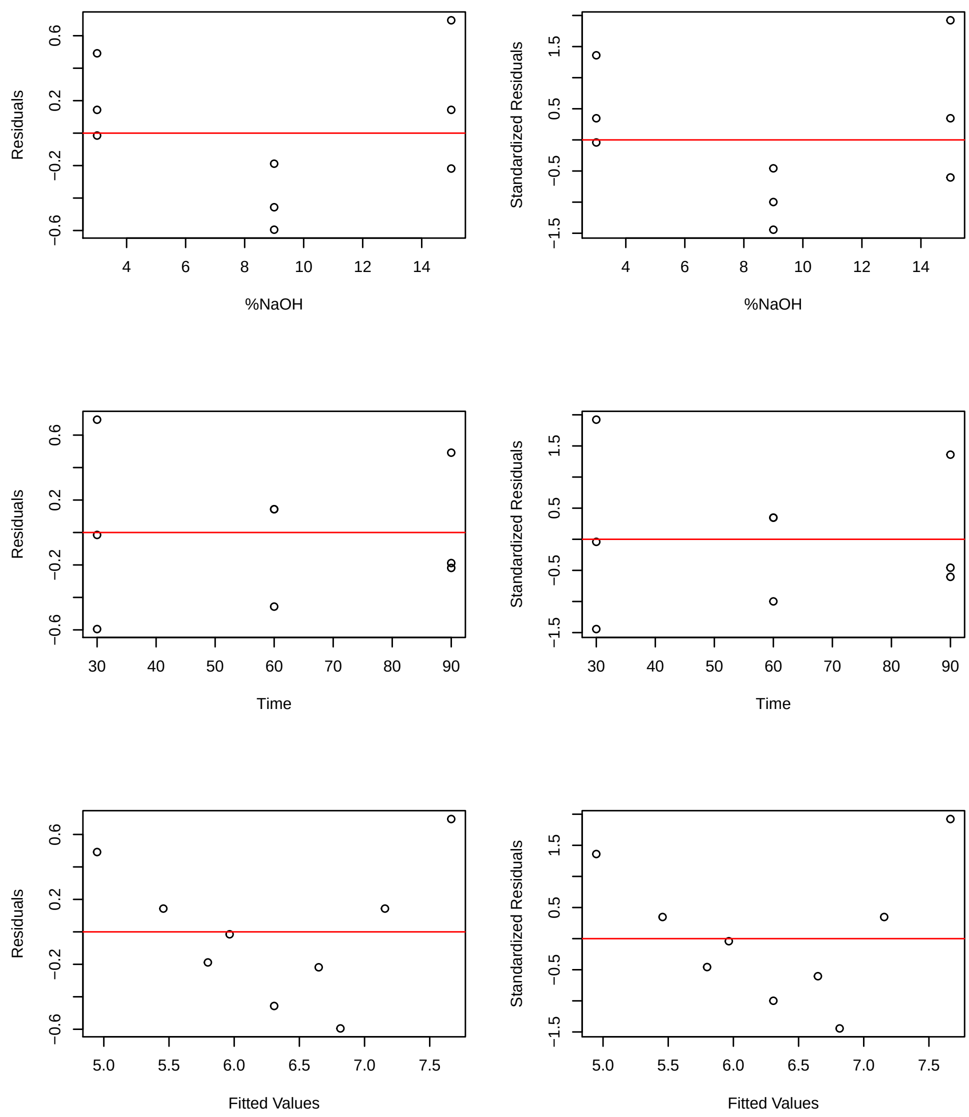
Homework 12

Due April 23, 2020 at 11:59 PM

1. P. 697: 2 (ignore (e), (g)) (5 points for (b), 4 points for others), dataset: `pulp.jmp`

   (a) From the JMP output, $s_{SF} = 0.4851$. Assuming that the model is appropriate, this measures the variation in Specific Surface Area for a fixed NaOH/Time condition.

   (b)



For each of teh three types of plots, the residuals and standardized residuals look almost exactly the same.

(c) The quantile used is $t_{6,0.95} = 1.943$.
The interval for β_0 is

$$b_0 \pm t_{6,0.95} \cdot se(b_0) = 6.0483 \pm 1.943(0.5208) = 6.043 \pm 1.011914 = (5.04, 7.06)$$

The interval for β_1 is

$$b_1 \pm t_{6,0.95} \cdot se(b_1) = 0.14167 \pm 1.943(0.03301) = 0.14167 \pm 0.6413843 = (0.078, 0.206)$$

The interval for β_2 is

$$b_2 \pm t_{6,0.95} \cdot se(b_2) = -0.016944 \pm 1.943(0.006601) = -0.016944 \pm 0.01282574 = (-0.0298, -0.0041)$$

(d) The quantile used is $t_{6,0.95} = 1.943$.
For $x_1 = 9, x_2 = 60$, from the JMP saved column we have $\hat{\mu}_{y|\boldsymbol{x}} = 6.30667$, $se(\hat{\mu}_{y|\boldsymbol{x}}) = 0.162$. So the interval is

$$\hat{\mu}_{y|\boldsymbol{x}} \pm t_{6,0.95} se(\hat{\mu}_{y|\boldsymbol{x}}) = 6.30667 \pm 1.943(0.162) = 6.30667 \pm 0.314766 = (5.99, 6.62)$$

For $x_1 = 10, x_2 = 70$, from the JMP saved column we have $\hat{\mu}_{y|\boldsymbol{x}} = 6.279$, $se(\hat{\mu}_{y|\boldsymbol{x}}) = 0.178$. So the interval is

$$\hat{\mu}_{y|\boldsymbol{x}} \pm t_{6,0.95} se(\hat{\mu}_{y|\boldsymbol{x}}) = 6.279 \pm 1.943(0.178) = 6.279 \pm 0.345854 = (5.93, 6.62)$$

(f) The quantile for the one-sided 90% prediction interval is $t_{6,0.9} = 1.440$. From the JMP output, the lower prediction bound for $x_1 = 9, x_2 = 60$

$$\begin{aligned}
&\hat{\mu}_{y|\boldsymbol{x}} - t_{6,0.9}\sqrt{s_{SF}^2 + (se(\hat{\mu}_{y|\boldsymbol{x}}))^2} \\
=&6.30667 - 1.440\sqrt{(0.4851)^2 + (0.162)^2} \\
=&6.30667 - 0.7364668 \\
=&5.57
\end{aligned}$$

For $x_1 = 10, x_2 = 70$

$$\begin{aligned}
&\hat{\mu}_{y|\boldsymbol{x}} - t_{6,0.9}\sqrt{s_{SF}^2 + (se(\hat{\mu}_{y|\boldsymbol{x}}))^2} \\
=&6.279 - 1.440\sqrt{(0.4851)^2 + (0.178)^2} \\
=&6.279 - 0.7440858 \\
=&5.53
\end{aligned}$$

(h) The ANOVA table is shown below.

Source	df	SS	MS	F
Regression	2	5.8854	2.9427	12.51
Error	6	1.4118	0.2353	
Total	8	7.2972		

The p-value is

$$P(F_{2,6} > 12.51) = 0.007242$$

The p-value is very small, so this is very strong evidecne that the model used is an improvement over a model which does not depend at all on NaOH and Time ($y = \beta_0 + \epsilon$).

2. P. 724: 7 (ignore (f)) (5 points for each question) , dataset: `grain_thresher.jmp`

(a) From the JMP output, $s_{SF} = 0.3471$. Assuming that the model is appropriate, this measures the variation in Weights for a fixed Spacing. Using equation (7.7) in the textbook for the pooled sample variance, $s_P = 0.3448$. These two estimates are very close, giving no indication that the model is inappropriate.

(b) 1. $H_0 : \beta_1 = \beta_2 = 0$, H_a : not all of β_1, β_2 are 0

2. The test statistic is

$$F = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{SSR/2}{SSE/77}$$

Assuming H_0 is true and the model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ is correct, the test statistic $F \sim F_{2,77}$.

3. From the JMP output, the observed F is 18.64 and the p-value is

$$P(F_{2,77} > 18.64) < 0.0001.$$

4. The p-value is very small, so we reject H_0 and conclude H_a.

5. There is overwhelming evidence that the model used is an improvement over a model in which Weight does not depend on Spacing ($y = \beta_0 + \epsilon$).

(c) 1. $H_0 : \beta_2 = 0$, $\beta_2 \neq 0$

2. The test statistic is

$$T = \frac{b_2 - 0}{se(b_2)}$$

Assuming H_0 is true and the model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ is correct, the test statistic $T \sim t_{77}$.

3. The observed test statistic $t = -6.10$. The p-value is

$$P(|T| > |-6.10|) = 2P(T > 6.10) < 0.0001$$

4. The p-value is very small, so we reject H_0 and conclude H_a.

5. There is overwhelming evidence that the quadratic model is an improment over the straight-line model $y = \beta_0 + \beta_1 x + \epsilon$.

(d) For the one-sided 90% confidence interval, we use $t_{77,0.9} = 1.2926$. For $x = 1$, the 90% lower confidence bound is

$$\begin{aligned}
&\hat{\mu}_{y|\boldsymbol{x}} - t_{77,0.9} se(\hat{\mu}_{y|\boldsymbol{x}} \\
=&13.0944 - 1.2926(0.526) \\
=&13.0944 - 0.6799076 \\
=&13.03
\end{aligned}$$

(e) The quantile used is still $t_{77,0.9} = 1.2926$. The lower prediction bound at $x = 1$ is

$$\begin{aligned}
&\hat{\mu}_{y|\boldsymbol{x}} - t_{77,0.9}\sqrt{s_{SF}^2 + (se(\hat{\mu}_{y|\boldsymbol{x}}))^2} \\
=&13.0944 - 1.2926\sqrt{(0.3471)^2 + (0.0526)^2} \\
=&13.0944 - 0.4537839 \\
=&12.64
\end{aligned}$$