

Announcement**The Extra Sum of Squares Principle**

Gerard E. Dallal, Ph.D.

The *Extra Sum of Squares Principle* allows us to compare two models for the same response where one model (the full model) contains all of the predictors in the other model (the reduced model) and more. For example, the reduced model might contain m predictors while the full model contains p predictors, where p is greater than m and all of the m predictors in the reduced model are among the p predictors of the full model, that is,

$$\begin{aligned}\text{full: } Y &= \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \cdots + \beta_p X_p + \varepsilon \\ \text{reduced: } Y &= \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \varepsilon\end{aligned}$$

The extra sum of squares principle allows us to determine whether there is statistically significant predictive capability in the set of additional variables. The specific hypothesis it tests is

$$H_0: \beta_{m+1} = \cdots = \beta_p = 0$$

The method works by looking at the reduction in the Residual Sum of Squares (or, equivalently, at the increase in Regression Sum of Squares) when the set of additional variables is added to the model. This change is divided by the number of degrees of freedom for the additional variables to produce a mean square. This mean square is compared to the Residual mean square from the full model. Most full featured software packages will handle the arithmetic for you. All the analyst need do is specify the two models.

Example: An investigator wanted to know, in this set of cross-sectional data, whether muscle strength was predictive of bone density after adjusting for age and measures of body composition. She had eight strength measures and no prior hypothesis about which, if any, might be more useful than the others. In such situations, it is common practice to ask whether there is any predictive capability in the set of strength measures.

Two models will be fitted, one containing all of the predictors and the other containing everything but the strength measures. The extra sum of squares principle can then be used to assess whether there is any predictive capability in the set of strength measures.

** ** * Full Model ** ** *

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	0.33038	0.02541	4.86	0.0003
Error	26	0.13582	0.00522		
Corrected Total	39	0.46620			

** ** * Reduced Model ** ** *

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.18929	0.03786	4.65	0.0024
Error	34	0.27691	0.00814		
Corrected Total	39	0.46620			

** ** * Extra Sum of Squares ** ** *

Source	DF	Mean Square	F Value	Pr > F
Numerator	8	0.01764	3.38	0.0087
Denominator	26	0.00522		

Adding the strength measures to the model increases the Regression Sum of Squares by 0.14109 ($=0.33038-0.18929$). Since there are eight strength measures, the degrees of freedom for the extra sum of squares is 8 and the mean square is 0.01764 ($=0.14109/8$). The ratio of this means square to the Error mean square from the full model is 3.38. When compared to the percentiles of the F distribution with 8 numerator degrees of freedom and 26 denominator degrees of freedom, the ratio of mean squares gives an observed significance level of 0.0087. From this we conclude that muscle strength is predictive of bone density after adjusting for various measures of body composition.

The next natural question is "which measures are predictive?" This is a difficult question, which we will put off for the moment. There are two issues. The first is the general question of how models might be simplified. This will be discussed in detail, but there is no satisfactory answer. The second is that there are too many predictors in this model--thirteen--to hope to be able to isolate individual effects with only 40 subjects.

Copyright © 1998 [Gerard E. Dallal](#)
 Last modified: 05/23/2012 02:51:55.