BIOS 6612
Lecture 16

Mixed Models for Repeated Measures
Covariance Structures

Review (15) / Current (Lecture 16)/ Preview (Lecture 17)

General Linear Mixed Effects Models (Mixed Models) are the most flexible method for analyzing repeated measures / correlated / longitudinal data.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \text{ and } V = Var(\mathbf{Y}) = \mathbf{ZGZ}' + \mathbf{R}$$

- Lecture 15: specify the random effect **b** (i.e. random intercept, random slope)
- Lecture 16: specify Var(Y)
- Lecture 17: RMANOVA and random intercept

	Repeated Measures ANOVA	**Covariance Pattern Model**	**Random Coefficients Model**
Time	Categorical	Categorical -effect of time (i.e. linear, quadratic) does not need to be specified	Continuous -must model the effect of time
Covariance Structure	Compound Symmetry (i.e. sphericity)	Attempts to account for all the potential sources of variability that have an impact on the covariance among repeated measures on the same individual	Usually assumes random effects account for most of the variation in the data and the remaining error components, ε_i, have a very simple covariance structure (e.g., $\sigma^2 I_N$)
Distinguish between-subject and within-subject sources of variability	Yes	No	Yes
Highly Unbalanced Data	No	No	Yes

Example: Dental Data (Wolfinger & Chang; Pothoff & Roy)

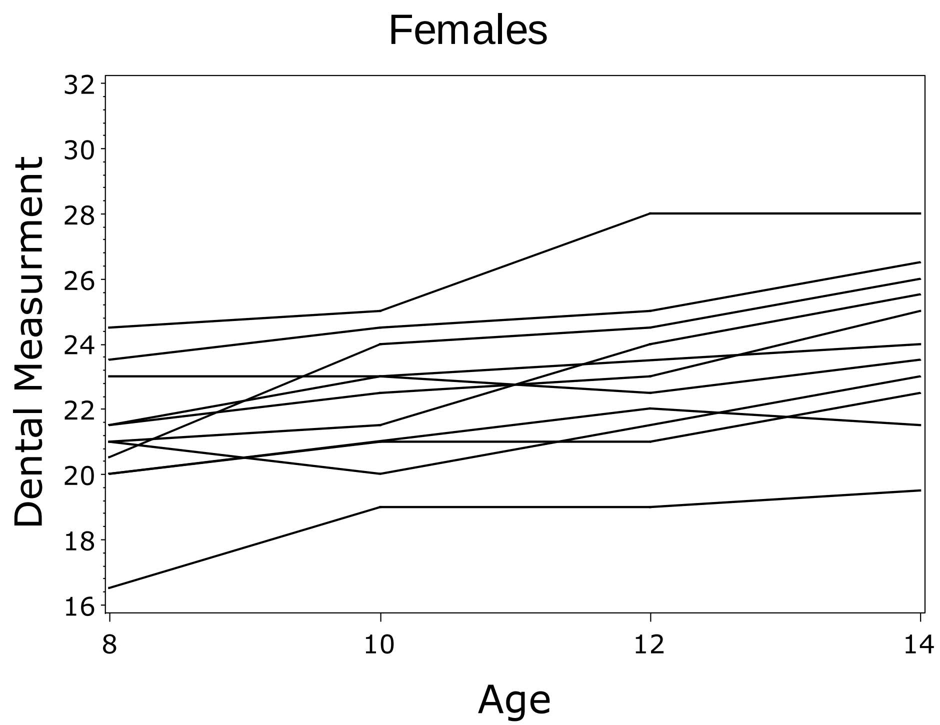
Dental measurements from the center of the pituitary to the pterygomaxillary fissure (measured in mm) for 11 girls and 16 boys at ages 8, 10, 12, and 14. The subjects are individual children, and there are four repeated measurements on each child.

Dental Measurements Data

Person	Gender	Age 8	Age 10	Age 12	Age 14
1	F	21.0	20.0	21.5	23.0
2	F	21.0	21.5	24.0	25.5
3	F	20.5	24.0	24.5	26.0
4	F	23.5	24.5	25.0	26.5
5	F	21.5	23.0	22.5	23.5
6	F	20.0	21.0	21.0	22.5
7	F	21.5	22.5	23.0	25.0
8	F	23.0	23.0	23.5	24.0
9	F	20.0	21.0	22.0	21.5
10	F	16.5	19.0	19.0	19.5
11	F	24.5	25.0	28.0	28.0
12	M	26.0	25.0	29.0	31.0
13	M	21.5	22.5	23.0	26.5
14	M	23.0	22.5	24.0	27.5
15	M	25.5	27.5	26.5	27.0
16	M	20.0	23.5	22.5	26.0
17	M	24.5	25.5	27.0	28.5
18	M	22.0	22.0	24.5	26.5
19	M	24.0	21.5	24.5	25.5
20	M	23.0	20.5	31.0	26.0
21	M	27.5	28.0	31.0	31.5
22	M	23.0	23.0	23.5	25.0
23	M	21.5	23.5	24.0	28.0
24	M	17.0	24.5	26.0	29.5
25	M	22.5	25.5	25.5	26.0
26	M	23.0	24.5	26.0	30.0
27	M	22.0	21.5	23.5	25.0

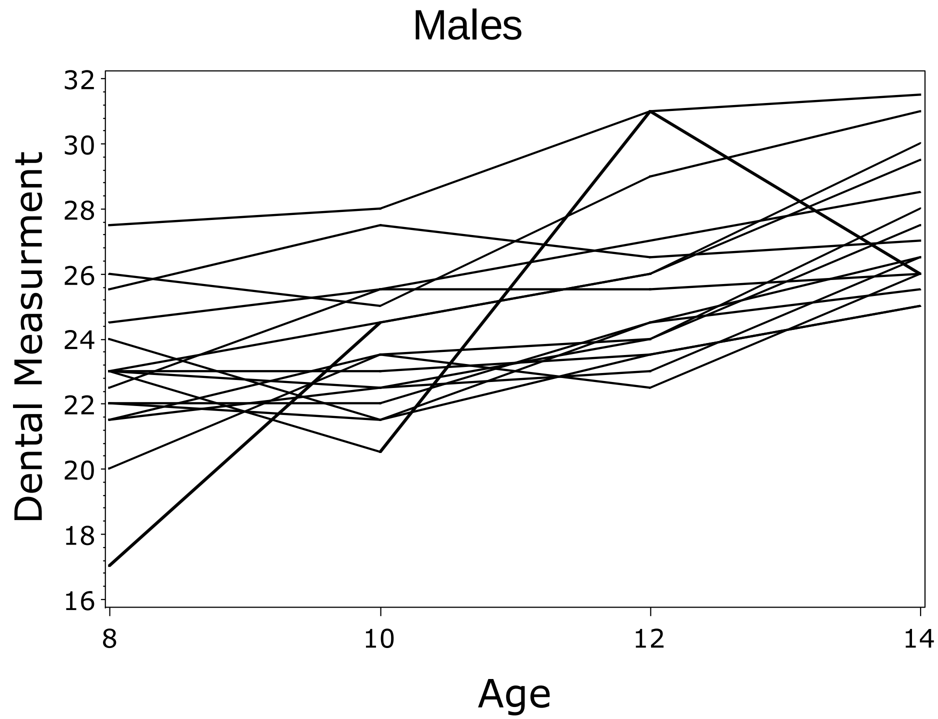
```
DATA forglm(keep=person gender y1-y4)
     formixed(keep=person gender age y);
INPUT person gender$ y1-y4;
OUTPUT forglm;
y=y1; age=8;  OUTPUT formixed;
y=y2; age=10; OUTPUT formixed;
y=y3; age=12; OUTPUT formixed;
y=y4; age=14; OUTPUT formixed;
DATALINES;
 1 F 21.0 20.0 21.5 23.0
 2 F 21.0 21.5 24.0 25.5
 3 F 20.5 24.0 24.5 26.0
 4 F 23.5 24.5 25.0 26.5
 5 F 21.5 23.0 22.5 23.5
 6 F 20.0 21.0 21.0 22.5
 7 F 21.5 22.5 23.0 25.0
 8 F 23.0 23.0 23.5 24.0
 9 F 20.0 21.0 22.0 21.5
10 F 16.5 19.0 19.0 19.5
11 F 24.5 25.0 28.0 28.0
12 M 26.0 25.0 29.0 31.0
13 M 21.5 22.5 23.0 26.5
14 M 23.0 22.5 24.0 27.5
15 M 25.5 27.5 26.5 27.0
16 M 20.0 23.5 22.5 26.0
17 M 24.5 25.5 27.0 28.5
18 M 22.0 22.0 24.5 26.5
19 M 24.0 21.5 24.5 25.5
20 M 23.0 20.5 31.0 26.0
21 M 27.5 28.0 31.0 31.5
22 M 23.0 23.0 23.5 25.0
23 M 21.5 23.5 24.0 28.0
24 M 17.0 24.5 26.0 29.5
25 M 22.5 25.5 25.5 26.0
26 M 23.0 24.5 26.0 30.0
27 M 22.0 21.5 23.5 25.0
;


*** For GPLOT symbols ***;
 %macro symbols;
 %do i = 1 %to 27;
 SYMBOL&i INTERPOL=JOIN COLOR=black LINE=1
          WIDTH=2;
 %end;
 %mend;
 %symbols;
```



Means: 21.1818 22.2273 23.0909 24.0909

```
PROC GPLOT DATA=formixed;
 PLOT (y)*age =person
/VAXIS=AXIS1 HAXIS=AXIS2 NOLEGND;
 AXIS1 LABEL = (FONT='Verdana' HEIGHT= 3 ANGLE=90 POSITION=top "Dental Measurment" )
                VALUE=(FONT='Verdana' HEIGHT=2) ORDER=(16 to 32 by 2) ;
 AXIS2  LABEL = (FONT='Verdana' HEIGHT=3 "Age") MINOR=NONE ORDER=(8 to 14 by 2)
                VALUE=(FONT='Verdana' HEIGHT=2);
 WHERE male = 0;
 LEGEND1 ;
 TITLE1 HEIGHT=3 FONT=ARIAL "Females";
 RUN;
```



Means: 22.8750 23.8125 25.7188 27.4688

```
PROC GPLOT DATA=formixed;
 PLOT (y)*age =person
/VAXIS=AXIS1 HAXIS=AXIS2 NOLEGND;
 AXIS1 LABEL = (FONT='Verdana' HEIGHT= 3 ANGLE=90 POSITION=top "Dental Measurment" )
              VALUE=(FONT='Verdana' HEIGHT=2) ORDER=(16 to 32 by 2) ;
 AXIS2 LABEL = (FONT='Verdana' HEIGHT=3 "Age") MINOR=NONE ORDER=(8 to 14 by 2)
               VALUE=(FONT='Verdana' HEIGHT=2);
 WHERE male = 1;
 LEGEND1 ;
 TITLE1 HEIGHT=3 FONT=ARIAL "Males";
```

Can compare models with different covariance structures (but same fixed effects) using REML or ML.

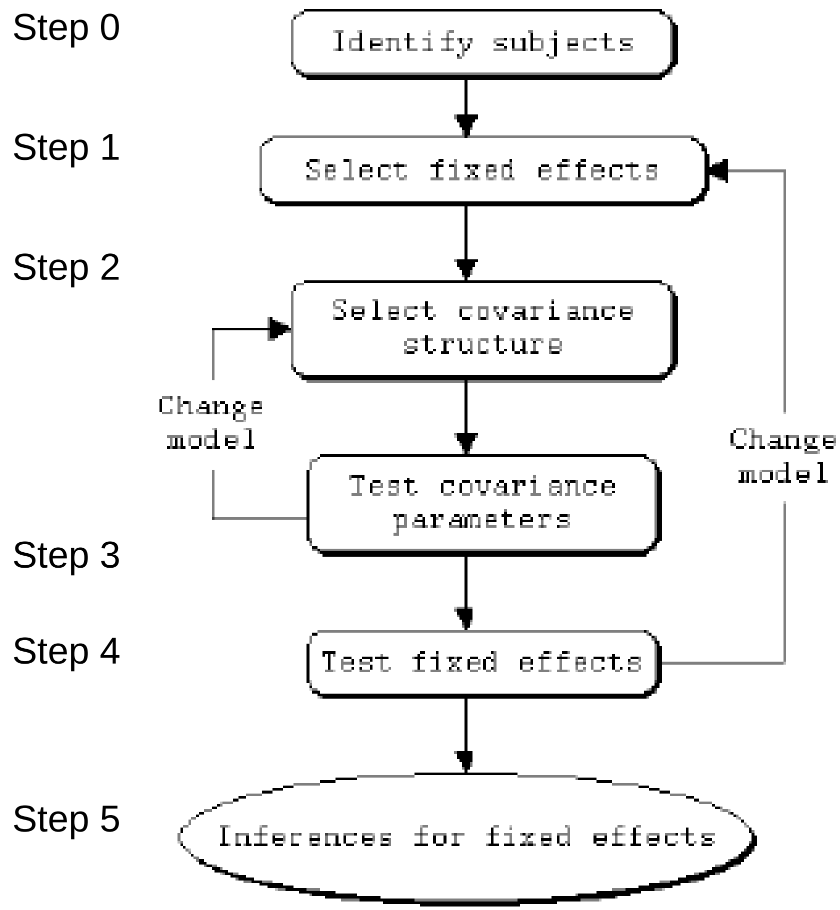


Figure 2. Repeated Measures Analysis in PROC MIXED

Must use ML to compare models with different fixed effects.

(REML depends upon the fixed-effects design matrix)

Step 1: Select Fixed Effects

- Age: continuous or categorical? Try categorical first (most flexible option)
- Gender
- Age × Gender

Step 2: Select Covariance Structure

Use the `TYPE =` option in SAS or `correlation=` option in R package ‘nlme’ function `gls()`

- Compound Symmetry
- Huynh-Feldt (type H)
- Unstructured

SAS uses the information from `TYPE=` and `SUB=` to construct the appropriate variance-covariance matrix.

- In this case, a 108×108 block diagonal with 27 4×4 blocks.
- Each of these blocks has the covariance structure given by the `TYPE=` option.

R syntax: `correlation=corCompSymm(form=~1|Person)` (for example)

- Correlation structure here is compound symmetry
- No covariates affect the variance components
- `Person` is the variable denoting independent subjects in the data set

Covariance Structures

- For a full list of covariance structures in SAS (Type=covariance-structure) http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect019.htm
  - Including autoregressive (AR) / autoregressive moving average (ARMA) [time series] and spatial covariance structures

Independence (Type=SIMPLE):

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

1 covariance parameter
(residual variance)

Compound Symmetry (Type=CS, `correlation=corCompSymm()`):

$$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}$$

2 covariance
parameters

Toeplitz (Type=TOEP, `correlation=corARMA(form=~1|id,p=3,q=0)`):

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

4 covariance parameters
= number of repeated measurements

First-Order Autoregressive (Type=AR(1), `correlation=corARMA(form=~1|id,p=1,q=0)`): Not used in this lecture example, just here for your reference

- Note that Toeplitz is a special case of autoregressive with order equal to number of time points minus 1: AR(3) for this example

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

2 covariance parameters

First-Order Autoregressive Moving-Average (Type=ARMA(1,1), `correlation= corARMA(form=~1|id,p=1,q=1)`): Not used in this lecture example

$$\sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$$

3 covariance parameters

Huynh-Feldt (Type= HF):

$$\begin{bmatrix} \sigma_1 & \frac{\sigma_1+\sigma_2}{2}-\lambda & \frac{\sigma_1+\sigma_3}{2}-\lambda & \frac{\sigma_1+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_2}{2}-\lambda & \sigma_2 & \frac{\sigma_2+\sigma_3}{2}-\lambda & \frac{\sigma_2+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_3}{2}-\lambda & \frac{\sigma_2+\sigma_3}{2}-\lambda & \sigma_3 & \frac{\sigma_3+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_4}{2}-\lambda & \frac{\sigma_2+\sigma_4}{2}-\lambda & \frac{\sigma_3+\sigma_4}{2}-\lambda & \sigma_4 \end{bmatrix}$$

5 covariance parameters
= number of repeated measurements plus 1 (λ)
i.e. 4+1=5

Unstructured (Type=UN):

- Syntax in R ‘nlme’ a bit more complicated: need to use two options
- `correlation=corSymm(form=~1|id)`
- `weights=varIdent(form=~1|time)`

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}$$

10 covariance parameters
For k repeated measurements,
then the number of covariance parameters= k+ (k-1)+ …+1

i.e. for k=4 => 4+3+2+1=10

SAS/R CODE

```
*** Mixed Model, CS Covariance ***;
PROC MIXED DATA=formixed METHOD=reml;
 CLASS gender age person;
 MODEL y = gender age gender*age;
 REPEATED age /TYPE=CS SUBJECT=person R RCORR;
RUN;
```

```
mod2 <- gls(distance ~ Gender*as.factor(age),
            correlation=corCompSymm(form=~1|Person),
            data=dental.long)
```

```
*** Mixed Model, HF Covariance ***;
PROC MIXED DATA=formixed METHOD=reml;
 CLASS gender age person;
 MODEL y = gender age gender*age;
 REPEATED age /TYPE=HF SUBJECT=person R RCORR;
RUN;
```

```
# Not directly available in R (would have to construct this manually)
```

```
*** Mixed Model, Unstructured Covariance ***;
PROC MIXED DATA=formixed METHOD=reml;
 CLASS gender age person;
 MODEL y = gender age gender*age;
 REPEATED age /TYPE=UN SUBJECT=person R RCORR;
RUN;
```

```r
mod3 <- gls(distance ~ Gender*as.factor(age),
            correlation=corSymm(form=~1|Person),
            weights=varIdent(form=~1|as.numeric(age)),
            data=dental.long)
```

```sas
*** Mixed Model, Toeplitz Covariance ***;
PROC MIXED DATA=formixed METHOD=reml;
 CLASS gender age person;
 MODEL y = gender age gender*age;
 REPEATED age /TYPE=TOEP SUBJECT=person R RCORR;
RUN;
```

```r
mod4 <- gls(distance ~ Gender*as.factor(age),
            correlation=corARMA(form=~1|Person,p=3,q=0),
            data=dental.long)
```

TYPE=CS

Model Information

Data Set	WORK.FORMIXED
Dependent Variable	y
Covariance Structure	Compound Symmetry
Subject Effect	person
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information

Class	Levels	Values
gender	2	F M
age	4	8 10 12 14
person	27	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Dimensions

Covariance Parameters	**2**
Columns in X	15
Columns in Z	0
Subjects	27
Max Obs Per Subject	4

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	470.49084642	
1	1	423.40853283	0.00000000

Convergence criteria met.

3.2854+1.9750=5.2604

$$\begin{bmatrix} 5.2604 & 3.2854 & 3.2854 & 3.2854 \\ 3.2854 & 5.2604 & 3.2854 & 3.2854 \\ 3.2854 & 3.2854 & 5.2604 & 3.2854 \\ 3.2854 & 3.2854 & 3.2854 & 5.2604 \end{bmatrix}$$

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	
CS	person	3.2854	σ_1
Residual		1.9750	σ^2

Estimated R Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	5.2604	3.2854	3.2854	3.2854
2	3.2854	5.2604	3.2854	3.2854
3	3.2854	3.2854	5.2604	3.2854
4	3.2854	3.2854	3.2854	5.2604

Estimated R Correlation Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6245	0.6245	0.6245
2	0.6245	1.0000	0.6245	0.6245
3	0.6245	0.6245	1.0000	0.6245
4	0.6245	0.6245	0.6245	1.0000

Fit Statistics

-2 Res Log Likelihood	423.4
AIC (smaller is better)	427.4
AICC (smaller is better)	427.5
BIC (smaller is better)	430.0

AIC, AICC, BIC are used to compare models. They penalize for the number of covariance parameters.

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	47.08	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.29	0.0054
age	3	75	35.35	<.0001
gender*age	3	75	2.36	0.0781

Don't make inference yet. Select covariance structure first!

```
summary(mod2)

## Generalized least squares fit by REML
##   Model: distance ~ Gender * as.factor(age)
##   Data: dental.long
##        AIC      BIC    logLik
##   443.4085 469.4602 -211.7043
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | Person
##  Parameter estimate(s):
##       Rho
## 0.6245472
##
## Coefficients:
##                                Value Std.Error   t-value p-value
## (Intercept)                21.181818 0.6915345 30.630168  0.0000
## GenderM                     1.693182 0.8983297  1.884811  0.0624
## as.factor(age)10            1.045455 0.5992479  1.744611  0.0841
## as.factor(age)12            1.909091 0.5992479  3.185812  0.0019
## as.factor(age)14            2.909091 0.5992479  4.854570  0.0000
## GenderM:as.factor(age)10 -0.107955 0.7784458 -0.138680  0.8900
## GenderM:as.factor(age)12  0.934659 0.7784458  1.200673  0.2327
## GenderM:as.factor(age)14  1.684659 0.7784458  2.164131  0.0328
##
##  Correlation:
##                          (Intr) GendrM a.()10 a.()12 a.()14 GM:.()10
## GenderM                  -0.770
## as.factor(age)10         -0.433  0.334
## as.factor(age)12         -0.433  0.334  0.500
## as.factor(age)14         -0.433  0.334  0.500  0.500
## GenderM:as.factor(age)10  0.334 -0.433 -0.770 -0.385 -0.385
## GenderM:as.factor(age)12  0.334 -0.433 -0.385 -0.770 -0.385  0.500
## GenderM:as.factor(age)14  0.334 -0.433 -0.385 -0.385 -0.770  0.500
##                          GM:.()12
## GenderM
## as.factor(age)10
## as.factor(age)12
## as.factor(age)14
```

```
## GenderM:as.factor(age)10
## GenderM:as.factor(age)12
## GenderM:as.factor(age)14  0.500
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.56151951 -0.60972339 -0.07927333  0.61436832  2.30264254
##
## Residual standard error: 2.293561
## Degrees of freedom: 108 total; 100 residual
```

```
getVarCov(mod2)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 5.2604 3.2854 3.2854 3.2854
## [2,] 3.2854 5.2604 3.2854 3.2854
## [3,] 3.2854 3.2854 5.2604 3.2854
## [4,] 3.2854 3.2854 3.2854 5.2604
##   Standard Deviations: 2.2936 2.2936 2.2936 2.2936
```

```
corMatrix(mod2$modelStruct$corStruct)[[1]] # just use Person=1 here since model is the same for all
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6245472 0.6245472 0.6245472
## [2,] 0.6245472 1.0000000 0.6245472 0.6245472
## [3,] 0.6245472 0.6245472 1.0000000 0.6245472
## [4,] 0.6245472 0.6245472 0.6245472 1.0000000
```

```
anova(mod2)
```

```
## Denom. DF: 100
##                         numDF  F-value p-value
## (Intercept)                 1 4123.164  <.0001
## Gender                      1    9.292  0.0029
## as.factor(age)              3   40.032  <.0001
## Gender:as.factor(age)       3    2.362  0.0759
```

Summary of Resulting Covariance Structures

- These covariance and correlation matrices can be obtained in SAS using the `R` and `RCORR` commands.
  `REPEATED age /TYPE=CS SUBJECT=person R RCORR;`

Covariance Structures / Correlation Matrix

Compound Symmetry:

$$\begin{bmatrix} 5.2604 & 3.2854 & 3.2854 & 3.2854 \\ 3.2854 & 5.2604 & 3.2854 & 3.2854 \\ 3.2854 & 3.2854 & 5.2604 & 3.2854 \\ 3.2854 & 3.2854 & 3.2854 & 5.2604 \end{bmatrix} \quad \begin{bmatrix} 1.0000 & 0.6245 & 0.6245 & 0.6245 \\ 0.6245 & 1.0000 & 0.6245 & 0.6245 \\ 0.6245 & 0.6245 & 1.0000 & 0.6245 \\ 0.6245 & 0.6245 & 0.6245 & 1.0000 \end{bmatrix}$$

$\boxed{\leftarrow \text{Most restrictive}}$

Huynh-Feldt (Type H):

$$\begin{bmatrix} 5.0264 & 2.7357 & 3.6251 & 3.1806 \\ 2.7357 & 4.3951 & 3.3095 & 2.8649 \\ 3.6251 & 3.3095 & 6.1739 & 3.7543 \\ 3.1806 & 2.8649 & 3.7543 & 5.2848 \end{bmatrix} \quad \begin{bmatrix} 1.0000 & 0.5820 & 0.6508 & 0.6171 \\ 0.5820 & 1.0000 & 0.6353 & 0.5944 \\ 0.6508 & 0.6353 & 1.0000 & 0.6573 \\ 0.6171 & 0.5944 & 0.6573 & 1.0000 \end{bmatrix}$$

Unstructured:

$$\begin{bmatrix} 5.4155 & 2.7168 & 3.9102 & 2.7102 \\ 2.7168 & 4.1848 & 2.9272 & 3.3172 \\ 3.9102 & 2.9272 & 6.4557 & 4.1307 \\ 2.7102 & 3.3172 & 4.1307 & 4.9857 \end{bmatrix} \quad \begin{bmatrix} 1.0000 & 0.5707 & 0.6613 & 0.5216 \\ 0.5707 & 1.0000 & 0.5632 & 0.7262 \\ 0.6613 & 0.5632 & 1.0000 & 0.7281 \\ 0.5216 & 0.7262 & 0.7281 & 1.0000 \end{bmatrix}$$

$\boxed{\leftarrow \text{No restrictions}}$

Covariance Structures **Correlation Matrix**

Toeplitz:

$$\begin{bmatrix} 5.3195 & 3.3325 & 3.7210 & 2.4870 \\ 3.3325 & 5.3195 & 3.3325 & 3.7210 \\ 3.7210 & 3.3325 & 5.3195 & 3.3325 \\ 2.4870 & 3.7210 & 3.3325 & 5.3195 \end{bmatrix} \quad \begin{bmatrix} 1.0000 & 0.6265 & 0.6995 & 0.4675 \\ 0.6265 & 1.0000 & 0.6265 & 0.6995 \\ 0.6995 & 0.6265 & 1.0000 & 0.6265 \\ 0.4675 & 0.6995 & 0.6265 & 1.0000 \end{bmatrix}$$

Simple:

$$\begin{bmatrix} 5.2604 & 0 & 0 & 0 \\ 0 & 5.2604 & 0 & 0 \\ 0 & 0 & 5.2604 & 0 \\ 0 & 0 & 0 & 5.2604 \end{bmatrix} \quad \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

- Obtaining covariance and correlation matrices can be a little more involved in R ‘nlme’
  - Can get correlation matrix with `corMatrix(model$modelStruct$corStruct)` (but need to specify a subject index with `[[1]]`)
  - In theory, `getVarCov()` will return the covariance matrix, but this sometimes doesn’t line up with results from SAS, so need to be careful with this (possibly a bug: https://www.jepusto.com/bug-in-nlme-getvarcov/)

Step 3: Test Covariance Structure

How do we select the 'best' covariance structure? (KKMN pp. 767-769).

- Use what makes sense clinically/biologically
- Try several correlation structures
- Use an unstructured correlation matrix
- Start with an unstructured correlation matrix and simplify
- Start with a Toeplitz (similar to treating time as categorical) structure and simplify.
- Use Goodness of Fit (GOF) procedures AIC and/or BIC
- Use the estimated correlation matrix of the Y's as a fixed (numerical) structure.
- Start with an independence correlation structure, fit the model, obtain residuals, then use the correlation matrix of the residuals as a fixed (numerical) structure.
- Use a random effects model.

Comparing models with different covariance structures

Likelihood Ratio Test / Likelihood Ratio Statistic

- Can be used for “nested” covariance structures

- The Likelihood Ratio Test Statistic is based on the change in value of the log-likelihood (Log L) or restricted log-likelihood (REML Log L) between two <u>nested</u> models.

$$LR = -2LN\left[\frac{Likelihood(reduced)}{Likelihood(full)}\right] \sim \chi^2_{df(full)-df(reduced)}$$

$$LR = [-2LogL(reduced)] - [-2LogL(full)]$$

Comparing models with different covariance structures

Likelihood-Based Information Criteria

- The relative fit of different covariance structures can be compared using the AIC or BIC, which are based on the REML log-likelihood.

- By adding $2p$ to $-2\times$REML log-likelihood (where p is the number of parameters), you can calculate *Akaike Information Criterion* (*AIC*).

- By adding $p\times\ln(n)$ to $-2\times$REMLlog-likelihood (**where *n* is the number of subjects, <u>not</u> the number of observations**), you get the *Bayesian information criteria* (*BIC*). [also called Schwartz criterion (SC)].

- Both the AIC and BIC statistics "penalize" the log-likelihood for estimating more parameters (with the BIC criterion producing the more severe penalization).

    - For ML, p is the number of covariance parameters plus the number of beta coefficients for the fixed effects.

    - FOR REML, p is the number of covariance parameters estimated.

- <u>Lower</u> values of AIC and BIC correspond to more desirable models.

Nested Models

Independence (1) Compound Symmetry (2) Toeplitz (4) Huynh-Feldt (5) Unstructured (10)

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}
\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}
\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}
\begin{bmatrix} \sigma_1 & \frac{\sigma_1+\sigma_2}{2}-\lambda & \frac{\sigma_1+\sigma_3}{2}-\lambda & \frac{\sigma_1+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_2}{2}-\lambda & \sigma_2 & \frac{\sigma_2+\sigma_3}{2}-\lambda & \frac{\sigma_2+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_3}{2}-\lambda & \frac{\sigma_2+\sigma_3}{2}-\lambda & \sigma_3 & \frac{\sigma_3+\sigma_4}{2}-\lambda \\ \frac{\sigma_1+\sigma_4}{2}-\lambda & \frac{\sigma_2+\sigma_4}{2}-\lambda & \frac{\sigma_3+\sigma_4}{2}-\lambda & \sigma_4 \end{bmatrix}
\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}$$

Reduced Model	Full Model	Nested?
Simple	Compound Symmetry	Yes (σ_1=0 for CS)
	Toeplitz	Yes ($\sigma_1 = \sigma_2 = \sigma_3 = 0$ for Toeplitz)
	Huynh-Feldt	Yes ($\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \lambda$ for Huynh-Feldt)
	Unstructured	Yes ($\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44}$ and $\sigma_{12} = \sigma_{13} = \sigma_{14} = \sigma_{23} = \sigma_{24} = \sigma_{34} = 0$)
Compound Symmetry	Toeplitz	Yes ($\sigma_1 = \sigma_2 = \sigma_3$ for Toeplitz)
	Huynh-Feldt	Yes ($\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$ for Huynh-Feldt)
	Unstructured	Yes ($\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44}$ and $\sigma_{12} = \sigma_{13} = \sigma_{14} = \sigma_{23} = \sigma_{24} = \sigma_{34}$)
Toeplitz	Huynh-Feldt	No ($\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44} \Rightarrow (\sigma_{11}+\sigma_{22})/2-\lambda = (\sigma_{22}+\sigma_{33})/2-\lambda$) i.e. all the off diagonal is the same!
	Unstructured	Yes ($\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma_{44}$ and $\sigma_{12} = \sigma_{23} = \sigma_{34}$ and $\sigma_{13} = \sigma_{24}$)
Huynh-Feldt	Unstructured	Yes

Example: Likelihood Ratio

Likelihood Ratio Test Statistic (Type=HF vs Type=UN):

LR =421.7206 – 414.0348= 7.6858 ~ χ_5^2 , p = 0.1744

-2LogL (reduced)	-2LogL (full)

Conclusion: No significant difference between full and reduced. We can use the reduced (simpler) covariance structure (HF).

Likelihood Ratio Test Statistic (Type=CS vs Type=HF):

LR = 423.4085– 421.7206= 1.6879~ χ_3^2 , p = 0.6396

Conclusion: No significant difference between full and reduced. We can use the reduced (simpler) covariance structure (CS).

Example: BIC and AIC

	Covariance Parameters	-2 REML Log L	AIC	BIC
SIMPLE	1	470.4908	472.5	473.8
CS	2	423.4085	427.4	430.0
TOEP	4	418.9499	426.9	432.1
HF	5	421.7206	431.7	438.2
UN	10	414.0348	434.0	447.0

TOEP is best using AIC

Conclusion: CS covariance is best using BIC

Note: R will compute AIC and BIC based on total number of parameters (regression and covariance parameters)

- since we are only comparing models with different covariance structures here, this distinction isn't important
- i.e., differences in AIC and BIC (all that is meaningful for these statistics anyway) will be the same regardless whether we include the regression parameters in our total parameter count
- can use `-2*logLik(model) + 2*(1+ length(model$modelStruct$corStruct)` to get just the covariance parameters in AIC

Step 4: Test Fixed Effects (based on Step 3, use TYPE=CS)

TYPE=TOEP

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.19	0.0056
age	3	75	29.97	<.0001
gender*age	3	75	2.34	0.0801

No significant gender × age interaction. Consider dropping interaction term.

TYPE=CS

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.29	0.0054
age	3	75	35.35	<.0001
gender*age	3	75	2.36	0.0781

TYPE=HF

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.39	0.0052
age	3	75	35.35	<.0001
gender*age	3	75	2.36	0.0781

TYPE=UN

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.29	0.0054
age	3	25	34.45	<.0001
gender*age	3	25	2.93	0.0532

Test Fixed Effects (R code/output)

```
# get ANOVA tables
# Toeplitz
anova(mod4)

## Denom. DF: 100
##                        numDF  F-value p-value
## (Intercept)                1 4133.718  <.0001
## Gender                     1    9.911  0.0022
## as.factor(age)             3   34.138  <.0001
## Gender:as.factor(age)      3    2.340  0.0779

# CS
anova(mod2)

## Denom. DF: 100
##                        numDF  F-value p-value
## (Intercept)                1 4123.164  <.0001
## Gender                     1    9.292  0.0029
## as.factor(age)             3   40.032  <.0001
## Gender:as.factor(age)      3    2.362  0.0759
```

```
# unstructured
anova(mod3)

## Denom. DF: 100
##                          numDF  F-value p-value
## (Intercept)                  1 4246.072  <.0001
## Gender                       1    7.719  0.0065
## as.factor(age)               3   39.441  <.0001
## Gender:as.factor(age)        3    2.930  0.0373
```

Why are these results different?

- SAS and R use different degrees of freedom for *F* tests
- ... but results are still quite similar

Return to Step 1: Select Fixed Effects

- Age
- Gender

Step 2/3: Select/Test Covariance Structure

	Covariance Parameters	-2 REML Log L	AIC	BIC
CS	2	433.4069	437.7	440.3
TOEP	4	428.7481	436.7	441.9
HF	5	432.0197	442.0	448.5
UN	10	424.8186	444.8	457.8

Conclusion: CS or TOEP depending on AIC/BIC!

NOTE: Do not compare these -2REML Log L/AIC/BIC to those from the interaction model. We cannot use REML estimates to compare models with different fixed effects!

Step 4: Test Fixed Effects

TYPE=CS

```
Type 3 Tests of Fixed Effects

                 Num      Den
Effect            DF       DF    F Value    Pr > F

gender             1       25       9.29    0.0054
age                3       78      38.04    <.0001
```

TYPE=TOEP

```
Type 3 Tests of Fixed Effects

                 Num      Den
Effect            DF       DF    F Value    Pr > F

gender             1       25      10.06    0.0040
age                3       78      29.60    <.0001
```

Step 5: Inference for Fixed Effects

NOTE: Although the interaction wasn't significant, I am going to perform inference on the model with the interaction term to be consistent with the Wolfinger article.

```
*** Mixed Model, CS Covariance ***;
*** With parameter estimates ***;
PROC MIXED DATA=formixed;
 CLASS gender age person;
 MODEL y = gender age gender*age /SOLUTION DDFM=SAT;
 REPEATED age /TYPE=CS SUBJECT=person R RCORR;
 LSMEANS gender / PDIFF CL;
 LSMEANS age / PDIFF CL;
RUN;
```

```
                            The Mixed Procedure

                             Model Information
           Data Set                     WORK.FORMIXED
           Dependent Variable           y
           Covariance Structure         Compound Symmetry
           Subject Effect               person
           Estimation Method            REML
           Residual Variance Method     Profile
           Fixed Effects SE Method      Model-Based
           Degrees of Freedom Method    Satterthwaite


                          Class Level Information

             Class      Levels    Values
             gender          2    F M
             age             4    8 10 12 14
             person         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                                  14 15 16 17 18 19 20 21 22 23
                                  24 25 26 27
```

Dimensions

Covariance Parameters	2
Columns in X	15
Columns in Z	0
Subjects	27
Max Obs Per Subject	4

Number of Observations

Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	470.49084642	
1	1	423.40853283	0.00000000

Convergence criteria met.

The Mixed Procedure

Estimated R Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	5.2604	3.2854	3.2854	3.2854
2	3.2854	5.2604	3.2854	3.2854
3	3.2854	3.2854	5.2604	3.2854
4	3.2854	3.2854	3.2854	5.2604

Estimated R Correlation Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6245	0.6245	0.6245
2	0.6245	1.0000	0.6245	0.6245
3	0.6245	0.6245	1.0000	0.6245
4	0.6245	0.6245	0.6245	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	person	3.2854
Residual		1.9750

Fit Statistics

-2 Res Log Likelihood	423.4
AIC (smaller is better)	427.4
AICC (smaller is better)	427.5
BIC (smaller is better)	430.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	47.08	<.0001

Solution for Fixed Effects

Effect	gender	Age	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			27.4687	0.5734	46.1	47.91	<.0001
gender	F		-3.3778	0.8983	46.1	-3.76	0.0005
gender	M		0	.	.	.	.
age		8	-4.5937	0.4969	75	-9.25	<.0001
age		10	-3.6562	0.4969	75	-7.36	<.0001
age		12	-1.7500	0.4969	75	-3.52	0.0007
age		14	0	.	.	.	.
gender*age	F	8	1.6847	0.7784	75	2.16	0.0336
gender*age	F	10	1.7926	0.7784	75	2.30	0.0241
gender*age	F	12	0.7500	0.7784	75	0.96	0.3384
gender*age	F	14	0	.	.	.	.
gender*age	M	8	0	.	.	.	.
gender*age	M	10	0	.	.	.	.
gender*age	M	12	0	.	.	.	.
gender*age	M	14	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	9.29	0.0054
age	3	75	35.35	<.0001
gender*age	3	75	2.36	0.0781

Least Squares Means

Effect	gender	Age	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
gender	F		22.6477	0.5861	25	38.64	<.0001	0.05
gender	M		24.9687	0.4860	25	51.38	<.0001	0.05
age		8	22.0284	0.4492	46.1	49.04	<.0001	0.05
age		10	23.0199	0.4492	46.1	51.25	<.0001	0.05
age		12	24.4048	0.4492	46.1	54.33	<.0001	0.05
age		14	25.7798	0.4492	46.1	57.39	<.0001	0.05

Least Squares Means

Effect	gender	Age	Lower	Upper
gender	F		21.4406	23.8549
gender	M		23.9678	25.9697
age		8	21.1243	22.9325
age		10	22.1158	23.9240
age		12	23.5007	25.3089
age		14	24.8757	26.6839

Differences of Least Squares Means

Effect	gender	Age	_gender	Age	Estimate	Standard Error	DF	t Value
gender	F		M		-2.3210	0.7614	25	-3.05
age		8		10	-0.9915	0.3892	75	-2.55
age		8		12	-2.3764	0.3892	75	-6.11
age		8		14	-3.7514	0.3892	75	-9.64
age		10		12	-1.3849	0.3892	75	-3.56
age		10		14	-2.7599	0.3892	75	-7.09
age		12		14	-1.3750	0.3892	75	-3.53

Differences of Least Squares Means

Effect	gender	Age	_gender	Age	Pr > |t|	Alpha	Lower	Upper
gender	F		M		0.0054	0.05	-3.8892	-0.7529
age		8		10	0.0129	0.05	-1.7668	-0.2161
age		8		12	<.0001	0.05	-3.1518	-1.6010
age		8		14	<.0001	0.05	-4.5268	-2.9760
age		10		12	0.0007	0.05	-2.1603	-0.6096
age		10		14	<.0001	0.05	-3.5353	-1.9846
age		12		14	0.0007	0.05	-2.1504	-0.5996

```r
# new model with male as reference group for gender and oldest as reference group for age
dental.long$Gender <- relevel(dental.long$Gender,ref='M')
dental.long$age <- relevel(as.factor(dental.long$age),ref='14')
mod2 <- gls(distance ~ Gender*age, # don't need as.factor() now because of recoding
            correlation=corCompSymm(form=~1|Person),
            data=dental.long)
# summary(mod2)

getVarCov(mod2)

## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 5.2604 3.2854 3.2854 3.2854
## [2,] 3.2854 5.2604 3.2854 3.2854
## [3,] 3.2854 3.2854 5.2604 3.2854
## [4,] 3.2854 3.2854 3.2854 5.2604
##   Standard Deviations: 2.2936 2.2936 2.2936 2.2936

corMatrix(mod2$modelStruct$corStruct)[[1]] # just use Person=1 here since all are the same

##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6245472 0.6245472 0.6245472
## [2,] 0.6245472 1.0000000 0.6245472 0.6245472
## [3,] 0.6245472 0.6245472 1.0000000 0.6245472
## [4,] 0.6245472 0.6245472 0.6245472 1.0000000

# L matrix for group-specific means
coef(mod2) # check the ordering of estimates

##   (Intercept)       GenderF          age8         age10         age12
##     27.468750     -3.377841     -4.593750     -3.656250     -1.750000
##  GenderF:age8 GenderF:age10 GenderF:age12
##      1.684659      1.792614      0.750000

L <- rbind(c(1,1,1/4,1/4,1/4,1/4,1/4,1/4), # female mean
           c(1,0,1/4,1/4,1/4,0,0,0), # male mean
           c(1,1/2,1,0,0,1/2,0,0), # age 8 mean
           c(1,1/2,0,1,0,0,1/2,0), # age 10 mean
           c(1,1/2,0,0,1,0,0,1/2), # age 12 mean
           c(1,1/2,0,0,0,0,0,0) # age 14 mean
)
```

```r
means.est <- L %*% coef(mod2)
means.se <- sqrt(diag(L %*% vcov(mod2) %*% t(L)))

ls.means <- data.frame(estimate=means.est,std.err=means.se)
rownames(ls.means) <- c('female','male',seq(8,14,by=2))
ls.means
```

```
##        estimate   std.err
## female 22.64773 0.5861384
## male   24.96875 0.4860003
## 8      22.02841 0.4491648
## 10     23.01989 0.4491648
## 12     24.40483 0.4491648
## 14     25.77983 0.4491648
```

```r
# get L matrix for mean differences from rows of L matrix for means
L.diffs <- rbind(L[1,]-L[2,], # female - male
                 L[3,]-L[4,], # 8 - 10
                 L[3,]-L[5,], # 8 - 12
                 L[3,]-L[6,], # 8 - 14
                 L[4,]-L[5,], # 10 - 12
                 L[4,]-L[6,], # 10 - 14
                 L[5,]-L[6,] # 12 - 14
)
diffs.est <- L.diffs %*% coef(mod2)
diffs.se <- sqrt(diag(L.diffs %*% vcov(mod2) %*% t(L.diffs)))

ls.diffs <- data.frame(estimate=diffs.est,std.err=diffs.se)
rownames(ls.diffs) <- c('female - male',
                        '8 - 10',
                        '8 - 12',
                        '8 - 14',
                        '10 - 12',
                        '10 - 14',
                        '12 - 14')
ls.diffs
```

```
##                 estimate   std.err
## female - male -2.3210227 0.7614161
## 8 - 10        -0.9914773 0.3892229
```

```
## 8 - 12         -2.3764205 0.3892229
## 8 - 14         -3.7514205 0.3892229
## 10 - 12        -1.3849432 0.3892229
## 10 - 14        -2.7599432 0.3892229
## 12 - 14        -1.3750000 0.3892229

# Alternative method using lsmeans or emmeans package

library(lsmeans)
## Loading required package: emmeans
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
emm_options(opt.digits = FALSE) # to force more digits to print
# means for gender
emmeans(mod2,spec='Gender')
## NOTE: Results may be misleading due to involvement in interactions

##  Gender   emmean        SE    df lower.CL upper.CL
##  M      24.96875 0.4860003 34.60 23.98171 25.95579
##  F      22.64773 0.5861384 40.49 21.46354 23.83191
##
## Results are averaged over the levels of: age
## d.f. method: satterthwaite
## Confidence level used: 0.95
```

Effect	gender	Age	Estimate	Error	DF	t Value	Pr > |t|	Alpha
gender	F		22.6477	0.5861	25	38.64	<.0001	0.05
gender	M		24.9687	0.4860	25	51.38	<.0001	0.05

```
# means for age
emmeans(mod2,spec='age')

## NOTE: Results may be misleading due to involvement in interactions

##  age   emmean        SE    df lower.CL upper.CL
##  14  25.77983 0.4491648 21.29 24.84651 26.71315
##  8   22.02841 0.4491648 18.33 21.08596 22.97086
##  10  23.01989 0.4491648 35.74 22.10871 23.93106
##  12  24.40483 0.4491648 33.93 23.49195 25.31771
##
## Results are averaged over the levels of: Gender
## d.f. method: satterthwaite
## Confidence level used: 0.95
```

Effect	gender	Age	Estimate	Error	DF	t Value	Pr > |t|	Alpha
age		8	22.0284	0.4492	46.1	49.04	<.0001	0.05
age		10	23.0199	0.4492	46.1	51.25	<.0001	0.05
age		12	24.4048	0.4492	46.1	54.33	<.0001	0.05
age		14	25.7798	0.4492	46.1	57.39	<.0001	0.05

```
*** CELL MEANS Mixed Model, CS Covariance ***;
*** With parameter estimates ***;

PROC MIXED DATA=formixed ;
 CLASS gender age person;
 MODEL y = gender*age/SOLUTION NOINT DDFM=SAT;
 REPEATED age /TYPE=CS SUBJECT=person R RCORR;
 CONTRAST 'Males vs Females'
              gender*age .25 .25 .25 .25 -.25 -.25 -.25 -.25;
 ESTIMATE 'Males vs Females'
               gender*age .25 .25 .25 .25 -.25 -.25 -.25 -.25;
 ESTIMATE 'M v F: Age  8' gender*age  1  0  0  0 -1  0  0  0;
 ESTIMATE 'M v F: Age 10' gender*age  0  1  0  0  0 -1  0  0;
 ESTIMATE 'M v F: Age 12' gender*age  0  0  1  0  0  0 -1  0;
 ESTIMATE 'M v F: Age 14' gender*age  0  0  0  1  0  0  0 -1;
RUN;
```

$$H_0 : \frac{\mu_{F,Age8} + \mu_{F,Age10} + \mu_{F,Age12} + \mu_{F,Age14}}{4} = \frac{\mu_{M,Age8} + \mu_{M,Age10} + \mu_{M,Age12} + \mu_{M,Age14}}{4}$$

```
                  The Mixed Procedure

                   Model Information

     Data Set                     WORK.FORMIXED
     Dependent Variable           y
     Covariance Structure         Compound Symmetry
     Subject Effect               person
     Estimation Method            REML
     Residual Variance Method     Profile
     Fixed Effects SE Method      Model-Based
     Degrees of Freedom Method    Satterthwaite
```

Class Level Information

Class	Levels	Values
gender	2	F M
age	4	8 10 12 14
person	27	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Dimensions

Covariance Parameters	2
Columns in X	8
Columns in Z	0
Subjects	27
Max Obs Per Subject	4

Number of Observations

Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	470.49084642	
1	1	423.40853283	0.00000000

Convergence criteria met.

Estimated R Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	5.2604	3.2854	3.2854	3.2854
2	3.2854	5.2604	3.2854	3.2854
3	3.2854	3.2854	5.2604	3.2854
4	3.2854	3.2854	3.2854	5.2604

Estimated R Correlation Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6245	0.6245	0.6245
2	0.6245	1.0000	0.6245	0.6245
3	0.6245	0.6245	1.0000	0.6245
4	0.6245	0.6245	0.6245	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	person	3.2854
Residual		1.9750

Fit Statistics

-2 Res Log Likelihood	423.4
AIC (smaller is better)	427.4
AICC (smaller is better)	427.5
BIC (smaller is better)	430.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	47.08	<.0001

Cell Means Model Replicates observed means

Solution for Fixed Effects

Effect	gender	Age	Estimate	Standard Error	DF	t Value	Pr > |t|
gender*age	F	8	21.1818	0.6915	46.1	30.63	<.0001
gender*age	F	10	22.2273	0.6915	46.1	32.14	<.0001
gender*age	F	12	23.0909	0.6915	46.1	33.39	<.0001
gender*age	F	14	24.0909	0.6915	46.1	34.84	<.0001
gender*age	M	8	22.8750	0.5734	46.1	39.89	<.0001
gender*age	M	10	23.8125	0.5734	46.1	41.53	<.0001
gender*age	M	12	25.7188	0.5734	46.1	44.85	<.0001
gender*age	M	14	27.4688	0.5734	46.1	47.91	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender*age	8	49.3	532.45	<.0001

```r
# reset reference groups
dental.long$Gender <- relevel(dental.long$Gender,ref='F')
dental.long$age <- as.numeric(as.character(dental.long$age))
mod2cm <- gls(distance ~ 0+as.factor(age):Gender, # only include interaction term and 0 for no intercept
              correlation=corCompSymm(form=~1|Person),
              data=dental.long)
summary(mod2cm)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ 0 + as.factor(age):Gender
##   Data: dental.long
##        AIC      BIC    logLik
##   443.4085 469.4602 -211.7043
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | Person
##  Parameter estimate(s):
##       Rho
## 0.6245472
##
## Coefficients:
##                              Value Std.Error  t-value p-value
## as.factor(age)8:GenderF  21.18182 0.6915345 30.63017       0
## as.factor(age)10:GenderF 22.22727 0.6915345 32.14196       0
## as.factor(age)12:GenderF 23.09091 0.6915345 33.39083       0
## as.factor(age)14:GenderF 24.09091 0.6915345 34.83689       0
## as.factor(age)8:GenderM  22.87500 0.5733901 39.89430       0
## as.factor(age)10:GenderM 23.81250 0.5733901 41.52932       0
## as.factor(age)12:GenderM 25.71875 0.5733901 44.85384       0
## as.factor(age)14:GenderM 27.46875 0.5733901 47.90586       0
## Residual standard error: 2.293561
## Degrees of freedom: 108 total; 100 residual
```

```r
anova(mod2cm)
```

```
## Denom. DF: 100
##                        numDF  F-value p-value
## as.factor(age):Gender     8 532.4544  <.0001
```

Results of ESTIMATE Statements

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Males vs Females	-2.3210	0.7614	25	-3.05	0.0054
M v F: Age 8	-1.6932	0.8983	46.1	-1.88	0.0658
M v F: Age 10	-1.5852	0.8983	46.1	-1.76	0.0843
M v F: Age 12	-2.6278	0.8983	46.1	-2.93	0.0053
M v F: Age 14	-3.3778	0.8983	46.1	-3.76	0.0005

Results of CONTRAST Statement

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Males vs Females	1	25	9.29	0.0054

Repeat of ESTIMATE Statements from p.37

```
ESTIMATE 'Males vs Females'
                  gender*age .25 .25 .25 .25 -.25 -.25 -.25 -.25;
ESTIMATE 'M v F: Age  8' gender*age  1  0  0  0 -1  0  0  0;
ESTIMATE 'M v F: Age 10' gender*age  0  1  0  0  0 -1  0  0;
ESTIMATE 'M v F: Age 12' gender*age  0  0  1  0  0  0 -1  0;
ESTIMATE 'M v F: Age 14' gender*age  0  0  0  1  0  0  0 -1;
```

Repeat of CONTRAST Statement from p.37

```
CONTRAST 'Males vs Females'
                  gender*age .25 .25 .25 .25 -.25 -.25 -.25 -.25;
```

CONCLUSION: Overall there is a significant difference in dental measurements between boys and girls ages 8-14 (p = 0.0054). There difference between boys and girls is not significant at age 8 (p = 0.0658) and age 10 (p = 0.0843) but is significant at age 12 (p = 0.0053) and age 14 (p = 0.0005).

One Final Covariance Structure to Consider

Allow different covariance matrix for males & females

Display matrix for observations 1 and 12

```
*** Mixed Model, Unstructured Covariance ***;
**** BY GENDER *****;
PROC MIXED DATA=formixed;
 CLASS gender age person;
 MODEL y = gender age gender*age;
 REPEATED age /TYPE=UN SUBJECT=person R=1,12 RCORR=1,12 GROUP=gender;
RUN;
```

```
                    The Mixed Procedure

                     Model Information
      Data Set                       WORK.FORMIXED
      Dependent Variable             y
      Covariance Structure           Unstructured
      Subject Effect                 person
      Group Effect                   gender
      Estimation Method              REML
      Residual Variance Method       None
      Fixed Effects SE Method        Model-Based
      Degrees of Freedom Method      Between-Within

                  Class Level Information

        Class      Levels    Values
        gender          2    F M
        age             4    8 10 12 14
        person         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                             14 15 16 17 18 19 20 21 22 23
                             24 25 26 27

                          Dimensions
           Covariance Parameters              20
           Columns in X                       15
           Columns in Z                        0
```

20 covariance parameters to be estimated!

Subjects	27
Max Obs Per Subject	4

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	470.49084642	
1	1	392.85396403	0.00000000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	person	gender F	4.5136
UN(2,1)	person	gender F	3.3545
UN(2,2)	person	gender F	3.6182
UN(3,1)	person	gender F	4.3318
UN(3,2)	person	gender F	4.0273
UN(3,3)	person	gender F	5.5909
UN(4,1)	person	gender F	4.3568
UN(4,2)	person	gender F	4.0773
UN(4,3)	person	gender F	5.4659
UN(4,4)	person	gender F	5.9409
UN(1,1)	person	gender M	6.0167
UN(2,1)	person	gender M	2.2917
UN(2,2)	person	gender M	4.5625
UN(3,1)	person	gender M	3.6292
UN(3,2)	person	gender M	2.1938
UN(3,3)	person	gender M	7.0323
UN(4,1)	person	gender M	1.6125
UN(4,2)	person	gender M	2.8104
UN(4,3)	person	gender M	3.2406
UN(4,4)	person	gender M	4.3490

Different covariance matrix (TYPE=UN) estimated for boys and girls.

Estimated R Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	4.5136	3.3545	4.3318	4.3568
2	3.3545	3.6182	4.0273	4.0773
3	4.3318	4.0273	5.5909	5.4659
4	4.3568	4.0773	5.4659	5.9409

Estimated R Correlation Matrix for person 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8301	0.8623	0.8414
2	0.8301	1.0000	0.8954	0.8794
3	0.8623	0.8954	1.0000	0.9484
4	0.8414	0.8794	0.9484	1.0000

Estimated R Matrix for person 12

Row	Col1	Col2	Col3	Col4
1	6.0167	2.2917	3.6292	1.6125
2	2.2917	4.5625	2.1938	2.8104
3	3.6292	2.1938	7.0323	3.2406
4	1.6125	2.8104	3.2406	4.3490

Estimated R Correlation Matrix for person 12

Row	Col1	Col2	Col3	Col4
1	1.0000	0.4374	0.5579	0.3152
2	0.4374	1.0000	0.3873	0.6309
3	0.5579	0.3873	1.0000	0.5860
4	0.3152	0.6309	0.5860	1.0000

Fit Statistics

-2 Res Log Likelihood	392.9
AIC (smaller is better)	432.9
AICC (smaller is better)	443.5
BIC (smaller is better)	458.8

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	25	8.80	0.0065
age	3	75	41.47	<.0001
gender*age	3	75	3.44	0.0210

Is this covariance structure better (according to AIC)?

No, not according to AIC

Can run this analysis in R as well, but need to split the data first into males and females and fit separate models

- 'nlme' `gls()` function doesn't allow covariance structure to depend on factors that vary between subjects
- Splitting the data works in this case because we have a saturated model for the mean (all possible covariate patterns are allowed to have separate means)

```
# unstructured covariance with different covariance matrix for males and females
# have to split this into models for males and females separately
mod5male <- gls(distance ~ as.factor(age),
            correlation=corSymm(form=~1|Person),
            weights=varIdent(form=~1|as.numeric(age)),
            data=subset(dental.long,Gender=='M'))
# summary(mod5male)
corMatrix(mod5male$modelStruct$corStruct)[[1]]

##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.4373933 0.5579296 0.3152311
## [2,] 0.4373933 1.0000000 0.3872902 0.6309231
## [3,] 0.5579296 0.3872902 1.0000000 0.5859863
## [4,] 0.3152311 0.6309231 0.5859863 1.0000000

# have to split this into models for males and females separately
mod5female <- gls(distance ~ as.factor(age),
                correlation=corSymm(form=~1|Person),
                weights=varIdent(form=~1|as.numeric(age)),
                data=subset(dental.long,Gender=='F'))
# summary(mod5female)
corMatrix(mod5female$modelStruct$corStruct)[[1]]

##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.8300897 0.8623149 0.8413559
## [2,] 0.8300897 1.0000000 0.8954158 0.8794235
## [3,] 0.8623149 0.8954158 1.0000000 0.9484072
## [4,] 0.8413559 0.8794235 0.9484072 1.0000000
```