1) 4.23 a from the book

HW 6

a) Given $\bar{x}_1 = 9.85$ $s_1^2 = 6.79$

$n_1 = 10$

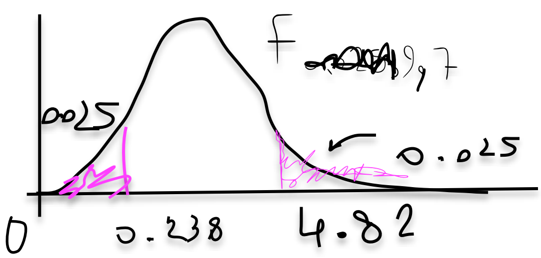
$\alpha = 0.05$

$\bar{x}_2 = 8.08$ $s_2^2 = 6.18$

$n_2 = 8$

Are the variances equal?

1.) Parameters of Interest: variances on % of impurity (σ_1^2, σ_2^2) before and after a new purification unit was installed



2) $H_0 : \sigma_1^2 = \sigma_2^2$

3) $H_1 : \sigma_1^2 \neq \sigma_2^2$

Test statistic:

4) $f_0 = \frac{s_1^2}{s_2^2} = \frac{6.79}{6.18} = 1.099 = \boxed{1.10}$

5) Rejection Criteria: if $f_0 > f_{\alpha/2, n_1-1, n_2-1}$ or

if $f_0 < f_{1-\alpha/2, n_1-1, n_2-1}$

p. 718 of book

$f_{0.025, 9, 7} = 4.82$

p-value: must be > 0.25 according to the tables

c) Calculations:

$f_0 < f_{\alpha/2, n_1-1, n_2-1}$ $\therefore$ fail to reject the null

7) <u>Critical value</u>: there is not enough evidence to conclude $\sigma_1^2 \neq \sigma_2^2$

<u>P-value</u>: p-value > 0.05 therefore we fail to reject the null hypothesis

Remember, p-value is the probability of obtaining f_0 if the null is true → in this case it is large indicating we don't have enough evidence to say it is not true!

b) p-value from Minitab: 0.922

Method

σ_1^2: variance of Sample 1
σ_2^2: variance of Sample 2
Ratio: σ_1^2/σ_2^2
F method was used. This method is accurate for normal data only.

Descriptive Statistics

Sample	N	StDev	Variance	95% CI for σ
Sample 1	10	2.606	6.790	(1.792, 4.757)
Sample 2	8	2.486	6.180	(1.644, 5.060)

Ratio of Standard Deviations

Estimated Ratio	95% CI for Ratio using F
1.04819	(0.477, 2.147)

Test

Null hypothesis $H_0: \sigma_1 / \sigma_2 = 1$
Alternative hypothesis $H_1: \sigma_1 / \sigma_2 \neq 1$
Significance level $\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
F	1.10	9	7	0.922

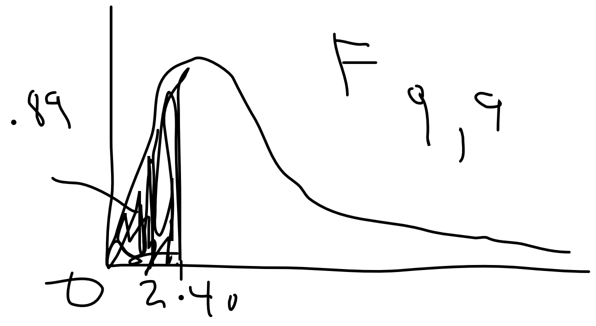
2)

a)

Test

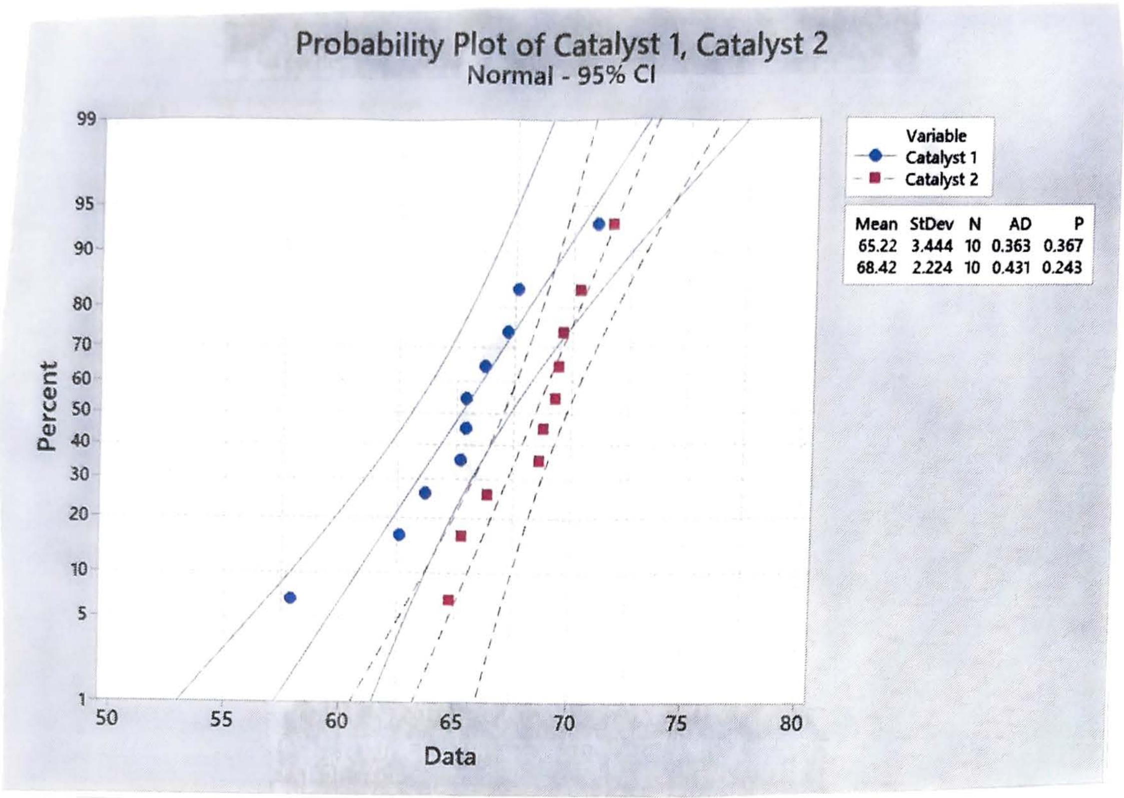
Null hypothesis	$H_0: \sigma_1^2 / \sigma_2^2 = 1$
Alternative hypothesis	$H_1: \sigma_1^2 / \sigma_2^2 < 1$
Significance level	$\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
F	2.40	9	9	0.896



Since the p-value > 0.05 we would fail to reject the null hypothesis that the variances are ~~eual~~ equal. We therefore could potentially assume equal variance for a t-test (though, if any doubt exists, it isn't a bad idea to go with unequal.)

b)



Both lines are straight, normality seems like a good assumption

3) a) Conduct ANOVA:

1) Parameter of interest: multiple means of baked anode density

2) Null hypothesis $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

3) Alternative hypothesis $H_1: \tau_i \neq 0$ for at least one i

4) Test statistic: $F_0 = \dfrac{\frac{SS_{Treatments}}{a-1}}{\frac{SS_E}{a(n-1)}} = \dfrac{MS_{Treatments}}{MS_E}$

5) Rejection Criteria: Reject H_0 if $F_0 > F_{\alpha, a-1, a(n-1)}$

6) Calculations: $a=4$, $n=6$, $N=24$

$$SS_T = SS_{Treatments} + SS_E$$

$$SS_T = \sum_{i=1}^{4} \sum_{j=1}^{6} (y_{ij} - \bar{y}_{..})^2$$

$$SS_{Treatments} = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\bar{y}_{..} = \text{grand average} = \frac{41.8 + 41.9 \text{ ... all points ... } + 41.5}{24} = 41.5167$$

$$= \frac{\sum_{i=1}^{4} \sum_{j=1}^{6} y_{ij}}{N}$$

calc in Excel

$$SS_T = (41.8-41.52)^2 + (41.9-41.52)^2 + \ldots \text{ all points! } \ldots + (41.5-41.52)^2$$

$$= 2.55$$

$$SS_{Treatments} = 6\left[(41.7-41.52)^2+(41.58-41.52)^2+(41.45-41.52)^2+(41.33-41.52)^2\right]$$

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{n} y_{ij}}{n} \quad i=1,2,\ldots a) \rightarrow \text{average at each treatment}$$

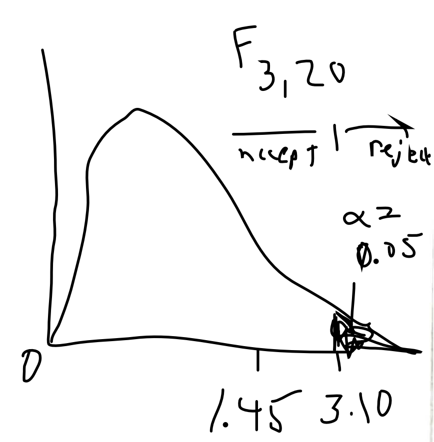
$$= 0.457$$

$$SS_E = SS_T - SS_{Treatments}$$

$$= 2.55 - 0.457 = 2.097$$

$$F_0 = \frac{\frac{0.457}{4-1}}{\frac{2.097}{4(6-1)}} = \boxed{1.45}$$

$$F_{\alpha, a-1, a(n-1)} = F_{0.05, 3, 20} = 3.10$$



Excel:

■ **TABLE 4E.6**

Baked Density Data for Exercise 4.39

Temperature (°C)	Density					
500	41.8	41.9	41.7	41.6	41.5	41.7
525	41.4	41.3	41.7	41.6	41.7	41.8
550	41.2	41.0	41.6	41.9	41.7	41.3
575	41.0	40.6	41.8	41.2	41.9	41.5

Temp	1	2	3	4	5	6	yi.	yi.
500	41.8	41.9	41.7	41.6	41.5	41.7	250.2	41.7
525	41.4	41.3	41.7	41.6	41.7	41.8	249.5	41.58333333
550	41.2	41	41.6	41.9	41.7	41.3	248.7	41.45
575	41	40.6	41.8	41.2	41.9	41.5	248	41.33333333
							996.4	41.51666667
							y..	y..

1st way	**SST**	**2.553333**						**Sstreatment**	**0.456667**
	0.080278	0.146944	0.033611	0.006944	0.000278	0.033611		0.033611111	
	0.013611	0.046944	0.033611	0.006944	0.033611	0.080278		0.004444444	
	0.100278	0.266944	0.006944	0.146944	0.033611	0.046944		0.004444444	
	0.266944	0.840278	0.080278	0.100278	0.146944	0.000278		0.033611111	
SSE	**2.096667**								
2nd way	**SST**	**2.553333**						**Sstreatment**	**0.456667**
	1747.24	1755.61	1738.89	1730.56	1722.25	1738.89		10433.34	
	1713.96	1705.69	1738.89	1730.56	1738.89	1747.24		10375.04167	
	1697.44	1681	1730.56	1755.61	1738.89	1705.69		10308.615	
	1681	1648.36	1747.24	1697.44	1755.61	1722.25		10250.66667	

Conclusions

$F_0 < F_{\alpha, a-1, a(n-1)}$

therefore, we fail to reject the null hypothesis, there is not enough evidence to say the temperature affects anode density

b) Residuals: $e_{ij} = y_{ij} - \bar{y}_{i.}$

e_{ij} → residual
y_{ij} → all data points
$\bar{y}_{i.}$ → average of treatments

e_{ij} shown below:

Residuals:

0.1	0.2	0	-0.1	-0.2	0
-0.18333	-0.28333	0.116667	0.016667	0.116667	0.216667
-0.25	-0.45	0.15	0.45	0.25	-0.15
-0.33333	-0.73333	0.466667	-0.13333	0.566667	0.166667

To make the plot:

1) rank data (Residuals)

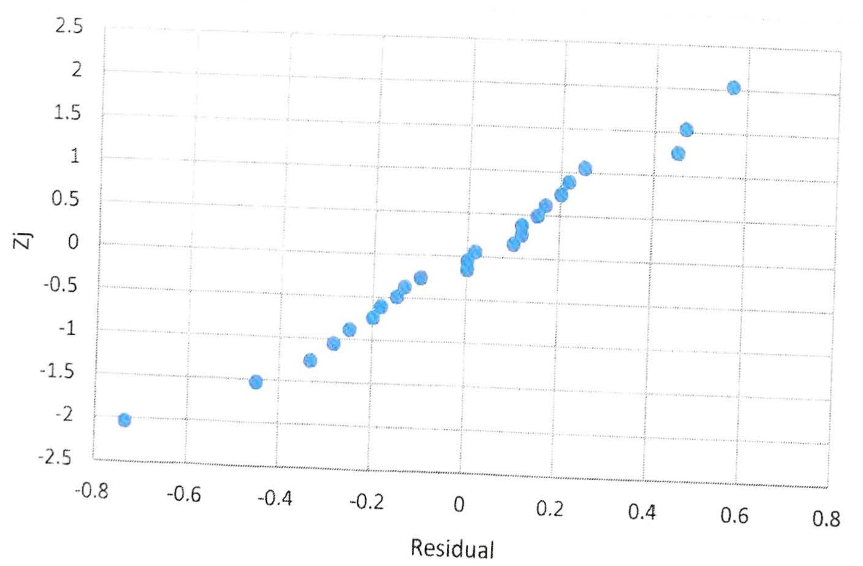
2) Calculate the cumulative frequency $\left(\frac{j-0.5}{n}\right)$ for each point

3) Find z_j → can input the frequencies into Minitab to get z_j's

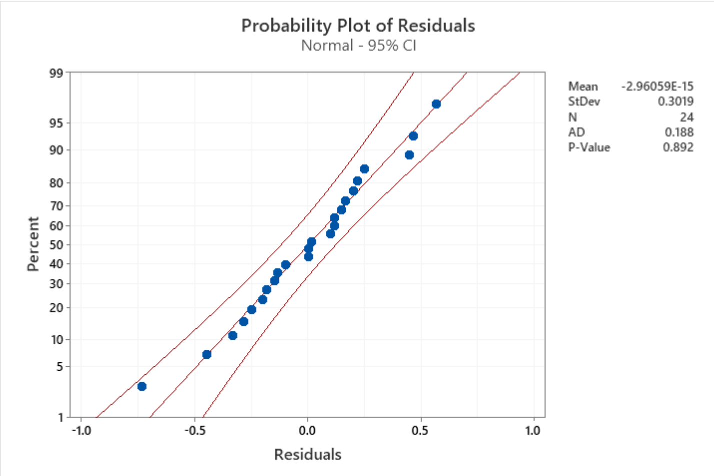
$$P(Z \le z) = \frac{j-0.5}{n} \text{ for each point}$$

4) Plot residual vs. z_j

	Ranked Residuals	Cumulative Frequency	Zj
1	-0.733333333	0.020833333	-2.03683
2	-0.45	0.0625	-1.53412
3	-0.333333333	0.104166667	-1.25816
4	-0.283333333	0.145833333	-1.05447
5	-0.25	0.1875	-0.88715
6	-0.2	0.229166667	-0.74159
7	-0.183333333	0.270833333	-0.61029
8	-0.15	0.3125	-0.48878
9	-0.133333333	0.354166667	-0.3741
10	-0.1	0.395833333	-0.26415
11	0	0.4375	-0.15731
12	0	0.479166667	-0.05225
13	0.016666667	0.520833333	0.05225
14	0.1	0.5625	0.15731
15	0.116666667	0.604166667	0.26415
16	0.116666667	0.645833333	0.3741
17	0.15	0.6875	0.48878
18	0.166666667	0.729166667	0.61029
19	0.2	0.770833333	0.74159
20	0.216666667	0.8125	0.88715
21	0.25	0.854166667	1.05447
22	0.45	0.895833333	1.25816
23	0.466666667	0.9375	1.53412
24	0.566666667	0.979166667	2.03683



The plot has a straight line, so ~~[illegible]~~ normality is a good assumption.



Note: directions say you can just cut and paste the residuals into Minitab to get this plot (don't have to do by hand in excel)

c. Since ~~[illegible]~~ we don't have enough evidence to say temperature matters, a low temperature might be recamended to save maitenence costs and energy.

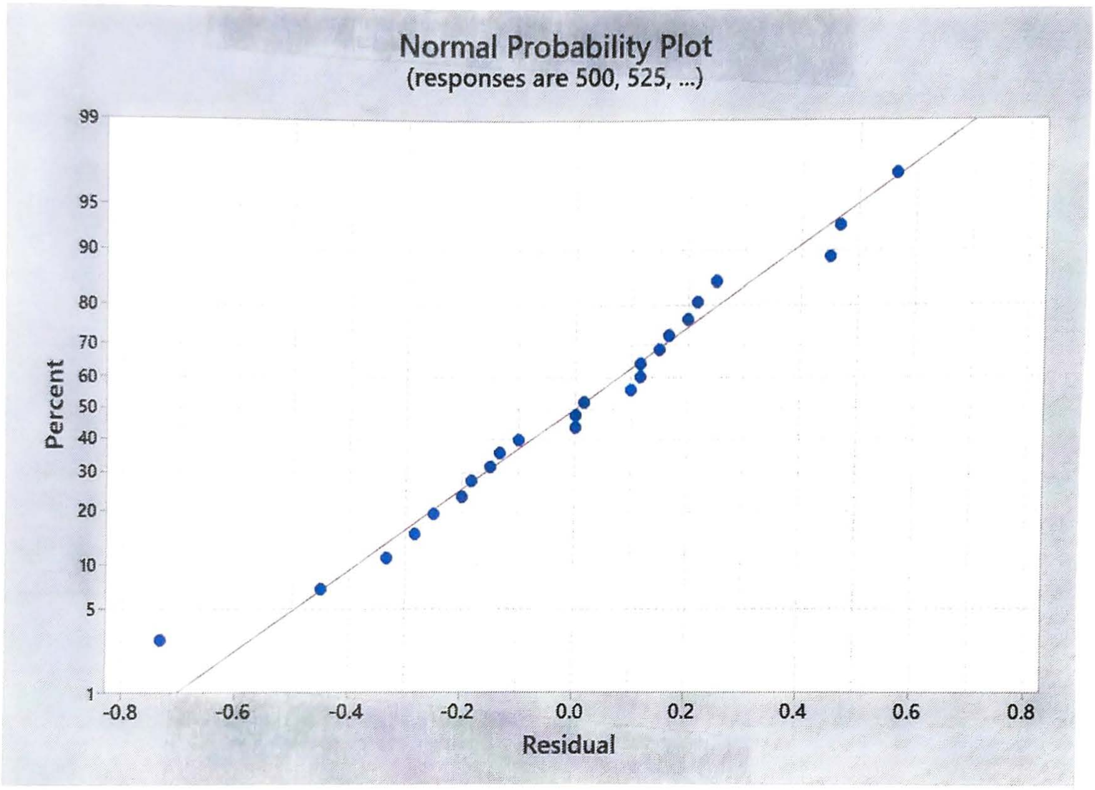
It also has lowest standard deviation of the samples.

Finally consider processing time in this decision → a lean princliple!

a) Minitab:

You can see the P-value is $0.258 > 0.05 = \alpha$

So the conclusion made previously was correct (fail to reject the null)



Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	500, 525, 550, 575

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	0.4567	0.1522	1.45	0.258
Error	20	2.0967	0.1048		
Total	23	2.5533			

1.)

a) Given the Mint tab output below, we conclude that we have enough evidence to reject the null hypothesis and say that the size of the orfice does affect the mean percentage of radon released.

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.05

Equal variances were assumed for the analysis.

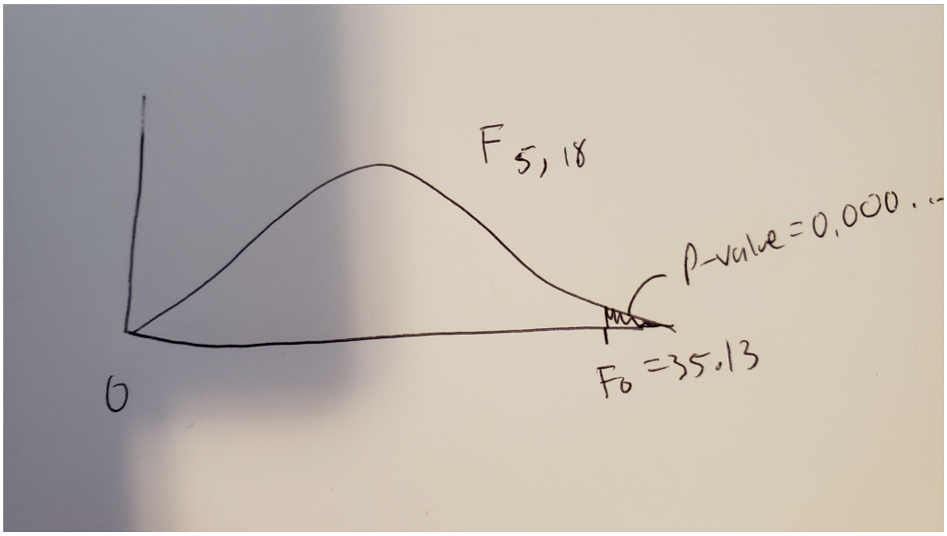
Factor Information

Factor	Levels	Values
Factor	6	0.37, 0.51, 0.71, 1.02, 1.4, 1.99

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	5	1163.7	232.742	35.13	0.000
Error	18	119.2	6.625		
Total	23	1283.0			

P-value < 0.05



a) 1 factor (orifice diameter)

2-a) 6 levels (or treatments)

3-a) $F_{critical\ value} = F_{\alpha, a-1, a(n-1)}$

$$\cancel{= F_{0.05, 5, 15} = \boxed{2.90}}$$

a = 6

n = 4

$$= F_{0.05, 5, 18} = \boxed{2.77}$$

F-Value = 35.13 which is >> than ~~2.90~~ 2.77

b) Residuals: $e_{ij} = y_{ij} - \bar{y}_{i.}$

b-2)

e_{ij}: residual

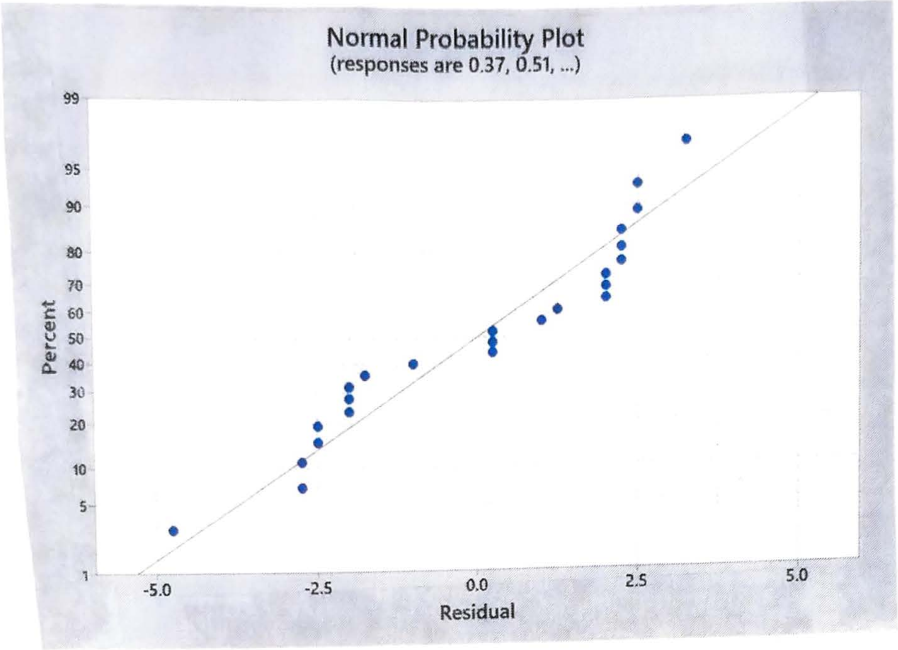
y_{ij}: all data points

$\bar{y}_{i.}$: average of treatments

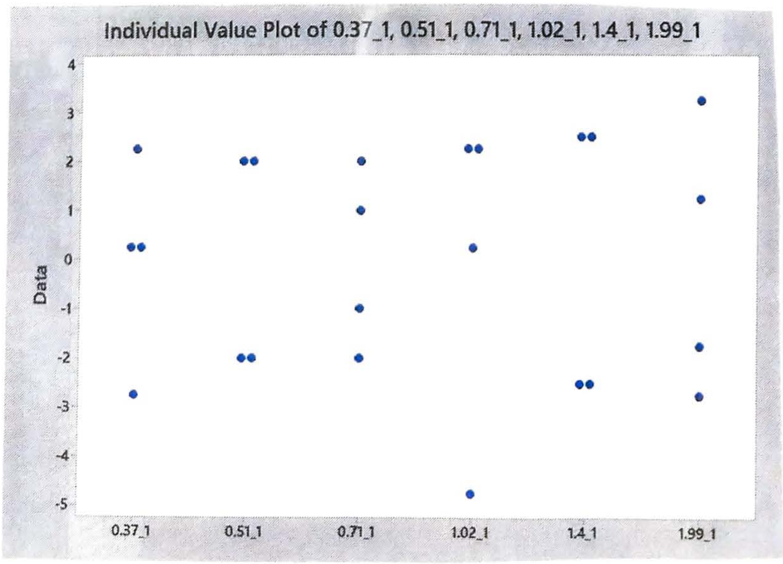
Residuals calculated in Excel

	0.37	0.51	0.71	1.02	1.4	1.99
	80	75	74	67	62	60
	83	75	73	72	62	61
	83	79	76	74	67	64
	85	79	77	74	67	66
$y\bar{i}.$	82.75	77	75	71.75	64.5	62.75
	RESIDUALS:					
	-2.75	-2	-1	-4.75	-2.5	-2.75
	0.25	-2	-2	0.25	-2.5	-1.75
	0.25	2	1	2.25	2.5	1.25
	2.25	2	2	2.25	2.5	3.25

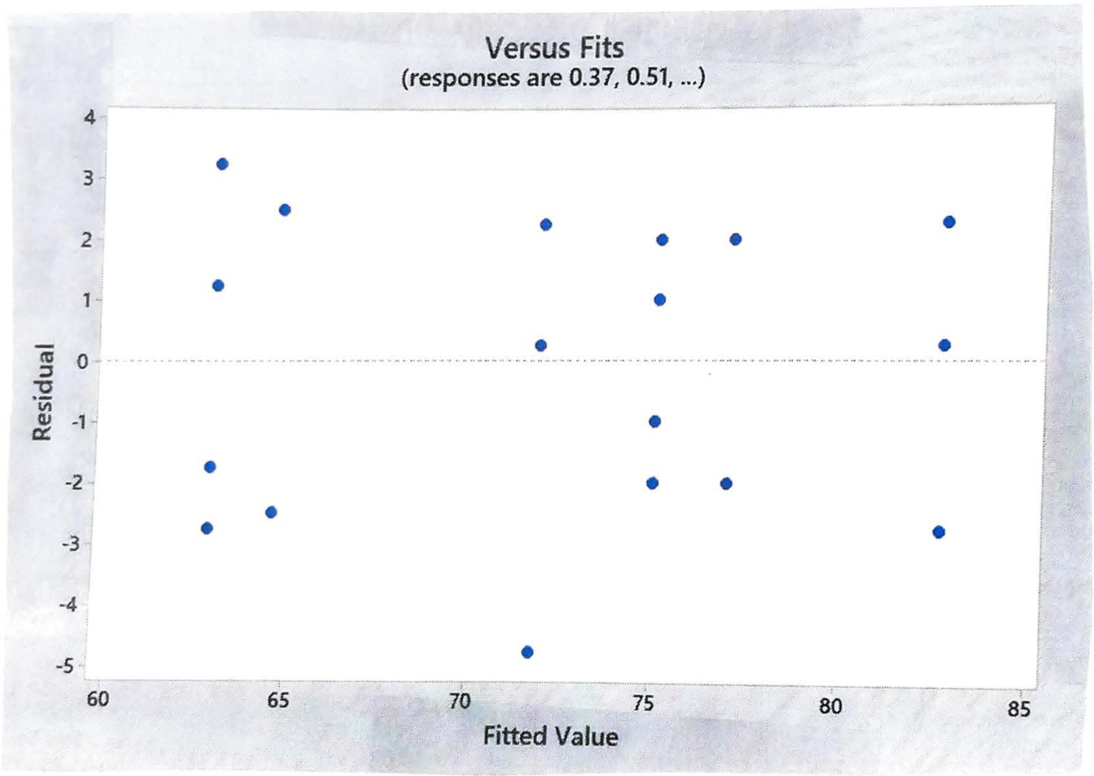
b-1) Normality seems appropriate



b-3) Checks the assumption that the variances are equal at each factor level. This assumption appears to be valid as no pattern emerges.



4) No pattern emerges so the variability in residuals does not depend on in magnitude of predicted values ($\bar{y}_{i.}$)
If a pattern emerged, we could try transforming the data.



$$LSD = t_{\alpha/2,\, a(n-1)} \sqrt{\frac{2MS_E}{n}}$$

$$LSD = t_{\frac{0.05}{2},\, 6(3)} = t_{0.025,\, 18} = 2.101$$

$$MS_E = 6.625 \qquad LSD = 2.101\sqrt{\frac{2(6.625)}{4}} = 3.824$$

$$\cancel{LSD = 2.101\sqrt{\frac{6.625}{6}}} = \cancel{2.207}$$

Comparisons: Difference greater than LSD?

.37 vs 0.51 : 5.75 yes A

.37 vs .71 : 7.75 yes B

.51 vs .71 : 2.0 no B

Group 0.37 and 0.51 are different, as well as 0.37 and 0.71, but 0.71 and 0.51 are not

Grouping Information Using the Fisher LSD Method and 95% Confidence

Factor	N	Mean	Grouping
0.37	4	82.75	A
0.51	4	77.00	B
0.71	4	75.000	B C
1.02	4	71.75	C
1.4	4	64.50	D
1.99	4	62.75	D

Means that do not share a letter are significantly different.