STAT 730 Chapter 12: MANOVA

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 730: Multivariate Data Analysis

A special case of the multivariate regression model in Chapter 6 is one-way and multi-way multivariate analysis of variance (MANOVA). To start, assume the one-way model

$$\mathbf{x}_{ij} \overset{ind.}{\sim} N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}),$$

where $i = 1, \ldots, k$ denotes the group having $j = 1, \ldots, n_i$ observations. Let $n = \sum_{i=1}^{k} n_i$ be the total number of observations.

We derived the LRT for this model in Chapter 5. Note also that the model is in the form of multivariate regression $\mathbf{Y} = \mathbf{ZB} + \mathbf{U}$ (Chapter 6) by taking $\mathbf{Y} \in \mathbb{R}^{n \times p}$ as

$$\mathbf{Y}' = [\underbrace{\mathbf{x}_{11} \cdots \mathbf{x}_{1n_1}}_{\text{group 1}} \underbrace{\mathbf{x}_{21} \cdots \mathbf{x}_{2n_2}}_{\text{group 2}} \cdots \underbrace{\mathbf{x}_{k1} \cdots \mathbf{x}_{kn_k}}_{\text{group k}}],$$

$\mathbf{B} = [\boldsymbol{\mu}_1 \boldsymbol{\mu}_2 \cdots \boldsymbol{\mu}_k]' \in \mathbb{R}^{k \times p}$, and $\mathbf{Z} = \text{block-diag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_k})$.

LRT of $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$

Recall

$$\lambda(\mathbf{X})^{2/n} = \frac{|\mathbf{W}|}{|n\mathbf{S}|} = |\mathbf{W}\mathbf{T}^{-1}|,$$

where $\mathbf{T} = n\mathbf{S}$ is the total sums of squares and cross products (SSCP) and $\mathbf{W} = \sum_{i=1}^{k} n_i \mathbf{S}_i$ is the SSCR for error, or within groups SSCP. The SSCP for regression is $\mathbf{B} = \mathbf{T} - \mathbf{W}$, or between groups. Then

$$\lambda^{n/2} = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{1}{|\mathcal{I}_p + \mathbf{W}^{-1}\mathbf{B}|}.$$

In Chapter 5 we showed

$$\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n - k) \text{ indep. } \mathbf{B} \sim W_p(\boldsymbol{\Sigma}, k - 1),$$

so that

$$\lambda^{2/n} \sim \Lambda(p, n - k, k - 1),$$

under H_0 provided $n \geq p + k$.

MANOVA decomposition

We can write the within SSCP matrix as
$\mathbf{W} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$, the total SSCP
$\mathbf{T} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})'$ and the between SSCP
$\mathbf{B} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'$. Here, $\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ and
$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$.

As in univariate one-way ANOVA $\mathbf{T} = \mathbf{W} + \mathbf{B}$, and a MANOVA table displays these along with their df.

Source	df	SSCP
Treatments	$k - 1$	**B**
Error	$n - k$	**W**
Total	$n - 1$	**T**

There are four common test statistics, including Wilks lambda from the LRT, computed for $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$.

```
data(iris)
f=manova(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
summary(f,test="Wilks") # "Pillai","Wilks","Hotelling-Lawley","Roy"
library(car)
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
model.matrix(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)
 ~Species,data=iris) # can use ~0+Species for no intercept
summary(f)            # but ANOVA test will be incorrect
?Manova # can also do type III tests, e.g. summary(Manova(f,type=3))
summary(Manova(f))
```

Compare **B** and **W** to pp. 344–345.

Fixed contrast of means

Say we want to test $H_0 : a_1\boldsymbol{\mu}_1 + \cdots + a_k\boldsymbol{\mu}_k = \mathbf{0}$. Assuming the linear regression setup $\mathbf{Y} = \mathbf{ZB} + \mathbf{U}$ described on slide 2, this is accomplished by taking $\mathbf{C} = [a_1 \cdots a_k]$, $\mathbf{M} = \mathcal{I}_p$ and $\mathbf{D} = \mathbf{0}_{1 \times p}$. For example, in the iris data if we wanted to test $H_0 : 0.5\boldsymbol{\mu}_1 + 0.5\boldsymbol{\mu}_2 - \boldsymbol{\mu}_3 = \mathbf{0}$ we could use

```
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~0+Species,
 data=iris)
linearHypothesis(f,hypothesis.matrix=matrix(c(0.5,0.5,-1),1,3))
```

If we do not force the intercept to be zero, then the first group (setosa) becomes baseline and the the regression effects are offsets for versicolor and virginica. You should verify that the correct matrix is in general $\mathbf{C} = \left[(\sum_{i=1}^{k} a_i)a_2 \cdots a_k\right]_{1 \times k}$. If $\mathbf{a}$ is a *contrast* then the first term is zero and the rest stay the same.

```
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
linearHypothesis(f,hypothesis.matrix=matrix(c(0,0.5,-1),1,3))
```

Recall the univariate test statistic

$$F^* = \sum_{i=1}^{k} \frac{n_i(\bar{y}_i - \bar{y})^2}{\sum_{j=1}^{k} n_i s_i^2} \sim F_{k-1,n-p} \text{ under } H_0.$$

The UIT considers this test for $y_{ij} = \mathbf{a}'\mathbf{x}_{ij}$. This is the same as considering $\mathbf{y} = \mathbf{X}\mathbf{a}$. Then $H_{0\mathbf{a}}$ is tested

$$F_{\mathbf{a}}^* = \frac{\sum_{i=1}^{k} n_i[\mathbf{a}'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})]^2}{\sum_{i=1}^{k} n_i \mathbf{a}'\mathbf{S}_i\mathbf{a}} = \frac{\mathbf{a}'[\sum_{i=1}^{k}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})']\mathbf{a}}{\mathbf{a}'[\sum_{i=1}^{k} n_i\mathbf{S}_i]\mathbf{a}} = \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}}.$$

From chapter 4 we know the maximum of this is λ_1, the largest e-value of $\mathbf{W}^{-1}\mathbf{B}$. The maximum occurs when $\mathbf{a} = \boldsymbol{\gamma}_{(1)}$, the corresponding e-vector.

Best standardized linear combination

The e-vector $\mathbf{a} = \boldsymbol{\gamma}_{(1)}$ is the standardized linear combination (SLC) most responsible for rejecting H_0 using Roy's greatest root. If we reject using the root statistic, we will also reject in the univariate case using $y_{ij} = \mathbf{a}'\mathbf{x}_{ij}$. We will revisit optimal SLCs in principal components analysis.

```
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris) # ~0 produces wrong B matrix
f=Manova(f) # computes W and B matrices
eigen(solve(f$SSPE)%*%f$SSP$Species)$vectors[,1]  # may need Re(...)
iris[1,] # interpretation?
```

The interpretation of $\mathbf{a} = \boldsymbol{\gamma}_{(1)}$ is exactly the same in the regression setting for general linear hypotheses; one simply needs $\mathbf{E}$ (same as $\mathbf{W}$ in MANOVA) and $\mathbf{H}$ (same as $\mathbf{B}$ in MANOVA). The first r linear combinations of $\mathbf{x}$, e.g. $y_1 = \mathbf{a}_1'\mathbf{x}, \ldots, y_r = \mathbf{a}_r'\mathbf{x}$, most responsible for rejecting H_0 are given by the first r canonical variables, described next.

The simplest possible structure is $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$. Here the means have dimension $r = 0$.

A slightly more complex possibility is that the means all lie on a straight line, e.g. $H_0 : \boldsymbol{\mu}_i = \boldsymbol{\mu} + a_i \boldsymbol{\gamma}_{(1)}$. This subspace of $\mathbb{R}^p$ has dimension $r = 1$: each mean is described by a one-dimensional $a_i \in \mathbb{R}$, given $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}_{(1)}$.

If the means lie on a two-dimensional hyperplane in $\mathbb{R}^p$ then $\boldsymbol{\mu}_i = \boldsymbol{\mu} + [\boldsymbol{\gamma}_{(1)} \boldsymbol{\gamma}_{(2)}] \mathbf{a}_i$ where $\mathbf{a}_i \in \mathbb{R}^2$.

In general, the means can all lie in an r-dimensional hyperplane of $\mathbb{R}^p$ where $r \leq \min\{k - 1, p\}$ (rank($\mathbf{B}$) $= k - 1$). We want to test H_0 : means lie in $\mathbb{R}^r$ dimensional hyperplane vs. H_a : they lare unrestricted for $r = 0, 1, \dots, \min\{k - 1, p\}$; $r = 0$ is the usual MANOVA test.

MKB (pp. 338–348) derive an approximate LRT of H_0:

$$(n - \tfrac{1}{2}(p + k)) \sum_{j=r+1}^{p} \log(\lambda_j + 1) \sim \chi^2_{(p-r)(k-r)}.$$

Book has typo for test statistic according to Bartlett (1947, eq. 14). Upon accepting this hypothesis, we can project the data and estimated $\hat{\mu}_i$ onto a lower dimensional subspace. If we restrict ourselves to $\mathbb{R}^2$, we can plot the results.

Let $\gamma_{(j)}$ be the jth e-vector of $\mathbf{W}^{-1}\mathbf{B}$ corresponding to λ_j but normalized s.t. $\gamma_{(j)}' \hat{\boldsymbol{\Sigma}}_u \gamma_{(j)} = 1$. The projection of any $\mathbf{x} \in \mathbb{R}^p$ onto the estimated r-dimensional plane is $\mathbf{P_R}\mathbf{x} = (\gamma_{(1)}'\mathbf{x}, \cdots, \gamma_{(r)}'\mathbf{x})'$ where $\mathbf{P_R} = [\gamma_{(1)} \cdots \gamma_{(r)}]'$. Often, most of the variability in the $\{\mathbf{x}_{ij}\}$ are explained by the first one or two canonical variables.

$\mathbf{P_R}\bar{\mathbf{x}}_1, \ldots, \mathbf{P_R}\bar{\mathbf{x}}_k$ are canonical means. Plot these w/ $\{\mathbf{P_R}\mathbf{x}_{ij}\}$ to show maximal separation in means from SLC.

```
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
f=Manova(f) # computes W and B matrices
e=eigen(solve(f$SSPE)%*%f$SSP$Species)
e$values=Re(e$values); e$vectors=Re(e$vectors)
p=length(e$values); k=f$df+1; n=f$error.df+k
for(r in 0:min(p,f$df-1)){
 stat=(n-0.5*(p+k))*sum(log(1+e$values[(r+1):p])); df=(p-r)*(k-1-r)
 cat("Test of ",r," dimensional subspace stat =",stat,", df = ",df,
 ", p-val =",1-pchisq(stat,df),"\n")
} # r=2 is necessary
Su=f$SSPE/(n-k)
g1=e$vectors[,1]; g1=g1/sqrt(g1%*%Su%*%g1)
g2=e$vectors[,2]; g2=g2/sqrt(g2%*%Su%*%g2)
d=as.matrix(iris[,1:4])
G=cbind(g1,g2); d12=d%*%G
plot(d12,pch=c(rep(1,50),rep(2,50),rep(3,50)),
 xlab="1st canon. variate",ylab="2nd canon. variate")
sp=matrix(0,3,2)
sp[1,]=colMeans(as.matrix(iris[1:50,1:4]))%*%G
sp[2,]=colMeans(as.matrix(iris[51:100,1:4]))%*%G
sp[3,]=colMeans(as.matrix(iris[101:150,1:4]))%*%G
ellipse(c(sp[1,]),shape=diag(2),radius=sqrt(qchisq(0.99,2)/50))
ellipse(c(sp[2,]),shape=diag(2),radius=sqrt(qchisq(0.99,2)/50))
ellipse(c(sp[3,]),shape=diag(2),radius=sqrt(qchisq(0.99,2)/50))
legend(-3,9.5,legend=c("setosa","versicolor","virginica"),pch=c(1,2,3),bty="n")
```

Note that these SLC are standardized so that $V(\mathbf{x}'_{ij}\gamma_{(j)}) = 1$.

Two-way MANOVA

Now we have

$$\mathbf{x}_{ijk} \sim N_p(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\tau}_j + (\boldsymbol{\alpha\tau})_{ij}, \boldsymbol{\Sigma}),$$

where $i = 1, \ldots, r$, $j = 1, \ldots, c$, and $k = 1, \ldots, n_{ij}$.

If $n_{ij} = m$ for all i and j the data are balanced and the total SSCP can be decomposed as

$$\mathbf{T} = \underbrace{\mathbf{R} + \mathbf{C} + \mathcal{I}}_{\mathbf{B}} + \mathbf{W}.$$

These matrices are defined on p. 351 and correspond to row, column, interaction, and within SSCP. Each matrix on the right-hand side can be respectively be written $\mathbf{X}'\mathbf{A}_i\mathbf{X}$ where $\mathbf{A}_i\mathbf{A}_4 = \mathbf{0}$ for $i = 1, 2, 3$ and each $\mathbf{A}_i$ is a projection matrix. One can show under each of $H_0 : \boldsymbol{\alpha}_i = \mathbf{0}$, $H_0 : \boldsymbol{\tau}_j = \mathbf{0}$, $H_0 : (\boldsymbol{\alpha\tau})_{ij} = \mathbf{0}$ that the LRT each boils down to a Wilk's lambda.

Tests

Within interaction model $\mathbf{x}_{ijk} \sim N_p(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\tau}_j + (\boldsymbol{\alpha\tau})_{ij}, \boldsymbol{\Sigma})$ we have Type III tests

$$H_0 : (\boldsymbol{\alpha\tau})_{ij} = \mathbf{0} \Rightarrow |\mathbf{W}(\mathbf{W}+\mathcal{I})^{-1}| \sim \Lambda(p, rc(m-1), (r-1)(c-1)).$$

$$H_0 : \boldsymbol{\tau}_j = \mathbf{0} \Rightarrow |\mathbf{W}(\mathbf{W} + \mathbf{C})^{-1}| \sim \Lambda(p, rc(m-1), c-1).$$

$$H_0 : \boldsymbol{\alpha}_j = \mathbf{0} \Rightarrow |\mathbf{W}(\mathbf{W} + \mathbf{R})^{-1}| \sim \Lambda(p, rc(m-1), r-1).$$

Within additive model $\mathbf{x}_{ijk} \sim N_p(\boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\tau}_j, \boldsymbol{\Sigma})$ we have Type III tests

$$H_0 : \boldsymbol{\tau}_j = \mathbf{0} \Rightarrow |(\mathbf{W}+\mathcal{I})(\mathbf{W}+\mathcal{I}+\mathbf{C})^{-1}| \sim \Lambda(p, rcm-r-c+1, c-1).$$

$$H_0 : \boldsymbol{\alpha}_j = \mathbf{0} \Rightarrow |(\mathbf{W}+\mathcal{I})(\mathbf{W}+\mathcal{I}+\mathbf{R})^{-1}| \sim \Lambda(p, rcm-r-c+1, r-1).$$

All tests come with "characteristic roots" – largest e-vector of $\mathbf{E}^{-1}\mathbf{B}$ in each of the above – which are the coefficients of the SLC given by the canonical variables.

Balance is nice when you've got it, but, in general...

If the data are unbalanced then one can first test for no interaction $H_0 : (\alpha\tau)_{ij} = \mathbf{0}$ using a general LRT test. If the interaction is not necessary, it can be dropped and attention turned toward the main effects. In R, all models are fit via, e.g. for $p = 3$

```
f1=lm(cbind(y1,y2,y3)~a*b,data=data) # interaction model
f2=lm(cbind(y1,y2,y3)~a+b,data=data) # additive model
```

regardless of balance. LRT tests are carried out via anova(f1,f2). Once you've decided on a model you can test contrasts as usual via linearHypothesis, obtain SSCP from Manova(f2), etc.

On p. 354. we have the "rat data:" the weight loss (in grams) of 24 rats in 1st and 2nd weeks given one of three drugs. Factors are sex (male/female) and drug (A/B/C).

```
library(car)
rats=read.table("http://www.stat.sc.edu/~hansont/stat730/rats.txt",header=T)
f1=lm(cbind(week1,week2)~sex*drug,data=rats) # interaction model
f2=lm(cbind(week1,week2)~sex+drug,data=rats) # additive model
summary(Manova(f1)) # gives table bottom p. 354
summary(Manova(f2)) # soak interaction into E=W+I
```

Compare to book pp. 353–355.

```
data rats;
input sex$ drug$ week1 week2 @@;
datalines;
  m a  5  6 m a  5  4 m a  9  9 m a  7  6
  m b  7  6 m b  7  7 m b  9 12 m b  6  8
  m c 21 15 m c 14 11 m c 17 12 m c 12 10
  f a  7 10 f a  6  6 f a  9  7 f a  8 10
  f b 10 13 f b  8  7 f b  7  6 f b  6  9
  f c 16 12 f c 14  9 f c 14  8 f c 10  5
;

proc glm data=rats;
 class sex drug;
 model week1 week2=sex|drug / ss3;
  manova h=sex*drug / printh printe; % mstat=exact gives exact p-values;
  manova h=sex / printh printe;
  manova h=drug / printh printe;
run;
```

SAS gives "characteristic roots" $\gamma_{(j)}$ for no charge. What is that percentage?

Proportion of variability explained

All MANOVA hypothesis tests are based on $\mathbf{E}\mathbf{B}^{-1}$. For example in the two-way model we might look at $\mathbf{E} = \mathbf{W} + \mathcal{I}$ and $\mathbf{B} = \mathbf{R}$ for $H_0 : \alpha_1 = \cdots = \alpha_r = \mathbf{0}$. First note that rank$(\mathbf{E}\mathbf{B}^{-1}) = \min\{\text{rank}(\mathbf{E}), \text{rank}(\mathbf{B}^{-1})\}$, and the ranks of each are equal to the projection matrices that define the matrix.

Consider the one-way model. Then rank$(\mathbf{W}\mathbf{B}^{-1}) \leq \min\{p, k-1\}$ as $\mathbf{W} = \mathbf{X}'[\mathcal{I}_n - \mathbf{P_Z}]\mathbf{X}$ and $\mathbf{B} = \mathbf{X}'[\mathbf{P_Z} - \mathbf{P_{1_n}}]\mathbf{X}$. Write $\mathbf{W}\mathbf{B}^{-1} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}\hat{\mathbf{\Gamma}}'$. Now consider the transformation $\mathbf{Y} = \mathbf{X}\hat{\mathbf{\Gamma}}$. The new data matrix $\mathbf{Y}$ has $\tilde{\mathbf{W}} = \hat{\mathbf{\Gamma}}'\mathbf{W}\hat{\mathbf{\Gamma}}$, $\tilde{\mathbf{B}} = \hat{\mathbf{\Gamma}}'\mathbf{B}\hat{\mathbf{\Gamma}}$, and $\tilde{\mathbf{W}}^{-1}\tilde{\mathbf{B}} = \hat{\mathbf{\Lambda}}$.

The Hotelling-Lawley trace statistic is $\sum_{i=1}^{p} \hat{\lambda}_i$. Then $\frac{\hat{\lambda}_r}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_p}$ is the proportion of this test statistic that is explained by the rth cononical covariate. Furthermore, Roys greatest root $\hat{\lambda}_1$ reduces to an F-test for the first canonical covariate $\mathbf{y}_{(1)}$, $\hat{\lambda}_2$ reduces to an F-test for $\mathbf{y}_{(2)}$, etc.

...NEED TO FIGURE THIS OUT...R GETS THE HYPOTHESIS
MATRIX WRONG WITHOUT AN INTERCEPT. IT GIVES THE
HYPOTHESIS FOR THE MODEL *WITH* AN INTERCEPT
PLUS THE SSCP TOTAL.