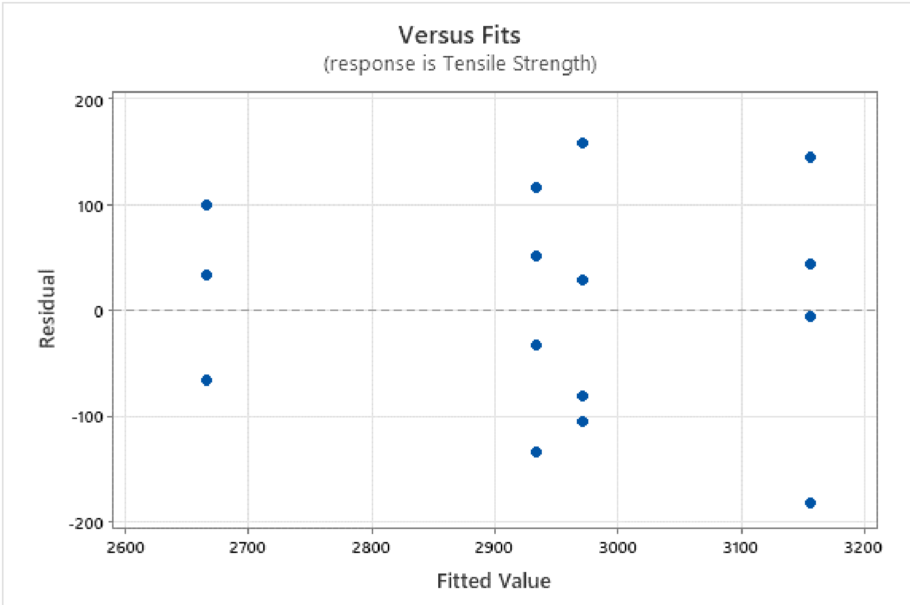


Figure 4: The residuals versus the predicted tensile strength from Minitab.

This graph indicates that there is not heteroskedasticity present, that along with the conclusion that the data is drawn from a normal distribution indicates that the model assumptions are verified and the hypothesis test is valid.

Question 3

a)

Tukeys's test at $FWE\alpha = 0.05$ is as follows. Note that there are $k = \binom{4}{2} = 6$ pairs, and the significance level is $\frac{\alpha}{k} = \frac{.05}{6} = 0.008\bar{3}$.

Grouping Information Using the Tukey Method and 95% Confidence

Technique	N	Mean	Grouping	
2	4	3156.3	A	
1	4	2971.0	A	
3	4	2933.8	A	
4	4	2666.3		B

Means that do not share a letter are significantly different.

Figure 5: The output of the One-way ANOVA: Tensile Strength versus Technique at $\alpha = 0.0083$ from Minitab.

The graphical results are:

Mixing Technique **(4)** **(3)** **(1)** **(2)**

These results are different from The Fisher LSD method as they indicate that technique 1, 2, and 3 are not not significantly different.

Question 4

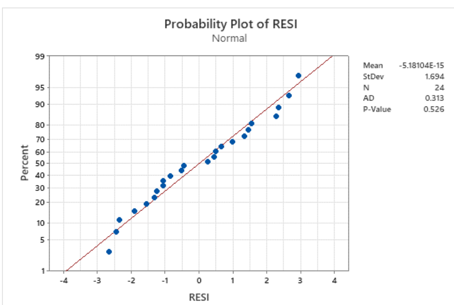
a)

H_0: All means are equal.
H_a: Not all means are equal.
Performing an F test at we can see that the P-value $= .052 > \alpha = .05$, therefore we can conclude the following. The evidence from the data is consistent with the hypothesis that all means are equal.

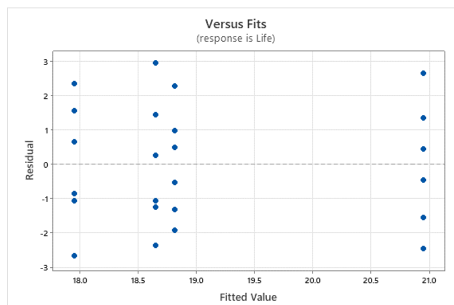
b)

Given that the F test concluded that all means are equal, It should not matter what fluid is selected given that the objective is long life. That being said, if forced to choose, I would pick fluid 3, as it had the highest sample mean at $\bar{y}_{3\cdot} = 20.95$.

c)



(a) Minitab output showing normal probability plot.



(b) Minitab output showing the residuals vs fitted value.

H_0: The data are drawn from a normal disribution.
H_a: The data are not drawn from a normal disribution.

Performing an Anderson Darling normality test at $\alpha = 0.05$ on the residuals gives a P-value of 0.526. Since the P-value is larger than α, we can conclude the following. The evidence of the data is consistent with the data being drawn from a normal disribution.

Observing the residuals shows no indication of heteroskedasticity. This, along with the normality test indicates that the basic analysis of variance assumptions are satisfied.

d)

In order to test whether the average of the mean life of the two fluid types 1 and 3 is greater than the mean life of fluid type 4. We set up the following right-tailed test at $\alpha = 0.05$.

H_0: $\frac{1}{2}(\mu_1 + \mu_3) - \mu_4 = 0$
H_a: $\frac{1}{2}(\mu_1 + \mu_3) - \mu_4 > 0$

The contrast of intrest, where $c_1 = \frac{1}{2}, c_2 = 0, c_3 = \frac{1}{2}, c_4 = -1$, and $a = 4$ is:

$$\begin{aligned}\hat{\Gamma} &= \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \\ &= \frac{1}{2}(18.65) + 0(17.95) + \frac{1}{2}(20.95) - (18.817) \\ &= 0.983\end{aligned}$$

Noting that $MSE = 3.3$, and $n = 6$ the test statistic is:

$$\begin{aligned}t &= \frac{\sum_{i=1}^{a} c_i \bar{y}_{i\cdot}}{\sqrt{\frac{MSE}{n} \cdot \sum_{i=1}^{a} c_i^2}} \\ &= \frac{0.983}{\sqrt{\frac{3.3}{6} \cdot 1.5}} \\ &= 1.032\end{aligned}$$

Since $t_{\alpha,N-a} = t_{.05,24-4} = 1.725$ the test is not in the right-tailed critical region, thus we can conclude the following. There is not enough statistical evidence to support the hypothesis that the average of the mean life of the two fluid types 1 and 3 is greater than the mean life of fluid type 4.

e)

Bonferroni's method to do a pairwise comparison at $FWE\alpha = 0.05$ is as follows. Note that there are $k = \binom{4}{2} = 6$ pairs, and the significance level is $\frac{\alpha}{k} = \frac{.05}{6} = 0.008\bar{3}$. The test statistic, where $MSE = 3.3$, $n_i = n_j = 6$, and $N - a = 24 - 4 = 20$, is:

$$\begin{aligned}t_{\alpha/2k,N-a}\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})} &= t_{0.00069,20}\sqrt{3.3(\frac{1}{6} + \frac{1}{6})} \\ &= 4.008\sqrt{3.3(\frac{1}{6} + \frac{1}{6})} \\ &= 4.203\end{aligned}$$

The differences, where $\bar{y}_{1\cdot} = 18.65$, $\bar{y}_{2\cdot} = 17.95$, $\bar{y}_{3\cdot} = 20.95$, and $\bar{y}_{4\cdot} = 18.817$, are:

Difference	|Difference|	Significantly Different?
$\bar{y}_{1\cdot} - \bar{y}_{2\cdot}$	0.700	No
$\bar{y}_{1\cdot} - \bar{y}_{3\cdot}$	2.300	No
$\bar{y}_{1\cdot} - \bar{y}_{4\cdot}$	0.167	No
$\bar{y}_{2\cdot} - \bar{y}_{3\cdot}$	3.000	No
$\bar{y}_{2\cdot} - \bar{y}_{4\cdot}$	0.867	No
$\bar{y}_{3\cdot} - \bar{y}_{4\cdot}$	2.153	No

Therefore the graphical for Bonferroni's method to do a pairwise comparison are:

Fluid Type **(2)** **(1)** **(4)** **(3)**

Question 5

a)

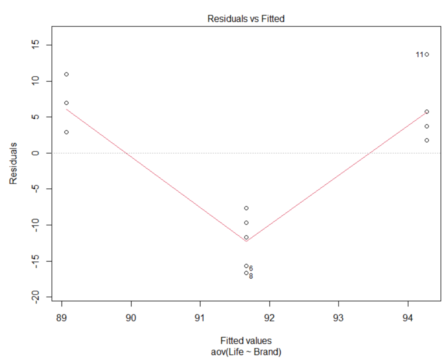
Listing 1: R output of the summary of the aov function

```
             Df  Sum Sq  Mean Sq  F value  Pr(>F)
Brand         1    67.6     67.6    0.668   0.429
Residuals    13  1315.7    101.2
```

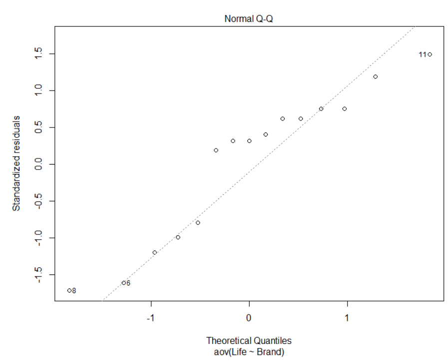
H_0: All means are equal.
H_a: Not all means are equal.

Performing an F test at we can see that the P-value $= 0.429 > \alpha = .05$, therefore we can conclude the following. The evidence from the data is consistent with the hypothesis that all means are equal.

b)



(a) R output showing the residuals.



(b) R output showing the normal probability plot.

c)

95 percent confidence interval estimate on the mean life of battery brand 2.

In order to construct a 95 percent confidence interval estimate on the mean life of battery brand 2. The following equation can be used.

$$\sum_{i=1}^{a} c_i \bar{y}_{i\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}\sum_{i=1}^{a} c_i^2} \leq \sum_{i=1}^{a} c_i \mu_i \leq \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} + t_{\alpha/2,N-a}\sqrt{\frac{MSE}{n}\sum_{i=1}^{a} c_i^2}$$

The contrast of intrest, where $c_1 = 0, c_2 = 1, c_3 = 0$, and $a = 3$ is:

$$\begin{aligned}\Gamma &= \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \\ &= 0(95.2) + 1(79.4) + 0(100.4) \\ &= 79.4\end{aligned}$$

Using R, the following values were obtained. $MSE = 1315.7$, and $t_{\alpha/2,N-a} = t_{0.025,12} = 2.178$. Thus, the 95 percent confidence interval estimate on the mean life of battery brand 2 is:

$$\begin{aligned}79.4 - 2.178\sqrt{\frac{1315.7}{5} \cdot 1} \leq \sum_{i=1}^{a} c_i \mu_i &\leq 79.4 + 2.178\sqrt{\frac{1315.7}{5} \cdot 1} \\ 44.06 \leq \sum_{i=1}^{a} c_i \mu_i &\leq 114.73\end{aligned}$$

99 percent confidence interval estimate on the mean difference between the lives of battery brands 2 and 3.

The contrast of intrest, where $c_1 = 0, c_2 = 1, c_3 = -1$, and $a = 3$ is:

$$\begin{aligned}\Gamma &= \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \\ &= 0(95.2) + 1(79.4) + -1(100.4) \\ &= -21\end{aligned}$$

Using R, the following values were obtained. $t_{\alpha/2,N-a} = t_{0.005,12} = 3.054$. Thus, the 99 percent confidence interval estimate on the mean difference between the lives of battery brands 2 and 3 is:

$$-21 - 3.054\sqrt{\frac{1315.7}{5} \cdot 2} \leq \sum_{i=1}^{a} c_i\mu_i \leq -21 + 3.054\sqrt{\frac{1315.7}{5} \cdot 2}$$

$$-91.06 \leq \sum_{i=1}^{a} c_i\mu_i \leq 49.06$$

d)

The percentage of batteries expected to fail before 85 weeks can be obtained in R using the following code.

Listing 2: Calculating the percentage of batteries expected to fail before 85 weeks.

```
# Calculate percentage of batteries expected to fail
percentage <- pnorm(85, mean = 100.4, sd = 36.27) * 100

# Print the result
percentage
```

Listing 3: Code output.

```
33.5566
```

Question 6

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$.
H_a: At least one σ_i^2 is different.

Listing 4: R output of the bartlett test.

```
                Bartlett test of homogeneity of variances

data:  Life by Fluid.Type
Bartletts K-squared = 0.26691, df = 3,
p-value = 0.9661
```

Since the P-value is large (0.9661), we fail to reject H_0, and can conclude the following. There is not enough statistical evidence to support the hypotheis that at least one σ_i^2 is different.

This is the same conclusion met when observing the residual plot regarding equality of variances in question 4.

Question 7

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$.
H_a: At least one σ_i^2 is different.

Tests

Method	Test Statistic	P-Value
Multiple comparisons	—	0.835
Levene	0.07	0.933

Figure 8: The output of the Test for Equal Variances: Life versus Brand from Minitab.

Since the P-value is large (0.933), we fail to reject H_0, and can conclude the following. There is not enough statistical evidence to support the hypotheis that at least one σ_i^2 is different.

This is the same conclusion met when observing the residual plot regarding equality of variances in question 5.