CHAPTER 11: COMPARING A SEQUENCE OF MODELS USING THE ANALYSIS OF VARIANCE

11.1 INTRODUCTION

A common problem faced by researchers is the comparison of models for a system. The question often arises when considering two or more models. Say you have two or more competing models and wish to determine which one is the best one for the system you are investigating; for example, for two competing models, call one the ***complete model*** and the other the ***reduced model***. For the Gebotys and Roberts (1989) example, consider model 1 the problem of predicting seriousness (y) from age (x_1) amount of TV news watched (x_2) and whether the person has been a victim of a crime (x_3). Model 1 or the *complete* model is given by

$$E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Suppose this model is a reasonable one, as determined by the F test in the ANOVA. The question that now arises is whether a *reduced* model can describe/predict seriousness (y) as well as the complete model. Researchers, for the sake of parsimony, prefer the model with the fewest variables. The reduced model must be a subset of the complete model or nested within the complete model. For example, consider model 2 predicting seriousness from age given by

$$E(y|x) = \beta_0 + \beta_1 x_1$$

Clearly, this is a subset of the complete model. Consider the model

$$E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$$

This model is not contained in the complete model, since the complete model does not include x_1^2. This is called a *non-nested* model, which can be compared using the techniques described in Gebotys (1987), but will not be discussed here.

The test of hypothesis comparing Model 1 and Model 2 is

$H_0 : \beta_2 = \beta_3 = 0$

(TV news and victimization not necessary)

$H_a : \beta_2 \neq \beta_3 \neq 0$

(TV news and victimization necessary)

If H_0 is rejected, then the complete model is necessary for modelling *y*. However, if we are unable to reject H_0, then we conclude the reduced model is as effective as the complete model in modelling *y*. From an intuitive point of view, the sums of squares for the model component for both models is compared in an *F* ratio. If they are equal, then the reduced model is seen to be equivalent to the complete model. The formal method of comparison is called the ***Analysis of Variance***. We will fully discuss this method in the following section.

11.2 THE ANALYSIS OF VARIANCE

The analysis of variance is a testing procedure for the normal linear model, and in particular, hypotheses concerning the parameter vector β. The technique is more general than the method discussed in previous chapters.

Suppose we know that $E(y|x) = X_1\beta_1$. In other words, this model (call this the complete model) is adequate.

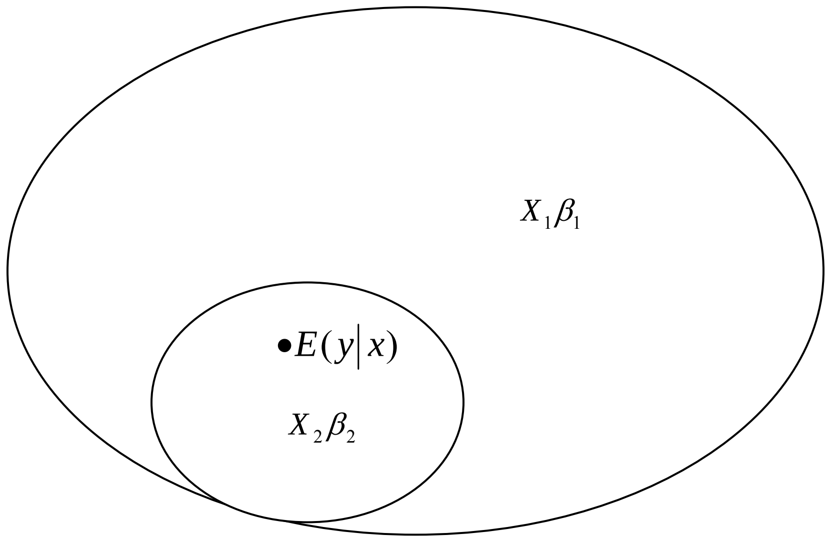
Consider the hypotheses:

$H_0 : E[y] = X_2\beta_2$ given that $E[y] \in X_1\beta_1$
and $X_2\beta_2 \in X_1\beta_1$ where X_i is of dimension $n \times k_i$ (read $\in$ as contained in).

where $X_2\beta_2$ is the reduced model and $X_1\beta_1$ is the complete model. The above hypothesis in words is given below. The researcher has evidence that the complete model (X_1) is adequate; however, s/he would like to determine whether the reduced model (X_2)

can predict *y* as well as the complete. Note the reduced model must be contained or *nested* in the complete model.

Graphically below we have $X_2\beta_2 \in$ (contained or nested) in $X_1\beta_1$. $E(y|x)$ may be in either section. If it is contained in $X_2\beta_2$ then the reduced model is just as adequate as the complete. This is seen in the figure below.



Continuing our example: Y = crime seriousness

x_2 = age

x_3 = TV news

x_4 = victimization

Model 1: $E[y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

$$X_1\beta_1 = \begin{bmatrix} 1 & 20 & 4 & 1 \\ 1 & 25 & 5 & 1 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & 35 & 5.5 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

where the number of parameters is equal to 4.

Model 1 is the complete model. The complete model is adequate as measured by the *F* test in the ANOVA.

Model 2: $E[y] = \beta_0 + \beta_1 x_1$

$$= \begin{bmatrix} 1 & 20 \\ 1 & 25 \\ . & . \\ . & . \\ . & . \\ 1 & 35 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$= X_2 \beta_2$$

where the number of parameters is equal to 2.

Model 2 is the reduced model. Note it is nested or contained in the complete model.

$$\boxed{\boldsymbol{E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}}$$

We now consider how to assess the evidence for or against H_0.

If H_0 were true, we would use the reduced model to estimate

$E[y]$ by $b_2 = (X_2' X_2)^{-1} X_2' y$

Model 1, or the complete model, would give us the following estimate of *E(y)*

$$b_1 = (X_1' X_1)^{-1} X_1' y$$

A natural approach to assessing the evidence for or against H_0 is to compare $X_2 b_2$ with $X_1 b_1$ ($X_i b_i$ is just a measure of fit for the model, see section 8.5). If H_0 is true, then $X_1 b_1$ should be close to $X_2 b_2$ in the sense that the difference can be explained by random variation.

We measure the closeness of $X_1 b_1$ to $X_2 b_2$ by the squared length of the vector $X_1 b_1 - X_2 b_2$. The squared length or *sums of squares* of $X_1 b_1 - X_2 b_2$ is $y' X_1 b_1 - y' X_2 b_2$ (Model Sums of Squares

Complete – *MSS* Reduced) and the test of significance will be used to compare this distance to error to see whether we obtain an unrealistic value under H_0.

Under H_0 the sum of squares $\frac{(y'X_1b_1 - y'X_2b_2)}{\sigma^2} \sim \chi_{n-k_1}$

Therefore, under H_0 we have that

$$\boxed{F = \frac{(y'X_1b_1 - y'X_2b_2)/(k_1 - k_2)}{s^2}}$$

is distributed as a $F(k_1 - k_2, n - k_1)$ variable.

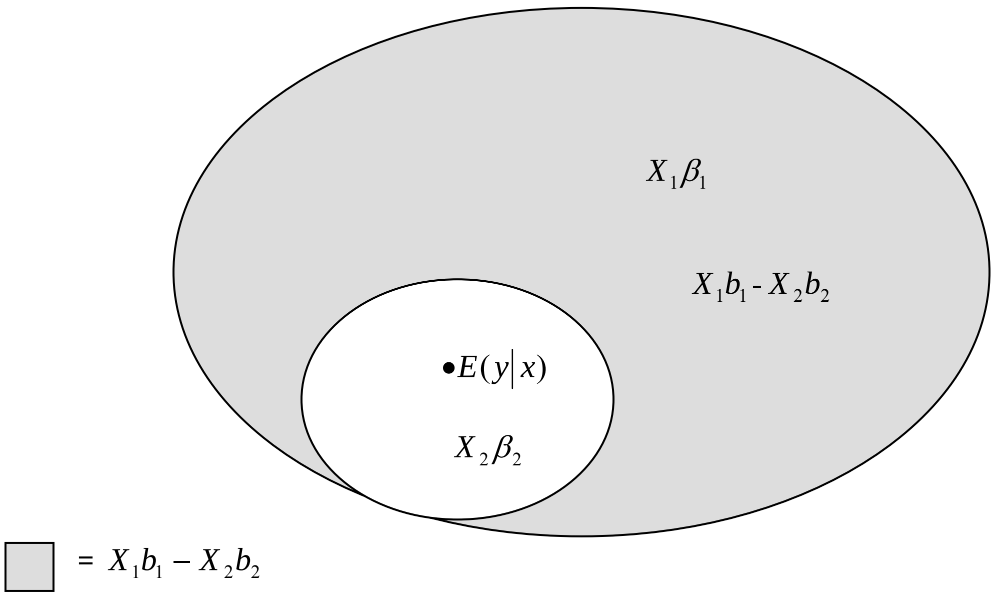
Thus, the test of significance is to compute the OLS or *p*-value and assess accordingly. We record this in an ANOVA table.

ANOVA

Source	Degrees of Freedom # parameters fit	Sums of Squares	Mean Square	F
Model x_2 (reduced)	k_2	$y'x_2b_2$	$y'x_2b_2 / k_2$	
Model x_1 – Model x_2 (difference part)	$k_1 - k_2$	$y'x_1b_1 - y'x_2b_2$	$(y'x_1b_1 - y'x_2b_2) / (k_1 - k_2)$	$/ s^2$
Residual (that left over after fitting Model X_1)	$n - k_1$	$y'y - y'x_1b_1$	s^2	
Total	n	$y'y$		

The ANOVA *F* test allows the researcher to test whether any subset of β's from the complete model is equal to zero.

Graphically, the difference part (Model 1 – Model 2) being tested is the shaded part below.



Example 1 (continued)

Gebotys and Roberts compared both models (complete and reduced) in terms of prediction of seriousness (y). The following variables are included, age (x_2), TV news (x_3), and victimization (x_4) scores. The data from 20 subjects are given below.

y serious	x_1 age	x_2 amount of TV news watched (hrs/wk)	x_3 previous victim of crime (1=yes, 0=no)
21	20	4.0	1
28	25	5.0	1
27	26	5.0	1
26	25	4.5	1
33	30	6.0	0
36	34	7.0	0
31	40	5.5	1
35	40	6.0	0
41	40	7.0	0
95	80	9.0	0
30	30	5.0	0
25	31	5.0	0

y serious	x_1 age	x_2 amount of TV news watched (hrs/wk)	x_3 previous victim of crime (1=yes, 0=no)
40	60	7.5	1
40	50	7.0	1
40	35	6.0	1
22	24	4.5	0
40	43	7.3	1
55	40	7.0	1
48	37	6.0	1
30	35	5.5	0

First, fit Model 1 the complete model.

$$\text{Model 1: } E[y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$
$$= X_1 \beta_1$$

The number of parameters is 4.

The ANOVA output of SPSS is listed below. Clearly, the complete model is adequate.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3764.448	3	1254.816	17.042	.000[a]
	Residual	1178.102	16	73.631		
	Total	4942.550	19			

a. Predictors: (Constant), Previous victim of crime, Age, Amount of TV news watched (hrs/wk)

b. Dependent Variable: Serious

where

$SS_1 = 3764.45$ (sum of squares of regression for Model 1)

$k_1 = 4$

$s_1^2 = 73.63$ (Mean Square Error (*MSE*) or Residual)

The researchers would like to know whether a model including only age will predict seriousness as well as Model 1 above. In order to answer this question, the next step is to fit Model 2 or the reduced model.

Model 2: $E[y] = \beta_0 + \beta_1 x_2$ is a model which only includes x_1

$$= X_2 \beta_2$$

The model has 2 parameters.

The ANOVA output below indicates the reduced model's adequacy.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3640.752	1	3640.752	50.341	.000[a]
	Residual	1301.798	18	72.322		
	Total	4942.550	19			

a. Predictors: (Constant), Age

b. Dependent Variable: Serious

where

$SS_2 = 3640.75$ (sum of square of regression for Model 2)

$k_2 = 2$

$s_2^2 = 72.32$ (MSE)

With the above information from the two models, we can answer the research question.

Question: do we need x_3 and x_4 in the model?

i.e. $H_0 : E[y] = X_2 \beta_2$ given that $E[y] \in X_1 \beta_1$

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

(x_3 and x_4 not necessary)

$H_a : \beta_2 \neq \beta_3 \neq 0$

(x_3 and x_4 necessary)

Therefore, $$F(2,16) = \frac{(SS_1 - SS_2)\big/(k_1 - k_2)}{s_1^2}$$

$$= \frac{3764.45 - 3640.75\big/2}{73.63}$$

$$= .84$$

The above result is usually displayed in an ANOVA table. Note that the table is corrected for the mean or intercept. It has been removed and is calculated using the ANOVA table from the computer output above.

Source	*DF*	*SS*	*MS*	*F*
Model 2 (intercept removed)	1	3640.75		
Difference (Model1 – Model2)	4-2 = 2	3764.45-3640.75 = 123.7	61.85	61.85 / 73.63 = .84
Residual (Model1)	16		73.63	
Total (corrected)	19			

Clearly the *F* statistic .84 with 2,16 degrees of freedom is not significant, $p > .05$. We do not have evidence against $H_0 : \beta_2 = \beta_3 = 0$. We conclude that the amount of TV news and victimization do not contribute significantly over and above the contribution of age. In other words, the reduced model (Model 2) is as adequate at predicting seriousness as the complete model.

11.3 COMPARING A SEQUENCE OF HYPOTHESES

Now suppose we have a sequence of hypotheses or models concerning $E[y]$; namely,

$$H_{01} : E[y] = X_2\beta_2 \text{ given that } E[y] \in X_1\beta_1$$

$$H_{02} : E[y] = X_3\beta_3 \text{ given that } E[y] \in X_2\beta_2$$

.

.

.

$$H_{0m} : E[y] = X_{m-1}\beta_{m-1} \text{ given that } E[y] \in X_m\beta_m$$

where $X_m \le X_{m-1} \le ... \le X_1$ is nested or contained within one another.

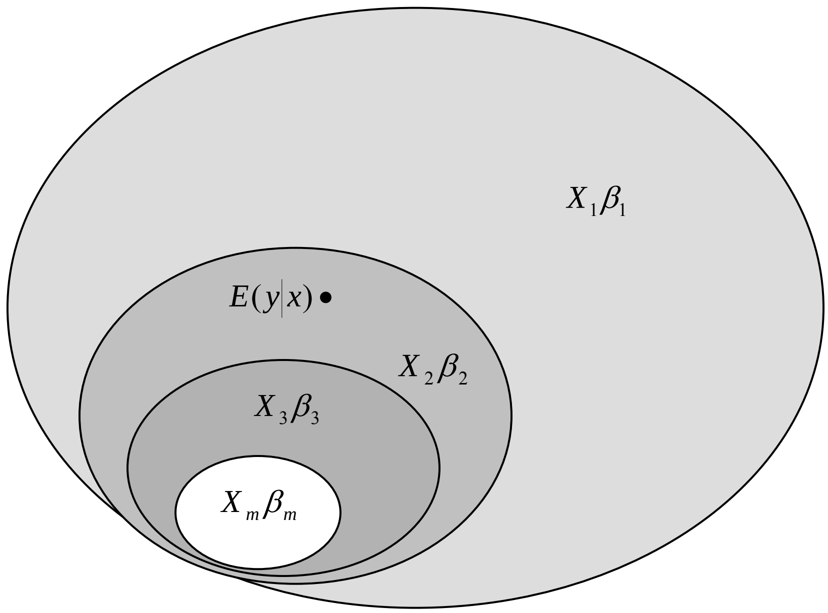
We first test the hypothesis H_{01}. If we obtain evidence against H_{01}, then naturally we test no further. In other words, the complete model does not adequately model the data. If there is no evidence against H_{01}, then we assume $E[y] \in x_2\beta_2$. We then test H_{02} in the obvious way. We obtain the following ANOVA table where k_i represents the number of parameters in Model x_i.

Note the table is not corrected for the mean.

Source	*DF*	*SS*	*MS*	*F*
Model X_m	k_m	$y'X_m b_m$	$y'X_m b_m / k_m$	
Difference X_{m-1} and X_m	$k_m - k_{m-1}$	$y'X_1 b_m - y'X_{m-1} b_{m-1}$		
.	.	.		
.	.	.		
.	.	.		
Difference Model X_1 and X_2	$k_1 - k_2$	$y'X_1 b_1 - y'X_2 b_2$		ms/s^2
Residual after fitting model X_1	$n - k_m$	$y'y - y'X_1 b_1$	s^2	
Total	n	$y'y$		

Note the error term for testing differences is always from the most complex model (complete model).

Graphically we have



$= X_1b_1 - X_2b_2$

$= X_2b_2 - X_3b_3$

$= X_3b_3 - X_mb_m$

This type of nested modeling is the basis for the ANOVA tables in the experimental design section that will be discussed in Part III of the text.

11.4 Exercises

1. Using the data from the Clinical Psychology problem in Section 9.7, perform the following analysis.

   Fit the model $E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

What is the estimated regression equation? Test at $\alpha = .05$ that $\beta_3 = \beta_4 = \beta_5 = 0$ in an ANOVA table. The complete model is

$$E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

2. Using the Tucher (1987) data from Section 9.7, compare the model given in part a)

$E(y|x) = \beta_0 + \beta_1 x_1 + ... + \beta_8 x_8$ with $E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Clearly state your conclusions.