MANOVA: Part 2
Profile Analysis and 2-way MANOVA

Edps/Soc 584, Psych 594

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Spring 2017

Outline

Extensions of 1–Way MANOVA:

- Profile Analysis
  - Two groups
  - Two or more groups
- 2–way MANOVA
  - Review 2-way ANOVA
  - 2-way MANOVA
  - Example
- Validity of Assumptions
- Unbalanced Designs

Reading: Johnson & Wichern pages 312-328.

Profile Analysis

Profile analysis is an extension of 1–way MANOVA involving p response variables administered to g groups of individuals or cases.

e.g., A battery of personality tests (sub-scales) where the p test scores are measured in the same units (or at least similar or commensurate units).

Profile Analysis gives you more specific hypotheses to test than the standard

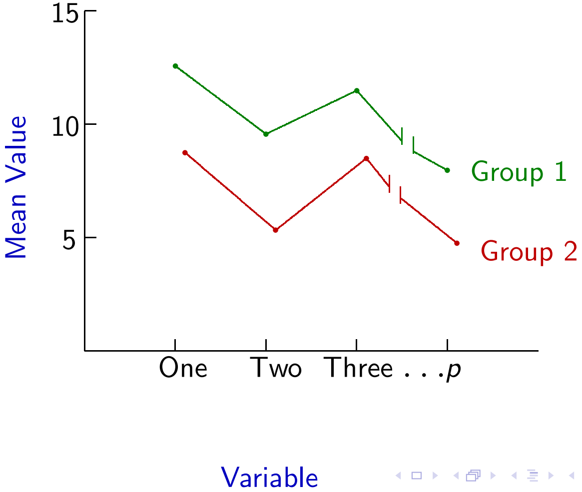
$$H_o : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g.$$

Consider $p \times 1$ mean vector for each group

$$\boldsymbol{\mu}_l' = (\mu_{l1}, \mu_{l2}, \ldots, \mu_{lp})$$

and plot the means as "profiles" for each group.

Example of Profiles for Two Groups



Profile Analysis: 2 Groups

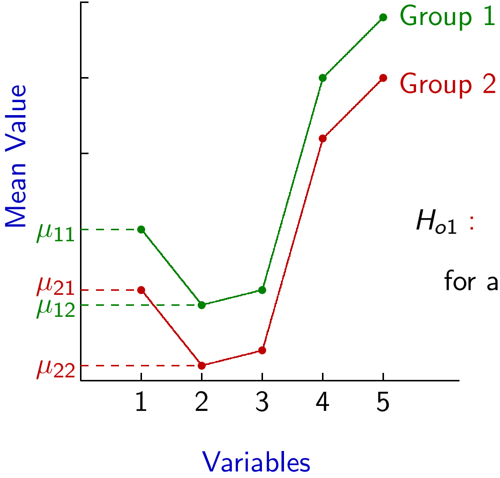
The usual null hypothesis in a 1-way MANOVA is

$$H_o : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

In profile analysis, we break this down into three sequential tests.

- Question 1: Are the profiles parallel?
  Are successive (adjacent) differences between means equal?
  If the answer is "no", then stop.
  If the answer is "yes", then go onto the next question/step.
- Question 2: Are the profiles the same?
  Are the population means equal for the two groups?
  If the answer is "no", then stop.
  If the answer is "yes", then go onto the next question/step.
- Question 3: Are the profiles flat?
  Are the population means equal over the variables?

Question 1: Parallel Profiles

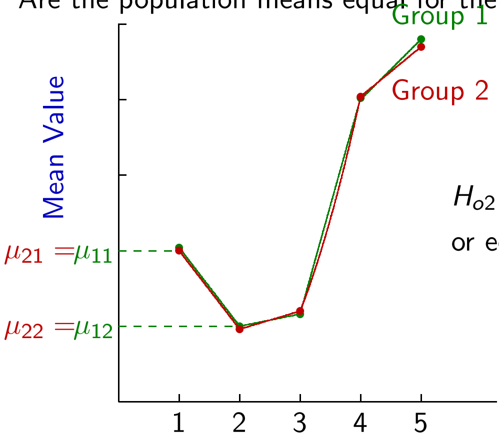


$$H_{o1}: \quad \mu_{1i} - \mu_{1,i-1} = \mu_{2i} - \mu_{2,i-1}$$

for all $i = 2, \ldots, p$

Question 2: Are Profiles Coincident?

Assuming that profiles are parallel, are they coincident?
Are the population means equal for the two groups?

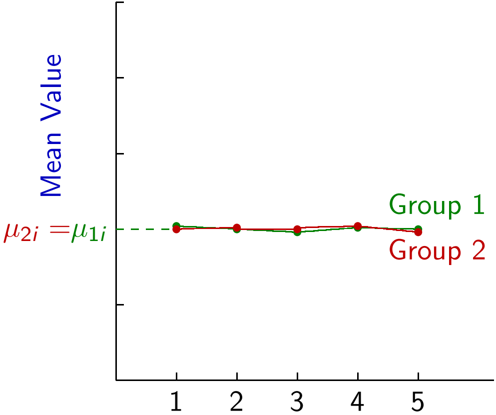


$H_{o2} : \mu_{1i} = \mu_{2i}$ for all $i = 1, \ldots p$

or equivalently $H_{o2} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$

Question 3

Assuming that the profiles are coincident, are they Level (flat)? Are all the means equal? (over groups and variables).



$$H_{o3}:\ \mu_{11} = \mu_{12} = \cdots = \mu_{1p}$$
$$= \mu_{21} = \mu_{22} = \cdots = \mu_{2p}$$

Testing Question 1: Parallel

Assuming $\mathbf{X}_{1j} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{X}_{2j} \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ for $j = 1, \ldots, n_l$, and independent.

Question 1: Are the profiles parallel?

We can write the hypothesis are $H_{o1} : \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$ where $\mathbf{C}_{(p-1)\times p}$ contrast matrix. e.g.,

$$\mathbf{C}_{(p-1)\times p} = \left(\begin{array}{rrrrcrr} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{array}\right)$$

What we're doing is linearly transforming our original p variables into $(p-1)$ new variables.

Question 1 continued

When we take

$$\mathbf{CX} = \begin{pmatrix} -X_1 + X_2 \\ -X_2 + X_3 \\ \vdots \\ -X_{p-1} + X_p \end{pmatrix},$$

which are measured on both groups (populations, etc).
Random variable $\mathbf{CX}_{1j} \sim \mathcal{N}_{p-1}(\mathbf{C}\boldsymbol{\mu}_1, \mathbf{C\Sigma C}')$,
& Random variable $\mathbf{CX}_{2j} \sim \mathcal{N}_{p-1}(\mathbf{C}\boldsymbol{\mu}_2, \mathbf{C\Sigma C}')$.
To estimate the covariance matrix $\mathbf{\Sigma}$, use $\mathbf{S}_{pool}$,

$$\mathbf{S}_{pool} \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}$$

and the estimate of $\mathbf{C\Sigma C}'$ equals $\mathbf{CS}_{pool}\mathbf{C}'$.

Finishing Question 1

How to test $H_{o1}: \mathbf{C}\boldsymbol{\mu}_1 - \mathbf{C}\boldsymbol{\mu}_2 = \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$?
Hotelling's T^2 for Two independent samples.
Reject H_{o1} if

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

where

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

- If you reject $H_{o1} \longrightarrow$ STOP.
  ... You can do any follow-up to examine differences.
- If you retain $H_{o2} \longrightarrow$ Conclude profiles are parallel and PROCEED to next question.

Testing Question 2: Coincident profiles

Assuming that the the profiles are parallel, are the profiles coincident?

If profiles are parallel, then one will be "above" the other for all $i = 1, \ldots, p$; that is,

$$\mu_{ij} > \mu_{ij} \qquad \text{for all} \quad i = 1, \ldots, p$$

or

$$\mu_{ij} < \mu_{ij} \qquad \text{for all} \quad i = 1, \ldots, p$$

So, profiles will be coincident only if the total "heights" are equal

$$\begin{aligned} (\mu_{11} + \mu_{12} + \cdots + \mu_{1p}) &= (\mu_{21} + \mu_{22} + \cdots + \mu_{2p}) \\ \mathbf{1}_p' \boldsymbol{\mu}_1 &= \mathbf{1}_p' \boldsymbol{\mu}_2 \end{aligned}$$

where $\mathbf{1}_p$ is a $(p \times 1)$ vector of ones.

Question 2 continued

The null hypothesis for question 2 is

$$H_{o2} : \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$$

We are forming a new variable, $\mathbf{1}'\mathbf{X}$, which is a simple sum and test whether this variable is equal for the two groups.

$$\mathbf{1}'\mathbf{X}_{1j} \sim \mathcal{N}_1(\mathbf{1}'\boldsymbol{\mu}_1, \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}) \quad j = 1, \ldots, n_1$$

$$\mathbf{1}'\mathbf{X}_{2j} \sim \mathcal{N}_1(\mathbf{1}'\boldsymbol{\mu}_2, \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}) \quad j = 1, \ldots, n_2$$

We estimate

$$\begin{array}{ccc} \mathbf{1}'\boldsymbol{\mu}_1 & \text{by} & \mathbf{1}'\bar{\mathbf{x}}_1 \\ \mathbf{1}'\boldsymbol{\mu}_2 & \text{by} & \mathbf{1}'\bar{\mathbf{x}}_2 \\ & \text{and} & \\ \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} & \text{by} & \mathbf{1}'\mathbf{S}_{pool}\mathbf{1} \end{array}$$

Finishing Up Question 2

To test

$$H_{o2:} \quad \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2,$$

we can do a simple univariate 2 independent sample t-test.
We will reject H_{o2} at the α-level if

$$t = \frac{\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \mathbf{1}'\mathbf{S}_{pool}\mathbf{1}}} > t_{n_1+n_2-2}(\alpha)$$

- If Reject $H_{o2} \longrightarrow$ STOP and conclude that the profiles are parallel but not coincident.
- If Retain $H_{o2} \longrightarrow$ PROCEED and test whether profiles are flat.

Question 3: Flat

If the profiles are coincident, do they all have the same mean? i.e.,

$$H_{o3} : \mu_{11} = \mu_{12} = \cdots = \mu_{1p} = \mu_{21} = \mu_{22} = \cdots = \mu_{2p}$$

We can test this using a contrast matrix **C**, such as the one we used for Question 1 (testing parallel profiles)

$$H_{o3} : \mathbf{C}(\mu_1 + \mu_2) = \mathbf{0}$$

Note that we are adding rather than subtracting as we did in H_{o1}. This tests

$$\mathbf{C}\boldsymbol{\mu}_1 + \mathbf{C}\boldsymbol{\mu}_2 = \mathbf{0}$$

$$\begin{pmatrix} (\mu_{12} - \mu_{11}) + (\mu_{22} - \mu_{21}) \\ (\mu_{13} - \mu_{12}) + (\mu_{23} - \mu_{22}) \\ \vdots \\ (\mu_{1p} - \mu_{1,p-1}) + (\mu_{2p} - \mu_{2,p-1}) \end{pmatrix} = \mathbf{0}$$

Question 3 continued

Testing

$$\begin{aligned} \mathbf{C}\boldsymbol{\mu}_1 + \mathbf{C}\boldsymbol{\mu}_2 &= \mathbf{0} \\ \begin{pmatrix} (\mu_{12}-\mu_{11}) + (\mu_{22}-\mu_{21}) \\ (\mu_{13}-\mu_{12}) + (\mu_{23}-\mu_{22}) \\ \vdots \\ (\mu_{1p}-\mu_{1,p-1}) + (\mu_{2p}-\mu_{2,p-1}) \end{pmatrix} &= \mathbf{0} \end{aligned}$$

Is the same as this, which is what we want to test,

$$\begin{aligned} (\mu_{12}+\mu_{22}) &= (\mu_{11}+\mu_{21}) \\ (\mu_{13}+\mu_{23}) &= (\mu_{12}+\mu_{22}) \\ \vdots \quad & \quad \vdots \\ (\mu_{1p}+\mu_{2p}) &= (\mu_{1,p-1}+\mu_{2,p-1}) \end{aligned}$$

Because If the profiles are coincident, then $\mu_{1i} = \mu_{2i}$ for all $i = 1, \ldots, p$. So if H_{o3} is true, then $\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$.

Question 3 continued

For $H_{o3}: \mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$, flat profiles, we are taking linear combinations in two ways simultaneously

- Over variables via the use of $\mathbf{C}$.
- Over groups via an additive function.

The mean $\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$ is estimated by taking the grand mean

$$\bar{\mathbf{X}} = \frac{\sum_{j=1}^{n_1} \mathbf{X}_{1j} + \sum_{j=1}^{n_2} \mathbf{X}_{2j}}{n_1 + n_2} = \frac{n_1 \bar{\mathbf{X}}_1 + n_2 \bar{\mathbf{X}}_2}{n_1 + n_2}$$

and the distribution is of $\mathbf{C}\bar{\mathbf{X}}$ is

$$\mathbf{C}\bar{\mathbf{X}} \sim \mathcal{N}_{p-1}\left(\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}'\left(\frac{1}{n_1 + n_2}\right)\right)$$

How do we get this?

Finishing up Question 3

Reject H_{o3} : $\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$ (profiles are flat) at α-level if

$$(n_1+n_2)\bar{\mathbf{X}}'\mathbf{C}'(\mathbf{CSC})^{-1}\mathbf{C}\bar{\mathbf{X}} > \frac{(n_1+n_2-1)(p-1)}{n_1+n_2-p+1}\mathcal{F}_{(p-1),(n_1+n_2-p+1}(\alpha)$$

where $\mathbf{S}$ is the total sample covariance matrix

$$\mathbf{S} = \frac{1}{n_1+n_2-1}\left[\sum_{j=1}^{n_1}(\mathbf{X}_{1j}-\bar{\mathbf{X}})(\mathbf{X}_{1j}-\bar{\mathbf{X}})' + \sum_{j=1}^{n_2}(\mathbf{X}_{2j}-\bar{\mathbf{X}})(\mathbf{X}_{2j}-\bar{\mathbf{X}})'\right]$$

Note that we use the total sample mean in computing $\mathbf{S}$.

Example: WAIS data

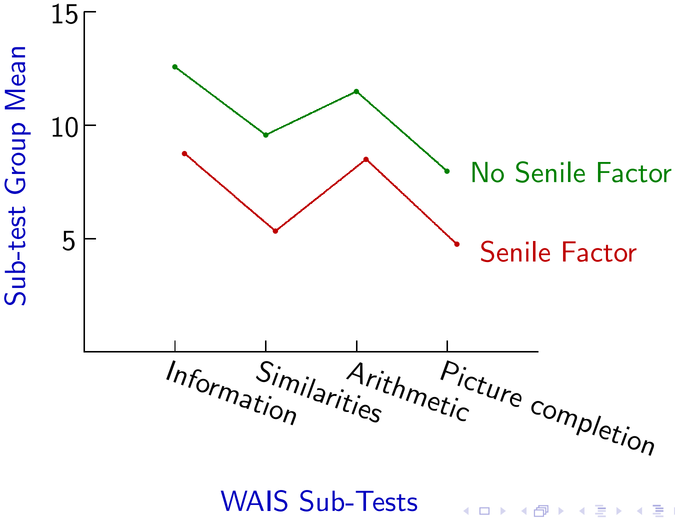
This example is from Morrison (2005): 49 elderly men in a study of human aging were classified into the diagnostic categories "senile factor present" and "no senile factor" on the basis of an intensive psychiatric examination. The Wechsler Adult Intelligence Scale (WAIS) was administered to all subjects by an independent investigator.

Mean scores by group on some of the WAIS subtests.

	Not Senile ($n = 37$)		Senile ($n = 12$)	
Sub-Test	$\bar{x}$	std dev	$\bar{x}$	std dev
Information	12.566	3.387	8.750	3.251
Similarities	9.486	3.380	5.333	4.271
Arithmetic	11.514	3.363	8.500	3.631
Picture	7.973	1.922	4.750	3.571

Note: My results differ slightly from Morrison's. There is either a typo in the text or in the data set. (no way to find out which).

WAIS Profiles



WAIS: Are Profiles Parallel?

The within group (pooled) sample covariance matrix:

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

$$= \begin{pmatrix} 11.262 & 8.995 & 7.164 & 3.379 \\ 8.995 & 13.019 & 7.037 & 2.308 \\ 7.164 & 7.037 & 11.750 & 2.639 \\ 3.379 & 2.308 & 2.639 & 5.813 \end{pmatrix}$$

Test for parallel profiles: The hypothesis is $H_o : \mu_{1i}-\mu_{1,i-1}=\mu_{2,i}-\mu_{2,i-1}$ for $i=2,3,4$; that is, the following three qualities simultaneously hold in the population,

$$\mu_{12}-\mu_{11} = \mu_{22}-\mu_{21}$$
$$\mu_{13}-\mu_{12} = \mu_{23}-\mu_{22}$$
$$\mu_{14}-\mu_{13} = \mu_{24}-\mu_{23}$$

WAIS: Are Profiles Parallel?

To test this hypothesis, we need a (3×4) contrast matrix $\mathbf{C}$

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Our hypothesis can now be,expressed as $H_o : \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$, or $H_o : \mathbf{C}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{0}$. This hypothesis is tested by Hotelling's T^2, and we need

$$\mathbf{C}\mathbf{S}_{pool}\mathbf{C}' = \begin{pmatrix} 6.291 & -4.150 & -0.944 \\ -4.151 & 10.694 & -4.382 \\ -0.944 & -4.382 & 12.286 \end{pmatrix},$$

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' = (3.818, 4.153, 3.014, 3.223),$$

and

$$(\mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2))' = (0.336, -1.140, 0.209)$$

WAIS: Are Profiles Parallel?

Putting all of this into our equation for T^2 gives us

$$\begin{aligned} T^2 &= \frac{n_1 n_2}{n_1+n_2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{C}'(\mathbf{C}\mathbf{S}_{pool}\mathbf{C}')^{-1}\mathbf{C}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2) \\ &= 1.224 \end{aligned}$$

and compare the following statistic to the $\mathcal{F}$ distribution,

$$F = \frac{n_1+n_2-p}{(n_1+n_2-2)(p-1)}T^2 = \frac{45}{(47)(3)}1.224 = 0.391$$

Since $\mathcal{F}_{3,45}(\alpha = .05) = 2.8115$ is greater than our observed statistic (or the p-value of F equals .76), we do not reject the hypothesis that the profiles are parallel.

Since we Retained the null, we will PROCEED to test whether the profiles are coincident.

WAIS: Are Profiles Coincident?

Test for Equal (coincident) Profiles.

We now test $H_o : \mu_{1i} = \mu_{2i}$ for $i = 1, 2, 3, 4$ variables. Since we concluded that the profiles are parallel, then we can test this hypothesis by testing whether

$$(\mu_{11} + \mu_{12} + \mu_{13} + \mu_{14}) = (\mu_{21} + \mu_{22} + \mu_{23} + \mu_{24})$$
$$\mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$$

where $\mathbf{1}$ is a (4×1) vector of ones. This is just a univariate, 2 independent sample t.

To test $H_o : \mathbf{1}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 0$, we need for each group the sums of the means over all variables, which are

$$\mathbf{1}'\bar{\mathbf{x}}_1 = 41.541 \qquad \text{and} \qquad \mathbf{1}'\bar{\mathbf{x}}_2 = 27.333$$

Estimate of the variance of the difference $= \mathbf{1}'\mathbf{S}_{pool}\mathbf{1} = 104.891$.

WAIS: Are Profiles Coincident?

Putting all of these statistics together for our test statistics gives us

$$t = \frac{41.541 - 27.333}{\sqrt{104.890\left(\frac{1}{12} + \frac{1}{37}\right)}} = 4.176$$

Since $t_{47}(\alpha = .025) = 2.012$ (or the p-value of $t_{47} = 4.15$ is $< .001$), we reject the null hypothesis. The profiles are not coincident.

At this point we STOP and should not go on to test whether profiles are level (flat).

Profiles Analysis for $g \geq 2$ Groups

To deal with two or more groups, we follow the same logic:

1. Using assumptions (including results from earlier tests), determine what the hypothesis implies for population parameters. This has implications for data.
2. Compute a statistic that reflects the implications for data.
3. Find a transformation of the statistic with a known sampling distribution.

It is easier to do profile analysis using the GLM framework. We will use

$$\mathbf{X} = \mathbf{A}\mathcal{B} + \epsilon,$$

where

- $\mathbf{X}$ is $(n \times p)$ data matrix.
- $\mathbf{A}$ is the $(n \times g)$ design matrix where g = number groups.
- $\mathcal{B}$ is the $(g \times p)$ parameter matrix.
- ϵ is the $(n \times p)$ residual matrix.

This is a multivariate regression model.

Two or More Groups

Our multivariate regression model is

$$\mathbf{X} = \mathbf{A}\mathcal{B} + \epsilon,$$

Test hypotheses of the form

$$H_o : \mathbf{L}\mathcal{B}\mathbf{M} = \mathbf{0}$$

where

- **L** is ($g\times$ number of tests between groups).
- **M** is ($p\times$ number of tests over variables).

Appropriate definitions of **L** and **M** will lead to tests for parallelism, coincidence, and level means.

Two or More: Parallel

Parallel profiles imply

$$H_{o1} : (\mu_{1i} - \mu_{1,i-1}) = (\mu_{2i} - \mu_{2,i-1}) = \cdots = (\mu_{gi} - \mu_{g,i-1})$$

for all $i = 2, \ldots, p$. This will be a MANOVA on differences.
This hypothesis takes linear combinations of the means within groups and then compare the groups. For the linear transformations of the regression parameters (i.e., the τ or μ's), define

$$\mathbf{M}_{p\times(p-1)} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

For the comparisons of groups on these "new" variables, define
$\mathbf{L} = \mathbf{C}_{(g-1)\times g}$ contrast matrix.

GLM: Parallel and WAIS

For the WAIS example, H_{o1} is

$$\begin{aligned}
\mathbf{L}\mathcal{B}\mathbf{M} &= (0,1,-1)\begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} & \beta_{04} \\ \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{0} \\
&= (0,1,-1)\begin{pmatrix} (\beta_{02}-\beta_{01}) & (\beta_{03}-\beta_{02}) & (\beta_{04}-\beta_{03}) \\ (\beta_{12}-\beta_{11}) & (\beta_{13}-\beta_{12}) & (\beta_{14}-\beta_{13}) \\ (\beta_{22}-\beta_{21}) & (\beta_{23}-\beta_{22}) & (\beta_{24}-\beta_{23}) \end{pmatrix} = \mathbf{0} \\
&= [(\beta_{12}-\beta_{11})-(\beta_{22}-\beta_{21}), (\beta_{13}-\beta_{12})-(\beta_{23}-\beta_{22}), \\
&\qquad (\beta_{14}-\beta_{13})-(\beta_{24}-\beta_{23})] = (0,0,0)
\end{aligned}$$

So we have

$$\begin{pmatrix} (\beta_{12}-\beta_{11})-(\beta_{22}-\beta_{21}) \\ (\beta_{13}-\beta_{12})-(\beta_{23}-\beta_{22}) \\ (\beta_{14}-\beta_{13})-(\beta_{24}-\beta_{23}) \end{pmatrix} \Longrightarrow \begin{pmatrix} (\mu_{12}-\mu_{11})-(\mu_{22}-\mu_{21}) \\ (\mu_{13}-\mu_{12})-(\mu_{23}-\mu_{22}) \\ (\mu_{14}-\mu_{13})-(\mu_{24}-\mu_{23}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Two or More: Coincident?

Assuming parallel, then coincidence implies

$$H_{o2} : \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2 = \cdots = \mathbf{1}'\boldsymbol{\mu}_g$$

So we're testing equivalence of a single variable over g groups: 1-way ANOVA.

The linear transformation of the variables is given by

$$\mathbf{M}_{p\times 1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

For the groups,

$$\mathbf{L} = \mathbf{I}$$

See SAS for WAIS example

Two or More: Level Profiles

If profiles are coincidence, then flat or level profiles means

$$H_{o3} : \mu_{11} = \mu_{12} = \cdots = \mu_{1p} = \mu_{21} = \cdots = \mu_{gp}$$

That is, sums of two variables are the same for pairs of groups. The required linear transformation,

$$\mathbf{M}_{(p\times(p-1)} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

The required contrasts of the groups,

$$\mathbf{L} = \mathbf{C}_{(g-1)\times g}$$

which is the same one used for testing parallel profiles.

2-way MANOVA

Mini Outline:

1. Review 2–way ANOVA
2. 2–way MANOVA for balanced data
3. Example: Distributed versus Massed Practice/instruction
4. Unbalanced designs.
5. Multivariate GLM and further extensions (MANCOVA, longitudinal)

Two-Way ANOVA

"2–way" ANOVA $\longrightarrow$ 2 Factors (qualitative variables). Notation:

Factor B

Factor A		1	$\cdots$	k	$\cdots$	b
	1	X_{11r}	$\cdots$	X_{1kr}	$\cdots$	X_{1br}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Factor A	l	X_{l1r}	$\cdots$	X_{lkr}	$\cdots$	X_{lbr}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	g	X_{gkr}	$\cdots$	X_{gkr}	$\cdots$	X_{gbr}

$\longleftarrow$ n observations in each of the gb combinations of levels of the factors.

- Observation: X_{lkr} = the r^{th} observation at level l of Factor A and level k of Factor B.
- Factor A: $l = 1, \ldots, g$
- Factor B: $k = 1, \ldots, b$
- Replications: $r = 1, \ldots, n$ For now, a balanced design.

ANOVA Model for Two Factors

$$X_{lkr} = \underbrace{\mu}_{\text{overall level}} + \underbrace{\tau_l}_{\text{fixed effect factor A at level } l} + \underbrace{\beta_k}_{\text{fixed effect factor B at level } k} + \underbrace{\gamma_{lk}}_{\text{interaction between factors A \&B at levels } l \text{ \& , } k} + \underbrace{\epsilon_{lkr}}_{\text{residual}}$$

where $$\sum_{l=1}^{g}\tau_g = \sum_{k=1}^{b}\beta_k + \sum_{l=1}^{g}\gamma_{lk} = \sum_{k=1}^{b}\gamma_{lk} = 0$$

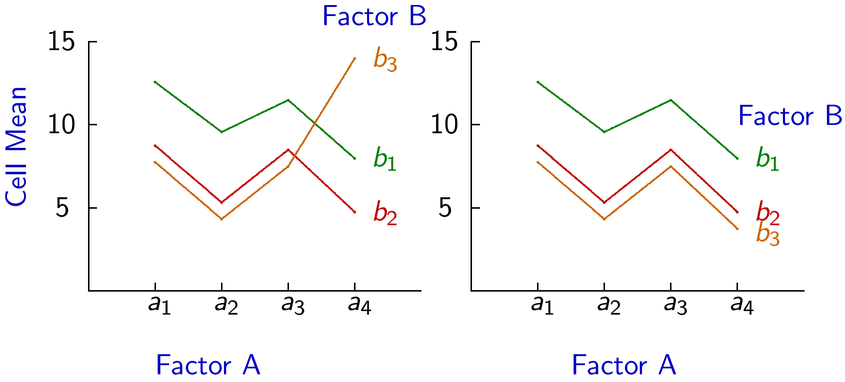
and $\epsilon_{lkr} \sim \mathcal{N}(0, \sigma^2)$ and all independent.

Model for the mean:

$$E(X_{lkr}) = \mu_{lk} = \mu + \tau_l + \beta_k + \gamma_{lk}$$

TThe effects are not additive; there is an interaction.

Examples: Interaction & No Interaction



Observation as Sum of Means

$$x_{lkr} = \underbrace{\bar{x}}_{\hat{\mu}} + \underbrace{(\bar{x}_{l\cdot\cdot} - \bar{x})}_{\hat{\tau}_l} + \underbrace{(\bar{x}_{\cdot k\cdot} - \bar{x})}_{\hat{\beta}_k} + \underbrace{(\bar{x}_{lk\cdot} - \bar{x}_{l\cdot\cdot} - \bar{x}_{\cdot k\cdot} + \bar{x})}_{\hat{\gamma}_{lk}} + \underbrace{(x_{lkr} - \bar{x}_{lk\cdot})}_{\hat{\epsilon}_{lkr}}$$

where

- $\bar{x}$ = overall sample mean.
- $\bar{x}_{l\cdot\cdot}$ = mean for l^{th} level of Factor A (or "row" mean).
- $\bar{x}_{\cdot k\cdot}$ = mean for k^{th} level of Factor B (or "column" mean).
- $\bar{x}_{lk\cdot}$ = mean for the l^{th} level of Factor A and k^{th} level of Factor B (or the "cell" mean).
- The decomposition of the observations above shows the estimates of the population effects.

Decomposition of Sums of Squares

$$x_{lkr} = \underbrace{\bar{x}}_{\hat{\mu}} + \underbrace{(\bar{x}_{l..} - \bar{x})}_{\hat{\tau}_l} + \underbrace{(\bar{x}_{.k.} - \bar{x})}_{\hat{\beta}_k} + \underbrace{(\bar{x}_{lk.} - \bar{x}_{l..} - \bar{x}_{.k.} + \bar{x})}_{\hat{\gamma}_{lk}} + \underbrace{(x_{lkr} - \bar{x}_{lk.})}_{\hat{\epsilon}_{lkr}}$$

To get the sums of squares:

1. Subtract $\bar{x}$ from both sides.
2. Square both sides.
3. Sum over observations within cells, levels of Factor A, and levels of Factor B.

   After a little algebra,

Decomposition of Sums of Squares

The sum of squares decomposition is

$$\sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{lkr}-\bar{x})^2 = \sum_{l=1}^{g} bn(\bar{x}_{l..}-\bar{x})^2 + \sum_{k=1}^{b} gn(\bar{x}_{.k.}-\bar{x})^2$$
$$+\sum_{l=1}^{g}\sum_{k=1}^{b} n(\bar{x}_{lk.}-\bar{x}_{l..}-\bar{x}_{.k.}+\bar{x})^2$$
$$+\sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{lkr}-\bar{x}_{lk.})^2$$

Sums of Squares Decomposition

We get an orthogonal decomposition of the sums of squares:

$$SS_{total} = SS_A + SS_B + SS_{AB} + SS_{residual}$$

Note: SS_{total} is corrected for the grand mean.

Also, we get decomposition of degrees of freedom,

$$\underbrace{(gbn-1)}_{\text{total}} = \underbrace{(g-1)}_{\text{Factor A}} + \underbrace{(b-1)}_{\text{Factor B}} + \underbrace{(g-1)(b-1)}_{\text{Interaction}} + \underbrace{gb(n-1)}_{\text{Residual}}$$

ANOVA Summary Table

Source of variation	Degrees of freedom	SS Sums of Squared
Factor A	$(g-1)$	$SS_A = \sum_{l=1}^{g} n(\bar{x}_{l\cdot\cdot} - \bar{x})^2$
Factor B	$(b-1)$	$SS_B = \sum_{k=1}^{b} gn(\bar{x}_{\cdot k\cdot} - \bar{x})^2$
Interaction	$(g-1)(b-1)$	$SS_{AB} = \sum_{l=1}^{g}\sum_{k=1}^{b} n(\bar{x}_{lk\cdot} - \bar{x}_{l\cdot\cdot} - \bar{x}_{\cdot k\cdot} + \bar{x})^2$
Residual (error)	$gb(n-1)$	$SS_{res} = \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{lkr} - \bar{x}_{lk\cdot})^2$
Total	$gbn-1$	$SS_{total} = \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{lkr} - \bar{x})^2$

Hypothesis Tests

- Interaction: $H_o : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{gb} = 0$. If the null is true and all assumptions valid,

$$F = \frac{SS_{AB}/((g-1)(b-1))}{SS_{res}/(gb(n-1))} \sim \mathcal{F}_{(g-1)(b-1),gb(n-1)}$$

- Factor A: $H_o : \tau_1 = \tau_2 = \ldots = \tau_g$. If the null is true and all assumptions valid,

$$F = \frac{SS_A/(g-1)}{SS_{res}/(gb(n-1))} \sim \mathcal{F}_{(g-1),gb(n-1)}$$

- Factor B: $H_o : \beta_1 = \beta_2 = \cdots = \beta_b$. If the null is true and all assumptions valid,

$$= \frac{SS_B/(b-1)}{SS_{res}/(gb(n-1))} \sim \mathcal{F}_{(b-1),gb(n-1)}$$

Start with the Interaction before the main effects.

Two-Way MANOVA

For p variables, $\mathbf{X}_{lkr}$ is a $p \times 1$ vector of measurements on p variables.

The model

$$\mathbf{X}_{lkr} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{lk} + \boldsymbol{\epsilon}_{lkr}$$

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

- $l = 1, \ldots, g$
- $k = 1, \ldots, b$
- $r = 1, \ldots, n_{lk}$
- And for identification

$$\sum_{l=1}^{g} \boldsymbol{\tau}_l = \sum_{k=1}^{b} \boldsymbol{\beta}_k = \sum_{l=1}^{g} \boldsymbol{\gamma}_{lk} = \sum_{k=1}^{b} \boldsymbol{\gamma}_{lk} = \mathbf{0}$$

Decomposition of Observation

We decompose the observation vector $(\mathbf{x}_{lkr} - \bar{\mathbf{x}})$ into sums of various vectors,

$$(\mathbf{x}_{lkr} - \bar{\mathbf{x}}) = \underbrace{(\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})}_{\hat{\boldsymbol{\tau}}_l} + \underbrace{(\bar{\mathbf{x}}_{.k.} - \bar{\mathbf{x}})}_{\hat{\boldsymbol{\beta}}_k} + \underbrace{(\bar{\mathbf{x}}_{lk.} - \bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}}_{.k.} + \bar{\mathbf{x}})}_{\hat{\boldsymbol{\gamma}}_{lk}}$$
$$+ \underbrace{(\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk.})}_{\hat{\boldsymbol{\epsilon}}_{lkr}}$$

If we take the sums of squares and cross-products of this, we obtain the SSCP decomposition,

$$\underbrace{\sum_l \sum_k \sum_r (\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})'}_{SSCP_{total}} = \ldots$$

which will take more space than is left on this slide.

Decomposition of SSCP

$$\underbrace{\sum_l \sum_k \sum_r (\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})'}_{SSCP_{total}}$$

$$= \underbrace{\sum_l bn(\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}})'}_{SSCP_{FactorA}}$$

$$+ \underbrace{\sum_k gn(\bar{\mathbf{x}}_{.k.} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{.k.} - \bar{\mathbf{x}})'}_{SSCP_{FactorB}}$$

$$+ \underbrace{\sum_l \sum_k n(\bar{\mathbf{x}}_{lk.} - \bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}}_{.k.} + \bar{\mathbf{x}})(\bar{\mathbf{x}}_{lk.} - \bar{\mathbf{x}}_{l..} - \bar{\mathbf{x}}_{.k.} + \bar{\mathbf{x}})'}_{SSCP_{interaction}}$$

$$+ \underbrace{\sum_l \sum_k \sum_r (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk.})(\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk.})'}_{SSCP_{residual}}$$

Decomposition of SSCP

So,

$$SSCP_{total} = SSCP_{FactorA} + SSCP_{FactorB} + SSCP_{Interaction} + SSCP_{residual}$$

which is orthogonal so long as the design is balanced (i.e., $n_{lk} = n$) or proportional. . . what to do with unbalanced is discussed a bit later.

And

$$\begin{aligned} df_{total} &= df_{FactorA} + df_{FactorB} + df_{Interaction} + df_{residual} \\ (gbn-1) &= (g-1) + (b-1) + (g-1)(b-1) + gb(n-1) \end{aligned}$$

Hypothesis Testing: Interaction First

$$H_o : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{gp}$$

- Test Statistic is Wilk's Lambda

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

- $\nu_h = \nu_{interaction} = (g-1)(n-1)$
- $\nu_e = \nu_{residual} + gb(n-1)$
- Distribution of Λ^* are given on the following slide for various cases.
- If you don't have $p \leq 2$ or $\nu_h \leq 2$, but you have large n then the approximate sampling distribution of

$$-\left(\nu_e - \frac{p+1-\nu_h}{2}\right)\ln(\Lambda^*) \qquad \text{is} \qquad \chi^2_{p\nu_h}$$

Distribution of Wilk's Lambda Λ^*

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

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 ν_h = degrees of freedom for hypothesis, and
ν_e = degrees of freedom for error (residual).

Hypothesis Testing: Interaction First

You should test the interaction first.

- If H_o for the interaction is rejected, then you don't so tests for main effects because the Factor main effects do not have a clear interpretation.
- J&W recommend doing p-univariate 2-way ANOVAs to find out "where" the interactions exist. (i.e., are there interactions for just some of the p variables and not others?).
- For variables with no interaction, you can interpret main effects.
- Note: In some cases you can go on to test main effects even though the interaction is significant (we'll do this in our example).
- In general, if you retain the null hypothesis for interaction, we continue and test main effects. . .

Hypothesis Testing: Main Effects

Factor A: $H_o : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ versus $H_a :$ at least one $\boldsymbol{\tau}_l \neq \mathbf{0}$

$$\Lambda^*_A = \frac{\det(SSCP_{residual})}{\det(SSCP_{residual} + SSCP_{FactorA})} = \frac{|E|}{|E + H_A|}$$

and $\nu_h = g - 1$. For $p \leq 2$ or $\nu_A \leq 2$ (i.e., $g \leq 3$), use the exact distribution given in the table.

Factor B: $H_o : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b = \mathbf{0}$ versus $H_a :$ at least one $\boldsymbol{\beta}_k \neq \mathbf{0}$

$$\Lambda^*_B = \frac{\det(SSCP_{residual})}{\det(SSCP_{residual} + SSCP_{FactorB})} = \frac{|E|}{|E + H_B|}$$

and $\nu_h = b - 1$. For $p \leq 2$ or $\nu_B \leq 2$ (i.e., $b \leq 3$), use the exact distribution given in the table.

Large n: $$-\left(\nu_e - \frac{p + 1 - \nu_h}{2}\right)\ln(\Lambda^*_h) \approx \chi^2_{p\nu_h}$$

Following Rejection

If a null hypothesis is rejected, you should perform additional analyses to figure out where the effects are. These could include

- Multivariate T^2; that is, generalized squared distances

$$D^2(\mathbf{X})_{lk|l'k'} = (\bar{\mathbf{X}}_{lk} - \bar{\mathbf{X}}_{l'k'})'\mathbf{S}_{pool}^{-1}(\bar{\mathbf{X}}_{lk} - \bar{\mathbf{X}}_{l'k'})$$

- Simultaneous confidence statements/intervals.
- When interactions are not significant, concentrate on contrasts of levels of Factor A and of Factor B (if they're significant). These are the same as those described in 1-way MANOVA notes.

  For example,

$$(\tau_{li} - \tau_{mi}) \quad \text{or} \quad (\mu_{li} - \mu_{mi})$$

  Use Bonferroni if planned.
- 2-way ANOVAs with multiple comparisons.
- Discriminant Analysis.

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 word processing to 60 female seniors in high school. 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. Note: $n_l = 10$ per cell of the design

Two variables (dependent measures):

X_1 = speed, and X_2 = accuracy

The Design and Data

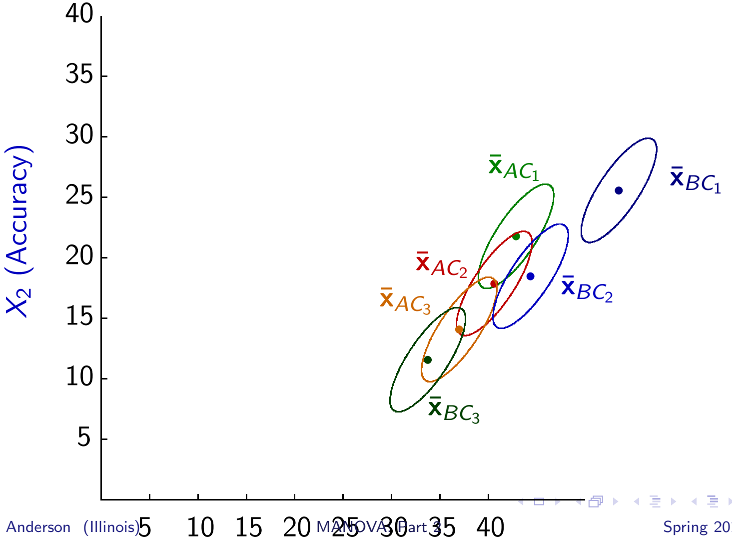
Various Means:

	C_1		C_2		C_3			
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
A	34.30	21.80	32.50	17.90	29.60	14.10	32.13	17.93
B	42.80	25.60	35.50	18.59	27.00	11.60	35.10	18.57
	38.55	23.79	34.00	18.20	28.30	12.85	33.62	18.25

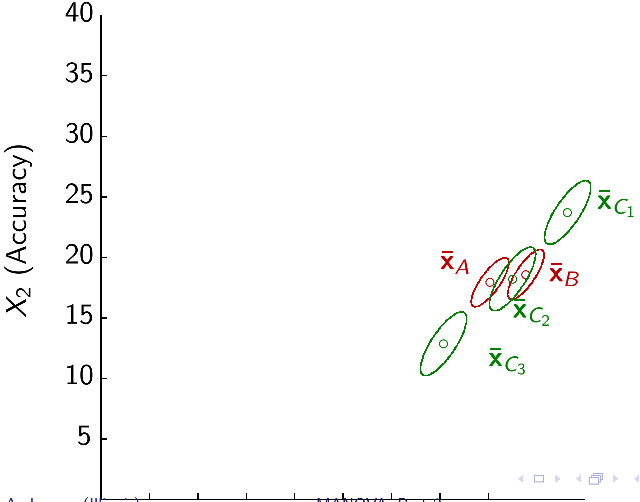
Notes:

- Cell means (black) based on $n_{lk} = 10$.
- Means for method (green) are based on $n_{l\cdot} = 30$
- Means for condition (blue) are based on $n_{\cdot k} = 20$.
- Total sample means (red) are based on $n_{\cdot\cdot} = 60$.

Cell Means & 95% Confidence Ellipses



Factor Means & 95% Confidence Ellipses



Test for Interaction

$$H_o : \gamma_{AC_1} = \gamma_{AC_2} = \gamma_{AC_3} = \gamma_{BC_1} = \gamma_{BC_2} = \gamma_{BC_3} = \mathbf{0}$$

$$H_a : \text{ not all } \gamma_{lk} = 0$$

$$\mathbf{W} = \mathbf{E} = \begin{pmatrix} 1333.10 & 211.20 \\ 211.20 & 184.70 \end{pmatrix} \qquad \mathbf{H}_{AB} = \begin{pmatrix} 308.04 & 174.82 \\ 174.82 & 99.23 \end{pmatrix}$$

$$\Lambda^* = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H}_{AB})} = \frac{201.618}{316.957} = .6361$$

and

$$F = \left(\frac{53}{2}\right)\left(\frac{1 - \sqrt{.6361}}{\sqrt{.6361}}\right) = 6.73$$

Comparing $F = 6.73$ to the $\mathcal{F}_{4,106}$ distribution, F has $p < .001$.

Test for Method Main Effect

. . . For purposes of illustration. . .

$$H_o : \boldsymbol{\tau}_A = \boldsymbol{\tau}_B = \mathbf{0} \qquad \text{versus} \qquad H_a : \text{ not all } \gamma_l = 0$$

$$\mathbf{W} = \mathbf{E} = \begin{pmatrix} 1333.10 & 211.20 \\ 211.20 & 184.70 \end{pmatrix} \qquad \mathbf{H}_{Method} = \begin{pmatrix} 132.02 & 28.18 \\ 28.18 & 6.02 \end{pmatrix}$$

$$\Lambda^* = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H}_{Method})} = \frac{201.618}{222.13} = .9077$$

and

$$F = \left(\frac{53}{1}\right)\left(\frac{1 - \sqrt{.9077}}{\sqrt{.9077}}\right) = 2.63$$

Comparing $F = 2.63$ to the $\mathcal{F}_{2,106}$ distribution, F has $p = .08$.

Test for Condition Main Effect

$$H_o : \boldsymbol{\beta}_{C_1} = \boldsymbol{\beta}_{C_2} = \boldsymbol{\beta}_{C_3} = \mathbf{0} \qquad \text{versus} \qquad H_a : \text{ not all } \boldsymbol{\beta}_k = 0$$

$$\mathbf{W} = \mathbf{E} = \begin{pmatrix} 1333.10 & 211.20 \\ 211.20 & 184.70 \end{pmatrix} \qquad \mathbf{H}_{Condition} = \begin{pmatrix} 1055.03 & 1111.55 \\ 1111.55 & 1177.30 \end{pmatrix}$$

$$\Lambda^* = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H}_{Condition})} = \frac{201.618}{1502.966} = .1341$$

and

$$F = \left(\frac{53}{2}\right)\left(\frac{1-\sqrt{.1341}}{\sqrt{.1341}}\right) = 45.85$$

Comparing $F = 45.85$ to the $\mathcal{F}_{4,106}$ distribution, F has $p < .001$.

Interpretation & Conclusion

- The effectiveness of the teaching methods (A and B) depends on which of the three conditions of distributed practice that a student used.
  (i.e., There is a significant interaction).
- The method by condition cell means indicate that method B is more effective in terms of both speed and accuracy under conditions C_1 and C_2, but A is more effective in terms of speed and accuracy under condition C_3.
- Statistical tests of these conclusions should be performed in follow up analyses. Potentially useful supplement analyses include ANOVAs for each of the dependent variables, simultaneous confidence intervals for various treatment effects, and discriminant analysis.

Validity of Assumptions

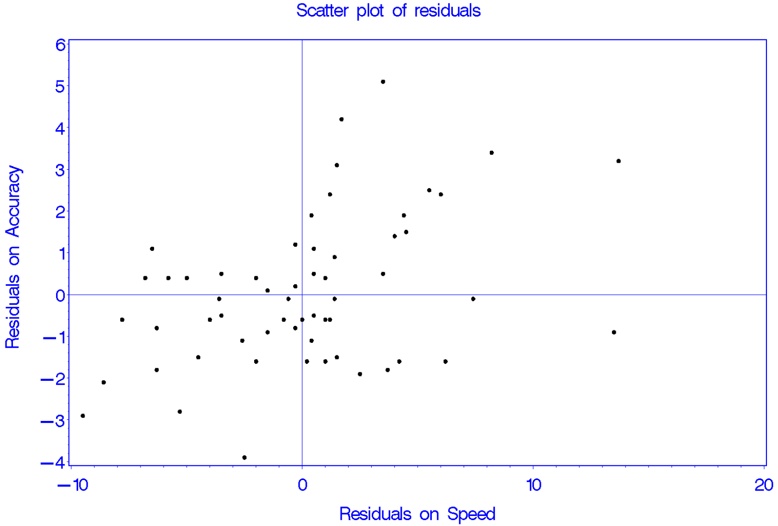
The assumptions are

- Multivariate normality.
- Equality of $\mathbf{\Sigma}$'s.
- Independence between and within groups.

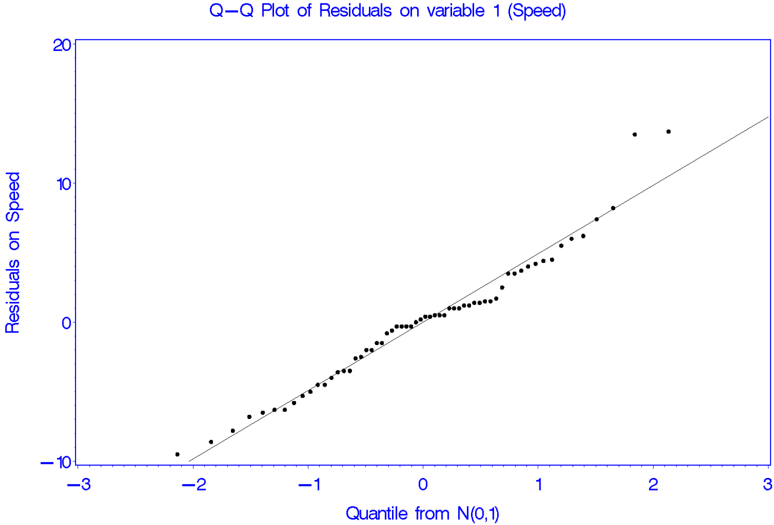
Checks on multivariate normality:

- Scatter plots of residuals.
- QQ plots of residuals.
- Scatter & QQ plots of principal components.
- Tests of univariate normality for each residuals.
- Tests of multivariate normality (see text).

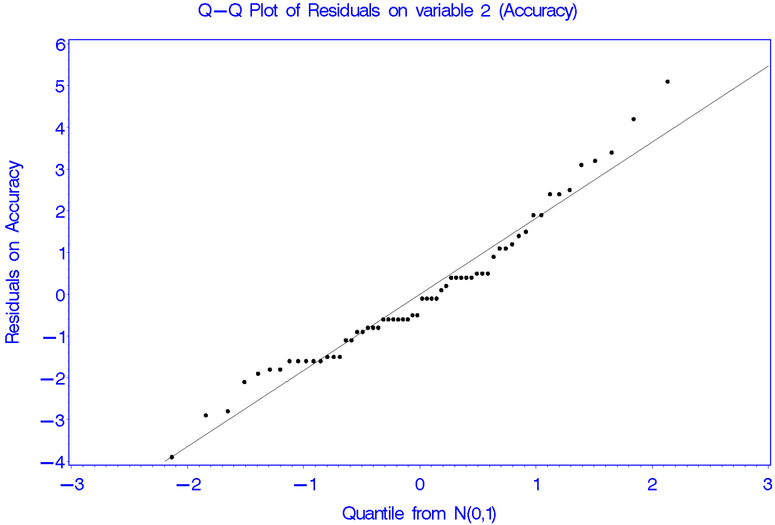
Shorthand Example: Scatter Plot



Shorthand Example: QQ Plot 1



Shorthand Example: QQ Plot 2



Equality of Covariance Matrices

- When there are equal n_{lk}'s (balanced), violating equality of $\mathbf{\Sigma}_{lk}$'s probably doesn't hurt too much.
- When there are unequal n_{lk}'s and $|\mathbf{\Sigma}_{lk}|$'s differ substantially, then will tend to make more errors.
  Where you make more Type I or Type II errors depends on how different the n_{lk}'s and $|\mathbf{\Sigma}_{lk}|$'s are from each other.
- Test of Equality of $\mathbf{\Sigma}$'s — this gets into ideas of tests and inference for covariance matrices, which leads into topics on canonical correlation and discriminate analysis.

Assume:

$$\mathbf{X}_{lj} \sim \mathcal{N}_p(\boldsymbol{\mu}_l, \mathbf{\Sigma}_l) \qquad \text{for } l = 1, \ldots, g \quad \text{and } j = 1, \ldots, n_l$$

and independent over the g groups and with groups.

The null hypothesis:

$$H_o : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_g \equiv \mathbf{\Sigma}$$

Testing Equality of Covariance Matrices

If $H_o : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_g \equiv \mathbf{\Sigma}$ is true, then

$$\mathbf{S}_{pool} = \mathbf{S} = \left(\frac{1}{\sum_{l=1}^{g}\nu_l}\right)\sum_{l=1}^{g}\nu_l \ln(|\mathbf{S}_i|)$$

where $\nu_l = (n_i - 1) = df(\mathbf{S}_i)$. When H_o is true, then $\mathbf{S}_{pool}$ is an unbiased estimate of $\mathbf{\Sigma}$.

Box's test (1949ish):

$$M = \left(\sum_{l=1}^{g}\nu_l\right)\ln(|\underbrace{\mathbf{S}}_{\mathbf{S}_{pool}}|) - \sum_{l=1}^{g}\nu_l \ln(|\mathbf{S}_l|)$$

Test statistic is asymptotically

$$\underbrace{\left[1 - \frac{2p^2+3p-1}{6(p+1)(g-1)}\left(\sum_{l=1}^{g}\left(\frac{1}{\nu_l}\right) - \frac{1}{\sum_{l=1}^{g}\nu_l}\right)\right]}_{\text{Correction factor}} M \approx \chi^2_{(g-1)p(p+1)/2}$$

Example: Box's Test of Equality of Σ_{lk}

$H_o : \mathbf{\Sigma}_{A,C_1} = \mathbf{\Sigma}_{A,C_2} = \mathbf{\Sigma}_{A,C_3} = \mathbf{\Sigma}_{B,C_1} = \mathbf{\Sigma}_{B,C_2} = \mathbf{\Sigma}_{B,C_3}$ versus
H_a : not all $\mathbf{\Sigma}_{lk}$ are equal.
The method by condition sample covariance matrices:

$$\mathbf{S}_{A,C_1} = \begin{pmatrix} 34.68 & 8.40 \\ 8.40 & 5.07 \end{pmatrix} \qquad \mathbf{S}_{B,C_1} = \begin{pmatrix} 29.73 & 1.47 \\ 1.47 & 2.93 \end{pmatrix}$$

$$\mathbf{S}_{A,C_2} = \begin{pmatrix} 38.50 & 4.17 \\ 4.17 & 7.43 \end{pmatrix} \qquad \mathbf{S}_{B,C_2} = \begin{pmatrix} 14.50 & 3.72 \\ 3.72 & 1.83 \end{pmatrix}$$

$$\mathbf{S}_{A,C_3} = \begin{pmatrix} 19.16 & 3.78 \\ 3.78 & 1.66 \end{pmatrix} \qquad \mathbf{S}_{B,C_3} = \begin{pmatrix} 11.56 & 2.33 \\ 2.33 & 1.60 \end{pmatrix}$$

Pooled Covariance Matrix:

$$\mathbf{S}_{pool} = \begin{pmatrix} 25.69 & 3.91 \\ 3.91 & 3.42 \end{pmatrix}$$

Example of Box's Test continued

$$\begin{aligned} M &= \Big(\sum_{l,k}(n_{lk}-1)\Big)\ln|\mathbf{S}_{pool}| - \sum_{l,k}(n_{lk}-1)|\mathbf{S}_{lk}| \\ &= 6(10-1)\ln(69.14) - (10-1)(\ln(105.14 + \ln(268.82) + \ln(20.30) \\ &\quad + \ln(85.07) + \ln(12.73) + \ln(13.04)) \\ &= 228.75 - 205.34 \\ &= 23.41 \end{aligned}$$

Since $n_{lk} = 10$ for all $K = (2)(3) = 6$ combinations of method and condition (i.e., equal cell sizes), the correction factor simplifies to

$$1 - \frac{(2p^2 + 3p - 1)(g+1)}{6(p+1)gn} = 1 - \frac{(2(2)^2 + 3(2) - 1)(6+1)}{6(2+1)10(6)} = .9157$$

Test statistic $= .9158(23.41) = 21.44$ with $df = (K-1)p(p+1)/2 = (6-1)2(2+1)/2 = 15$, which using the χ^2 distribution has p-value $= .123$. Retain null hypothesis; that is, the equal covariance matrix assumption is reasonable.

Box's Test— the easier method

```
data shorthand;
 input speed accuracy method $ conditin $;
  if conditin='C1' and method='A' then group=1;
  else if conditin='C1' and method='B' then group=2;
  else if conditin='C2' and method='A' then group=3;
  else if conditin='C2' and method='B' then group=4;
  else if conditin='C3' and method='A' then group=5;
  else if conditin='C3' and method='B' then group=6;
proc discrim simple pool=test Wcov Pcov list;
  class group;
  var speed accuracy;
```

...and the output...

PROC DISCRIM output

```
Notation:

K       =   Number of Groups

P       =   Number of Variables

N       =   Total Number of Observations - Number of Groups

N(i)    =   Number of Observations in the i'th Group - 1

                                              N(i)/2
            Π |Within SS Matrix(i)|
V       =   -----------------------------------
                                              N/2
                 |Pooled SS Matrix|
```

PROC DISCRIM output

```
                    -                      -    2
                    |       1         1    |  2P + 3P - 1
RHO   =   1.0 -     |   SUM -----  -  ---  |  -------------
                    |       N(i)      N    |  6(P+1)(K-1)
                    -                      -

DF = .5(K-1)P(P+1)
                                          |  PN/2            |
                                          | N   V            |
Under the null hypothesis:  -2 RHO ln     | ---------------- |
                                          |   PN(i)/2        |
                                          | Π N(i)           |
is distributed approximately as Chi-Square(DF).
 Chi-Square   DF   Pr > ChiSq

  21.218963   15     0.1300
```

PROC DISCRIM output

```
Since the Chi-Square value is not significant at the 0.1
level, a pooled covariance matrix will be used in the
discriminant function.  Reference:  Morrison, D.F. (1976)
Multivariate Statistical Methods p252
```

There is a more recent version of Morrison.

Checking Independence Assumption

- MANOVA is generally relatively robust to violations of multivariate normality and equal covariance matrices; however, it is not robust to violations of
  - Independence of observations within groups.
  - Independence of observations between groups.
- Violation of independence generally increases the Type I error rate (i.e., it's higher than what you think it is, and possibly much larger).
- You can check for possible violation by computing the intra-class correlation for each of the variables using

$$r_{intra} = \frac{MS_{between} - MS_{error}}{MS_{between} + (n-1)MS_{error}}$$

Checking Independence Assumption (continued)

- You can also consider the data collection procedure/method, because this is where dependence can slip in.
- It's the independence within that is most likely to be violated. In some cases this can be dealt with.

Unbalanced Designs

Similar to 1-way ANOVA.

- With balanced designs (or proportional)

$$SSCP_{total} = SSCP_A + SSCP_B + SSCP_{AB} + SSCP_{residual}$$

  or

$$SSCP_{model} = SSCP_A + SSCP_B + SSCP_{AB}$$

  and interpretation is straight forward.
- With unbalanced designs, the partitioning of the is not longer unique. It depends on
  - The model.
  - Various sub-models of model as specified by the order in which various SSCP are extracted. . .

These notes are based on Khattree & Naik.

Order in which SSCP are extracted

Suppose

$$\mathbf{X}_{lkr} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{lk} + \boldsymbol{\epsilon}_{lkr}$$

where $l = 1, \ldots, g$, $k = 1, \ldots, b$, and $r = 1, \ldots, n_{lk}$.
We could partition total SSCP as

$$SSCP_{\text{total corrected}} = SSCP_{A|\mu} + SSCP_{B|\mu,A} + SSCP_{AB|\mu,A,B}$$

or

$$SSCP_{\text{total corrected}} = SSCP_{B|\mu} + SSCP_{A|\mu,B} + SSCP_{AB|\mu,A,B}$$

When the design is unbalanced, $SSCP_{A|\mu} \neq SSCP_{A|\mu,B}$
There are 4 different ways of computing SSCP's, which can lead to different results and different interpretation.

Type I SSCP: "Sequential SSCP"

The partitioning of the model SSCP into component SSCP due to each variable or effect (including interactions) as it's added sequentially to the model as given in the model statement in proc glm in SAS.
e.g.,

$$\textit{model} \;\; x1\; x2 = A \;\; B \;\; C \;\; A*B \;\; A*C \;\; B*C \;\; A*B*C$$

That is all main effects, 2–way interactions and a 3-way interaction. The SSCP for $A*C = SSCP(A*C|\mu, A, B, C, A*B)$; that is, SSCP is adjusted for all previously entered terms,

$$\textit{model} \;\; x1\; x2 = \underbrace{A \;\; B \;\; C \;\; A*B}_{\text{already entered}} \;\; A*C \;\; B*C \;\; A*B*C$$

Example Type I SSCP: "Sequential SSCP"

To demonstrate this, I deleted some observations from the shorthand data set to make it a bit unbalanced. The sample sizes are now

	Practice Condition			
Method	C1	C2	C3	Total
A	5	10	8	23
B	10	7	6	23
Total	15	17	14	46

Below are the hypothesis SSCP matrices for two different orders:

Order: $M \quad C \quad C*M$ $\qquad$ $C \quad M \quad C*M$

$$\mathbf{H}_m = \begin{pmatrix} 164.54 & 83.22 \\ 83.22 & 42.09 \end{pmatrix} \qquad \mathbf{H}_m = \begin{pmatrix} 43.09 & 0.72 \\ 0.72 & 0.01 \end{pmatrix}$$

Type I: "Sequential SSCP" continued

- The sum of $\mathbf{H}_m + \mathbf{H}_c$ will be equivalent regardless of whether method order condition was entered first.
- With Type I SSCP, the sum of SSCP matrices of all the effects in the model will sum to the total for the model, regardless of the order in which they are entered into the model.
- The total SSCP is completely partitioned.

Type II or "Partial SSCP"

With Type II, an SSCP matrix equals the increase in model SSCP for a particular variable (effect). e.g.,

$$\text{model} \;\; x1\; x2 = A \;\; B \;\; C \;\; A*B \;\; A*C \;\; B*C \;\; A*B*C$$

That is all main effects, 2–way interactions and a 3-way interaction.
Consider the SSCP for $A*C$. The SSCP for $A*C$ is the increase in model SSCP by adding $A*C$ to the model that has A, B, $A*C$ and $B*C$ in it.
The three-way interaction $A*B*C*$ contains $A*C$ in it, so the SSCP for $A*C$ is "in the model". The effects $A*B*C$ and $A*C$ are correlated.
With Type II, the sum of the SSCP's do not equal the total model SSCP.

Example: Type II or "Partial SSCP"

Below are the SSCP matrices for Type I and Type II and the totals:

	Type I: Sequential		Type II: Partial
	$M \quad C \quad C*M$	$C \quad M \quad C*M$	
$\mathbf{H}_m$	$\begin{pmatrix} 164.54 & 83.22 \\ 83.22 & 42.09 \end{pmatrix}$	$\begin{pmatrix} 43.09 & 0.72 \\ 0.72 & 0.01 \end{pmatrix}$	$\begin{pmatrix} 43.09 & 0.72 \\ 0.72 & 0.01 \end{pmatrix}$
$\mathbf{H}_c$	$\begin{pmatrix} 886.58 & 925.59 \\ 925.59 & 966.37 \end{pmatrix}$	$\begin{pmatrix} 1008.02 & 1008.08 \\ 1008.08 & 1008.45 \end{pmatrix}$	$\begin{pmatrix} 886.58 & 925.59 \\ 925.59 & 966.37 \end{pmatrix}$
$\mathbf{H}_{cm}$	$\begin{pmatrix} 280.31 & 137.06 \\ 37.06 & 67.64 \end{pmatrix}$	$\begin{pmatrix} 280.31 & 137.06 \\ 37.06 & 67.64 \end{pmatrix}$	$\begin{pmatrix} 280.31 & 137.06 \\ 37.06 & 67.64 \end{pmatrix}$
Sum	$\begin{pmatrix} 1331.43 & 1145.87 \\ 1145.87 & 1076.10 \end{pmatrix}$		$\begin{pmatrix} 1209.98 & 1063.38 \\ 1063.38 & 1034.0 \end{pmatrix}$

Type III & Type IV SSCP

- These are similar to Type II in terms of the idea; that is, they are a kind of partial SSCP, but what you condition on is different.
- Type III SSCP give the increase in model SSCP due to adding a particular variable or interaction (effect) to a model that contains all other variables and effects listed in the model.
- For example, if

$$\textit{model} \quad x1\ x2 = A \quad B \quad C \quad A*B \quad A*C \quad B*C \quad A*B*C$$

That is all main effects, 2–way interactions and a 3-way interaction.

The Type III SSCP for $A*C$ is the increase in model SSCP that is obtained by adding $A*C$ to a model that has A, B, C, $A*B$, $B*C$ and $A*B*C$.

In general these will tend to be smaller than Types I and II

Type IV and When to Use Which

Type IV are designed for situations where there are empty cells.

- For balanced designs, the 4 ways to compute SSCP yield the same results.
- For unbalanced designs, the 4 ways can lead to different conclusions.
- Type III is appropriate when interest is in comparing the effects of experimental variables.
- For model building purposes (finding a good model for prediction), Types I and/or II might be preferable.
- If there are design parameters, make sure that tests do not involve these.
- The choice should be made on a case-by-case basis with thought given to what it is you want to test.

Conclusions

- Any design that you can form for univariate case can be formed for the multivariate situation.
- This can be done by creating an appropriate design matrix in the general linear model context.
- Models can include fixed (as we have done), but also random effects (as in the univariate case).
- It is possible to reformulated MANOVA as a linear mixed model, and hence provides a natural way to handle missing observations (be sure there are systematic reasons for missing values). See Snijders & Bosker. ...see Goldstein.