Extra Sums of Squares

- **Definitions.**
  1. $SSTO = SSR + SSE$.
  2. $SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$
     $SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1)$.
  3. $SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$.
  4. $SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2)$
     $SSR(X_3|X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$.
  5. $SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$.

- **Body Fat Example.** The values stored in 'BodyFat.csv' file contains the data for a study of the relation of amount of body fat (Y) to several possible predictor variables, based on a sample of 20 healthy females 25-34 years old. The possible predictor variables are *triceps skinfold thickness* (X_1), *thigh circumference* (X_2), and *midarm circumference* (X_3). The amount of body fat for each of the 20 persons was obtained by a cumbersome and expensive procedure requiring the immersion of the person in water. It would therefore be very helpful if a regression model with some or all of these predictor variables could provide reliable estimates of the amount of body fat since the measurements needed for the predictor variables are easy to obtain.

```
data.body=read.csv("BodyFat.csv",header=T)
attach(data.body)
y=fat
x1=triceps
x2=thigh
x3=midarm

results.12=lm(y~x1+x2)
summary(results.12)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.1742     8.3606  -2.293   0.0348 *
x1            0.2224     0.3034   0.733   0.4737
x2            0.6594     0.2912   2.265   0.0369 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

anova(results.12)
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 352.27  352.27 54.4661 1.075e-06 ***
x2         1  33.17   33.17  5.1284    0.0369 *
Residuals 17 109.95    6.47

results.1=lm(y~x1)
anova(results.1)
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 352.27  352.27  44.305 3.024e-06 ***
Residuals 18 143.12    7.95

# ssr(x2|x1)=ssr(x1,x2)-ssr(x1)=sse(x1)-sse(x1,x2)  # 33.17
# f_obs=(ssr(x2|x1)/1)/(sse(x1,x2)/(n-3))   # Ho: beta_2=0
```

```
results.123=lm(y~x1+x2+x3)
anova(results.123)
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 352.27  352.27 57.2768 1.131e-06 ***
x2         1  33.17   33.17  5.3931   0.03373 *
x3         1  11.55   11.55  1.8773   0.18956
Residuals 16  98.40    6.15

# ssr(x3|x1,x2)=ssr(x1,x2,x3)-ssr(x1,x2)=sse(x1,x2)-sse(x1,x2,x3)   # 11.55
# f_obs=(ssr(x3|x1,x2)/1)/(sse(x1,x2,x3)/(n-4)) # Ho: beta_3=0









# ssr(x2,x3|x1)=sse(x1)-sse(x1,x2,x3)=ssr(x1,x2,x3)-ssr(x1)   # 33.17+11.55
# f_obs=(ssr(x2,x3|x1)/2)/(sse(x1,x2,x3)/(n-4)) # Ho: beta_2=0 and beta_3=0









##Or you can use these commands:
red=lm(y~x1)
full=lm(y~x1+x2+x3)
anova(red,full) # Partial F-Test for Ho: beta_2=0 and beta_3=0

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
```