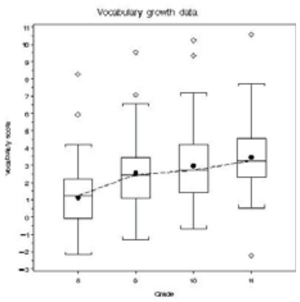
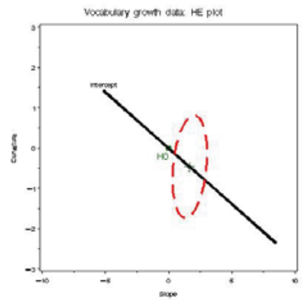
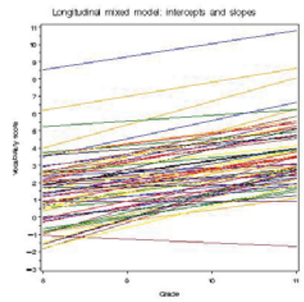






Repeated measures: ANOVA and MANOVA

Psychology 6140

Repeated measures designs

- Learning/longitudinal designs
  - Each subject measured on the same task over multiple occasions

Subj	Trial 1	Trial2	…	Trial p
S1	12	16	…	29
S2	15	18	…	32

  - Or, there can also be 1 or more between-subject factors

Group	Subj	Trial 1	Trial 2	…	Trial p
Control	S1	12	16	…	29
Control	S2	15	18	…	32
Treated	S3	21	26	…	47
Treated	S4	19	24	…	38

Repeated measures designs

- Within-subject designs
  - Each subject tested on different tasks or under different conditions

	A1			A2		
	B1	B2	B3	B1	B2	B3
S1	14	18	10	21	28	27
S2	19	22	16	25	30	29

  - NB: Scores for same S are dependent; scores for different Ss are independent
  - Dependence must be taken into account in analysis – Why ??

Repeated measures designs

- Pre-post designs
  - Pre-test(s) – Treatment – Post-test(s)

  - Each S serves as his/her own 'control'
    - Sometimes treated as ANCOVA (pretest as a covariate)
    - Sometimes treated using 'gain' scores: post-pre, followup-pre
  - There can also be multiple outcome measures at each occasion
    - E.g., depression, anxiety, self-worth x (pre, post)
    - These are "doubly-multivariate" designs

Why use repeated measures designs?

- Control for individual differences
  - When individuals vary widely, within-S comparisons may be more sensitive than between-S comparisons
  - Between-S designs assume random assignment, making groups equivalent, but only on average, in the long run

Subject	Control	Treatment
Subject 1	12	14
Subject 2	25	28
Subject 3	29	32
Subject 4	54	57

Diff between control & treatment is small, but every subject did better under treatment

Within-S test can be far more powerful

Why use repeated measure designs?

- Change or learning – little choice!
  - Vocabulary growth in 2^{nd} language learning
  - Student math achievement in grades 1-6
  - Therapy outcomes over sessions
- Special populations, few available subjects
  - Eye-hand coordination in astronaut trainees
  - Motor skill relearning in stroke patients
  - Perception studies
    - Many trials, many combinations of stimulus factors
    - Often n=2, 3, … (authors)

Caveats: Carryover, order effects

- Effect of a given treatment may depend on what happened before
  - Practice effect – better over time regardless of treatment
  - Fatigue – worse over time
  - Priming – A, then B different from B, then A
- Counter-balance: vary order over subjects
  - E.g., latin squares

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

Each treatment in each position equally often

Analysis methods: Overview

- Univariate, repeated measures
  - Different error term for each effect
  - Strong assumptions (Σ: compound symmetry) for validity of within-S effects
- MANOVA
  - No additional assumptions (Σ: unstructured)
  - Test all hypotheses via GLH-- $H_0 : \mathbf{L\,B\,M} = \mathbf{0}$
- Mixed model
  - Most flexible (Σ: unstructured, CS, AR(1), …)
  - Allows missing data, drop-out, unequal time points
  - Also handles fixed and random factors

Univariate approach: Hypothesis tests

- Between-S effects: tested on sums (means) over repeated measures
- Different error terms for different effects

2 Within

(i) Source of variation	df
Between subjects	$n-1 = 9$
Within subjects	$n(pq-1) = 50$
A	$(p-1) = 1$
B	$(q-1) = 2$
AB	$(p-1)(q-1) = 2$
$A \times$ subjects	$(p-1)(n-1) = 9$
$B \times$ subjects	$(q-1)(n-1) = 18$
$AB \times$ subjects	$(p-1)(q-1)(n-1) = 18$
Total	$npq-1 = 59$

A between/B within

(ii) Source of variation	df
Between subjects	$np-1 = 19$
A	$p-1 = 1$
Subjects (a_1)	$n-1 = 9$
Subjects (a_2)	$n-1 = 9$
Within subjects	$np(q-1) = 40$
B	$(q-1) = 2$
AB	$(p-1)(q-1) = 2$
$B \times$ subjects (a_1)	$(q-1)(n-1) = 18$
$B \times$ subjects (a_2)	$(q-1)(n-1) = 18$
Total	$npq-1 = 59$

Proper error terms based on E(MS): To test H_0: term=0

$$E(F) = \frac{\sigma^2_{error} + E(MS_{term})}{\sigma^2_{error}}$$

Univariate approach

- Contrast this with completely randomized, between-S design

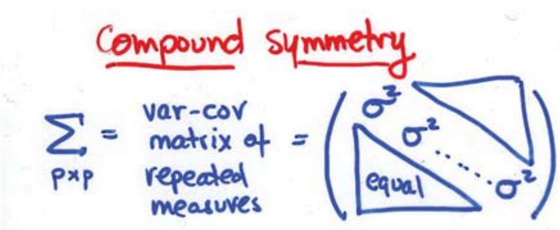
Completely randomized Source of variation	df
Between treatment means	$pq-1 = 5$
A	$(p-1) = 1$
B	$(q-1) = 2$
AB	$(p-1)(q-1) = 2$
Within treatments "error"	$pq(n-1) = 54$
Total	$npq-1 = 59$

Each effect here is tested against within-cells Error MS

No need to worry about the proper error term

Univariate approach: assumptions

- The validity of these tests depends on assumption about the pattern of correlations among the repeated measures
  - Only applies to within-S effects
  - Strongest form: compound symmetry



This implies:
- Equal variances
- Equal correlations
- Unlikely for longitudinal data
- Possible for split-plot designs

Univariate approach: assumptions

- Huynh-Feldt conditions (weaker): Sphericity
  - Variances and covariances may differ, as long as they can be expressed as:

$$\begin{cases} \sigma_i^2 = \alpha_i + \alpha_i + \tau \\ \sigma_{ij} = \alpha_i + \alpha_j \end{cases}$$

  - True, iff $(y_1\text{-}y_2)$, $(y_2\text{-}y_3)$, … $(y_{p-1} - y_p)$ have constant variance and are uncorrelated
  - → Orthogonal contrasts of repeated measures, **Y M**, have

$$\mathbf{\Sigma}_{YM} = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}$$

Spherical covariance

(Maunchly's test)

Ways around these problems

Effect of violations:
- P-values biased downward (actual α much larger than nominal α)

(a) Univariate Fixups

Box: Null dist^n of F^* for within-S effects can be approximated by adjusting df for test

$$F^*_{\text{within-S}} \sim F(\epsilon \cdot df_H, \epsilon \cdot df_E)$$

where $\frac{1}{p-1} < \epsilon \le 1$

worst case violation — HF conditions met

- Box-test (1 within-S factor): $F(\overset{1}{\cancel{p-1}}, \cancel{(p-1)}(N-1)) \longrightarrow F(1, N-1)$

very conservative → low power

GG and HF corrections

Key idea: degree of departure from sphericity can be assessed by estimating ε

- Geiser-Greenhouse
- Hughn-Feldt

} different ways of estimating ϵ → use $F(\hat{\epsilon} \cdot df_H, \hat{\epsilon}\, df_E)$

G-G estimate	H-F estimate
• more conservative	• more liberal
• good when ϵ is small ($\epsilon < .5$)	• good when ϵ is large ($\epsilon \ge .5$)
[H-F conditions very wrong]	[H-F approximately met]

GG & HF corrections

G-G $\hat{\epsilon}$ H-F $\tilde{\epsilon}$

Monte Carlo study of effect of Box correction for degrees of freedom on level of significance

				$\hat{\varepsilon}$				$\tilde{\varepsilon}$		
ε	n	k (p)	α:	0.10	0.05	0.01	α:	0.10	0.05	0.01
0.363	10	5		0.096	0.052	0.012		0.105	0.060	0.018
	15	5		0.096	0.051	0.012		0.101	0.054	0.015
	20	5		0.098	0.054	0.012		0.101	0.033	0.015
0.752	10	5		0.080	0.034	0.007		0.102	0.055	0.013
	15	5		0.082	0.038	0.008		0.096	0.051	0.013
	20	5		0.094	0.044	0.011		0.102	0.051	0.014
0.831	10	5		0.078	0.036	0.006		0.101	0.053	0.013
	15	5		0.085	0.040	0.009		0.101	0.053	0.014
	20	5		0.091	0.046	0.080		0.103	0.053	0.012
1.00	10	5		0.071	0.029	0.003		0.095	0.046	0.009
	15	5		0.081	0.034	0.005		0.095	0.047	0.009
	20	5		0.093	0.046	0.006		0.105	0.050	0.010

From Huynh and Feldt (1979).

How the correction affects critical F-values (& therefore p-values)

Eg. w/ $n=5$, $p=3$

	ϵ	$\epsilon\, df_h, \epsilon\, df_e$	$\alpha = .05$ critical F	
	1.0	2,9	4.46	liberal
	0.8	1.6, 6.4	5.21	
$\frac{1}{p-1} =$	0.5	1, 4	7.71	conservative

Note: if $F^* > F(\epsilon = \frac{1}{p-1})$, it will be significant for any larger ϵ.

Summary:

- Only matters when ε << 1
- HF better when ε > 0.5
- GG better when ε < 0.5

MANOVA approach

- Between-S effects are tested in the same way
  - Same results as in univariate approach
- NO assumption required about structure of Σ
- Actual α error rates are approximately correct or are exact
- However, smallish sample sizes may have lower power
  - Power increases with ratio N/p
  - Univariate approach “buys power” with stronger assumptions
- Statistical tests: based on Wilks’ Λ, HLT, Roy, …

How to decide?

(1) [Not recommended]

Mauchley's test (for transformed repeated measures) $\xrightarrow{\text{reject sphericity}}$ MANOVA

Mauchley's test $\xrightarrow{\text{accept}}$ Univariate mixed model

(2) [Recommended]

(a) Examine univariate tests (& ϵ values) & MANOVA tests (if $s > 1$)

(b) $\epsilon \approx 1 \rightarrow$ univariate (no adjustment necessary)

$\frac{2}{p-1} < \epsilon < 1 \rightarrow$ univariate, w/ adj df

$\frac{1}{p-1} < \epsilon < \frac{2}{p-1} \rightarrow$ MANOVA

Mixed model approach (proc mixed)

- Univariate analysis w/ choice of various covariance structures

UNSTRUCTURED $\Sigma =$ anything

Compound Symmetry $\Sigma = \begin{pmatrix} \sigma^2 & c & c \\ c & \sigma^2 & c \\ c & c & \sigma^2 \end{pmatrix}$

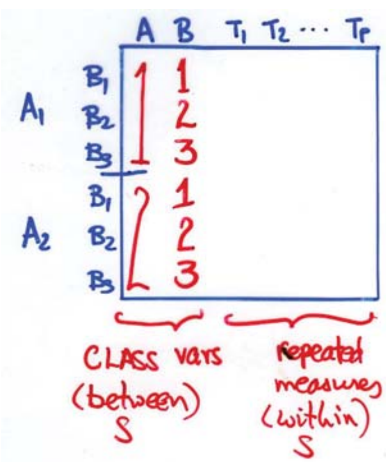
Uncorrelated $\Sigma = \begin{pmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \sigma^2 \end{pmatrix}$

⋮

(many choices)

Mixed models are far more general. The point here is that the covariance structure can be part of the model

Analysis: data structures (SAS)

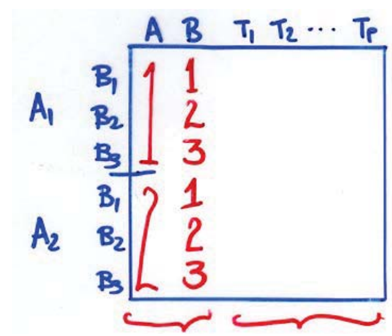


"Wide" format: 1 obs / experimental unit
- repeated measures: `repeated` stmt
- MANOVA: `manova` statement

```
proc glm;
  class A B;
  model T1-T8 = A B A*B /nouni; (or A|B)
  manova H=A B A*B;
  repeated trials 8 polynomial /summary;
           name  # levels  contrasts
or repeated hand 2, trials 4 polynomial;   several crossed within-S factors
```

NB: **any** missing data removes the whole case!

Analysis: data structures (R)



Between factors

Repeated measures (Within)

"Wide" format: 1 obs / experimental unit
- `idata` data frame: repeated factors
- MANOVA: car::`Anova()` function

between-s design

```
mod <- lm(cbind(T1, T2, ..., T8) ~ A * B)
within <- expand.grid(hand=c("L", "R"),
                      trial=ordered(1:4))
Anova(mod, idata=within, idesign= ~ hand * trial)
```

within-s design

```
> within
  hand trial
1    L     1
2    R     1
3    L     2
4    R     2
5    L     3
6    R     3
7    L     4
8    R     4
```

Relates the variables (T1:T8) to within-S factors

Analysis: data structures

Subj	A	B	Trial	SCORE
1	1	1	1	–
			2	–
			3	–
2	1	1	1	–
			2	–
			3	–
3	1	2	1	–
			2	–
			3	–
4	1	2	1	–
			2	–
			3	–

"Long" format: 1 obs / response

- Needed for plotting
- Allows missing data (use available)
- PROC GLM: can specify error terms
- PROC MIXED: can specify covariance structure
- R: **`aov()`**, like GLM, with error terms
- R: `nmle` package for mixed models

Example: Vocabulary growth study

- Vocabulary scores for a cohort of n=64 children were assessed in Grades 8-11 at the University of Chicago Lab. School
- Interest is focused on the form of vocabulary growth in this age range.
- e.g., does it decelerate, like physical growth?

subject	grade8	grade9	grade10	grade11
1	1.75	2.60	3.76	3.68
2	0.90	2.47	2.44	3.43
3	0.80	0.93	0.40	2.27
4	2.42	4.15	4.56	4.21
5	-1.31	-1.31	-0.66	-2.22
6	-1.56	1.67	0.18	2.33
7	1.09	1.50	0.52	2.33
8	-1.92	1.03	0.50	3.04
9	-1.61	0.29	0.73	3.24
10	2.47	3.64	2.87	5.38
...	...	...	...	...

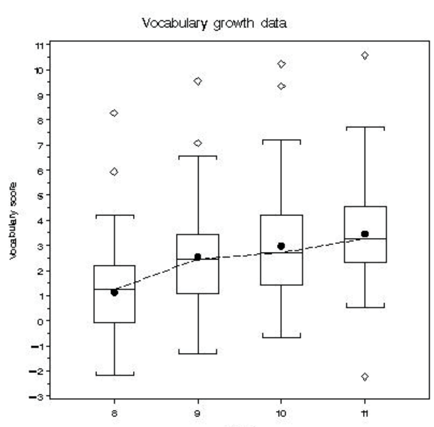
Data in wide format


```
data vlong; set vocab;
  keep subject grade vocab;
  grade=8;  vocab=grade8; output;
  grade=9;  vocab=grade9; output;
  grade=10; vocab=grade10; output;
  grade=11; vocab=grade11; output;
  run;
%boxplot(data=vlong, var=vocab,
      class=Grade);
```

For plotting, reshape to long format, e.g., to

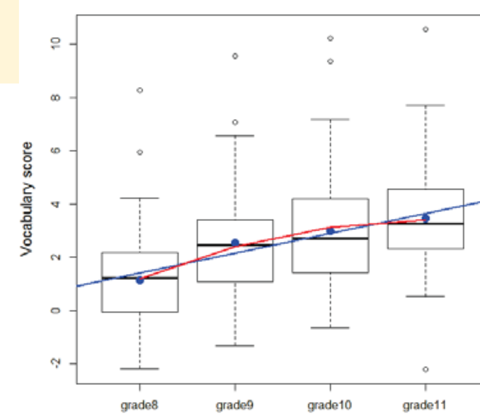
plot vocab * grade

subject	grade	vocab
1	8	1.75
1	9	2.60
1	10	3.76
1	11	3.68
2	8	0.90
2	9	2.47
2	10	2.44
2	11	3.43
3	8	0.80
3	9	0.93
3	10	0.40
3	11	2.27
...	...	...



In R, no need to reshape for this (boxplot() can plot a data frame)

```
data(VocabGrowth, package="heplots")
boxplot(VocabGrowth, ylab="Vocabulary score")
means <- colMeans(VocabGrowth)
points(1:4, means, pch=16, cex=1.5, col="blue")
    # plot linear trend
abline(lm(means ~ I(1:4)), col="blue", lwd=2)
    #  plot quadratic model
quad.mod <- lm(means ~ poly(I(1:4),2))
lines(1:4, predict(quad.mod), col="red")
```



Analysis: Univariate & MANOVA

- PROC GLM:
  - REPEATED statement gives **both** univariate and multivariate tests
  - Can specify type of contrasts for repeated factor(s)

# of levels

No between factors

```
title 'Multivariate Repeated Measures Analysis';
proc glm data=vocab;
    model grade8-grade11 = /nouni;
    repeated grade 4 (8 9 10 11) polynomial /
        short summary printh printe;
run;
```

Specific tests of within-S contrasts, e.g., Linear, Quadradic, Cubic

For a quantitative factor

MANOVA output

Sphericity:

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.9030496	6.2942969	0.2786
Orthogonal Components	5	0.9030496	6.2942969	0.2786

OK!

Overall grade effect:

MANOVA Test Criteria and Exact F Statistics
for the Hypothesis of no **grade** Effect
H = Type III SSCP Matrix for grade
E = Error SSCP Matrix

S=1 M=0.5 N=29.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.17422126	96.38	3	61	<.0001
Pillai's Trace	0.82577874	96.38	3	61	<.0001
Hotelling-Lawley Trace	4.73982748	96.38	3	61	<.0001
Roy's Greatest Root	4.73982748	96.38	3	61	<.0001

Univariate tests

Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Adj Pr > F G - G	Adj Pr > F H - F
grade	3	193.9456531	64.6485510	78.77	<.0001	<.0001	<.0001
Error(grade)	189	155.1194469	0.8207378				

Greenhouse-Geisser Epsilon 0.9428
Huynh-Feldt Epsilon 0.9917

Summary so far:

- Mean vocabulary scores differ significantly over grade ✓
- Supported both by multivariate and univariate tests ✓
- GG and HF ε indicate univariate assumptions are valid ✓
- What about trends over grade?

Univariate tests: Within-S contrasts

Repeated Measures Analysis of Variance
Analysis of Variance of Contrast Variables

grade_N represents the nth degree polynomial contrast for grade

Contrast Variable: grade_1 Linear

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	177.1464003	177.1464003	221.59	<.0001
Error	63	50.3654197	0.7994511		

✓

Contrast Variable: grade_2 Quadratic

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	13.65302500	13.65302500	19.53	<.0001
Error	63	44.03157500	0.69891389		

✓

Contrast Variable: grade_3 Cubic

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	3.14622781	3.14622781	3.26	0.0756
Error	63	60.72245219	0.96384845		

×

$\sum_{contrasts} =$

Source	DF	Type III SS
grade	3	193.9456531
Error(grade)	189	155.1194469

Overall test of Grade

Where these tests come from: H & E matrices

H matrix:

	grade_1	grade_2	grade_3
grade_1	177.1464	-49.1791	23.60811
grade_2	-49.1791	13.65	-6.55404
grade_3	23.60811	-6.55404	3.146227

Overall test of Grade

trace(H) = 193.94
$= SS_{Grade}$

E matrix:

	grade_1	grade_2	grade_3
grade_1	50.3654	12.0796	-3.1401
grade_2	12.0796	44.0316	-4.0826
grade_3	-3.1401	-4.0826	60.7225

trace(E) = 155.12
$= SS_{Error(grade)}$

Trend tests: Linear Quadratic Cubic

Source	DF	Type III SS (Linear)	Type III SS (Quadratic)	Type III SS (Cubic)
Mean	1	177.1464003	13.65302500	3.14622781
Error	63	50.3654197	44.03157500	60.72245219

MANOVA test:

$| H - \lambda E | = 0$

Same analysis in R

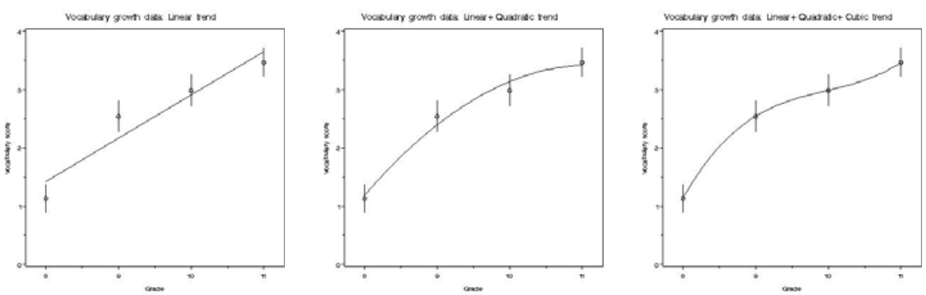
Within-s factor (poly contrasts)   Repeated measures   No between factors

```
# std Multivariate & Univariate repeated measures analysis
Vocab.mod <- lm(cbind(grade8,grade9,grade10,grade11) ~ 1,
                           data=VocabGrowth)

within <-data.frame(grade=ordered(8:11))
# basic short summary: multivariate tests
(Vocab.aov <- Anova(Vocab.mod, idata=within, idesign=~grade))
# detailed summary: Univariate & multivariate tests
summary(Vocab.aov)
```

Multivariate test:

```
Type III Repeated Measures MANOVA Tests: Pillai test statistic
            Df test stat approx F num Df den Df    Pr(>F)
(Intercept)  1   0.65289  118.498      1     63 4.115e-16 ***
grade        1   0.82578   96.376      3     61 < 2.2e-16 ***
```


Visualizing results: meanplots

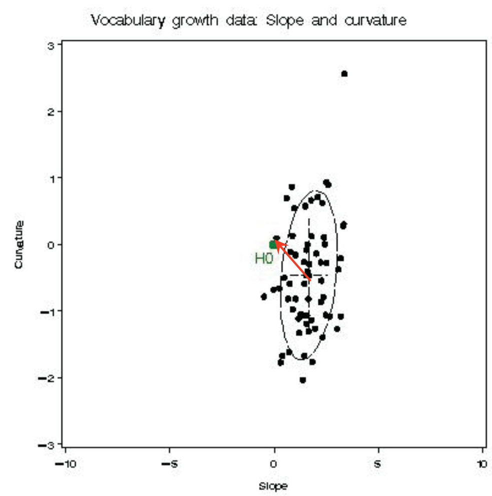


%meanplot(data=vlong, var=vocab, class=Grade, interp=rl); Linear

%meanplot(data=vlong, var=vocab, class=Grade, interp=rq); Quadratic

%meanplot(data=vlong, var=vocab, class=Grade, interp=rc); Cubic

Visualizing results: HE plots

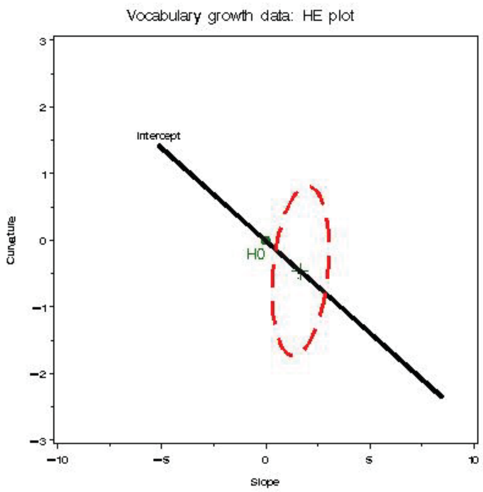


The MANOVA is based on analysis of **Y M**, where **M** gives within-S contrasts

$$\mathbf{M} = \begin{array}{c} \begin{pmatrix} -3 & 1 & -1 \\ -1 & -1 & 3 \\ 1 & -1 & -3 \\ 3 & 1 & 1 \end{pmatrix} \\ \begin{array}{ccc} \text{Lin} & \text{Quad} & \text{Cubic} \end{array} \end{array}$$

The plot shows the slopes and curvatures for individuals, with a 68% data ellipse

Visualizing results: HE plots



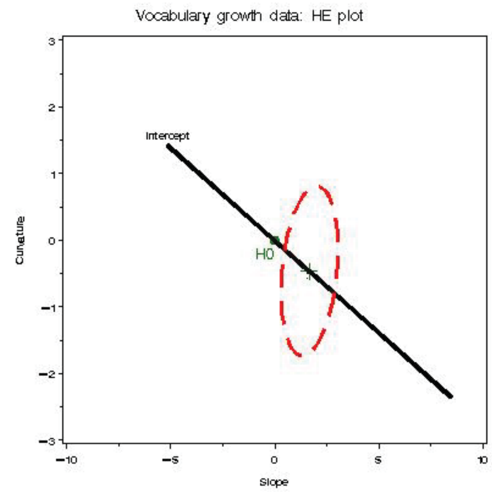
The MANOVA test for Grade is testing

$$H_0: \mu_{Lin} = \mu_{Quad} = \mu_{Cubic} = 0$$

The **H** matrix measures the distance between the actual means and (0, 0, 0)

The **E** matrix shows the covariation of slope, curvature and cubic effects

Visualizing results: HE plots



Interpretation:

H ellipse:

- mean slope > 0, mean curvature < 0
- more variation against H_0 in slope than curvature

E ellipse: those with larger slopes tend to have slightly larger curvature --- flatter trajectories

Alternative analyses: polynomial regression

```
title 'Polynomial regression, ignoring subject';
proc glm data=vlong;
   model vocab = grade|grade|grade / ss1;
   run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	193.945653	64.648551	15.83	<.0001
Error	252	1029.105122	4.083750		
Corrected Total	255	1223.050775			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
grade	1	177.1464003	177.1464003	43.38	<.0001
grade*grade	1	13.6530250	13.6530250	3.34	0.0687
grade*grade*grade	1	3.1462278	3.1462278	0.77	0.3809

The SS for grade, $grade^2$, $grade^3$ are correct, but the error term is wrong

Alternative analyses: polynomial regression

```
title 'Polynomial regression, including subject';
proc glm data=vlong;
   class subject;
   model vocab = subject grade|grade|grade / ss1;
   run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	66	1067.931328	16.180778	19.71	<.0001
Error	189	155.119447	0.820738		
Corrected Total	255	1223.050775			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
subject	63	873.9856750	13.8727885	16.90	<.0001
grade	1	177.1464003	177.1464003	215.84	<.0001
grade*grade	1	13.6530250	13.6530250	16.64	<.0001
grade*grade*grade	1	3.1462278	3.1462278	3.83	0.0517

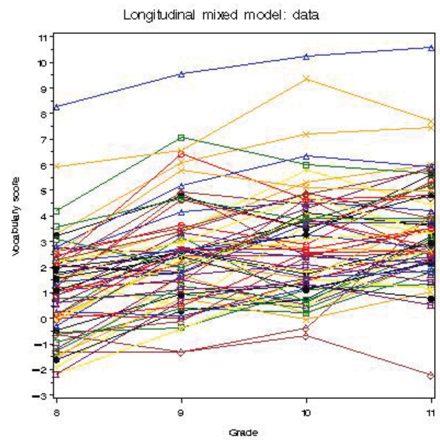
This gives results identical to the repeated measures univariate results (except that the pooled Error(grade) is used for all tests)

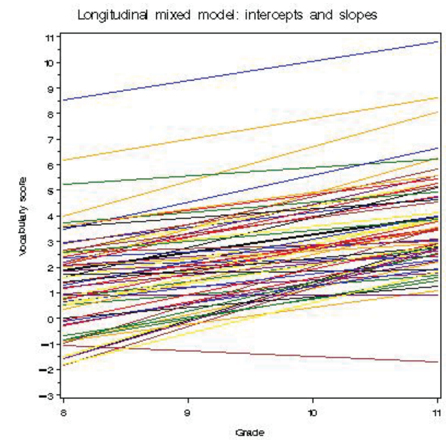
Preview: longitudinal mixed models

- Level 1 model: individual growth
  - Constant: $y_{it} = \beta_{i0} + \varepsilon_{it}$
  - Linear growth: $y_{it} = \beta_{i0} + \beta_{i1}\,(\text{Grade-8}) + \varepsilon_{it}$
  - Quadratic growth: $y_{it} = \beta_{i0} + \beta_{i1}\,(\text{Grade-8}) + \beta_{i2}\,(\text{Grade-8})^2 + \varepsilon_{it}$
- Interpretation:
  - β_{i0} is the true initial status of person i at grade 8
  - β_{i1} is the true slope of person i growth trajectory at grade 8
  - β_{i2} is the true curvature (change in slope) for person i per year
- These differ from traditional linear models in that we regard individual coefficients as random effects that can also be modeled

Longitudinal models: Individual growth curves

We can think of intercepts & slopes as random effects – how do they vary?





Preview: longitudinal mixed models

- Level 2 model: Random effects (intercepts & slopes as outcomes)
  - $\beta_{i0} = \gamma_{00} + \zeta_{0i}$
  - $\beta_{i1} = \gamma_{10} + \zeta_{1i}$

  where $\begin{pmatrix}\zeta_{0i}\\ \zeta_{1i}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{10} & \sigma_1^2\end{pmatrix}\right]$

- Level 2 model: Between individual effects
  - e.g., if individuals had been given different treatments
  - $\beta_{i0} = \gamma_{00} + \gamma_{01}\,\text{TREAT}_i + \zeta_{0i}$
  - $\beta_{i1} = \gamma_{10} + \gamma_{11}\,\text{TREAT}_i + \zeta_{1i}$

Mixed models allow us to model random effects in flexible ways and test hypotheses regarding pop^n variance components

PROC MIXED for longitudinal growth

```
%include data(vocab);
*-- Define grade so 0 = Grade 8 (initial status);
data vlong; set vocab;
     keep subject grade vocab;
     grade=0; vocab=grade8;  output;
     grade=1; vocab=grade9;  output;
     grade=2; vocab=grade10; output;
     grade=3; vocab=grade11; output;
          run;

*-- Linear growth;
proc mixed data=vlong noinfo method=ml covtest;
     class subject;
     model vocab = grade / solution;
     random intercept grade / subject=subject type=un;
run;

*-- Quadratic growth;
proc mixed data=vlong noinfo method=ml covtest;
     class subject;
     model vocab = grade|grade / solution;
     random intercept grade|grade / subject=subject type=un;
run;
```

Example: Pre-post design (2B, 1W)

```
subj     group       sex     pre      post      fol

  1      Control      M       2        3         3
  2      Control      M       4        3         4
  3      Control      M       6        5         7
  4      Control      F       5        3         4
  5      Control      F       4        6         4

  6      Treat_A      M       8        9         9
  7      Treat_A      M       5        8         9
  8      Treat_A      F       3        5         6
  9      Treat_A      F       4        4         5

 10      Treat_B      M       4        7         8
 11      Treat_B      M       3        5         6
 12      Treat_B      M       6        9         8
 13      Treat_B      F       6        6         8
 14      Treat_B      F       2        5         6
 15      Treat_B      F       3        7         7
 16      Treat_B      F       5        7         8
```

Data in wide format

NB: the Between-S design is unbalanced

```
              group
sex     control  A  B
  F           2  2  4
  M           3  2  3
```

Type II tests preferred for unbalanced designs

Plotting means

```
data  long;
   set repmes;
   phase ='1:Pre      ' ;  response=pre;  output;
   phase ='2:Post     ' ;  response=post; output;
   phase ='3:FollowUp' ;  response=fol;  output;
```

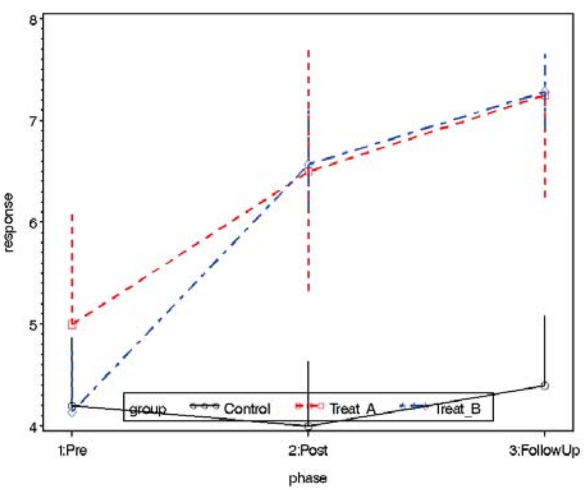
```
subj   group    sex  phase         response

  1    Control   M   1:Pre              2
  1    Control   M   2:Post             3
  1    Control   M   3:FollowUp         3
  2    Control   M   1:Pre              4
  2    Control   M   2:Post             3
  2    Control   M   3:FollowUp         4
  3    Control   M   1:Pre              6
  3    Control   M   2:Post             5
  3    Control   M   3:FollowUp         7
  4    Control   F   1:Pre              5
  4    Control   F   2:Post             3
  4    Control   F   3:FollowUp         4
  5    Control   F   1:Pre              4
  5    Control   F   2:Post             6
  5    Control   F   3:FollowUp         4
 ...    ...
```

Need to transpose data to the long format, to plot

response * factor(s)

NB: phase is prefixed by a number to sort properly on an axis

Plotting means




```
%meanplot(data=long,
    var=response,
    class=phase group);
```

This plot summarizes some of what we'll see in the statistical tests

Both univariate and multivariate tests are carried out with the REPEATED statement

- Use CONTRAST statements for between-S effects
- Specify within-S contrasts on REPEATED statement

```
title 'Group x Phase analysis';
proc glm data=repmes;
   class group;
   model pre post fol = group|sex / ss2 nouni;
   contrast 'Trt vs Control' group  -2  1  1 ;
   contrast 'Treat A vs B'   group   0  1 -1 ;
   REPEATED phase 3  contrast(1)
            / short summary printM printH printE;
run;
```

$$M^T = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{array}{l} \text{Post – Pre} \\ \text{Fol - Pre} \end{array}$$

Between-S tests

- Tests of Between-S effects appear separately, because they use within group SS as the error term
- The same Between-S tests are used with a MANOVA

Repeated Measures Analysis of Variance
Tests of Hypotheses for **Between Subjects Effects**

Source	DF	Type II SS	Mean Square	F Value	Pr > F
group	2	42.25729927	21.12864964	4.63	0.0377
sex	1	11.65729927	11.65729927	2.56	0.1410
group*sex	2	26.04825629	13.02412814	2.86	0.1045
Error	10	45.61111111	4.56111111		

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Trt vs Control	1	35.86440678	35.86440678	7.86	0.0187
Treat A vs B	1	0.37426901	0.37426901	0.08	0.7804

Within-S tests: Univariate

Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for **Within Subject Effects**

Source	DF	Type II SS	Mean Square	F Value	Pr > F	Adj Pr > F G - G	Adj Pr > F H - F
phase	2	33.50000000	16.75000000	20.87	<.0001	<.0001	<.0001
phase*group	4	15.73357664	3.93339416	4.90	0.0064	0.0122	0.0064
phase*sex	2	0.33357664	0.16678832	0.21	0.8141	0.7662	0.8141
phase*group*sex	4	2.04420114	0.51105028	0.64	0.6424	0.6116	0.6424
Error(phase)	20	16.05555556	0.80277778				

Greenhouse-Geisser Epsilon 0.7995
Huynh-Feldt Epsilon 1.4037

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F	Adj Pr > F G - G	Adj Pr > F H - F
phase*Trt vs Control	2	12.57909605	6.28954802	7.83	0.0031	0.0063	0.0031
phase*Treat A vs B	2	1.42397661	0.71198830	0.89	0.4275	0.4092	0.4275

Within-S tests: Multivariate

M matrix, from CONTRAST(1)

phase_N represents the contrast between the nth level of phase and the 1st

M Matrix Describing Transformed Variables

	pre	post	fol
phase_2	-1.000000000	1.000000000	0.000000000
phase_3	-1.000000000	0.000000000	1.000000000

Sphericity tests:

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	2	0.4349367	7.4929926	0.0236
Orthogonal Components	2	0.7492726	2.5978712	0.2728

Within-S tests: Multivariate

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no **phase** Effect
H = Type II SSCP Matrix for phase
E = Error SSCP Matrix

S=1 M=0 N=3.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.14947512	25.61	2	9	0.0002
Pillai's Trace	0.85052488	25.61	2	9	0.0002
Hotelling-Lawley Trace	5.69007670	25.61	2	9	0.0002
Roy's Greatest Root	5.69007670	25.61	2	9	0.0002

MANOVA Test Criteria and F Approximations for the Hypothesis of no **phase*group** Effect
H = Type II SSCP Matrix for phase*group
E = Error SSCP Matrix

S=2 M=-0.5 N=3.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.31773492	3.48	4	18	0.0283
Pillai's Trace	0.68518291	2.61	4	20	0.0667
Hotelling-Lawley Trace	2.13809442	4.69	4	9.8537	0.0221
Roy's Greatest Root	2.13379071	10.67	2	10	0.0033

Within-S tests: Multivariate

```
MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no phase*sex Effect
                    H = Type II SSCP Matrix for phase*sex
                           E = Error SSCP Matrix

                            S=1    M=0    N=3.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.95685724       0.20         2         9    0.8200
Pillai's Trace              0.04314276       0.20         2         9    0.8200
Hotelling-Lawley Trace      0.04508798       0.20         2         9    0.8200
Roy's Greatest Root         0.04508798       0.20         2         9    0.8200
```

```
MANOVA Test Criteria and F Approximations for the Hypothesis of no phase*group*sex
                 H = Type II SSCP Matrix for phase*group*sex
                           E = Error SSCP Matrix

                          S=2    M=-0.5    N=3.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.69426103       0.90         4        18    0.4841
Pillai's Trace              0.31059765       0.92         4        20    0.4721
Hotelling-Lawley Trace      0.43338209       0.95         4    9.8537    0.4747
Roy's Greatest Root         0.41658268       2.08         2        10    0.1753
```


Repeated measures as GLH: H_0: LBM=0

- **L**: specifies between-S effects: selection of coefficients tested
- **M**: specifies within-S effects: linear combinations of responses
- → **LBM=0** tests between-S diff[ces] in the transformed responses

	Between-S effects tested using **M** for factor C			
Within-S effects	Intercept	**L** = L_A	**L** = L_B	**L** = L_{A*B}
M = (1 1 1)	--	A	B	A*B
$\mathbf{M} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$	C	A*C	B*C	A*B*C

Repeated measures: GLH approach

```
proc glm data=repmes;
   class group;
   model pre post fol = group|sex / nouni;
   contrast 'Trt vs Control' group  -2  1  1 ;
   contrast 'Treat A vs B'   group   0  1 -1 ;

   *-- Group x Time effect;
   manova h=group      M = ( -1  1  0,
                              0 -1  1 ) / short;
   *-- Sex x Time effect;
    manova h=sex       M = ( -1  1  0,
                              0 -1  1 ) / short;

   *-- Time effect;
   manova h=intercept  M = ( -1  1  0,
                              0 -1  1 ) / short;

   *-- Group, Sex and Group*Sex effects;
   manova h=group|sex  M = (  1  1  1 ) / short;
   run;
```

Between (**L**)   Within (**M**)

Use this method for testing contrasts of repeated measures *not provided* by the REPEATED stmt

Repeated measures in R

- Fitting: **`mod <-lm(Y ~ A*B)`** [between formula]→ an mlm object
- Tests: **`aov <-Manova(mod, idesign=~within)`**
- **`print(aov); summary(aov)`**– univ & multiv tests

```
library(car)       # for Anova() functions
# MANOVA model
mod.OBK <- lm(cbind(pre, post, fup) ~  treatment*gender,  data=OBK)

# for linear and quadratic effects of 'Time'
phase <- ordered(c("pretest", "posttest", "followup"),
    levels=c("pretest", "posttest", "followup"))
idata <- data.frame(phase)

# Multivariate tests for repeated measures
aov.OBK <- Manova(mod.OBK, idata=idata, idesign=~phase, type="III")
aov.OBK
```

```
> idata
     phase
1  pretest
2 posttest
3 followup
```

Repeated measures in R

Multivariate tests: `print(aov, test="Pillai")` – compact display for one statistic

```
> aov.OBK

Type III Repeated Measures MANOVA Tests: Pillai test statistic
                      Df test stat approx F num Df den Df    Pr(>F)
(Intercept)            1     0.967  296.389      1     10 9.241e-09 ***
treatment              2     0.441    3.940      2     10 0.0547069 .
gender                 1     0.268    3.659      1     10 0.0848003 .
treatment:gender       2     0.364    2.855      2     10 0.1044692
phase                  1     0.814   19.645      2      9 0.0005208 ***
treatment:phase        2     0.696    2.670      4     20 0.0621085 .
gender:phase           1     0.066    0.319      2      9 0.7349696
treatment:gender:phase 2     0.311    0.919      4     20 0.4721498
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `summary()` method for Anova.mlm objects gives more detail

Univariate tests

```
> summary(aov.OBK, multivariate=FALSE)

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity

                            SS num Df Error SS den Df        F    Pr(>F)
(Intercept)            1351.86      1    45.61     10 296.3888 9.241e-09 ***
treatment                35.95      2    45.61     10   3.9405  0.054707 .
gender                   16.69      1    45.61     10   3.6591  0.084800 .
treatment:gender         26.05      2    45.61     10   2.8555  0.104469
phase                    25.90      2    16.06     20  16.1329 6.732e-05 ***
treatment:phase          15.58      4    16.06     20   4.8510  0.006723 **
gender:phase              0.45      2    16.06     20   0.2828  0.756647
treatment:gender:phase    2.04      4    16.06     20   0.6366  0.642369
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Mauchly Tests for Sphericity

                       Test statistic p-value
phase                         0.74927 0.27282
treatment:phase               0.74927 0.27282
gender:phase                  0.74927 0.27282
treatment:gender:phase        0.74927 0.27282

Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity

                        GG eps Pr(>F[GG])
phase                  0.79953  0.0002814 ***
treatment:phase        0.79953  0.0126909 *
gender:phase           0.79953  0.7089599
treatment:gender:phase 0.79953  0.6116209
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                        HF eps Pr(>F[HF])
phase                  0.92786  0.0001125 ***
treatment:phase        0.92786  0.0084388 **
gender:phase           0.92786  0.7408568
treatment:gender:phase 0.92786  0.6319975
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

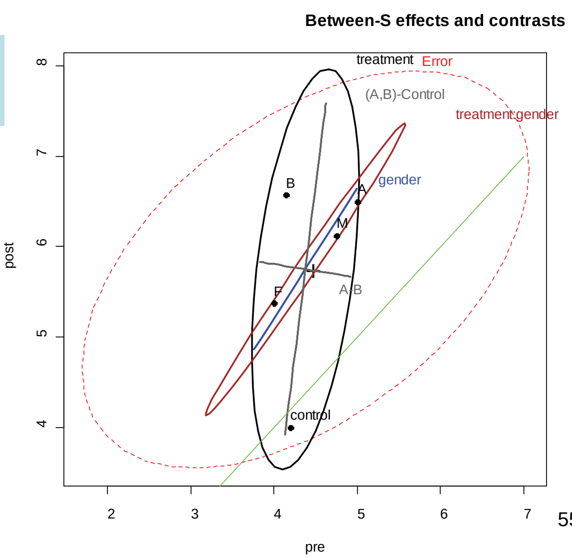
between

within

HE plots for between effects

```
# HE plots: Between-S effects
library(heplots)
heplot(mod.OBK)
```

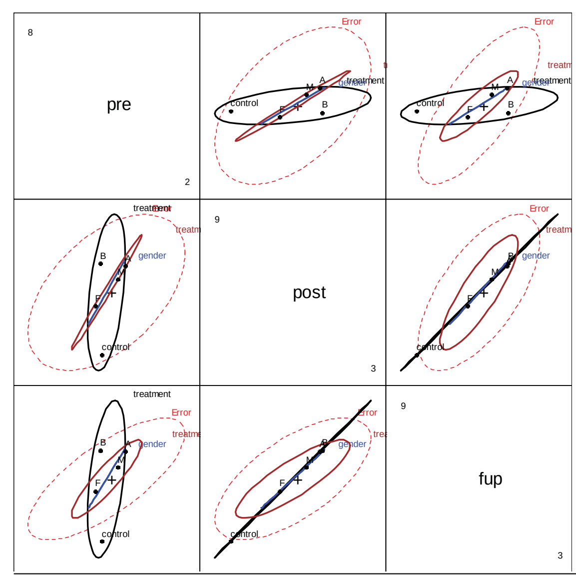
Treatments differ, largely in (A,B)-Control contrast on post-test



HE plots for between effects

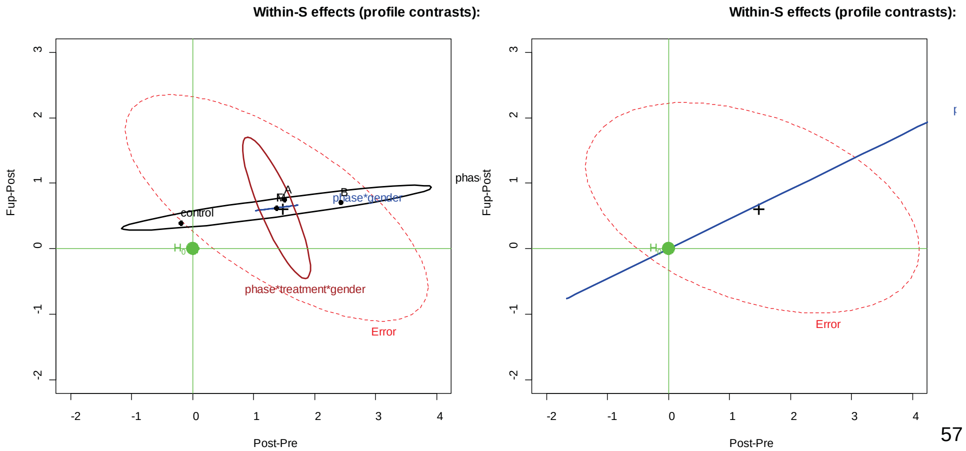
```
pairs(mod.OBK)
```

Treatment effects are nearly the same at post-test and follow-up



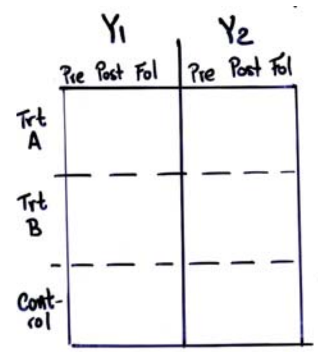
HE plots for within effects

```
OBK$phase.1 <- OBK$post - OBK$pre        # profile contrasts
OBK$phase.2 <- OBK$fup - OBK$post
mod1.OBK <- lm(cbind(phase.1, phase.2) ~ treatment*gender,  data=OBK)
heplot(mod1.OBK)
```



Doubly-multivariate designs

- Repeated measures
- Two (or more) separate criteria



- For ordinary repeated measure designs w/ 2 or more repeated factors, REPEATED stmt generates proper M matrix from one-way contrasts

REPEATED Y (2) TIME (3) ;

$$M = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_{M_Y} \otimes \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}}_{M_{time}}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 0 & -1 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 0 & 1 & 0 \\ 1 & -1 & \underbrace{0 \;\; -1}_{} & & \underbrace{0 \;\; 1}_{} & \end{bmatrix}$$

int Y time Y*time

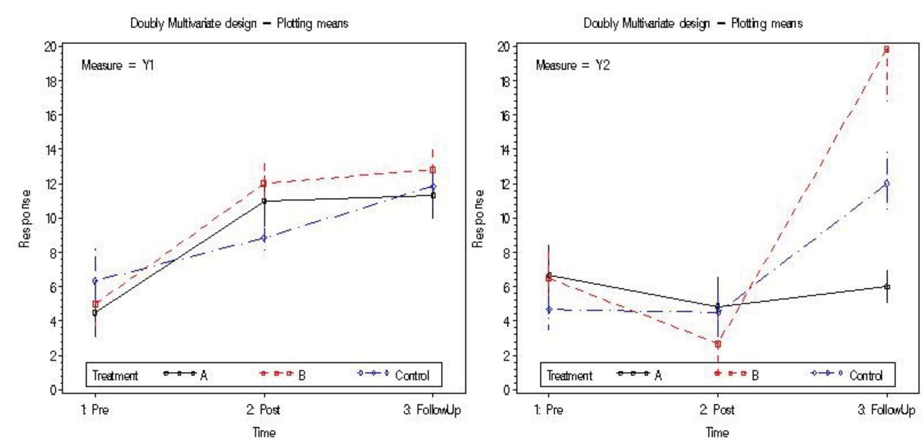
But this doesn't quite do the right tests (why?)

Doubly-multivariate designs: Example

trt	reps	Pre Y1	Post Y1	Fol Y1	Pre Y2	Post Y2	Fol Y2
TreatA	1	3	13	9	0	0	9
TreatA	2	0	14	10	6	6	3
TreatA	3	4	6	17	8	2	6
TreatA	4	7	7	13	7	6	4
TreatA	5	3	12	11	6	12	6
TreatA	6	10	14	8	13	3	8
TreatB	1	9	11	17	8	11	27
TreatB	2	4	16	13	9	3	26
TreatB	3	8	10	9	12	0	18
TreatB	4	5	9	13	3	0	14
TreatB	5	0	15	11	3	0	25
TreatB	6	4	11	14	4	2	9
Control	1	10	12	15	4	3	7
Control	2	2	8	12	8	7	20
Control	3	4	9	10	2	0	10
Control	4	10	8	8	5	8	14
Control	5	11	11	11	1	0	11
Control	6	1	5	15	8	9	10


```
data reshape;
  set exam9;
  if trt ^= 'Control' then trt=substr(trt,6);
  measure = 'Y1';
    time = '1: Pre     '; response = PreY1 ; output;
    time = '2: Post    '; response = PostY1; output;
    time = '3: FollowUp'; response = FolY1 ; output;
  measure = 'Y2';
    time = '1: Pre     '; response = PreY2 ; output;
    time = '2: Post    '; response = PostY2; output;
    time = '3: FollowUp'; response = FolY2 ; output;
```

```
%meanplot(data=reshape,
    class=Time Trt Measure,
    response=response);
```



Doubly-multivariate design: Example

```
proc glm data=exam9;
  class trt;
  model PreY1 PostY1 FolY1 PreY2 PostY2 FolY2 = trt / NOUNI;
  contrast 'Treat A vs B'    trt   0  1 -1;
  contrast 'Trt vs Control' trt  -2  1  1;
  repeated  measure 2 identity, time 3 profile / printM summary;
  run;
```

	Y1	Y2
Y1	1	0
Y2	0	1

⊗

	Mean	Pre-Post	Post-Fol
Pre	1	1	0
Post	1	-1	1
Fol	1	0	-1

=

	M1	M2	PP1	PF1	PP2	PF2
PreY1	1	0	1	0	0	0
PostY1	1	0	-1	1	0	0
FolY1	1	0	0	-1	0	0
PreY2	0	1	0	0	1	0
PostY2	0	1	0	0	-1	1
FolY2	0	1	0	0	0	-1

“identity” contrast does the right thing

Equal means on measures – not applicable here (why?)

MANOVA Test Criteria for the Hypothesis of no **measure** Effect
H = Type III SSCP Matrix for measure
E = Error SSCP Matrix
S=1 M=0 N=6

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02165587	316.24	2	14	<.0001
Pillai's Trace	0.97834413	316.24	2	14	<.0001
Hotelling-Lawley Trace	45.17686368	316.24	2	14	<.0001
Roy's Greatest Root	45.17686368	316.24	2	14	<.0001

MANOVA test of treatment

MANOVA Test Criteria for the Hypothesis of no **measure*trt** Effect
H = Type III SSCP Matrix for measure*trt
E = Error SSCP Matrix
S=2 M=-0.5 N=6

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.72215797	1.24	4	28	0.3178
Pillai's Trace	0.27937444	1.22	4	30	0.3240
Hotelling-Lawley Trace	0.38261660	1.31	4	15.818	0.3074
Roy's Greatest Root	0.37698780	2.83	2	15	0.0908

MANOVA test of time

MANOVA Test Criteria for the Hypothesis of no **measure*time** Effect
H = Type III SSCP Matrix for measure*time
S=1 M=1 N=5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.14071380	18.32	4	12	<.0001
Pillai's Trace	0.85928620	18.32	4	12	<.0001
Hotelling-Lawley Trace	6.10662362	18.32	4	12	<.0001
Roy's Greatest Root	6.10662362	18.32	4	12	<.0001

MANOVA test of time x treatment interaction

MANOVA Test Criteria for the Hypothesis of no **measure*time*trt** Effect
H = Type III SSCP Matrix for measure*time*trt
S=2 M=0.5 N=5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.22861451	3.27	8	24	0.0115
Pillai's Trace	0.96538785	3.03	8	26	0.0151
Hotelling-Lawley Trace	2.52557514	3.64	8	15	0.0149
Roy's Greatest Root	2.12651905	6.91	4	13	0.0033

Univariate Between-S tests (ignore: not a sensible hypothesis)

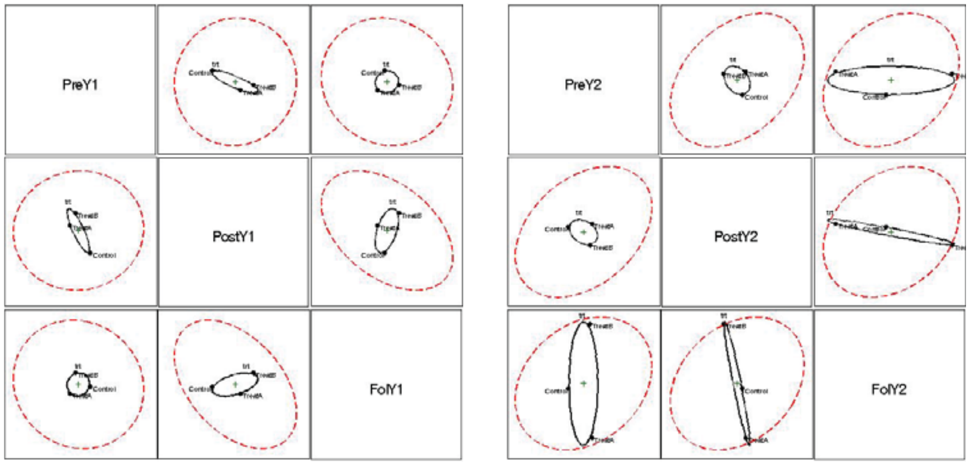
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	2	112.9074074	56.4537037	2.83	0.0908
Error	15	299.5000000	19.9666667		

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Treat A vs B	1	105.1250000	105.1250000	5.27	0.0366
Trt vs Control	1	7.7824074	7.7824074	0.39	0.5418

Visualizing: HE plots

Y1 scores Y2 scores



Summary

- Repeated measure designs:
  - more sensitive tests for within-S factors
  - allow study of growth and change
- Univariate approach
  - strong assumptions, but GG and HF can correct for violation
- MANOVA
  - NO assumption about structure of **Σ**
  - Tests based on Wilks' Λ, HLT, Roy, …
- Mixed model
  - Allows missing data, variable time points
  - Can model individual's coefficients in a Level 2 model
- Visualization: meanplots, HE plots