STAT 8120 – Module 5 Homework

Due 3/1/2020

Connor Armstrong

***5.13*** *The factors that influence the breaking strength of a synthetic fiber are being studied. Four production machines and three operators are chosen and a factorial experiment is run using fiber from the same production batch. The results are as follows:*

**5.13 Conditions**

|  |
| --- |
| *5.133 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.*  *5.134 Example 3.1 and 3.6. Replace largest 725 by 250 and compare KW and standard methods.* |

**Table 5.13.1**

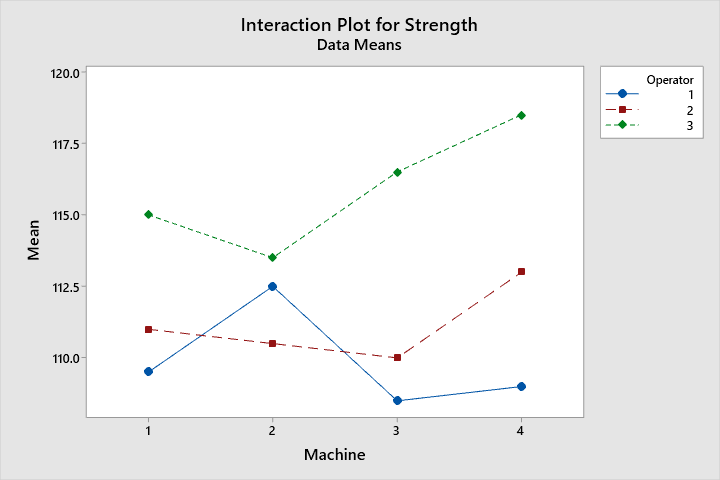
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Machine | | | |
| Operator | | 1 | 2 | 3 | 4 |
| 1 | | 109  110 | 110  115 | 108  109 | 110  108 |
| 2 | | 110  112 | 110  111 | 111  109 | 114  112 |
| 3 | | 116  114 | 112  115 | 114  119 | 120  117 |

***5.13.a*** *Analyze the data and draw conclusions. Use 𝛼 = 0.05.*

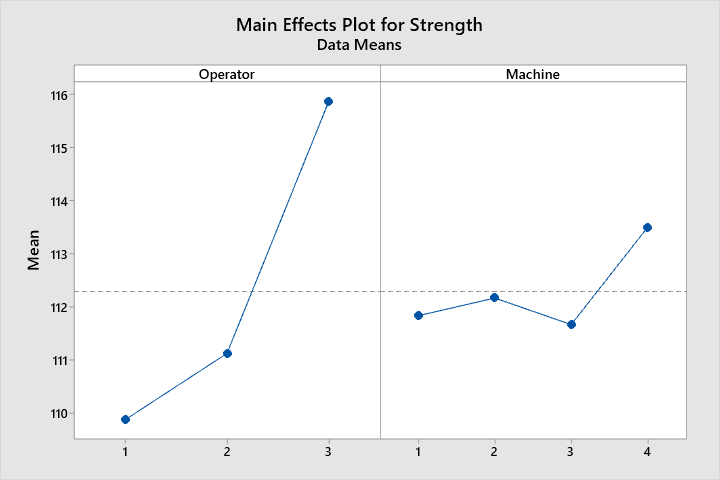
**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Operator | 2 | 160.33 | 80.167 | 21.14 | 0.000 |
| Machine | 3 | 12.46 | 4.153 | 1.10 | 0.389 |
| Operator\*Machine | 6 | 44.67 | 7.444 | 1.96 | 0.151 |
| Error | 12 | 45.50 | 3.792 |  |  |
| Total | 23 | 262.96 |  |  |  |

Only the operator factor is significant, with the standard significance level of 0.05. Therefore, pending the validation of the assumptions, there is enough evidence to reject the null hypothesis that the mean strength of synthetic fiber is equivalent between operators. There is insufficient evidence to reject the null hypotheses that mean strength by machine and the interaction term differ.

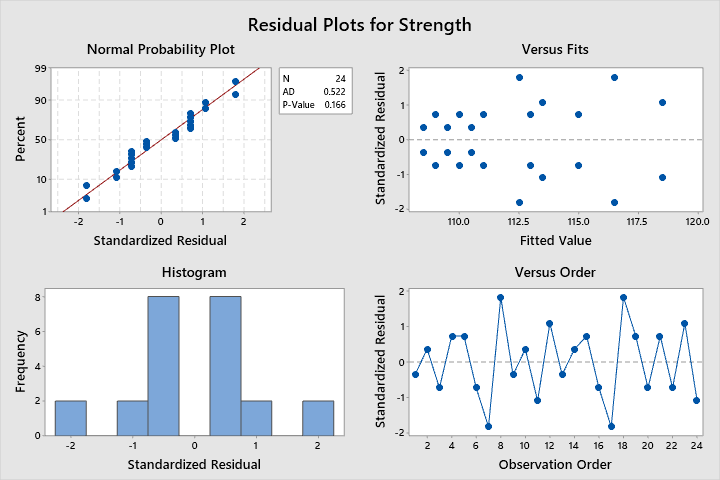


The interaction plot does not demonstrate significant interaction between factors.

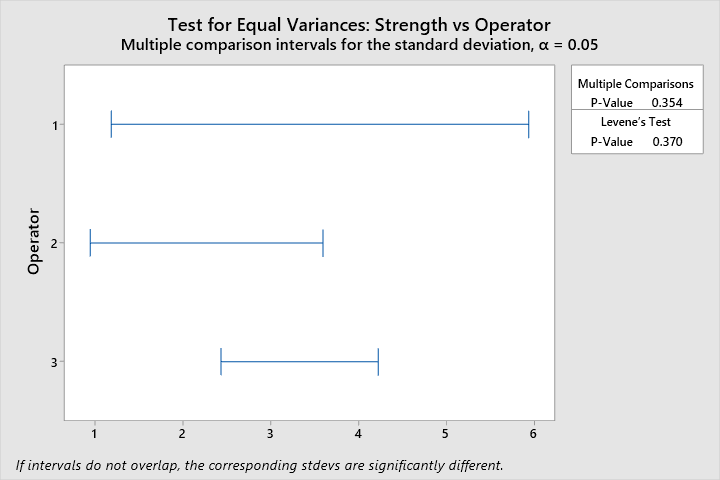


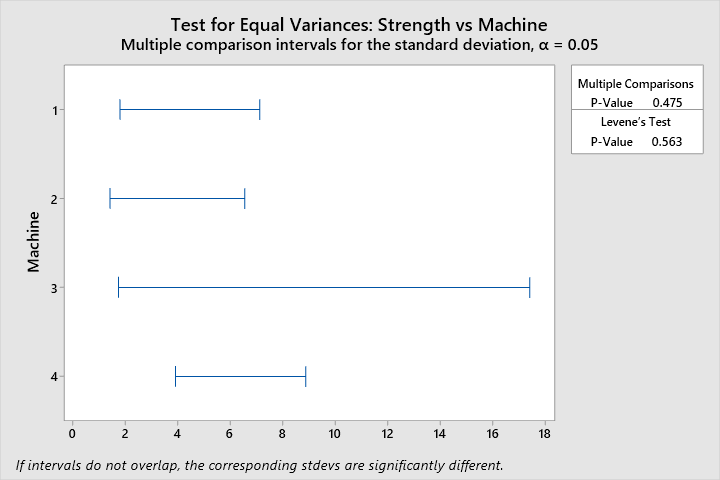
The main effect plot indicates a strong variance between the different Operators, and a relatively small variance among machines.

Inspection of the residual plots for strength (on the next page) shows a p-value of 0.166 for the Anderson-Darling test for normality, which satisfies the normality assumption for this analysis. In the absence of run order data, the independence assumption will remain unverified. The versus fits plot does not show standardized residuals beyond 2 standard deviations from the expected value, therefore there are no outliers.



Levene’s test of homogeneity of variance for strength vs. operator was computed using Minitab, resulting in a p-value of 0.370 which validates the homogeneity of variance assumption for strength vs operator. A p-value (plot on next page) of 0.563 for strength vs. machine validates the homogeneity of variance assumption for strength vs. machine.





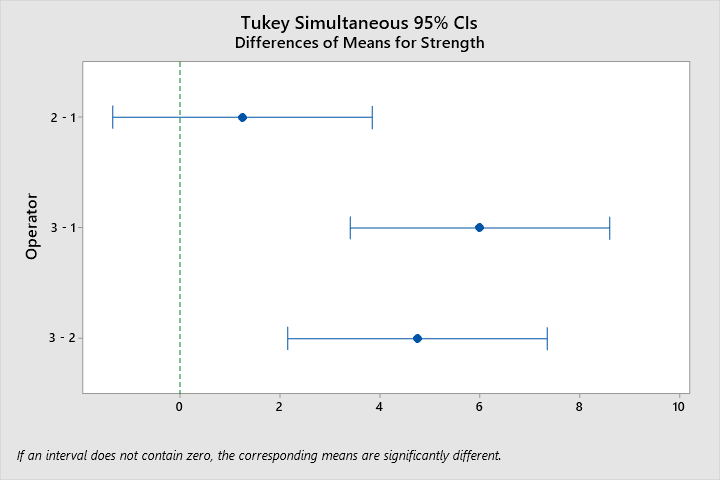
Tukey’s pairwise comparison method was utilized to determine which operator had the highest mean strength. The mean strength of 115.875 for operator 3 is significantly higher than operators 2 and 1, whose mean strengths are not significantly different.

**Tukey Pairwise Comparisons: Operator**

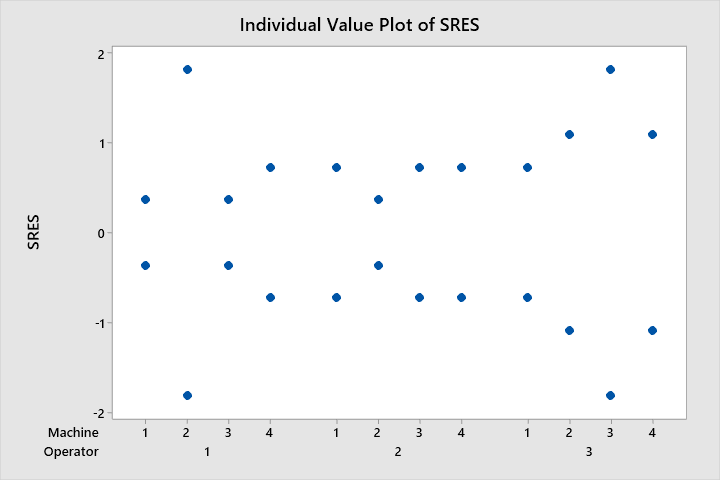
**Grouping Information Using the Tukey Method and 95% Confidence**

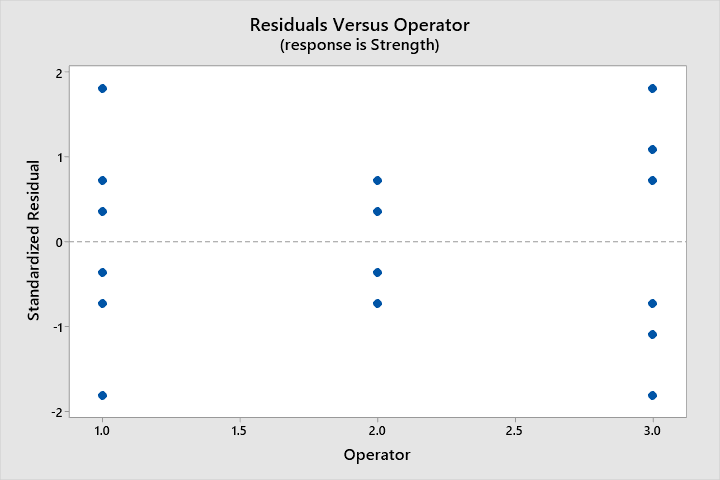
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Operator** | **N** | **Mean** | **Grouping** | |
| 3 | 8 | 115.875 | A |  |
| 2 | 8 | 111.125 |  | B |
| 1 | 8 | 109.875 |  | B |

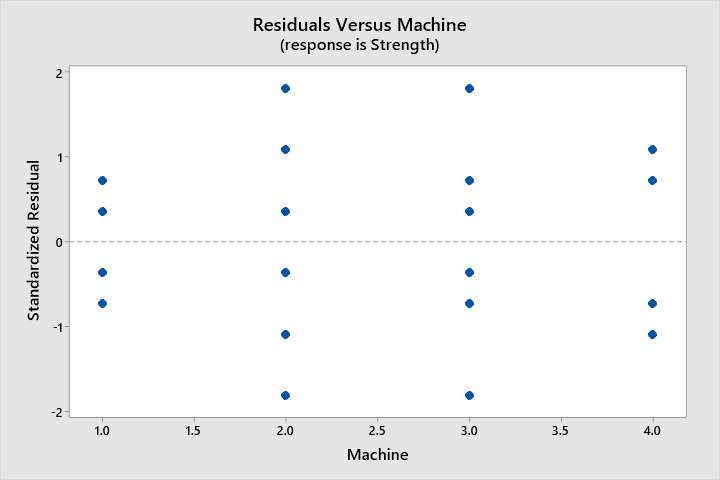
*Means that do not share a letter are significantly different.*



***5.13.b*** *Prepare appropriate residual plots and comment on the model’s adequacy.*

**





The residual plots do not indicate a significant difference in variances between factor levels for operator or machine. Refer to paragraph under the analysis of variance table for the discussion of model adequacy.

***5.23*** *The percentage of hardwood concentration in raw pulp, the vat pressure, and the cooking time of the pulp are being investigated for their effects on the strength of paper. Three levels of hardwood concentration, three levels of pressure, and two cooking times are selected. A factorial experiment with two replicates is conducted, and the following data are obtained:*

|  |  |  |  |
| --- | --- | --- | --- |
| **Cooking Time 3.0 Hours** | | | |
| **% Hardwood Concentration** | **Pressure** | | |
| **400** | **500** | **650** |
| **2** | 196.6  196.0 | 197.7  196.0 | 199.8  199.4 |
| **4** | 198.5  197.2 | 196.0  196.9 | 198.4  197.6 |
| **8** | 197.5  196.6 | 195.6  196.2 | 197.4  198.1 |

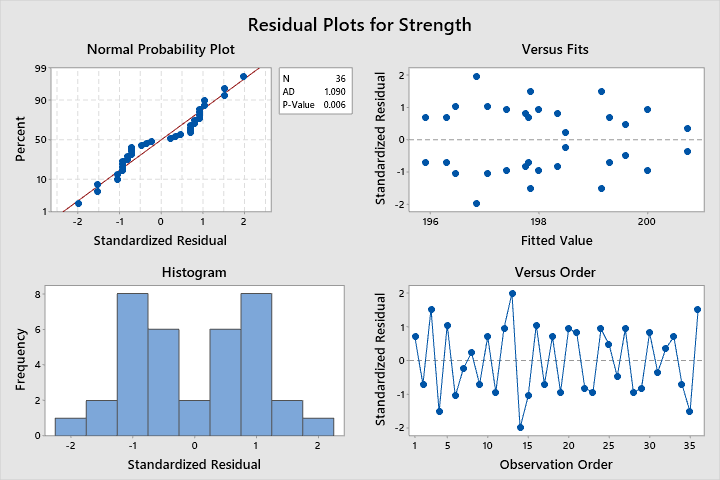
|  |  |  |  |
| --- | --- | --- | --- |
| **Cooking Time 4.0 Hours** | | | |
| **% Hardwood Concentration** | **Pressure** | | |
| **400** | **500** | **650** |
| **2** | 198.4  198.6 | 199.6  200.4 | 200.6  200.9 |
| **4** | 197.5  198.1 | 198.7  198.0 | 199.6  199.0 |
| **8** | 197.6  198.4 | 197.0  197.8 | 198.5  199.8 |

***5.23.a*** *Analyze the data and draw conclusions. Use 𝛼 = 0.05.*

**Analysis of Variance**

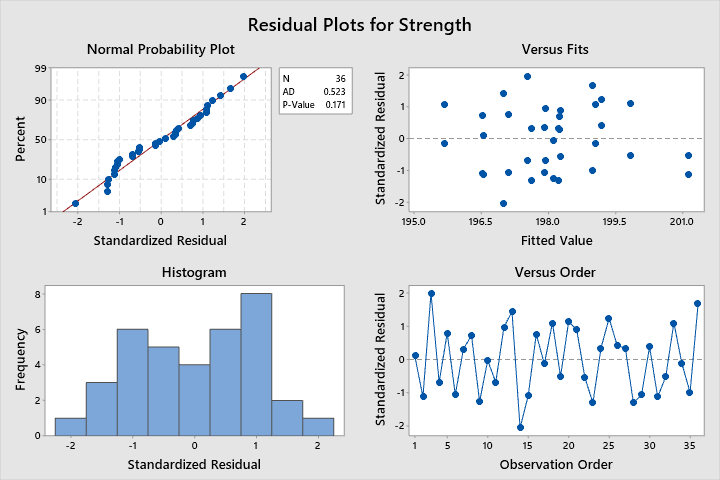
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Hardwood | 2 | 7.764 | 3.8819 | 10.62 | 0.001 |
| CookingTime | 1 | 20.250 | 20.2500 | 55.40 | 0.000 |
| Pressure | 2 | 19.374 | 9.6869 | 26.50 | 0.000 |
| Hardwood\*CookingTime | 2 | 2.082 | 1.0408 | 2.85 | 0.084 |
| Hardwood\*Pressure | 4 | 6.091 | 1.5228 | 4.17 | 0.015 |
| CookingTime\*Pressure | 2 | 2.195 | 1.0975 | 3.00 | 0.075 |
| Hardwood\*CookingTime\*Pressure | 4 | 1.973 | 0.4933 | 1.35 | 0.290 |
| Error | 18 | 6.580 | 0.3656 |  |  |
| Total | 35 | 66.309 |  |  |  |

All factors, including interaction terms, except for the 3rd order term Hardwood\*CookingTime\*Pressure (with a p-value of 0.290>0.25), are significant. The model does not pass the AD test for normality with a p-value of 0.006 as demonstrated in the following figure. Eliminating this term from the model results in a model which passes the AD normality test with a p-value of 0.171 as indicated in the second figure on the next page.



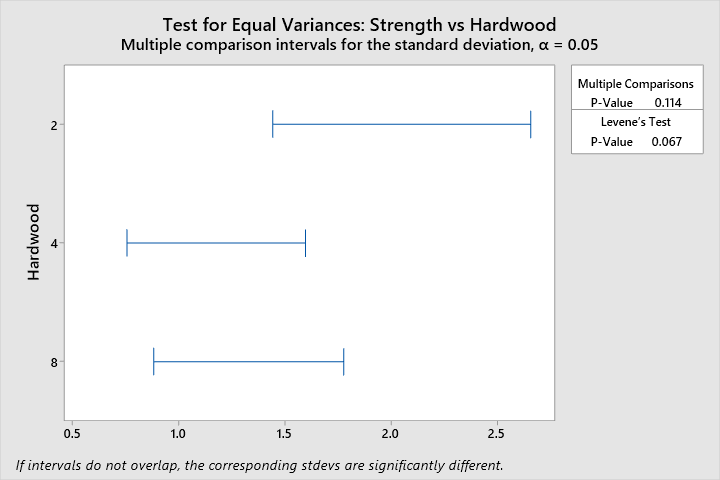
**Analysis of Variance Table for Model without 3rd Order Interaction Term**

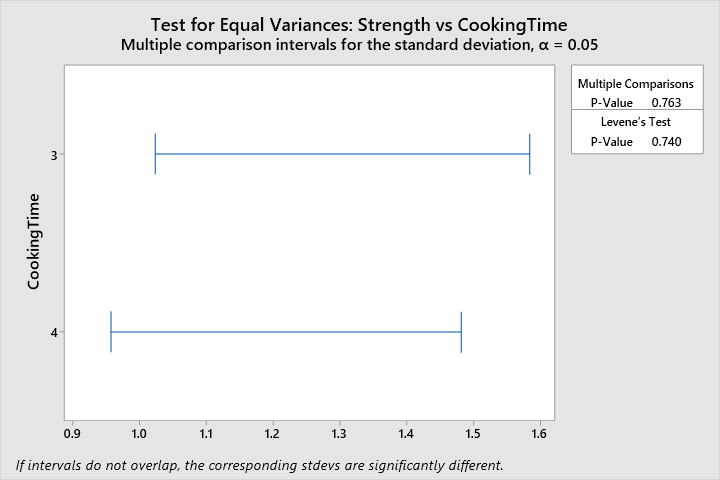
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Hardwood | 2 | 7.764 | 3.8819 | 9.98 | 0.001 |
| CookingTime | 1 | 20.250 | 20.2500 | 52.08 | 0.000 |
| Pressure | 2 | 19.374 | 9.6869 | 24.92 | 0.000 |
| Hardwood\*CookingTime | 2 | 2.082 | 1.0408 | 2.68 | 0.091 |
| Hardwood\*Pressure | 4 | 6.091 | 1.5228 | 3.92 | 0.015 |
| CookingTime\*Pressure | 2 | 2.195 | 1.0975 | 2.82 | 0.081 |
| Error | 22 | 8.553 | 0.3888 |  |  |
| Lack-of-Fit | 4 | 1.973 | 0.4933 | 1.35 | 0.290 |
| Pure Error | 18 | 6.580 | 0.3656 |  |  |
| Total | 35 | 66.309 |  |  |  |

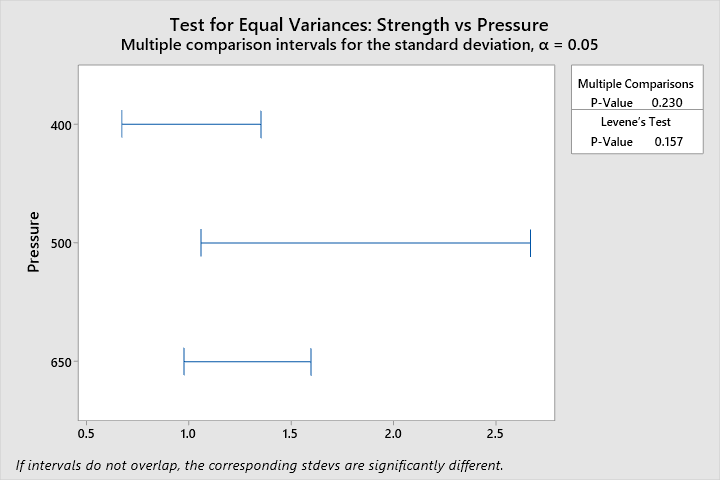


Referring to the ANOVA table for the altered model, all terms except for Hardwood\*CookingTime and CookingTime\*Pressure are significant with p-values <0.05. This indicates that Hardwood, CookingTime, Pressure, and Hardwood\*Pressure influence the mean strength and that the means between the different factor levels are likely not equivalent. The null hypotheses that means between factor levels for these factors are equivalent ought to be rejected, pending the validation of the assumptions.

As mentioned previously, the modified model without the 3rd order interaction term passes the AD normality test with a p-value of 0.171. There is one potential outlier, observation 14, with a standardized residual of -2.05. It is expected to find 5% of values outside 2 standard deviations from the mean in a normal distribution, so this value does not present a significant concern. Without run-order information, it is not possible to validate the assumption of independence. Levene’s test of homogeneity of variance will be implemented to evaluate the homogeneity of variance assumption for each factor.







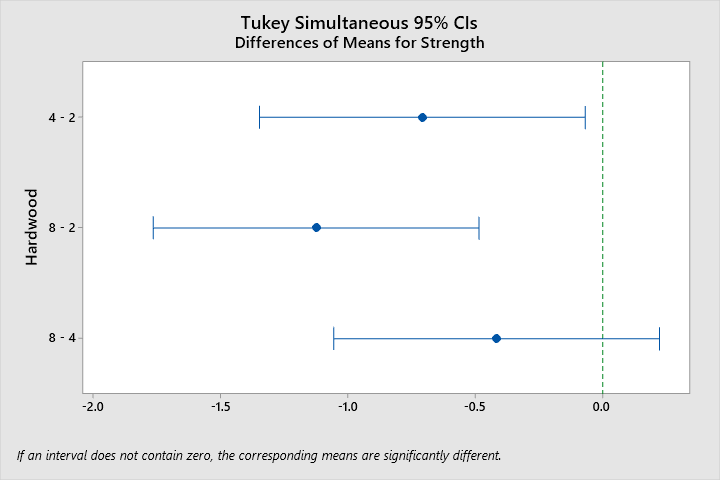
Having p-values >0.05 for Levene’s test of homogeneity of variance validates the homogeneity of variance assumption. Tukey’s Pairwise comparison method will be implemented to determine which combinations of factors means are different from the others.

**Tukey Pairwise Comparisons: Hardwood**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hardwood** | **N** | **Mean** | **Grouping** | |
| 2 | 12 | 198.667 | A |  |
| 4 | 12 | 197.958 |  | B |
| 8 | 12 | 197.542 |  | B |

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

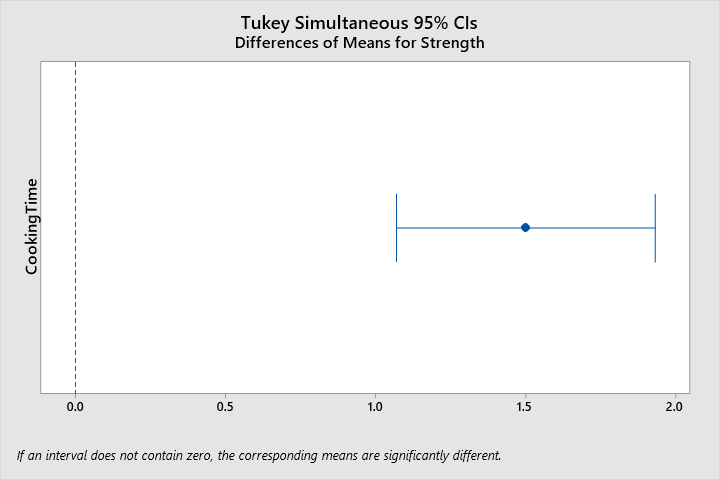


**Tukey Pairwise Comparisons: CookingTime**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CookingTime** | **N** | **Mean** | **Grouping** | |
| 4 | 18 | 198.806 | A |  |
| 3 | 18 | 197.306 |  | B |

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

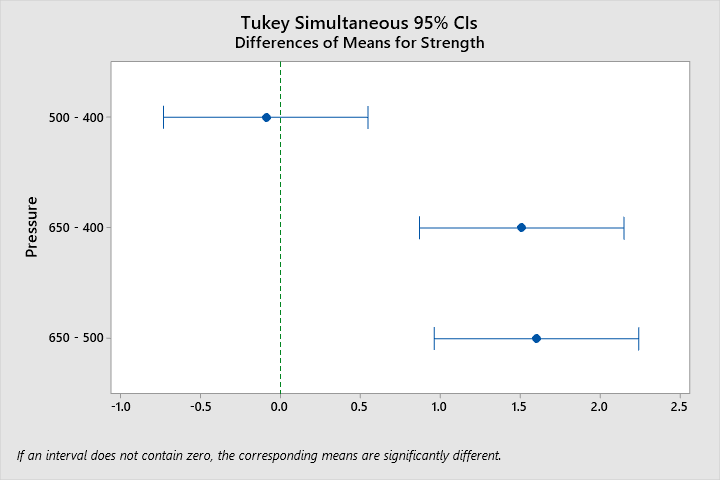


**Tukey Pairwise Comparisons: Pressure**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pressure** | **N** | **Mean** | **Grouping** | |
| 650 | 12 | 199.092 | A |  |
| 400 | 12 | 197.583 |  | B |
| 500 | 12 | 197.492 |  | B |

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

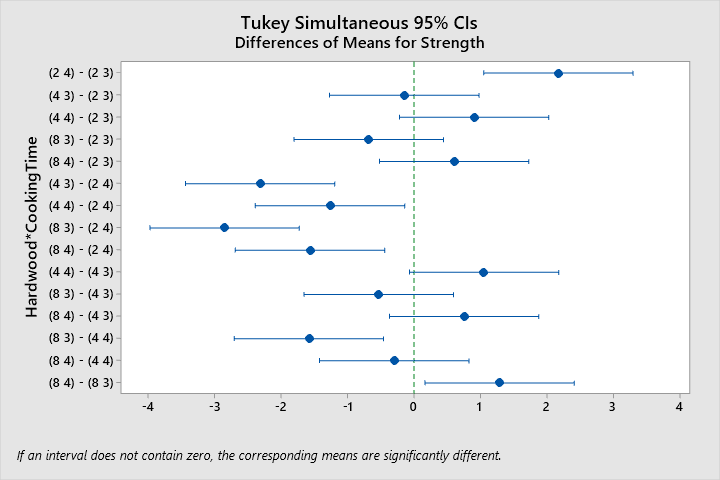


**Tukey Pairwise Comparisons: Hardwood\*CookingTime**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hardwood\*CookingTime** | **N** | **Mean** | **Grouping** | | |
| 2 4 | 6 | 199.750 | A |  |  |
| 4 4 | 6 | 198.483 |  | B |  |
| 8 4 | 6 | 198.183 |  | B |  |
| 2 3 | 6 | 197.583 |  | B | C |
| 4 3 | 6 | 197.433 |  | B | C |
| 8 3 | 6 | 196.900 |  |  | C |

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

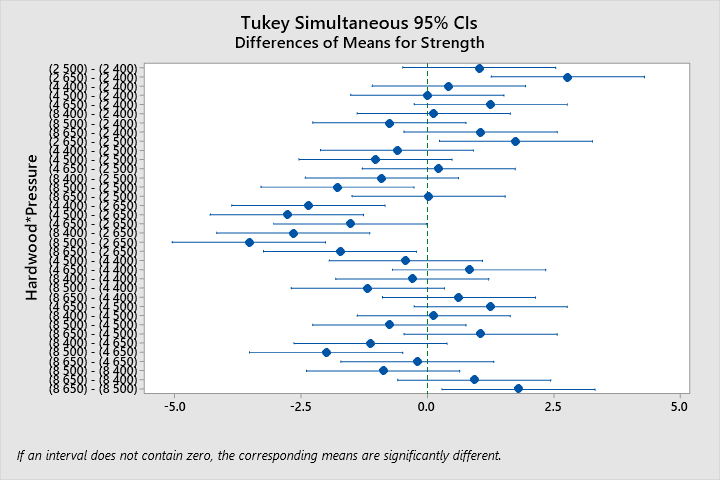


**Tukey Pairwise Comparisons: Hardwood\*Pressure**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hardwood\*Pressure** | **N** | **Mean** | **Grouping** | | |
| 2 650 | 4 | 200.175 | A |  |  |
| 4 650 | 4 | 198.650 |  | B |  |
| 8 650 | 4 | 198.450 |  | B |  |
| 2 500 | 4 | 198.425 |  | B |  |
| 4 400 | 4 | 197.825 |  | B | C |
| 8 400 | 4 | 197.525 |  | B | C |
| 2 400 | 4 | 197.400 |  | B | C |
| 4 500 | 4 | 197.400 |  | B | C |
| 8 500 | 4 | 196.650 |  |  | C |

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

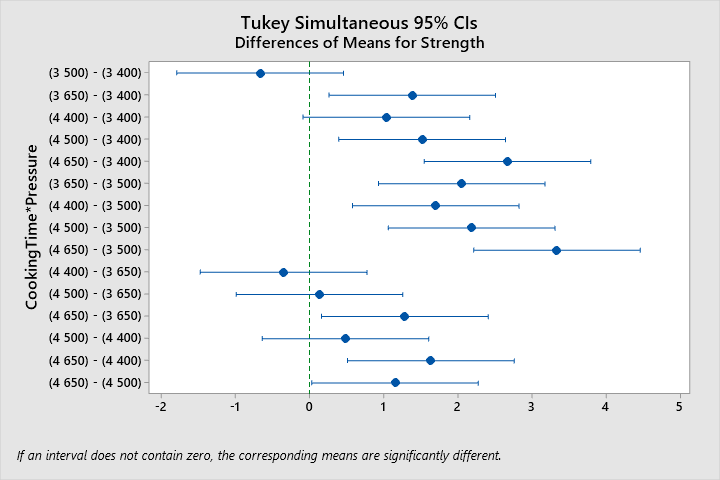


**Tukey Pairwise Comparisons: CookingTime\*Pressure**

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

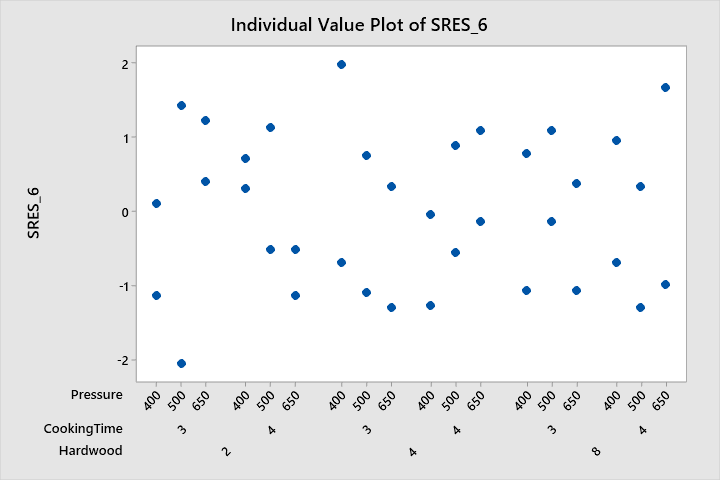
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CookingTime\*Pressure** | **N** | **Mean** | **Grouping** | | | |
| 4 650 | 6 | 199.733 | A |  |  |  |
| 4 500 | 6 | 198.583 |  | B |  |  |
| 3 650 | 6 | 198.450 |  | B |  |  |
| 4 400 | 6 | 198.100 |  | B | C |  |
| 3 400 | 6 | 197.067 |  |  | C | D |
| 3 500 | 6 | 196.400 |  |  |  | D |

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

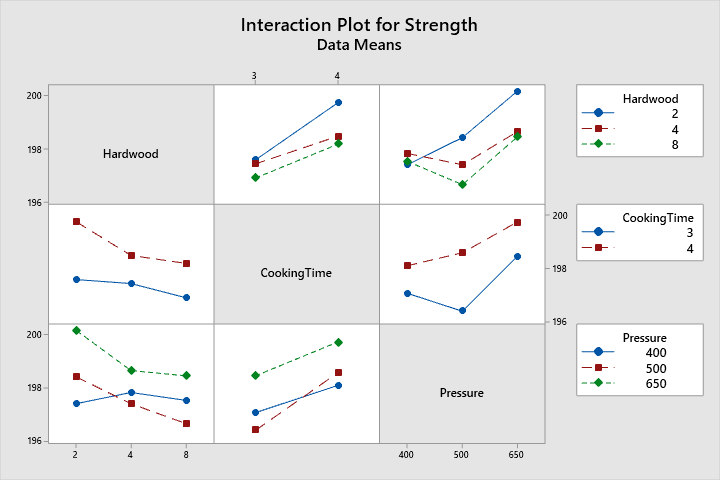


***5.23.b*** *Prepare appropriate residual plots and comment on the model’s adequacy.*

See above for commentary on the model adequacy. The residuals plotted against the factors does not indicate a concerning pattern or demonstrate a lack of homogeneity of variance.



***5.23.c*** *Under what set of conditions would you operate this process? Why?*



The interaction plot matrix does not present significant interaction effects. For the highest value of strength, this analysis indicates that the best factor levels are: Cooking time of 4, pressure of 650, and a Hardwood value of 2. This combination should be selected for the process because the Tukey pairwise comparison results support the conclusion that this combination results in the highest mean strength.

The analysis was also completed in SAS per the note (+) requirement. See below for code and relevant output. Discussions will not be repeated.

/\*STAT 8120 - Module 5 HW\*/

libname hw5 "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 5";

**run**;

**proc** **import** datafile = "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 5\S8120Ch5Data122317.xlsx"

out = hw5.q2

DBMS = xlsx

Replace;

sheet = "P5.23";

**run**;

ods rtf;

ods graphics on;

**proc** **glm** data = hw5.q2 plots=diagnostics;

class hardwood cookingtime pressure;

model strength = hardwood | cookingtime | pressure @**2**; \*@2 indicates a reduced model with highest order of 2;

output out = stdres student = stdresidual;

Title "SAS Factorial Design ANOVA for Problem 5.23";

**run**;

LSMeans hardwood | cookingtime | pressure @**2** / Pdiff = All;

**run**;

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

var stdresidual;

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

histogram/normal;

**run**;

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

scatter x=region y=stdresidual;

**run**;

ods graphics off;

ods rtf close;

**quit**;

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 13 | 57.75555556 | 4.44273504 | 11.43 | <.0001 |
| **Error** | 22 | 8.55333333 | 0.38878788 |  |  |
| **Corrected Total** | 35 | 66.30888889 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Strength Mean** |
| --- | --- | --- | --- |
| 0.871008 | 0.314825 | 0.623529 | 198.0556 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Hardwood** | 2 | 7.76388889 | 3.88194444 | 9.98 | 0.0008 |
| **CookingTime** | 1 | 20.25000000 | 20.25000000 | 52.08 | <.0001 |
| **Hardwood\*CookingTime** | 2 | 2.08166667 | 1.04083333 | 2.68 | 0.0911 |
| **Pressure** | 2 | 19.37388889 | 9.68694444 | 24.92 | <.0001 |
| **Hardwood\*Pressure** | 4 | 6.09111111 | 1.52277778 | 3.92 | 0.0150 |
| **CookingTime\*Pressure** | 2 | 2.19500000 | 1.09750000 | 2.82 | 0.0810 |



Least Square means Tukey multiple comparisons output:

| **Hardwood** | **CookingTime** | **Strength LSMEAN** | **LSMEAN Number** |
| --- | --- | --- | --- |
| **2** | **3** | 197.583333 | 1 |
| **2** | **4** | 199.750000 | 2 |
| **4** | **3** | 197.433333 | 3 |
| **4** | **4** | 198.483333 | 4 |
| **8** | **3** | 196.900000 | 5 |
| **8** | **4** | 198.183333 | 6 |

| **Least Squares Means for effect Hardwood\*CookingTime Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: Strength** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** | **4** | **5** | **6** |
| **1** |  | <.0001 | 0.9982 | 0.1670 | 0.4291 | 0.5662 |
| **2** | <.0001 |  | <.0001 | 0.0208 | <.0001 | 0.0031 |
| **3** | 0.9982 | <.0001 |  | 0.0753 | 0.6789 | 0.3313 |
| **4** | 0.1670 | 0.0208 | 0.0753 |  | 0.0027 | 0.9579 |
| **5** | 0.4291 | <.0001 | 0.6789 | 0.0027 |  | 0.0188 |
| **6** | 0.5662 | 0.0031 | 0.3313 | 0.9579 | 0.0188 |  |



| **Hardwood** | **Pressure** | **Strength LSMEAN** | **LSMEAN Number** |
| --- | --- | --- | --- |
| **2** | **400** | 197.400000 | 1 |
| **2** | **500** | 198.425000 | 2 |
| **2** | **650** | 200.175000 | 3 |
| **4** | **400** | 197.825000 | 4 |
| **4** | **500** | 197.400000 | 5 |
| **4** | **650** | 198.650000 | 6 |
| **8** | **400** | 197.525000 | 7 |
| **8** | **500** | 196.650000 | 8 |
| **8** | **650** | 198.450000 | 9 |

| **Least Squares Means for effect Hardwood\*Pressure Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: Strength** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **1** |  | 0.3698 | <.0001 | 0.9856 | 1.0000 | 0.1606 | 1.0000 | 0.7398 | 0.3406 |
| **2** | 0.3698 |  | 0.0154 | 0.9007 | 0.3698 | 0.9998 | 0.5334 | 0.0135 | 1.0000 |
| **3** | <.0001 | 0.0154 |  | 0.0007 | <.0001 | 0.0468 | 0.0001 | <.0001 | 0.0174 |
| **4** | 0.9856 | 0.9007 | 0.0007 |  | 0.9856 | 0.6384 | 0.9986 | 0.2167 | 0.8793 |
| **5** | 1.0000 | 0.3698 | <.0001 | 0.9856 |  | 0.1606 | 1.0000 | 0.7398 | 0.3406 |
| **6** | 0.1606 | 0.9998 | 0.0468 | 0.6384 | 0.1606 |  | 0.2616 | 0.0042 | 0.9999 |
| **7** | 1.0000 | 0.5334 | 0.0001 | 0.9986 | 1.0000 | 0.2616 |  | 0.5683 | 0.4989 |
| **8** | 0.7398 | 0.0135 | <.0001 | 0.2167 | 0.7398 | 0.0042 | 0.5683 |  | 0.0119 |
| **9** | 0.3406 | 1.0000 | 0.0174 | 0.8793 | 0.3406 | 0.9999 | 0.4989 | 0.0119 |  |



| **CookingTime** | **Pressure** | **Strength LSMEAN** | **LSMEAN Number** |
| --- | --- | --- | --- |
| **3** | **400** | 197.066667 | 1 |
| **3** | **500** | 196.400000 | 2 |
| **3** | **650** | 198.450000 | 3 |
| **4** | **400** | 198.100000 | 4 |
| **4** | **500** | 198.583333 | 5 |
| **4** | **650** | 199.733333 | 6 |

| **Least Squares Means for effect CookingTime\*Pressure Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: Strength** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** | **4** | **5** | **6** |
| **1** |  | 0.4555 | 0.0100 | 0.0826 | 0.0042 | <.0001 |
| **2** | 0.4555 |  | 0.0001 | 0.0013 | <.0001 | <.0001 |
| **3** | 0.0100 | 0.0001 |  | 0.9219 | 0.9990 | 0.0188 |
| **4** | 0.0826 | 0.0013 | 0.9219 |  | 0.7589 | 0.0020 |
| **5** | 0.0042 | <.0001 | 0.9990 | 0.7589 |  | 0.0423 |
| **6** | <.0001 | <.0001 | 0.0188 | 0.0020 | 0.0423 |  |



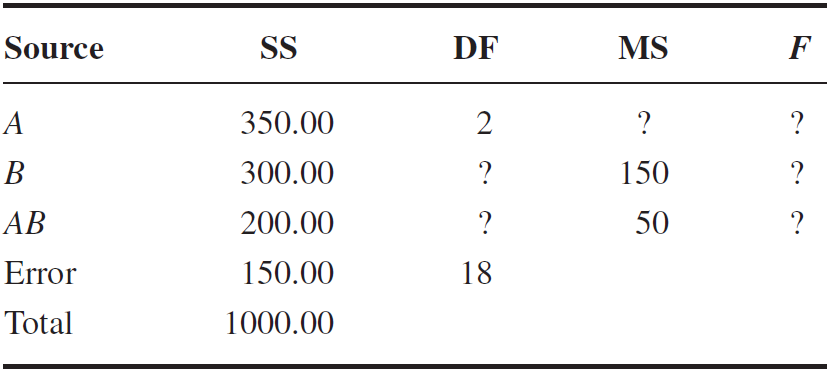
| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.963905 | **Pr < W** | 0.2828 |
| **Kolmogorov-Smirnov** | **D** | 0.113671 | **Pr > D** | >0.1500 |
| **Cramer-von Mises** | **W-Sq** | 0.084143 | **Pr > W-Sq** | 0.1835 |
| **Anderson-Darling** | **A-Sq** | 0.523492 | **Pr > A-Sq** | 0.1782 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| -2.05156 | 14 | 1.12836 | 20 |
| -1.29932 | 28 | 1.23093 | 25 |
| -1.29932 | 23 | 1.43609 | 13 |
| -1.26513 | 9 | 1.67544 | 36 |
| -1.12836 | 2 | 1.98317 | 3 |





***5.48*** *In Problem 4.58 you met physics PhD student Laura Van Ertia who had conducted a single-factor experiment in her pursuit of the unified theory. She is at it again, and this time she has moved on to a two-factor factorial conducted as a completely randomized design. From her experiment, Laura has constructed the following incomplete ANOVA display:*



***5.48.a*** *How many levels of factor B did she use in the experiment?*

***5.48.b*** *How many degrees of freedom are associated with interaction?*

***5.48.c*** *The error mean square is \_\_\_\_\_\_\_\_.*

***5.48.d*** *The mean square for factor A is \_\_\_\_\_\_\_\_.*

***5.48.e*** *How many replicates of the experiment were conducted?*

***5.48.f*** *What are your conclusions about interaction and the two main effects?*

The critical F-values at the standard significance level of determined from an F-table are:

The associated F statistics for factors A, B, and the interaction effect are all greater than the critical F-values determined above, therefore the factors A, B, and the interaction effect are all significant at the significance level.

***5.48.g*** *An estimate of the standard deviation of the response variable is \_\_\_\_\_\_\_\_.*

***5.48.h*** *If this experiment had been run in blocks there would have been \_\_\_\_\_\_\_\_ degrees of freedom for blocks.*

Per table 4.2 in Montgomery Design and Analysis of Experiments 9th ed (2017),