

Ch7 Body Fat Example

Adam McQuistan

Monday, April 04, 2016

Body Fat Data

Y = Body Fat X1 = Tricep Skinfold Thickness X2 = Thigh Circumference X3 = Midarm Circumference

```
setwd("C:\\Users\\AdamMcQuistan\\Documents\\ISQA 8340\\ch7")
df <- read.table(file="data/bodyfat.txt", header=T, as.is=T, sep="\t")
str(df)
```

```
## 'data.frame':    20 obs. of  4 variables:
## $ X1: num  19.5 24.7 30.7 29.8 19.1 25.6 31.4 27.9 22.1 25.5 ...
## $ X2: num  43.1 49.8 51.9 54.3 42.2 53.9 58.5 52.1 49.9 53.5 ...
## $ X3: num  29.1 28.2 37 31.1 30.9 23.7 27.6 30.6 23.2 24.8 ...
## $ Y : num  11.9 22.8 18.7 20.1 12.9 21.7 27.1 25.4 21.3 19.3 ...
```

Create function to produce a full anova table from an anova model built off a lm model

```
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)
}
```

Evaluate the Effects of Square Errors as β_k enter the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

```
result_lm <- lm(Y ~ X1 + X2 + X3, data=df)
result_aov <- fullRegressionAnova(anova(result_lm))
```

```
## VariationSource DF      SS      MS  F_stats
## 1      Regression  3 396.98461 132.328204 21.515712
```

```
## 2          X1  1 352.26980 352.269797 57.276796
## 3          X2  1  33.16891  33.168913  5.393051
## 4          X3  1  11.54590  11.545902  1.877289
## 5    Residuals 16  98.40489   6.150306        NA
## 6          Total 19 495.38950        NA        NA
```

Test whether a single $\beta_k = 0$

To test if a single 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

t-test - if $|t^*| \leq t(1 - \alpha; n - p)$ conclude H_0

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 = result_aov$F_stats[4]
F_crit = qf(0.99, result_aov$DF[4], result_aov$DF[5])
conclusion <- ifelse(F_stat <= F_crit, "Conclude Null Hypthosis", "Reject Null Hypthosis")
cat("F* =", F_stat, "F crit =", F_crit, "\n", conclusion, sep=" ")
```

```
## F* = 1.877289 F crit = 8.530965
## Conclude Null Hypthosis
```

Test whether multiple $\beta_K = 0$

Test whether both X2 and X3 should be removed from the model

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_a: \text{not both } \beta_2 \text{ and } \beta_3 \text{ equal zero}$$

Do a simple anova test of reduced against full model and assess partial F test to test whether at $\alpha = 0.05$ the models are different

```
full <- result_lm
reduced <- lm(Y ~ X1, data=df)
anova(reduced, full)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1
## Model 2: Y ~ X1 + X2 + X3
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      18 143.120
## 2      16  98.405  2    44.715 3.6352 0.04995 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You can see that the p-value is essentially equal to 0.05 so it probably would not be a good idea to remove both from the model.