Tests of General Linear Hypotheses

BIOS 6611

CU Anschutz

Week 12

Introduction to General Linear Hypotheses

Introduction

Previously, we discussed tests for a single parameter or group of parameters. Now, we will learn how to test any linear combination of parameters (i.e., increased flexibility).

A *general linear hypothesis* (GLH) is any hypothesis that tests a linear combination of regression parameters. For example:

$$\begin{aligned} &1. \quad H_0: \quad \beta_2 = 0 \\ &2. \quad H_0: \quad \beta_1 - \beta_2 = 0 \Rightarrow \beta_1 = \beta_2 \\ &3. \quad H_0: \quad \beta_1 = \beta_2 = 0 \end{aligned}$$

are all examples of hypotheses that fit into the general linear framework.

Introduction

These kinds of tests can put into the more general framework:

$$
\begin{aligned}
H_0: &\quad \mathbf{c}\beta = \mathbf{d} \\
H_1 &\quad : \mathbf{c}\beta \neq \mathbf{d}
\end{aligned}
$$

$\mathbf{c}$ is an $r \times p^*$ matrix that is of rank r and $r \leq p^*$, where

- r is the number of linear combinations of parameters we wish to test
- $p^* = p$ when an intercept is included in the model
- $p^* = p - 1$ for a no intercept model

$\mathbf{d}$ is an $r \times 1$ vector, often will be a vector of zeros (but not necessarily).

GLH Examples

We can use a GLH to test a single parameter:

$$
\begin{aligned}
& H_0: \quad \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0 \\
\Rightarrow \quad & H_0: \quad \beta_2 = 0
\end{aligned}
$$

To compare two or more parameters:

$$
\begin{aligned}
& H_0: \quad \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0 \\
\Rightarrow \quad & H_0: \quad \beta_1 - \beta_2 = 0 \quad \Rightarrow \quad H_0: \beta_1 = \beta_2
\end{aligned}
$$

GLH Examples (cont.)

To conduct simultaneous tests:

$$H_0: \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad H_0: \quad \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad H_0: \beta_1 = 0 \text{ and } \beta_3 = 0$$

To conduct simultaneous tests that compare two or more parameters:

$$H_0: \quad \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad H_0: \quad \begin{pmatrix} \beta_1 - \beta_2 \\ \beta_1 - \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad H_0: \beta_1 = \beta_2 \text{ and } \beta_1 = \beta_3$$

Testing a General Linear Hypothesis

Testing a GLH

The F-test can be used to test the GLH:

$$F = \frac{[(\mathbf{c}\hat{\beta} - \mathbf{d})^T(\mathbf{c}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}^T)^{-1}(\mathbf{c}\hat{\beta} - \mathbf{d})/r]}{\hat{\sigma}^2_{Y|X}} \sim F_{r,n-p-1}$$

This reduces to our Partial F-test for testing a group of variables, because

$$(\mathbf{c}\hat{\beta})^T(\mathbf{c}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}^T)^{-1}(\mathbf{c}\hat{\beta}) = SS_{model}(full) - SS_{model}(reduced)$$

(Therefore, a test of a single linear hypothesis for a single parameter reduces to our t-test.)

Example

Tests of General Linear Hypothesis: Example

Recall our birth weight and smoking data set from the categorical variables lecture (non-smoker is the reference category):

$$E[birthweight] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Say we want to test the hypotheses:

$$H_0 : \beta_{heavy} = \beta_{light}, \text{ or equivalently } H_0 : \beta_{heavy} - \beta_{light} = 0$$

This can written in the General Linear Hypothesis framework as:

$$H_0 : \begin{pmatrix} 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = 0$$

Example: R and the glh.test function

Testing GLH's can be implemented easily with the glh.test function in R:

```
BWT <- read.csv("birthweight_smoking_dataset.csv", header=T)
# Create indicator variables
BWT$Xf <- BWT$momsmoke=='Former'
BWT$Xl <- BWT$momsmoke=='Light'
BWT$Xh <- BWT$momsmoke=='Heavy'
BWT$Xn <- BWT$momsmoke=='Non'
```

```
library(gmodels)
mod <- lm(birthwt~Xf+Xl+Xh,data=BWT)
c_matrix <- rbind( c(0,0,-1,1) ) # Construct c matrix to test B_heavy-B_light=0
glh.test(mod, c_matrix, d=rep(0,nrow(c_matrix))) # test GLH
```

```
##
##   Test of General Linear Hypothesis
## Call:
## glh.test(reg = mod, cm = c_matrix, d = rep(0, nrow(c_matrix)))
## F = 0.4225, df1 =  1, df2 = 23, p-value = 0.5221
```

Simultaneous Hypothesis Test Example

Say we want to test the hypothesis:

$$H_0 : \beta_{heavy} = \beta_{light} = 0$$

This can be written in the GLH framework as:

$$H_0 : \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Simultaneous Hypothesis Test Example (cont.)

```r
 # Construct c matrix to test B_heavy=B_light=0
c_matrix <- rbind( c(0,0,1,0),c(0,0,0,1) )
glh.test(mod, c_matrix, d=rep(0,nrow(c_matrix)))
```

```
##
##   Test of General Linear Hypothesis
## Call:
## glh.test(reg = mod, cm = c_matrix, d = rep(0, nrow(c_matrix)))
## F = 5.7061, df1 =  2, df2 = 23, p-value = 0.00972
```

Example: SAS's PROC REG

SAS can also conduct these tests within `PROC REG`:

```
proc reg data=bw;

    model birthwt = former light heavy;

    /* test 1 */
    test light=heavy;

    /* test 2 */
    test light=heavy=0;

run;
```

Test 1 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.37296	0.42	0.5221
Denominator	23	0.88277		

Test 2 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	5.03715	5.71	0.0097
Denominator	23	0.88277		

Contrasts

Contrasts

Contrasts are most often used to test linear combinations of group means in a Cell Means Model, but can also be utilized in other modeling approaches as well.

A **contrast** (or **linear contrast**) (L) is any linear combination of parameters whose coefficients add up to 0. Specifically,

$$L = \sum_{i=1}^{k} c_i \mu_i \text{ where } \sum_{i=1}^{k} c_i = 0$$

Orthogonal contrasts are a set of contrasts such that for all pairs of contrasts, the cross-product of the coefficients is zero (assuming equal sample sizes):

$$\sum_{i=1}^{k} c_{Ai} c_{Bi} = 0$$

Orthogonality is desirable because the Model Sums of Squares is partitioned into statistically independent sums of squares.

Orthogonal polynomial contrasts are a special set of orthogonal contrasts that test for polynomial patterns in the data. Useful because they remove the inherent multicollinearity associated with polynomial terms.

These are described in more detail in the Supplemental Material slides.