Lecture 8: Testing Hypotheses in Multiple Regression

Reading Assignment:

- Muller and Fetterman, Chapter 5: "Testing Hypotheses in Multiple Regression"

After fitting a model, one seeks to draw inferences about parameters. Correlations and confidence intervals measure scientific importance, while tests and p-values assess statistical "significance". The two concepts are not necessarily the same!

Review of GLH Concepts

- Assume HILE Gauss and let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

- We estimate the $\boldsymbol{\beta}$ as $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ or $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$.

- Compute $\widehat{\boldsymbol{y}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}}$, $SSE = (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) = \mathbf{y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{y}$, and $\widehat{\sigma}^2 = MSE = \frac{SSE}{dfE}$, with $r = \mathsf{rank}(\boldsymbol{X})$ and $dfE = n - r$.

- Define $\boldsymbol{C}_{a \times p}$, which implies $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$, and state the GLH as $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. If $\boldsymbol{X}$ is less than full rank, check estimability.

- Estimate secondary parameters: $\widehat{\boldsymbol{\theta}} = \boldsymbol{C}\tilde{\boldsymbol{\beta}}$.

- Compute $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'$ and

$$SSH = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\boldsymbol{M}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

- Compute $F_{obs} = \frac{SSH/dfH}{SSE/dfE}$ for $dfH = a$. reject H_0 if $F_{obs} > F_F^{-1}(1 - \alpha, a, n - r)$.

Although all the tests we consider fall into the framework of the General Linear Hypothesis (GLH), we will discuss several different subclasses of hypothesis tests in the GLH framework, including the following.

1. *Overall tests*, which measure the contribution of the entire set of predictors

2. *Addition of one variable tests*, which measure the contribution of one variable beyond others

3. *Intercept tests*, which measure the value of the intercept in predicting the response

4. *Addition of a group of variables tests*, which measure the contribution of a group of variables beyond that of others

5. GLH tests, which include the above tests and others not included in the above categories.

Calculating Test Statistic:

A GLM test statistic can be expressed in terms of SSE's (and SSR's) or $\{(\boldsymbol{X}'\boldsymbol{X})^{-1}, \widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2, \boldsymbol{C}, \boldsymbol{\theta}_0\}$. Under HILE Gauss, the likelihood ratio test statistic follows an F distribution.

All tests compare two models: the full model and the reduced model (this is the basic idea of likelihood ratio tests, called the *likelihood ratio principle*).

We say that the reduced model is *nested* in the full model (so that its parameters are a subset of the parameters in the full model). Sometimes, investigators wish to compare non-nested models. (Example: An investigator wants to compare a model that uses weight as a continuous predictor to a model that uses indicator variables for overweight, underweight, or normal weight.) This is much more difficult and involves model-selection criteria other than F-tests. We

will only consider *nested* models for the present.

It is important to recognize that tests decompose a fixed amount of variance, σ_y^2, so that sums of squares are not created or erased but are simply moved into different locations (from SSE to SSR or vice versa). An important consequence of this fact is that SSR increases $\Leftrightarrow SSE$ decreases.

Choosing an Error Term

In many cases, a series of nested models are compared. Tests are conducted for each model in the series, comparing that model to another.

An important question is "What error term should be used?".

- It is safest to use the error term from the largest model in the entire series instead of the larger of any given pair because $SSE/(n-r)$ from the largest model "guarantees" (by assumption) an unbiased estimate of σ^2.

- If terms beyond those in the smaller model of a pair are unimportant, then the SSE from the smaller model allows a more powerful test because the error df are $(n - r + d)$ rather than $(n - r)$, where d indicates the number predictors by which the models differ.

- If terms beyond those in the smaller model of a pair are important, but we use the SSE from the smaller, inadequate model, then we will inflate (bias upwards) the estimate of σ^2. This may substantially reduce power.

- A model that is too small (under-fitting) gives biased $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ and possibly leads to a large power loss for any n.

- A model that is too large gives unbiased $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ and usually only a small power loss that goes to zero as n goes to ∞.

- Because $(n - r + d)/(n - r)$ should be near 1.0 with a sufficient amount of data, the choice of error term should not really matter.

- As a general rule, we recommend using $\widehat{\sigma}^2$ from the largest model in a series. This largest model is often called (somewhat loosely) the *full model*.

Comparing Two Models

If two models differ only by the addition or deletion of one or more variables, then

$$
\begin{aligned}
USS(\text{larger}) &\geq USS(\text{smaller}) \\
CSS(\text{larger}) &\geq CSS(\text{smaller}) \\
SSE(\text{larger}) &\leq SSE(\text{smaller}),
\end{aligned}
$$

because the larger model explains more of the variability in the data and thus has a larger SS model (and smaller SSE) than the smaller model. The SS Total of the two models should be identical.

Note that CSS may not be defined for all models.

Test Class I: Overall Tests

Corrected Overall Test

The *corrected overall test* compares the full model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i,$$

to the intercept-only model,

$$y_i = \beta_0 + \varepsilon_i.$$

If the model does not span an intercept this test is not defined.

This test corresponds to

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

and

$$H_0 : \rho_c^2 = 0.$$

Let $SSE(\text{reduced}) = SSE(\beta_0)$ be the sum of squares for error for the intercept-only model, and let $SSE(\text{full}) = SSE(\beta_0, \beta_1, \ldots, \beta_{p-1})$ be the SSE for the full model. (Generally, $SSE(\beta_i, \beta_j)$ is the SSE for a model including the two parameters β_i and β_j.) Then the sum of squares for the hypothesis is given by

$$
\begin{aligned}
SSH &= SSE(\beta_0) - SSE(\beta_0, \beta_1, \ldots, \beta_{p-1}) \\
&= SSE(\text{reduced model}) - SSE(\text{full model}) \\
&= CSS(\text{full model}) - CSS(\text{reduced model}) \\
&= CSS(\text{full model}) - 0.
\end{aligned}
$$

The F statistic is

$$
F_{obs} = \dfrac{\dfrac{SSE(\text{reduced}) - SSE(\text{full})}{df\,E(\text{reduced}) - df\,E(\text{full})}}{SSE(\text{full})/df\,E(\text{full})}
$$

$$
= \dfrac{\dfrac{SSE(\beta_0) - SSE(\beta_0, \beta_1, \ldots \beta_{p-1})}{df\,E(\beta_0) - df\,E(\beta_0, \beta_1, \ldots \beta_{p-1})}}{SSE(\beta_0, \beta_1, \ldots \beta_{p-1})/df\,E(\beta_0, \beta_1, \ldots \beta_{p-1})}
$$

$$
= \dfrac{CSS(\beta_0, \ldots, \beta_{p-1})\big/(p-1)}{SSE(\beta_0, \ldots, \beta_{p-1})\big/(n-p)},
$$

for full rank $\boldsymbol{X}$.

The hypothesis is rejected if

$$
F_{obs} > F_F^{-1}(1 - \alpha, p - 1, n - p).
$$

We may also obtain the same test via the GLH formulation with

$$F_{obs} = \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{M}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/df H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/df E},$$

where

- $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' M^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is the SSH,

- $df H = a$ is the hypothesis degrees of freedom or the number of rows of $\boldsymbol{\theta}$ (unless we pick a redundant $\boldsymbol{\theta}$),

- $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ is the SSE, and

- $df E = \mathsf{rank}(\mathbf{I} - \mathbf{H})$ is the degrees of freedom for error.

To conduct the corrected overall test, we pick

$$\boldsymbol{C} = \begin{bmatrix} \mathbf{0}_{p-1} & | & \boldsymbol{I}_{p-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \end{bmatrix}_{(p-1)\times p}$$

and $\boldsymbol{\theta}_0 = \mathbf{0}$, which imply that $\boldsymbol{\theta} = [\beta_1, \ldots \beta_{p-1}]'$ and $a = p - 1$.

The following R and SAS code may be used to construct the corrected overall test for the ozone data, which is a test of

$$H_0 : \beta_{OUTDOOR} = \beta_{HOME} = \beta_{TIMEOUT} = 0.$$

```
> ozone = read.table("ozone.txt", header = T, sep = " ") # space delimited
> n = nrow(ozone) # number of rows of ozone (obs)
> X = model.matrix(~ outdoor + home + time_out, data = ozone) # predictor matrix
> y = ozone$personal # response
> head(X)
```

```
  (Intercept)  outdoor  home time_out
1           1 35.87771 22.29     0.57
2           1 43.79189 13.97     0.90
3           1 49.81255 18.96     0.55
4           1 34.37366 22.27     0.17
5           1 45.95496 23.40     0.00
6           1 64.76558 39.62     0.30
```

```
> bhat = solve(t(X) %*% X) %*% t(X) %*% y # from notes
> yhat=X%*%bhat; # predicted values
> ehat=y-yhat; # residuals
> p=ncol(X); # num columns of X
> df=n-p; # df
```

```
> I_n = diag(rep(1, n))
> H=X%*%solve(t(X) %*% X) %*% t(X)
> sse = t(y) %*%(I_n - H)%*% y # SSE
> mse=sse/df; # MSE
> C=matrix(c(0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1), nrow = 3, byrow = T)
> print(C)

     [,1] [,2] [,3] [,4]
[1,]    0    1    0    0
[2,]    0    0    1    0
[3,]    0    0    0    1

> M=C%*%solve(t(X)%*%X)%*%t(C)
> thetahat=C%*%bhat
> ssh=t(thetahat)%*%solve(M)%*%thetahat
> msh=ssh/nrow(thetahat)
> f_obs=msh/mse
> p=1-pf(f_obs,3,60)
> print(f_obs)

          [,1]
[1,] 9.922767

> print(p)
```

```
          [,1]
[1,] 2.094332e-05
```

We may also obtain the same values in SAS after running the following

```
proc glm data=ozone;
model personal= outdoor home time_out;
run;
```

The GLM Procedure

Dependent Variable: personal   Personal Ozone Exposure (ppb)

Source	DF	Sum of Squares	Mean Square	F Value
Model	3	5034.90667	1678.30222	9.92
Error	60	10148.19129	169.13652	
Corrected Total	63	15183.09796		

Source	Pr > F
Model	<.0001
Error	
Corrected Total	

We reject the null hypothesis that outdoor concentration of ozone, home concentration of ozone, and proportion of time spent outdoors have no effect on personal ozone exposure. We conclude that at least one of our predictors (outdoor concentration of ozone, home concentration of ozone, or proportion of time spent outdoors) is related to personal ozone exposure.

Heuristics

For data that are exactly normal, this F test is a *likelihood ratio test*. You will learn more about the likelihood ratio test in BIOS 661, so our discussion will be brief.

Our general approach is to fit two models: one full model containing all the parameters of interest, and one smaller model with only an intercept. Then, we compare their maximized likelihoods (or log-likelihoods).

For a fixed value of σ^2, the maximized log-likelihood of the intercept-only model is

$$\max \log L(\beta_0) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2}\frac{SSE(\beta_0)}{\sigma^2},$$

and the maximized log-likelihood of the larger model is

$$\max \log L(\beta_0, \ldots, \beta_{p-1}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2}\frac{SSE(\beta_0, \ldots, \beta_{p-1})}{\sigma^2}.$$

To compare the log-likelihoods, we calculate minus twice their difference (denoted $-2\log\lambda$), which is given by

$$\frac{SSE(\beta_0) - SSE(\beta_0, \ldots, \beta_{p-1})}{\sigma^2}.$$

We see that this criterion is used to judge whether the additional predictors associated with $(\beta_1, \ldots, \beta_{p-1})$ result in a significant reduction in the SSE. This likelihood ratio criterion, $-2\log\lambda$, has an exact F distribution when our data are exactly normal. In BIOS 661, you will learn more about the asymptotic distribution of this criterion, useful when the data are *not* exactly normal.

Uncorrected Overall Test (rarely used)

The *uncorrected overall test* is a test of whether all parameters (including an intercept, if there is one) are equal to zero. This test is defined whether or not the model includes an intercept. It compares the null model,

$$y_i = \varepsilon_i,$$

with the full model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i$$

and tests both

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

and

$$H_0 : \rho_{\mathsf{u}}^2 = 0.$$

The test statistic and p-value are location dependent.

The F statistic is

$$F_{obs} = \frac{\dfrac{SSE(\text{reduced}) - SSE(\text{full})}{df\,E(\text{reduced}) - df\,E(\text{full})}}{SSE(\text{full})/df\,E(\text{full})}$$

$$= \frac{\dfrac{SSE(\text{null}) - SSE(\beta_0, \beta_1, \ldots \beta_{p-1})}{df\,E(\text{null}) - df\,E(\beta_0, \beta_1, \ldots \beta_{p-1})}}{SSE(\beta_0, \beta_1, \ldots \beta_{p-1})/df\,E(\beta_0, \beta_1, \ldots \beta_{p-1})}$$

$$= \frac{USS(\beta_0, \ldots, \beta_{p-1})\big/p}{SSE(\beta_0, \ldots, \beta_{p-1})\big/(n-p)},$$

for full rank $\boldsymbol{X}$, the hypothesis is rejected if
$F_{obs} > F_F^{-1}(1 - \alpha, p, n - p)$.

We may also obtain the same test via the GLH formulation with

$$F_{obs} = \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{M}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/df\,H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/df\,E},$$

where $df\,H = p$ and $df\,E = n - p$ when $\mathbf{X}$ is full rank.

To conduct the uncorrected overall test, we pick

$$C = \begin{bmatrix} \mathbf{I}_p \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ \vdots & \ddots & \\ 0 & & 1 \end{bmatrix}_{p \times p}$$

and $\boldsymbol{\theta}_0 = \mathbf{0}$. Thus $\boldsymbol{\theta} = [\beta_0, \beta_1, \ldots \beta_{p-1}]'$ and $a = p$.

Model Pools

Added-Last Pool

With p parameters, 2^p distinct models can be defined. We usually only consider subsets formed by some rule. For added-last tests, model j has β_j deleted and contains $p-1$ parameters.

An *added-last test* compares the full model to a reduced model that is obtained by deleting the jth variable from the full model. In this subset, there exist p reduced models (corresponding to deleting each of the p parameters). Each reduced model has $p-1$ parameters, and the full model has p parameters. The model pool for added-last testing is provided below.

Added-Last Model Pool

Model				
-0	$y_i=$	$\beta_1 x_{i1}+\beta_2 x_{i2}$	$\cdots$	$\cdots+\beta_{p-1}x_{ip-1}+\varepsilon_i$
-1	$y_i=\beta_0$	$+\beta_2 x_{i2}$	$\cdots$	$\cdots+\beta_{p-1}x_{ip-1}+\varepsilon_i$
$\vdots$				
$-j$	$y_i=\beta_0+\beta_1 x_{i1}$	$\cdots$ $+\beta_{j-1}x_{ij-1}$		$\cdots+\beta_{p-1}x_{ip-1}+\varepsilon_i$
$\vdots$				
full	$y_i=\beta_0+\beta_1 x_{i1}$	$\cdots$ $+\beta_{j-1}x_{ij-1}+\beta_j x_{ij}$	$\cdots+\beta_{p-1}x_{ip-1}+\varepsilon_i$	

Added-In-Order Pool

An *added-in-order test* compares parameters added in sequence. In this subset, there exist $p+1$ models (the null model and p models corresponding to adding each of the p parameters). The $(p+1)$th model is the full model and has p parameters, the pth model has one fewer parameter, the $(p-1)$th model has two fewer parameters, etc.

The model pool for added-in-order testing is provided below.

$$\text{Added-in-Order Model Pool}$$

Model

$\emptyset$	$y_i = \varepsilon_i$
0	$y_i = \beta_0 + \varepsilon_i$
1	$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$
$\vdots$	
$j-1$	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{j-1} x_{ij-1} + \varepsilon_i$
j	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{j-1} x_{ij-1} + \beta_j x_{ij} + \varepsilon_i$
$\vdots$	
$p-1$	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{j-1} x_{ij-1} + \beta_j x_{ij} \ldots + \beta_{q-1} x_{iq-1} + \varepsilon_i$

Model $p-1$ is the full model.

Test Class 2: Addition of One Variable

Added-Last Test

The *added-last test* seeks to assess the usefulness of one predictor, above and beyond all others. An added-last test for variable j compares the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{j-1} x_{ij-1} + \beta_{j+1} x_{ij+1} + \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i$$

to the full model and corresponds to testing the hypothesis $H_0 : \beta_j = 0$ when β_j is the final variable added to the model.

In this test class, there are $(p-1)$ reduced models (eliminating each of the $(p-1)$ non-intercept parameters) with $(p-1)$ parameters each (including the intercept). We compare each reduced model to the full model to produce an added-last-test for each predictor.

Added-last tests do not depend on the order of variables in the model. Added-last tests are printed in SAS as the Type III SS tests.

Denote the sum of squares for error for the reduced model (without the jth variable) by

$$SSE(\beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{p-1}).$$

Then the F statistic is

$$
\begin{aligned}
F_{obs} &= \frac{\dfrac{SSE(\text{reduced}) - SSE(\text{full})}{df E(\text{reduced}) - df E(\text{full})}}{SSE(\text{full})/df E(\text{full})} \\[2em]
&= \frac{\dfrac{SSE(\beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{p-1}) - SSE(\beta_0, \ldots, \beta_{p-1})}{df E(\beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{p-1}) - df E(\beta_0, \ldots, \beta_{p-1})}}{SSE(\beta_0, \ldots, \beta_{p-1})\big/ df E(\beta_0, \ldots, \beta_{p-1})} \\[2em]
&= \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{M}^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \big/ df H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \big/ df E},
\end{aligned}
$$

where

$$\mathbf{C} = \begin{bmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}_{1 \times p}$$

has a 1 in the position corresponding to the variable we wish to test, and 0 elsewhere. If we let $\theta_0 = 0$, then we have $\theta = \beta_j$ and $a = 1$.

$T_{obs} = F_{obs}^{\frac{1}{2}}$ follows a Student's T distribution with $(n - p)$ df if $\boldsymbol{X}$ is full rank. Many regression programs report T_{obs} as a test of the jth regression coefficient equaling zero, so the usual T test for a regression coefficient provides an added last test.

We reject the hypothesis if $F_{obs} \geq F_F^{-1}(1 - \alpha, 1, n - p)$.

SAS code and output for an added-last test of the variable $O_{OUTDOOR}$ is provided.

```
> C=matrix(c(0, 1, 0, 0), nrow = 1, byrow = T)
> print(C)

     [,1] [,2] [,3] [,4]
[1,]    0    1    0    0

> M=C%*%solve(t(X)%*%X)%*%t(C)
> thetahat=C%*%bhat
> ssh=t(thetahat)*solve(M)%*%thetahat
> msh=ssh/nrow(thetahat)
> f_obs=msh/mse
> p=1-pf(f_obs,1,60)
> print(f_obs)

          [,1]
[1,] 1.022314

> print(p)

          [,1]
[1,] 0.3160309

proc glm data=ozone;
model personal= outdoor home time_out;
run;
```

```
                    The GLM Procedure


Source                   DF      Type I SS     Mean Square    F Value

outdoor                   1    2419.043105    2419.043105      14.30
home                      1    2086.106656    2086.106656      12.33
time_out                  1     529.756909     529.756909       3.13


           Source                    Pr > F

           outdoor                   0.0004
           home                      0.0009
           time_out                  0.0818



Source                   DF     Type III SS    Mean Square    F Value

outdoor                   1     172.910655     172.910655       1.02
home                      1    2208.421320    2208.421320      13.06
time_out                  1     529.756909     529.756909       3.13


           Source                    Pr > F

           outdoor                   0.3160
           home                      0.0006
```

```
                 time_out                    0.0818


                              Standard
     Parameter          Estimate             Error     t Value      Pr > |t|


     Intercept        3.78348593        4.34205547        0.87        0.3870
     outdoor          0.09142005        0.09041683        1.01        0.3160
     home             0.59543659        0.16478332        3.61        0.0006
     time_out        13.64453832        7.70973105        1.77        0.0818
```

Again, note that the t value squared is equal to the F statistic for the added-last test (Type III SS). We fail to reject the null hypothesis and conclude that outdoor ozone concentration does not make a significant contribution to personal exposure above and beyond the intercept, home ozone, and the proportion of time spent outdoors.

Caution: Added-Last Tests

Suppose you want to estimate a person's weight given the length of their legs. (Legs account for 30-35% of human body weight.) First, you fit the model

$$E(\text{weight}) = \beta_0 + \beta_1 \text{ right leg length} .$$

The results of this model fit are provided below.

```
proc glm;
model weight=rleg;
run;
*********************************************************************************
                          The GLM Procedure


Dependent Variable: weight


                                    Sum of
Source                    DF        Squares      Mean Square    F Value
Model                      1     3627.670184     3627.670184     125.75
Error                     98     2827.099916       28.847958
Corrected Total           99     6454.770100


              Source                        Pr > F
```

```
                Model                            <.0001


        R-Square       Coeff Var        Root MSE       weight Mean
        0.562016        6.969661        5.371030          77.06300


Source                        DF       Type I SS       Mean Square    F Value
rleg                           1     3627.670184       3627.670184     125.75

                Source                            Pr > F
                rleg                             <.0001


Source                        DF      Type III SS      Mean Square    F Value
rleg                           1     3627.670184       3627.670184     125.75

                Source                            Pr > F
                rleg                             <.0001


                               Standard
    Parameter          Estimate            Error      t Value     Pr > |t|
    Intercept       -3.731074382       7.22481246       -0.52       0.6067
    rleg             1.008948511       0.08997309       11.21       <.0001
```

The added-last test, given by the "Type III" results, leads us to conclude that right leg length is significantly related to weight.

Now consider the results from fitting the model

$$E(\text{weight}) = \beta_0 + \beta_1 \text{ right leg length } + \beta_2 \text{ left leg length.}$$

```
proc glm;
model weight=rleg lleg;
run;
*************************************************************************
                         The GLM Procedure


Dependent Variable: weight


                                       Sum of
Source                      DF         Squares     Mean Square    F Value
Model                        2     3698.009886     1849.004943      65.06
Error                       97     2756.760214       28.420208
Corrected Total             99     6454.770100


                  Source                        Pr > F
                  Model                         <.0001


        R-Square        Coeff Var        Root MSE      weight Mean
        0.572911         6.917795        5.331061        77.06300


Source                      DF       Type I SS      Mean Square    F Value
```

rleg	1	3627.670184	3627.670184	127.64
lleg	1	70.339702	70.339702	2.47

Source	Pr > F
rleg	<.0001
lleg	0.1189

Source	DF	Type III SS	Mean Square	F Value
rleg	1	88.17311509	88.17311509	3.10
lleg	1	70.33970223	70.33970223	2.47

Source	Pr > F
rleg	0.0813
lleg	0.1189

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-1.828644701	7.27229423	-0.25	0.8020
rleg	9.433109580	5.35550484	1.76	0.0813
lleg	-8.447148369	5.36937194	-1.57	0.1189

The added-last tests of right leg length (p=0.08) and left leg length (0.12) indicate that neither significantly predicts weight! This occurs because after you adjust for the effect of the length of one leg, the other provides no additional useful information. Such situations arise often in biostatistics, especially when predictors may be highly correlated.

With an added-in-order test ("Type I" in SAS), which we will discuss next, instead of evaluating each variable in the presence of all others, we will evaluate variables in a fixed order. Using that type of test for the leg data, you find that the leg added first (here, the right leg) does significantly predict weight, while the addition of the second leg to the model does not sharpen our prediction over a model already containing one leg. Due to the high correlation of left and right leg lengths (here, greater than 0.99), this result is not surprising.

Added-in-Order Test

The *added-in-order test* seeks to assess the contribution of predictor j above and beyond all of the preceding $j-1$ predictors (without the $j+1$, $j+2$, etc. predictors in the model).

Added-in-order SS are mutually exclusive and together exhaustive pieces of the model SS (i.e.,if you add the SS for each predictor, the resulting sum is the model SS). Results do depend on order of inclusion except when all predictors are uncorrelated. If all predictors are uncorrelated, then added-last SS coincide with added-in-order.

Added-in-order tests are available from SAS type I SS. The sum of squares for the hypothesis for the added-in-order test for the variable j is given by

$$
\begin{aligned}
SSH &= SSE(\beta_0, \ldots, \beta_{j-1}) - SSE(\beta_0, \ldots, \beta_{j-1}, \beta_j) \\
&= SSE(\text{smaller}) - SSE(\text{larger}).
\end{aligned}
$$

The recommended F statistic is given by

$$
F_{obs} = \frac{\dfrac{SSE(\text{smaller}) - SSE(\text{larger})}{df\,E(\text{smaller}) - df\,E(\text{larger})}}{SSE(\text{full})/df\,E(\text{full})}
$$

$$
= \frac{\dfrac{SSE(\beta_0,\beta_1,\ldots,\beta_{j-1}) - SSE(\beta_0,\ldots,\beta_j)}{df\,E(\beta_0,\beta_1,\ldots,\beta_{j-1}) - df\,E(\beta_0,\ldots,\beta_j)}}{SSE(\beta_0,\ldots,\beta_{p-1})\big/(n-p)}
$$

$$
= \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'M^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big/df\,H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}\big/df\,E},
$$

where

$$
\mathbf{C} = \begin{bmatrix} 0 & \ldots & 0 & 1 \end{bmatrix}_{1\times(j+1)}
$$

and multiplies the $\boldsymbol{\beta}_{(j+1)\times 1}$ vector corresponding to the larger model. If we let $\theta_0 = 0$, then we have $\theta = \beta_j$ and $a = 1$.

(Although it is possible to use the MSE from the larger of the two models being compared as the denominator of the F test, the MSE from the full model is preferred.)

Test Class 3: Tests of the Intercept

General Features

While tests of the intercept are generally not recommended, they may be obtained by treating the intercept variable (a column of 1's) just like any other variable in Class 2 (Addition of One Variable Tests).

Muller and Fetterman discuss examples of situations in which tests of the intercept are valid.

Test Class 4: Addition of a Group of Variables

The tests in this class are conducted just like tests in Class 2, with the exception that now we are testing more than one variable (and thus have more than one degree of freedom for the hypothesis).

Group Added-Last Tests

We assume predictors fall into two groups $\{X_1, \ldots X_{g_1}\}$ and $\{X_{g_1+1}, \ldots X_{p-1}\}$, with g_1 variables in the first group and $g_2 = [(p-1) - g_1]$ variables in the second group. Note that $p = 1 + g_1 + g_2$.

Group Added-Last Pool

Model	
1	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{g_1} x_{ig_1} \hspace{4cm} + \varepsilon_i$
2	$y_i = \beta_0 \hspace{2.5cm} + \beta_{g_1+1} x_{ig_1+1} \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i$
3	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{g_1} x_{ig_1} + \beta_{g_1+1} x_{ig_1+1} \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i$

The added-last test of the first group compares model 2 and model 3, with corresponding hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{g_1} = 0.$$

The added last test of the second group compares model 1 and model 3, with corresponding hypothesis

$$H_0 : \beta_{g_1+1} = \beta_{g_1+2} = \ldots = \beta_{p-1} = 0.$$

Always compare the appropriate reduced model to the full model. The methods generalize to any number of groups.

These tests are often called *lack-of-fit* tests and measure the adequacy of a smaller model compared to a larger model.

The F statistic for an added-last test of the first group is given by

$$
\begin{aligned}
F_{obs} &= \frac{\dfrac{SSE(\text{reduced}) - SSE(\text{full})}{df\,E(\text{reduced}) - df\,E(\text{full})}}{SSE(\text{full})/df\,E(\text{full})} \\[2em]
&= \frac{\dfrac{SSE(\beta_{g_1+1}, \beta_{g_1+2}, \ldots, \beta_{p-1}) - SSE(\beta_0, \ldots, \beta_{p-1})}{df\,E(\beta_{g_1+1}, \beta_{g_1+2}, \ldots, \beta_{p-1}) - df\,E(\beta_0, \ldots, \beta_{p-1})}}{SSE(\beta_0, \ldots, \beta_{p-1})/(n-p)} \\[2em]
&= \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{M}^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/df\,H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/df\,E},
\end{aligned}
$$

where

$$
\mathbf{C} = \begin{bmatrix} \mathbf{0}_{\mathbf{g_1} \times \mathbf{1}} & \mathbf{I}_{g_1} & \mathbf{0}_{\mathbf{g_1} \times \mathbf{g_2}} \end{bmatrix}
$$

and $\boldsymbol{\theta}_0 = \mathbf{0}$, which implies $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_{g_1})$ and $a = g_1$.

Exercise: Computing a Group Added Last Test

Given the following SAS code, compute the group added-last test of no effect of outdoor ozone concentration and home ozone concentration for the ozone data.

```
proc glm data=ozone;
model personal=outdoor home time_out;
run;
proc glm data=ozone;
model personal=time_out;
run;
```

```
********************************************************************************

                           The GLM Procedure


Dependent Variable: personal   Personal Ozone Exposure (ppb)


                                   Sum of
Source                      DF     Squares      Mean Square    F Value
Model                        3    5034.90667     1678.30222      9.92
Error                       60   10148.19129      169.13652
Corrected Total             63   15183.09796
```

```
                  Source                      Pr > F
                  Model                      <.0001


       R-Square      Coeff Var        Root MSE      personal Mean
       0.331613      55.23389        13.00525          23.54578
```

The GLM Procedure

Dependent Variable: personal   Personal Ozone Exposure (ppb)

```
                                  Sum of
Source                   DF       Squares      Mean Square    F Value
Model                     1      574.49202       574.49202       2.44
Error                    62    14608.60594       235.62268
Corrected Total          63    15183.09796


                  Source                      Pr > F
                  Model                      0.1235



       R-Square      Coeff Var        Root MSE      personal Mean
       0.037838      65.19217        15.35001          23.54578
```

Group Added-In-Order Tests

The *group added-in-order test* seeks to assess the usefulness of a group of predictors above and beyond that of preceding predictors.

Group Added-in-Order Pool

Model	
0	$y_i = \beta_0 \hspace{10em} +\varepsilon_i$
1	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{g_1} x_{ig_1} \hspace{6em} +\varepsilon_i$
2	$y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_{g_1} x_{ig_1} + \beta_{g_1+1} x_{ig_1+1} \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i$

An added-in-order test of the first group compares models 0 and 1, with corresponding hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{g_1} = 0.$$

Listing order strongly affects the tests except when all predictors are uncorrelated (then added-last tests coincide with added-in-order tests).

Added-in-order sums of squares are mutually exclusive and together exhaustive, while added-last sums of squares typically does not overlap.

The F statistic for an added-in-order test of the first group is given by

$$
\begin{aligned}
F_{obs} &= \frac{\dfrac{SSE(\text{smaller}) - SSE(\text{larger})}{df\,E(\text{smaller}) - df\,E(\text{larger})}}{SSE(\text{full})/df\,E(\text{full})} \\[2em]
&= \frac{\dfrac{SSE(\beta_0) - SSE(\beta_0, \ldots, \beta_{g_1})}{df\,E(\beta_0) - df\,E(\beta_0, \ldots, \beta_{g_1})}}{SSE(\beta_0, \ldots, \beta_{p-1})/(n - p)} \\[2em]
&= \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{M}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/df\,H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/df\,E},
\end{aligned}
$$

where

$$
\mathbf{C} = \begin{bmatrix} \mathbf{0}_{\mathbf{g_1} \times \mathbf{1}} & \mathbf{I}_{g_1} \end{bmatrix}
$$

and $\boldsymbol{\theta}_0 = \mathbf{0}$, which implies $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_{g_1})$ and $a = g_1$.

As with a single variable test, the in-order test for the last group always coincides with the corresponding added-last test.

Exercise: Computing a Group Added-In-Order Test

Given the additional SAS code and output below, compute the group added-in-order test of no effect of outdoor ozone concentration and home ozone concentration for the ozone data.

```
proc glm data=ozone;
model personal=/;
run;
proc glm data=ozone;
model personal=outdoor home;
run;
```


```
*******************************************************************************


                          The GLM Procedure


Dependent Variable: personal    Personal Ozone Exposure (ppb)


                                Sum of
Source                 DF       Squares      Mean Square    F Value
Model                   1    35481.84414    35481.84414     147.23
Error                  63    15183.09796      241.00155
Uncorrected Total      64    50664.94210
```

```
                         Source                      Pr > F
                         Model                       <.0001
                         Error
                     Uncorrected Total


               The GLM Procedure


Dependent Variable: personal    Personal Ozone Exposure (ppb)


                               Sum of
Source                    DF       Squares     Mean Square    F Value
Model                      2    4505.14976     2252.57488      12.87
Error                     61   10677.94820      175.04833
Corrected Total           63   15183.09796


                         Source                      Pr > F
                         Model                       <.0001


          R-Square      Coeff Var       Root MSE     personal Mean
          0.296721      56.19089       13.23058         23.54578


Source                    DF       Type I SS     Mean Square    F Value
outdoor                    1    2419.043105     2419.043105      13.82
home                       1    2086.106656     2086.106656      11.92


                         Source                      Pr > F
```

```
                        outdoor                  0.0004
                        home                     0.0010


Source                      DF     Type III SS    Mean Square   F Value
outdoor                      1      240.247837     240.247837      1.37
home                         1     2086.106656    2086.106656     11.92

                        Source                  Pr > F
                        outdoor                 0.2459
                        home                    0.0010
```

Test Class 5: Other GLH Tests

As we have seen so far, the GLH test is a powerful tool to test many hypotheses.

Example: Laboratory Instrument Validation

A hospital laboratory has obtained a new machine to measure CO bound to hemoglobin. Investigators wish to know if the new machine is producing the same readings as the old one.

If so, the predicted regression line should be a line with slope one through the origin.

Ideally,

$$
\begin{aligned}
\text{new}_i &= \text{old}_i + \varepsilon_i \\
&= 0 + 1 \cdot \text{old}_i + \varepsilon_i \\
&= \beta_0 + \beta_1 \text{old}_i + \varepsilon_i \\
\textbf{new} &= \begin{bmatrix} \mathbf{1} & \textbf{old} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \boldsymbol{\varepsilon}
\end{aligned}
$$

The equivalence of the new and old machines implies $\beta_0 = 0$ and $\beta_1 = 1$. So our null hypothesis is

$$H_0: \quad \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Setting $\boldsymbol{C} = \boldsymbol{I}_2$ yields $\boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$.

Choosing $\boldsymbol{\theta}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ defines a GLH which compares two models,

$$1 \quad \mathsf{new}_i = \mathsf{old}_i + \varepsilon_i$$

$$2 \quad \mathsf{new}_i = \beta_0 + \beta_1 \cdot \mathsf{old}_i + \varepsilon_i \,,$$

with test statistic

$$F_{obs} = \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{M}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/df H}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/df E},$$

where $df H = a = 2$.

We note that PROC GLM will only test a GLH with $\boldsymbol{\theta}_0 = \mathbf{0}$. In order to conduct this test, we must either use R or PROC REG with the TEST statement.

Multiple Testing

In multiple regression, we may wish to conduct $p - 1$ tests (one for each predictor in the model). This multiple testing often inflates α above the desired nominal level, which is generally $\alpha = 0.05$.

The Bonferroni inequality says that

$$\Pr\{A_1 \cup A_2\} = \Pr\{A_1\} + \Pr\{A_2\} - \Pr\{A_1 \cap A_2\}$$
$$\Pr\{A_1 \cup A_2\} \ \leq \ \Pr\{A_1\} + \Pr\{A_2\} \ .$$

If $\Pr\{A_1\} = \Pr\{A_2\} = \alpha_k$, then $\Pr\{A_1 \cup A_2\} \leq 2\alpha_k$.
More generally

$$\Pr\{A_1 \cup A_2 \cup A_3 \ \ldots \ \cup A_K\} \leq K\alpha_k \ .$$

To ensure an upper bound of α overall, we use $\alpha_k = \alpha/K$.

For independent events (this rarely applies to multiple comparisons!)

$$\Pr\{A_1 \cup A_2 \cup A_3 \ \ldots \ \cup A_K\} = 1 - (1 - \alpha_k)^K \ .$$

The Bonferroni correction is accurate for small K, is almost always accurate for independent events, but is least accurate for highly correlated events. The poor accuracy results in less powerful tests, but the Bonferroni correction does guarantee control of the type I error rate.

Exercise: Bonferroni Correction

Suppose we will test 10 predictors and want an overall $\alpha = 0.05$. At which level should we conduct each test?

Interaction

X_1 and X_2 interact (in predicting Y) if one must know the level of X_1 in order to describe relationship of Y to X_2 (and hence must know the level of X_2 in order to describe relationship of Y to X_1).

For example, one cancer treatment may be beneficial to patients with a certain genotype of a gene but not to other patients.

Interaction Model Pool

Model		
1	$y_i = \beta_0 \qquad\qquad\qquad +\beta_3 x_{i1} x_{i2} +\varepsilon_i$	$avoid$
2	$y_i = \beta_0 \qquad +\beta_2 x_{i2} +\beta_3 x_{i1} x_{i2} +\varepsilon_i$	$avoid$
3	$y_i = \beta_0 +\beta_1 x_{i1} \qquad +\beta_3 x_{i1} x_{i2} +\varepsilon_i$	$avoid$
4	$y_i = \beta_0 \qquad\qquad\qquad\qquad +\varepsilon_i$	
5	$y_i = \beta_0 +\beta_1 x_{i1} \qquad\qquad +\varepsilon_i$	
6	$y_i = \beta_0 \qquad +\beta_2 x_{i2} \qquad +\varepsilon_i$	
7	$y_i = \beta_0 +\beta_1 x_{i1} +\beta_2 x_{i2} \qquad +\varepsilon_i$	
8	$y_i = \beta_0 +\beta_1 x_{i1} +\beta_2 x_{i2} +\beta_3 x_{i1} x_{i2} +\varepsilon_i$	

Analysis plan:

First, test the interaction, H_0: $\beta_3 = 0$, with $C = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$ and $\theta_0 = 0$, to compare model 8 to model 7. If significant, quit.

The interaction test asks if $y_i = \alpha_0 + \alpha_1 x_{i1} + \varepsilon_i$ is the same for all x_{i2}.

For $x_{i2} \neq x'_{i2}$, the null hypothesis corresponds to parallel lines with distinct intercepts (for $\beta_3 = 0$).

If β_3 is significant, both variables are important, even if β_1 and/or β_2 is not significant. If β_3 is significant, retain both variables, even if β_1 and/or β_2 is not significant.

Always include *main effects* (here, x_1 and x_2) in the model when an interaction term between them is also in the model.

Interpreting Interaction Terms

With an interaction term in the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i,$$

it is difficult to interpret β_1 and β_2 in isolation. However, we may express the effect of a one-unit increase in x_1 with x_2 held constant as

$$
\begin{aligned}
E(y_i | x_{i1} + 1, x_{i2}) \quad - \quad & E(y_i | x_{i1}, x_{i2}) = \\
& \beta_0 + \beta_1(x_{i1} + 1) + \beta_2 x_{i2} \\
& + \beta_3(x_{i1} + 1)x_{i2} \\
& - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2}) \\
= \quad & \beta_1 + \beta_3 x_{i2}.
\end{aligned}
$$

Similarly, we can find the effect of a one-unit increase in x_2 with x_1 held constant.

Next: Correlation

Reading Assignment:

• Muller and Fetterman, Chapter 6: "Correlations" (Required)