# Comparison of Several Multivariate Means

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication/
Graduate Institute of Networking and Multimedia

1

## Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)

2

## Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

3

## Outline

- Profile Analysis
- ANOVA for Repeated Measures
- Repeated Measures Designs and Growth Curves
- Perspectives and Strategy for Analyzing Multivariate Models

4

## Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)

5

## Introduction

- Extend previous ideas to handle problems involving the comparison of several mean vectors

6

## Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)

7

## Questions

- What is the paired comparison?
- How to design experiments for paired comparison?
- How to test if the population means of paired groups are different?
- How to compute the confidence interval for the difference of population means of paired groups?

8

## Questions

- How to compare population means of two populations without paired experiments?
- In such a case, how to estimate the common variance?

9

## Scenarios

- To test if the differences are significant between
  - Teaching using Power Point vs. using chalks and blackboard only
  - Drug vs. placebos
  - Processing speed of MP3 player model I of brand A vs. model G of brand B
  - Performance of students going to cram schools vs. those not

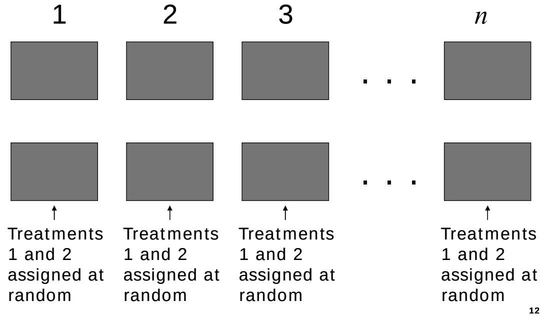
10

## Paired Comparisons

- Measurements are recorded under different sets of conditions
- See if the responses differ significantly over these sets
- Two or more treatments can be administered to the same or similar experimental units
- Compare responses to assess the effects of the treatments

11

## Experiment Design for Paired Comparisons




12

## Single Response (Univariate) Case

$$D_j = X_{j1} - X_{j2},\ j = 1, 2, \cdots, n$$
$$D_j : N(\delta, \sigma_d^2)$$
$$t = \frac{\overline{D} - \delta}{s_d / \sqrt{n}} : t_{n-1}$$
Reject $H_0 : \delta = 0$ in favor of $H_1 : \delta \neq 0$ if $|t| > t_{n-1}(\alpha/2)$

$100(1-\alpha)\%$ confidence interval for $\delta$
$$\overline{d} - t_{n-1}(\alpha/2)\frac{s_d}{\sqrt{n}} \leq \delta \leq \overline{d} + t_{n-1}(\alpha/2)\frac{s_d}{\sqrt{n}}$$

13

## Comparing Means from Two Populations

- Without explicitly controlling for unit-to-unit variability, as in the paired comparison case
- Experimental units are randomly assigned to populations
- Applicable to a more general collection of experimental units

14

## Assumptions Concerning the Structure of Data

$X_{11}, X_{12}, \cdots, X_{1n_1}$ : random sample from univariate population with mean $\mu_1$ and variance $\sigma_1^2$

$X_{21}, X_{22}, \cdots, X_{2n_2}$ : random sample from univariate population with mean $\mu_2$ and variance $\sigma_2^2$

$X_{11}, X_{12}, \cdots, X_{1n_1}$ are independent of $X_{21}, X_{22}, \cdots, X_{2n_2}$

Further assumptions when $n_1$ and $n_2$ small :

Both populations are univariate normal

$\sigma_1^2 = \sigma_2^2$

15

## Pooled Estimate of Population Variance

$$\sum_{j=1}^{n_1}(x_{j1} - \overline{x}_1)(x_{j1} - \overline{x}_1)' \approx (n_1 - 1)\sigma^2$$
$$\sum_{j=1}^{n_2}(x_{j2} - \overline{x}_2)(x_{j2} - \overline{x}_2)' \approx (n_2 - 1)\sigma^2$$
$$s_{pooled}^2 = \frac{\sum_{j=1}^{n_1}(x_{j1} - \overline{x}_1)(x_{j1} - \overline{x}_1)' + \sum_{j=1}^{n_2}(x_{j2} - \overline{x}_2)(x_{j2} - \overline{x}_2)'}{n_1 + n_2 - 2}$$
$$= \frac{n_1 - 1}{n_1 + n_2 - 2}s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2}s_2^2$$

16

## *t*-Statistics for Comparing Two Populations

$$X_{11}, X_{12}, \cdots, X_{1n_1} : N(\mu_1, \sigma^2)$$
$$X_{21}, X_{22}, \cdots, X_{2n_2} : N(\mu_2, \sigma^2)$$
$$\overline{X}_1 - \overline{X}_2 = \frac{1}{n_1}X_{11} + \cdots + \frac{1}{n_1}X_{1n_1} - \frac{1}{n_2}X_{21} + \cdots - \frac{1}{n_2}X_{2n_2}$$
$$: N(\mu_1 - \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2)$$
$$\Rightarrow t = \left(\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)\right) / \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{pooled}^2}$$

17

## Test of Hypothesis

Reject $H_0 : \mu_1 - \mu_2 = \delta_0$ in favor of $H_1 : \mu_1 - \mu_2 \neq \delta_0$

$$\text{if} \quad \left| \frac{\overline{x}_1 - \overline{x}_2 - \delta_0}{s_{pooled}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| > t_{n_1+n_2-2}\left(\frac{\alpha}{2}\right)$$

18

## Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)



## Questions

- How to make paired comparison for multivariate data?
- How to use the contrast matrix to carry out paired comparison for multivariate data?
- What is the repeated measures?
- How to test for equality of treatments in a repeated measures?



## Example 6.1: Effluent Data from Two Labs

| | Commercial lab | | State lab of hygiene | |
|---|---|---|---|---|
| Sample $j$ | $x_{1j1}$ (BOD) | $x_{1j2}$ (SS) | $x_{2j1}$ (BOD) | $x_{2j2}$ (SS) |
| 1 | 6 | 27 | 25 | 15 |
| 2 | 6 | 23 | 28 | 13 |
| 3 | 18 | 64 | 36 | 22 |
| 4 | 8 | 44 | 35 | 29 |
| 5 | 11 | 30 | 15 | 31 |
| 6 | 34 | 75 | 44 | 64 |
| 7 | 28 | 26 | 42 | 30 |
| 8 | 71 | 124 | 54 | 64 |
| 9 | 43 | 54 | 34 | 56 |
| 10 | 33 | 30 | 29 | 20 |
| 11 | 20 | 14 | 39 | 21 |

Source: Data courtesy of S. Weber.



## Multivariate Extension: Notations

$X_{1j1}$ = variable 1 under treatment 1

$X_{1j2}$ = variable 2 under treatment 1

$\vdots \qquad \vdots$

$X_{1jp}$ = variable $p$ under treatment 1

-------------------------------

$X_{2j1}$ = variable 1 under treatment 2

$X_{2j2}$ = variable 2 under treatment 2

$\vdots \qquad \vdots$

$X_{2jp}$ = variable $p$ under treatment 2



## Result 6.1

$$D_{j1} = X_{1j1} - X_{2j1}$$
$$D_{j2} = X_{1j2} - X_{2j2}$$
$$\vdots \qquad \vdots$$
$$D_{jp} = X_{1jp} - X_{2jp}$$
$$\mathbf{D}_j = [D_{j1}, D_{j2}, \cdots, D_{jp}]'$$
$$\mathbf{D}_j : N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d), \quad j = 1, 2, \cdots, n$$
$$T^2 = n(\overline{\mathbf{D}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\overline{\mathbf{D}} - \boldsymbol{\delta}) : \frac{(n-1)p}{(n-p)} F_{p,n-p}$$
$$\overline{\mathbf{D}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{D}_j, \quad \mathbf{S}_d = \frac{1}{n-1}\sum_{j=1}^{n} (\mathbf{D}_j - \overline{\mathbf{D}})(\mathbf{D}_j - \overline{\mathbf{D}})'$$



## Test of Hypotheses and Confidence Regions

$\mathbf{d}_j' = [d_{j1}, d_{j2}, \cdots, d_{jp}]$: observed differences

Reject $H_0 : \boldsymbol{\delta} = 0$ in favor of $H_1 : \boldsymbol{\delta} \neq 0$ if

$$T^2 = n\bar{\mathbf{d}}' \mathbf{S}_d^{-1} \bar{\mathbf{d}} > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

Confidence regions: $(\bar{\mathbf{d}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\bar{\mathbf{d}} - \boldsymbol{\delta}) \le \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$

$$\delta_i : \bar{d}_i \pm \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}}, \quad \delta_i : \bar{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{s_{d_i}^2}{n}}$$

## Example 6.1: Check Measurements from Two Labs

$$\bar{\mathbf{d}}=\begin{bmatrix}\bar{d}_1\\ \bar{d}_2\end{bmatrix}=\begin{bmatrix}-9.36\\ 13.27\end{bmatrix},\quad \mathbf{S}_d=\begin{bmatrix}199.26 & 88.38\\ 88.38 & 418.61\end{bmatrix}$$

$$T^2=11\begin{bmatrix}-9.36 & 13.27\end{bmatrix}\begin{bmatrix}0.0055 & -0.0012\\ -0.0012 & 0.0026\end{bmatrix}\begin{bmatrix}-9.36\\ 13.27\end{bmatrix}$$

$$=13.6>\frac{2\times 10}{9}F_{2,9}(0.05)=9.47$$

Reject $H_0:\boldsymbol{\delta}=0$

$\delta_1: -9.36\pm\sqrt{9.47}\sqrt{199.26/11}$ or $(-22.46, 3.74)$

$\delta_2: 13.27\pm\sqrt{9.47}\sqrt{418.61/11}$ or $(-5.71, 32.25)$

Both includes zero

25

## Alternative View

$$\bar{\mathbf{x}}'=\left[\bar{x}_{11},\bar{x}_{12},\cdots,\bar{x}_{1p},\bar{x}_{21},\bar{x}_{22},\cdots,\bar{x}_{2p}\right]$$

$$\mathbf{S}=\begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix}$$

$$\underset{(p\times 2p)}{\mathbf{C}}=\begin{bmatrix}1 & 0 & \cdots & 0 & | & -1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0 & | & 0 & -1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1 & | & 0 & 0 & \cdots & -1\end{bmatrix}$$

$$\mathbf{d}_j=\mathbf{C}\mathbf{x}_j,\quad \bar{\mathbf{d}}=\mathbf{C}\bar{\mathbf{x}},\quad \mathbf{S}_d=\mathbf{CSC}',\quad T^2=n\bar{\mathbf{x}}'\mathbf{C}'(\mathbf{CSC}')^{-1}\mathbf{C}\bar{\mathbf{x}}$$

26

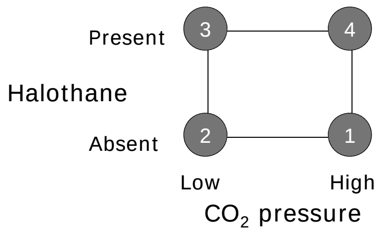
## Repeated Measures Design for Comparing Measurements

- $q$ treatments are compared with respect to a single response variable
- Each subject or experimental unit receives each treatment once over successive periods of time

27

## Example 6.2: Treatments in an Anesthetics Experiment

- 19 dogs were initially given the drug pentobarbitol followed by four treatments




28

## Example 6.2: Sleeping-Dog Data

| Dog | Treatment 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 426 | 609 | 556 | 600 |
| 2 | 253 | 236 | 392 | 395 |
| 3 | 359 | 433 | 349 | 357 |
| 4 | 432 | 431 | 522 | 600 |
| 5 | 405 | 426 | 513 | 513 |
| 6 | 324 | 438 | 507 | 539 |
| 7 | 310 | 312 | 410 | 456 |
| 8 | 326 | 326 | 350 | 504 |
| 9 | 375 | 447 | 547 | 548 |
| 10 | 286 | 286 | 403 | 422 |
| 11 | 349 | 382 | 473 | 497 |
| 12 | 429 | 410 | 488 | 547 |
| 13 | 348 | 377 | 447 | 514 |
| 14 | 412 | 473 | 472 | 446 |
| 15 | 347 | 326 | 455 | 468 |
| 16 | 434 | 458 | 637 | 524 |
| 17 | 364 | 367 | 432 | 469 |
| 18 | 420 | 395 | 508 | 531 |
| 19 | 397 | 556 | 645 | 625 |

Source: Data courtesy of Dr. J. Atlee.

29

## Contrast Matrix

$$\mathbf{X}_j=\begin{bmatrix}X_{j1}\\ X_{j2}\\ \vdots\\ X_{jq}\end{bmatrix},\quad j=1,2,\cdots,n\quad \boldsymbol{\mu}=E(\mathbf{X}_j)$$

$$\begin{bmatrix}\mu_1-\mu_2\\ \mu_1-\mu_3\\ \vdots\\ \mu_1-\mu_q\end{bmatrix}=\begin{bmatrix}1 & -1 & 0 & \cdots & 0\\ 1 & 0 & -1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 0 & 0 & \cdots & -1\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_q\end{bmatrix}=\mathbf{C}\boldsymbol{\mu}$$

30

## Test for Equality of Treatments in a Repeated Measures Design

$\mathbf{X}: N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{C}: \text{contrast matrix}$

Test of $H_0: \mathbf{C}\boldsymbol{\mu} = 0$ vs. $H_1: \mathbf{C}\boldsymbol{\mu} \neq 0$

Reject $H_0$ if

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}\mathbf{C}\bar{\mathbf{x}} > \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha)$$

31

## Example 6.2: Contrast Matrix

$$(\mu_3 + \mu_4) - (\mu_1 + \mu_2) = (\text{Halothane contrast})$$
$$(\mu_1 + \mu_3) - (\mu_2 + \mu_4) = (\text{CO}_2 \text{ contrast})$$
$$(\mu_1 + \mu_4) - (\mu_2 + \mu_3) = (\text{H} - \text{CO}_2 \text{ interaction})$$

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

32

## Example 6.2: Test of Hypotheses

$$\bar{\mathbf{x}} = \begin{bmatrix} 368.21 \\ 404.63 \\ 479.26 \\ 502.89 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 2819.29 & & & \\ 3568.42 & 7963.14 & & \\ 2943.49 & 5303.98 & 6851.32 & \\ 2295.35 & 4065.44 & 4499.63 & 4878.99 \end{bmatrix}$$

$$\mathbf{C}\bar{\mathbf{x}} = \begin{bmatrix} 209.31 \\ -60.05 \\ -12.79 \end{bmatrix}, \quad \mathbf{CSC}' = \begin{bmatrix} 9432.32 & 1098.92 & 927.62 \\ 1098.92 & 5195.84 & 914.54 \\ 927.62 & 914.54 & 7557.44 \end{bmatrix}$$

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{x}}) = 116$$

$$\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(0.05) = 10.94$$

Reject $H_0: \mathbf{C}\boldsymbol{\mu} = 0$

33

## Example 6.2: Simultaneous Confidence Intervals

Contrast of halothane influence

$$(\bar{x}_3 + \bar{x}_4) - (\bar{x}_1 + \bar{x}_2) \pm \sqrt{\frac{18(3)}{16} F_{3,16}(0.05)} \sqrt{\frac{\mathbf{c}_1'\mathbf{S}\mathbf{c}_1}{19}} = 209.31 \pm 73.70$$

$\text{CO}_2$ pressure influence

$$-60.05 \pm \sqrt{10.94}\sqrt{\frac{5195.84}{19}} = -60.5 \pm 54.70$$

H - $\text{CO}_2$ "interaction"

$$-12.79 \pm \sqrt{10.94}\sqrt{\frac{7557.44}{19}} = -12.79 \pm 65.97$$

34

## Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)

35

## Questions

- How to compare mean vectors from two populations, not forming paired comparison groups?
- How to pool covariance matrices from two populations?
- How to find simultaneous confidence intervals for comparing mean vectors from two populations?

36

## Questions

- What is the multivariate Behrens-Fisher problem and how to solve it?



## Comparing Mean Vectors from Two Populations

- Populations: Sets of experiment settings
- Without explicitly controlling for unit-to-unit variability, as in the paired comparison case
- Experimental units are randomly assigned to populations
- Applicable to a more general collection of experimental units



## Assumptions Concerning the Structure of Data

$\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1n_1}$ : random sample from $p-$variate population with mean vector $\boldsymbol{\mu}_1$ and covariance $\boldsymbol{\Sigma}_1$

$\mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2n_2}$ : random sample from $p-$variate population with mean vector $\boldsymbol{\mu}_2$ and covariance $\boldsymbol{\Sigma}_2$

$\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1n_1}$ are independent of $\mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2n_2}$

Further assumptions when $n_1$ and $n_2$ small:

Both populations are multivariate normal

$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$



## Pooled Estimate of Population Covariance Matrix

$$\sum_{j=1}^{n_1} \left(\mathbf{x}_{j1} - \bar{\mathbf{x}}_1\right)\left(\mathbf{x}_{j1} - \bar{\mathbf{x}}_1\right)' \approx (n_1 - 1)\boldsymbol{\Sigma}$$

$$\sum_{j=1}^{n_2} \left(\mathbf{x}_{j2} - \bar{\mathbf{x}}_2\right)\left(\mathbf{x}_{j2} - \bar{\mathbf{x}}_2\right)' \approx (n_2 - 1)\boldsymbol{\Sigma}$$

$$\mathbf{S}_{pooled} = \frac{\sum_{j=1}^{n_1} \left(\mathbf{x}_{j1} - \bar{\mathbf{x}}_1\right)\left(\mathbf{x}_{j1} - \bar{\mathbf{x}}_1\right)' + \sum_{j=1}^{n_2} \left(\mathbf{x}_{j2} - \bar{\mathbf{x}}_2\right)\left(\mathbf{x}_{j2} - \bar{\mathbf{x}}_2\right)'}{n_1 + n_2 - 2}$$

$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2$$



## Result 6.2

$$\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1n_1} : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

$$\mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2n_2} : N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

$$\Rightarrow T^2 = \left[\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{pooled}\right]^{-1} \left[\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]$$

is distributed as

$$\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}$$



## Proof of Result 6.2

$$\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 = \frac{1}{n_1}\mathbf{X}_{11} + \cdots + \frac{1}{n_1}\mathbf{X}_{1n_1} - \frac{1}{n_2}\mathbf{X}_{21} + \cdots - \frac{1}{n_2}\mathbf{X}_{2n_2}$$

$$: N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right)$$

$$(n_1 - 1)\mathbf{S}_1 : W_{n_1 - 1}(\boldsymbol{\Sigma}), \quad (n_2 - 1)\mathbf{S}_2 : W_{n_2 - 1}(\boldsymbol{\Sigma})$$

$$(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 : W_{n_1 + n_2 - 2}(\boldsymbol{\Sigma})$$

$$T^2 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \left[\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]' \mathbf{S}_{pooled}^{-1}$$

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \left[\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]$$

$$= N_p(0, \boldsymbol{\Sigma})' \left[\frac{W_{n_1 + n_2 - 2}(\boldsymbol{\Sigma})}{n_1 + n_2 - 2}\right]^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma}) : \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}$$

# Wishart Distribution

$$w_{n-1}(\mathbf{A}\,|\,\mathbf{\Sigma}) = \frac{|\mathbf{A}|^{(n-p-2)/2} e^{-\mathrm{tr}\left[\mathbf{A}\mathbf{\Sigma}^{-1}\right]/2}}{2^{p(n-1)/2}\pi^{p(p-1)/4}|\mathbf{\Sigma}|^{(n-1)/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-i)\right)}$$

$\mathbf{A}$ : positive definite

Properties :

$\mathbf{A}_1 : W_{m_1}(\mathbf{A}_1\,|\,\mathbf{\Sigma}), \quad \mathbf{A}_2 : W_{m_2}(\mathbf{A}_2\,|\,\mathbf{\Sigma}) \Rightarrow$

$\quad \mathbf{A}_1 + \mathbf{A}_2 : W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2\,|\,\mathbf{\Sigma})$

$\mathbf{A} : W_m(\mathbf{A}\,|\,\mathbf{\Sigma}) \Rightarrow \mathbf{CAC}' : W_m(\mathbf{CAC}'\,|\,\mathbf{C\Sigma\Sigma}')$

43

# Test of Hypothesis

Reject $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$ in favor of $H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \boldsymbol{\delta}_0$

if $\quad T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{pooled}\right]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0)$

$\quad > \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)$

Note $E(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$

$\quad \mathrm{Cov}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$

$\quad = \mathrm{Cov}(\overline{\mathbf{X}}_1) - \mathrm{Cov}(\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2) - \mathrm{Cov}(\overline{\mathbf{X}}_2, \overline{\mathbf{X}}_1) + \mathrm{Cov}(\overline{\mathbf{X}}_2)$

$\quad = \mathrm{Cov}(\overline{\mathbf{X}}_1) + \mathrm{Cov}(\overline{\mathbf{X}}_2) = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{\Sigma}$

44

# Example 6.3: Comparison of Soaps Manufactured in Two Ways

$n_1 = n_2 = 50$

$\bar{\mathbf{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$

$\mathbf{S}_{pooled} = \frac{49}{98}\mathbf{S}_1 + \frac{49}{98}\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}, \quad \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 = \begin{bmatrix} -1.9 \\ 0.2 \end{bmatrix}$

Eigenvalues and eigenvectors of $\mathbf{S}_{pooled}$ :

$\quad \lambda_1 = 5.303, \quad \mathbf{e}_1 = [0.290 \quad 0.957]'$

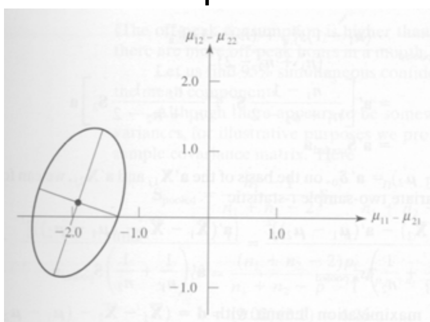
$\quad \lambda_2 = 1.697, \quad \mathbf{e}_1 = [0.957 \quad -0.290]'$

$\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(0.05) = 0.25$

$\sqrt{\lambda_1}\sqrt{0.25} = 1.15, \quad \sqrt{\lambda_2}\sqrt{0.25} = 0.65$

45

# Example 6.3




46

# Result 6.3: Simultaneous Confidence Intervals

$c^2 = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)$

$\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{pooled}\mathbf{a}}$

will cover $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ for all $\mathbf{a}$

In particular, $\mu_{1i} - \mu_{2i}$ will be covered by

$(\overline{X}_{1i} - \overline{X}_{2i}) \pm c\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{ii,pooled}}$

47

# Example 6.4: Electrical Usage of Homeowners with and without ACs

$\bar{\mathbf{x}}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, \quad n_1 = 45$

$\bar{\mathbf{x}}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, \quad n_1 = 55$

$\mathbf{S}_{pooled} = \frac{n_1 - 1}{n_1+n_2-2}\mathbf{S}_1 + \frac{n_2-1}{n_1+n_2-2}\mathbf{S}_2$

$\quad = \begin{bmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{bmatrix}$

$c^2 = \frac{98(2)}{97}F_{2,97}(0.05) = 6.26$

48

## Example 6.4: Electrical Usage of Homeowners with and without ACs

95% simultaneous confidence intervals

$$\mu_{11}-\mu_{21}:(204.4-130.0)\pm\sqrt{6.26}\sqrt{\left(\frac{1}{45}+\frac{1}{55}\right)10963.7}$$

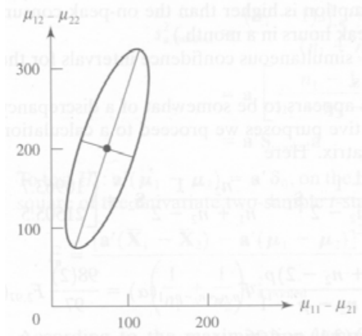
or $21.7 \le \mu_{11}-\mu_{21} \le 127.1$

$$\mu_{12}-\mu_{22}:(556.6-355.0)\pm\sqrt{6.26}\sqrt{\left(\frac{1}{45}+\frac{1}{55}\right)63661.3}$$

or $74.7 \le \mu_{12}-\mu_{22} \le 328.5$

49

## Example 6.4: 95% Confidence Ellipse




50

## Bonferroni Simultaneous Confidence Intervals

$$\mu_{1i}-\mu_{2i}:(\bar{x}_1-\bar{x}_2)\pm t_{n_1+n_2-2}\left(\frac{\alpha}{2p}\right)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{ii,pooled}}$$

51

## Result 6.4

$n_1-p$ and $n_2-p$ are large

100% confidence ellipsoid for $\boldsymbol{\mu}_1-\boldsymbol{\mu}_2$ :

$$[\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]'\left[\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right]^{-1}$$
$$[\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]\le\chi_p^2(\alpha)$$

Simultaneous confidence intervals for $\mathbf{a}'(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$:

$$\mathbf{a}'(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)\pm\sqrt{\chi_p^2(\alpha)}\sqrt{\mathbf{a}'\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)\mathbf{a}}$$

52

## Proof of Result 6.4

$$E(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2)=\boldsymbol{\mu}_1-\boldsymbol{\mu}_2$$

$$\mathrm{Cov}(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2)=\mathrm{Cov}(\bar{\mathbf{X}}_1)+\mathrm{Cov}(\bar{\mathbf{X}}_2)=\frac{1}{n_1}\boldsymbol{\Sigma}_1+\frac{1}{n_2}\boldsymbol{\Sigma}_2$$

$$\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2:\text{nearly } N_p\left(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2,\frac{1}{n_1}\boldsymbol{\Sigma}_1+\frac{1}{n_2}\boldsymbol{\Sigma}_2\right)$$

$$[\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]'\left(\frac{1}{n_1}\boldsymbol{\Sigma}_1+\frac{1}{n_2}\boldsymbol{\Sigma}_2\right)^{-1}$$
$$[\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]:\chi_p^2$$

$$\boldsymbol{\Sigma}_1\sim\mathbf{S}_1,\quad\boldsymbol{\Sigma}_2\sim\mathbf{S}_2$$

53

## Remark

If $n_1=n_2=n$

$$\frac{n-1}{n+n-2}=\frac{1}{2}$$

$$\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2=\frac{1}{n}(\mathbf{S}_1+\mathbf{S}_2)$$

$$=\frac{(n-1)\mathbf{S}_1+(n-1)\mathbf{S}_2}{n+n-2}\left(\frac{1}{n}+\frac{1}{n}\right)=\mathbf{S}_{pooled}\left(\frac{1}{n}+\frac{1}{n}\right)$$

54

# Example 6.5

Example 6.4 Data

$$\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2=\begin{bmatrix}464.17 & 886.08\\ 886.08 & 2642.15\end{bmatrix}$$

$$\mu_{11}-\mu_{21}: 74.4\pm\sqrt{5.99}\sqrt{464.17} \text{ or } (21.7, 127.1)$$

$$\mu_{12}-\mu_{22}: 201.6\pm\sqrt{5.99}\sqrt{2642.15} \text{ or } (75.8, 327.4)$$

$$H_0: \boldsymbol{\mu}_1-\boldsymbol{\mu}_2=0$$

$$T^2=[\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2]'\left[\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right]^{-1}[\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2]=15.66>\chi_2^2(0.05)=5.99$$

$$\text{Critical linear combination}: \left[\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right]^{-1}[\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2]=\begin{bmatrix}0.041\\0.063\end{bmatrix}$$

55

# Multivariate Behrens-Fisher Problem

- Test $H_0$: $\mu_1$-$\mu_2$=0
- Population covariance matrices are unequal
- Sample sizes are not large
- Populations are multivariate normal
- Both sizes are greater than the number of variables

56

# Approximation of $T^2$ Distribution

$$T^2=\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)'\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)$$

$$=\frac{\nu p}{\nu-p+1}F_{p,\nu-p+1}$$

$$\nu=\frac{p+p^2}{\sum_{i=1}^{2}\frac{1}{n_i}\left\{\operatorname{tr}\left[\left(\frac{1}{n_i}\mathbf{S}_i\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right)^2\right]+\left(\operatorname{tr}\left[\frac{1}{n_i}\mathbf{S}_i\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right]\right)^2\right\}}$$

$$\min(n_1,n_2)\le\nu\le n_1+n_2$$

57

# Confidence Region

$$\left(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)'\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\left(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)$$

$$\le\frac{\nu p}{\nu-p+1}F_{p,\nu-p+1}(\alpha)$$

58

# Example 6.6

- Example 6.4 data

$$\frac{1}{n_1}\mathbf{S}_1\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}=\begin{bmatrix}0.776 & -0.060\\ -0.092 & 0.646\end{bmatrix}$$

$$\frac{1}{n_2}\mathbf{S}_2\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}=\begin{bmatrix}0.224 & -0.060\\ 0.092 & 0.354\end{bmatrix}$$

$$\nu=77.6$$

$$\frac{\nu p}{\nu-p+1}F_{p,\nu-p+1}(0.05)=\frac{155.2}{76.6}\times 3.12=6.32$$

$$T^2=15.66>6.32,\quad H_0:\boldsymbol{\mu}_1-\boldsymbol{\mu}_2=0 \text{ is rejected}$$

59

# Outline

- Introduction
- Comparison of Univariate Mean
- Paired Comparisons and a Repeated Measures Design
- Comparing Mean Vectors from Two Populations
- Comparison of Several Univariate Population Mean (One-Way ANOVA)

60

## Questions

- Why paired comparisons are not good ways to compare several population means?
- How to compute summed squares (between)?
- How to compute summed squares (within)?
- How to compute summed squares (total)?

61

## Questions

- How to calculate the degrees of freedom for summed squares (between)?
- How to calculate the degrees of freedom for summed squares (within)?
- How to calculate the degrees of freedom for summed squares (total)?

62

## Questions

- How to compute the F value for testing of the null hypothesis?
- How are the three kinds of summed squares related?
- How to explain the geometric meaning of the degrees of freedom for a treatment vector?
- What is an ANOVA table?

63

## Scenarios

- To test if the following statements are plausible
  - Music compressed by four MP3 compressors are with the same quality
  - Three new drugs are all as effective as a placebo
  - Four brands of beer are equally tasty
  - Lectures, group studying, and computer assisted instruction are equally effective for undergraduate students

64

## Comparing Four MP3 Compressors

- Test four brands, *A*, *B*, *C*, *D*
- 10 subjects each brand (40 in total) to provide a satisfaction rating on a 10-point scale
- Assume that the rating to each brand is a normal distribution, but all four distributions are with the same variance

65

## Hypotheses

- Null hypothesis

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

- Alternative hypothesis

$$H_1 : \textit{Not} \text{ all the } \mu\text{s are equal}$$

66

## Problem of Using a *t*-Test

- Must compare two brands at a time
- There are 6 possible comparisons
- Each has a 0.05 chance of being significant by chance
- Overall chance of significant result, even when no difference exist, approaches $1-(0.95)^6 \sim 0.26$



## Sample Data

| Subject | A | B | C | D |
|---|---|---|---|---|
| 1 | 4 | 5 | 7 | 2 |
| 2 | 4 | 5 | 8 | 1 |
| 3 | 5 | 6 | 7 | 2 |
| 4 | 5 | 6 | 9 | 3 |
| 5 | 6 | 7 | 6 | 3 |
| 6 | 3 | 6 | 3 | 4 |
| 7 | 4 | 4 | 2 | 5 |
| 8 | 4 | 5 | 2 | 4 |
| 9 | 3 | 6 | 2 | 4 |
| 10 | 4 | 3 | 3 | 3 |
| Mean | 4.2 | 5.3 | 4.9 | 3.1 |

Grand mean: 4.375

*Adapted from: G. R. Norman and D. L. Streiner, *Biostatistics*, 3rd ed.



## Thinking in Terms of Signals and Noises

- Signals
  - Overall difference among the means of the groups
  - Sum of all the squared differences between group means and the overall means
- Noises
  - Overall variability within the groups
  - Sum of all the squared differences between individual data and their group means



## Sum of Squares (Between)

$$SS(between) = n\sum\left(\bar{x}_\ell - \bar{x}\right)^2$$

$$SS(between) = 10[(4.2-4.375)^2 + (5.3-4.375)^2 + (4.9-4.375)^2 + (3.1-4.375)^2] = 27.875$$



## Sum of Squares (Within)

$$SS(within) = \sum_\ell \sum_j \left(x_{\ell j} - \bar{x}_\ell\right)^2$$

$$\begin{aligned} SS(within) = {} & (4-4.2)^2 + (4-4.2)^2 + \cdots + (4-4.2)^2 + \\ & (5-5.3)^2 + (5-5.3)^2 + \cdots + (3-5.3)^2 + \\ & (7-4.9)^2 + (8-4.9)^2 + \cdots + (3-4.9)^2 + \\ & (2-3.1)^2 + (1-3.1)^2 + \cdots + (3-3.1)^2 \\ & [40 \text{ terms}] \\ = {} & 101.50 \end{aligned}$$



## Sum of Squares (Total)

$$SS(total) = \sum_\ell \sum_j (x_{\ell j} - \bar{x})^2$$

$$x_{\ell j} - \bar{x} = (x_{\ell j} - \bar{x}_\ell) + (\bar{x}_\ell - \bar{x})$$

$$\left(x_{\ell j} - \bar{x}\right)^2 = (x_{\ell j} - \bar{x}_\ell)^2 + 2(x_{\ell j} - \bar{x}_\ell)(\bar{x}_\ell - \bar{x}) + (\bar{x}_\ell - \bar{x})^2$$

$$\sum_j (x_{\ell j} - \bar{x}_\ell) = 0$$

$$\sum_j \left(x_{\ell j} - \bar{x}\right)^2 = \sum_j (x_{\ell j} - \bar{x}_\ell)^2 + n(\bar{x}_\ell - \bar{x})^2$$

$$SS(total) = SS(within) + SS(between)$$

## Sum of Squares (Total)

$$
\begin{aligned}
SS(total) = {} & (4-4.375)^2+(4-4.375)^2+\cdots+(4-4.375)^2+ \\
& (5-4.375)^2+(5-4.375)^2+\cdots+(3-4.375)^2+ \\
& (7-4.375)^2+(8-4.375)^2+\cdots+(3-4.375)^2+ \\
& (2-4.375)^2+(1-4.375)^2+\ldots+(3-4.375)^2 \\
& [40 \text{ terms}] \\
= {} & 129.375 = 101.50 + 27.875
\end{aligned}
$$

73

## $\chi^2$ Distribution

$X_1 : N(\mu_1, \sigma_1^2),\quad X_2 : N(\mu_2, \sigma_2^2),\quad \cdots,$

$X_\nu : N(\mu_\nu, \sigma_\nu^2);\quad Z_i = \dfrac{X_i - \mu_i}{\sigma_i} : N(0,1)$

$$\chi^2 = \sum_{i=1}^{\nu} \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2, \quad \nu : \text{degrees of freedom (d.f.)}$$

$$f_\nu(\chi^2) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} \left(\chi^2\right)^{\nu/2-1} e^{-\chi^2/2}, & \chi^2 > 0 \\ 0, & \chi^2 \le 0 \end{cases}$$

(Gamma distribution with $\alpha = \nu/2$)

74

## Distribution of Sum of Squares

$X : N(\mu, \sigma^2)$

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left( X_j - \overline{X} \right)^2$$

$(n-1)\dfrac{S^2}{\sigma^2} : \chi^2$ distribution with $n-1$ degrees of freedom

[proved by moment generating function, see
P. G. Hoel, *Introduction to Mathematical Statistics,* 5th ed.,
John Wiley & Sons, 1984, p. 281]

75

## Distribution of Sum of Squares

$$SS(within) = \sum_{\ell=1}^{g} \sum_{j=1}^{n} \left( x_{\ell j} - \overline{x} \right)^2$$

$\dfrac{SS(within)}{\sigma^2} : \chi^2$ distribution with

degree of freedom $df(within) = gn - 1$

$$SS(between) = \sum_{\ell=1}^{g} \left( \overline{x}_\ell - \overline{x} \right)^2$$

$$\overline{x} = \frac{1}{gn} \sum_{\ell=1}^{g} \sum_{j=1}^{n} x_{\ell j} = \frac{1}{g} \sum_{\ell=1}^{g} \left( \frac{1}{n} \sum_{j=1}^{n} x_{\ell j} \right) = \frac{1}{g} \sum_{\ell=1}^{g} \overline{x}_\ell$$

$\dfrac{SS(between)}{\sigma^2} : \chi^2$ distribution with

degree of freedom $df(between) = g - 1$

76

## *F*-Distribution

$\chi_1^2, \chi_2^2$ : independent, with d.f. $f_1$ and $f_2$, respectively

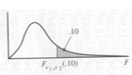
$$F = \frac{\chi_1^2 / f_1}{\chi_2^2 / f_2}, F > 0$$

$$f(F) = \frac{\Gamma\left( \dfrac{f_1 + f_2}{2} \right)}{\Gamma\left( \dfrac{f_1}{2} \right) \Gamma\left( \dfrac{f_2}{2} \right)} \left( \frac{f_1}{f_2} \right)^{\frac{f_1}{2}} \frac{F^{\frac{f_1}{2}-1}}{\left( 1 + \dfrac{f_1 F}{f_2} \right)^{(f_1+f_2)/2}}$$

$: F_{f_1, f_2}$

77

## *F*-Distribution




| $\nu_2$ \ $\nu_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 25 | 30 | 40 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 39.86 | 49.50 | 53.59 | 55.83 | 57.24 | 58.20 | 58.91 | 59.44 | 59.86 | 60.19 | 60.71 | 61.22 | 61.74 | 62.05 | 62.26 | 62.53 | 62.79 |
| 2 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 | 9.39 | 9.41 | 9.42 | 9.44 | 9.45 | 9.46 | 9.47 | 9.47 |
| 3 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 | 5.23 | 5.22 | 5.20 | 5.18 | 5.17 | 5.17 | 5.16 | 5.15 |
| 4 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 | 3.92 | 3.90 | 3.87 | 3.84 | 3.83 | 3.82 | 3.80 | 3.79 |
| 5 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 | 3.30 | 3.27 | 3.24 | 3.21 | 3.19 | 3.17 | 3.16 | 3.14 |
| 6 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 | 2.94 | 2.90 | 2.87 | 2.84 | 2.81 | 2.80 | 2.78 | 2.76 |
| 7 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 | 2.70 | 2.67 | 2.63 | 2.59 | 2.57 | 2.56 | 2.54 | 2.51 |
| 8 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 | 2.54 | 2.50 | 2.46 | 2.42 | 2.40 | 2.38 | 2.36 | 2.34 |
| 9 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 | 2.42 | 2.38 | 2.34 | 2.30 | 2.27 | 2.25 | 2.23 | 2.21 |
| 10 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 | 2.28 | 2.24 | 2.20 | 2.17 | 2.16 | 2.13 | 2.11 |
| 11 | 3.23 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 | 2.21 | 2.17 | 2.12 | 2.10 | 2.08 | 2.05 | 2.03 |

78

## Distribution of *F*

$$F = \frac{SS(between)/df(between)}{SS(within)/df(within)}:$$

$F$ distribution of degree of freedoms

$df(between)$ and $df(within)$



## Expected Values of Sum of Squares

if no difference between groups

$E[SS(between)/df(between)] = \sigma_{err}^2$

if no difference within groups

$E[SS(between)/df(between)] = n\sigma_{bet}^2$

if both differences can happen

$E[SS(between)/df(between)] = n\sigma_{bet}^2 + \sigma_{err}^2$

Thus, if $H_0$ is invalid

$$\frac{E[SS(between)/df(between)]}{E[SS(within)/df(within)]} = \frac{n\sigma_{bet}^2 + \sigma_{err}^2}{\sigma_{err}^2} > 1$$

$F > 1$



## Degrees of Freedom

$$\begin{aligned} df(between) &= g-1 = 4-1 = 3 \\ df(within) &= g(n-1) = 4(10-1) = 36 \\ df(total) &= gn-1 = gn-g+g-1 \\ &= df(within) + df(between) \\ &= 40-1 = 39 = 36+3 \end{aligned}$$



## ANOVA Summary

| Source | Sum of Squares | *df* | Mean square | F |
|---|---|---|---|---|
| Between | 27.875 | 3 | 9.292 | 3.296 |
| Within | 101.500 | 36 | 2.819 | |
| Total | 129.375 | 39 | | |



## Hypothesis Testing

$F = 3.296 > F_{3,36}(0.05) = 2.86$

reject $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$

at 0.05 significance level



## Univariate ANOVA

$X_{\ell 1}, X_{\ell 2}, \cdots, X_{\ell n_\ell}$ : random sample from $N(\mu_\ell, \sigma^2)$

$\ell = 1, 2, \cdots, g$

Null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_g$

Reparameterization

$\mu_\ell = \mu + \tau_\ell$

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$

$X_{\ell j} = \mu + \tau_\ell + e_{\ell j}, \quad e_{\ell j} : N(0, \sigma^2), \quad \sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$

$x_{\ell j} = \bar{x} + (\bar{x}_\ell - \bar{x}) + (x_{\ell j} - \bar{x}_\ell)$

# Univariate ANOVA

$$\left(x_{\ell j}-\bar{x}\right)^2=\left(\bar{x}_\ell-\bar{x}\right)^2+\left(x_{\ell j}-\bar{x}_\ell\right)^2+2\left(\bar{x}_\ell-\bar{x}\right)\left(x_{\ell j}-\bar{x}_\ell\right)$$

$$\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}_\ell\right)=0$$

$$\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}\right)^2=n_\ell\left(\bar{x}_\ell-\bar{x}\right)^2+\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}_\ell\right)^2$$

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}\right)^2=\sum_{\ell=1}^{g}n_\ell\left(\bar{x}_\ell-\bar{x}\right)^2+\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}_\ell\right)^2$$

$$(\mathrm{SS}_{cor})=(\mathrm{SS}_{tr})+(\mathrm{SS}_{res})$$

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}x_{\ell j}^2=\left(n_1+n_2+\cdots+n_\ell\right)\bar{x}^2+\sum_{\ell=1}^{g}n_\ell\left(\bar{x}_\ell-\bar{x}\right)^2+\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}\left(x_{\ell j}-\bar{x}_\ell\right)^2$$

$$(\mathrm{SS}_{obs})=(\mathrm{SS}_{mean})+(\mathrm{SS}_{tr})+(\mathrm{SS}_{res})$$



# Concept of Degrees of Freedom

$\mathbf{y}'=\left[x_{11},\cdots,x_{1n_1},x_{21},\cdots,x_{2n_2},\cdots,x_{gn_g}\right]$: d.f. $n=n_1+n_2+\cdots+n_g$

Treatment vector

$$\left(\bar{x}_1-\bar{x}\right)\begin{bmatrix}1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\\ 0\\ \vdots\\ 0\end{bmatrix}+\left(\bar{x}_2-\bar{x}\right)\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}+\cdots+\left(\bar{x}_g-\bar{x}\right)\begin{bmatrix}0\\ \vdots\\ 0\\ 0\\ \vdots\\ 0\\ 1\\ \vdots\\ 1\end{bmatrix}$$

$$=\left(\bar{x}_1-\bar{x}\right)\mathbf{u}_1+\left(\bar{x}_2-\bar{x}\right)\mathbf{u}_2+\cdots+\left(\bar{x}_g-\bar{x}\right)\mathbf{u}_g$$



# Concept of Degrees of Freedom

$\mathbf{1}=[1,\cdots,1]=\mathbf{u}_1+\mathbf{u}_2+\cdots+\mathbf{u}_g$

Treatment vector and $\mathbf{1}$ are all on the hyperplane

spanned by $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_g$ : d.f. $g$

$\mathbf{1}$ is perpendicular to the treatment vector

$\therefore$ mean vector $\bar{x}\mathbf{1}$ : d.f. $g-1$

Residual vector

$\mathbf{e}=\mathbf{y}-\bar{x}\mathbf{1}-\left[\left(\bar{x}_1-\bar{x}\right)\mathbf{u}_1+\left(\bar{x}_2-\bar{x}\right)\mathbf{u}_2+\cdots+\left(\bar{x}_g-\bar{x}\right)\mathbf{u}_g\right]$

perpendicular to the hyperplane spanned by $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_g$

$\therefore$ d.f. of $\mathbf{e}$ : $n-g$



# Univariate ANOVA

ANOVA TABLE FOR COMPARING UNIVARIATE POPULATION MEANS

| Source of variation | Sum of squares (SS) | Degrees of freedom (d.f.) |
|---|---|---|
| Treatments | $\mathrm{SS}_{tr}=\sum_{\ell=1}^{g}n_\ell(\bar{x}_\ell-\bar{x})^2$ | $g-1$ |
| Residual (Error) | $\mathrm{SS}_{res}=\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(x_{\ell j}-\bar{x}_\ell)^2$ | $\sum_{\ell=1}^{g}n_\ell-g$ |
| Total (corrected for the mean) | $\mathrm{SS}_{cor}=\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(x_{\ell j}-\bar{x})^2$ | $\sum_{\ell=1}^{g}n_\ell-1$ |



# Univariate ANOVA

Reject $H_0:\tau_1=\tau_2=\cdots=\tau_g=0$ at level $\alpha$ if

$$F=\frac{\mathrm{SS}_{tr}/(g-1)}{\mathrm{SS}_{res}/\left(\sum_{\ell=1}^{g}n_\ell-g\right)}>F_{g-1,\sum n_\ell-g}(\alpha)$$

$$\frac{1}{1+\mathrm{SS}_{tr}/\mathrm{SS}_{res}}=\frac{\mathrm{SS}_{res}}{\mathrm{SS}_{res}+\mathrm{SS}_{tr}}$$



# Examples 6.7 & 6.8

$$\begin{pmatrix}9&6&9\\0&2&\\3&1&2\end{pmatrix}=\begin{pmatrix}4&4&4\\4&4&\\4&4&4\end{pmatrix}+\begin{pmatrix}4&4&4\\-3&-3&\\-2&-2&-2\end{pmatrix}+\begin{pmatrix}1&-2&1\\-1&1&\\1&-1&0\end{pmatrix}$$

$\mathrm{SS}_{obs}=216,\ \mathrm{SS}_{mean}=128$

$\mathrm{SS}_{tr}=78,$ d.f. $=3-1=2$

$\mathrm{SS}_{res}=10,$ d.f. $=(3+2+3)-3=5$

$$F=\frac{\mathrm{SS}_{tr}/(g-1)}{\mathrm{SS}_{res}/\left(\sum n_\ell-g\right)}=\frac{78/2}{10/5}=19.5>F_{2,5}(0.01)=13.27$$

$H_0:\tau_1=\tau_2=\tau_3=0$ is rejected at the 1% level

## Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

91

## Questions

- What is the one-way MANOVA table?
- How to compute Wilk's lambda for MANOVA?
- How to test the equality of several mean vectors from the Wilk's lambda?
- How to test the equality of several mean vectors for large sample size?

92

## Questions

- What are other statistics used in statistical software package for one-way MANOVA?

93

## Scenario: Example 6.10, Nursing Home Data

- Nursing homes can be classified by the owners: private (271), non-profit (138), government (107)
- Costs: nursing labor, dietary labor, plant operation and maintenance labor, housekeeping and laundry labor
- To investigate the effects of ownership on costs

94

## One-Way MANOVA

Population 1: $\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1n_1}$

Population 2: $\mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2n_2}$

$\vdots \qquad \vdots$

Population $g$: $\mathbf{X}_{g1}, \mathbf{X}_{g2}, \cdots, \mathbf{X}_{gn_g}$

MANOVA (Multivariate ANalysis Of VAriance) is used to investigate whether the population mean vectors are the same, and, if not, which mean components differ significantly

95

## Assumptions about the Data

$\mathbf{X}_{\ell 1}, \mathbf{X}_{\ell_2}, \cdots, \mathbf{X}_{\ell n_\ell}$ : random sample from a population with mean $\boldsymbol{\mu}_\ell$, $\ell = 1, 2, \cdots, g$

Random sample from different populations are independent

All populations have a common covariance matrix $\boldsymbol{\Sigma}$

Each population is multivariate normal

96

## MANOVA

$$\mathbf{X}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \mathbf{e}_{\ell j};\ j = 1, 2, \cdots, n_\ell;\ \ell = 1, 2, \cdots, g$$

$$\mathbf{e}_{\ell j} : N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu} : \text{overall mean (level)}$$

$$\boldsymbol{\tau}_\ell : \ell\text{th treatment effect, } \sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \mathbf{0}$$

$$\mathbf{x}_{\ell j} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}) + (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell) = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\tau}}_\ell + \hat{\mathbf{e}}_{\ell j}$$

$$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})' = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$$

$$+\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' = \mathbf{B} + \mathbf{W}$$

97

## MANOVA

MANOVA TABLE FOR COMPARING POPULATION MEAN VECTORS

| Source of variation | Matrix of sum of squares and cross products (SSP) | Degrees of freedom (d.f.) |
|---|---|---|
| Treatment | $\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$ | $g - 1$ |
| Residual (Error) | $\mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'$ | $\sum_{\ell=1}^{g} n_\ell - g$ |
| Total (corrected for the mean) | $\mathbf{B} + \mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'$ | $\sum_{\ell=1}^{g} n_\ell - 1$ |

98

## MANOVA

$$\mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'$$

$$= (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \cdots + (n_g - 1)\mathbf{S}_g$$

Reject $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = 0$ if Wilk's lambda

$$\Lambda^* = \frac{|W|}{|B+W|} = \frac{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'\right|}{\left|\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'\right|}$$

is too small

99

## Distribution of Wilk's Lambda

| No. of variables | No. of groups | Sampling distribution for multivariate normal data |
|---|---|---|
| $p = 1$ | $g \ge 2$ | $\left(\frac{\Sigma n_\ell - g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$ |
| $p = 2$ | $g \ge 2$ | $\left(\frac{\Sigma n_\ell - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$ |
| $p \ge 1$ | $g = 2$ | $\left(\frac{\Sigma n_\ell - p - 1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$ |
| $p \ge 1$ | $g = 3$ | $\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\Sigma n_\ell - p - 2)}$ |

100

## Test of Hypothesis for Large Size

If $H_0$ is true and $\sum n_\ell = n$ is large,

$$-\left(n - 1 - \frac{p+g}{2}\right)\ln \Lambda^* : \chi^2_{p(g-1)}$$

Reject $H_0$ at significance level $\alpha$ if

$$-\left(n - 1 - \frac{p+g}{2}\right)\ln\left(\frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|}\right) > \chi^2_{p(g-1)}(\alpha)$$

101

## Popular MANOVA Statistics Used in Statistical Packages

$$\text{Wilk's lambda } \Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|}$$

$$\text{Lawley-Hotelling trace} = tr\left[\mathbf{B}\mathbf{W}^{-1}\right]$$

$$\text{Pillai trace} = tr\left[\mathbf{B}(\mathbf{B}+\mathbf{W})^{-1}\right]$$

Roy's largest root =

maximum eigenvalue of $\mathbf{W}(\mathbf{B}+\mathbf{W})^{-1}$

102

## Example 6.9

$$\left(\begin{matrix} \begin{bmatrix}9\\3\end{bmatrix} & \begin{bmatrix}6\\2\end{bmatrix} & \begin{bmatrix}9\\7\end{bmatrix} \\ \begin{bmatrix}0\\4\end{bmatrix} & \begin{bmatrix}2\\0\end{bmatrix} & \\ \begin{bmatrix}3\\8\end{bmatrix} & \begin{bmatrix}1\\9\end{bmatrix} & \begin{bmatrix}2\\7\end{bmatrix} \end{matrix}\right) \bar{\mathbf{x}}_1 = \begin{bmatrix}8\\4\end{bmatrix}, \bar{\mathbf{x}}_2 = \begin{bmatrix}1\\2\end{bmatrix}, \bar{\mathbf{x}}_3 = \begin{bmatrix}4\\5\end{bmatrix}, \bar{\mathbf{x}} = \begin{bmatrix}4\\5\end{bmatrix}$$

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix}, SS_{obs} = SS_{mean} + SS_{tr} + SS_{res} = 128 + 78 + 10 = 216$$

$$\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix}, SS_{obs} = SS_{mean} + SS_{tr} + SS_{res} = 200 + 48 + 24 = 272$$

103

## Example 6.8

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & \\ 5 & 5 & 5 \end{pmatrix} + \begin{pmatrix} -1 & -1 & -1 \\ -3 & -3 & \\ 3 & 3 & 3 \end{pmatrix} + \begin{pmatrix} -1 & -2 & 3 \\ 2 & -2 & \\ 0 & 1 & -1 \end{pmatrix}$$

Cross products

Mean: $8 \times 4 \times 5 = 160$

Treatment: $3 \times 4 \times (-1) + 2 \times (-3) \times (-3) + 3 \times (-2) \times 3 = -12$

Residual: $1 \times (-1) + (-2) \times (-2) + 1 \times 3 + \cdots + 0 \times (-1) = 1$

Total: $9 \times 3 + 6 \times 2 + 9 \times 7 + \cdots + 2 \times 7 = 149$

104

## Example 6.9

| Source of variation | Matrix of sum of squares and cross products | Degrees of freedom |
|---|---|---|
| Treatment | $\begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$ | $3 - 1 = 2$ |
| Residual | $\begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$ | $3 + 2 + 3 - 3 = 5$ |
| Total (corrected) | $\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$ | 7 |

105

## Example 6.9

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} = \frac{\begin{vmatrix} 10 & 1 \\ 1 & 24 \end{vmatrix}}{\begin{vmatrix} 88 & -11 \\ -11 & 72 \end{vmatrix}} = 0.0385$$

$$\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)\frac{\sum n_\ell - g - 1}{g-1} = \left(\frac{1-\sqrt{0.0385}}{\sqrt{0.0385}}\right)\frac{8-3-1}{3-1}$$

$$= 8.19 > F_{2(g-1), 2\sum n_\ell - g - 1}(0.01) = F_{4,8}(0.01) = 7.01$$

Reject $H_0$

106

## Example 6.10: Nursing Home Data

- Nursing homes can be classified by the owners: private (271), non-profit (138), government (107)
- Costs: nursing labor, dietary labor, plant operation and maintenance labor, housekeeping and laundry labor
- To investigate the effects of ownership on costs

107

## Example 6.10

| Group | Number of observations | Sample mean vectors |
|---|---|---|
| $\ell = 1$ (private) | $n_1 = 271$ | |
| $\ell = 2$ (nonprofit) | $n_2 = 138$ | $\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix}; \bar{\mathbf{x}}_2 = \begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix}; \bar{\mathbf{x}}_3 = \begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix}$ |
| $\ell = 3$ (government) | $n_3 = 107$ | |
| | $\sum_{\ell=1}^{3} n_\ell = 516$ | |

Sample covariance matrices

$$\mathbf{S}_1 = \begin{bmatrix} .291 & & & \\ -.001 & .011 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{bmatrix}; \quad \mathbf{S}_2 = \begin{bmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{bmatrix};$$

$$\mathbf{S}_3 = \begin{bmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{bmatrix}$$

Source: Data courtesy of State of Wisconsin Department of Health and Social Services.

108

# Example 6.10

$$\mathbf{W} = (n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + (n_3-1)\mathbf{S}_3$$
$$= \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & 0.633 & 1.484 & \\ 9.581 & 2.428 & 0.394 & 6.538 \end{bmatrix}$$
$$\bar{\mathbf{x}} = \frac{n_1\bar{\mathbf{x}}_1 + n_2\bar{\mathbf{x}}_2 + n_3\bar{\mathbf{x}}_3}{n_1+n_2+n_3} = [2.136 \quad 0.519 \quad 0.102 \quad 0.380]'$$
$$\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})' = \begin{bmatrix} 3.475 & & & \\ 1.111 & 1.225 & & \\ 0.821 & 0.453 & 0.235 & \\ 0.584 & 0.610 & 0.230 & 0.304 \end{bmatrix}$$

109

# Example 6.10

$$\Lambda^* = \frac{|W|}{|B+W|} = 0.7714$$
$$\left(\frac{\sum n_\ell - p - 2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = 17.67$$
$$F_{2\times 4, 2\times 510}(0.01) \approx \chi_8^2(0.01)/8 = 2.51$$
or, approximate analysis
$$-(n-1-(p+g)/2)\ln\left(\frac{|W|}{|B+W|}\right) = 132.76$$
$$\chi^2_{p(g-1)}(0.01) = \chi_8^2(0.01) = 20.09$$
Reject $H_0$ by both analyses

110

# Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

111

# Questions

- What are the Bonferroni Intervals for Treatment Effects?

112

# Bonferroni Intervals for Treatment Effects

$$\hat{\tau}_{ki} = \bar{x}_{ki} - \bar{x}_i, \quad \hat{\tau}_{ki} - \hat{\tau}_{\ell i} = \bar{x}_{ki} - \bar{x}_{\ell i}$$
$$\mathrm{Var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) = \mathrm{Var}(\bar{x}_{ki} - \bar{x}_{\ell i}) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\sigma_{ii}$$
$$\mathbf{W} = (n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + \cdots + (n_g-1)\mathbf{S}_g$$
$$= (n-g)\mathbf{S}_{pooled} \approx (n-g)\mathbf{\Sigma}$$
$$\mathrm{Var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) \approx \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\frac{w_{ii}}{(n-g)}$$
$$m = pg(g-1)/2$$

113

# Result 6.5: Bonferroni Intervals for Treatment Effects

With confidence at least $(1-\alpha)$

$\tau_{ki} - \tau_{\ell i}$ belongs to

$$\bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)}$$

114

## Example 6.11: Example 6.10 Data

$\hat{\boldsymbol{\tau}}_1 = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}} = [-0.070 \quad -0.039 \quad -0.020 \quad -0.020]'$

$\hat{\boldsymbol{\tau}}_3 = \bar{\mathbf{x}}_3 - \bar{\mathbf{x}} = [0.137 \quad 0.002 \quad 0.023 \quad 0.003]'$

$\hat{\tau}_{13} - \hat{\tau}_{33} = -0.20 - 0.023 = -0.043,\ n = 516$

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_3}\right)\frac{w_{33}}{n-g}} = \sqrt{\left(\frac{1}{271} + \frac{1}{107}\right)\frac{1.484}{516-3}} = 0.00614$$

$t_{513}(0.05/4\times 3\times 2) = 2.87$

95% simultaneous confidence interval for $\tau_{13} - \tau_{33}$

$-0.043 \pm 2.87 \times 0.00614$ or $(-0.061, -0.025)$

95% simultaneous confidence intervals for

$\tau_{13} - \tau_{23}$ and $\tau_{23} - \tau_{33}$ : $(-0.058, -0.026)$, $(-0.021, 0.019)$

115

## Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

116

## Questions

- How to test if the covariance matrices of several populations are equal? (Box's *M*-Test)

117

## Test for Equality of Covariance Matrices

- With $g$ populations, null hypothesis

  $H_0$: $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_g = \Sigma$
- Assume multivariate normal populations
- Likelihood ratio statistic for testing $H_0$

$$\Lambda = \prod_\ell \left(\frac{|\mathbf{S}_\ell|}{|\mathbf{S}_{pooled}|}\right)^{(n_\ell - 1)/2}$$

$$\mathbf{S}_{pooled} = \frac{1}{\sum_\ell (n_\ell - 1)} \{(n_1 - 1)\mathbf{S}_1 + \cdots + (n_g - 1)\mathbf{S}_g\}$$

118

## Box's *M*-Test

$$M = -2\ln\Lambda = \left[\sum_\ell (n_\ell - 1)\right]\ln|\mathbf{S}_{pooled}| - \sum_\ell [(n_\ell - 1)\ln|\mathbf{S}_\ell|]$$

$$u = \left[\sum_\ell \frac{1}{(n_\ell - 1)} - \frac{1}{\sum_\ell (n_\ell - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right]$$

$C = (1-u)M$ : approximate $\chi^2_\nu$

$\nu = \frac{1}{2}p(p+1)(g-1)$

Reject $H_0$ if $C > \chi^2_\nu(\alpha)$

119

## Example 6.12

- Example 6.10 - nursing home data

$g = 3,\quad p = 4,\quad n_1 = 271,\quad n_2 = 138,\quad n_3 = 107$

$\ln|\mathbf{S}_1| = -17.397,\quad \ln|\mathbf{S}_2| = -13.926$

$\ln|\mathbf{S}_3| = -15.741,\quad \ln|\mathbf{S}_{pooled}| = -15.564$

$$u = \left[\frac{1}{270} + \frac{1}{137} + \frac{1}{106} - \frac{1}{270+137+106}\right]\left[\frac{2(4)^2 + 3(4) - 1}{6(4+1)(3-1)}\right]$$

$= 0.0133$

$M = [270+137+106](-15.564) -$

$[270(-17.397) + 137(-13.926) + 106(-15.741)] = 289.3$

$C = (1 - 0.0133)\times 289.3 = 285.5$

$\nu = 4(4+1)(3-1)/2 = 20$

$H_0$ is rejected at any reasonable lelel of significance from $\chi^2_\nu$ table for comparison with $C$

120

## Example 6.13: Plastic Film Data

$x_1$ = tear resistance, $x_2$ = gloss, and $x_3$ = opacity

| | | Factor 2: Amount of additive | | | | | |
|---|---|---|---|---|---|---|---|
| | | Low (1.0%) | | | High (1.5%) | | |
| | | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| Factor 1: Change in rate of extrusion | Low (−10)% | [6.5 | 9.5 | 4.4] | [6.9 | 9.1 | 5.7] |
| | | [6.2 | 9.9 | 6.4] | [7.2 | 10.0 | 2.0] |
| | | [5.8 | 9.6 | 3.0] | [6.9 | 9.9 | 3.9] |
| | | [6.5 | 9.6 | 4.1] | [6.1 | 9.5 | 1.9] |
| | | [6.5 | 9.2 | 0.8] | [6.3 | 9.4 | 5.7] |
| | | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| | High (10%) | [6.7 | 9.1 | 2.8] | [7.1 | 9.2 | 8.4] |
| | | [6.6 | 9.3 | 4.1] | [7.0 | 8.8 | 5.2] |
| | | [7.2 | 8.3 | 3.8] | [7.2 | 9.7 | 6.9] |
| | | [7.1 | 8.4 | 1.6] | [7.5 | 10.1 | 2.7] |
| | | [6.8 | 8.5 | 3.4] | [7.6 | 9.2 | 1.9] |

121

## Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

122

## Questions

- How to determine if a factor and its interaction with the other factor is significant if two factors are involved in an experiment?
- What are the four types of interactions of two factors?
- What is the two-way ANOVA table?

123

## Scenarios

- To observe if effects of factors in the following scenarios are significant
  - Ratings of music compressed by MP3 compressors: brands vs. ages of the subjects
  - Performance of Teaching: methods (Lectures, group studying, and computer assisted instruction) vs. genders of undergraduate students

124

## Teaching Methods vs. Gender: Knowing only Overall Mean

| Gender | CAI | Lecture | Group Studying | Mean |
|---|---|---|---|---|
| Boys | 50 | 50 | 50 | 50 |
| Girls | 50 | 50 | 50 | 50 |
| Mean | 50 | 50 | 50 | 50 |

125

## Teaching Methods vs. Gender:
Knowing Overall Mean and Row Effects

| Gender | CAI | Lecture | Group Studying | Mean |
|---|---|---|---|---|
| Boys | 40 | 40 | 40 | 40 |
| Girls | 60 | 60 | 60 | 60 |
| Mean | 50 | 50 | 50 | 50 |

126

## Teaching Methods vs. Gender:
Knowing Overall Mean, Row Effects, and Column Effects

| Gender | CAI | Lecture | Group Studying | Mean |
|---|---|---|---|---|
| Boys | 50 | 40 | 30 | 40 |
| Girls | 70 | 60 | 50 | 60 |
| Mean | 60 | 50 | 40 | 50 |

127

## Teaching Methods vs. Gender: Including Interaction Terms

| Gender | CAI | Lecture | Group Studying | Mean |
|---|---|---|---|---|
| Boys | 65 | 40 | 15 | 40 |
| Girls | 55 | 60 | 65 | 60 |
| Mean | 60 | 50 | 40 | 50 |

128

## Comparing Four MP3 Compressors

- Test four brands, *A*, *B*, *C*, *D*
- 10 subjects, 5 young and 5 senior, each brand (40 in total) to provide a satisfaction rating on a 10-point scale
- Assume that the rating to each brand is a normal distribution, but all four distributions are with the same variance

129

## Sample Data

| | | A | B | C | D | Mean |
|---|---|---|---|---|---|---|
| Young Subjects | 1~4 | 4 | 5 | 7 | 2 | 5.05 |
| | 5~8 | 4 | 5 | 8 | 1 | |
| | 9~12 | 5 | 6 | 7 | 2 | |
| | 13~16 | 5 | 6 | 9 | 3 | |
| | 17~20 | 6 | 7 | 6 | 3 | |
| | Mean | 4.8 | 5.8 | 7.4 | 2.2 | |

*Adapted from: G. R. Norman and D. L. Streiner, *Biostatistics*, 3rd ed.

130

## Sample Data

| | | A | B | C | D | Mean |
|---|---|---|---|---|---|---|
| Senior Subjects | 21~24 | 3 | 6 | 3 | 4 | 3.70 |
| | 25~28 | 4 | 4 | 2 | 5 | |
| | 29~32 | 4 | 5 | 2 | 4 | |
| | 33~36 | 3 | 6 | 2 | 4 | |
| | 37~40 | 4 | 3 | 3 | 3 | |
| | Mean | 3.6 | 4.8 | 2.4 | 4.0 | |

| | A | B | C | D | Mean |
|---|---|---|---|---|---|
| Brand Mean | 4.2 | 5.3 | 4.9 | 3.1 | 4.375 |

*Adapted from: G. R. Norman and D. L. Streiner, *Biostatistics*, 3rd ed.

131

## Sum of Squares (Young/Senior)

$$SS(young\,/\,senior) = bn\sum_{\ell=1}^{g}(\bar{x}_{\ell\bullet} - \bar{x})^2$$

$$SS(young\,/\,senior) = 20[(5.05-4.375)^2 + (3.70-4.375)^2] = 18.225$$

132

## Sum of Squares (Brands)

$$SS(brands) = gn\sum_{k=1}^{b}(\bar{x}_{\bullet k} - \bar{x})^2$$

$$SS(brands) = 10[(4.2-4.375)^2 + (5.3-4.375)^2 + (4.9-4.375)^2 + (3.1-4.375)^2] = 27.875$$

133

## Sum of Squares (Within)

$$SS(within) = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{\gamma=1}^{n}(x_{\ell k\gamma} - \bar{x}_{\ell k})$$

$$\begin{aligned} SS(within) = {} & (4-4.8)^2 + (4-4.8)^2 + \cdots + (6-4.8)^2 + \\ & (5-5.8)^2 + (5-5.8)^2 + \cdots + (7-5.9)^2 + \\ & \cdots + \\ & (4-4.0)^2 + (5-4.0)^2 + \cdots + (3-4.0)^2 \\ & [40 \text{ terms}] \\ = {} & 24.80 \end{aligned}$$

134

## Sum of Squares (Total)

$$x_{\ell kr} = \bar{x} + (\bar{x}_{\ell\bullet} - \bar{x}) + (\bar{x}_{\bullet k} - \bar{x}) + (\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x}) + (x_{\ell kr} - \bar{x}_{\ell k})$$

$$\sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell kr} - \bar{x})^2 = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell\bullet} - \bar{x})^2 + \sum_{k=1}^{b} gn(\bar{x}_{\bullet k} - \bar{x})^2 + \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x})^2 + \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell kr} - \bar{x}_{\ell k})^2$$

$$\mathrm{SS}(total) = \mathrm{SS}(young/senior) + \mathrm{SS}(brand) + \mathrm{SS}(interactions) + \mathrm{SS}(within)$$

135

## Sum of Squares (Interactions)

$$SS(interactions) = n\sum_{\ell=1}^{g}\sum_{k=1}^{b}(\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x})^2$$

$$\begin{aligned} SS(interactions) = {} & 5[(4.8-4.875)^2 + (3.6-3.525)^2 + \cdots + (4.0-2.425)^2] \\ & [8 \text{ terms}] \\ = {} & 58.475 \end{aligned}$$

136

## Sum of Squares (Total)

$$\begin{aligned} SS(total) = {} & (4-4.375)^2 + (4-4.375)^2 + \cdots + (4-4.375)^2 + \\ & (5-4.375)^2 + (5-4.375)^2 + \cdots + (3-4.375)^2 + \\ & (7-4.375)^2 + (8-4.375)^2 + \cdots + (3-4.375)^2 + \\ & (2-4.375)^2 + (1-4.375)^2 + \ldots + (3-4.375)^2 \\ & [40 \text{ terms}] \\ = {} & 129.375 = 18.225 + 58.475 + 24.80 + 27.875 \end{aligned}$$

137

## Expected Values of Sum of Squares

$E[SS(brand)/df(brand)]$ contains $\sigma^2_{brand}, \sigma^2_{interactions}, \sigma^2_{err}$

$E[SS(within)/df(within)] = \sigma^2_{err}$

Thus, if brand effect is significant

$$\frac{E[SS(brand)/df(brand)]}{E[SS(within)/df(within)]} > 1$$

$F_{brand} > 1$

138

## Degrees of Freedom

$$df(young/senior) = g-1 = 2-1 = 1$$
$$df(brand) = b-1 = 4-1 = 3$$
$$df(within) = bg(n-1) = 8(5-1) = 32$$
$$df(interactions) = (b-1)(g-1) = (4-1)(2-1) = 3$$
$$\begin{aligned} df(total) &= bgn-1 = bg(n-1)+(b-1)(g-1)+b-1+g-1 \\ &= df(within)+df(interactions)+ \\ &\quad df(brand)+df(young/senior) \\ &= 40-1 = 39 = 32+3+3+1 \end{aligned}$$

139

## Two-way ANOVA Summary

| Source | Sum of Squares | *df* | Mean square | F |
|---|---|---|---|---|
| Brand | 27.875 | 3 | 9.29 | 11.99 |
| Young/ Senior | 18.225 | 1 | 18.23 | 23.52 |
| Brand X Y/S | 58.475 | 3 | 19.49 | 25.15 |
| Within | 24.80 | 32 | 0.78 | |
| Total | 129.375 | 39 | | |

140

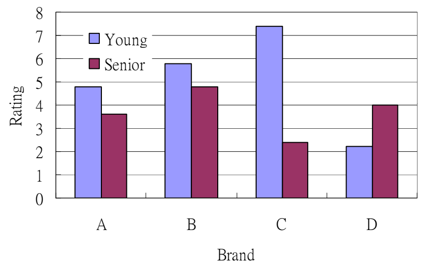
## Hypothesis Testing

$$F_{brand} = 11.99 > F_{3,32}(0.05) \approx 2.92$$
$$F_{Y/S} = 23.52 > F_{1,32}(0.05) \approx 4.17$$
$$F_{interactions} = 25.15 > F_{3,32}(0.05) \approx 2.92$$

All factors and interactions are significant at 0.05 significance level

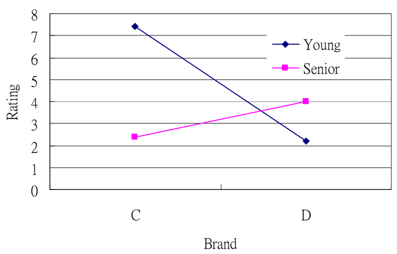
141

## Histogram of Means



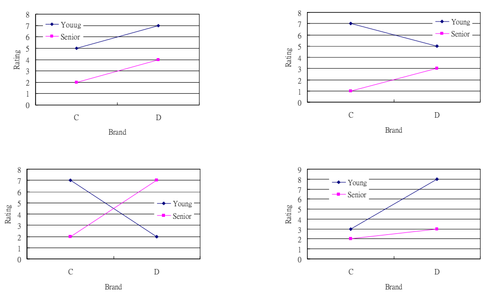

142

## Effect of Interaction




143

## Possible Types of Interactions



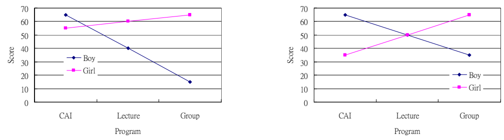
144

# de Groot's Experiment (1965)

- Observed the ability of chess masters and novices to recall piece positions
- Experts
  - Recalled about 90% of the pieces in a typical mid-game
- Novices
  - Recalled about 20%
- Many factors might have been introduced
- Randomized piece positions
  - Everybody recalled about 20%
  - No effect of expertise

145

# Interpretation by Adjusted Data




146

# Two-Way ANOVA

$$X_{\ell kr} = \mu + \tau_\ell + \beta_k + \gamma_{\ell k} + e_{\ell kr}$$

$$\ell = 1, 2, \cdots, g; \quad k = 1, 2, \cdots, b; \quad r = 1, 2, \cdots, n$$

$$\sum_{\ell=1}^{g} \tau_\ell = \sum_{k=1}^{b} \beta_k = \sum_{\ell=1}^{g} \gamma_{\ell k} = \sum_{k=1}^{b} \gamma_{\ell k} = 0, \quad e_{\ell kr} : N(0, \sigma^2)$$

$$E(X_{\ell kr}) = \mu + \tau_\ell + \beta_k + \gamma_{\ell k}$$

$$x_{\ell kr} = \bar{x} + (\bar{x}_{\ell\bullet} - \bar{x}) + (\bar{x}_{\bullet k} - \bar{x}) + (\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x}) + (x_{\ell kr} - \bar{x}_{\ell k})$$

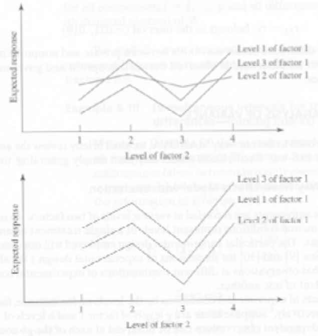
$$\sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (x_{\ell kr} - \bar{x})^2 = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell\bullet} - \bar{x})^2 + \sum_{k=1}^{b} gn(\bar{x}_{\bullet k} - \bar{x})^2$$

$$+ \sum_{\ell=1}^{g} \sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x})^2 + \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (x_{\ell kr} - \bar{x}_{\ell k})^2$$

$$SS_{cor} = SS_{fac1} + SS_{fac2} + SS_{int} + SS_{res}$$

147

# Effect of Interactions




148

# Two-Way ANOVA

ANOVA TABLE FOR COMPARING EFFECTS OF TWO FACTORS AND THEIR INTERACTION

| Source of variation | Sum of squares (SS) | Degrees of freedom (d.f.) |
|---|---|---|
| Factor 1 | $SS_{fac1} = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell\cdot} - \bar{x})^2$ | $g - 1$ |
| Factor 2 | $SS_{fac2} = \sum_{k=1}^{b} gn(\bar{x}_{\cdot k} - \bar{x})^2$ | $b - 1$ |
| Interaction | $SS_{int} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x})^2$ | $(g - 1)(b - 1)$ |
| Residual (Error) | $SS_{res} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (x_{\ell kr} - \bar{x}_{\ell k})^2$ | $gb(n - 1)$ |
| Total (corrected) | $SS_{cor} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (x_{\ell kr} - \bar{x})^2$ | $gbn - 1$ |

149

# Two-Way ANOVA

$F$ – ratio tests

$$\frac{SS_{fac1} / (g-1)}{SS_{res} / \left(gb(n-1)\right)} : \text{for effects of factor 1}$$

$$\frac{SS_{fac2} / (b-1)}{SS_{res} / \left(gb(n-1)\right)} : \text{for effects of factor 2}$$

$$\frac{SS_{int} / (g-1)(b-1)}{SS_{res} / \left(gb(n-1)\right)} : \text{for effects of}$$

factor 1 - factor 2 intercation

150

## Outline

- Comparing Several Multivariate Population Means (One-Way Manova)
- Simultaneous Confidence Intervals for Treatment Effects
- Testing for Equality of Covariance Matrices
- Two-Way ANOVA
- Two-Way Multivariate Analysis of Variance

151

## Questions

- What is the two-way MANOVA table?
- How to determine if the interaction effect exists?
- How to test the effect of each factor by the two-way MANOVA?
- How to determine the Bonferroni confidence intervals if the interaction effect is negligible?

152

## Two-Way MANOVA

$$\mathbf{X}_{\ell kr} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{\ell k} + \mathbf{e}_{\ell kr}$$

$$\ell = 1, 2, \cdots, g; \quad k = 1, 2, \cdots, b; \quad r = 1, 2, \cdots, n$$

$$\sum_{\ell=1}^{g} \boldsymbol{\tau}_\ell = \sum_{k=1}^{b} \boldsymbol{\beta}_k = \sum_{\ell=1}^{g} \boldsymbol{\gamma}_{\ell k} = \sum_{k=1}^{b} \boldsymbol{\gamma}_{\ell k} = 0, \quad \mathbf{e}_{\ell kr} : N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

$$\mathbf{x}_{\ell kr} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{\bullet k} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}}_{\bullet k} + \bar{\mathbf{x}}) + (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})$$

$$\sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})' =$$

$$\sum_{\ell=1}^{g} bn(\bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}})' + \sum_{k=1}^{b} gn(\bar{\mathbf{x}}_{\bullet k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\bullet k} - \bar{\mathbf{x}})'$$

$$+ \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}}_{\bullet k} + \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\bullet} - \bar{\mathbf{x}}_{\bullet k} + \bar{\mathbf{x}})'$$

$$+ \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})'$$

153

## Two-Way MANOVA

MANOVA TABLE FOR COMPARING FACTORS AND THEIR INTERACTION

| Source of variation | Matrix of sum of squares and cross products (SSP) | Degrees of freedom (d.f.) |
|---|---|---|
| Factor 1 | $\mathrm{SSP}_{fac1} = \sum_{\ell=1}^{g} bn(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})'$ | $g - 1$ |
| Factor 2 | $\mathrm{SSP}_{fac2} = \sum_{k=1}^{b} gn(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})'$ | $b - 1$ |
| Interaction | $\mathrm{SSP}_{int} = \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})'$ | $(g-1)(b-1)$ |
| Residual (Error) | $\mathrm{SSP}_{res} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})'$ | $gb(n-1)$ |
| Total (corrected) | $\mathrm{SSP}_{cor} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})'$ | $gbn - 1$ |

154

## Two-Way MANOVA

Test for interaction

For large samples, reject $H_0 : \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \cdots = \boldsymbol{\gamma}_{gb} = 0$ if

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2}\right]\ln \Lambda^* > \chi^2_{(g-1)(b-1)}(\alpha)$$

$$\text{Wilk's lambda } \Lambda^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{int} + \mathrm{SSP}_{res}|}$$

If interaction effects exist, the factor effects do not have a clear interpretation

155

## Two-Way MANOVA

Test for factor 1 effect

For large samples, reject $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = 0$ if

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2}\right]\ln \Lambda^* > \chi^2_{(g-1)p}(\alpha)$$

$$\text{Wilk's lambda } \Lambda^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{fac1} + \mathrm{SSP}_{res}|}$$

Test for factor 2 effect

For large samples, reject $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b = 0$ if

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2}\right]\ln \Lambda^* > \chi^2_{(b-1)p}(\alpha)$$

$$\text{Wilk's lambda } \Lambda^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{fac2} + \mathrm{SSP}_{res}|}$$

156

## Bonferroni Confidence Intervals

With negligible interactions,
the simultaneus confidence intervals are

$$(\bar{x}_{\ell \bullet i} - \bar{x}_{m \bullet i}) \pm t_p\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{bn}} \quad \text{for } \tau_{\ell i} - \tau_{mi}$$

and

$$(\bar{x}_{\bullet ki} - \bar{x}_{\bullet qi}) \pm t_p\left(\frac{\alpha}{pb(b-1)}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{gn}} \quad \text{for } \beta_{ki} - \beta_{qi}$$

$$\nu = gb(n-1), \quad \mathbf{E} = \mathrm{SSP}_{res}$$

157

## Example 6.13: MANOVA Table

| Source of variation | SSP | d.f. |
|---|---|---|
| Factor 1: change in rate of extrusion | $\begin{bmatrix} 1.7405 & -1.5045 & .8555 \\ & 1.3005 & -.7395 \\ & & .4205 \end{bmatrix}$ | 1 |
| Factor 2: amount of additive | $\begin{bmatrix} .7605 & .6825 & 1.9305 \\ & .6125 & 1.7325 \\ & & 4.9005 \end{bmatrix}$ | 1 |
| Interaction | $\begin{bmatrix} .0005 & .0165 & .0445 \\ & .5445 & 1.4685 \\ & & 3.9605 \end{bmatrix}$ | 1 |
| Residual | $\begin{bmatrix} 1.7640 & .0200 & -3.0700 \\ & 2.6280 & -.5520 \\ & & 64.9240 \end{bmatrix}$ | 16 |
| Total (corrected) | $\begin{bmatrix} 4.2655 & -.7855 & -.2395 \\ & 5.0855 & 1.9095 \\ & & 74.2055 \end{bmatrix}$ | 19 |

158

## Example 6.13: Interaction

$$\Lambda^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{int} + \mathrm{SSP}_{res}|} = 0.7771$$

$$(g-1)(b-1) = 1$$

$$F = \left(\frac{1-\Lambda^*}{\Lambda^*}\right)\frac{(gb(n-1)-p+1)/2}{(|(g-1)(b-1)-p|+1)/2} : F_{\nu_1,\nu_2}$$

$$\nu_1 = |(g-1)(b-1)-p|+1 = 3$$

$$\nu_2 = gb(n-1)-p+1 = 14$$

$$F = 1.34 < F_{3,14}(0.05) = 3.34$$

$H_0 : \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \boldsymbol{\gamma}_{21} = \boldsymbol{\gamma}_{22} = 0$ (no interaction) is not rejected

159

## Example 6.13: Effects of Factors 1 & 2

$$\Lambda_1^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{fac1} + \mathrm{SSP}_{res}|} = 0.3819$$

$$\Lambda_2^* = \frac{|\mathrm{SSP}_{res}|}{|\mathrm{SSP}_{fac2} + \mathrm{SSP}_{res}|} = 0.5230$$

$$F_1 = \left(\frac{1-\Lambda_1^*}{\Lambda_1^*}\right)\frac{\nu_2/2}{\nu_1/2} = 7.55, \quad \nu_1 = |(g-1)-p|+1 = 3$$

$$F_2 = \left(\frac{1-\Lambda_2^*}{\Lambda_2^*}\right)\frac{\nu_2/2}{\nu_1/2} = 4.26, \quad \nu_1 = |(b-1)-p|+1 = 3$$

$$\nu_2 = gb(n-1)-p+1 = 14$$

$F_1 > F_{3,14}(0.05) = 3.34$, reject $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = 0$

$F_2 > F_{3,14}(0.05) = 3.34$, reject $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 0$

160

## Outline

- Profile Analysis
- ANOVA for Repeated Measures
- Repeated Measures Designs and Growth Curves
- Perspectives and Strategy for Analyzing Multivariate Models

161

## Questions

- What is the profile analysis?
- How to carry out the profile analysis?

162

## Profile Analysis

- A battery of p treatments (tests, questions, etc.) are administered to two or more group of subjects
- The question of equality of mean vectors is divided into several specific possibilities
  - Are the profiles parallel?
  - Are the profiles coincident?
  - Are the profiles level?

163

## Example 6.14: Love and Marriage Data

Extremely negative | Very negative | Moderately negative | Slightly negative | Slightly positive | Moderately positive | Very positive | Extremely positive

1 2 3 4 5 6 7 8

1. All things considered, how would you describe *your contributions* to the marriage?
2. All things considered, how would you describe *your outcomes* from the marriage?

Subjects were also asked to respond to the following questions, using the 5-point scale shown.

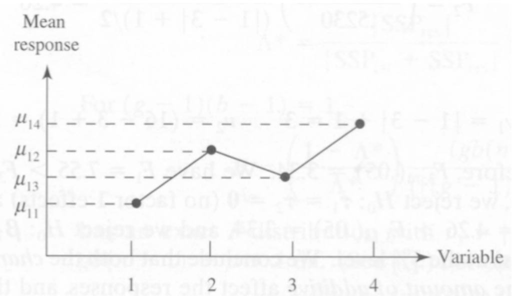
3. What is the level of *passionate* love that you feel for your partner?
4. What is the level of *companionate* love that you feel for your partner?

None at all | Very little | Some | A great deal | Tremendous amount

1 2 3 4 5

164

## Population Profile




165

## Profile Analysis

Assume two populations

Are the profiles parallel?

$$H_{01}: \mu_{1i} - \mu_{1i-1} = \mu_{2i} - \mu_{2i-1}, i = 2, 3, \cdots, p$$

Are the profiles coincident?

$$H_{02}: \mu_{1i} = \mu_{2i}, i = 1, 2, \cdots, p$$

Are the profiles level?

$$H_{03}: \mu_{11} = \mu_{12} = \cdots = \mu_{1p} = \mu_{21} = \mu_{22} = \cdots = \mu_{2p}$$

166

## Test for Parallel Profiles

$$\underset{(p-1)\times p}{\mathbf{C}} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

$$\mathbf{CX}_{1j} : N_{p-1}(\mathbf{C\mu}_1, \mathbf{C\Sigma\Sigma'}), \mathbf{CX}_{2j} : N_{p-1}(C\mathbf{\mu}_2, \mathbf{C\Sigma\Sigma'})$$

Reject $H_{01}: \mathbf{C\mu}_1 = \mathbf{C\mu}_2$ at level $\alpha$ if

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{C}'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{CS}_{pooled}\mathbf{C}'\right]^{-1}\mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > c^2$$

$$c^2 = \frac{(n_1 + n_2 - 2)(p-1)}{n_1 + n_2 - p} F_{p-1, n_1+n_2-p}(\alpha)$$

167

## Test for Coincident Profiles

Given parallel profiles

Reject $H_{02}: \mathbf{1}'\mathbf{\mu}_1 = \mathbf{1}'\mathbf{\mu}_2$ at level $\alpha$ if

$$T^2 = \mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{1}'\mathbf{S}_{pooled}\mathbf{1}\right]^{-1}\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

$$= \left(\frac{\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{1}'\mathbf{S}_{pooled}\mathbf{1}\right]}}\right)^2 > t^2_{n_1+n_2-2}\left(\frac{\alpha}{2}\right) = F_{1, n_1+n_2-2}(\alpha)$$

168

## Test for Level Profiles

Given coincident profiles

Reject $H_{03}: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ at level $\alpha$ if

$$(n_1+n_2)\bar{\mathbf{x}}'\mathbf{C}'\left[\mathbf{CSC}'\right]^{-1}\mathbf{C}\bar{\mathbf{x}} > c^2$$

$$c^2 = \frac{(n_1+n_2-1)(p-1)}{n_1+n_2-p-1} F_{p-1,n_1+n_2-p+1}(\alpha)$$

$$\bar{\mathbf{x}} = \frac{n_1}{n_1+n_2}\bar{\mathbf{x}}_1 + \frac{n_2}{n_1+n_2}\bar{\mathbf{x}}_2$$

169

## Example 6.14

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 6.833 \\ 7.033 \\ 3.967 \\ 4.700 \end{bmatrix}, \bar{\mathbf{x}}_2 = \begin{bmatrix} 6.633 \\ 7.000 \\ 4.000 \\ 4.533 \end{bmatrix}$$

$$\mathbf{S}_{pooled} = \begin{bmatrix} 0.606 & 0.262 & 0.066 & 0.161 \\ 0.262 & 0.637 & 0.173 & 0.143 \\ 0.066 & 0.173 & 0.810 & 0.029 \\ 0.161 & 0.143 & 0.029 & 0.306 \end{bmatrix}$$

170

## Example 6.14: Test for Parallel Profiles

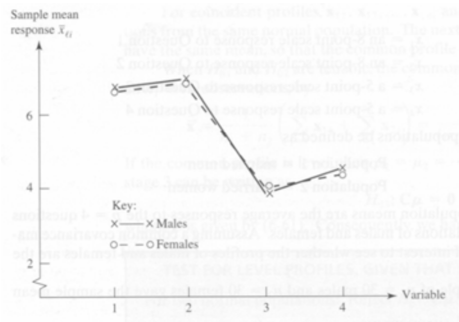
$$\mathbf{CS}_{pooled}\mathbf{C}' = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \mathbf{S}_{pooled} \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.200 \\ 0.033 \\ -0.033 \\ 0.167 \end{bmatrix} = \begin{bmatrix} -0.167 \\ -0.066 \\ 0.200 \end{bmatrix}$$

$$T^2 = 1.005 < \frac{(30+30-2)(4-1)}{30+30-4} F_{3,56}(0.05) = 8.7$$

171

## Example 6.14: Sample Profiles




172

## Example 6.14: Test for Coincident Profiles

$$\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = 0.367$$
$$\mathbf{1}'\mathbf{S}_{pooled}\mathbf{1} = 4.207$$

$$T^2 = \left( \frac{0.367}{\sqrt{\left(\frac{1}{30}+\frac{1}{30}\right)4.207}} \right)^2 = 0.501 < F_{1,58}(0.05) = 4.0$$

173

## Outline

- Profile Analysis
- ANOVA for Repeated Measures
- Repeated Measures Designs and Growth Curves
- Perspectives and Strategy for Analyzing Multivariate Models

174

## Questions

- What are repeated measures?
- How to view the data for repeated measures in a two-way ANOVA view?
- How to test the null hypothesis in repeated measures?

175

## Repeated-Measures ANOVA

- Drugs A, B, C are tested to see if they are equally effective for pain relief
- Subjects are to take all of the drugs, in turn, suitably blinded and after a suitable washout period
- Subjects rate the degree of pain belief on a 1 to 6 scale (1: no relief, 6 complete relief)

176

## Avoiding Order Effects

- Randomize the order of treatment
  - 1/3 get drug A first, 1/3 get drug B first, 1/3 get drug C first
- People in a long, natural healing course may grow tolerant of the irritant and learn to tune them out
  - The last medication may work the best
  - Order effects

177

## Sample Data

| Subject | A | B | C | Average |
|---|---|---|---|---|
| 1 | 5 | 3 | 2 | 3.33 |
| 2 | 5 | 4 | 3 | 4.00 |
| 3 | 5 | 6 | 5 | 5.33 |
| 4 | 6 | 4 | 2 | 4.00 |
| 5 | 6 | 6 | 6 | 6.00 |
| 6 | 4 | 2 | 1 | 2.33 |
| 7 | 4 | 4 | 3 | 3.67 |
| 8 | 4 | 5 | 5 | 4.67 |
| 9 | 4 | 2 | 2 | 2.67 |
| 10 | 5 | 3 | 1 | 3.00 |
| Means | 4.80 | 3.90 | 3.00 | 3.90 |

178

*Adapted from: G. R. Norman and D. L. Streiner, *Biostatistics*, 3rd ed.

## Two-Way ANOVA View

- Individual subjects as one factor
- Pain reliever as a second factor
- Cells are defined by
  - Subjects: 10 levels
  - Drug: 3 levels
- One observation per cell
- Special case of two-way ANOVA
  - $n = 1, g = 10, b = 3$

179

## Sum of Squares (Drug)

$$SS(drug) = g\sum_{k=1}^{b}(\bar{x}_{\bullet k} - \bar{x})^2$$

$$SS(drug) = 10[(4.8-3.9)^2 + (3.9-3.9)^2 + (3.0-3.9)^2]$$
$$= 16.2$$

180

## Sum of Squares (Subjects)

$$SS(subjects) = b\sum_{\ell=1}^{g}(\bar{x}_{\ell\bullet} - \bar{x})^2$$

$$SS(subjects) = 3[(3.33-3.90)^2 + (4.00-3.90)^2 + \cdots + (3.00-3.90)^2] = 36.7$$

181

## Sum of Squares (Interaction)

$$SS(interaction) = \sum_{\ell=1}^{g}\sum_{k=1}^{b}(\bar{x}_{\ell k} - \bar{x}_{\ell\bullet} - \bar{x}_{\bullet k} + \bar{x})^2$$

$$SS(interaction) = [(5-4.23)^2 + (3-3.33)^2 + \cdots + (1-2.10)^2]$$

[30 terms]

$$= 15.8$$

182

## Sum of Squares (Within)

$$SS(within) = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\left(x_{\ell k\gamma} - \bar{x}_{\ell k}\right) = 0$$

183

## Degrees of Freedom

$$\begin{aligned}
df(subject) &= g-1 = 10-1 = 9 \\
df(drug) &= b-1 = 3-1 = 2 \\
df(within) &= bg(n-1) = 0 \\
df(interaction) &= (b-1)(g-1) = (3-1)(10-1) = 18 \\
df(total) &= bg-1 = (b-1)(g-1) + b-1 + g-1 \\
&= df(interaction) + df(drug) + df(subject) \\
&= 30-1 = 29 = 18+2+9
\end{aligned}$$

184

## Signal vs. Noise

- To determine if there is any significant difference in relief from different pain relievers
  - Main effect of Drug
- $SS(within) = 0$
- Choose $SS(interaction)$ as error term
  - Reflects the extent to which different subjects respond differently to the different drug types

185

## ANOVA Table

| Source | Sum of Squares | *df* | Mean square | F |
|---|---|---|---|---|
| Drug | 16.2 | 2 | 8.100 | 9.225 |
| Subject | 36.7 | 9 | 4.078 | |
| Drug X Subject | 15.8 | 18 | 0.878 | |
| Totals | 68.7 | 29 | | |

186

## Hypothesis Testing

$$F_{Drug} = 9.225 > F_{2,18}(0.05) \approx 3.55$$

Drug effect is significant (i.e., difference exists) at 0.05 significance level



## ANOVA Table for Same Data as a One-Way ANOVA Test

| Source | Sum of Squares | *df* | Mean square | F |
|---|---|---|---|---|
| Drug | 16.2 | 2 | 8.100 | 4.107 |
| Error | 52.5 | 27 | 1.944 | |
| Totals | 68.7 | 29 | | |



## Outline

- Profile Analysis
- ANOVA for Repeated Measured
- Repeated Measures Designs and Growth Curves
- Perspectives and Strategy for Analyzing Multivariate Models



## Questions

- How to compare growth curves?



## Example 6.15: Ulna Data, Control Group

| Subject | Initial | 1 year | 2 year | 3 year |
|---|---|---|---|---|
| 1 | 87.3 | 86.9 | 86.7 | 75.5 |
| 2 | 59.0 | 60.2 | 60.0 | 53.6 |
| 3 | 76.7 | 76.5 | 75.7 | 69.5 |
| 4 | 70.6 | 76.1 | 72.1 | 65.3 |
| 5 | 54.9 | 55.1 | 57.2 | 49.0 |
| 6 | 78.2 | 75.3 | 69.1 | 67.6 |
| 7 | 73.7 | 70.8 | 71.8 | 74.6 |
| 8 | 61.8 | 68.7 | 68.2 | 57.4 |
| 9 | 85.3 | 84.4 | 79.2 | 67.0 |
| 10 | 82.3 | 86.9 | 79.4 | 77.4 |
| 11 | 68.6 | 65.4 | 72.3 | 60.8 |
| 12 | 67.8 | 69.2 | 66.3 | 57.9 |
| 13 | 66.2 | 67.0 | 67.0 | 56.2 |
| 14 | 81.0 | 82.3 | 86.8 | 73.9 |
| 15 | 72.3 | 74.6 | 75.3 | 66.1 |
| Mean | 72.38 | 73.29 | 72.47 | 64.79 |

Source: Data courtesy of Everett Smith.



## Example 6.15: Ulna Data, Treatment Group

| Subject | Initial | 1 year | 2 year | 3 year |
|---|---|---|---|---|
| 1 | 83.8 | 85.5 | 86.2 | 81.2 |
| 2 | 65.3 | 66.9 | 67.0 | 60.6 |
| 3 | 81.2 | 79.5 | 84.5 | 75.2 |
| 4 | 75.4 | 76.7 | 74.3 | 66.7 |
| 5 | 55.3 | 58.3 | 59.1 | 54.2 |
| 6 | 70.3 | 72.3 | 70.6 | 68.6 |
| 7 | 76.5 | 79.9 | 80.4 | 71.6 |
| 8 | 66.0 | 70.9 | 70.3 | 64.1 |
| 9 | 76.7 | 79.0 | 76.9 | 70.3 |
| 10 | 77.2 | 74.0 | 77.8 | 67.9 |
| 11 | 67.3 | 70.7 | 68.9 | 65.9 |
| 12 | 50.3 | 51.4 | 53.6 | 48.0 |
| 13 | 57.7 | 57.0 | 57.5 | 51.5 |
| 14 | 74.3 | 77.7 | 72.6 | 68.0 |
| 15 | 74.0 | 74.7 | 74.5 | 65.7 |
| 16 | 57.3 | 56.0 | 64.7 | 53.0 |
| Mean | 69.29 | 70.66 | 71.18 | 64.53 |

Source: Data courtesy of Everett Smith.

## Comparison of Growth Curves

$\mathbf{X}_{\ell j}$ : vector of $p$ measurements on subject $j$ in group $\ell$

$j = 1, 2, \cdots, n_\ell; \quad \ell = 1, 2, \cdots, g$

$\mathbf{X}_{\ell j}$ : Multivariate normal with covariance $\mathbf{\Sigma}$

Putthoff - Roy model

$$E(\mathbf{X}_{\ell j}) = \begin{bmatrix} \beta_{\ell 0} + \beta_{\ell 1} t_1 + \cdots + \beta_{\ell q} t_1^q \\ \beta_{\ell 0} + \beta_{\ell 1} t_2 + \cdots + \beta_{\ell q} t_2^q \\ \vdots \\ \beta_{\ell 0} + \beta_{\ell 1} t_p + \cdots + \beta_{\ell q} t_p^q \end{bmatrix} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^q \\ 1 & t_2 & \cdots & t_2^q \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & \cdots & t_p^q \end{bmatrix} \begin{bmatrix} \beta_{\ell 0} \\ \beta_{\ell 1} \\ \vdots \\ \beta_{\ell q} \end{bmatrix}$$

$$= \mathbf{B}\boldsymbol{\beta}_\ell$$

193

## Comparison of Growth Curves

Maximum likelihood estimators of $\boldsymbol{\beta}_\ell$ :

$$\hat{\boldsymbol{\beta}}_\ell = \left(\mathbf{B}'\mathbf{S}_{pooled}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}'\mathbf{S}_{pooled}^{-1}\overline{\mathbf{X}}_\ell$$

$$\mathbf{S}_{pooled} = \frac{1}{N-g}\left((n_1-1)\mathbf{S}_1 + \cdots + (n_g-1)\mathbf{S}_g\right) = \frac{\mathbf{W}}{N-g}$$

$$N = \sum_{\ell=1}^{g} n_\ell, \quad \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}_\ell) = \frac{k}{n_\ell}\left(\mathbf{B}'\mathbf{S}_{pooled}^{-1}\mathbf{B}\right)^{-1}$$

$$k = (N-g)(N-g-1)/(N-g-p+q)(N-g-p+q+1)$$

$$\mathbf{W}_q = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}\left(\mathbf{X}_{\ell j} - \mathbf{B}\hat{\boldsymbol{\beta}}_\ell\right)\left(\mathbf{X}_{\ell j} - \mathbf{B}\hat{\boldsymbol{\beta}}_\ell\right)', \quad \Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W}_q|}$$

Reject the null hypothesis that the polynomial is adequate if

$$-\left(N - (p-q+g)/2\right)\ln \Lambda^* > \chi^2_{(p-q-1)g}(\alpha)$$

194

## Example 6.15

Use quadratic growth model

$$\begin{bmatrix}\hat{\beta}_1 & \hat{\beta}_2\end{bmatrix} = \begin{bmatrix} 73.0701\ (2.58) & 70.1387\ (2.50) \\ 3.6444\ (0.83) & 4.0900\ (0.80) \\ -2.0274\ (0.28) & -1.8534\ (0.27) \end{bmatrix}$$

Control Group : $73.07 + 3.64t - 2.03t^2$

Treatment Group : $70.14 + 4.09t - 1.85t^2$

$\Lambda^* = 0.7627$

$-\left(N - (p-q+g)/2\right)\ln \Lambda^* = 7.86 < \chi^2_{(4-2-1)2}(0.01) = 9.21$

195

## Outline

- Profile Analysis
- ANOVA for Repeated Measured
- Repeated Measures Designs and Growth Curves
- Perspectives and Strategy for Analyzing Multivariate Models

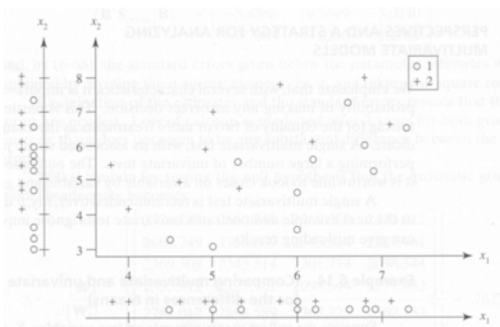
196

## Questions

- What are the strategies in multivariate analysis?
- Why is the experimental design important?

197

## Example 6.16: Comparing Multivariate and Univariate Tests




198

## Example 6.16: Comparing Multivariate and Univariate Tests

Univariate test on $x_1: F = 2.46 < F_{1,18}(0.10) = 3.01$

Univariate test on $x_2: F = 2.68 < F_{1,18}(0.10) = 3.01$

Accept $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$

Hotelling's test :

$$T^2 = 17.29 > c^2 = \frac{18 \times 2}{17} F_{2,17}(0.01) = 12.94$$

Reject $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$

199

## Strategy for Multivariate Comparison of Treatments

- Try to identify outliers
  - Perform calculations with and without the outliers
- Perform a multivariate test of hypothesis
- Calculate the Bonferroni simultaneous confidence intervals
  - For all pairs of groups or treatments, and all characteristics

200

## Importance of Experimental Design

- Differences could appear in only one of the many characteristics or a few treatment combinations
- Differences may become lost among all the inactive ones
- Best preventative is a good experimental design
  - Do not include too many other variables that are not expected to show differences

201