Exam 2 - Question 2

Adam McQuistan

Tuesday, April 05, 2016

Problem 2 - Do problem 6.16 on page 251.

- Do not do part (b)

Part A. Test whether there is a regression relation, use alpha = 0.10. State alternative and decision rules. What does the test imply about the predictors and their usefullness?

```
df <- read.csv("data/6.15-6.16.csv")
names(df) = c("Satisfaction","Age", "Severity", "Anxiety")
result <- lm(Satisfaction ~ Age + Severity + Anxiety, data=df)
result_smry <- summary(result)
F_crit <- round(qf(0.90, df1=3, df2=42),1)
F_stat <- round(as.numeric(result_smry$fstatistic["value"]),1)
```

Analyze and test for the relatedness of Satisfaction to Age, Severity, and Anxiety.

F critical F(0.90, 3, 42) is 2.2.

H_0 : $\beta_1 = 0$ and $\beta_2 = 0$ for F* <= F critical

H_a : not both β_1 and β_2 equal zero for F* > F critical

Since F* (30.1) > F critical (2.2) we conclude that Satisfaction are related to Age, Severity, and Anxiety, the usefullness of the model for making predictions have yet to be determined.

Part C. Calculation coefficient of multople determination.

```
summary(result)
```

```
##
## Call:
## lm(formula = Satisfaction ~ Age + Severity + Anxiety, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## Age          -1.1416     0.2148  -5.315 3.81e-06 ***
## Severity     -0.4420     0.4920  -0.898   0.3741
```

```
## Anxiety     -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

The Multiple R^2 value is 0.6822 which translates into the model explaining about 68% of the variation in satisfaction. The Adjusted R^2 value of 0.6595 shows the correction for more than on predictor variable overfitting which is a very small change from the standard multiple r-squared value indicating little effects of overfitting.

To analyze and test the individual parameters of the regression model we use the students t-test such that:

$H_0 : \beta_p = 0$ indicating no relationship

$H_a : \beta_p \neq 0$ indicating there is a relationship

Since the p-values for both parameters (Disposable income and Target population) are less than 0.05 we conclude each individual parameter is significant and related to sales.