STAT 8120 – Module 4 Homework

Due 2/9/2020

Connor Armstrong

***4.12*** *Consider the hardness testing experiment described in Section 4.1. Suppose that the experiment was conducted as described and that the following Rockwell C-scale data (coded by subtracting 40 units) obtained:*

**4.12 Conditions**

|  |
| --- |
| *4.123 For the REST OF THE SEMESTER, always give Residual Analysis for any problem in SG3 Table 7 order. Use RA comparisons to justify the need for a transformation (with vs. without). Report multiple Tukey comparisons using a PC Bar Plot, discuss sigma distances. Give factor levels producing an optimum.* |

**Table 4.12.1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coupon | | | |
| Tip | | 1 | 2 | 3 | 4 |
| 1 | | 9.3 | 9.4 | 9.6 | 10.0 |
| 2 | | 9.4 | 9.3 | 9.8 | 9.9 |
| 3 | | 9.2 | 9.4 | 9.5 | 9.7 |
| 4 | | 9.7 | 9.6 | 10.0 | 10.2 |

***4.12.a*** *Analyze the data from this experiment.*

The data in Table 4.12.1 was processed using Minitab. The following ANOVA table was included in the results:

**Analysis of Variance for 4.12.a**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Tip | 3 | 0.38500 | 0.128333 | 14.44 | 0.001 |
| Coupon | 3 | 0.82500 | 0.275000 | 30.94 | 0.000 |
| Error | 9 | 0.08000 | 0.008889 |  |  |
| Total | 15 | 1.29000 |  |  |  |

Given a P-value of 0.001 and 0.000 respectively from the ANOVA table above, and given a significance level of 0.05, the null hypothesis that Tip and Coupon do not affect the mean hardness ought to be rejected. Therefore, both Tip and Coupon influence the mean hardness, with a confidence level of 95%.

***4.12.b*** *Use the Fisher LSD method to make comparisons among the four tips to determine specifically which tips differ in mean hardness readings.*

According to the Tukey Method computed in Minitab below, the only tip which is significantly different from the others was tip #4.

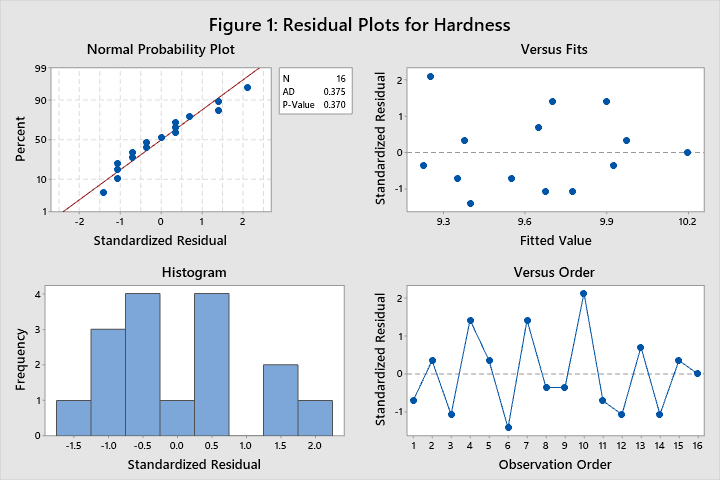
**Grouping Information Using the Tukey Method and 95% Confidence**

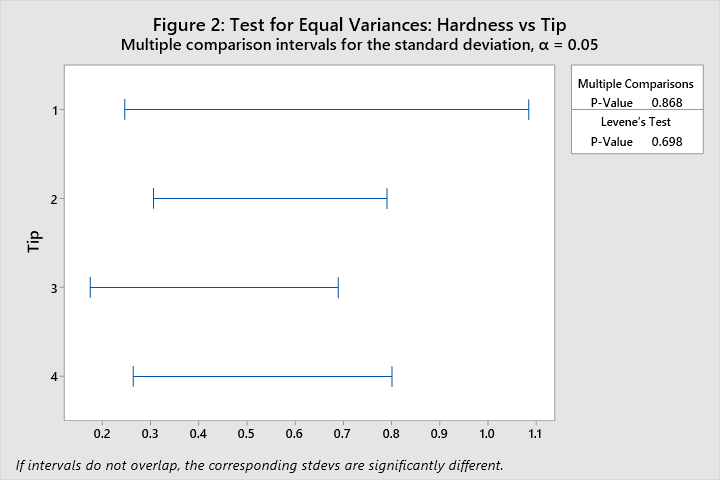
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tip** | **N** | **Mean** | **Grouping** | |
| 4 | 4 | 9.875 | A |  |
| 2 | 4 | 9.600 |  | B |
| 1 | 4 | 9.575 |  | B |
| 3 | 4 | 9.450 |  | B |

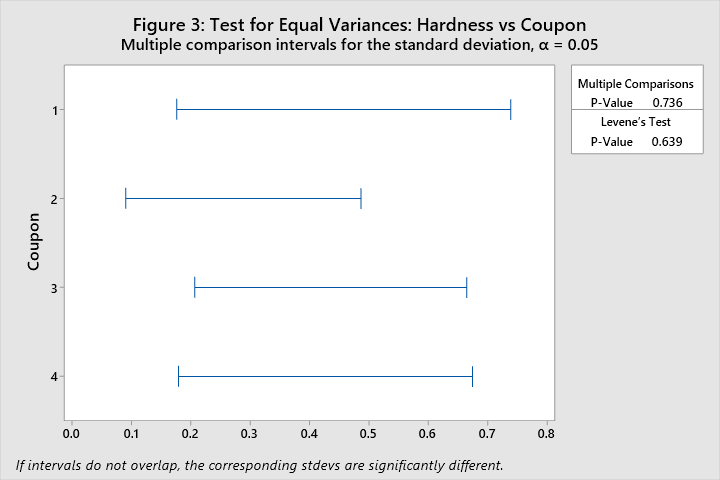
*Means that do not share a letter are significantly different.*

***4.12.c*** *Analyze the residuals from this experiment.*

The assumptions need to be verified to accept the above conclusion. The data passes the A.D. normality test with a p-value of 0.37. There is one outlier just outside 2 sigma. There can be expected that approximately 5% of data points will be found more than 2 standard deviations from the mean. The outlier will be noted, but not removed. Run order was not provided for this dataset, so independence will remain unverified for this exercise. Levene’s test was used to test for equal variance by tip and coupon separately (Figures 2 and 3), verifying the assumption of homogeneity of variance.







***4.27*** *The effect of five different ingredients (A, B, C, D, E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately 1 ½ hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use 𝛼 = 0.05) and draw conclusions.*

**Table 4.27.1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Day | | | | |
| Batch | | 1 | 2 | 3 | 4 | 5 |
| 1 | | A = 8 | B = 7 | D = 1 | C = 7 | E = 3 |
| 2 | | C = 11 | E = 2 | A = 7 | D = 3 | B = 8 |
| 3 | | B = 4 | A = 9 | C = 10 | E = 1 | D = 5 |
| 4 | | D = 6 | C = 8 | E = 6 | B = 6 | A = 10 |
| 5 | | E = 4 | D = 2 | B = 3 | A = 8 | C = 8 |

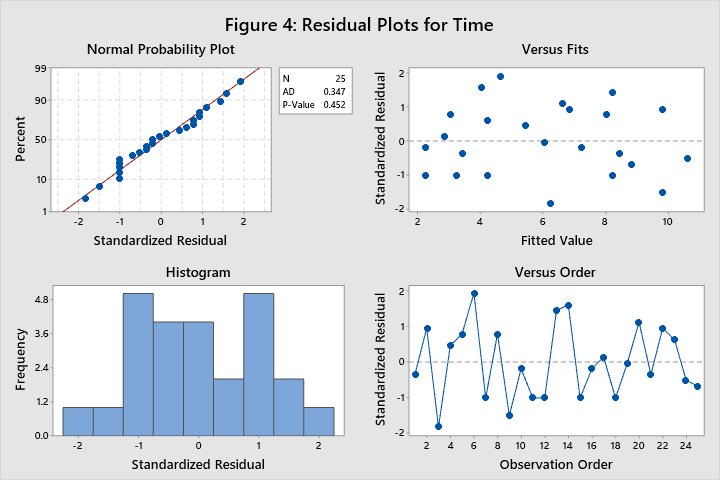
The data from table 4.27.1 was evaluated using Minitab Latin Square Analysis. See results below:

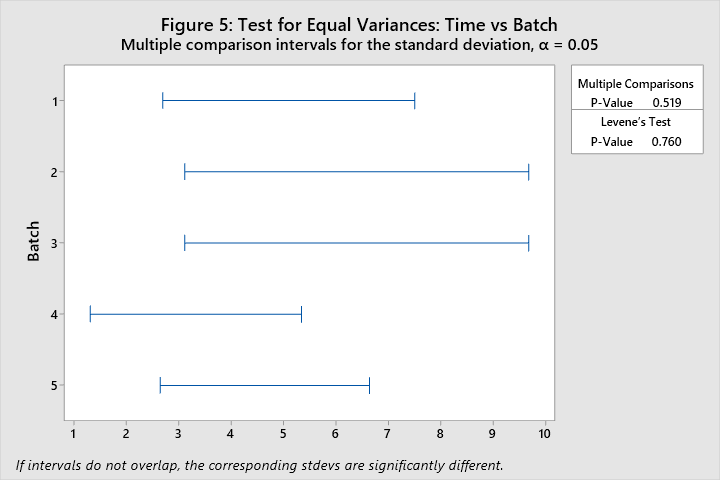
**Analysis of Variance for 4.27**

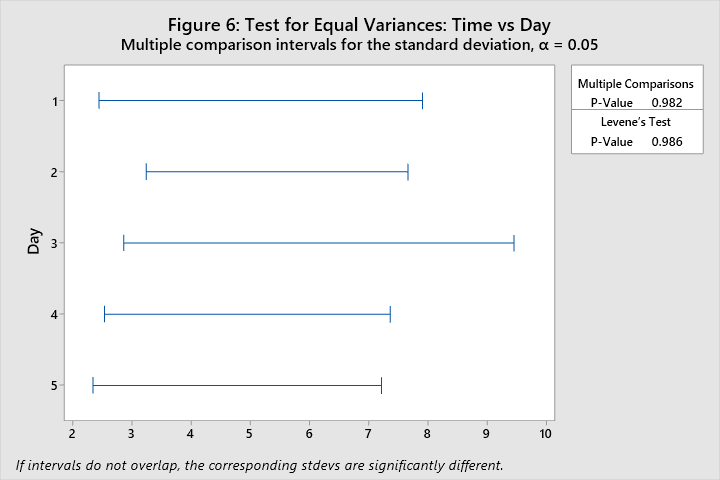
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Batch | 4 | 15.44 | 3.860 | 1.23 | 0.348 |
| Day | 4 | 12.24 | 3.060 | 0.98 | 0.455 |
| Catalyst | 4 | 141.44 | 35.360 | 11.31 | 0.000 |
| Error | 12 | 37.52 | 3.127 |  |  |
| Total | 24 | 206.64 |  |  |  |

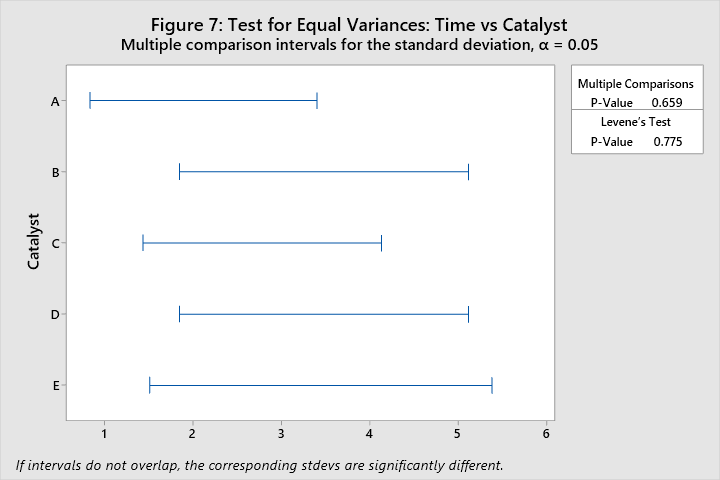
Given a P-value of 0.348, 0.455, and 0.000 respectively from the ANOVA table above, and given a significance level of 0.05, the variables Batch (P = 0.348) and Day (P = 0.455) do not affect the mean, but catalyst (P = 0.000) does affect the mean, with a confidence level of 95%. In other words, the means for different catalysts were different, but the means for different days and different batches were not. Acceptance of the conclusion of this analysis requires that the assumptions be validated.

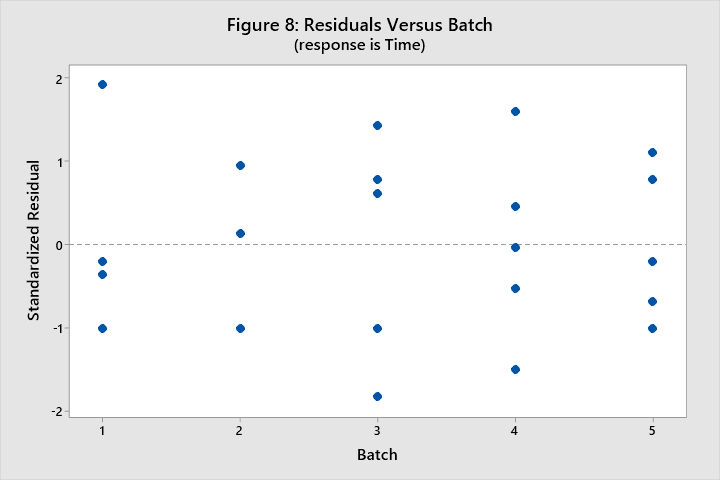
Per the A.D. test in figure 4, with a p-value of 0.452, the data is sufficiently normal. Refer to figures 5, 6, and 7 for verification of the homogeneity of variance assumption utilizing Levene’s test for batch, day, and catalyst respectively. There are no values outside +/- 2 sigma from the mean, upon inspection of figure 4 versus fits plot, satisfying the no outlier assumption. There is no pattern between batch number and the residuals of time in figure 8, which sufficiently satisfies the independence assumption.











The table below details the results of the Fisher Pairwise Comparison computation performed in Minitab. Mean times for catalysts C and A are significantly greater than the times for catalysts B, D, and E.

**Fisher Pairwise Comparisons: Catalyst**

**Grouping Information Using Fisher LSD Method and 95% Confidence**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Catalyst** | **N** | **Mean** | **Grouping** | |
| C | 5 | 8.8 | A |  |
| A | 5 | 8.4 | A |  |
| B | 5 | 5.6 |  | B |
| D | 5 | 3.4 |  | B |
| E | 5 | 3.2 |  | B |

*Means that do not share a letter are significantly different.*

***4.45*** *An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use 𝛼 = 0.05) and draw conclusions.*

**Table 4.45.1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Car | | | | |
| Additive | | 1 | 2 | 3 | 4 | 5 |
| 1 | |  | 17 | 14 | 13 | 12 |
| 2 | | 14 | 14 |  | 13 | 10 |
| 3 | | 12 |  | 13 | 12 | 9 |
| 4 | | 13 | 11 | 11 | 12 |  |
| 5 | | 11 | 12 | 10 |  | 8 |

The data were processed using Minitab to produce the ANOVA table below:

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Additive | 4 | 35.73 | 8.9333 | 9.81 | 0.001 |
| Car | 4 | 35.23 | 8.8083 | 9.67 | 0.001 |
| Error | 11 | 10.02 | 0.9106 |  |  |
| Total | 19 | 76.95 |  |  |  |

The incomplete block ANOVA analysis has determined that both the additive and the car influence the mean mileage observations, with p-values of 0.001 for each factor. With a confidence level of 95%, the means for the different factors are not necessarily all equal. A pairwise comparison method must be utilized to determine which groups are different from the others.

There are 4 groupings of means for the 5 additives. The Table below can be interpreted as follows: Additive 1 (Grouping A) is significantly different from Additives 3, 4, and 5, but not significantly different from Additive 2. Similarly, Additive 4 is significantly different from Additives 1 and 2, but not significantly different from Additives 3 and 5. Mean mileage for the additives in the experiment are similar to those which share the same Grouping letter, and significantly different from those which they do not share a grouping number with. Additive 1 has the highest mileage on average, but it has not been determined statistically to be significantly different from additive 2.

**Fisher Pairwise Comparisons: Additive**

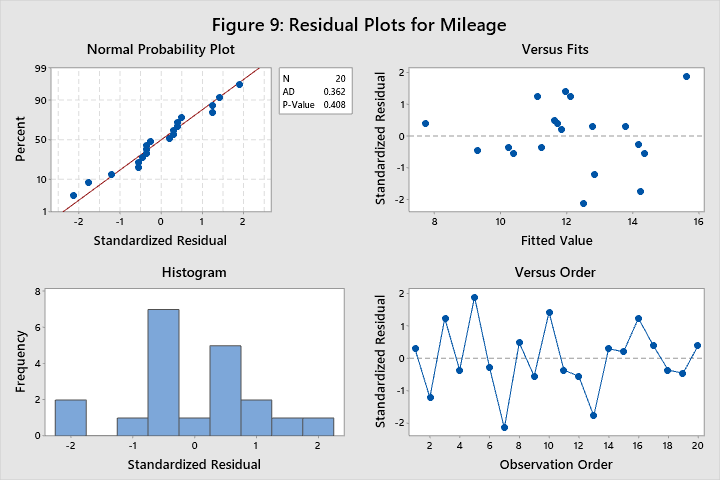
**Grouping Information Using Fisher LSD Method and 95% Confidence**

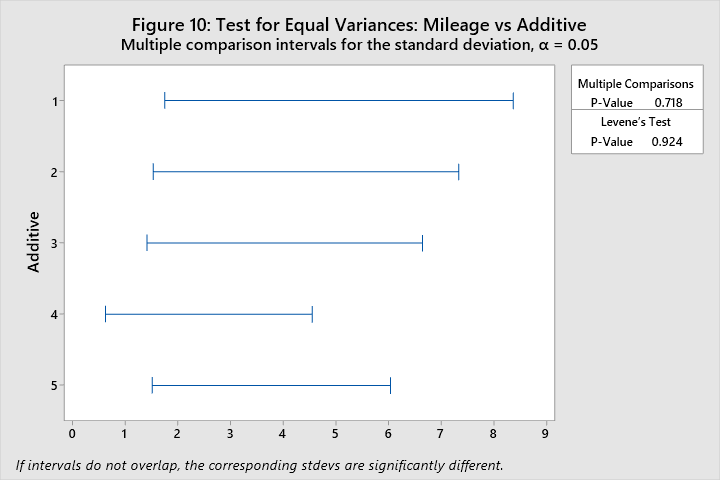
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Additive** | **N** | **Mean** | **Grouping** | | | |
| 1 | 4 | 14.2500 | A |  |  |  |
| 2 | 4 | 12.7833 | A | B |  |  |
| 3 | 4 | 11.8500 |  | B | C |  |
| 4 | 4 | 11.1167 |  |  | C | D |
| 5 | 4 | 10.2500 |  |  |  | D |

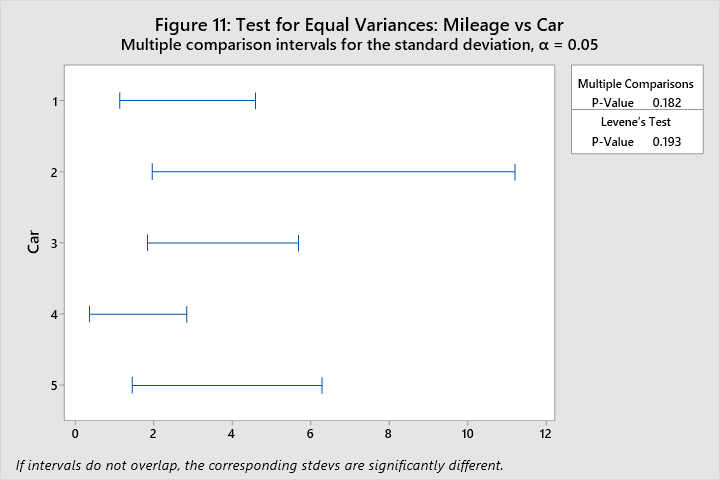
*Means that do not share a letter are significantly different.*

In order to accept these conclusions, the assumptions of ANOVA must be validated, which will be completed on the next page.

With a p-value of 0.408, the data passes the A.D. test for normality seen in figure 9 below. The variance by factor must be analyzed, this is achieved using Levene’s test of homogeneity for Additive and Car and can be seen in Figures 10 and 11. Having p-values of >0.05, both factors pass the verification of the homogeneity of variance assumption. There is one outlier just outside 2 sigma (-2.14 σ). It is expected that approximately 5% of data points will be found more than 2 standard deviations from the mean. The outlier will be noted, but not removed.







SAS Output and Code is below.

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Additive** | 4 | 31.70000000 | 7.92500000 | 8.70 | 0.0020 |
| **Car** | 4 | 35.23333333 | 8.80833333 | 9.67 | 0.0013 |





| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.964921 | **Pr < W** | 0.6461 |
| **Kolmogorov-Smirnov** | **D** | 0.148767 | **Pr > D** | >0.1500 |
| **Cramer-von Mises** | **W-Sq** | 0.063358 | **Pr > W-Sq** | >0.2500 |
| **Anderson-Darling** | **A-Sq** | 0.362444 | **Pr > A-Sq** | >0.2500 |

/\*

STAT 8120 - Module 3 Lab

\*/

libname mod4 "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 4";

**run**;

**proc** **import** datafile = "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 4\S8120Ch4Data122317.xlsx"

out = mod4.q3

DBMS = xlsx

Replace;

sheet = "Lab 3";

**run**;

**proc** **import** datafile = "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 4\S8120Ch4Data122317.xlsx"

out = mod4.q45

DBMS = xlsx

Replace;

sheet = "Prob4.45";

**run**;

ods rtf;

ods graphics on;

**proc** **glm** data = mod4.q45 plots=diagnostics;

class additive car;

model mileage = additive car;

means additive / HOVTEST=levene;

means car / HOVTEST=levene;

output out=stdres student=stdresidual;

Title "SAS RCBD for Question 4.45";

**run**;

means additive/tukey;

**run**;

contrast "Lowest vs. All Others" Additive **3** -**1** -**1** -**1**;

**run**;

**proc** **univariate** data = stdres normal;

var stdresidual;

qqplot stdresidual/normal(mu=est sigma=est);

histogram/normal;

**run**;

**proc** **sgplot** data = stdres;

scatter x=car y=stdresidual;

**run**;

ods graphics off;

ods rtf close;

**quit**;