STAT 8120 – Applied Experimental Design

Lab 5 Report – March 15, 2020

Connor Armstrong

*The purpose of this report is to fulfill the requirement for Module 6 Lab, according to the supplied lab documentation, LabModule6.docx. Minitab is utilized as an analytical tool to address the questions in the lab document.*

***Problem 1:*** *Assume that a study identified 3 potentially important factors for influencing Y=check processing time (minutes), a key customer satisfaction variable for a bank service department. The three variables listed in Table 5 are to being by selecting logical extremes of the three factors.*

**Table 5. Check Processing Study Factors for Question 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Factor** | **Name** | **Low Level (-)** | **Current** | **High Level (+)** |
| A | Preprocessing Steps | 0 | 1 | 5 |
| B | Scanning Time (sec.) | 5 | 10 | 20 |
| C | Check Design | Current | Current | New |

*To gain an understanding of the 23 methodology, assume the impact on Y is given in Table 6. Typically, the main point of an experiment is to estimate these effects.*

**Table 6. Factor Effects for Question 1(a)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Factor** | **-** | **+** | **Effect** |
| A | 3 | -3 | -6 |
| B | 1 | 1 | 0 |
| C | 0 | -4 | -4 |

**Table 7. Design Matrix for Bank Check Processing**

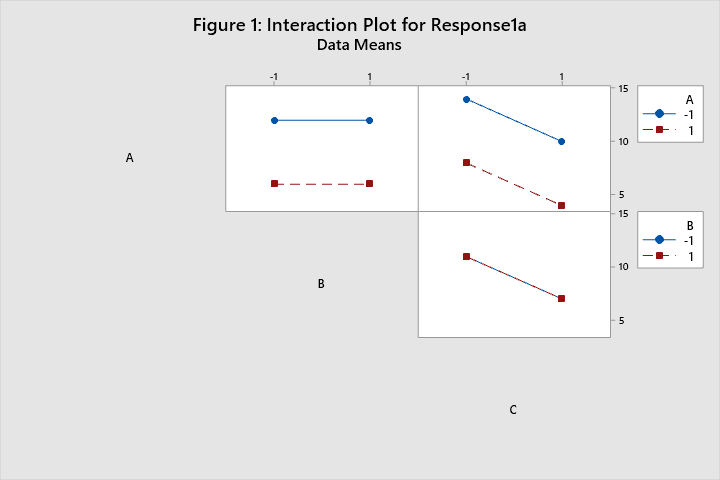
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Run** | **StdOrder** | **A** | **B** | **C** | **Response 1a**  **Ya=A+B+C+10** | **10** | **A** | **B** | **C** | **Response 1b** | | |
| **N(0,2)** | **Yb** | |
| 1 | (1) | - | - | - | 14 | 10 | 3 | 1 | 0 | 2 | | 16 |
| 2 | a | + | - | - | 8 | 10 | -3 | 1 | 0 | 2 | | 10 |
| 3 | b | - | + | - | 14 | 10 | 3 | 1 | 0 | 1 | | 15 |
| 4 | ab | + | + | - | 8 | 10 | -3 | 1 | 0 | -2 | | 6 |
| 5 | c | - | - | + | 10 | 10 | 3 | 1 | -4 | -2 | | 8 |
| 6 | ac | + | - | + | 4 | 10 | -3 | 1 | -4 | 0 | | 4 |
| 7 | bc | - | + | + | 10 | 10 | 3 | 1 | -4 | -3 | | 7 |
| 8 | abc | + | + | + | 4 | 10 | -3 | 1 | -4 | 0 | | 4 |

*Response 1a: Ya=10+A+B+C, Response 1b: Yb=10+A+B+C+N(0,2) with error.*

*Minitab Commands for N(0,σ=2) error*

1. *Calc>Random Data>Normal> rows=8,μ=0,σ=2,Store in c1 >ok*
2. *Calc>Store in c2>enter round(c1) →column N(0,2)*

***1.a*** *Assume μ=10 minutes is the current base for processing with NO factor effects. Construct responses using Table 6 factor effects and complete Table 7 for Response 1a. Estimate the factor main factor effects given in Table 6 and interactions using contrasts by hand and with Minitab. Review appropriate plots. Write a brief summary to management on the study conclusions and suggested follow-up.*

**

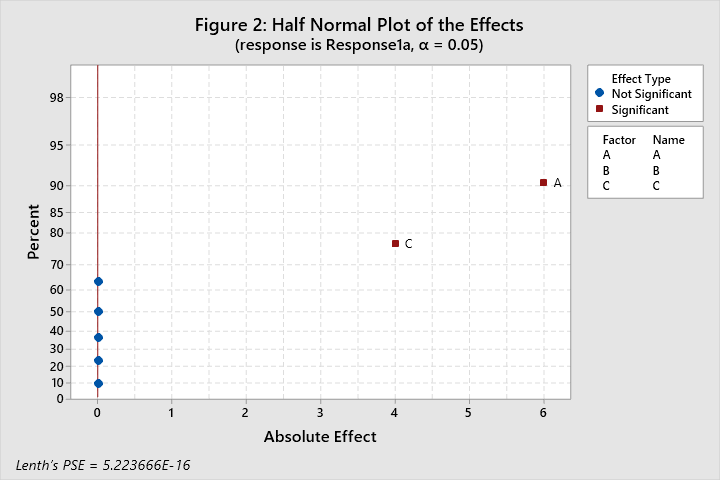
The data was processed using Minitab, and the subsequent model is displayed below. The degrees of freedom for error is 0, therefore the p-values for coefficients and in the ANOVA table cannot be computed. The factor effects can be computed, and are displayed in the coded coefficients table. The effect for the factor B cannot be seen, because it is constant. Therefore, it is accounted for in the constant coefficient. There are no significant interactions between the 3 variables, as confirmed by the interaction plot above (Figure 1) and the half-normal probability plot of the effects of the coefficients in Figure 2.

**Table 8: Coded Coefficients for the Model for Question 1.a**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **Effect** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant |  | 9.000 | \* | \* | \* |  |
| A | -6.000 | -3.000 | \* | \* | \* | 1.00 |
| B | 0.000000 | 0.000000 | \* | \* | \* | 1.00 |
| C | -4.000 | -2.000 | \* | \* | \* | 1.00 |
| A\*B | 0.000000 | 0.000000 | \* | \* | \* | 1.00 |
| A\*C | 0.000000 | 0.000000 | \* | \* | \* | 1.00 |
| B\*C | 0.000000 | 0.000000 | \* | \* | \* | 1.00 |
| A\*B\*C | 0.000000 | 0.000000 | \* | \* | \* | 1.00 |

**Regression Equation in Uncoded Units**

|  |  |  |
| --- | --- | --- |
| Response1a | = | 9.000 - 3.000 A + 0.000000 B - 2.000 C + 0.000000 A\*B + 0.000000 A\*C + 0.000000 B\*C + 0.000000 A\*B\*C |



The following is a summary for management.

Analysis of the results of the experiment indicate that the high level of 5 for preprocessing steps results in a net 6 minute decrease in check processing time compared to the low level of 0 steps. Scanning time did not result in a significant difference in check processing time. The new check design resulted in a net 4 minute decrease in check processing time. Accordingly, the recommendation of this analysis is to select 5 preprocessing steps and use the new check design to decrease check processing time. More testing, particularly between the number of preprocessing steps and check design, will help to confirm these conclusions.

***1.b*** *Assume Error ∼ N( is added to the factor effects. Generate 8 random deviates in Minitab, and place values in the left Response 1b column of Table 7. Use Minitab commands: Calc>Random Data>Normal. Round values to the nearest integer minute (Calc>Round). Use these 8 random responses and generate the factor effects as in part (a). Estimate the factor effects using only Minitab (no hand calculations). Pool interactions where appropriate and estimate effects. Note: Re-simulate error if your analysis produces no significant effects or an effect larger than the A effect.*

Refer to Table 7 to view the simulated and calculated data for error and the subsequent response for part 1.b. The factorial regression model was computed with the following results:

**Table 9: Coded Coefficients for the Full Model for Question 1.b**

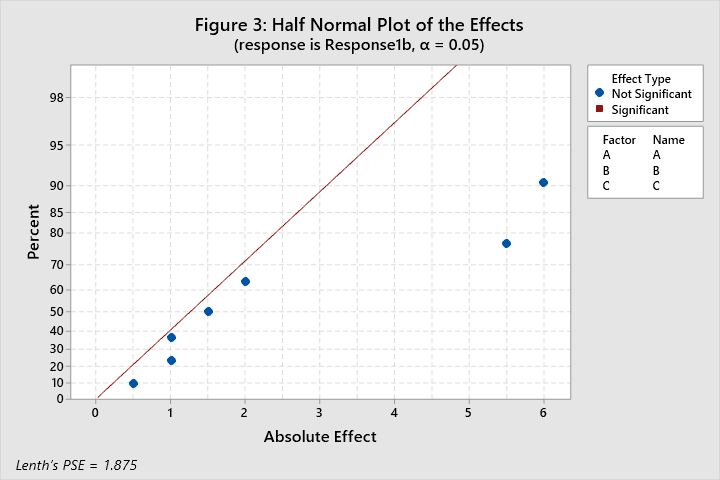
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **Effect** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant |  | 8.750 | \* | \* | \* |  |
| A | -5.500 | -2.750 | \* | \* | \* | 1.00 |
| B | -1.5000 | -0.7500 | \* | \* | \* | 1.00 |
| C | -6.000 | -3.000 | \* | \* | \* | 1.00 |
| A\*B | -0.5000 | -0.2500 | \* | \* | \* | 1.00 |
| A\*C | 2.000 | 1.000 | \* | \* | \* | 1.00 |
| B\*C | 1.0000 | 0.5000 | \* | \* | \* | 1.00 |
| A\*B\*C | 1.0000 | 0.5000 | \* | \* | \* | 1.00 |

**Table 10: Analysis of Variance for the Full Model for Question 1.b**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Model | 7 | 149.500 | 21.3571 | \* | \* |
| Linear | 3 | 137.000 | 45.6667 | \* | \* |
| A | 1 | 60.500 | 60.5000 | \* | \* |
| B | 1 | 4.500 | 4.5000 | \* | \* |
| C | 1 | 72.000 | 72.0000 | \* | \* |
| 2-Way Interactions | 3 | 10.500 | 3.5000 | \* | \* |
| A\*B | 1 | 0.500 | 0.5000 | \* | \* |
| A\*C | 1 | 8.000 | 8.0000 | \* | \* |
| B\*C | 1 | 2.000 | 2.0000 | \* | \* |
| 3-Way Interactions | 1 | 2.000 | 2.0000 | \* | \* |
| A\*B\*C | 1 | 2.000 | 2.0000 | \* | \* |
| Error | 0 | \* | \* |  |  |
| Total | 7 | 149.500 |  |  |  |

**Regression Equation in Uncoded Units**

|  |  |  |
| --- | --- | --- |
| Response1b | = | 8.750 - 2.750 A - 0.7500 B - 3.000 C - 0.2500 A\*B + 1.000 A\*C + 0.5000 B\*C + 0.5000 A\*B\*C |





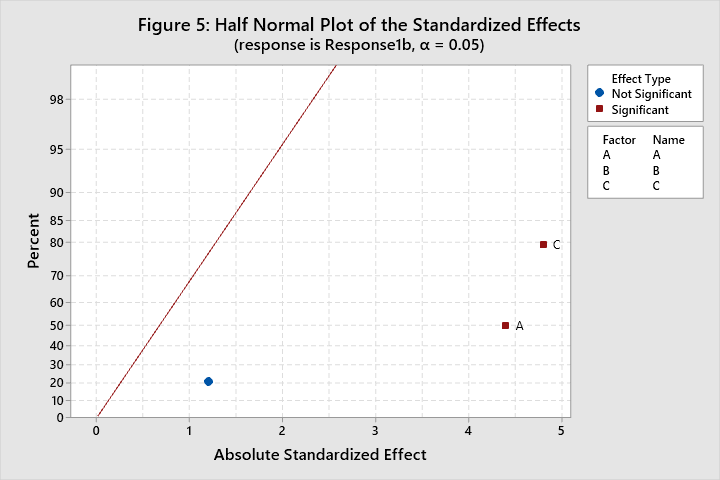
Having 0 degrees of freedom for error, the p-values cannot be computed. The half-normal probability plot of the effects in figure 3 indicates that no effects are significant. The interaction matrix in Figure 4 indicates no significant interaction. Removing the interaction terms will allow the p-values to be computed. As expected, the p-values for factors A and C are significant at the 0.05 significance level, and the factor B is not significant. Figure 5 supports the conclusion that factors A and C are significant while B is not.

**Table 11: Coded Coefficients for the Reduced Model for Question 1.b**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **Effect** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant |  | 8.750 | 0.625 | 14.00 | 0.000 |  |
| A | -5.500 | -2.750 | 0.625 | -4.40 | 0.012 | 1.00 |
| B | -1.500 | -0.750 | 0.625 | -1.20 | 0.296 | 1.00 |
| C | -6.000 | -3.000 | 0.625 | -4.80 | 0.009 | 1.00 |

**Table 12: Analysis of Variance for the Reduced Model for Question 1.b**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Model | 3 | 137.000 | 45.667 | 14.61 | 0.013 |
| Linear | 3 | 137.000 | 45.667 | 14.61 | 0.013 |
| A | 1 | 60.500 | 60.500 | 19.36 | 0.012 |
| B | 1 | 4.500 | 4.500 | 1.44 | 0.296 |
| C | 1 | 72.000 | 72.000 | 23.04 | 0.009 |
| Error | 4 | 12.500 | 3.125 |  |  |
| Total | 7 | 149.500 |  |  |  |



Figures 6-9 are the residual plots for the model. The residuals are normally distributed per the Anderson-Darling normality test with a p-value of 0.836. There are no standardized residuals beyond 2 sigma, validating the absence of outliers assumption. The residual plots versus the factors do not demonstrate a significant non-homogeneity of variance concern. Absence of run-order data prevents validation of the assumption of independence.

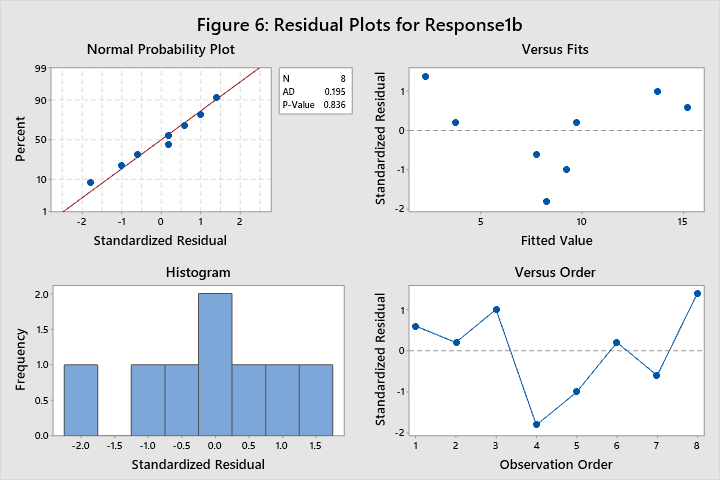
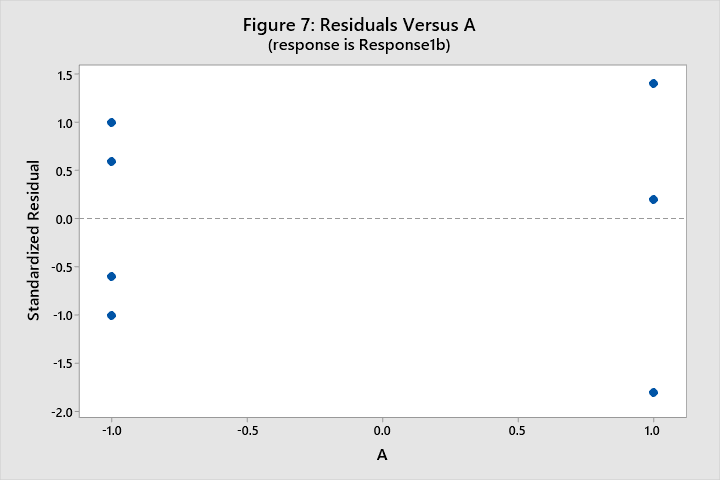
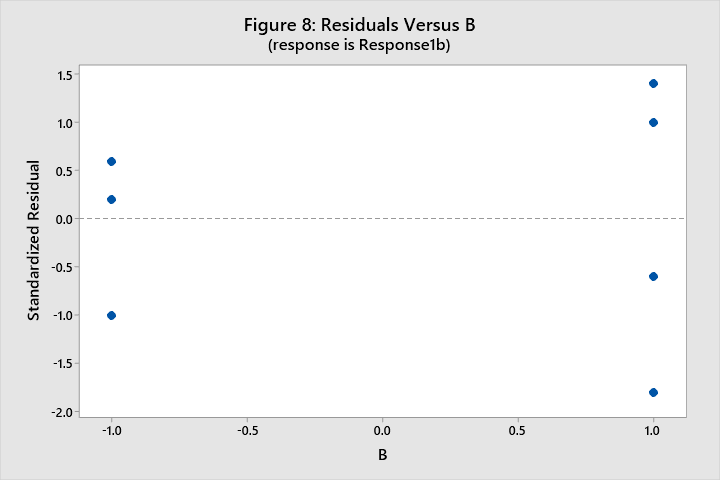
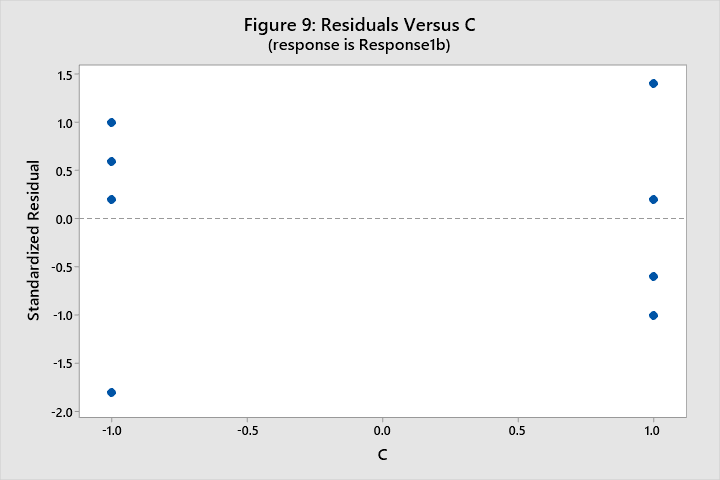
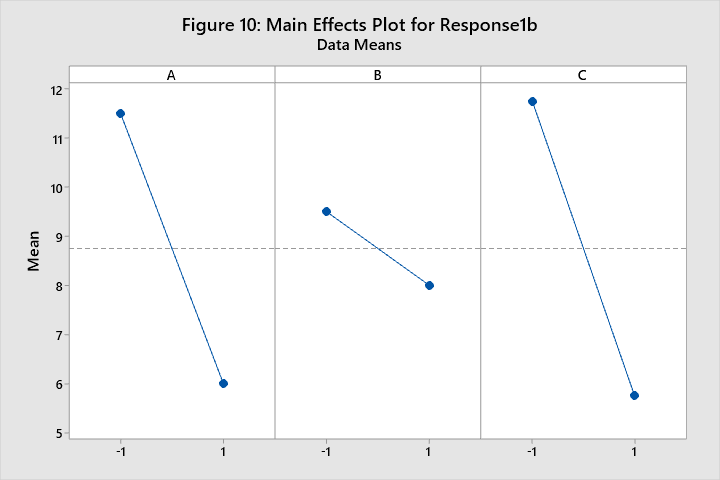
   

Figure 10 displays the main effects for the factors in the model. The high values for A and C result in lower mean processing time, while B does not make a significant difference. Figure 4 supports the conclusion of an absence of significant interactions.



***1.c*** *Compare the results in parts (a) and (b) in a table and suggest conclusions on the impact of error on factorial experiments. Consider σ-distance of effects.*

**Table 13: Summary and Comparison of Conclusions for Questions 1.a and 1.b**

|  |  |  |
| --- | --- | --- |
|  | **Question 1.a** | **Question 1.b** |
| *Calculated Effect for A* | -6 | -5.5 |
| *Calculated Effect for B* | 0 | -1.5 |
| *Calculated Effect for C* | -4 | -6 |
| *Significance of Interaction Terms* | No | No |
| *Able to compute p-values?* | No | Yes, without full model |
| *Presence of Error* | No | Yes, N(0, 2) |
| *Which Factor levels yield best processing time?* | A=1, B doesn’t matter, C=1 | A=1, B doesn’t matter, C=1 |

The presence of error in the data did not fundamentally alter the conclusions of the analysis, but did mask the complete truth of the significance and effects of the model terms. For example, while it is known for this exercise that the true effect for factor A is -6, the calculated effect for the model with error happened to be less than that of factor C, which is known to be -4. If the standard deviation of the error were larger the impact to the viability of the model would be more significant. Analysis of the residuals in a factorial model allows quantification of the impact of error on experimental analyses.