Repeated Measures Anova: the Wide, the Long and the Long

2007 West Coast Stata Users Group

Phil Ender

Statistical Consulting Group
UCLA Academic Technology Services

October 2007

ATS gets numerous questions on

- Analysis of variance, in general

ATS gets numerous questions on

- Analysis of variance, in general
- Repeated measures analysis of variance, in particular

ATS gets numerous questions on

- Analysis of variance, in general
- Repeated measures analysis of variance, in particular
- With repeated measures on some or on all factors

ATS gets numerous questions on

- Analysis of variance, in general
- Repeated measures analysis of variance, in particular
- With repeated measures on some or on all factors
- Most common stat packages used for anova at UCLA:
  SPSS
  SAS

Three approaches to repeated measures anova

- The Wide – Multivariate Models

Three approaches to repeated measures anova

- The Wide – Multivariate Models
- the Long – Anova Models

Three approaches to repeated measures anova

- The Wide – Multivariate Models
- the Long – Anova Models
- and the Long – Mixed Models

Example Design

- To illustrate the different approaches to repeated measures anova, I will use the following example design:

Example Design

- To illustrate the different approaches to repeated measures anova, I will use the following example design:
- Two factors:
- One between-subjects factor (A) with 2 levels
- One repeated (within-subjects) factor (B) with 4 levels

Design Schematic

Split-Plot Factorial 2·4

		B			
		b_1	b_2	b_3	b_4
A	a_1	s_1	s_1	s_1	s_1
	a_2	s_2	s_2	s_2	s_2

¡Warning!

There will be some SAS and SPSS code in this presentation.

Sorry.

The Wide – Multivariate Models

SAS proc glm

```
proc glm data=wide;
  class a;
  model y1 y2 y3 y4 = a;
  repeated b 4;
run;
quit;
```

Output

- SAS includes two kinds of output:
- Multivariate tests of b and a*b

Output

- SAS includes two kinds of output:
- Multivariate tests of b and a*b
- Univariate tests of a, b, and a*b

Assumptions

Multivariate assumptions:

Observations are multivariate normal
Covariance structure – unstructured

Assumptions

Multivariate assumptions:

Observations are multivariate normal
Covariance structure – unstructured

Univariate assumptions:

Nonadditivity assumption; no subject by treatment interaction
Covariance structure – compound symmetric
Plus all the standard ones concerning normality and homogeneity of variance, etc

Now For Some SPSS Code

Now For Some SPSS Code

point ... click

point ... point ... click ... click

point ... click ... click

Now For Some SPSS Code

point ... click

point ... point ... click ... click

point ... click ... click

Actually, the SPSS syntax code looks a lot like the SAS code

Multivariate Results in Stata

```
manova y1 y2 y3 y4 = a
```

Multivariate Results in Stata

```
manova y1 y2 y3 y4 = a
```

/ multivariate test of a*b interaction */*

```
matrix ymat = (1,0,0,-1\0,1,0,-1\0,0,1,-1)
manovatest a, ytransform(ymat)
```

Multivariate Results in Stata

```
manova y1 y2 y3 y4 = a
```

/ multivariate test of a*b interaction */*

```
matrix ymat = (1,0,0,-1\0,1,0,-1\0,0,1,-1)
manovatest a, ytransform(ymat)
```

/ multivariate test of b main effect */*

```
matrix xmat = (1, .5, .5)
manovatest, test(xmat) ytransform(ymat)
```

The Downside of Wide

- Missing data can be a killer

The Downside of Wide

- Missing data can be a killer
- Manova analyses use listwise deletion for missing data

The Downside of Wide

- Missing data can be a killer
- Manova analyses use listwise deletion for missing data
- If a single value for a subject is missing the whole subject is deleted

The Downside of Wide

- Missing data can be a killer
- Manova analyses use listwise deletion for missing data
- If a single value for a subject is missing the whole subject is deleted
- There are single imputation methods based on row and column means

The Long – Anova Models

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

- This linear model is fixed-effects only, there are no random-effects

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

- This linear model is fixed-effects only, there are no random-effects
- Can be more easily seen when expressed as a regression model, $\mathbf{y} = \mathbf{X}\beta + \epsilon$

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

- This linear model is fixed-effects only, there are no random-effects
- Can be more easily seen when expressed as a regression model, $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- Random effects are computed using different denominators for the various F-ratios

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

- This linear model is fixed-effects only, there are no random-effects
- Can be more easily seen when expressed as a regression model, $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- Random effects are computed using different denominators for the various F-ratios
- nonadditivity assumption - no block (subject) treatment interaction

The Anova Linear Model

$$Y_{ijk} = \mu + \alpha_j + \pi_{i(j)} + \beta_k + \alpha * \beta_{jk} + \beta * \pi_{ki(j)} + \epsilon_{ijk}$$

- This linear model is fixed-effects only, there are no random-effects
- Can be more easily seen when expressed as a regression model, $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- Random effects are computed using different denominators for the various F-ratios
- nonadditivity assumption - no block (subject) treatment interaction
- For this model ϵ_{ijk} and $\beta * \pi_{ki(j)}$ are not separately estimable

SAS convert wide to long

```
data long;
set wide;
y=y1; b=1; output;
y=y2; b=2; output;
y=y3; b=3; output;
y=y4; b=4; output;
drop y1 y2 y3 y4;
run;
```

SAS proc glm

```
proc glm data=long;
class a b s;
model y = a s(a) b a*b / ss3;
test h=a e=s(a);
run;
quit;
```

Stata convert wide to long

```
reshape long y, i(s) j(b)
```

Stata anova

```
anova y a / s|a b a*b /, repeated(b)
```

Stata anova

```
anova y a / s|a b a*b /, repeated(b)
```

/ inspect pooled-within covariance matrix */*

```
matrix list e(Srep)
```

anova repeated option

- Along with the e(Srep) matrix ...
- Allows evaluation of compound symmetry assumption
- Gives conservative p-values if assumption is not met

Stata manova also works with long data

```
manova y = a / s|a b a*b /
```

Stata manova also works with long data

```
manova y = a / s|a b a*b /
```

- Does not have a repeated option

Stata manova also works with long data

```
manova y = a / s|a b a*b /
```

- Does not have a repeated option
- Displays univariate F-ratios in multivariate format
- Output is a bit cluttered

Stata manova also works with long data

```
manova y = a / s|a b a*b /
```

- Does not have a repeated option
- Displays univariate F-ratios in multivariate format
- Output is a bit cluttered
- Useful for multivariate repeated measures

There is also a user written ado

```
wsanova y b, id(s) between(a) epsilon
```

There is also a user written ado

```
wsanova y b, id(s) between(a) epsilon
```

- -wsanova- (John Gleason) – findit wsanova

Using regression

```
recode s 5=1 6=2 7=3 8=4, generate(ss)
```

Using regression

```
recode s 5=1 6=2 7=3 8=4, generate(ss)

xi3: regress y e.a*e.b r.ss@i.a
```

Using regression

```
recode s 5=1 6=2 7=3 8=4, generate(ss)

xi3: regress y e.a*e.b r.ss@i.a

test _Ib_2 _Ib_3 _Ib_4
test _Ia2Xb2 _Ia2Xb3 _Ia2Xb4
```

Using regression

```
recode s 5=1 6=2 7=3 8=4, generate(ss)

xi3: regress y e.a*e.b r.ss@i.a

test _Ib_2 _Ib_3 _Ib_4
test _Ia2Xb2 _Ia2Xb3 _Ia2Xb4

test2 _Ia_2 / _Iss2Wa1 _Iss2Wa2 _Iss3Wa1 ///
      _Iss3Wa2  _Iss4Wa1 _Iss4Wa2 _Iss4Wa2
```

Regression comments

- Don't use dummy coded variables

Regression comments

- Don't use dummy coded variables
- Use -xi3- or -desmat- to create effect coded variables

Regression comments

- Don't use dummy coded variables
- Use -xi3- or -desmat- to create effect coded variables
- -xi3- can create effect coding on the fly

Regression comments

- Don't use dummy coded variables
- Use -xi3- or -desmat- to create effect coded variables
- -xi3- can create effect coding on the fly
- -desmat- (John Hendrickx) – findit desmat
- -xi3- (Mitchell & Ender) – findit xi3
- -test2- (Ender) – findit test2

Dummy Coding versus Effect Coding

F-ratios

	Dummy	*Effect*
A main effect	*15.78*	*2.00*
B main effect	*35.96*	*127.89*
*A*B interaction*	*12.74*	*12.74*

The Long again – Mixed Models

The Linear Mixed Model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon$$

The Linear Mixed Model

$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon$

where

$\mathbf{y}$ is the nx1 response vector

$\mathbf{X}$ is the nxp fixed-effects design matrix

β is the px1 vector of fixed effects

$\mathbf{Z}$ is the nxq random-effects design matrix

$\mathbf{u}$ is the qx1 vector of random effects

ϵ is the nx1 vector of errors

SAS proc mixed 1

```
proc mixed data=long;
class a b;
model y = a b a*b;
random intercept / subject=s;
run;
```

SAS proc mixed 2

```
proc mixed data=long;
class a b;
model y = a b a*b;
repeated b / subject=s type=cs;
run;
```

SAS Output

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
a	1	6	2.00	0.2070
b	3	18	127.89	<.0001
a*b	3	18	12.74	0.0001

Stata xtmixed

```
xi3 e.a*e.b
```

Stata xtmixed

```
xi3 e.a*e.b

xtmixed y _Ia_2 _Ib_2 _Ib_3 _Ib_4 ///
          _Ia2Xb2 _Ia2Xb3 _Ia2Xb4 || s:
```

Stata xtmixed

```
xi3 e.a*e.b

xtmixed y _Ia_2 _Ib_2 _Ib_3 _Ib_4 ///
          _Ia2Xb2 _Ia2Xb3 _Ia2Xb4 || s:

test _Ia_2
test _Ib_2 _Ib_3 _Ib_4
test _Ia2Xb2 _Ia2Xb3 _Ia2Xb4
```

Stata xtmixed comments

- Once again, don't use dummy coding

Stata xtmixed comments

- Once again, don't use dummy coding
- -xi3- does not work directly with -xtmixed-

Stata xtmixed comments

- Once again, don't use dummy coding
- -xi3- does not work directly with -xtmixed-
- Test command displays results as chi-square
- F approximation equals χ^2/df

SAS Covariance Types

SAS allows for a number of covariance structures:

autoregressive, heterogeneous autoregressive,
compound symmetry, heterogeneous CS,
Toeplitz, unstructured, and over a dozen more

SAS proc mixed revisited

```
proc mixed data=long;
class a b;
model y = a b a*b;
repeated b / subject=s type=ar(1);
run;
```

Stata xtregar

```
tsset s b

xi3: xtregar y e.a*e.b

test _Ia_2
test _Ib_2 _Ib_3 _Ib_4
test _Ia2Xb2 _Ia2Xb3 _Ia2Xb4
```

Comparing proc mixed w/ ar(1) and xtregar

- Results are close but not identical

Comparing proc mixed w/ ar(1) and xtregar

- Results are close but not identical
- proc mixed uses a REML estimator

Comparing proc mixed w/ ar(1) and xtregar

- Results are close but not identical
- proc mixed uses a REML estimator
- xtregar uses a GLS estimator

proc mixed w/ ar(1) vs xtregar

F-ratios

	proc mixed	xtregar
A main effect	1.65	1.94
B main effect	91.57	102.78333
A*B interaction	14.69	13.866667

Closing Comments

For balanced designs the repeated anova and mixed models yield the same results

Closing Comments

For balanced designs the repeated anova and mixed models yield the same results

Results are different for unbalanced designs

Closing Comments

For balanced designs the repeated anova and mixed models yield the same results

Results are different for unbalanced designs

For missing data within subjects use mixed with FIML

Closing Comments

For balanced designs the repeated anova and mixed models yield the same results

Results are different for unbalanced designs

For missing data within subjects use mixed with FIML

Repeated measures anova is not necessarily the best way to study change over time

Web Page

A web page with all of the Stata commands
and complete output can be found at

http://www.ats.ucla.edu/stat/stata/library/2007wcsug.htm

The End