

## The general linear model

Univariate case:

Data:  $Y_1, \dots, Y_n$ ; covariate  $r$ -vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$   
(each written as row vector).

Model

$$Y_i = \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i$$

for iid  $N(0, \sigma^2)$  errors  $\epsilon_i$  and unknown coefficient vector  $\boldsymbol{\beta}$  a  $r \times 1$  (column) vector.

Commonly: first entry in each  $\mathbf{x}_i$  is 1; corresponding  $\beta_1$  is "intercept". In text: use  $r + 1$  for  $r$  so that  $r$  is number of coefficients other than intercept.

Model fitting: Maximum Likelihood = Least squares.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

where  $\mathbf{Y}$  is  $n \times 1$ , entries  $Y_i$  and  $\mathbf{X}$  is  $n \times r$  with  $i$ th row  $\mathbf{x}_i$ .

Fitted response vector

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Fitted residual vector

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \left\{ \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \right\} \mathbf{Y}$$

What sort of hypotheses can we test:

$$\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$$

for a given  $q \times r$  matrix of constants  $\mathbf{L}$ .

**Example:** Test all  $\beta_i$  but  $\beta_1$  are 0:

$$\mathbf{L} = \left[ \mathbf{0}_{(r-1) \times 1} \mid \mathbf{I}_{r-1 \times r-1} \right]$$

**Example:** One way ANOVA with  $m$  groups:  $\mathbf{X}$  has  $m$  columns. Entry  $j$  for observation  $i$  is 1 or 0 according as case number  $i$  belongs to group  $j$  or not. For our data on 3 test scores:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

To test  $\mu_1 = \mu_2 = \mu_3 = \mu_4$  could use

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

or

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

or many others!

To test  $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ :

Fit model twice:

Full model: get Error sum of squares:

$$ESS = \|\hat{\boldsymbol{\epsilon}}\|^2 = \mathbf{Y}^T \left\{ \mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right\} \mathbf{Y}$$

Restricted model: minimize

$$\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

subject to  $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .

Method: assume rank of  $\mathbf{L}$  is  $q < r$ . (Full rank.)

Find  $\mathbf{L}^*$  a  $(r - q) \times r$  so that

$$\mathbf{P} = \begin{bmatrix} \mathbf{L}^* \\ \mathbf{L} \end{bmatrix}$$

is invertible. Write

$$\boldsymbol{\beta}^* = \mathbf{P}\boldsymbol{\beta}$$

and

$$\mathbf{X}^* = \mathbf{X}\mathbf{P}^{-1}$$

Model is

$$\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\epsilon}$$

Partition  $\boldsymbol{\beta}^*$  into  $\boldsymbol{\beta}_1^*$  of length  $r - q$  and  $\boldsymbol{\beta}_2^*$  of length  $q$ .

Then

$$\mathbf{L}\boldsymbol{\beta} = \mathbf{0} \equiv \boldsymbol{\beta}_2^* = \mathbf{0}$$

Restricted ESS is

$$ESS_R \equiv \|(\mathbf{I} - \mathbf{X}_1^* \{(\mathbf{X}_1^*)^T (\mathbf{X}_1^*)\}^{-1} (\mathbf{X}_1^*)^T) \mathbf{Y}\|^2$$

$F$  test based on

$$\frac{(ESS_R - ESS)/q}{ESS/(n - r)}$$

which has  $F_{q, n-r}$  distribution on null.

**Multivariate Case** Now  $\mathbf{Y}_i$  is a vector of  $p$  responses. Assemble, row by row, in data matrix  $\mathbf{Y}$  which is  $n \times p$ . Write model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where now  $\boldsymbol{\beta}$  is  $r \times p$ .

Model for errors: rows are iid  $MVN_p(\mathbf{0}, \boldsymbol{\Sigma})$ .

NOTE: fitting same regression model to each different response variable but with errors correlated.

Least squares estimates: as before

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Fitted response matrix

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Fitted residual matrix

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \left\{ \mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right\} \mathbf{Y}$$

What sort of hypotheses can we test:

$$\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$$

AND more general:

$$\mathbf{L}\boldsymbol{\beta}\mathbf{M} = \mathbf{0}$$

**Example:** Back to one way manova profile analysis: Design matrix  $\mathbf{X}$  has entries

$$X_{ij} = \begin{cases} 1 & \text{Subject } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$$

(Number of rows = number of observations, number of columns = number of groups.)

Matrix  $\boldsymbol{\beta}$  has row  $j$  equal to  $\boldsymbol{\mu}_j^T$ . Number of rows = number of groups, number of columns = number of response variables.

Hypotheses of interest:

1) Parallel profiles:

$$\mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'}$$

for all  $i, i', j, j'$ . Do case of 4 groups and 3 variables. So  $\beta$  is  $4 \times 3$ :

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$$

and

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}$$

2) Coincident profiles:  $\mu_i = \mu_{i'}$  for all  $i$  and  $i'$ . Now  $\mathbf{M}$  is the identity and  $\mathbf{L}$  is as above.

3) Constant profiles:  $\mathbf{L}$  is the identity,  $\mathbf{M}$  as above.



Testing  $L\beta M = 0$ .

1) Multiply model equation by  $M$ . Get

$$Y^* = X\beta^* + \epsilon^*$$

with

$$Y^* = YM$$

$$\beta^* = \beta M$$

and

$$\epsilon^* = \epsilon M.$$

Reduces problem to testing  $L\beta = 0$ .

2) So now: drop \*s and discuss test of  $L\beta = 0$ . Reparametrize again as in univariate: Find  $L^*$  a  $(p - q) \times p$  so that

$$P = \begin{bmatrix} L^* \\ L \end{bmatrix}$$

is invertible. Write

$$\beta^* = P\beta$$

and

$$X^* = XP^{-1}$$

Model is

$$Y = X^* \beta^* + \epsilon$$

Partition  $\beta^*$  into  $\beta_1^*$  with dimensions  $r - q \times p$  and  $\beta_2^*$  of dimension  $q \times p$ .

Then

$$L\beta = 0 \equiv \beta_2^* = 0$$

3) Again drop \*s and discuss test of  $\beta_2 = 0$  in model

$$Y = [X_1 | X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon$$

Union Intersection Principle:  $\beta_2 = 0$  if and only if  $\beta_2 a = 0$  for all  $p \times 1$  vectors  $a$ . So multiply model equation by  $a$  get

$$Ya = X\beta a + \epsilon a$$

or

$$Y^* = X\beta^* + \epsilon^*$$

which is now a univariate regression model.

Get corresponding  $F$  statistic:

$$F = \frac{\mathbf{a}^T \mathbf{Y}^T \mathbf{Q} \mathbf{Y} \mathbf{a} / q}{\mathbf{a}^T \mathbf{Y}^T (\mathbf{I} - \mathbf{C}) \mathbf{Y} \mathbf{a} / (n - r)}$$

where

$$\mathbf{C} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

and

$$\mathbf{Q} = \mathbf{C} - \mathbf{X}_1(\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1$$

To test intersection of null hypotheses use critical region of form union of corresponding rejection regions.

I.e. maximize  $F$  over  $\mathbf{a}$ .

Maximize

$$\frac{\mathbf{a}^T \mathbf{A} \mathbf{a}}{\mathbf{a}^T \mathbf{B} \mathbf{a}}$$

over  $\mathbf{a} \neq \mathbf{0}$ :

Solution for  $\mathbf{B}$  not singular is largest eigenvalue of

$$\mathbf{B}^{-1} \mathbf{A}$$

as earlier in course. This gives Roy's test.

**Example:** MANACOVA: 45 patients, 4 groups (defined by body type and obesity status).

Record 4 biochemical measurements on urine of each subject: three response variables, one covariate: specific gravity of the urine.

Compare groups adjusting for specific gravity.

```
run;
  data long;
  infile 'dq5.1';
  input group creatin chloride choline specgrav;
run;
proc print;
run;
proc glm;
  class group;
  model creatin chloride choline = group|specgrav;
  manova h= group*specgrav / printh printe;
run;
proc glm;
  class group;
  model creatin chloride choline = group specgrav;
  manova h= group / printh printe;
  means group / bon scheffe tukey;
run;
proc glm;
  class group;
  model creatin chloride choline = group;
  manova h= group / printh printe;
  means group / bon scheffe tukey;
run;
```

Begin by examining group by specific gravity interaction:

Characteristic Roots and Vectors of:

E Inverse \* H, where

H = Type III SS&CP Matrix for SPECGRAV\*GROUP

E = Error SS&CP Matrix

Manova Test Criteria and F Approximations for  
the Hypothesis of no Overall SPECGRAV\*GROUP Effect

S=3 M=-0.5 N=16.5

Statistic	Value	F	NumDF	DenDF	Pr > F
Wilks' Lambda	0.6153	2.0939	9	85.33	0.0387
Pillai's Trace	0.3899	1.8426	9	111	0.0683
Hotg-Lly Trace	0.6167	2.3070	9	101	0.0211
Roy's Greatest	0.6026	7.4320	3	37	0.0005

NOTE: F Statistic for Roy's Greatest Root  
is an upper bound.

The group by specific gravity interaction appears to be marginally significant so that adjusting for specific gravity comparisons is difficult.

In what follows we ignore the possibility of any such interaction.

Look at model with no interaction term:

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall group Effect

H = Type III SSCP Matrix for group

E = Error SSCP Matrix

S=3 M=-0.5 N=18

Statistic	Value	F	NumDF	Den DF	Pr > F
Wilks' Lambda	0.5972	2.43	9	92.633	0.0159
Pillai's Trace	0.4398	2.29	9	120	0.0208
Hotg-Lly Trace	0.6127	2.54	9	56.616	0.0159
Roy's Greatest	0.4877	6.50	3	40	0.0011

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Conclusion: Pretty clear evidence of differences between groups. Should examine simultaneous confidence intervals.

In same model: do we need covariate?

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall specgrav Effect

H = Type III SSCP Matrix for specgrav

E = Error SSCP Matrix

S=1 M=0.5 N=18

Statistic	Value	F	NumDF	DenDF	Pr > F
Wilks' Lambda	0.420	17.51	3	38	<.0001
Pillai's Trace	0.580	17.51	3	38	<.0001
Hotg-Lly Trace	1.382	17.51	3	38	<.0001
Roy's Greatest	1.382	17.51	3	38	<.0001

So covariate is significant.

But what if we had not measured it?

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall group Effect

H = Type III SSCP Matrix for group

E = Error SSCP Matrix

S=3      M=-0.5      N=18.5

Statistic	Value	F	NumDF	DenDF	Pr > F
Wilks' Lambda	0.608	2.40	9	95.066	0.0171
Pillai's Trace	0.431	2.30	9	123	0.0203
Hotg-Lly Trace	0.580	2.47	9	58.185	0.0185
Roy's Greatest	0.434	5.93	3	41	0.0019

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Same conclusion: difference in group means.



Why is specgrav needed in model but irrelevant to conclusions? No real differences between groups in terms of specgrav

```
data long;
  infile 'dq5.1';
  input group creatin chloride choline specgrav;
run;
proc glm;
  class group;
  model specgrav = group ;
run;
```

which gives

The GLM Procedure

Dependent Variable: specgrav

Source	DF	Sum of Squares	MeanSquare	F	Pr > F
Model	3	120.30	40.100241	1.19	0.3269
Error	41	1386.28	33.811636		
Corrd	44	1506.58			

Adjustment for covariates:

Suppose  $x$  is a covariate and  $Y$  a response. In order to compare the effect of a treatment on  $Y$  we might have a treatment and a control group.

Example:  $x$  is blood pressure before and  $Y$  is blood pressure after treatment (real or placebo depending on group).

Both  $x$  and  $Y$  must be regarded as random in this example.

Models:  $(X, Y)$  has  $BVN(\mu, \Sigma)$  distribution with  $\mu$  and  $\Sigma$  which depend on whether the subject is in treatment or control.

Control group:

$$\mu = (\mu_1, \mu_2)$$
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Notice assumptions: means for measurements 1 and 2 might be different (allows for placebo effect); no assumption of equal variability as yet.

Treatment group:

$$\mu = (\mu_1, \mu_2^*)$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho^* \sigma_1 \sigma_2^* \\ \rho \sigma_1 \sigma_2^* & (\sigma_2^*)^2 \end{bmatrix}$$

Notice: baseline means, variances assumed equal (should be true for double-blind randomized trial).

If we knew parameters how would we summarize treatment effect?

No easy way.

Simplifying assumptions: several possible.

Usual: Equal covariance matrices. Treatment effect is

$$\delta \equiv \mu_2^* - \mu_2$$

If so: how can we estimate  $\delta$ ? Notice:

$$\bar{X}_T - \bar{X}_C$$

has mean 0 so

$$\hat{\delta}_a \equiv \bar{Y}_T - \bar{Y}_C + a(\bar{X}_T - \bar{X}_C)$$

is unbiased for  $\delta$  for any choice of  $a$ .

Variance is

$$\left( \frac{1}{n_T} + \frac{1}{n_C} \right) \begin{bmatrix} 1 & a \end{bmatrix} \Sigma \begin{bmatrix} 1 \\ a \end{bmatrix}$$

Minimized when  $a = -\rho\sigma_1/\sigma^2$ ; variance becomes

$$\left( \frac{1}{n_T} + \frac{1}{n_C} \right) \sigma_1^2 (1 - \rho^2)$$

Smaller than ignoring baseline covariate. Improvement important if  $\rho^2$  is not too small.

Other treatment effect models?

Effect of treatment: “after” measurement is

$$Y + T$$

where  $T$  is a  $N(\delta, \sigma_T^2)$  effect.

Could assume  $T$  and  $X, Y$  independent. (Not too realistic.)

Could ask: what does joint  $X, Y, T$  distribution have to be to get same as previous model.

Potential problem: joint distribution of  $X, Y, T$  not estimable from this experimental design.