Alexander Caines
MATH-310
3-29-2021

EXAM 2

1. The following are the results of running an ANOVA on the three factors (temperature, fabric denier, and air pressure) at each of their respective three levels. *I was not able to factor the data for Fabric, so I will provide the results for temperature and pressure.* Note that as temperature increases, the affect on the response becomes more significant as the p-value decreases. For pressure, the same phenomenon occurs. As the pressure increases, the p-value decreases, and thus the significance of the affect increases.

```
          Df Sum Sq Mean Sq F value  Pr(>F)
Pressure   1 170118  170118  16.843 0.00107 **
Fabric     2  26012   13006   1.288 0.30665
Residuals 14 141400   10100
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1: Results ANOVA at 17.2 kPa

```
          Df Sum Sq Mean Sq F value   Pr(>F)
Pressure   1 135509  135509  245.05 2.89e-10 ***
Fabric     2  37337   18669   33.76 4.41e-06 ***
Residuals 14   7742     553
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Results ANOVA at 34.4 kPa

```
          Df Sum Sq Mean Sq F value   Pr(>F)
Pressure   1 137362  137362  400.61 1.06e-11 ***
Fabric     2  18707    9354   27.28 1.48e-05 ***
Residuals 14   4800     343
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: Results ANOVA at 103.4 kPa

```
                    Df Sum Sq Mean Sq F value Pr(>F)
Temperature..Degrees.  1  25936   25936   3.167 0.0968 .
Fabric                 2   8785    4393   0.536 0.5964
Residuals             14 114644    8189
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: Results of ANOVA for temperature at 8 Degrees

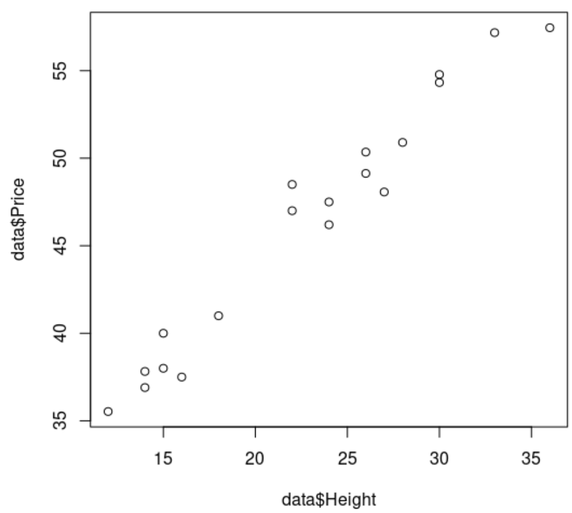
```
                     Df Sum Sq Mean Sq F value   Pr(>F)
Temperature..Degrees.  1   5993    5993   15.01  0.00168 **
Fabric                 2  27563   13782   34.52 3.87e-06 ***
Residuals             14   5590     399
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: Results ANOVA for temperature at 50 degrees

```
                     Df Sum Sq Mean Sq F value   Pr(>F)
Temperature..Degrees.  1  29960   29960   27.93 0.000115 ***
Fabric                 2  70940   35470   33.07 4.97e-06 ***
Residuals             14  15018    1073
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6: Results ANOVA for temperature at 75 degrees

2. (a) Given that the points in the dataset (without having plotted them) seem to increase linearly when ordered, it seems as if the prices are deterministically related to the height.
   (b) The scatter plot indicates a positive, linear relatonship between height and price.



   (c) The equation of least squares is $y = 0.98715x + 23.77215$.
   (d) At a height of 27, the corresponding price is \$50.43. Given the summary below, the corresponding residual is the result of the least squares (50.4252) plus the median residual (-0.8796), which is 49.55.
   (e) In referring to the above figure, one will see that the adjusted R-squared is 0.96. So, 96% of the variation in sales price can be attributed to the approximate linear relationship between truss height and price.

3. (a) Below is the result of the following line of code, given the appropriate datset, *data*:

$$model \leftarrow lm(NO3\ CO, data = data);\ summary(model)$$

```
Call:
lm(formula = Price ~ Height, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35522 -0.63584 -0.08796  0.92263  3.01053 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 23.77215    1.11347   21.35 1.03e-13 ***
Height       0.98715    0.04684   21.07 1.27e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.416 on 17 degrees of freedom
Multiple R-squared:  0.9631,	Adjusted R-squared:  0.961 
F-statistic: 444.1 on 1 and 17 DF,  p-value: 1.271e-13
```

Figure 7: Summary of linear model.

```
Call:
lm(formula = NO3 ~ CO, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5398 -0.7906  0.2801  0.9139  2.5317 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.220414   0.991656  -0.222     0.83    
CO           0.043620   0.003429  12.719  4.3e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.024 on 7 degrees of freedom
Multiple R-squared:  0.9585,	Adjusted R-squared:  0.9526 
F-statistic: 161.8 on 1 and 7 DF,  p-value: 4.298e-06
```

Figure 8: Model of NO3 and CO data.

(b) The equation produced by the linear regression in R was

$$\hat{y} = 13.63x + 21.97$$

So, the prediction for $x = 400$ is 5473.97. Given that the mean of CO is 211.89, the following work shows that the prediction interval is $(-139.5264, 563)$.

$$\hat{y} = mean(data\$CO);\ t = 0.627;\ error = 21.734;\ r^2 = deviance(model)$$

(c) Without the largest value of CO, the fitted equation is $\hat{y} = 15.21x + 21.70$, yielding 6105.7 as the prediction: a number far outside the previous interval. So, the largest CO value has a substantial affect on the model.

4. Below I show that $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$.

$$
\begin{aligned}
SSE &= \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i \\
&= \sum y_i^2 - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum y_i - \hat{\beta}_1 \sum x_i y_i \\
&= \sum y_i^2 - \bar{y} \sum y_i + \hat{\beta}_1 \bar{x} \sum y_i - \hat{\beta}_1 \sum x_i y_i \\
&= \sum y_i^2 - \frac{1}{n} \sum y_i \sum y_i + \hat{\beta}_1 (\sum \bar{x} y_i - \sum x_i y_i) \\
&= \sum y_i^2 - \frac{1}{n} (\sum y_i)^2 - \hat{\beta}_1 ((\sum x_i y_i) - \sum \bar{x} y_i) \\
&= S_{yy} - \hat{\beta}_1 ((\sum x_i y_i) - \sum \bar{x} y_i) \\
&= S_{yy} - \hat{\beta}_1 S_{xy}
\end{aligned}
$$

5. Although I was not able to format the data to run the ANOVAs properly, I would have gone about this problem by running an ANOVA to see if there is any significant affect on setosa by any of the listed factors. If the ANOVA indicated that one of them did by, returning a p-value less than 0.05, then I would have run four pairwise t-tests to determine which one had the most significant affect. After determining this, I would have tested my assumptions by checking whether or not the dataset I was using satisifed homogeneity of variances. in order to deter