Exam 2 - Question 4

Adam McQuistan

Tuesday, April 05, 2016

Problem 4 - Do problem 7.7 on page 289 - Refers to the commercial realestate data from above

Part A. Obtain an analysis of variance table for decomposing the regression sum of squares into extra sums of squares associated with the following scenarios:

- X4
- X1 given X4
- X2 given X4
- X2 given X1 and X4
- X3 given X1 and X2 and X4

For this I have created a full (detailed) regression anova table

```
fullRegressionAnova <- function(lm_anova){
  VariationSource <- c("Regression", rownames(lm_anova), "Total")
  SSR <- sum(lm_anova$"Sum Sq"[1:(length(lm_anova$"Sum Sq")-1)])
  SST <- sum(lm_anova$"Sum Sq")
  DFReg <- sum(lm_anova$"Df"[1:(length(lm_anova$"Df")-1)])
  MSE <- lm_anova$"Mean Sq"[length(lm_anova$"Mean Sq")]
  MSR <- SSR / DFReg

  SS <- c(SSR,lm_anova$"Sum Sq",SST)
  MS <- c(MSR, lm_anova$"Mean Sq", NA)
  DF <- c(DFReg, lm_anova$"Df", sum(lm_anova$"Df"))
  F_stat <- MSR / MSE
  F_stats <-c(F_stat, lm_anova$"F value",NA)
  df_out <- data.frame(VariationSource, DF,SS, MS, F_stats)
  print(df_out)
  return(df_out)
}
```

Load Data Set and Recode the Variables

```
setwd("C:\\Users\\AdamMcQuistan\\Documents\\ISQA 8340\\Exam 2")
df <- read.csv("data/6.18.csv")
names(df)
```

```
## [1] "Case"    "Rental"  "Age"     "Expense" "Vacancy" "Footage"
```

```
names(df)[2:6] = c("Y", "X1", "X2", "X3", "X4")
names(df)
```

```
## [1] "Case" "Y"    "X1"   "X2"   "X3"   "X4"
```

X4

```
resultX4 <- lm(Y ~ X4, data=df)
resultX4_aov <- fullRegressionAnova(anova(resultX4))
```

```
##   VariationSource DF       SS        MS F_stats
## 1      Regression  1  67.7751 67.775098 31.7227
## 2              X4  1  67.7751 67.775098 31.7227
## 3       Residuals 79 168.7824  2.136486      NA
## 4           Total 80 236.5575        NA      NA
```

X1 given X4

```
resultX1_X4 <- lm(Y ~ X4 + X1, data=df)
resultX1_X4_aov <- fullRegressionAnova(anova(resultX1_X4))
```

```
##   VariationSource DF        SS        MS  F_stats
## 1      Regression  2 110.04967 55.024833 33.92625
## 2              X4  1  67.77510 67.775098 41.78759
## 3              X1  1  42.27457 42.274568 26.06492
## 4       Residuals 78 126.50783  1.621895       NA
## 5           Total 80 236.55750        NA       NA
```

X2 given X4

```
resultX2_X4 <- lm(Y ~ X4 + X2, data=df)
resultX2_X4_aov <- fullRegressionAnova(anova(resultX2_X4))
```

```
##   VariationSource DF         SS        MS  F_stats
## 1      Regression  2  77.066085 38.533043 18.84476
## 2              X4  1  67.775098 67.775098 33.14572
## 3              X2  1   9.290987  9.290987  4.54380
## 4       Residuals 78 159.491415  2.044762       NA
## 5           Total 80 236.557500        NA       NA
```

X2 given X1 and X4

```
resultX2_X1nX4 <- lm(Y ~ X1 + X4 + X2, data=df)
resultX2_X1nX4_aov <- fullRegressionAnova(anova(resultX2_X1nX4))
```

```
##   VariationSource DF        SS        MS  F_stats
## 1      Regression  3 137.90716 45.969053 35.88043
## 2              X1  1  14.81852 14.818520 11.56637
## 3              X4  1  95.23115 95.231147 74.33120
## 4              X2  1  27.85749 27.857493 21.74374
## 5       Residuals 77  98.65034  1.281173       NA
## 6           Total 80 236.55750        NA       NA
```

X3 given X1, X2, and X4

```
resultX3_X1nX2nX4 <- lm(Y ~ X1 + X2 + X4 + X3, data=df)
resultX3_X1nX2nX4_aov <- fullRegressionAnova(anova(resultX3_X1nX2nX4))
```

```
##   VariationSource DF          SS         MS    F_stats
## 1      Regression  4 138.3269061 34.5817265 26.7555260
## 2              X1  1  14.8185198 14.8185198 11.4649363
## 3              X2  1  72.8020109 72.8020109 56.3261669
## 4              X4  1  50.2866291 50.2866291 38.9062476
## 5              X3  1   0.4197463  0.4197463  0.3247534
## 6       Residuals 76  98.2305939  1.2925078         NA
## 7           Total 80 236.5575000         NA         NA
```

Part B. Test whether X3 can be dropped from the model give X1, X2, X3. Use F statistic and state p-value for leve $\alpha = 0.01$

To test if a singel β_k can be dropped from a multiple regression model, use the following formal test

$H_0 : \beta_k = 0$

$H_a : \beta_k \neq 0$

Use the test statistics

partial F test Note the partial F test is different in that it only tests whether one $\beta_k = 0$ not whether all $\beta_k = 0$ - if F* $\leq F$ critical conclude H_0

Test if X3 should be removed from the model at an $\alpha = 0.01$

```
F_stat <- resultX3_X1nX2nX4_aov$F_stats[5]
F_crit <- qf(0.99, resultX3_X1nX2nX4_aov$DF[5], resultX3_X1nX2nX4_aov$DF[6])
p_val <- 1-pf(q=F_stat, df1=resultX3_X1nX2nX4_aov$DF[5], df2=resultX3_X1nX2nX4_aov$DF[6])
conclusion <- ifelse(F_stat <= F_crit, "Conclude Null Hypthosis", "Reject Null Hypthosis")
cat("F* =", F_stat, ", F crit (0.99,",  resultX3_X1nX2nX4_aov$DF[5],
    ",",  resultX3_X1nX2nX4_aov$DF[6], ") =", F_crit, "\n", conclusion,", Pvalue = ", p_val, sep=" ")
```

```
## F* = 0.3247534 , F crit (0.99, 1 , 76 ) = 6.980578
##  Conclude Null Hypthosis , Pvalue =  0.5704457
```