

TUTORIAL 4
STA437 WINTER 2015

AL NOSEDAL

CONTENTS

1. Maximization of Quadratic Forms	1
2. One-Way Multivariate Analysis of Variance	2
2.1. Roy's Test	3
2.2. Pillai and Lawley-Hotelling Tests	3
3. Examples	4

1. MAXIMIZATION OF QUADRATIC FORMS

Result. Let $\mathbf{B}_{p \times p}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$. Then

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_1$$

(attained when $\mathbf{x} = \mathbf{e}_1$).

Proof. Let $\mathbf{P}_{p \times p}$ be the orthogonal matrix whose columns are the eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$ and \mathbf{D} be the diagonal matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ along the main diagonal. Let $\mathbf{B}^{1/2} = \mathbf{P} \mathbf{D}^{1/2} \mathbf{P}'$ and $\mathbf{y} = \mathbf{P}' \mathbf{x}$.

Consequently, $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{y} \neq \mathbf{0}$. Thus,

$$\frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\mathbf{x}' \mathbf{B}^{1/2} \mathbf{B}^{1/2} \mathbf{x}}{\mathbf{x}' \mathbf{P} \mathbf{P}' \mathbf{x}} = \frac{\mathbf{x}' \mathbf{P} \mathbf{D}^{1/2} \mathbf{P}' \mathbf{P} \mathbf{D}^{1/2} \mathbf{P}' \mathbf{x}}{\mathbf{y}' \mathbf{y}} = \frac{\mathbf{y}' \mathbf{D} \mathbf{y}}{\mathbf{y}' \mathbf{y}}$$

$$\frac{\mathbf{x}' \mathbf{B} \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^p y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1$$

Setting $\mathbf{x} = \mathbf{e}_1$ gives

$$\frac{\mathbf{e}_1' \mathbf{B} \mathbf{e}_1}{\mathbf{e}_1' \mathbf{e}_1} = \mathbf{e}_1' \mathbf{B} \mathbf{e}_1 = \lambda_1.$$

□

2. ONE-WAY MULTIVARIATE ANALYSIS OF VARIANCE

After being exposed to Hotelling's T^2 , it should come as no surprise to you that Multivariate Analysis of Variance (MANOVA) could be handled by reducing each subject's scores on p variables to a single number -a simple linear combination of his scores on those original p variables. Heuristically, one-way MANOVA consists of a search for that linear combination of the variables which maximally discriminates among the k groups in the sense of producing the largest possible univariate F-ratio, followed by comparison of this largest possible univariate F to a critical value appropriate to such a statistic. Formally, it should be fairly clear from your familiarity with the relationship between the variance of a linear combination of variables and the variances of the component variables that

$$F(\mathbf{a}) = \frac{MSB_V}{MSW_V} = \frac{\sum_{l=1}^g n_l (\bar{v}_l - \bar{v})^2 / (g-1)}{\sum_{l=1}^g n_l \sum_{j=1}^{n_l} (v_{lj} - \bar{v}_l)^2 / (\sum_{l=1}^g n_l - g)} = \frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{W} \mathbf{a}} \left(\frac{N-g}{g-1} \right)$$

where MSB_V is the mean square between groups for variable V ; MSW_V the within-group mean square for dependent variable V ;

$$V_j = a_1 X_{1j} + a_2 X_{2j} + \dots + a_p X_{pj}$$

where $j = 1, 2, \dots, n$ (number of subjects) and $n = n_1 + n_2 + \dots + n_g$.

The task is to choose an \mathbf{a} such that $F(\mathbf{a})$ is maximized. However, maximizing $F(\mathbf{a})$ is equivalent to maximizing $\frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{W} \mathbf{a}}$. Note that

$$\max_{\mathbf{a} \neq \mathbf{0}} \frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{W} \mathbf{a}} = \max_{\mathbf{a} \neq \mathbf{0}} \frac{\mathbf{a}' \mathbf{W}^{1/2} \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \mathbf{W}^{1/2} \mathbf{a}}{\mathbf{a}' \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{a}}$$

Since $\mathbf{W}^{1/2}$ is symmetric, letting $\mathbf{x} = \mathbf{W}^{1/2} \mathbf{a}$ and $\mathbf{B}^* = \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$, we have

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' \mathbf{B}^* \mathbf{x}}{\mathbf{x}' \mathbf{x}}$$

Therefore, using our result for Maximization of Quadratic forms

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}' \mathbf{B}^* \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \lambda_1^*$$

where λ_1^* is the largest eigenvalue of $\mathbf{B}^* = \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$. According to your textbook, the maximum of $F(\mathbf{a})$ corresponds to the largest eigenvalue of $\mathbf{W}^{-1} \mathbf{B}$. We will show that $\mathbf{B}^* = \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$ and $\mathbf{W}^{-1} \mathbf{B}$ have the same eigenvalues.

Definition. Two matrices \mathbf{A} and \mathbf{B} are called similar if there exists an invertible matrix \mathbf{X} such that $\mathbf{A} = \mathbf{X}^{-1} \mathbf{B} \mathbf{X}$.

Theorem. Suppose \mathbf{A} and \mathbf{B} are similar matrices. Then \mathbf{A} and \mathbf{B} have the same characteristic polynomial and hence the same eigenvalues.

Proof.

$$\begin{aligned}
 |\lambda\mathbf{I} - \mathbf{A}| &= |\lambda\mathbf{I} - \mathbf{X}^{-1}\mathbf{B}\mathbf{X}| \\
 &= |\lambda\mathbf{X}^{-1}\mathbf{I}\mathbf{X} - \mathbf{X}^{-1}\mathbf{B}\mathbf{X}| \\
 &= |\mathbf{X}^{-1}(\lambda\mathbf{I} - \mathbf{B})\mathbf{X}| \\
 &= |\mathbf{X}^{-1}||\lambda\mathbf{I} - \mathbf{B}||\mathbf{X}| \\
 &= |\lambda\mathbf{I} - \mathbf{B}|
 \end{aligned}$$

This is the characteristic polynomial of \mathbf{B} , so \mathbf{A} and \mathbf{B} have the same characteristic polynomial. Hence \mathbf{A} and \mathbf{B} have the same eigenvalues. \square

Hence, if we can show that $\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$ and $\mathbf{W}^{-1}\mathbf{B}$ are similar, we are done. Proof.

$$\mathbf{W}^{-1}\mathbf{B} = \mathbf{W}^{-1/2}\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}\mathbf{W}^{1/2}$$

Let $\mathbf{A}^* = \mathbf{W}^{-1}\mathbf{B}$, $\mathbf{X} = \mathbf{W}^{1/2}$, and $\mathbf{B}^* = \mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$. Clearly, \mathbf{A}^* and \mathbf{B}^* are similar. Therefore, they have the same eigenvalues. \square

2.1. Roy's Test. To test $H_0 : \mu_1 = \mu_2 = \dots = \mu_g$ (g number of groups or samples) based on λ_1 , the largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$, we use Roy's union-intersection test, also called Roy's largest root test. The test statistic is given by

$$\theta = \frac{\lambda_1}{1 + \lambda_1}$$

Critical values for θ are given in Table A.10. We reject $H_0 : \mu_1 = \mu_2 = \dots = \mu_g$ if $\theta > \theta_{\alpha, s, m, N}$. The parameters s , m , and N are defined as

$s = \min(\nu_H, p)$, $m = \frac{1}{2}(\nu_H - p) - 1$, $N = \frac{1}{2}(\nu_E - p - 1)$. Where p = number of variables, ν_H = degrees of freedom for hypothesis, ν_E = degrees of freedom for error.

2.2. Pillai and Lawley-Hotelling Tests. There are two additional test statistics for $H_0 : \mu_1 = \mu_2 = \dots = \mu_g$ based on the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_s$ of $\mathbf{W}^{-1}\mathbf{B}$. The Pillai statistic is given by

$$V^{(s)} = \text{tr}[(W + B)^{-1}B] = \sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i}.$$

We reject H_0 for $V^{(s)} \geq V_{\alpha}^{(s)}$. The upper percentage points, $V_{\alpha}^{(s)}$, are given in Table A.11, indexed by s , m , and N , which are defined as in Roy's test.

The Lawley-Hotelling statistic is defined as

$$U^{(s)} = \text{tr}(W^{-1}B) = \sum_{i=1}^s \lambda_i$$

and is also known as Hotelling's generalized T^2 -statistic. Table A.12 gives upper percentage points of the test statistic

$$\frac{\nu_E}{\nu_H} U^{(s)}.$$

We reject H_0 for large values of the test statistic.

3. EXAMPLES

Example 1. Four psychological tests were given to 32 men and 32 women. The data are recorded in file PSYCH.DAT. The variables are:

y_1 = pictorial inconsistencies,
 y_2 = paper from board,
 y_3 = tool recognition,
 y_4 = vocabulary.

The mean vectors are

$$\bar{\mathbf{y}}_1 = \begin{pmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{pmatrix}$$

$$\bar{\mathbf{y}}_2 = \begin{pmatrix} 12.34 \\ 13.91 \\ 16.66 \\ 21.94 \end{pmatrix}$$

The covariance matrices of the two samples are

$$\mathbf{S}_1 = \begin{pmatrix} 5.192 & 4.545 & 6.522 & 5.250 \\ 4.545 & 13.18 & 6.760 & 6.266 \\ 6.522 & 6.760 & 28.67 & 14.47 \\ 5.250 & 6.266 & 14.47 & 16.65 \end{pmatrix}$$

$$\mathbf{S}_2 = \begin{pmatrix} 9.136 & 7.549 & 4.864 & 4.151 \\ 7.549 & 18.60 & 10.22 & 5.446 \\ 4.864 & 10.22 & 30.04 & 13.49 \\ 4.151 & 5.446 & 13.49 & 28.00 \end{pmatrix}$$

Test the hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ at the 0.01 significance level.

```
## "Reading" data
```

```
data<-read.table(file="PSYCH.DAT")
```

```
## breaking down data

males<-data[1:32 , -1]

females<-data[33:64 , -1]

## sample sizes

n1<-dim(males)[1]

n2<-dim(females)[1]

## mean vectors

y.bar.1<-apply(males,2,FUN=mean)

y.bar.2<-apply(females,2,FUN=mean)

## covariance matrices

S.1<-cov(males)

S.2<-cov(females)

Sp<-(n1+n2-2)^(-1)*((n1-1)*S.1 + (n2-1)*S.2)

## Hotelling's T^2

T.2<-(n1*n2/(n1+n2))*t(y.bar.1-y.bar.2)%*%solve(Sp)%*%(y.bar.1-y.bar.2)

## Critical value

p<-dim(males)[2]

crit.val<-((n1+n2-2)*p/(n1+n2-p-1))*qf(0.99,p,n1+n2-p-1)

crit.val

## Again... but using MANOVA.
```

```
groups<-factor(data[,1])

groups

Y<-cbind(data[,2],data[,3],data[,4],data[,5])

Y

fit<-manova(Y~groups)

## showing MANOVA table

summary(fit,test="Roy")

## matrix W and B

sum.roy<-summary(fit,test="Roy")

## largest eigenvalue

lambda.1<-sum.roy$Eigen[1]

## N = total number of individuals

N<-n1 + n2

## g = number of groups or samples

g<-2

## largest univariate F

F.a<-(N-g)*lambda.1/(g-1)
```

Example 2. In a classical experiment carried out from 1918 to 1934, apple trees of different rootstocks were compared. The data for eight trees from each of six rootstocks are given in roots.DAT. The variables are

$$y_1 = \text{trunk girth at 4 years (mm x 100)}$$

$$y_2 = \text{extension growth at 4 years (m)}$$

y_3 = trunk girth at 15 years (mm x 100)

y_4 = weight of tree above ground at 15 years (lb x 1000)

In this case, the mean vectors represent six points in four-dimensional space. Compare the mean vectors for significant differences using all four MANOVA tests (use $\alpha = 0.05$). Compare your results with those shown in your textbook.

```
roots<-read.table(file='roots.DAT')

groups<-factor(roots[,1])

Y2<-cbind(roots[,2],roots[,3],roots[,4],roots[,5])

fit.2<-manova(Y2~groups)

## showing MANOVA table

summary(fit.2,test="Wilks")

## matrix W and B

sum.wilks<-summary(fit.2,test="Wilks")

sum.wilks

## B

sum.wilks$SS[1]

## W

sum.wilks$SS[2]

## Pillai's statistic

summary(fit.2,test="Pillai")

## Lawley-Hotelling statistic
```

```
summary(fit.2,test="Hotelling-Lawley")
```

```
## Roy's test statistic
```

```
summary(fit.2,test="Roy")
```

Example 3. Tests on Individual variables following rejection of H_0 by the overall MANOVA test.

In Example 2, the hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_6$ was rejected for the rootstock data of roots.DAT. We can therefore test the four individual variables using the 0.05 level of significance for a univariate F-test.

```
## Individuals anovas
```

```
anova1<-aov(Y2[ ,1]~groups)
```

```
summary(anova1)
```

```
anova2<-aov(Y2[ ,2]~groups)
```

```
summary(anova2)
```

```
anova3<-aov(Y2[ ,3]~groups)
```

```
summary(anova3)
```

```
anova4<-aov(Y2[ ,4]~groups)
```

```
summary(anova4)
```

Thus for three of the four variables, the six means differ significantly.