Exam 2 - Ch 7 Question 15

Adam McQuistan

Monday, April 18, 2016

Using the commerical properties data set calculate the the following coefficients of partial determination

- R^2Y1
- R^2Y2
- $R^2Y1|4$
- R^2Y14
- $R^2Y2|1,4$
- $R^2Y3|1,2,4$
- R^2

```
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)
}

df <- read.csv("data/6.18.csv")

str(df)
```

```
## 'data.frame':    81 obs. of  6 variables:
##  $ Case   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Rental : num  13.5 12 10.5 15 14 10.5 14 16.5 17.5 16.5 ...
##  $ Age    : int  1 14 16 4 11 15 2 1 1 8 ...
##  $ Expense: num  5.02 8.19 3 10.7 8.97 ...
##  $ Vacancy: num  0.14 0.27 0 0.05 0.07 0.24 0.19 0.6 0 0.03 ...
##  $ Footage: int  123000 104079 39998 57112 60000 101385 31300 248172 215000 251015 ...
```

```
names(df) = c("Case", "Y", "X1", "X2", "X3", "X4")
```

$X_1 = Age$

$X_2 = Expense$

$X_3 = Vacancy$

$X_4 = Footage$

R^2Y1

```
full <- lm(Y ~ X1, data=df)
full_smry <- summary(full)
full_aov <- fullRegressionAnova(anova(full))
```

```
##   VariationSource DF        SS        MS  F_stats
## 1      Regression  1  14.81852 14.818520 5.279464
## 2              X1  1  14.81852 14.818520 5.279464
## 3       Residuals 79 221.73898  2.806823       NA
## 4           Total 80 236.55750        NA       NA
```

```
result_pct <- round(full_smry$r.squared * 100, 1)

msg <- paste("X1 explains ",
             result_pct, " percent of the variation in Y", sep="")
cat(msg)
```

```
## X1 explains 6.3 percent of the variation in Y
```

R^2Y2

```
full <- lm(Y ~ X2, data=df)
full_smry <- summary(full)
full_aov <- fullRegressionAnova(anova(full))
```

```
##   VariationSource DF        SS        MS  F_stats
## 1      Regression  1  40.50333 40.503331 16.32081
## 2              X2  1  40.50333 40.503331 16.32081
## 3       Residuals 79 196.05417  2.481698       NA
## 4           Total 80 236.55750        NA       NA
```

```
result_pct <- round(full_smry$r.squared * 100, 1)

msg <- paste("X2 explains ",
             result_pct, " percent of the variation in Y", sep="")
cat(msg)
```

```
## X2 explains 17.1 percent of the variation in Y
```

$R^2Y1|4$

```r
full <- lm(Y ~ X4 + X1, data=df)
reduced <- lm(Y ~ X4, data = df)

full_aov <- fullRegressionAnova(anova(full))
```

```
##   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
```

```r
reduced_aov <- fullRegressionAnova(anova(reduced))
```

```
##   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
```

```r
SSR_X1_given_X4 <- full_aov[3,3]
SSE_reduced <- reduced_aov[3,3]

result_pct <- round(SSR_X1_given_X4 / SSE_reduced * 100, 1)

msg <- paste("Adding in X1 to the model containing X4 reduces the error sum of squares by ",
             result_pct, " percent", sep="")
cat(msg)
```

```
## Adding in X1 to the model containing X4 reduces the error sum of squares by 25 percent
```

$R^2Y1, 4$

```r
full <- lm(Y ~ X1 + X4, data=df)
full_smry <- summary(full)

full_aov <- fullRegressionAnova(anova(full))
```

```
##   VariationSource DF        SS        MS   F_stats
## 1      Regression  2 110.04967 55.024833 33.926255
## 2              X1  1  14.81852 14.818520  9.136545
## 3              X4  1  95.23115 95.231147 58.715964
## 4       Residuals 78 126.50783  1.621895        NA
## 5           Total 80 236.55750        NA        NA
```

```r
result_pct <- round(full_smry$r.squared * 100, 1)

msg <- paste("The variables X1 and X4 explains ",
             result_pct, " percent of the variation in Y", sep="")
cat(msg)
```

```
## The variables X1 and X4 explains 46.5 percent of the variation in Y
```

$R^2Y2|1,4$

```
full <- lm(Y ~ X4 + X1 + X2, data=df)
reduced <- lm(Y ~ X4 + X1, data = df)

full_aov <- fullRegressionAnova(anova(full))
```

```
##   VariationSource DF        SS        MS  F_stats
## 1      Regression  3 137.90716 45.969053 35.88043
## 2              X4  1  67.77510 67.775098 52.90081
## 3              X1  1  42.27457 42.274568 32.99676
## 4              X2  1  27.85749 27.857493 21.74374
## 5       Residuals 77  98.65034  1.281173       NA
## 6           Total 80 236.55750        NA       NA
```

```
reduced_aov <- fullRegressionAnova(anova(reduced))
```

```
##   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
```

```
SSR_X2_given_X1X4 <- full_aov[4,3]
SSE_reduced <- reduced_aov[4,3]

result_pct <- round(SSR_X2_given_X1X4 / SSE_reduced * 100, 1)

msg <- paste("Adding in X2 to the model containing X4 and X1 reduces the error sum of squares by \n",
             result_pct, " percent", sep="")
cat(msg)
```

```
## Adding in X2 to the model containing X4 and X1 reduces the error sum of squares by
## 22 percent
```

$R^2Y3|1,2,4$

```
full <- lm(Y ~ X1 + X2 + X4 + X3, data=df)
reduced <- lm(Y ~ X1 + X2 + X4, data = df)

full_aov <- fullRegressionAnova(anova(full))
```

```
##   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
```

```
reduced_aov <- fullRegressionAnova(anova(reduced))
```

```
##   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              X2  1  72.80201 72.802011 56.82449
## 4              X4  1  50.28663 50.286629 39.25045
## 5       Residuals 77  98.65034  1.281173       NA
## 6           Total 80 236.55750        NA       NA
```

```
SSR_X3_given_X1X2X4 <- full_aov[5,3]
SSE_reduced <- reduced_aov[5,3]

result_pct <- round(SSR_X3_given_X1X2X4 / SSE_reduced * 100, 1)

msg <- paste("Adding in X3 to the model containing X1, X2, and X4 reduces the error sum of squares by \n
             result_pct, " percent", sep="")
cat(msg)
```

```
## Adding in X3 to the model containing X1, X2, and X4 reduces the error sum of squares by
## 0.4 percent
```

R^2

```
full <- lm(Y ~ X1 + X2 + X3 + X4, data=df)
full_smry <- summary(full)

full_aov <- fullRegressionAnova(anova(full))
```

```
##   VariationSource DF         SS        MS   F_stats
## 1      Regression  4 138.326906 34.581727 26.755526
## 2              X1  1  14.818520 14.818520 11.464936
## 3              X2  1  72.802011 72.802011 56.326167
## 4              X3  1   8.381417  8.381417  6.484616
## 5              X4  1  42.324958 42.324958 32.746385
## 6       Residuals 76  98.230594  1.292508        NA
## 7           Total 80 236.557500        NA        NA
```

```
result_pct <- round(full_smry$r.squared * 100, 1)

msg <- paste("The variables X1, X2, X3, and X4 explains ",
             result_pct, " percent of the variation in Y", sep="")
cat(msg)
```

```
## The variables X1, X2, X3, and X4 explains 58.5 percent of the variation in Y
```