Multivariate Analysis Lecture 8: Eigenvalues, Covariance Matrices, and MANOVA

Zhaoxia Yu
Professor, Department of Statistics

2023-04-27

Outline of Lecture 08

- Eigenvalues of covariance matrices
- Examples of 2×2 covariance matrices
- Review of one-way ANOVA
- One-way MANOVA
- A heads up of the midterm project

Section 1

Eigenvalues and Cov Matrices

Subsection 1

Eigenvalues

Eigenvalues and Eigenvectors

- Let A be a $p \times p$ square matrix.
- We say $\nu \in \mathbb{R}^p$ is an eigenvector and $\lambda \in \mathbb{R}$ is the corresponding eigenvalue of A if

$$A\nu = \lambda\nu$$

- The eigenvalues λ of A are roots of the characteristic equation

$$\det(\lambda I - A) = 0$$

The Spectral Decomposition of A Symmetric Matrix

- Let $A_{p\times p}$ be a symmetric matrix

$$A = \Gamma\Lambda\Gamma^T = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i'$$

where

- $\lambda_1, \cdots, \lambda_p$, often ordered from the largest to the smallest, are the eigenvalues of A
- Γ is an orthogonal matrix, i.e., $\Gamma\Gamma^T = \Gamma^T\Gamma = \mathbf{I}$ and the columns of Γ are the eigenvectors of A.
- Λ is the diagonal matrix of the eigenvalues

Subsection 2

Covariance Matrices

A Covariance Matrix Has to be P.(S.)D.

- A covariance matrix has to be either positive definite (p.d.) or positive semidefinite (p.s.d.)
- What is the definition of p.d. or p.s.d.?
  - We say $\mathbf{A}_{p\times p}$ is p.d. (p.s.d.) is $x^T\mathbf{A}x > 0$ (≥ 0) for any $x \in \mathbb{R}^p$ and $x \neq 0$
- Why do covariance matrices has to be p.d. or p.s.d? The following page provides a nice explanation:

https://gowrishankar.info/blog/why-covariance-matrix-should-be-positive-semi-definite-tests-using-breast-cancer-dataset/

A Covariance Matrix Has to be P.(S.)D.

- Intuitively, the information of the pairwise covariance/correlation has to be consistent. Can the following matrix be a covariance matrix?

$$\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0 \\ 0.9 & 0 & 1 \end{pmatrix}$$

- The matrix indicates that
  - Variable 1 and 2 are highly correlated
  - Variable 1 and 3 are highly correlated
  - Variable 2 and 3 are not correlated
- The pairwise correlations do not seem to be consistent

A Covariance Matrix Has to be P.(S.)D.

- Examine the following matrix, which involves a parameter ρ

$$A = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & \rho \\ 0.9 & \rho & 1 \end{pmatrix}$$

- What values can ρ take in order for the matrix to be an appropriate covariance matrix?
- This is a linear algebra problem. In order for the matrix to be a covariance matrix, the eigenvalues should be non-negative.
- Recall that the eigenvalues are the roots to

$$0 = \det(\lambda I - A) = \left| \begin{pmatrix} \lambda - 1 & -0.9 & -0.9 \\ -0.9 & \lambda - 1 & -\rho \\ -0.9 & -\rho & \lambda - 1 \end{pmatrix} \right|$$

Subsection 3

Examples

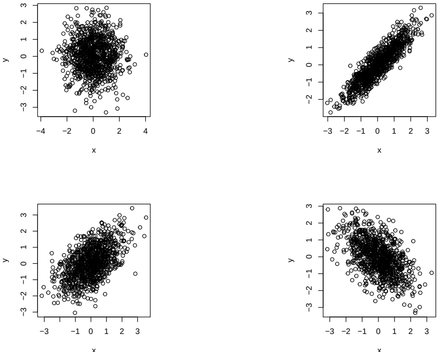
Eigenvalues of Covariance Matrix: Example 1

```
p=2; n=1000; rho1=0; rho2=0.9; rho3=0.5; rho4=-0.5
Sigma1=diag(1-rho1, p, p) + matrix(rho1, p, p)
Sigma2=diag(1-rho2, p, p) + matrix(rho2, p, p)
Sigma3=diag(1-rho3, p, p) + matrix(rho3, p, p)
Sigma4=diag(1-rho4, p, p) + matrix(rho4, p, p)
X1=data.frame(mvrnorm(n, rep(0,p), Sigma1)); names(X1)=c("x","y")
X2=data.frame(mvrnorm(n, rep(0,p), Sigma2)); names(X2)=c("x","y")
X3=data.frame(mvrnorm(n, rep(0,p), Sigma3)); names(X3)=c("x","y")
X4=data.frame(mvrnorm(n, rep(0,p), Sigma4)); names(X4)=c("x","y")
```

Eigenvalues of Covariance Matrix: Example 1

- Simulated data

```
par(mfrow=c(2,2),pty="s")
plot(X1);plot(X2);plot(X3);plot(X4)
```



Eigenvalues of Covariance Matrix: Example 1

- The true covariance matrices

```
Sigma1
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

```
Sigma2
```

```
##      [,1] [,2]
## [1,]  1.0  0.9
## [2,]  0.9  1.0
```

```
Sigma3
```

```
##      [,1] [,2]
## [1,]  1.0  0.5
## [2,]  0.5  1.0
```

```
Sigma4
```

```
##      [,1] [,2]
## [1,]  1.0 -0.5
## [2,] -0.5  1.0
```

Eigenvalues of Covariance Matrix: Example 1

- Eigenvalues of the true covariance matrices

```
eigen(Sigma1)$values
```

```
## [1] 1 1
```

```
eigen(Sigma2)$values
```

```
## [1] 1.9 0.1
```

```
eigen(Sigma3)$values
```

```
## [1] 1.5 0.5
```

```
eigen(Sigma4)$values
```

```
## [1] 1.5 0.5
```

Eigenvalues of Covariance Matrix: Example 1

- Eigenvalues of the estimated covariance matrices

```
eigen(cov(X1))$values
```

```
## [1] 1.149824 1.002009
```

```
eigen(cov(X2))$values
```

```
## [1] 2.0081797 0.1108885
```

```
eigen(cov(X3))$values
```

```
## [1] 1.5843622 0.4613252
```

```
eigen(cov(X4))$values
```

```
## [1] 1.5988678 0.4885592
```

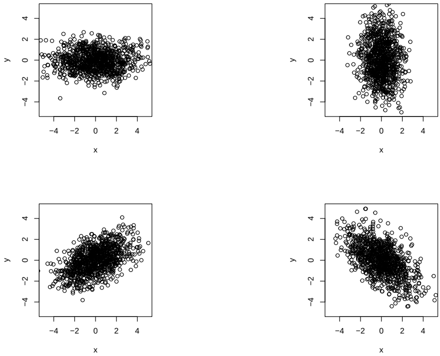
Eigenvalues of Covariance Matrix: Example 2

```
n=1000
Sigma1=diag(c(4,1), 2, 2)
Sigma2=diag(c(1,4), 2, 2)
theta=pi/6
R1=matrix(c(cos(theta), sin(theta), -sin(theta), cos(theta)), 2,2)
theta=pi/4+pi/2
R2=matrix(c(cos(theta), sin(theta), -sin(theta), cos(theta)), 2,2)
Sigma3=R1%*%Sigma1%*%t(R1)
Sigma4=R2%*%Sigma1%*%t(R2)
X1=data.frame(mvrnorm(n, rep(0,2), Sigma1)); names(X1)=c("x","y")
X2=data.frame(mvrnorm(n, rep(0,2), Sigma2)); names(X2)=c("x","y")
X3=data.frame(mvrnorm(n, rep(0,2), Sigma3)); names(X3)=c("x","y")
X4=data.frame(mvrnorm(n, rep(0,2), Sigma4)); names(X4)=c("x","y")
```

Eigenvalues of Covariance Matrix: Example 2

- Simulated data

```
par(mfrow=c(2,2),pty="s")
plot(X1, xlim=c(-5,5), ylim=c(-5,5));plot(X2, xlim=c(-5,5), ylim=c(-5,5));
plot(X3, xlim=c(-5,5), ylim=c(-5,5));plot(X4, xlim=c(-5,5), ylim=c(-5,5))
```



Eigenvalues of Covariance Matrix: Example 2

- The true covariance matrices

```
Sigma1
```

```
##      [,1] [,2]
## [1,]    4    0
## [2,]    0    1
```

```
Sigma2
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    4
```

```
Sigma3
```

```
##          [,1]     [,2]
## [1,] 3.250000 1.299038
## [2,] 1.299038 1.750000
```

```
Sigma4
```

```
##      [,1] [,2]
## [1,]  2.5 -1.5
## [2,] -1.5  2.5
```

Eigenvalues of Covariance Matrix: Example 2

- Eigenvalues of the true covariance matrices

```
eigen(Sigma1)$values
```

```
## [1] 4 1
```

```
eigen(Sigma2)$values
```

```
## [1] 4 1
```

```
eigen(Sigma3)$values
```

```
## [1] 4 1
```

```
eigen(Sigma4)$values
```

```
## [1] 4 1
```

Eigenvalues of Covariance Matrix: Example 2

- Eigenvalues of the estimate covariance matrices

```
eigen(cov(X1))$values
```

```
## [1] 3.8339388 0.9233912
```

```
eigen(cov(X2))$values
```

```
## [1] 3.8055355 0.9921622
```

```
eigen(cov(X3))$values
```

```
## [1] 3.5480606 0.9413241
```

```
eigen(cov(X4))$values
```

```
## [1] 3.937493 1.101251
```

Section 2

One-way ANOVA

One-way ANOVA: Notations and Assumptions

- g independent samples
  - $Y_{11}, \cdots Y_{1,n_1} \overset{iid}{\sim} N(\mu_1, \sigma^2)$
  - $Y_{21}, \cdots Y_{2,n_2} \overset{iid}{\sim} N(\mu_2, \sigma^2)$
  - $\cdots\cdots$
  - $Y_{g1}, \cdots Y_{g,n_g} \overset{iid}{\sim} N(\mu_g, \sigma^2)$
- Total sample size $n = n_1 + \cdots + n_g = \sum_{i=1}^{g} n_i$
- Group means: $\bar{Y}_{i.}$ for $i = 1, \cdots, g$
- Grand/overall mean: $\bar{Y}_{..}$

One-Way ANOVA: Partition the Sum of Squares of Total:

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{..}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i}[(Y_{ij}-\bar{Y}_{i.})+(\bar{Y}_{i.}-\bar{Y}_{..})]^2 \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2+\sum_{i=1}^{g}n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2+ \\
&\quad 2\sum_{i=1}^{g}\sum_{j=1}^{n_i}[(Y_{ij}-\bar{Y}_{i.})(\bar{Y}_{i.}-\bar{Y}_{..})] \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2+\sum_{i=1}^{g}n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2+ \\
&\quad 2\sum_{i=1}^{g}[(\bar{Y}_{i.}-\bar{Y}_{..})\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})] \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2+\sum_{i=1}^{g}n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2+2\sum_{i=1}^{g}[(\bar{Y}_{i.}-\bar{Y}_{..})\cdot 0] \\
&= \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}[(Y_{ij}-\bar{Y}_{i.})^2}_{SSE}+\underbrace{\sum_{i=1}^{g}n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2}_{SSTR}
\end{aligned}
$$

One-Way ANOVA: E(SSE)

- $SSTO = SSE + SSTR = SSW + SSB$:
- SSE is also known as SSW, the within-group variance

$$SSE = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^{g} (n_i - 1) s_i^2$$

where s_i^2 is the sample variance for the ith group. Recall that $E[s_i^2] = \sigma^2$. Therefore,

$$E[SSE] = \sum_{i=1}^{g} (n_i - 1)\sigma^2 = (n - g)\sigma^2$$

One-Way ANOVA: E(SSTR)

- *SSTR* is also known as *SSB*, the between-group variance
- The calculation of $E(SSTR)$ requires the following results

$$
\begin{aligned}
E[\bar{Y}_{i.}^2] &= Var[\bar{Y}_{i.}] + \mu_i^2 = \frac{\sigma^2}{n_i} + \mu_i^2 \\
E[\bar{Y}_{..}^2] &= Var[\bar{Y}_{..}] + \mu_{..}^2 = \frac{\sigma^2}{n} + \mu_{..}^2 \\
E[\bar{Y}_{i.}\bar{Y}_{..}] &= \frac{1}{n}E[\bar{Y}_{i.}\sum_{j=1}^{g} n_j \bar{Y}_{j.}] = \frac{1}{n}E[\bar{Y}_{i.}(n_i\bar{Y}_{i.} + \sum_{j\neq i}^{g} n_j \bar{Y}_{j.})] \\
&= \frac{n_i}{n}E[\bar{Y}_{i.}^2] + \mu_i \sum_{j\neq i} \frac{n_j}{n}\mu_j = \frac{1}{n}\sigma^2 + \frac{n_i}{n}\mu_i^2 + \mu_i \sum_{j\neq i} \frac{n_j}{n}\mu_j \\
&= \frac{1}{n}\sigma^2 + \mu_i \bar{\mu}_{..}
\end{aligned}
$$

One-Way ANOVA: E(SSTR) (continued)

$$\begin{aligned}
E[SSTR] &= \sum_{i=1}^{g} n_i E[(\bar{Y}_{i.} - \bar{Y}_{..})^2] \\
&= \sum_{i=1}^{g} n_i E[\bar{Y}_{i.}^2 + \bar{Y}_{..}^2 - 2\bar{Y}_{i.}\bar{Y}_{..}] \\
&= \sum_{i=1}^{g} n_i E[\bar{Y}_{i.}^2 + \bar{Y}_{..}^2 - 2\bar{Y}_{i.}\bar{Y}_{..}] \\
&= \sum_{i=1}^{g} n_i (\frac{\sigma^2}{n_i} + \mu_i^2 + \frac{\sigma^2}{n} + \mu_{..}^2 - 2[\frac{1}{n}\sigma^2 + \mu_i \bar{\mu}_{..}]) \\
&= (g-1)\sigma^2 + \sum_{i=1}^{g} n_i (\mu_i - \bar{u}_{..})^2
\end{aligned}$$

where $\bar{\mu}_{..} = \frac{1}{n}\sum_{i=1}^{g} n_i \mu_i$.

One-Way ANOVA: Mean of Sum of Squares

- The null hypothesis

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_g$$

- The alternative hypothesis

$$H_a\colon \mu_i \neq \mu_j \text{ for at least one pair of } (i, j)$$

- Mean sum of squares: $MSE = \frac{1}{n-g} SSE$, $MSTR = \frac{1}{g-1} SSTR$
- $E(MSE) = \sigma^2$
- $E(MSTR) = \sigma^2 + \frac{1}{g-1} \sum_{i=1}^{g} n_i(\mu_i - \bar{u}_{..})^2 \overset{H_0}{=} \sigma^2$
- Thus, a reasonable statistic is the ratio of the two mean sum of squares

One-Way ANOVA: F-statistic

- The F-statistic is defined as

$$F = \frac{MSTR}{MSE}$$

- The null distribution

$$F = \frac{\frac{SSTR}{\sigma^2}/(g-1)}{\frac{SSE}{\sigma^2}/(n-g)} \overset{H_0}{\sim} F_{g-1,n-g}$$

- To derive the null distribution of F, we need to show that
  - $\frac{SSTR}{\sigma^2} \overset{H_0}{\sim} \chi^2_{g-1}$
  - $\frac{SSE}{\sigma^2} \sim \chi^2_{n-g}$
  - $SSTR$ and SSE are independent

ANOVA Table and Distributions

Source	SS	MS	DF	F
Treatment	$SSTR = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MSTR = \frac{SSTR}{g-1}$	g-1	$F = \frac{MSTR}{MSE}$
Error	$SSE = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$	$MSE = \frac{SSE}{n-g}$	n-g	
Total	$SSTO = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$			

Section 3

MANOVA Sample Cov

ANOVA to MANOVA

- So far we have learned
    - One-sample Hotelling's T^2
    - Two-sample Hotelling's T^2
- The next logical extension is to multiple samples, i.e., the multivariate version anova, or MANOVA
- Compared to one-sample or two-sample multivariate analysis, there are many choices for comparing multiple samples for multivariate data
- We will cover the following methods:

Notations and Assumptions

- g independent random samples
  - $\mathbf{Y}_{11}, \cdots \mathbf{Y}_{1,n_1} \overset{iid}{\sim} N(\boldsymbol{\mu}_1, \mathbf{\Sigma})$
  - $\mathbf{Y}_{21}, \cdots \mathbf{Y}_{2,n_2} \overset{iid}{\sim} N(\boldsymbol{\mu}_2, \mathbf{\Sigma})$
  - $\cdots\cdots$
  - $\mathbf{Y}_{g1}, \cdots \mathbf{Y}_{g,n_g} \overset{iid}{\sim} N(\boldsymbol{\mu}_g, \mathbf{\Sigma})$
- n_i: the number of observations in group i
- The ith random sample is from $N(\boldsymbol{\mu}_i, \mathbf{\Sigma})$

Notations and Assumptions (continued)

- Each $\mathbf{Y}_{ij} \in \mathbb{R}^p$, i.e,

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijp} \end{pmatrix}$$

- The null hypothesis

$$H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g$$

- The alternative hypothesis

$$H_1\colon \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \text{ for at least one pair of } (i, j)$$

The Total Covariance Matrix

- The total covariance matrix is the covariance matrix if the group information is ignored. If we pool all the $n = n_1 + \cdots + n_g = \sum_{i=1}^{g} n_g$ observations together, what is the covariance matrix?

$$
\begin{aligned}
\mathbf{T} &= \textstyle\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})' \\
&= \textstyle\sum_{i=1}^{g}\sum_{j=1}^{n_i}\{(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i}) + (\bar{\mathbf{Y}}_{i} - \bar{\mathbf{Y}}_{..})\}\{(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i}) + (\bar{\mathbf{Y}}_{i} - \bar{\mathbf{Y}}_{..})\}' \\
&= \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})'}_{W} + \underbrace{\sum_{i=1}^{g} n_i(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})'}_{B} + 0
\end{aligned}
$$

The Within-Group Sample Covariance Matrix

- In the definition of the total covariance matrix, we compare each observation to the grand mean vector
- In within-group covariance matrix, we compare each observation to its group mean

$$\mathbf{W} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})'$$

The Within-Group Sample Covariance Matrix (continued)

- Let $\mathbf{W}_i = \sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})'$. We can show that

$$\mathbf{W}_1, \cdots, \mathbf{W}_g \overset{independent}{\sim} Wishart_p(n_i - 1, \mathbf{\Sigma})$$

- Recall that the sum of independently distributed chi-squared distributed random variables also follows a chi-squared distribution. Similarly,

$$\mathbf{W} = \sum_{\mathbf{i=1}}^{\mathbf{g}} \mathbf{W_i} \sim \mathbf{Wishart_p(n - g, \Sigma)}$$

The Between-Group Sample Covariance Matrix

- The between-group sample covariance matrix captures the difference in mean vectors between groups

$$\mathbf{B} = \sum_{i=1}^{g} n_i(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})'$$

- When the null hypothesis is true, **B** follows a Wishart distribution

$$\mathbf{B} \overset{H_0}{\sim} Wishart_p(g-1, \mathbf{\Sigma})$$

Outline of Proof

- Let $\mathbf{Y}_{n\times p}$ denote the data matrix
- Let $\mathbf{X}_{n\times g}$ denote the design matrix, which consists of dummy variables of the group membership. We also define $\mathbf{P_{n\times n}} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$. Then

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{n_g} & \mathbf{0}_{n_g} & \cdots & \mathbf{1}_{n_g} \end{pmatrix}, \mathbf{P} = \begin{pmatrix} \frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{n_g}\mathbf{1}_{n_g}\mathbf{1}_{n_g}^T \end{pmatrix}$$

Outline of Proof (continued)

- It can be shown that

$$\mathbf{W} = \mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y} \ , \ \mathbf{B} = \mathbf{Y}^T(\mathbf{P} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{Y}$$

- It is not difficult to verify that $\mathbf{P}$, $\mathbf{I} - \mathbf{P}$, and $\mathbf{P} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ are all projection matrices. For a projection matrix, its rank equals its trace. Thus, it is not difficult to show that

$$rank(\mathbf{P}) = g, rank(\mathbf{I} - \mathbf{P}) = n - g, rank(\mathbf{P} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T) = g - 1$$

- In Lecture 06 we showed that $(n-1)\mathbf{S} \sim Wishart_p(n-1, \mathbf{\Sigma})$. Use similar methods, we can show that

$$\mathbf{W} \sim Wishart_p(n - g, \mathbf{\Sigma}) \ , \ \mathbf{B} \overset{H_0}{\sim} Wishart_p(g - 1, \mathbf{\Sigma})$$
$$\mathbf{W} \perp \mathbf{B}$$

MANOVA Table

- We can also construct a table

Source	Sample Cov	DF
Treatment	$\mathbf{B} = \sum_{i=1}^{g} n_i(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.} - \bar{\mathbf{Y}}_{..})'$	g-1
Error	$\mathbf{W} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i.})'$	n-g
Total	$\mathbf{T} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})'$	

Section 4

MANOVA Test Statistics

MANOVA Test Statistics

- In one-way ANOVA, we understand that SSB should be large relative to SSW when the null hypothesis is not true
- In the multivariate version, we also expect that $\mathbf{B}$ should be large relative to $\mathbf{W}$ when the null hypothesis is not true
- However, $\mathbf{B}$ and $\mathbf{W}$ are matrices. How to define “large”?

Iris Data: B and W

```
#rearrange the data such as the response matrix is
#an n-by-p matrix
Y=cbind(SepalL=c(iris3[,1,1],iris3[,1,2],iris3[,1,3]),
SepalW=c(iris3[,2,1],iris3[,2,2],iris3[,2,3]),
PetalL=c(iris3[,3,1],iris3[,3,2],iris3[,3,3]),
PetalW=c(iris3[,4,1],iris3[,4,2],iris3[,4,3]))
#for unknown reasons, data.frame won't work but cbind works
#alternatively, we can use the following way to define y
#Y=aperm(iris3,c(1,3,2));dim(y)=c(150,4)

#define the covariate variable X, which is vector of labels
iris.type=rep(c("Setosa","Versicolor","Virginica"),each=50)
T=(150-1)*cov(Y)
W=(50-1)*cov(iris3[,,1]) +(50-1)*cov(iris3[,,2])+(50-1)*cov(iris3[,,3])
B=T-W
```

Iris Data: B and W

```
B
```

```
##           SepalL    SepalW   PetalL    PetalW
## SepalL  63.21213 -19.95267 165.2484  71.27933
## SepalW -19.95267  11.34493 -57.2396 -22.93267
## PetalL 165.24840 -57.23960 437.1028 186.77400
## PetalW  71.27933 -22.93267 186.7740  80.41333
```

```
W
```

```
##          Sepal L. Sepal W. Petal L. Petal W.
## Sepal L.  38.9562  13.6300  24.6246   5.6450
## Sepal W.  13.6300  16.9620   8.1208   4.8084
## Petal L.  24.6246   8.1208  27.2226   6.2718
## Petal W.   5.6450   4.8084   6.2718   6.1566
```

MANOVA in R

```
summary.manova {stats}  R Documentation
Summary Method for Multivariate Analysis of Variance
Description
A summary method for class "manova".

Usage
## S3 method for class 'manova'
summary(object,
        test = c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"),
        intercept = FALSE, tol = 1e-7, ...)
```

MANOVA - Method 1: Pallai Trace

$$V = trace(\mathbf{B}(\mathbf{B} + \mathbf{W})^{-\mathbf{1}}) = trace(\mathbf{B}\mathbf{T}^{-1})$$

MANOVA - Method 2: Wilk's Lambda

- Wilk's Lambda distribution Let $\mathbf{A} \sim Wishart_p(m_1, \mathbf{I})$ and $\mathbf{B} \sim Wishart_p(m_2, \mathbf{I})$ be independent with $m_1 > p$. We say

$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A}+\mathbf{B}|} \sim \Lambda(p, m_1, m_2)$$

- Test Statistic

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\boldsymbol{\Sigma}^{-1/2}\mathbf{W}\boldsymbol{\Sigma}^{-1/2}|}{|\boldsymbol{\Sigma}^{-1/2}(\mathbf{B}+\mathbf{W})\boldsymbol{\Sigma}^{-1/2}|}$$

- By the definition of Wilk's Lambda distribution,

$$\Lambda^* \overset{H_0}{\sim} \Lambda(p, n-g, g-1)$$

MANOVA - Method 3: Lawley-Hotelling Trace

$$T_0^2 = trace(\mathbf{B}\mathbf{W}^{-1})$$

MANOVA - Method 4: Roy's Largest Root

- Two equivalent test statistics have been used

$$\lambda_{max}(\mathbf{B}\mathbf{W}^{-1})$$
$$\lambda_{max}(\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1})$$

Test Statistics and the Eigenvalues of $\mathbf{BW}^{-1}$

- One interesting observation is that all the test statistics can be expressed in terms of eigenvalues of $\mathbf{BW}^{-1}$. Let $\lambda_1, \cdots, \lambda_p$ denote the eigenvalues, from the largest to the smallest. We have
- Pillai trace

$$tr(\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}) = \sum_{i=1}^{min(p,g-1)} \frac{\lambda_i}{1+\lambda_i}$$

- Wilk's Lambda

$$\Lambda^* = \prod_{i=1}^{min(p,g-1)} \frac{1}{1+\lambda_i}$$

Test Statistics and the Eigenvalues of $\mathbf{BW}^{-1}$

- Lawley-Hotelling trace

$$trace(\mathbf{BW}^{-1}) = \sum_{i=1}^{min(p,g-1)} \lambda_i$$

- Roy's largest root

$$\lambda_{max}(\mathbf{BW}^{-1}) = \lambda_1$$
$$\lambda_{max}(\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}) = \frac{\lambda_1}{1+\lambda_1}$$

Proof

- If you are curious about how to prove these results, I provide an example.

$$
\begin{aligned}
(\Lambda^*)^{-1} &= \frac{|\mathbf{B}+\mathbf{W}|}{|\mathbf{W}|} = |\mathbf{I}+\mathbf{W}^{-1}\mathbf{B}| \\
&= \prod_{i=1}^{p}[\text{the ith eigenvalue of } \mathbf{I}+\mathbf{W}^{-1}\mathbf{B}] \\
&= \prod_{i=1}^{p}[1+\text{the ith eigenvalue of } \mathbf{W}^{-1}\mathbf{B}] \\
&= \prod_{i=1}^{p}(1+\lambda_i) \overset{1}{=} \prod_{i=1}^{min(p,g-1)}(1+\lambda_i)
\end{aligned}
$$

As a result, $\Lambda^* = \prod_{i=1}^{p} \frac{1}{1+\lambda_i}$

Proof (continued)

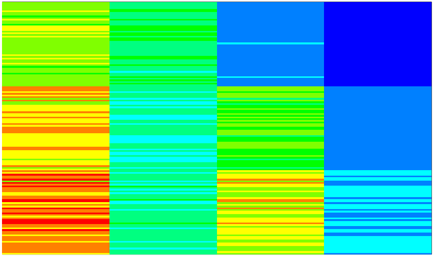
- The last step implies that $rank(\mathbf{W}^{-1}\mathbf{B}) = min(p, g-1)$
- Why is it true?
  - First, $rank(\mathbf{W}^{-1}\mathbf{B}) = rank(\mathbf{B})$
  - Second, because $\mathbf{B} \overset{H_0}{\sim} Wishart_p(g-1, \mathbf{\Sigma})$, there exists $\mathbf{Y}_{(g-1)\times p}$ such that

    $$\mathbf{B} = \mathbf{Y}^T\mathbf{Y} \ ,$$

    and $\mathbf{Y}$ is a random sample of size $g-1$ from $N(\mathbf{0}, \mathbf{\Sigma})$. The rank of $\mathbf{Y}$ is $rank(\mathbf{Y}) = min(p, g-1)$.
  - Third, $rank(\mathbf{B}) = rank(\mathbf{Y}^T\mathbf{Y}) = rank(\mathbf{Y}) = min(p, g-1)$.

Example - Iris Data

```
mycolors=rainbow(12)[9:1]
image(t(iris[150:1, 1:4]),col = mycolors, xaxt="n", yaxt="n")
```



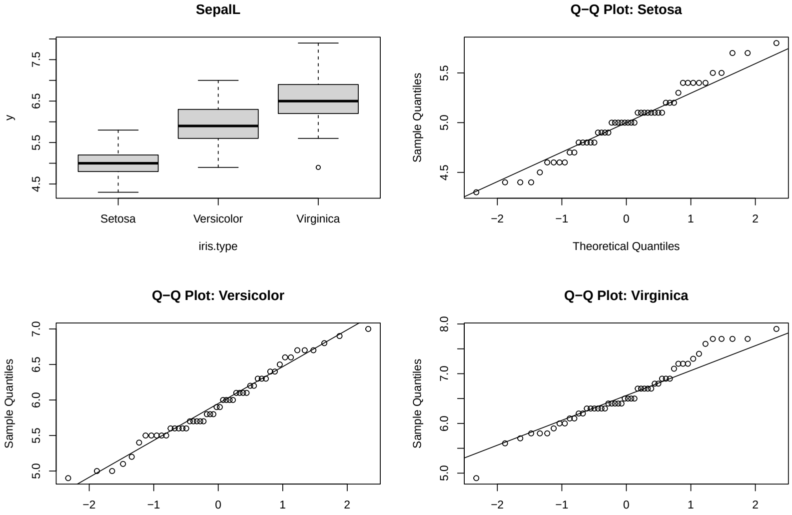
Iris Data: Univariate One-way ANOVA: Data Formatting

```
#rearrange the data in the (X,Y) format
y=c(iris3[,1,1], iris3[,1,2], iris3[,1,3])
#alternatively, you may use:
#y=aperm(iris3[,1,], c(1,2)); dim(y)=c(150,1)

#define the covariate variable X,
#which is vector of labels
iris.type=rep(c("Setosa","Versicolor","Virginica"),each=50)
```

Iris Data: Univariate One-way ANOVA: Exploratory

```
#visual checking
par(mfrow=c(2,2))
#box plot
boxplot(y~iris.type, main="SepalL")
#alternatively, you may use: boxplot(iris3[,1,],
#main="SepalL")

#qq plots
qqnorm(iris3[,1,1], main="Q-Q Plot: Setosa");
qqline(iris3[,1,1])
qqnorm(iris3[,1,2], main="Q-Q Plot: Versicolor");
qqline(iris3[,1,2])
qqnorm(iris3[,1,3], main="Q-Q Plot: Virginica");
qqline(iris3[,1,3])
```

Iris Data: Univariate One-way ANOVA: Exploratory



Iris Data: Univariate One-way ANOVA: Analysis

```
obj.aov=aov(y~iris.type)
summary(obj.aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## iris.type     2  63.21  31.606   119.3 <2e-16 ***
## Residuals   147  38.96   0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

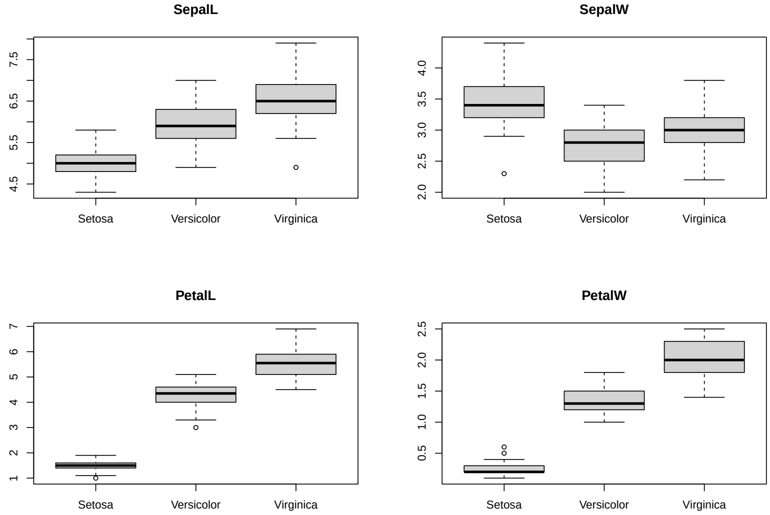
Iris Data: MANOVA: Data Formatting

```
#rearrange the data such as the response matrix is
#an n-by-p matrix
Y=cbind(SepalL=c(iris3[,1,1],iris3[,1,2],iris3[,1,3]),
SepalW=c(iris3[,2,1],iris3[,2,2],iris3[,2,3]),
PetalL=c(iris3[,3,1],iris3[,3,2],iris3[,3,3]),
PetalW=c(iris3[,4,1],iris3[,4,2],iris3[,4,3]))
#for unknown reasons, data.frame won't work but cbind works
#alternatively, we can use the following way to define y
#Y=aperm(iris3,c(1,3,2));dim(y)=c(150,4)

#define the covariate variable X, which is vector of labels
iris.type=rep(c("Setosa","Versicolor","Virginica"),each=50)
```

Iris Data: MANOVA: Exploratory

```
#visual investigation
par(mfrow=c(2,2))
boxplot(iris3[,1,],main="SepalL")
boxplot(iris3[,2,],main="SepalW")
boxplot(iris3[,3,],main="PetalL")
boxplot(iris3[,4,],main="PetalW")
```

Iris Data: MANOVA: Exploratory



Conducting MANOVA "Manually"

```
T=(150-1)*cov(Y)
W=(50-1)*cov(iris3[,,1]) +(50-1)*cov(iris3[,,2])+(50-1)*cov(iris3[,,3])
B=T-W
Lambda=prod(1/(1+ eigen(B%*%solve(W))$values))
(150-3-2)/3*(1-sqrt(Lambda))/sqrt(Lambda)
```

```
## [1] 267.3711
```

```
# Using relationship between Wilk's lambda and F-distribution
# (see wikipedia about "Wilks's lambda distribution")
1-pf((150-3-2)/3*(1-sqrt(Lambda))/sqrt(Lambda), 2*3, 150-3-2)
```

```
## [1] 0
```

Conducting MANOVA using "manova" in R

```
obj=manova(Y~iris.type)
obj.aov
```

```
## Call:
##    aov(formula = y ~ iris.type)
##
## Terms:
##                  iris.type Residuals
## Sum of Squares    63.21213  38.95620
## Deg. of Freedom          2       147
##
## Residual standard error: 0.5147894
## Estimated effects may be unbalanced
```

Conducting MANOVA using "manova" in R

```
summary(obj, test="Pillai")
```

```
##            Df Pillai approx F num Df den Df    Pr(>F)
## iris.type   2 1.1919   53.466      8    290 < 2.2e-16 ***
## Residuals 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(obj, test="Wilks")
```

```
##            Df    Wilks approx F num Df den Df    Pr(>F)
## iris.type   2 0.023439   199.15      8    288 < 2.2e-16 ***
## Residuals 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conducting MANOVA using “manova” in R

```r
summary(obj, test="Hotelling-Lawley")
```

```
##            Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## iris.type   2           32.477   580.53      8    286 < 2.2e-16 ***
## Residuals 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(obj, test="Roy")
```

```
##            Df    Roy approx F num Df den Df    Pr(>F)
## iris.type   2 32.192     1167      4    145 < 2.2e-16 ***
## Residuals 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Too Many Choices?

- Due to the nature of multivariate analysis, we have seen many choices for conducting one-way MANOVA
- Do they work equally well?
- Does their performance depend on the true distribution?
- You will be asked to compare the methods in your midterm project