

# Comparisons of Several Multivariate Populations

## Edps/Soc 584 and Psych 594 Applied Multivariate Statistics

Carolyn J. Anderson

Department of Educational Psychology



# Outline

## ● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

- 1–way ANOVA
  - ◆ Classic Treatment
  - ◆ As a general linear model
- 1–way MANOVA
  - ◆ The Model: Generalization of ANOVA to multivariate.
  - ◆ Hypothesis Testing
  - ◆ Example 1: Massed vs distributed practice
- Multivariate General Linear Model and Example 2: Increased survival
- Following up to a significant result
  - ◆ Multivariate contrasts
  - ◆ Simultaneous confidence intervals
  - ◆ Discriminant function
- Summary of PCA, MANOVA, DA
- SAS IML and PROC GLM.

Reading: Johnson & Wichern pages 296–323.

# Generalizing 1-way ANOVA to Multivariate Data

and Generalizing multivariate  $T^2$  to more than two populations. Suppose that we have random samples from  $g$  populations and measures on  $p$  variables:

$$\begin{array}{ll} \text{Population 1:} & \mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1} \\ \text{Population 2:} & \mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2} \\ & \vdots \\ \text{Population } g: & \mathbf{x}_{g1}, \mathbf{x}_{g2}, \dots, \mathbf{x}_{gn_g} \end{array}$$

where each  $\mathbf{x}_{lj}$  is a  $(p \times 1)$  vector.

Examples:

- Are 5 standardized tests scores the same for high school students who attend different high school programs (i.e., general, vo/tech, academic).
- Are survival times measured in two ways different between those treated with supplemental vitamin C the over six types of cancer?

## ● Outline

Introduction to MANOVA

● Generalizing 1-way ANOVA to Multivariate Data

● Basic Assumptions

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# Basic Assumptions

These are needed for Statistical Inference.

1.  $\mathbf{X}_{l1}, \mathbf{X}_{l2}, \dots, \mathbf{X}_{ln_l}$  is a random sample of size  $n_l$  from a population with means  $\mu_l$  for  $l = 1, \dots, g$  (i.e., **observations within populations are independent** and representative of their populations).
2. Random samples from **different populations are independent**.
3. All populations have the **same covariance matrix,  $\Sigma$** .
4.  $\mathbf{X}_{lj} \sim \mathcal{N}(\mu_l, \Sigma)$ ; that is, **each population is multivariate normal**.

If a population is **not** multivariate normal, then for large  $n_l$  central limit theorem may “kick-in”.

- Outline

- Introduction to MANOVA

- Generalizing 1-way ANOVA to Multivariate Data

- Basic Assumptions

- One-way ANOVA

- GLM

- One-Way MANOVA

- Hypothesis Testing

- Example: Distributed vs Massed Practice

- Following up a Significant Result

- Multivariate GLM

- Simultaneous Confidence Intervals

- Summary: PCA, MANOVA, DA

- SAS IML and GLM

# One-way ANOVA Review

The univariate case where  $p = 1$

**Assumptions:**  $X_{lj} \sim \mathcal{N}(\mu_l, \sigma^2)$  *i.i.d* for  $j = 1, \dots, n_l$  and  $l = 1, \dots, g$ .

**Hypotheses:**

$$H_o : \mu_1 = \mu_2 = \dots = \mu_g \quad \text{versus} \quad H_a : \text{not } H_o$$

We usually express  $\mu_l$  as the sum of a grand mean and deviations from the grand mean

$$\begin{aligned} \underbrace{\mu_l}_{l^{th} \text{ pop. mean}} &= \underbrace{\mu}_{\text{grand mean}} + \underbrace{\mu_l - \mu}_{l^{th} \text{ pop. treatment effect}} \\ &= \mu + \tau_l \end{aligned}$$

If  $\mu_1 = \mu_2 = \dots = \mu_g$ , then an equivalent way to write the null hypothesis is

$$H_o : \tau_1 = \tau_2 = \dots = \tau_g = 0$$

● Outline

Introduction to MANOVA

One-way ANOVA

- One-way ANOVA Review
- The Model for an Observation
- The Sums of Squares
- Sums of Squared Decomposition & Geometry
- ANOVA Summary Table

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# The Model for an Observation

$$X_{lj} = \mu + \tau_l + \epsilon_{lj}$$

where  $\epsilon_{lj} \sim \mathcal{N}(0, \sigma^2)$  and independent.

$\epsilon_{lj}$  is “random error”.

We typically impose the condition  $\sum_{l=1}^g \tau_l = 0$  as an **identification constraint**.

The decomposition of an observation is

$$\underbrace{X_{lj}}_{\text{observation}} = \underbrace{\bar{X}}_{\text{overall sample mean}} + \underbrace{(\bar{X}_l - \bar{X})}_{\text{estimated treatment effect}} + \underbrace{(X_{lj} - \bar{X}_l)}_{\text{residual “error”}}$$

$\bar{X}$  is the estimator of  $\mu$

$\hat{\tau}_l = (\bar{X}_l - \bar{X})$  is the estimator of  $\tau_l$

$(X_{lj} - \bar{X}_l)$  is the estimator of  $\epsilon_{lj}$ .

- Outline

- Introduction to MANOVA

- One-way ANOVA

- One-way ANOVA Review
- The Model for an Observation
- The Sums of Squares
- Sums of Squared
- Decomposition & Geometry
- ANOVA Summary Table

- GLM

- One-Way MANOVA

- Hypothesis Testing

- Example: Distributed vs Massed Practice

- Following up a Significant Result

- Multivariate GLM

- Simultaneous Confidence Intervals

- Summary: PCA, MANOVA, DA

- SAS IML and GLM

# The Sums of Squares

The sum of squared observations

$$SS_{obs} = SS_{total} = \sum_{l=1}^g \sum_{j=1}^{n_l} X_{lj}^2$$

We also take the three components of  $X_{lj}$  and form sums of squares

$$SS_{mean} = \sum_{l=1}^g \sum_{j=1}^{n_l} \bar{X}^2 = \left( \sum_{l=1}^g n_l \right) \bar{X}^2$$

$$SS_{treatment} = \sum_{l=1}^g \sum_{j=1}^{n_l} \hat{\tau}_l^2 = \sum_{l=1}^g \sum_{j=1}^{n_l} (\bar{X}_l - \bar{X})^2 = \sum_{l=1}^g n_l (\bar{X}_l - \bar{X})^2$$

$$SS_{residual} = \sum_{l=1}^g \sum_{j=1}^{n_l} \hat{\epsilon}_{lj}^2 = \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X}_l)^2$$

● Outline

Introduction to MANOVA

One-way ANOVA

- One-way ANOVA Review
- The Model for an Observation
- The Sums of Squares
- Sums of Squared Decomposition & Geometry
- ANOVA Summary Table

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# Sums of Squared Decomposition & Geometry

$$SS_{obs} = SS_{mean} + SS_{tr} + SS_{res}$$

or

$$SS_{corrected} = SS_{obs} - SS_{mean} = SS_{tr} + SS_{res}$$

This works because the components (sums of squares) are orthogonal.

**Geometry:** Consider the  $n = (\sum_{l=1}^g n_l)$  dimensional **observation space** where each observation defines a dimension.

- We break this space into three orthogonal sub-spaces corresponding to each component.
- The dimensionality of the sub-space corresponds to the degrees of freedom for the corresponding  $SS$ .

(see text for more details).

- Outline

Introduction to MANOVA

One-way ANOVA

- One-way ANOVA Review
- The Model for an Observation
- The Sums of Squares
- Sums of Squared Decomposition & Geometry
- ANOVA Summary Table

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# ANOVA Summary Table

Let  $n_+ = \sum_{l=1}^g n_l$ , the total sample size

Source of Variation	Sum of Squares	$df$
Treatment	$SS_{tr} = (\sum_{l=1}^g n_l) \bar{X}_l^2$	$g - 1$
Residual	$SS_{res} = \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X}_l)^2$	$n_+ - g$
Total (corrected for mean)	$(SS_{obs} - SS_{mean}) = \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X})^2$	$n_+ - 1$

**Test statistic** for  $H_o : \mu_1 = \dots = \mu_g$  (or  $H_o : \tau_1 = \dots = \tau_g$ ) and its sampling distribution are

$$F = \frac{SS_{tr}/(g-1)}{SS_{res}/(n_+ - g)} \sim \mathcal{F}_{(g-1), (n_+ - g)}$$

**Reject  $H_o$  for**

- “large” values of  $SS_{tr}/SS_{res}$ .
- “large” values of  $1 + SS_{tr}/SS_{res}$ .
- “small” values of  $(1 + SS_{tr}/SS_{res})^{-1} = \frac{SS_{res}}{SS_{tr} + SS_{res}}$

● Outline

Introduction to MANOVA

One-way ANOVA

- One-way ANOVA Review
- The Model for an Observation
- The Sums of Squares
- Sums of Squared Decomposition & Geometry
- ANOVA Summary Table

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# One-Way ANOVA as a GLM

- Outline
- Introduction to MANOVA

---

- One-way ANOVA

---

- GLM
- One-Way ANOVA as a GLM
- Least Squares Estimates of GLM
- Testing Using  $C$

---

- One-Way MANOVA

---

- Hypothesis Testing

---

- Example: Distributed vs Massed Practice

---

- Following up a Significant Result

---

- Multivariate GLM

---

- Simultaneous Confidence Intervals

---

- Summary: PCA, MANOVA, DA

---

- SAS IML and GLM

$$\begin{pmatrix} X_{11} \\ \vdots \\ X_{1n_1} \\ X_{21} \\ \vdots \\ X_{2n_2} \\ \vdots \\ X_{g1} \\ \vdots \\ X_{gn_g} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_2 \\ \vdots \\ \tau_{g-1} \\ \vdots \\ \tau_{g-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{g1} \\ \vdots \\ \epsilon_{gn_g} \end{pmatrix}$$

$$\underbrace{\mathbf{X}_{n_+ \times 1}}_{\text{"Dependent"}} = \underbrace{\mathbf{A}_{n_+ \times g}}_{\text{Design Matrix}} \underbrace{\boldsymbol{\beta}_{g \times 1}}_{\text{Parameters}} + \underbrace{\boldsymbol{\epsilon}_{n_+ \times 1}}_{\text{Residuals}}$$

# Least Squares Estimates of GLM

How we get parameter estimates depends on how the design matrix is set up. There are multiple ways of setting up the design matrix. We'll use the rank  $g$  matrix  $A$  on the previous slide.

$$\begin{aligned}\hat{\beta} &= (A' A)^{-1} A' X \\ \hat{x} &= A \hat{\beta} = \{\bar{x}_l\}_{n_+ \times 1} \\ \hat{\epsilon} &= X - \hat{X} \\ &= X - A(A' A)^{-1} A' X \\ &= (I - A(A' A)^{-1} A') X\end{aligned}$$

Our hypothesis test of equal population means,

$$H_o : \mu_1 = \mu_2 = \cdots = \mu_g \Leftrightarrow \tau_1 = \tau_2 = \cdots = \tau_g = 0$$

can be expressed as  $H_o : C\beta = 0$  where  $C$  is a contrast matrix.

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

- One-Way ANOVA as a GLM

- Least Squares Estimates of GLM

- Testing Using  $C$

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs Massed Practice

---

Following up a Significant Result

---

Multivariate GLM

---

Simultaneous Confidence Intervals

---

Summary: PCA, MANOVA, DA

---

SAS IML and GLM

# Testing Using $C$

For example,

$$C_{(g-1) \times g} = \begin{pmatrix} 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

So

$$H_o : C\beta = \mathbf{0} = \begin{pmatrix} \tau_1 - \tau_2 \\ \tau_2 - \tau_3 \\ \vdots \\ \tau_{g-2} - \tau_{g-1} \end{pmatrix}$$

Our  $F$ -test (given before) tests  $H_o : C\beta = \mathbf{0}$ .

- From GLM framework, you can introduce “continuous” (numerical) variables.
- ANOVA and multiple regression are essentially the same.
- We can generalize the GLM to the multivariate GLM.
- SAS PROC GLM will make more sense.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

● One-Way ANOVA as a GLM

● Least Squares Estimates of GLM

● Testing Using  $C$

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# One-Way MANOVA

MANOVA model for comparing  $g$  population mean vectors parallels univariate ANOVA:

$$\underbrace{\begin{pmatrix} \mathbf{X}_{lj} \\ \text{observation} \\ \text{vector} \\ p \times 1 \end{pmatrix}}_{\text{Random}} = \underbrace{\begin{pmatrix} \boldsymbol{\mu} \\ \text{overall} \\ \text{mean} \\ \text{vector} \\ p \times 1 \end{pmatrix}}_{\text{Fixed}} + \underbrace{\begin{pmatrix} \boldsymbol{\tau}_l \\ l^{\text{th}} \text{ treatment} \\ \text{effect vector} \\ p \times 1 \end{pmatrix}}_{\text{Fixed}} + \underbrace{\begin{pmatrix} \boldsymbol{\epsilon}_{lj} \\ \text{residual for} \\ l^{\text{th}} \text{ group,} \\ j^{\text{th}} \text{ case} \\ p \times 1 \end{pmatrix}}_{\text{Random}}$$

where  $\boldsymbol{\epsilon}_{lj} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$  and all independent.

for  $j = 1, \dots, n_l$  cases per group, and  $l = 1, \dots, g$  groups.

For Identification,  $\sum_{l=1}^g n_l \boldsymbol{\tau}_l = \mathbf{0}$ .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

● One-Way MANOVA

● Observation Vectors

● Sums-of-Squares and Cross-Products (SSCP)

● Sum of Squares

● A Closer Look at Within Groups SSCP

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# Observation Vectors

Each component of  $X_{lj}$  satisfies the 1-way ANOVA model, but now the model includes covariances among the components.

These covariances are assumed to be equal across populations.

A vector of observations can be decomposed as

$$\begin{pmatrix} X_{lj} \\ \text{Observation} \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \text{overall} \\ \text{sample} \\ \text{mean} \end{pmatrix} + \begin{pmatrix} (\bar{X}_l - \bar{X}) \\ \text{estimated} \\ \text{treatment} \\ \text{effect} \end{pmatrix} + \begin{pmatrix} (X_{lj} - \bar{X}_l) \\ \text{residual} \end{pmatrix}$$
$$= \hat{\mu} + \hat{\tau}_l + \hat{\epsilon}_{lj}$$

We also have a decomposition of **sums-of-squares and crossproducts**, or “SSCP” for short.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

● One-Way MANOVA

● Observation Vectors

● Sums-of-Squares and Cross-Products (SSCP)

● Sum of Squares

● A Closer Look at Within Groups SSCP

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# Sums-of-Squares and Cross-Products (SSCP)

First we'll find the total corrected squares and cross-products.

$$\begin{aligned}(\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})' &= [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})'] [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})']' \\ &= \underbrace{(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})'}_{\text{squares \& cross-products}} \\ &\quad + \underbrace{(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'}_{\text{cross-products}}\end{aligned}$$

and now sum all of this over cases and groups.

Since addition is distributive, we'll do this in pieces and look just at cross-product first. . .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

● One-Way MANOVA

● Observation Vectors

● Sums-of-Squares and Cross-Products (SSCP)

● Sum of Squares

● A Closer Look at Within Groups SSCP

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# Sum of Squares

$$\begin{aligned}
 \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' &= \left( \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) \right) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \\
 &= \left( \left( \sum_{j=1}^{n_l} \mathbf{x}_{lj} \right) - n_l \bar{\mathbf{x}}_l \right) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \\
 &= n_l \underbrace{(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})}_{\mathbf{0}} (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' = \mathbf{0}
 \end{aligned}$$

Now summing the rest over  $j$  and  $l$  we get

$$\sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})' = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' + \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'$$

$$\begin{aligned}
 \text{Total (corrected) SSCP} &= \text{Treatment} + \text{Residual} \\
 &= \text{Between Groups} + \text{Within Groups} \\
 &= \text{Hypothesis} + \text{Error}
 \end{aligned}$$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

● One-Way MANOVA

● Observation Vectors

● Sums-of-Squares and Cross-Products (SSCP)

● Sum of Squares

● A Closer Look at Within Groups SSCP

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# A Closer Look at Within Groups SSCP

$$\begin{aligned} \mathbf{W} = \mathbf{E} &= \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' \\ &= \sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)' + \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)' \\ &\quad \cdots + \sum_{j=1}^{n_g} (\mathbf{x}_{gj} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gj} - \bar{\mathbf{x}}_g)' \\ &= (n_1 - 1)\mathbf{S}_1 + (n_2 - 2)\mathbf{S}_2 + \cdots + (n_g)\mathbf{S}_g \end{aligned}$$

where  $\mathbf{S}_l$  is the sample covariance matrix for the  $l^{th}$  group (treatment, condition, etc).

$\mathbf{W}$  ( $\mathbf{E}$ ) is proportional to a pooled estimated of the common  $\Sigma$ .

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

- One-Way MANOVA
- Observation Vectors
- Sums-of-Squares and Cross-Products (SSCP)
- Sum of Squares
- A Closer Look at Within Groups SSCP

---

Hypothesis Testing

---

Example: Distributed vs Massed Practice

---

Following up a Significant Result

---

Multivariate GLM

---

Simultaneous Confidence Intervals

---

Summary: PCA, MANOVA, DA

---

SAS IML and GLM

# Between Groups SSCP & Test Statistic

With respect to between groups SSCP,

$$B = H = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' = \sum_{l=1}^g n_l \hat{\boldsymbol{\tau}}_l \hat{\boldsymbol{\tau}}_l'$$

If  $H_o : \tau_1 = \tau_2 = \dots = \tau_g = \mathbf{0}$  is true, Then  $B$  ( $H$ ) should be “close” to 0.

To test  $H_o$ , we consider the ratio of **generalized SSCPs**,

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}$$

where  $\mathbf{T} = \mathbf{W} + \mathbf{B}$  (i.e., the total corrected SSCP).

$\Lambda^*$  is known as **“Wilk’s Lambda”**. It’s equivalent to likelihood ratio statistic.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$

● Distribution of Wilk’s Lambda  $\Lambda^*$

● Other Test Statistics

● Pillai’s Trace and Roy’s

Largest Root

● How They are All Related to  $\Lambda^*$

● Which Test Statistic to Use

● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA

Summary: PCA, MANOVA, DA

# Hypothesis Testing with $\Lambda^*$

$\Lambda^*$  is a ratio of generalized sampling variances

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{\prod_{i=1}^p \lambda_i}{\prod_{i=1}^p \lambda_i^*}$$

where  $\lambda_i$  are eigenvalues of  $\mathbf{W}$ , and  $\lambda_i^*$  are eigenvalues of  $\mathbf{T}$ .

If  $H_o : \tau_1 = \tau_2 = \dots = \tau_g = 0$  is **true** then  $B$  is close to 0

$$\implies \mathbf{T} \approx \mathbf{W}$$

$$\implies \lambda_i \approx \lambda_i^*$$

$$\implies \Lambda^* \text{ close to } 1.$$

If  $H_o : \tau_1 = \tau_2 = \dots = \tau_g = 0$  is **false** then  $B$  is not close 0

$\implies$  values on diagonals of  $\mathbf{T}$ , which will be positive, will be large.

$$\implies \lambda_i < \lambda_i^*$$

$$\implies \Lambda^* \text{ is "small".}$$

The exact **distribution of  $\Lambda^*$**  can be derived for special cases of  $p$  and  $g$ .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$   
● Distribution of Wilk's Lambda  $\Lambda^*$

● Other Test Statistics  
● Pillai's Trace and Roy's

Largest Root

● How They are All Related to  $\Lambda^*$

● Which Test Statistic to Use  
● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA  
Summary: PCA, MANOVA, DA

# Distribution of Wilk's Lambda $\Lambda^*$

$$\text{Wilk's } \Lambda^* = \frac{|SSCP_e|}{|SSCP_e + SSCP_h|}$$

Number variables	$df$ for Hypothesis	Sampling distribution for multivariate data
------------------	---------------------	---

$p = 1$	$\nu_h \geq 1$	$\left(\frac{\nu_e}{\nu_h}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim \mathcal{F}_{\nu_h, \nu_e}$
---------	----------------	---

$p = 2$	$\nu_h \geq 1$	$\left(\frac{\nu_e-1}{\nu_h}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim \mathcal{F}_{2\nu_h, 2(\nu_e-1)}$
---------	----------------	---

$p \geq 1$	$\nu_h = 1$	$\left(\frac{\nu_e+\nu_h-p}{p}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim \mathcal{F}_{p, (\nu_e+\nu_h-p)}$
------------	-------------	---

$p \geq 2$	$\nu_h = 2$	$\left(\frac{\nu_e+\nu_h-p-1}{p}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim \mathcal{F}_{2p, 2(\nu_e+\nu_h-p-1)}$
------------	-------------	---

where  $\nu_h$  = degrees of freedom for hypothesis, and  
 $\nu_e$  = degrees of freedom for error (residual).

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

- Between Groups SSCP & Test Statistic

- Hypothesis Testing with  $\Lambda^*$
- Distribution of Wilk's Lambda  $\Lambda^*$

- Other Test Statistics
- Pillai's Trace and Roy's

Largest Root

- How They are All Related to  $\Lambda^*$

- Which Test Statistic to Use
- Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA  
 Summary: PCA, MANOVA, DA

# Other Test Statistics

There is more than one way to combine the information in  $B$  and  $W$  (or  $H$  and  $E$ ).

## ■ Wilk's $\Lambda^*$

$$\Lambda^* = \frac{|W|}{|W + B|} = \frac{|E|}{|E + H|} = \frac{\prod_{i=1}^p \lambda_i}{\prod_{i=1}^p \lambda_i^*}$$

where  $\lambda_i$  are eigenvalues of  $E$ , and  $\lambda_i^*$  are eigenvalues of  $(E + H)$ .

## ■ Hotelling-Lawley Trace Criteria

$$= \text{trace}(E^{-1}H) = \text{tr}(HE^{-1}) = \sum_{i=1}^g \tilde{\lambda}_i$$

where  $\tilde{\lambda}_i$  is eigenvalue of  $HE^{-1}$ .

- ◆ Reject  $H_o$  when  $\text{tr}(HE^{-1})$  is large.
- ◆ When  $H_o$  is true,  $n\text{tr}(HE^{-1}) \sim \chi_{p(g-1)}^2$ .
- ◆ Note:  $df = \text{rank of design matrix (GLM approach)}$ .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$

● Distribution of Wilk's Lambda  $\Lambda^*$

● Other Test Statistics

● Pillai's Trace and Roy's

Largest Root

● How They are All Related to  $\Lambda^*$

● Which Test Statistic to Use

● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA

Summary: PCA, MANOVA, DA

# Pillai's Trace and Roy's Largest Root

## ■ Pillai's Trace Criterion

$$= \text{trace}(\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}) = \text{trace}(\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1}) = \sum_{i=1}^p \frac{\tilde{\lambda}_i}{1 + \tilde{\lambda}_i}$$

where  $\tilde{\lambda}_i$  is the eigenvalue (root) of  $\mathbf{H}\mathbf{E}^{-1}$ .

## ■ Roy's Largest Root Criterion

$$\begin{aligned}\theta &= \text{largest root of } (\mathbf{E} + \mathbf{H})^{-1}\mathbf{H} \\ &= \text{largest root of } \mathbf{H}(\mathbf{E} + \mathbf{H})^{-1} \\ &= \left( \frac{\tilde{\lambda}_1}{1 + \tilde{\lambda}_1} \right)\end{aligned}$$

where  $\tilde{\lambda}_1$  is the largest root of  $\mathbf{E}^{-1}\mathbf{H} = \mathbf{H}\mathbf{E}^{-1}$ .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$

● Distribution of Wilk's Lambda  $\Lambda^*$

● Other Test Statistics

● Pillai's Trace and Roy's

Largest Root

● How They are All Related to

$\Lambda^*$

● Which Test Statistic to Use

● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA Summary: PCA, MANOVA, DA

# How They are All Related to $\Lambda^*$

Let  $\tilde{\lambda}_i$  be root of  $HE^{-1}$  (“eigenvalue of  $H$  relative to  $E$ ”) and if all  $\tilde{\lambda}_i$ ’s  $> 0$  (i.e.,  $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \dots \geq \tilde{\lambda}_p \geq 0$ ),

Then we can write

$$\begin{aligned}\Lambda^* &= \frac{|E|}{|E + H|} \\ &= \frac{1}{|I + HE^{-1}|} && \text{because } |AB| = |A||B| \\ &= \frac{1}{\prod_{i=1}^p (1 + \tilde{\lambda}_i)}\end{aligned}$$

So  $\Lambda^*$  is a decreasing function of  $\tilde{\lambda}_i$

Note:

$$\theta_j = \frac{\tilde{\lambda}_j}{1 + \tilde{\lambda}_j} \quad \text{and} \quad \tilde{\lambda}_i = \frac{\theta_i}{1 - \theta_i}$$

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

- Between Groups SSCP &

Test Statistic

- Hypothesis Testing with  $\Lambda^*$

- Distribution of Wilk's Lambda

$\Lambda^*$

- Other Test Statistics

- Pillai's Trace and Roy's

Largest Root

- How They are All Related to

$\Lambda^*$

- Which Test Statistic to Use

- Other Cases and Summary

MANOVA

---

Example: Distributed vs

Massed Practice

---

Following up a Significant

Result

---

Multivariate GLM

---

Simultaneous Confidence

Intervals

---

MANOVA

Summary: PCA, MANOVA, DA

# Which Test Statistic to Use

- Wilk's  $\Lambda^*$  = likelihood ratio statistic.
- If all statistics lead to the same conclusion, use  $\Lambda^*$ .
- If statistics lead to different conclusion, need to figure out why.
- From Simulation studies (power & robustness):
  - ◆ Roy's largest root found to be the least useful, except when the population structure is such that groups differ in one dimension and one group is much more different from the rest.
  - ◆ Others all do pretty good w/rt power (they use more information in  $E$  and  $H$  than Roys's).
  - ◆ Pillai's trace criterion is
    - Least affected by departures from usual population model (i.e., more robust against departures from normality).
    - Better for "diffuse" alternative hypotheses versus sharper ones.
    - When roots are approximately equal, it has best power.
  - ◆ Wilk's and Hotwelling-Lawley have about the same power for a wider-range (spectrum) of alternative hypotheses.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$

● Distribution of Wilk's Lambda  $\Lambda^*$

● Other Test Statistics

● Pillai's Trace and Roy's

Largest Root

● How They are All Related to  $\Lambda^*$

● Which Test Statistic to Use

● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA

Summary: PCA, MANOVA, DA

# Other Cases and Summary MANOVA

For cases not covered. . . If  $H_o$  is true and  $\sum_{l=1}^g n_l = n$  is large, then

$$-\ln \Lambda^* \left( n - 1 - \frac{1}{2}(p + g) \right) = (|\mathbf{W}|/|\mathbf{B} + \mathbf{W}|) \left( n - 1 - \frac{1}{2}(p + g) \right) \approx \chi_{p(g-1)}^2$$

You should examine the residual vectors for normality and outliers (i.e.,  $\hat{\epsilon}_{lj}$ 's). . . maybe use PCA or methods mentioned in the text.

Source of variation	SSCP	df	Wilk's $\Lambda^*$
Treatment (Between)	$B = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})'$	$g - 1$	$ \mathbf{W} / \mathbf{T} $
Residual (Within)	$W = \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{jl} - \bar{\mathbf{x}}_l)(\mathbf{x}_{jl} - \bar{\mathbf{x}}_l)'$	$n - g$	
Total (corrected for mean)	$T = W + B$ $= \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{jl} - \bar{\mathbf{x}})(\mathbf{x}_{jl} - \bar{\mathbf{x}})'$	$n - 1$	

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

● Between Groups SSCP & Test Statistic

● Hypothesis Testing with  $\Lambda^*$

● Distribution of Wilk's Lambda  $\Lambda^*$

● Other Test Statistics

● Pillai's Trace and Roy's Largest Root

● How They are All Related to  $\Lambda^*$

● Which Test Statistic to Use

● Other Cases and Summary MANOVA

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

MANOVA Summary: PCA, MANOVA, DA

# Example: Distributed vs Massed Practice

**1-Way MANOVA:** Data from Tatsuoka (1988), *Multivariate Analysis: Techniques for Educational and Psychological Research*, pp 273–279.

An experiment was conducted for comparing 2 methods (A & B) of teaching shorthand to 60 female seniors in a vocational high school (a dated example). Also of interest were the effects of distributed versus massed practice

$C_1$ : 2 hours of instruction/day for 6 weeks

$C_2$ : 3 hours of instruction/day for 4 weeks

$C_3$ : 4 hours of instruction/day for 3 weeks

So each subject received a total of 12 hours of instruction. For now, we'll just look the effect of distributed versus massed practice. Note:  $n_l = 20$  for  $l = 1, 2, 3$

Two variables (dependent measures):

$X_1 =$  speed

$X_2 =$  accuracy

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

- Example: Distributed vs  
Massed Practice

- Descriptive Statistics
- Means and Confidence  
Regions

- Hypothesis Test
- Hypothesis Test continued
- Test Statistic & Distribution

---

Following up a Significant  
Result

---

Multivariate GLM

---

Simultaneous Confidence  
Intervals

---

Summary: PCA, MANOVA, DA

---

SAS IML and GLM

MANOVA

# Descriptive Statistics

The overall mean vector and mean vectors for each condition:

$$\bar{\mathbf{x}} = \begin{pmatrix} 33.62 \\ 18.25 \end{pmatrix} \quad \bar{\mathbf{x}}_1 = \begin{pmatrix} 38.55 \\ 23.70 \end{pmatrix} \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} 34.00 \\ 18.20 \end{pmatrix} \quad \bar{\mathbf{x}}_3 = \begin{pmatrix} 28.30 \\ 12.85 \end{pmatrix}$$

The treatment effect vectors (i.e.,  $\hat{\boldsymbol{\tau}}_i = \bar{\mathbf{x}}_i - \bar{\mathbf{x}}$ )

$$\hat{\boldsymbol{\tau}}_1 = \begin{pmatrix} 4.93 \\ 5.45 \end{pmatrix} \quad \hat{\boldsymbol{\tau}}_2 = \begin{pmatrix} 0.38 \\ -0.05 \end{pmatrix} \quad \hat{\boldsymbol{\tau}}_3 = \begin{pmatrix} -5.32 \\ -5.40 \end{pmatrix}$$

Sample covariance matrices:

$$\mathbf{S}_1 = \begin{pmatrix} 49.52 & 13.17 \\ 13.17 & 7.59 \end{pmatrix} \quad \mathbf{S}_2 = \begin{pmatrix} 27.47 & 4.21 \\ 4.21 & 4.48 \end{pmatrix}$$

$$\mathbf{S}_3 = \begin{pmatrix} 16.33 & 4.42 \\ 4.42 & 3.19 \end{pmatrix}$$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

● Example: Distributed vs  
Massed Practice

● Descriptive Statistics

● Means and Confidence  
Regions

● Hypothesis Test

● Hypothesis Test continued

● Test Statistic & Distribution

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

MANOVA

Slide 27 of 66

# Means and Confidence Regions

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

● Example: Distributed vs  
Massed Practice

● Descriptive Statistics

● Means and Confidence  
Regions

● Hypothesis Test

● Hypothesis Test continued

● Test Statistic & Distribution

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

MANOVA

Accuracy

50

40

30

20

10

0

10

20

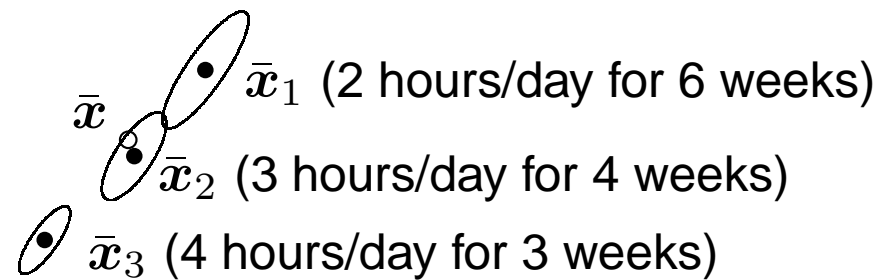
30

40

50

Speed

95% confidence regions for  $\mu_1$ ,  $\mu_2$  and  $\mu_3$   
where  $n_l = 20$



# Hypothesis Test

No difference between massed versus distributed practice on either speed or accuracy:

$$H_o : \tau_1 = \tau_2 = \tau_3 = 0 \quad \text{versus} \quad H_a : \tau_l \neq 0 \quad \text{for all } l = 1, 2, 3$$

The within groups (residual) sums of squares and cross-products matrix

$$\begin{aligned} \mathbf{W} &= (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + (n_3 - 1)\mathbf{S}_3 \\ &= 19 \begin{pmatrix} 49.52 & 13.17 \\ 13.17 & 7.59 \end{pmatrix} + 19 \begin{pmatrix} 27.47 & 4.21 \\ 4.21 & 4.48 \end{pmatrix} \\ &\quad + 19 \begin{pmatrix} 16.33 & 4.42 \\ 4.42 & 3.19 \end{pmatrix} \\ &= \begin{pmatrix} 1773.15 & 414.20 \\ 414.20 & 289.95 \end{pmatrix} \end{aligned}$$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

● Example: Distributed vs  
Massed Practice

● Descriptive Statistics

● Means and Confidence

Regions

● Hypothesis Test

● Hypothesis Test continued

● Test Statistic & Distribution

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

MANOVA

Slide 29 of 66

# Hypothesis Test continued

The between groups SSCP matrix:

$$\begin{aligned} \mathbf{B} &= \sum_{l=1}^3 n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \\ &= 20 \begin{pmatrix} 4.93 \\ 5.45 \end{pmatrix} (4.93, 5.45) + 20 \begin{pmatrix} 0.38 \\ -0.05 \end{pmatrix} (0.38, -0.05) \\ &\quad + 20 \begin{pmatrix} -5.32 \\ -5.40 \end{pmatrix} (-5.32, -5.40) \\ &= \begin{pmatrix} 1055.033 & 1111.55 \\ 1111.55 & 1177.30 \end{pmatrix} \\ \mathbf{T} = \mathbf{W} + \mathbf{B} &= \begin{pmatrix} 2828.18 & 1525.75 \\ 1525.75 & 1467.25 \end{pmatrix} \end{aligned}$$

Or  $\mathbf{T} = (n - 1)\mathbf{S}$  where  $\mathbf{S}$  is the covariance matrix computed over all groups and  $n$  is the total sample. Then  $\mathbf{B} = \mathbf{T} - \mathbf{W}$ .

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

● Example: Distributed vs  
Massed Practice

● Descriptive Statistics

● Means and Confidence

Regions

● Hypothesis Test

● Hypothesis Test continued

● Test Statistic & Distribution

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

MANOVA

Slide 30 of 66

# Test Statistic & Distribution

$$|\mathbf{W}| = (1773.15)(289.95) - (414.20)^2 = 342563.2$$

$$|\mathbf{T}| = (2828.18)(1467.25) - (1525.75)^2 = 1821738.9$$

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{342563.2}{1821738.9} = 0.188$$

For  $p = 2$  and  $g = 3$ , we can use the exact sample distribution:

$$\left( \frac{n - g - 1}{g - 1} \right) \left( \frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) = \left( \frac{n - p - 2}{p} \right) \left( \frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim \mathcal{F}_{2(g-1), 2(n-g-1)}$$

which for this example,

$$\frac{(60 - 3 - 1)}{(3 - 1)} \left( \frac{1 - \sqrt{.188}}{\sqrt{.188}} \right) = \frac{56}{2} \left( \frac{.566}{.434} \right) = 36.568$$

Since  $\mathcal{F}_{4,112}(\alpha = .05) \sim \mathcal{F}_{4,120}(\alpha = .05) = 2.45$ , reject  $H_0$  that treatment vectors are all equal to  $\mathbf{0}$ . The data support the conclusion that there is an effect of massed versus distributed practice.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

● Example: Distributed vs  
Massed Practice

● Descriptive Statistics

● Means and Confidence

Regions

● Hypothesis Test

● Hypothesis Test continued

● Test Statistic & Distribution

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

MANOVA

Slide 31 of 66

# Following up a Significant Result

1. Multivariate contrasts & confidence regions
2. “Tests” on individual variables (simultaneous confidence intervals for group/treatment differences).
3. Discriminant Analysis.

## Multivariate Contrasts

We need the multivariate generalization of the general linear model:

$$\mathbf{X}_{gn \times p} = \mathbf{A}_{gn \times (g+1)} \mathbf{B}_{(g+1) \times p} + \mathcal{E}_{gn \times p}$$

where  $\mathbf{A}$  is the design matrix (it could have  $g$  or  $g + 1$  columns depending on the parameterization), and  $\mathbf{B}$  is a matrix of coefficients (model parameters)... some examples...

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

● Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

# $A$ is $n_+ \times g$ with dummy codes

$$\mathbf{X}_{n_+ \times p} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \beta_{o1} & \beta_{o2} & \cdots & \beta_{op} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{g-1,1} & \beta_{g-1,2} & \cdots & \beta_{g-1,p} \end{pmatrix} + \mathcal{E}_{n_+ \times p}$$

Given the design matrix above,  $\beta_{ok} = \mu_{gk}$ , and  $\beta_{lk} = \mu_{lk} - \mu_{gk}$ .

If  $p = 1$ , we would have 1-way ANOVA.

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# SAS PROC GLM design: $A$ is $gn \times (g + 1)$

An alternative design matrix and parameter vector:

$$\mathbf{X}_{n_+ \times p} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}_{n_+ \times (g+1)} \begin{pmatrix} \beta_{o1} & \beta_{o2} & \cdots & \beta_{op} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{g-1,1} & \beta_{g-1,2} & \cdots & \beta_{g-1,p} \end{pmatrix} + \mathcal{E}_{n_+ \times p}$$

Normally,  $\hat{\beta} = (A' A)^{-1} A' X$ ; however, the rank of  $A$  defined above (and hence  $A' A$ ) is only  $g \implies$  There's no unique solution to  $A' X = A' A \beta$ .

- Outline
- Introduction to MANOVA
- One-way ANOVA
- GLM
- One-Way MANOVA
- Hypothesis Testing
- Example: Distributed vs Massed Practice
- Following up a Significant Result
- Multivariate GLM
- $A$  is  $n_+ \times g$  with dummy codes
- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$
- What's Interesting
- Estimable and Testable
- Example: Cameron & Pauling Data
- Example: Descriptive statistics
- Plot of Means and 95% Confidence Regions
- Plot of Means and 95% Confidence Regions
- Results of MANOVA
- Hypothesis Test
- MANOVA
- Estimated MANOVA parameters

# What's Interesting

We're interested in differences between group means; that is,

$$\mu_i - \mu_k = (\mu + \tau_i)_{p \times 1} - (\mu + \tau_k)_{p \times 1} = \tau_i - \tau_k$$

Even if we can't get unique estimates of elements of  $\mathcal{B}$ , we can get unique estimates of differences between parameter estimates, which correspond to differences between group means regardless of what inverse of  $(A' A)$  is used.

**Moore-Penrose** inverses of non-full rank square matrix  $(A' A)$  is denoted by  $(A' A)^-$ .

SAS PROC GLM uses the Moore-Penrose inverse of  $(A' A)$ .

In SAS/PROC IML, the Moore-Penrose inverse is obtained by the command `ginv( )`, for example

$$giAA = ginv(A'^*A);$$

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

MANOVA

- Estimated MANOVA  
parameters

# Estimable and Testable

What we can do is test linear combinations of elements of  $\mathcal{B}$  if the linear combination is a **contrast**.

**Estimable:** A linear function  $c'\mathcal{B}$  is estimable if

$$(A'A)(A'A)^{-}c = c.$$

**Testable:** A linear function is testable if it only involves the estimable **functions** of  $\mathcal{B}$ .

Contrasts of elements of  $\mathcal{B}$  are estimable and therefore testable. These correspond to differences between means.

We'll demonstrate multivariate General linear model by example...

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# Example: Cameron & Pauling Data

Increase in survival of cancer patients given supplemental treatment with vitamin C.

“Increase in survival” = the number of days a patient survives minus the number of days matched control survives.

$x_1 = d_1$  = increase in survival measured as days from first hospitalization.

$x_2 = d_2$  = increase in survival measured days from un-treatability.

type = type of cancer (1 =stomach, 2 =bronchus, 3 =colon, 4 =rectum, 5 =bladder, 6 =kidney).

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# Example: Descriptive statistics

$l$	Type	$n_l$	$\bar{x}_i = \bar{d}_i$	$S_l$	
1	Stomach	12	70.67	125869.88	15206.70
			94.92	15206.70	12235.36
2	Bronchus	16	49.50	23915.07	17461.47
			106.88	17461.47	16619.45
3	Colon	16	117.19	252027.76	147884.76
			293.19	147884.76	118274.56
4	Rectum	7	297.43	508715.62	155250.05
			226.57	155250.05	56340.62
5	Bladder	5	1304.20	3747663.20	214071.05
			129.80	214071.05	14697.70
6	Kidney	7	118.86	129018.48	3344.62
			101.71	3344.62	17392.91

$$S_{pool} = \sum_{l=1}^5 (n_l - 1) S_l = \begin{pmatrix} 419668.42 & 76815.85 \\ 76815.85 & 45848.122 \end{pmatrix}$$

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with dummy codes

- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling Data

- Example: Descriptive statistics

- Plot of Means and 95% Confidence Regions

- Plot of Means and 95% Confidence Regions

- Results of MANOVA

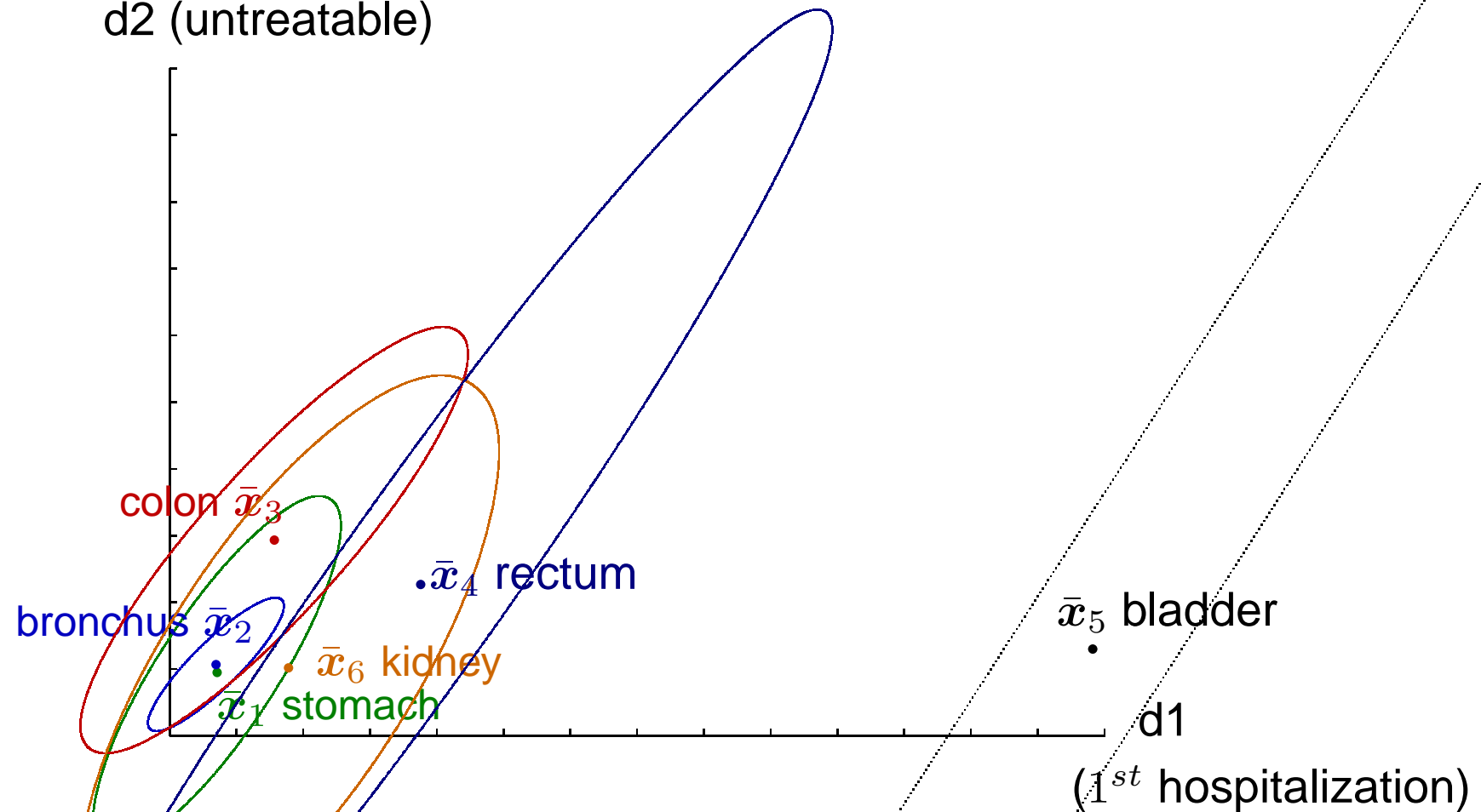
Hypothesis Test

MANOVA

- Estimated MANOVA parameters

# Plot of Means and 95% Confidence Regions

d2 (untreatable)



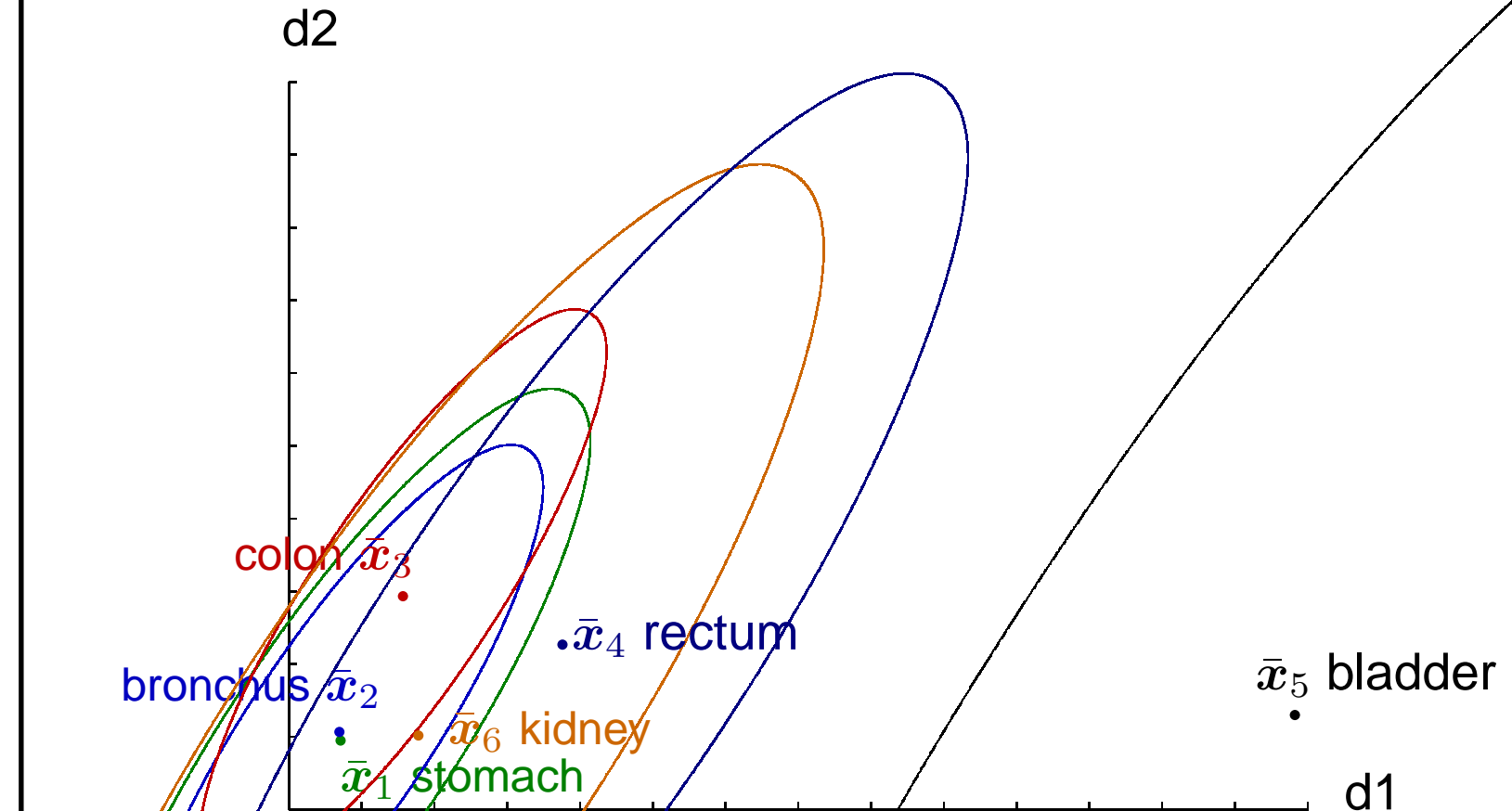
Using  $S_i$  to compute regions

$$n_1 = 12, n_2 = 16, n_3 = 16, n_4 = 7, n_5 = 5, n_6 = 7$$

- Outline
- Introduction to MANOVA
- One-way ANOVA
- GLM
- One-Way MANOVA
- Hypothesis Testing
- Example: Distributed vs Massed Practice
- Following up a Significant Result
- Multivariate GLM
- $A$  is  $n_+ \times g$  with dummy codes
- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$
- What's Interesting
- Estimable and Testable
- Example: Cameron & Pauling Data
- Example: Descriptive statistics
- Plot of Means and 95% Confidence Regions
- Plot of Means and 95% Confidence Regions
- Results of MANOVA Hypothesis Test
- Estimated MANOVA parameters

# Plot of Means and 95% Confidence Regions

- Outline
- Introduction to MANOVA
- One-way ANOVA
- GLM
- One-Way MANOVA
- Hypothesis Testing
- Example: Distributed vs Massed Practice
- Following up a Significant Result
- Multivariate GLM
- $A$  is  $n_+ \times g$  with dummy codes
- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$
- What's Interesting
- Estimable and Testable
- Example: Cameron & Pauling Data
- Example: Descriptive statistics
- Plot of Means and 95% Confidence Regions
- Plot of Means and 95% Confidence Regions
- Results of MANOVA Hypothesis Test
- Estimated MANOVA parameters



Using  $S_{pool}$  to the compute regions

# Results of MANOVA Hypothesis Test

$$H_o : \mu_{stomach} = \mu_{bronchus} = \mu_{colon} = \mu_{rectum} = \mu_{bladder} = \mu_{kidney}$$

or equivalently

$$H_o : \tau_{stomach} = \tau_{bronchus} = \tau_{colon} = \tau_{rectum} = \tau_{bladder} = \tau_{kidney}$$

- $df$  type of cancer (hypothesis)  $\nu_h = g - 1 = 6 - 1 = 5$
- $df$  within (error)  $= \nu_e = \sum_l n_l - g = 63 - 6 = 57$
- Wilk's  $\Lambda^* = \det(\mathbf{W}) / \det(\mathbf{T}) = 0.5817749$
- Since  $p = 2$  dependent variables, Wilk's  $\Lambda^*$  has an exact sampling distribution that is  $\mathcal{F}$ , in particular

$$F = \left( \frac{(\nu_e - 1)}{\nu_h} \right) \left( \frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim \mathcal{F}_{2\nu_h, 2\nu_e}$$

- $F = 3.4838$  and  $p$ -value = .0005.
- Reject  $H_o$ . The data support the conclusion that not all of the means (or  $\tau$ 's are equal).

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

●  $A$  is  $n_+ \times g$  with  
dummy codes

● SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

● What's Interesting

● Estimable and Testable

● Example: Cameron & Pauling  
Data

● Example: Descriptive  
statistics

● Plot of Means and 95%  
Confidence Regions

● Plot of Means and 95%  
Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA  
parameters

# Estimated MANOVA parameters

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 205.56 \\ 166.46 \end{pmatrix}$$

Type		hospitalization	untreatable
Stomach	$\hat{\boldsymbol{\tau}}'_1 =$	(-134.888,	-71.543)
Bronchus	$\hat{\boldsymbol{\tau}}'_2 =$	(156.055,	-59.585)
Colon	$\hat{\boldsymbol{\tau}}'_3 =$	(-88.368,	126.727)
Rectum	$\hat{\boldsymbol{\tau}}'_4 =$	(91.873,	60.111)
Bladder	$\hat{\boldsymbol{\tau}}'_5 =$	(1098.644,	-36.660)
Kidney	$\hat{\boldsymbol{\tau}}'_6 =$	(-86.698,	-64.746)

Recall that  $\boldsymbol{\mu}_l = \boldsymbol{\mu} + \boldsymbol{\tau}_l$

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- MANOVA

- Estimated MANOVA  
parameters

# MANOVA as a multivariate GLM

A main effect and six dummy variables (this is what PROC GLM does). So the design matrix looks like

$$A_{n_+ \times 7} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} \} n_{stomach} \\ \} n_{bronchus} \\ \} n_{colon} \\ \} n_{rectum} \\ \} n_{bladder} \\ \} n_{kidney} \end{matrix}$$

MANOVA (multivariate general linear model):

$$y_{n_+ \times 2} = A_{n_+ \times 7} B_{7 \times 2} + \epsilon_{n_+ \times 2}$$

Estimation:

$$B = (A' A)^{-1} A' y$$

Predicted values:

$$\hat{y} = AB \quad \text{where} \quad \hat{y}'_{jl} = (\bar{x}_{1l}, \bar{x}_{2l})$$

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# MANOVA as a multivariate GLM (continued)

For  $\hat{y}'_{jl} = (\bar{x}_{1l}, \bar{x}_{2l})$ , it is the the case that

$$\bar{x}_{j1l} = b_{o1} + b_{l1} \quad \text{and} \quad \bar{x}_{j2l} = b_{o2} + b_{l2}$$

So to compare two groups (types of cancer),

$$\bar{x}_{il} - \bar{x}_{il^*} = (b_{oi} + b_{li}) - (b_{oi} + b_{l^*i}) = b_{li} - b_{l^*i}$$

Consider a contrast between means for two types of cancer, for example, stomach and bronchial,

$$c'B = (0, 1, -1, 0, 0, 0, 0) \begin{pmatrix} b_{o1} & b_{o2} \\ b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \\ b_{61} & b_{62} \end{pmatrix} = ((b_{11} - b_{21}), (b_{12} - b_{22})) = ((\bar{x}_{11} - \bar{x}_{21}), (\bar{x}_{12} - \bar{x}_{22}))$$

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs Massed Practice

---

Following up a Significant Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with dummy codes

- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling Data

- Example: Descriptive statistics

- Plot of Means and 95% Confidence Regions

- Plot of Means and 95% Confidence Regions

---

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# With the Parameter Estimates

$$B = \begin{pmatrix} 326.31 & 158.84 \\ -255.64 & -63.93 \\ -276.81 & -51.97 \\ -209.12 & 134.34 \\ -28.88 & 67.73 \\ 977.89 & -29.04 \end{pmatrix}$$

stomach:  $\bar{x}_{11} = b_{o1} + b_{11} = 326.31 + (-255.64) = 70.67$

$\bar{x}_{12} = b_{o2} + b_{12} = 158.84 + (-63.93) = 94.91$

bronchus:  $\bar{x}_{21} = b_{o1} + b_{21} = 326.31 + (-276.81) = 49.50$

$\bar{x}_{22} = b_{o2} + b_{22} = 158.84 + (-51.97) = 106.87$

$$\begin{aligned} (0, 1, -1, 0, 0, 0, 0)B &= ((-255.64 + 276.81), (-63.93 - 51.97)) \\ &= (21.17, -11.96) = ((\bar{x}_{11} - \bar{x}_{21}), (\bar{x}_{12} - \bar{x}_{22})) = \mathbf{0}' \end{aligned}$$

i.e., Are effects for stomach and bronchus the same?

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

MANOVA

- Estimated MANOVA  
parameters

# Testing $H_o : CB = 0$

Our hypothesis tests can be of the form

$$H_o : C_{r \times g} B_{g \times p} M_{p \times s} = 0$$

$C$  defines hypotheses (contrasts) on the elements of columns of  $B$ ; that is, comparison between means on same variables over groups.

$M$  defines hypotheses (contrasts) on the elements of rows of  $B$ ; that is, comparison between means on same group over variables.

For now  $M = I$  and we'll consider hypotheses of the form

$$H_o : CB = 0$$

Specifically, we want to consider

$$H_o : 0b_{0k} + c_1b_{1k} + c_2b_{2k} + \dots + c_gb_{gk} \Leftrightarrow c_1\tau_1 + c_2\tau_2 + \dots + c_g\tau_g$$

where  $\sum_{l=1}^g c_l = 0$ .

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# Testing Contrasts: The H matrix

For a simple contrast, such as

$$c = (0, 1, -1, 0, \dots, 0)$$

we could do this as a multivariate  $T^2$  test for independent groups; however, we'll stay within the MANOVA and multivariate linear model framework (so we can test multiple ones).

Suppose that we have a contrast matrix  $C_{r \times (g+1)}$  where the rows are  $r$  orthogonal contrasts, the hypothesis matrix equals

$$H = (CB)'(C(A'A)^{-1}C')^{-1}(CB)$$

For a balanced design (i.e.,  $n_1 = n_2 = \dots = n_g = n$  and a single contrast (i.e.,  $r = 1$ ), this reduces to

$$H = \frac{n}{\sum_{l=1}^g c_l} \left( \sum_{l=1}^g c_l \bar{x}_l \right) \left( \sum_{l=1}^g c_l \bar{x}_l \right)'$$

Note:  $c_0 = 0$ .

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with dummy codes

- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling Data

- Example: Descriptive statistics

- Plot of Means and 95% Confidence Regions

- Plot of Means and 95% Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# Testing Contrasts

The “Error” matrix is  $W$ ; that is,

$$\begin{aligned} E = W &= \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_l)(\mathbf{x}_{jk} - \bar{\mathbf{x}}_l)' \\ &= \mathbf{X}'\mathbf{X} - \mathbf{B}'(\mathbf{A}'\mathbf{A})\mathbf{B} \end{aligned}$$

Wilk’s Lambda for the test  $H_o : \mathbf{C}\mathbf{B} = \mathbf{0}$  is

$$\Lambda^* = \frac{\det(\mathbf{E})}{\det(\mathbf{H} + \mathbf{E})}.$$

To find the transformation of this to an  $\mathcal{F}$  distribution:

$p = \text{anything}$

$\nu_{\text{hypothesis}} = r$

$\nu_{\text{error}} = \sum_l n_l - p$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

●  $A$  is  $n_+ \times g$  with  
dummy codes

● SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

● What’s Interesting

● Estimable and Testable

● Example: Cameron & Pauling  
Data

● Example: Descriptive  
statistics

● Plot of Means and 95%  
Confidence Regions

● Plot of Means and 95%  
Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA  
parameters

# Example: Five types the same?

Three equivalent forms for Hypothesis 1:

$$H_0 : \quad \mu_{bronchus} = \mu_{colon} = \mu_{kidney} = \mu_{rectum} = \mu_{stomach}$$

$$H_0 : \quad \tau_{bronchus} = \tau_{colon} = \tau_{kidney} = \tau_{rectum} = \tau_{stomach}$$

$$H_0 : \quad \beta_{bronchus} = \beta_{colon} = \beta_{kidney} = \beta_{rectum} = \beta_{stomach}$$

where  $\beta_l$  is a  $p \times 1$  column vector of  $B'$  (i.e., a row of  $B$  written as a column).

For the contrast matrix we need to know the order of the effects in the GLM.

- I re-order them so that they are in alphabetical order, because
- PROC GLM puts them in alphabetical order (or numerical if groups are coded this way).

...

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

●  $A$  is  $n_+ \times g$  with  
dummy codes

● SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

● What's Interesting

● Estimable and Testable

● Example: Cameron & Pauling  
Data

● Example: Descriptive  
statistics

● Plot of Means and 95%  
Confidence Regions

● Plot of Means and 95%  
Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA  
parameters

# $H_o$ : Four types the same?

$$H_o : CB = \mathbf{0} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \\ \beta_{51} & \beta_{52} \\ \beta_{61} & \beta_{61} \end{pmatrix} \begin{matrix} \text{intercept} \\ \text{bladder} \\ \text{bronchus} \\ \text{colon} \\ \text{kidney} \\ \text{rectum} \\ \text{stomach} \end{matrix}$$

So

$$H_o : \begin{pmatrix} (\beta_{21} - \beta_{31}) & (\beta_{22} - \beta_{32}) \\ (\beta_{21} - \beta_{41}) & (\beta_{22} - \beta_{42}) \\ (\beta_{21} - \beta_{51}) & (\beta_{22} - \beta_{52}) \\ (\beta_{21} - \beta_{61}) & (\beta_{22} - \beta_{62}) \end{pmatrix} = \begin{pmatrix} (\tau_{21} - \tau_{31}) & (\tau_{22} - \tau_{32}) \\ (\tau_{21} - \tau_{41}) & (\tau_{22} - \tau_{42}) \\ (\tau_{21} - \tau_{51}) & (\tau_{22} - \tau_{52}) \\ (\tau_{21} - \tau_{61}) & (\tau_{22} - \tau_{62}) \end{pmatrix} = \mathbf{0}$$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

●  $A$  is  $n_+ \times g$  with  
dummy codes

● SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

● What's Interesting

● Estimable and Testable

● Example: Cameron & Pauling  
Data

● Example: Descriptive  
statistics

● Plot of Means and 95%  
Confidence Regions

● Plot of Means and 95%  
Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA  
parameters

# Hypothesis Matrices

$$H = (CB)'(C(A'A)(ns + nb + nc + nr + nd + nk - 6); C')^{-1}(CB)$$

$$\begin{pmatrix} 324369.92785 & 180717.48204 \\ 180717.48204 & 429243.40815 \end{pmatrix}$$

The  $E$  error SSCP is the same as  $W$  that we used before, which equals

$$E = W = X'X - B(A'A)B'$$

$$= \begin{pmatrix} 24340768.476 & 4455319.3042 \\ 4455319.3042 & 2659191.047 \end{pmatrix}$$

Wilk's Lambda,

$$\Lambda^* = \frac{\det(E)}{\det(H + E)} = \frac{4.4877E13}{5.4684E13} = 0.820661$$

and

■  $\nu_h =$  number of rows of  $C = 4$

■  $\nu_e =$  number of rows of  $X - p = 57$

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

●  $A$  is  $n_+ \times g$  with  
dummy codes

● SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

● What's Interesting

● Estimable and Testable

● Example: Cameron & Pauling  
Data

● Example: Descriptive  
statistics

● Plot of Means and 95%  
Confidence Regions

● Plot of Means and 95%  
Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA  
parameters

# Results

Referring to the table for transformations of  $\Lambda^*$  that have sampling distributions that are  $\mathcal{F}$ , we use the one for  $p = 2$  and  $\nu_h \geq 1$ , which is

$$F = \left( \frac{\nu_e - 1}{\nu_h} \right) \left( \frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) = \left( \frac{56}{4} \right) \left( \frac{1 - \sqrt{0.820661}}{\sqrt{0.820661}} \right) = 1.4541861$$

If the null is true, then this should have a  $\mathcal{F}_{2\nu_h, 2(\nu_e - 1)}$  sampling distribution.

Comparing  $F = 1.45$  to the  $\mathcal{F}_{4, 112}$ , we find that the  $p$ -value is .18. Retain the null hypothesis. The data suggest not difference in increased survival of patients over different types of cancer (except bladder).

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with dummy codes
- SAS PROC GLM design:  $A$  is  $gn \times (g + 1)$
- What's Interesting
- Estimable and Testable
- Example: Cameron & Pauling Data
- Example: Descriptive statistics
- Plot of Means and 95% Confidence Regions
- Plot of Means and 95% Confidence Regions

● Results of MANOVA

Hypothesis Test

MANOVA

● Estimated MANOVA parameters

# Five versus the Rest

$$H_o : \tau_{bladder} = \tau_{bronchus} + \tau_{colon} + \tau_{kidney} + \tau_{rectum} + \tau_{stomach} / 5$$

or equivalently

$$H_o : C\mathcal{B} = \mathbf{0} = (0, -5, 1, 1, 1, 1, 1)$$

(	$\beta_{o1}$	$\beta_{o2}$	)	intercept
	$\beta_{11}$	$\beta_{12}$		bladder
	$\beta_{21}$	$\beta_{22}$		bronchus
	$\beta_{31}$	$\beta_{32}$		colon
	$\beta_{41}$	$\beta_{42}$		kidney
	$\beta_{51}$	$\beta_{52}$		rectum
	$\beta_{61}$	$\beta_{61}$		stomach

$E$  is the same as before, but now

$$H = \begin{pmatrix} 6266038.6574 & -186105.9247 \\ -186105.9247 & 5527.481893 \end{pmatrix}$$

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

---

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# The Test and Result

$$\Lambda^* = \frac{4.4877E13}{6.3332E13} = 0.7085934$$

$$\nu_e = \sum_{l=1}^g n_l - g = 57.$$

$\nu_h = 1$ , the number of rows of  $C$ .

So, for  $\nu_h = 1$ , use

$$F = \left( \frac{\nu_e + \nu_h - p}{p} \right) \left( \frac{1 - \Lambda^*}{\Lambda^*} \right) = 11.514901,$$

which if the null is true (and assumptions valid),  $F$  should have a sampling distribution that is  $\mathcal{F}_{p,(\nu_e + \nu_h - p)}$ . Comparing  $F$  to  $\mathcal{F}_{2,56}$ , we get a  $p$ -value  $< .01$ . Reject  $H_o$ .

**In summary**, the mean survival of patients with bladder cancer differs from that of those with other types of cancer; however, no support for differences between the other types.

**Question:** Are there differences for survival from first hospitalization and/or from time of unreadability?

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

- $A$  is  $n_+ \times g$  with  
dummy codes

- SAS PROC GLM design:  $A$   
is  $gn \times (g + 1)$

- What's Interesting

- Estimable and Testable

- Example: Cameron & Pauling  
Data

- Example: Descriptive  
statistics

- Plot of Means and 95%  
Confidence Regions

- Plot of Means and 95%  
Confidence Regions

- Results of MANOVA

Hypothesis Test

- Estimated MANOVA

parameters

# Simultaneous Confidence Intervals

We can construct simultaneous confidence intervals for components of differences

$$\tau_l - \tau_{l'} \quad (\text{which equal } \mu_l - \mu_{l'})$$

or other linear combinations such as

$$\tau_1 - (\tau_2 + \tau_3)/2$$

There are at least **three** ways of doing this

- **Specify a matrix  $M$**  in the hypothesis test  $H_o : CBM = 0$  that is a  $(p \times 1)$  vector with all

$$M' = (0, \dots, \underbrace{1}_{i^{th}}, \dots, 0)$$

- **Bonferroni-type**: Same as above but split the  $\alpha$  into pieces, on part for each of the **planned** comparisons.
- **Roy's method**, which is based on the “union-intersection” principle.

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

● Simultaneous Confidence  
Intervals

● Using  $CBM = 0$

● Confidence interval for  
 $CBM$

● Our example: CI for  
 $CBM$

● Plot of Means and 95%  
Confidence Regions

● Notes about These CIs

● Bonferroni Intervals

● Roy's Method

● A Truly Multivariate Follow-Up  
MANOVA

Summary: PCA, MANOVA, DA

# Using $CBM = 0$

- $C$  picks out which two (or more groups to compare).  
e.g., want to compare bladder with the rest,

$$C = (0, 1, -.2, -.2, -.2, -.2, -.2)$$

- $M$  picks out which variable (or linear combination of variables).  
e.g., Just compare  $d_1$ , increase in survival from first hospitalization,

$$M' = (1, 0)$$

Putting these together in our example gives us

$$(0, 1, -.2, -.2, -.2, -.2, -.2) \begin{pmatrix} \beta_{o1} & \beta_{o2} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \\ \beta_{51} & \beta_{52} \\ \beta_{61} & \beta_{61} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left( \beta_{11} - \frac{1}{5} \sum_{l=2}^6 \beta_{l1} \right)$$

# Confidence interval for $CBM$

$$\beta_{11} - \frac{1}{5} \sum_{l=2}^6 \beta_{l1} = \tau_{11} - \sum_{l=2}^6 \tau_{l1}$$

We need two things:

- A “fudge-factor”— a value from a probability distribution
- An estimate of the standard error

A  $(1 - \alpha)100\%$  confidence statement given vectors  $C_{1 \times (g+1)}$  and  $M_{p \times 1}$  is

$$CBM \pm \sqrt{\mathcal{F}_{1, \nu_e}(\alpha)} \sqrt{(M' S_{pool} M)(C(A' A)^{-1} C')}$$

Note: Consider two columns of  $B$ ,  $\beta_i$  and  $\beta_k$ , the covariance matrix between them is

$$\text{cov}(\beta_i, \beta_k) = s_{pool, ik} (A' A)^{-1}$$

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs Massed Practice

Following up a Significant Result

Multivariate GLM

Simultaneous Confidence Intervals

- Simultaneous Confidence Intervals

- Using  $CBM = 0$

- Confidence interval for  $CBM$

- Our example: CI for  $CBM$

- Plot of Means and 95% Confidence Regions

- Notes about These CIs

- Bonferroni Intervals

- Roy's Method

- A Truly Multivariate Follow-Up MANOVA

Summary: PCA, MANOVA, DA

# Our example: CI for $CBM$

$$\begin{aligned}\beta_{11} - \frac{1}{5} \sum_{l=2}^6 \beta_{l1} &= 1304.20 - \frac{1}{5} (49.50 + 117.19 + 118.86 + 297.43 + 70.67) \\ &= 1173.47\end{aligned}$$

$$\begin{aligned}(M' S_{pool} M)(C(A'A)^{-1} C') &= s_{pool,11} (0.2197619) \\ &= (427031.03)(0.2197619) = 93845.152\end{aligned}$$

And  $\mathcal{F}_{1,57}(.05) = 4.01$ . So our 95% confidence interval is

$$\begin{aligned}1304.20 \pm \sqrt{4.01} \sqrt{93845.152} \\ 1304.20 \pm \sqrt{4.01} (306.34) \longrightarrow (560.03, 1786.91)\end{aligned}$$

Since 0 is not in the interval, the mean increase in survival from first hospitalization due to bladder cancer is larger than the average of the others means.

Should we test whether the same is true for increase in survival from time of untreatability?

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

- Simultaneous Confidence  
Intervals

- Using  $CBM = 0$

- Confidence interval for  
 $CBM$

- Our example: CI for  
 $CBM$

- Plot of Means and 95%  
Confidence Regions

- Notes about These CIs

- Bonferroni Intervals

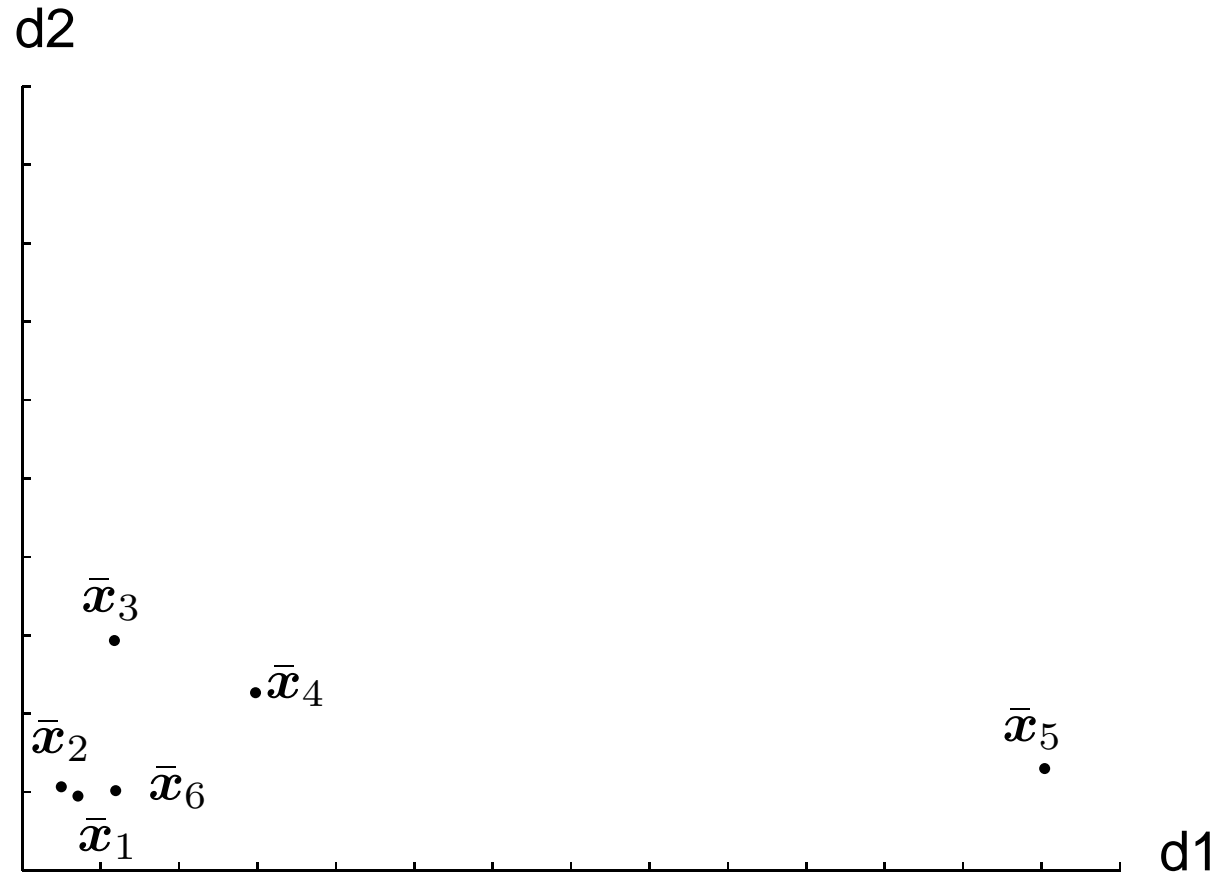
- Roy's Method

- A Truly Multivariate Follow-Up  
MANOVA

Summary: PCA, MANOVA, DA

# Plot of Means and 95% Confidence Regions

- Outline
- Introduction to MANOVA
- One-way ANOVA
- GLM
- One-Way MANOVA
- Hypothesis Testing
- Example: Distributed vs Massed Practice
- Following up a Significant Result
- Multivariate GLM
- Simultaneous Confidence Intervals
  - Simultaneous Confidence Intervals
  - Using  $CBM = 0$
  - Confidence interval for  $CBM$
  - Our example: CI for  $CBM$
  - Plot of Means and 95% Confidence Regions
  - Notes about These CIs
  - Bonferroni Intervals
- Roy's Method
- A Truly Multivariate Follow-Up MANOVA
- Summary: PCA, MANOVA, DA



# Notes about These CIs

- If you're only looking (testing) the difference between two means, e.g.,

$$\tau_{li} - \tau_{l'i}$$

Then the standard error is just  $s_{pool,ii} \sqrt{1/n_l + 1/n_{l'}}$

- When looking at a difference for a variable (e.g., above), these confidence statements are equivalent to what you would get from 1-way ANOVA using Fisher's least significant differences; that is, they are univariate CIs.
- When considering a linear combination of variables, these CIs are equivalent to univariate CIs where you've analyzed a new or composite variable defined by the linear combination.
- In our example, we don't have to worry too much about inflated Type I error rate, because we only did one CI after rejecting the overall test and using multivariate contrasts to narrow down where differences exist.
- If you do all pairwise differences, there are  $g(g-1)/2$  pairs times  $p$  variables. (e.g.,  $2(6)(5)/2 = 30$ ).

- Outline

---

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs Massed Practice

---

Following up a Significant Result

---

Multivariate GLM

---

Simultaneous Confidence Intervals

- Simultaneous Confidence Intervals

- Using  $CBM = 0$

- Confidence interval for  $CBM$

- Our example: CI for  $CBM$

- Plot of Means and 95% Confidence Regions

- Notes about These CIs

- Bonferroni Intervals

- Roy's Method

- A Truly Multivariate Follow-Up MANOVA

---

Summary: PCA, MANOVA, DA

# Bonferroni Intervals

If you have planned to look at all pairwise comparisons before looking at the data (i.e.,  $m = pg(g - 1)/2$ ), then you can use

$$t_{\nu_e}(\alpha/(2m))$$

as your fudge factor.

Let  $n_+ = \sum_{l=1}^g n_l$ . For the model  $X_{lj} = \mu + \tau_l + \epsilon_{lj}$  with  $j = 1, \dots, n_l$  and  $l = 1, \dots, g$  with confidence at least  $(1 - \alpha)$ ,  $(\tau_{li} - \tau_{l'i})$  belongs to

$$(\bar{x}_{li} - \bar{x}_{l'i}) \pm t_{\nu_e}(\alpha/(2m)) \sqrt{s_{pool,ii} \left( \frac{1}{n_l} + \frac{1}{n_{l'}} \right)}$$

for all components (variables)  $i = 1, \dots, p$  and all differences  $l < l' = 1, \dots, g$ .

No need to do this with our cancer data.

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

- Simultaneous Confidence

Intervals

- Using  $CBM = 0$

- Confidence interval for  
 $CBM$

- Our example: CI for  
 $CBM$

- Plot of Means and 95%  
Confidence Regions

- Notes about These CIs

- Bonferroni Intervals

- Roy's Method

- A Truly Multivariate Follow-Up  
MANOVA

Summary: PCA, MANOVA, DA

# Roy's Method

- This is based on the union intersection principle.
- This is more like the first method that we considered (i.e., *CBM*); however, we use a different distribution for our “fudge factor”.
- We use Greatest Root, of  $|H - \theta(E + H)| = 0$  where  $H$  (between groups or hypothesis SSCP matrix) and  $E = W$  (error or within groups SSCP matrix) are independent Wishart matrices.
- To apply this result, we need percentiles of the greatest root distribution of the largest root  $\lambda$  of the equation  $|H - \lambda E| = 0$ . Percentile can be found in tables
- This distribution does not depend on  $\Sigma$  but only on  $df = n - g - p - 1$ .

Recommendation: I'd suggest using Scheffé's method where you do 1-way ANOVA on a linear combination of variables and then specify the contrast that you want.

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

- Simultaneous Confidence

Intervals

- Using  $CBM = 0$

- Confidence interval for

$CBM$

- Our example: CI for

$CBM$

- Plot of Means and 95%

Confidence Regions

- Notes about These CIs

- Bonferroni Intervals

- Roy's Method

- A Truly Multivariate Follow-Up

MANOVA

Summary: PCA, MANOVA, DA

# Summary: PCA, MANOVA, DA

● Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

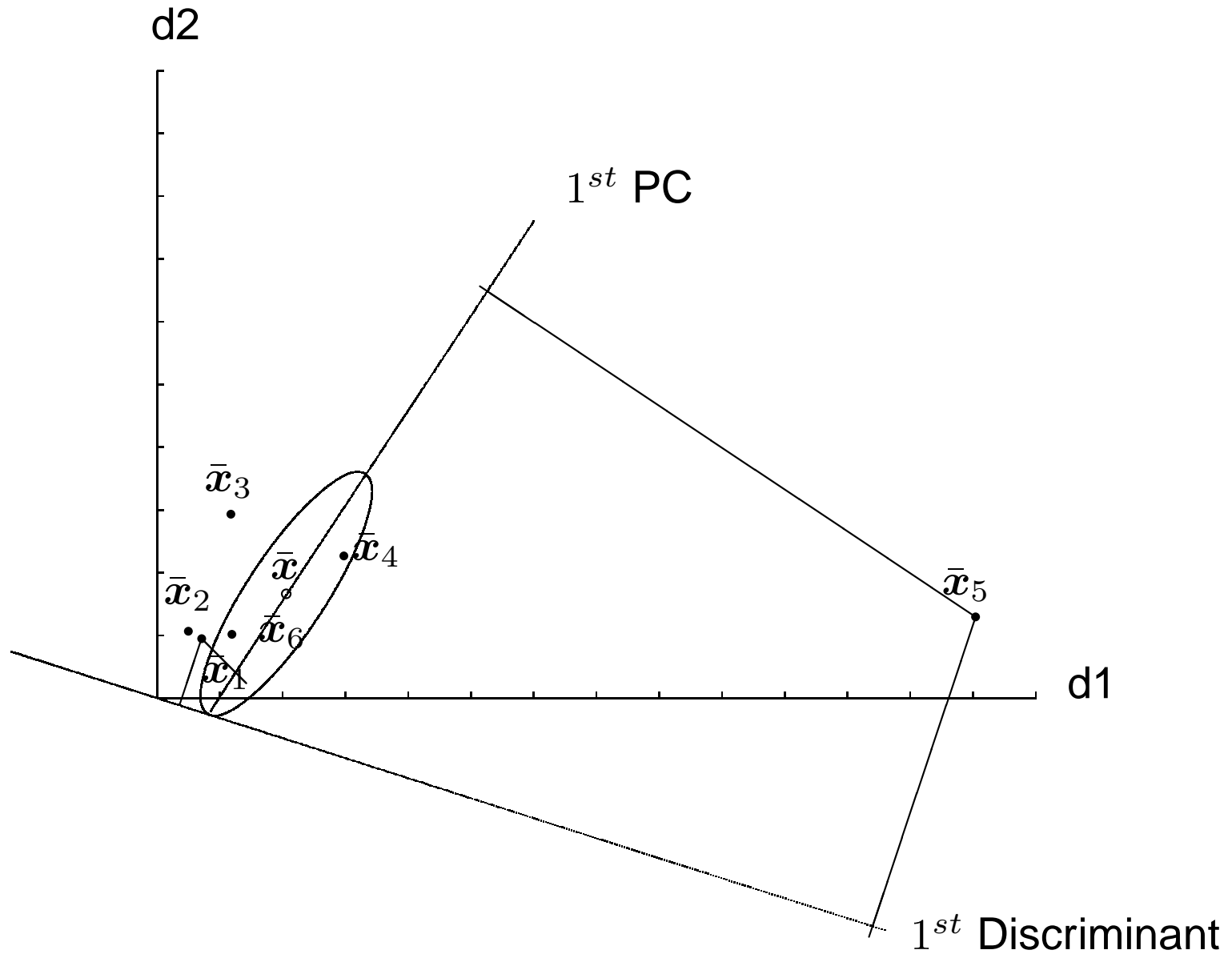
Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

● Summary: PCA, MANOVA,  
DA

SAS IML and GLM



# SAS IML and GLM

- Outline

Introduction to MANOVA

One-way ANOVA

GLM

One-Way MANOVA

Hypothesis Testing

Example: Distributed vs  
Massed Practice

Following up a Significant  
Result

Multivariate GLM

Simultaneous Confidence  
Intervals

Summary: PCA, MANOVA, DA

SAS IML and GLM

- SAS IML and GLM
- SAS GLM Output

- SAS IML code using “traditional” approach & GLM one.

- PROC GLM data=vitC;

```
class type ;
```

```
model d1 d2 = type /solution;
```

\* Note: The order of the values in the contrast are alphabetical, in this case order is bladder bronchus colon kidney rectum stomach;

```
contrast 'bronchus=colon=kidney=rectum=stomach'
```

```
type 0 1 0 0 0 -1,      type 0 0 1 -1 0 0,
```

```
type 0 1 -1 -1 0 1,      type 0 1 1 1 -4 1;
```

```
contrast 'bladder vs others' type -5 1 1 1 1 1 ;
```

```
manova h=type /printhe printe;
```

```
estimate 'b vs o' type 1 -.2 -.2 -.2 -.2 -.2;
```

```
lsmeans type;
```

```
title 'MANOVA of vitamin C and Cancer';
```

- Alternate MANOVA statement where  $M$  is entered as  $M'$ :  

```
manova h=type M=(1 0);
```

# SAS GLM Output

- Outline

Introduction to MANOVA

---

One-way ANOVA

---

GLM

---

One-Way MANOVA

---

Hypothesis Testing

---

Example: Distributed vs  
Massed Practice

---

Following up a Significant  
Result

---

Multivariate GLM

---

Simultaneous Confidence  
Intervals

---

Summary: PCA, MANOVA, DA

---

SAS IML and GLM

---

- SAS IML and GLM

- SAS GLM Output

- Univariate ANOVAs for each dependent variable.
- If requested,  $E$  (printe) and  $H$  (printh) SSCP matrices.
- $p$  characteristic roots and vectors of  $E^{-1}H$  (i.e., discriminant functions).
- Other requested statistics:
  - ◆ contrasts
  - ◆ estimates of contrasts
  - ◆ cell means
  - ◆ etc.
- Test statistics for “no overall effect” specified in MANOVA statement.