Supplement for Lecture 14: Assessing a Multiple Regression Model

Load and Clean Data

Variables of Interest in `fatal` - *adj_fatal* = Number of Vehicle Fatalities Per 1,000 People - *youngdrivers* = Percent of Drivers 15 - 24 - *year* = Year - *unemp* = Unemployment Rate - *beertax* = Tax on Case of Beer - *miles* = Average Miles Per Driver

```
data("Fatalities") # Load Data

fatal = Fatalities[,c("fatal","pop","youngdrivers","year",
                      "unemp","beertax","miles")]
fatal$adj_fatal = (fatal$fatal/fatal$pop)*1000
fatal$youngdrivers=fatal$youngdrivers*100
fatal$year = as.numeric(fatal$year) #Need to Convert to Numeric Variable (Currently a Factor Variable)

fatal$fatal=NULL
fatal$pop=NULL

#Rearrange Variables (Put "Y" variable at Beginning)
fatal=fatal[,c(6,1:5)]

#Preview Data
head(fatal)
```

```
##   adj_fatal youngdrivers year unemp  beertax    miles
## 1  0.212836      21.1572    1  14.4 1.539379 7233.887
## 2  0.234848      21.0768    2  13.7 1.788991 7836.348
## 3  0.233643      21.1484    3  11.1 1.714286 8262.990
## 4  0.219348      21.1140    4   8.9 1.652542 8726.917
## 5  0.266914      21.3400    5   9.8 1.609907 8952.854
## 6  0.271859      21.5527    6   7.8 1.560000 9166.302
```

Fit Linear Regression Model

```
#Fit Linear Regression Model
mod = lm(adj_fatal~youngdrivers + year + unemp + beertax + miles,data=fatal)

#Summary from Model
summary(mod)

#Pulling Out R-squared Adjusted R-squared
sum.out = summary(mod)

sum.out$r.squared
sum.out$adj.r.squared

#Calculate R-squared by hand
```

```
cor(x=fatal$adj_fatal,y=fitted(mod))^2
```

Interpretation of t-Tests: Only predictor variable where we don't have enough evidence to conclude that it's coefficient/slope is significantly different from 0 is *year*.

Interpretation of Slope for *miles*: Holding all other predictor variables (*youngdrivers, year, unemp*, and *beertax*) constant, the average number of vehicle fatalities per 1000 people in a state will increase by 0.0000015960 for every 1 unit increase in the average miles per driver in that state.

Alternative Interpretation of Slope for *miles*: Holding all other predictor variables (*youngdrivers, year, unemp*, and *beertax*) constant, the average number of vehicle fatalities per 1000 people in a state will increase by 0.0015960 if the average miles per driver in that state increased by 1,000.

ANOVA Table

```
#Run anova() function and notice how it is broken down by predictor variable. This is called sequential
#P-value in F-Test is Actually Testing if the Previous Model is significantly different than the Previo

anova(mod)

#Modified ANOVA function
#anova455
anova455(mod)

#Another Option for Getting F-statistic
mod.none=lm(adj_fatal~1,data=fatal) #Intercept Only
mod.full=lm(adj_fatal~., data=fatal) #Full Model Includes Intercept

#Table has all the same information but is presented differently.
anova(mod.none,mod.full)
```

#Confidence Intervals and Prediction Intervals

```
youngdrivers = 24
year  = 8 #Original Years 1982-1988 <=> New Years 1-7
unemp = 8.5
beertax = 1.8
miles = 8000

predict(mod,newdata=data.frame(youngdrivers,year,unemp,beertax,miles),interval="confidence")

predict(mod,newdata=data.frame(youngdrivers,year,unemp,beertax,miles),interval="prediction")
```