W&M CSCI 688: Design of Experiments
Homework 2

Megan Rose Bryant

September 25, 2014

3.5

The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted and the following data were collected:

Mixing Technique	Tensile Strength (lb/in^2)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

a.) Test the hypothesis that mixing techniques affect the strength of the cement. Use an $\alpha = 0.05$.

Based on the Minitab express ouput below, the model has a $F_0 = 12.73$ and a $p - value = 0.0005$. Now, to interpret this F-statistic properly, we need to find the fixed significance level region criteria for rejection. We have $a = 4$ treatments with $n = 4$ observations each, therefore the relevant rejection criteria is $F_{3,12,0.05} = 3.49$. Since $F_0 = 12.73 > 3.49$, we may reject the null in favor of the alternative and conclude that the mixing techniques have an affect on the strength of the cement.

One-Way ANOVA: T1, T2, T3, T4

9/15/2014 2:26:41 PM

Method

Null hypothesis	All means are equal
Alternative hypothesis	At least one mean is different
Significance level	α = 0.05

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	T1, T2, T3, T4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	489740.188	163246.729	12.73	0.0005
Error	12	153908.250	12825.688		
Total	15	643648.438			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
113.250552	76.09%	70.11%	57.49%

Means

Factor	N	Mean	StDev	95% CI
T1	4	2971.00	120.56	(2847.62, 3094.38)
T2	4	3156.25	135.98	(3032.87, 3279.63)
T3	4	2933.75	108.27	(2810.37, 3057.13)
T4	4	2666.25	80.97	(2542.87, 2789.63)

Pooled StDev = 113.250552

c.) Use the Fisher LSD method with $\alpha = 0.05$ to make comparisons between pairs of means.

The Fisher Least Significant Difference method uses the t-statistic for testing $H_0 : \mu_i \neq \mu_j$

$$t_0 = \frac{\bar{y_i}. - \bar{y_j}.}{\sqrt{MS_E(\frac{1}{n_i} + \frac{1}{n_j})}}$$

against the following rejection criteria

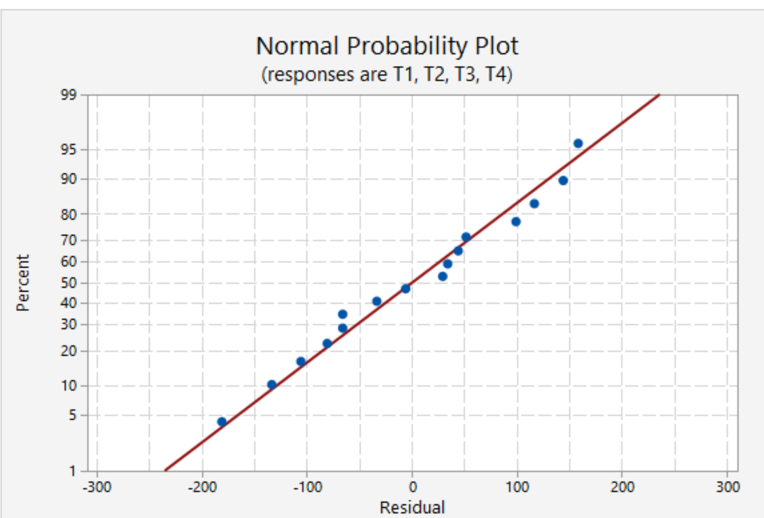
$$LSD = t_{\alpha/2, N-a}\sqrt{\frac{2MS_E}{n}} = t_{.025,}\sqrt{\frac{2*12825.688}{4}} = 2.1788 * 80.0802 = 174.479.$$

Comparison	Value	> 174.479?
T1 vs T2	$\lvert 2971.00 - 3156.25 \rvert = 185.25$	Yes
T1 vs T3	$\lvert 2971.00 - 2933.75 \rvert = 37.25$	No
T1 vs T4	$\lvert 2971.00 - 2666.25 \rvert = 304.75$	Yes
T2 vs T3	$\lvert 3156.25 - 2933.75 \rvert = 222.5$	Yes
T2 vs T4	$\lvert 3156.25 - 2666.25 \rvert = 490.00$	Yes
T3 vs T 4	$\lvert 2933.75 - 2666.25 \rvert = 267.5$	Yes

The Fisher LSD test reveals that the means of mixing technique 1 and 3 are not significantly different at the $\alpha = .05$ level. However, at the 5% significance level, all other pairs of means are significantly different.

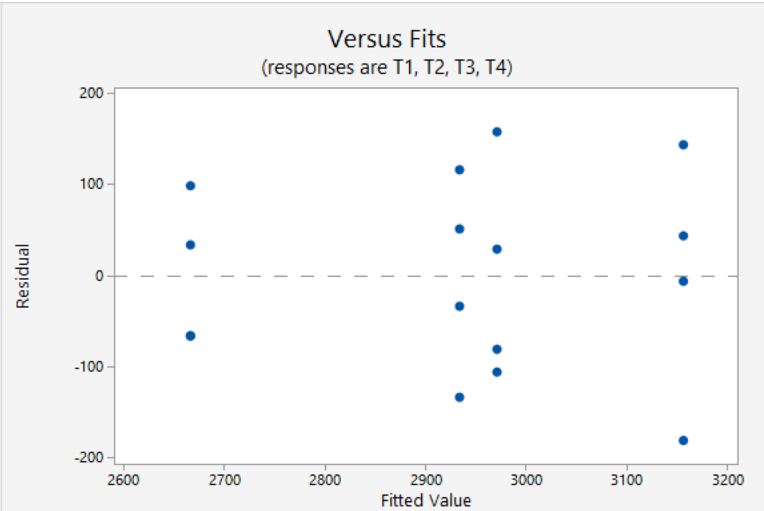
d.)Construct a normal probability plot of the residuals. What conclusions would you draw about the validity of the normality assumption?

Based on the derived normal probability plot, we cannot conclude that there is any reason to question the normality assumption.

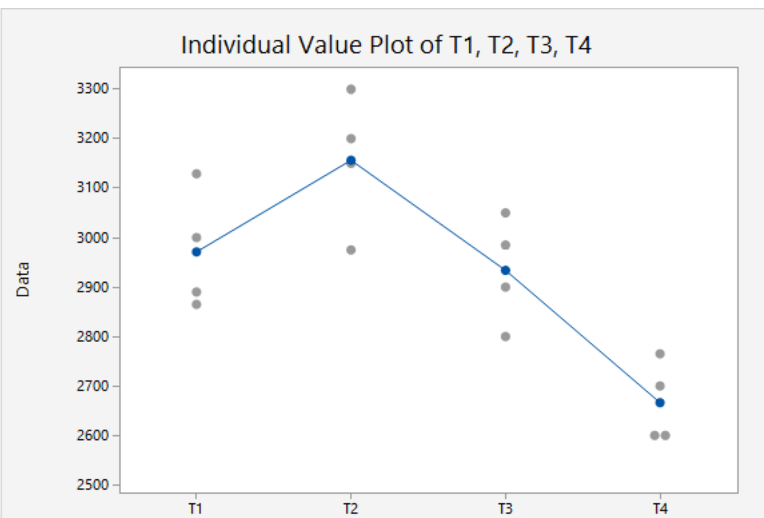


e.) Plot the residuals versus the predicted tensile strength. Comment on the plot.

The residual plot appears to be structureless and contains no obvious patterns.



f.)Prepare a scatter plot of the results to aid the interpretation of the results of this experiment.



3.16

A manufacturer of television sets is interested in the effet on tube conductivity of four different types of coatingfor color picture tubes. A completely randomized experiment is conducted and the following conductivity data are obtained:

Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	136	132	127
4	129	127	132	129

a.) Is there a difference in conductivity due to coating type? Use $\alpha = 0.05$.

Based on the Minitab express ouput below, the model has a $F_0 = 14.30$ and a $p - value = 0.0003$. Now, to interpret this F-statistic properly, we need to find the fixed significance level region criteria for rejection. We have $a = 4$ treatments with $n = 4$ observations each, therefore the relevant rejection criteria is $F_{3,12,0.05} = 3.49$. Since $F_0 = 14.30 > 3.49$, we may reject the null in favor of the alternative and conclude that the coating type has an effect on conductivity.

One-Way ANOVA: T1, T2, T3, T4

9/15/2014 9:42:53 PM

Method

Null hypothesis	All means are equal
Alternative hypothesis	At least one mean is different
Significance level	α = 0.05

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	T1, T2, T3, T4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	844.6875	281.5625	14.30	0.0003
Error	12	236.2500	19.6875		
Total	15	1080.9375			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.43705984	78.14%	72.68%	61.14%

Means

Factor	N	Mean	StDev	95% CI
T1	4	145.000	3.916	(140.166, 149.834)
T2	4	145.250	6.652	(140.416, 150.084)
T3	4	132.250	3.862	(127.416, 137.084)
T4	4	129.250	2.062	(124.416, 134.084)

Pooled StDev = 4.43705984

b.) Estimate the overall mean and the treatment effects.

$$\hat{\mu} = \frac{\sum_{i=1}^{4}\sum_{j=1}^{4}\mu_{i,j}}{4*4} = \frac{2207}{16} = 137.9375$$
$$\hat{\tau}_1 = \bar{y}_1 - \hat{\mu} = 145.00 - 137.9375 = 7.0625$$
$$\hat{\tau}_2 = \bar{y}_2 - \hat{\mu} = 145.250 - 137.9375 = 7.3125$$
$$\hat{\tau}_3 = \bar{y}_3 - \hat{\mu} = 132.250 - 137.9375 = -5.6875$$
$$\hat{\tau}_4 = \bar{y}_4 - \hat{\mu} = 129.250 - 137.9375 = -8.6875$$

c.)Compute a 95% confidence interval estimate of the mean of coating type 4. Compute a 99% confidence interval estimate of the mean difference between coating types 1 and 4.

95% Confidence Interval of T4: $\mu_4 \pm t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} = 129.250 \pm t_{.025,12}\sqrt{\frac{19.6875}{4}} = 129.250 \pm 2.1789 * 2.21855 = 129.250 \pm 4.834$

Therefore, the confidence interval is $(124.416, 134.084)$ i.e. (129.250 ± 4.834).

Note*: This CI is given in the Minitab Express Output.

99% Confidence Interval of Mean Differnce Between T1 and T4: $(\mu_1 - \mu_4) \pm t_{\alpha/2,N-a}\sqrt{\frac{2*MS_E}{n}} = 145.00 - 129.25 \pm t_{.005,12}\sqrt{\frac{2*19.6875}{4}} = 15.75 \pm 3.0545 * 3.1375 = 15.75 \pm 9.58.$

Therefore, the confidence interval is $(6.17, 25.33)$ i.e. (15.75 ± 9.58).

d.)Test all pairs of means using the Fisher LSD method with $\alpha = 0.05$.

Fisher Individual Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
T2-T1	0.250	3.137	(-6.586, 7.086)	0.08	0.9378
T3-T1	-12.750	3.137	(-19.586, -5.914)	-4.06	0.0016
T4-T1	-15.750	3.137	(-22.586, -8.914)	-5.02	0.0003
T3-T2	-13.000	3.137	(-19.836, -6.164)	-4.14	0.0014
T4-T2	-16.000	3.137	(-22.836, -9.164)	-5.10	0.0003
T4-T3	-3.000	3.137	(-9.836, 3.836)	-0.96	0.3578

Simultaneous confidence level = 81.57%

The Fisher Least Significant Difference method uses the t-statistic for testing $H_0 : \mu_i \neq \mu_j$

$$t_0 = \frac{\bar{y_i}. - \bar{y_j}.}{\sqrt{MS_E(\frac{1}{n_i} + \frac{1}{n_j})}}$$

against the following rejection criteria

$$LSD = t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} = t_{.025,}\sqrt{\frac{2*19.6875}{4}} = 2.1788 * 3.1375 = 6.835985.$$

Comparison	Value	> 6.835985?
T1 vs T2	$\|145 - 145.250\| = .250$	No
T1 vs T3	$\|145 - 132.250\| = 12.75$	Yes
T1 vs T4	$\|145 - 129.250\| = 15.75$	Yes
T2 vs T3	$\|145.250 - 132.250\| = 13.000$	Yes
T2 vs T4	$\|145.250 - 129.250\| = 16.000$	Yes
T3 vs T 4	$\|132.250 - 129.250\| = 3$	No

The Fisher LSD test reveals that the pairs means of the pairs T_1 &T_2 and T_3 & T_4 are not significantly different at the 5% significance level. However, all other pairs are significantly different at the $\alpha = 0.05$ level.

f.) Assuming that coating type 4 is currently in use, what are your recommendations to the manufacturer? We wish to minimize conductivity.

In order to minimize conductivity, we would continue to utilize coating 4 as it has the lowest mean conductivity and is not signficantly different than coating 3.

3.23

Four chemists are asked to determine the percentage of methyl alcohol in a certain chemical compound. Each chemist makes three determinations, and the results are the following:

Chemist	Percentage of Methyl Alcohol		
1	84.99	84.04	84.38
2	85.15	85.13	84.88
3	84.72	84.48	85.16
4	84.20	84.10	84.55

a.)Do chemists differ significantly? Use $\alpha = 0.05$.

Based on the Minitab express ouput below, the model has a $F_0 = 3.25$ and a $p - value = 0.0813$. Now, to interpret this F-statistic properly, we need to find the fixed significance level region criteria for rejection. We have $a = 4$ treatments with $n = 3$ observations each, therefore the relevant rejection criteria is $F_{3,8,0.05} = 4.066$. Since $F_0 = 3.25 \not> 4.066$, we fail to reject the null hypothesis and cannot conclude that the percentage of Methyl Alcohol differs significantly between chemists.

One-Way ANOVA: C1, C2, C3, C4

9/15/2014 10:46:14 PM

Method

Null hypothesis	All means are equal
Alternative hypothesis	At least one mean is different
Significance level	$\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	C1, C2, C3, C4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	1.04456667	0.348188889	3.25	0.0813
Error	8	0.85820000	0.107275000		
Total	11	1.90276667			

Model Summary

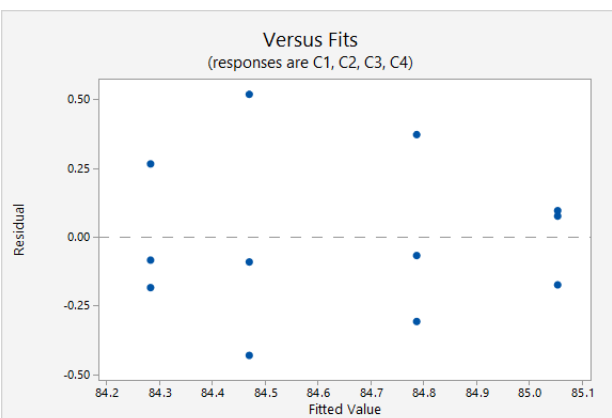
S	R-sq	R-sq(adj)	R-sq(pred)
0.327528625	54.90%	37.98%	0.00%

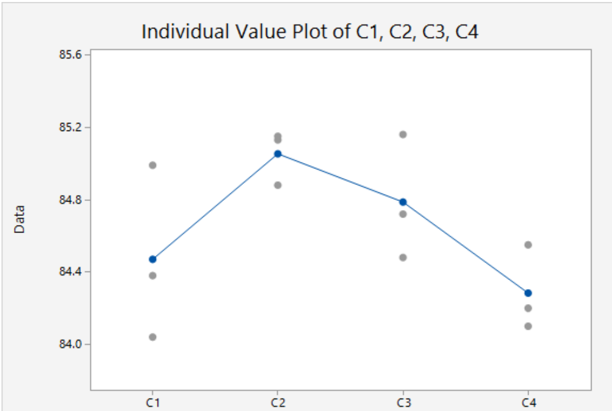
Means

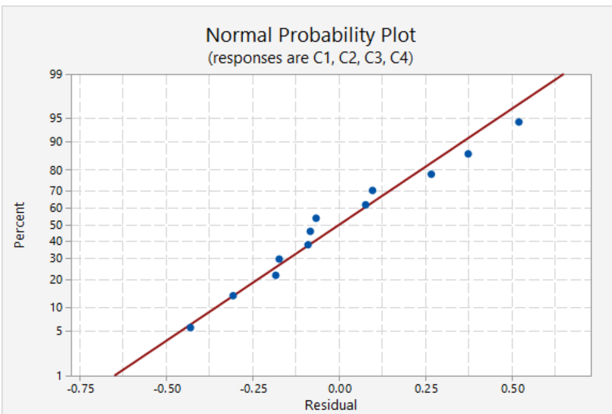
Factor	N	Mean	StDev	95% CI
C1	3	84.4700	0.4814	(84.0339, 84.9061)
C2	3	85.05333	0.15044	(84.61727, 85.48940)
C3	3	84.7867	0.3449	(84.3506, 85.2227)
C4	3	84.2833	0.2363	(83.8473, 84.7194)

Pooled StDev = 0.327528625

b.) Analyze the residuals from this experiment.







The residual plot doesn't have any obvious patterns, so there is no immediate cause for concern. The normal probability plot is fairly straight and passes the fat pencil test, so there is no reason to question the normality assumption. The scatterplot (individual plot) also doesn't contain any obvious patterns or cause for concern. Therefore, there is nothing in the residuals on which to base a challenge to the validity of the data.

c.) If chemist 2 is a new employee, construct a meaningful set of orthogonal constrasts that might have been useful at the start of the experiment.

We know that two contrast with coefficents $\{c_i\}$ and $\{d_i\}$ are orthogonal if $\sum_{i=1}^{4} c_i d_i = 0$, for a balanced design.

We also know that since we have 4 treatements, we need a set of $4 - 1 = 3$ orthogonal contrasts to partition the sum of squares due to treatements into 3 independent, single-degree-of-freedom components. If Chemist 2 is a new employee, then we could use that information to make additional comparisons, denoted $H_{0,1}$.

$H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_{0,1} = \mu_2 = (\mu_1 + \mu_3 + \mu_4)$

This implies that the coefficient of μ_2 must be -3. The other two contrast may be arbitrarily chosen so long as they meet the specified criteria.

Chemists	Total	C1	C2	C3
1	253.41	1	1	0
2	255.16	-3	0	0
3	254.36	1	1	1
4	252.85	1	-2	-1

To determine whether or not these contrasts would have been useful at the start of the experiment, we must determine whether or not the contrast is significant at the 5% significance level.

$SS_{C1} = \frac{(-4.86))^2}{3*12} = .6561$
$SS_{C2} = \frac{(2.07))^2}{6*3} = .23804$
$SS_{C3} = \frac{(1.51))^2}{2*3} = .38002$

Since the SSC has only one degree of freedom, we know that $SSC = MSC$.

Therefore, we need only determine the relevant test statistics.
$F_{C1} = \frac{MSC}{MSE} = \frac{5.9049}{.10725} = 6.11748$
$F_{C1} = \frac{MSC}{MSE} = \frac{2.14245}{.10725} = 2.21958$
$F_{C1} = \frac{MSC}{MSE} = \frac{3.42015}{.10725} = 3.54328$

For an ANOVA test involving contrast, the critical value is $F_{\alpha,a-1,N-a} = F_{0.05,3,11} = 3.59$. The only contrast test statistic that is greater than the critical value is C1. Therefore, the only contrast that would have been useful at the start of the test at a 5% significance level woul have been contrast 1.

3.28

Consider testing the equality of the means of two normal populations, where the variances are unknown but are assumed to be equal. The appropriate test procedure is the pooled t-test. Show that the pooled t-test is equivalent to the single factor analysis of variance.

We know that the test statistic for a pooled t-test is:

$$t_0 = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{S_p\sqrt{2n}}\ t_{2n-2}, \text{ given that the sample sizes are equal.}$$

And, we know that the fundamental ANOVA identity is $SST = SSTR + SSE$ which is equivalent to $\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$

In this instance, we know that $a = 2$ since we are doing a pooled t-test.

We want to show that the pooled t-test is equivalent to the ANOVA. The ANOVA uses the F_0.

$$t_0^2 = \frac{(\bar{y}_{1.} - \bar{y}_{2.})^2}{S_p^2 \frac{2}{n}} = \frac{1}{S_p^2}(\bar{y}_{1.} - \bar{y}_{2.})^2(\frac{n}{2})$$

Let us examine first the former part.

$$S_p^2 = \frac{\sum_{j=1}^{n}(y_{1,j} - \bar{y}_{1.})^2 + \sum_{j=1}^{n}(y_{2,j} - \bar{y}_{2.})^2}{2(n-1)} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{1.})^2}{2(n-1)} = \frac{SSE}{2(n-1)} = MSE, \text{ where } a = 2.$$

Now, we shall look at the remainder of the t_0^2 equation.

$$(\bar{y}_1 - \bar{y}_2)^2 \frac{n}{2}$$

$$\text{We know that } \sum_{j=1}^{n} y_{ij} = y_{i.}, \bar{y}_{i.} = \frac{y_{i.}}{n}, i = 1, \cdots, a \text{ and } y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}, \bar{y}_{..} = \frac{y_{..}}{N}$$

Therefore,

$$(\bar{y}_{1.} - \bar{y}_{2.})^2 \frac{n}{2} = (\bar{y}_{1.}^2 + \bar{y}_{2.}^2 - 2\bar{y}_{1.}\bar{y}_{2.})\frac{n}{2} = \sum_{j=1}^{2} \frac{\bar{y}_{i.}^2}{n} - \frac{y_{..}}{2n} = SSTR$$

Therefore, we have $t_0^2 = \frac{SSTR}{MSE}$. However, since the data is pooled, $SSTR = MSTR$ and we have $t_0^2 = \frac{MSTR}{MSE}$, which is the F_0 statistic for an ANOVA with $a = 2$.

3.31

Use Bartlett's test to determine if the assumption of equal vairances is satisfied in Problem 3.24. Use $\alpha = 0.05$. Did you reach the same conclusion regarding equality of variances by examining residual plots?

Bartlett's Test requires the calculation of a test statistic that has a sampling distribution which is closely approximated by the χ^2 distribution with $a - 1$ degrees of freedom (assuming that the a random samples are from independent normal populations).

The test statistic is $\chi_0^2 = 2.3026\frac{q}{c}$ where

$$q = (N - a)\log_{10} S_p^2 - \sum_{j=1}^{a} \log_{10} S_j^2$$

$$c = 1 + \frac{1}{3(a-1)}(\sum_{j=1}^{a}(n_j - 1)^{-1} - (N - a)^{-1})$$

$$\text{and } S_p^2 = \frac{\sum_{j=1}^{a}(n_j - 1)s_i^2}{N - a}$$

$s_1^2 = 11.2$
$s_2^2 = 14.8$
$s_3^2 = 20.80$

$$S_p^2 = \frac{\sum_{j=1}^{a}(n_j-1)s_i^2}{N-a} = \frac{4*11.2+4*14.8+4*20.8}{15-3} = 15.6$$

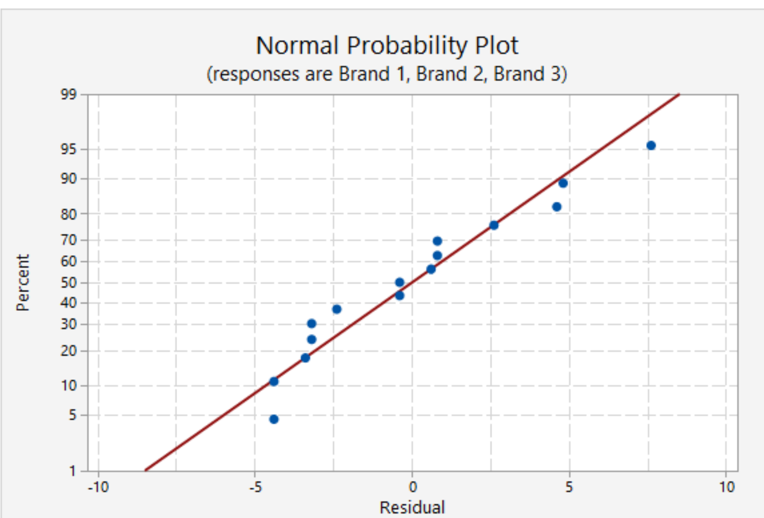
$$q = (N-a)\log_{10} S_p^2 - \sum_{j=1}^{a} \log_{10} S_j^2 =$$
$$(15-3)(\log_{10} 15.6 - 4*(\log_{10} 11.2 + \log_{10} 14.8 + \log_{10} 20.8) = .167323$$

$$c = 1 + \frac{1}{3(a-1)}(\sum_{j=1}^{a}(n_j-1)^{-1} - (N-a)^{-1}) = 1 + \frac{1}{6}(\sum_{j=1}^{3}(5-1)^{-1} - (15-3)^{-1}) = \frac{1}{6}(\frac{3}{4} + \frac{1}{12}) = \frac{41}{36} = 1.1388$$

$$\chi_0^2 = 2.3026(\frac{.167323}{1.1388}) = .3383$$

Our test statistic is $\chi^2_{0.05,4} = 9.49$. Since $\chi_0 = .3383 \not> 9.49 = \chi^2_{0.05,4}$, we fail to reject the null hypothesis in favor of the alternative and cannot conclude that the three brands of batteries are different. Therefore, we find nothing to contradict the assumption of equal variances.

We would reach the same conclusion by examining the residual plot below (for the data of 3-24). The plotted points are relatively linear and would form a sufficient basis for contradicting the equality assumption.



3.32

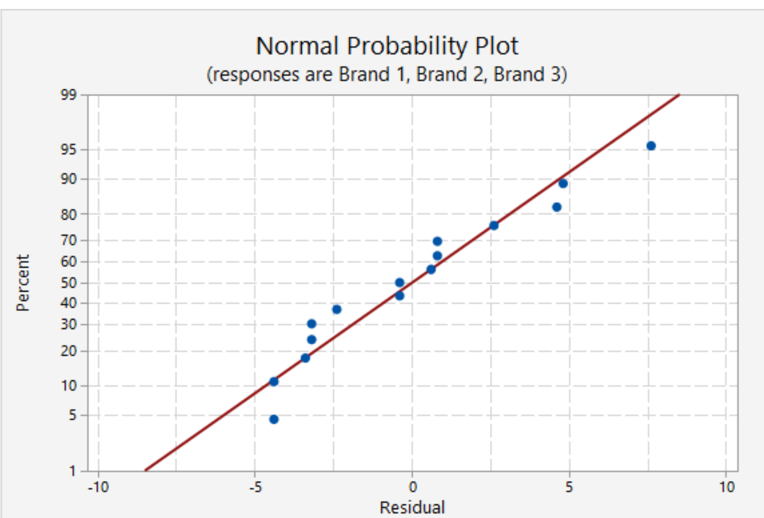
Use the modified Levene test to determine if the assumption of equal variances is satisfied in Problem 3.24. Use $\alpha = 0.05$*. Did you reach the same conclusion regarding the equality of the variances by examining residual plots?*

The modified Levene test evaluates the differences the deviations from each observation and the median of that treatment, so we must first calculate the $d_{ij} = |y_{ij} - \widetilde{y}_i|$ for each observation. Doing so generates the following table:

Brand 1	Brand 2	Brand 3
4	4	8
0	0	0
4	5	4
0	4	2
4	2	0

Based on the below ANOVA output using the derived differences, we cannot reject the null hypothesis at the 5% significance level. Therefore, we cannot conclude that there is a difference in the brands of batteries and we conclude that the assumption of equal variances is statisfied.

We would reach the same conclusion by examining the residual plot below (for the data of 3-24). The plotted points are relatively linear and would form a sufficient basis for contradicting the equality assumption.



One-Way ANOVA: D1, D2, D3

9/22/2014 5:40:17 PM

Method

Null hypothesis	All means are equal
Alternative hypothesis	At least one mean is different
Significance level	α = 0.05

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	3	D1, D2, D3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	0.9333333	0.46666667	0.07	0.9328
Error	12	80.0000000	6.66666667		
Total	14	80.9333333			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.58198890	1.15%	0.00%	0.00%

Means

Factor	N	Mean	StDev	95% CI
D1	5	2.4000	2.1909	(-0.1159, 4.9159)
D2	5	3.0000	2.0000	(0.4841, 5.5159)
D3	5	2.800	3.347	(0.284, 5.316)

Pooled StDev = 2.58198890