quiz_ohanian3

Jackson Ohanian

4/29/2020

MLR

QUIZ

QUESTION 1

I will use the data on airplanes that I found for the group project but was not used. Using this data, use MLR to find an appropriate model for total departures.

```
ap <- read.csv("airport.csv")
```

1.a Create a nested MLR of this data out of two SLRs, predicting departures. Construct a confidence interval for each non-nested (SLR) model. Use a signifigance level of 0.95. Interpret the difference between the two intervals.

```
model1 <- lm(departures~Passengers, data = ap)
model2 <- lm(departures~Freight, data = ap)
model3 <- lm(departures~Passengers + Freight, data = ap)
```

```
confint(model1, newdata = ap, interval = "prediction", level = 0.95)
```

```
##                    2.5 %       97.5 %
## (Intercept) 5.601827e+03 9.814085e+03
## Passengers  1.196259e-02 1.272227e-02
```

```
confint(model2, newdata = ap, interval = "prediction", level = 0.95)
```

```
##                    2.5 %       97.5 %
## (Intercept) 2.369427e+04 4.086551e+04
## Freight     3.229537e-01 5.197077e-01
```

While the interval for departures on passengers (model 1) is relatively tight, the interval fro freight (model 2) is rather loose. This could imply that freight is not as reliant of a predictor as Passengers is.

1.b Compare the r-squared values of these models

```
summary(model1)$r.squared
```

```
## [1] 0.9688073
```

```
summary(model2)$r.squared
```

```
## [1] 0.3504656
```

```
summary(model3)$r.squared
```

```
## [1] 0.9694209
```

While the MLR (model 3) and passengers model (model 1) have relatively similar R squared value, the freight model (model 2) has a rather low R squared value, further implying freight is not a good predictor, as it does not account for a large amount of the variance in departures.

1.c Compare these models using anova tests

```
null_model = lm(departures ~ 1, data = ap)
anova(null_model, model1)
```

```
## Analysis of Variance Table
##
## Model 1: departures ~ 1
## Model 2: departures ~ Passengers
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    134 4.4344e+11
## 2    133 1.3832e+10  1 4.2961e+11 4130.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(null_model, model2)
```

```
## Analysis of Variance Table
##
## Model 1: departures ~ 1
## Model 2: departures ~ Freight
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    134 4.4344e+11
## 2    133 2.8803e+11  1 1.5541e+11 71.762 3.973e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(null_model, model3)
```

```
## Analysis of Variance Table
##
## Model 1: departures ~ 1
## Model 2: departures ~ Passengers + Freight
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    134 4.4344e+11
## 2    132 1.3560e+10  2 4.2988e+11 2092.3 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model1, model3)
```

```
## Analysis of Variance Table
##
## Model 1: departures ~ Passengers
## Model 2: departures ~ Passengers + Freight
##   Res.Df        RSS Df Sum of Sq      F Pr(>F)
## 1    133 1.3832e+10
## 2    132 1.3560e+10  1 272066644 2.6484  0.106
```

```
anova(model2, model3)
```

```
## Analysis of Variance Table
##
## Model 1: departures ~ Freight
## Model 2: departures ~ Passengers + Freight
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    133 2.8803e+11
```

```
## 2    132 1.3560e+10  1 2.7447e+11 2671.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.d Which model appears to be the best fit? Why?

Comparison of all three models with a null model shows that all three are signifigant regressions. Comparison of all three models as well as the r squared values of the model, shows that model 2 does not do a great job of explaining the variance in departures.

ANOVA tests realize that, at a signifigance level of 0.05, model 3 is not a signifigant regression compared to the nested model model 1. Therefore, the null hypothesis is accepted, that modeling departures by passengers is no different than modeling by passengers + freight.

With this in mind, I would choose model 1, passengers as the only predictor, as the best model of the 3 I selected.