

# BIO 226, Lab 4: Fitting Parametric Curves and Modeling the Covariance

TA: Godwin Yung

Tuesday March 12, 2015

# Introduction

Longitudinal data requires modeling of both mean and covariance structure.

- ① Covariance between measurements (assuming model for  $\beta$ ):
  - Unstructured
  - Modeled according to specific pattern e.g. autoregressive, banded
  - Random effects (later in the course)
- ② Mean response over time (assuming model for  $\Sigma$ ):
  - Analysis of response profile
  - Parametric or semi-parametric curves

# Today's Lab

Our objective is to analyze and interpret a two group repeated measures study by considering an appropriate covariance structure and parametric curve models: quadratic, linear.

## Outline:

Part I: Data: weight loss study

Choose a covariance structure

Part II: Model the mean using analysis of response profiles

Model the mean using parametric curves

Interpret covariate parameter estimates in final model

## Data: Weight Loss Study

Sixty overweight males were randomly assigned to one of two weight loss programs. Both programs included a diet regimen and daily exercise routine. In addition, subjects in program 1 received daily encouragement from a representative from the study, while those in program 2 did not. Weight (in pounds) was measured at baseline (month 0) and at 3, 6, and 9 months.

We have a *balanced* design with all subjects measured at the same four time points. Also, our data are *complete* (no missing data).



# Inputting Data

```
proc format;  
    value prog 1="1:encourage"  
              2="2:none";  
run;  
  
/* Input data in wide format */  
data wtloss;  
    infile 'weightloss.dat';  
    input id y1-y4 program;  
    format program prog.;  
run;  
  
proc print data=wtloss;  
run;
```

# Data in Original (wide) Format

Obs	id	y1	y2	y3	y4	program
1	1	266.1	254.8	222.0	175.5	1: encourage
2	2	248.1	198.4	170.8	149.1	1: encourage
3	3	253.2	207.1	167.3	159.1	1: encourage
4	4	263.1	221.7	207.8	171.8	1: encourage
5	5	267.5	250.2	223.9	188.4	1: encourage
...						
35	35	244.6	232.8	218.6	207.8	2: none
36	36	265.9	251.0	237.6	237.8	2: none
37	37	209.3	196.6	192.5	190.4	2: none
38	38	258.8	233.7	233.8	208.5	2: none
39	39	232.5	224.8	221.4	199.7	2: none

# Reformatting Data and Calculating Means

```
/* Create dataset with a variable for month coded 0,3,6,9,  
a variable for month category coded 11 for month 0  
(so that month 0 will be the reference category in proc mixed),  
and a housekeeping variable t. Also change data from wide format  
to long format to be used in proc MIXED */
```

```
data wtloss2;  
    set wtloss;  
    weight=y1; month=0; monthcat=11; t=1; output;  
    weight=y2; month=3; monthcat=3; t=2; output;  
    weight=y3; month=6; monthcat=6; t=3; output;  
    weight=y4; month=9; monthcat=9; t=4; output;  
    format program prog.;  
    drop y1-y4;  
  
run;  
  
proc means data=wtloss2 n mean std nway;  
    var weight;  
    class program month;  
    output out=meandata mean=meanweight;  
  
run;
```

# Mean Weight by Program and Month

program	month	Obs	N	Mean	Std Dev
1: encourage	0	34	34	241.6088235	23.0151250
	3	34	34	219.2058824	21.1974547
	6	34	34	193.7705882	21.9254733
	9	34	34	169.0441176	21.9959665
2: none	0	26	26	244.9500000	21.7206676
	3	26	26	237.0038462	20.4058321
	6	26	26	227.1192308	24.6290076
	9	26	26	218.0346154	20.5998047

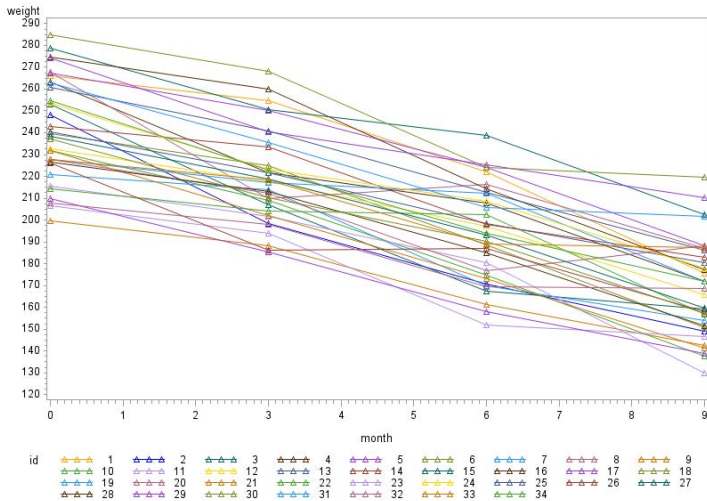


# Plotting Response Profiles

```
/* Plot all response profiles on 2 plots (one for each group) */  
  
/* Use different colors for each person */  
  
proc gplot data=wtloss2;  
symbol1 interpol=join value=triangle;  
symbol2 interpol=join value=triangle;  
symbol3 interpol=join value=triangle;  
symbol4 interpol=join value=triangle;  
symbol5 interpol=join value=triangle;  
plot weight*month=id;  
where program=1;  
Title 'Weight Profiles for each person in Program 1';  
run;  
  
proc gplot data=wtloss2;  
plot weight*month=id;  
where program=2;  
Title 'Weight Profiles for each person in Program 2';  
run;
```

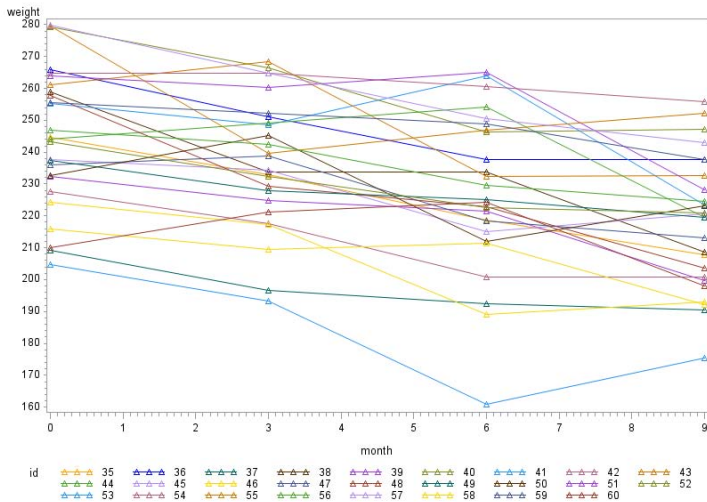
# Response Profile Plots

Weight Profiles for each person in Program 1



# Response Profile Plots

Weight Profiles for each person in Program 2

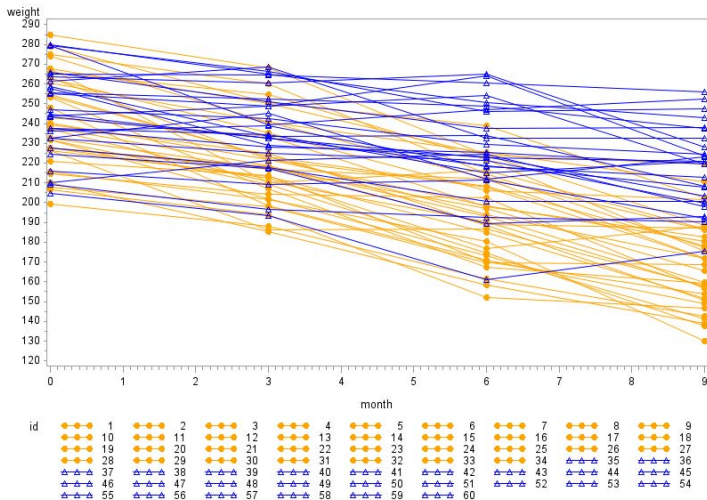


# Plotting Response Profiles

```
/* Plot all response profiles on one plot */  
/* Use 2 colors, one for each group */  
proc gplot data=wtloss2;  
symbol1 color=orange  
         interpol=join  
         value=dot  
         repeat=34;  
symbol2 color=blue  
         interpol=join  
         value=triangle  
         repeat=26;  
plot weight*month=id;  
Title 'Weight Profiles for all Subjects by Program';  
run;
```

# Response Profile Plots

Weight Profiles for all Subjects by Program

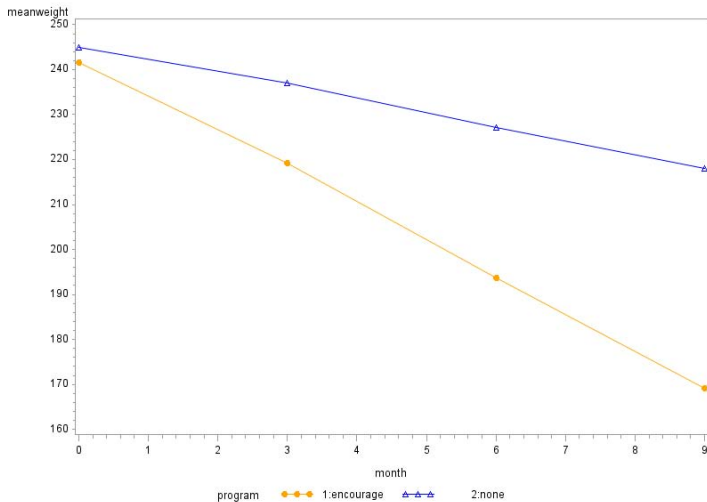


# Plotting Mean Response Profiles

```
/* Plot mean response profiles */  
proc means data=wtloss2 n mean std nway;  
    var weight;  
    class program month;  
    output out=meandata mean=meanweight;  
run;  
  
proc gplot data=meandata;  
    symbol1 color=orange  
        interpol=join  
        value=dot;  
    symbol2 color=blue  
        interpol=join  
        value=triangle;  
    plot meanweight*month=program;  
    format program prog.;  
    Title 'Mean Weight Profiles for each Program';  
run;
```

# Mean Response Profile Plots

Mean Weight Profiles for each Program



# Approach for choosing a Covariance Structure

- We looked at the data, now we want to model the covariance
- In selecting the model for the covariance, we need to temporarily assume some mean model.
- We assume a “maximal” mean model: this minimizes any potential misspecification.
- Once we have selected the model for the covariance, we will then consider simpler, more parsimonious models for the mean. Simpler models (if appropriate) will have more power to detect differences in the mean patterns



# Maximal Mean Model

Viewing time as categorical (“analysis of response profiles”) avoids putting assumptions about how the mean changes with time.

*Maximal Model:*

$$Y_{ij} = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij} + \beta_7 X_{7ij} + \beta_8 X_{8ij} + \epsilon_{ij}$$

- $X_{1ij} = 1$  for all measurements
- $X_{2ij} = 1$  if  $\text{Month}_{ij} = 3$ , 0 otherwise
- $X_{3ij} = 1$  if  $\text{Month}_{ij} = 6$ , 0 otherwise
- $X_{4ij} = 1$  if  $\text{Month}_{ij} = 9$ , 0 otherwise
- $X_{5ij} = 1$  if subject  $i$  is encouraged, 0 otherwise
- $X_{6ij} = X_{2ij} * X_{5ij} = 1$  if subject  $i$  is encouraged &  $\text{Month}_{ij} = 3$ , 0 otherwise
- $X_{7ij} = X_{3ij} * X_{5ij} = 1$  if subject  $i$  is encouraged &  $\text{Month}_{ij} = 6$ , 0 otherwise
- $X_{8ij} = X_{4ij} * X_{5ij} = 1$  if subject  $i$  is encouraged &  $\text{Month}_{ij} = 9$ , 0 otherwise
- $i = 1, \dots, 60$  subjects
- $j = 1, \dots, 4$  weight loss measurements

- $\epsilon_i = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \sim N(0, \Sigma)$ , where  $\Sigma$  is a 4x4 Covariance matrix

## Maximal Mean Model cont.

This model allows for a different mean between every program at every time occasion. It is called a “fully saturated” model because there are the same number of parameters (8) as there are covariate patterns (2 groups \* 4 time occasions). This model does not make any assumptions about the pattern of the mean changes.

Parameters correspond to the following means:

$\mu_{21} = \beta_1$	$\mu_{11} = \beta_1 + \beta_5$
$\mu_{22} = \beta_1 + \beta_2$	$\mu_{12} = \beta_1 + \beta_2 + \beta_5 + \beta_6$
$\mu_{23} = \beta_1 + \beta_3$	$\mu_{13} = \beta_1 + \beta_3 + \beta_5 + \beta_7$
$\mu_{24} = \beta_1 + \beta_4$	$\mu_{14} = \beta_1 + \beta_4 + \beta_5 + \beta_8$

Note: This is only possible with a balanced (equally-timed) design

# Covariance Structures

Now that we have temporarily assumed the fully saturated model as our maximal model, we will select the model for the covariance.

We will consider three covariance structures (there are certainly more)

- unstructured
- compound symmetry
- autoregressive-1

# Covariance Structures cont.

## 1 unstructured (UN)

$$\text{Cov}(\mathbf{Y}_i) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

## 2 compound symmetry (CS)

$$\text{Cov}(\mathbf{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

## 3 autoregressive (AR-1)

$$\text{Cov}(\mathbf{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

Q1: How many parameters are in each structure?

Q2: Which of these models is nested within another?

# Choice of Covariance Structure

To choose between the covariance structures that are nested, we can use a likelihood ratio test. Since each model has the SAME mean model, we can fit the models with REML and still perform this test.

Note: if the mean models are different, we must use standard ML estimation for a LRT. But since the REML penalty term is the same for models with the same mean model, REML is appropriate here.

# Covariance Structure Hypothesis Tests

Formally, we can test two hypotheses (due to nested relationship):

- 1  $H_0$ : CS is adequate vs.  $H_A$ : UN is necessary
- 2  $H_0$ : AR-1 is adequate vs.  $H_A$ : UN is necessary

Note:

- Our null hypothesis is always that the covariance structure with fewer parameters is adequate
- We cannot formally test between CS and AR-1 as they have the same number of parameters (2) to model the covariance and are not nested
- To choose between CS and AR-1, we could look at their AIC or likelihoods (since they have the same number of parameters).

# CS vs UN

First, we fit the model with an unstructured covariance.

```
/* Unstructured Covariance, Categorical Time, REML */  
proc mixed data=wtloss2 noclprint;  
  class id program monthcat t;  
  model weight=monthcat program*monthcat/ chisq s;  
  repeated t / type=un subject=id r rcorr;  
  title 'Unstructured Covariance Structure, month categorical, REML';  
run;
```

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	504.74	393.90	414.84	380.10
2	393.90	435.14	399.76	360.45
3	414.84	399.76	534.98	386.40
4	380.10	360.45	386.40	458.19

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8405	0.7983	0.7904
2	0.8405	1.0000	0.8285	0.8073
3	0.7983	0.8285	1.0000	0.7805
4	0.7904	0.8073	0.7805	1.0000

# CS vs UN

Then, we fit the model with a compound-symmetry covariance structure.

```
/* compound symmetry, categorical time, REML */  
proc mixed data=wtloss2 noclprint;  
  class id program monthcat t;  
  model weight=monthcat program*monthcat/ chisq s;  
  repeated t/ type=cs subject=id r rcorr;  
  title 'Compound Symmetry Covariance Structure, month categorical, REML';  
run;
```

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	483.26	389.24	389.24	389.24
2	389.24	483.26	389.24	389.24
3	389.24	389.24	483.26	389.24
4	389.24	389.24	389.24	483.26

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8054	0.8054	0.8054
2	0.8054	1.0000	0.8054	0.8054
3	0.8054	0.8054	1.0000	0.8054
4	0.8054	0.8054	0.8054	1.0000



## CS vs UN

And we perform the LRT:

CS vs. UN		
Structure	-2 REML Log-Likelihood	Number of Cov. Parameters
Compound Symmetry	1905.8	2
Unstructured	1900.9	10
Difference	4.9	8

LRT yields  $G^2 = 4.9$  ( $p = 0.768$ ), so we cannot reject the null hypothesis at  $\alpha = 0.05$  and conclude that the assumption of compound symmetry covariance structure is adequate for the data. Here, we are “happy” to not reject the null, as we obtain a more parsimonious (and efficient) model.

# AR-1 vs UN

For the second hypothesis test, we fit the model with an Autoregressive(1) covariance structure.

```
/* AR-1, categorical time, REML */  
proc mixed data=wtloss2 noclprint;  
  class id program monthcat t;  
  model weight=monthcat program program*monthcat/ chisq s;  
  repeated t/ type=AR(1) subject=id r rcorr;  
  title 'AR-1 Covariance Structure, month categorical, REML';  
run;
```

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	482.39	391.92	318.41	258.70
2	391.92	482.39	391.92	318.41
3	318.41	391.92	482.39	391.92
4	258.70	318.41	391.92	482.39

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8124	0.6601	0.5363
2	0.8124	1.0000	0.8124	0.6601
3	0.6601	0.8124	1.0000	0.8124
4	0.5363	0.6601	0.8124	1.0000

# AR-1 vs UN

And now we can perform the LRT test:

AR-1 vs. UN		
Structure	-2 REML Log-Likelihood	Number of Cov. Parameters
Autoregressive-1	1931.2	2
Unstructured	1900.9	10
Difference	30.3	8

LRT yields  $G^2 = 30.3$  with 8 df ( $p = 0.0002$ ), so we reject the null hypothesis at  $\alpha = 0.05$  and conclude that the assumption of autoregressive covariance structure is not adequate when compared to unstructured.

# Choice of Covariance Structure

Structure	-2 REML Log-Likelihood	number of parameters	AIC
Compound Symmetry	1905.8	2	1909.8
Autoregressive	1931.2	2	1935.2
Unstructured	1900.9	10	1920.9

- $AIC = -2 * (\text{log-likelihood}) + 2 * (\text{number of parameters})$ , where the most adequate model is the one with the smallest value for AIC
- AIC is a measure of the predictive ability of the model that penalizes for model complexity.
- CS is a good choice because our hypothesis testing shows it is adequate compared to UN
- CS may be more efficient since it only uses 2 parameter estimates
- CS also has a larger likelihood than AR-1.
- We will use a compound symmetry (CS) covariance structure for the remainder of the lab.

# Summary of Part I

- We looked at our data, profile plots
- To chose a covariance structure, we had to temporarily assume a maximal mean model, so we used the saturated model as our most general model for the mean
- We considered 3 different covariance structures: unstructured vs. two more parsimonious models
- We used LRTs and AIC to choose a covariance structure
- We chose compound symmetry (CS) covariance structure for this data

# Analysis of Profiles and Contrasts

Analysis of response profiles:

- Treat group and time categorically, using indicators for each group, time combination
- Ignores time-ordering of repeated measures in longitudinal study
- May have low power to detect group differences in mean response over time

## Question 1: Contrasts

Q1. Test the null hypothesis that the pattern of weight over time is identical (coincides) for both weight loss programs.

Which effects are involved in this the null hypothesis? Recall common null hypotheses for response profiles are:

- (a) no group  $\times$  time interaction effect
- (b) no time effect
- (c) no group effect

## Question 1: Contrasts

Q1. Test the null hypothesis that the pattern of weight over time is identical (coincides) for both weight loss programs.

Which effects are involved in this the null hypothesis? Recall common null hypotheses for response profiles are:

- (a) no group  $\times$  time interaction effect
- (b) no time effect
- (c) no group effect



## Question 1: Contrasts

We can use a LRT to test that the mean profiles coincide.

What is the full model?

What is the reduced model?

## Question 1: SAS Code for LRT

```
/* LRT for just time effect  
(no group effect or group—time interaction) */  
  
/* First: fit ML, Covariance CS, Full Model */  
proc mixed data=wtloss2 noclprint method=ml;  
    class id program monthcat t;  
    model weight = monthcat program program*monthcat / s;  
    repeated t / type=cs subject=id r rcorr;  
    title 'CS Covariance Structure, Full Model, ML';  
run;  
  
/* Next: fit ML, Covariance CS, Reduced Model */  
proc mixed data=wtloss2 noclprint method=ml;  
    class id program monthcat t;  
    model weight = monthcat / s;  
    repeated t / type=cs subject=id r rcorr;  
    title 'CS Covariance Structure, Reduced Model, ML';  
run;
```

## Question 1: SAS Output for LRT

Testing that the mean profiles coincide		
Structure	-2 Log Likelihood	Number of Parameters
Time only model	2085.0	4
Saturated model	1935.3	8
Difference	149.7	4

LRT yields  $G^2 = 149.7$  with 4 df ( $p < 0.0001$ ), so we reject the null hypothesis at  $\alpha = 0.05$  and conclude that the mean profiles do not coincide.

## Question 2: Contrasts

Q2. Test the null hypothesis that the mean response profiles of the two weight loss programs are parallel.

Which effects are involved in this the null hypothesis? Recall common null hypotheses for response profiles are:

- (a) no group  $\times$  time interaction effect
- (b) no time effect
- (c) no group effect

## Question 2: Contrasts

Q2. Test the null hypothesis that the mean response profiles of the two weight loss programs are parallel.

Which effects are involved in this the null hypothesis? Recall common null hypotheses for response profiles are:

- (a) no group  $\times$  time interaction effect
- (b) no time effect
- (c) no group effect

## Question 2: Contrasts

We want to test the null hypothesis that there is no group  $\times$  time interaction effect:

$$H_0 : \beta_6 = \beta_7 = \beta_8 = 0.$$

In Lab 3, we reviewed two tests that we can use for this

- 1 LRT
- 2 Multivariate Wald test

## Question 2: SAS Code for LRT

If we use the LRT, make sure to fit using ML.

```
/* LRT for no group x time interaction */  
/* First: fit ML, Covariance CS, Full Model */  
proc mixed data=wtloss2 noclprint method=ml;  
    class id program monthcat t;  
    model weight = monthcat program program*monthcat / s;  
    repeated t / type=cs subject=id r rcorr;  
    title 'CS Covariance Structure, Full Model, ML';  
run;  
/* Next: fit ML, Covariance CS, Reduced Model */  
proc mixed data=wtloss2 noclprint method=ml;  
    class id program monthcat t;  
    model weight = monthcat program / s;  
    repeated t / type=cs subject=id r rcorr;  
    title 'CS Covariance Structure, Reduced Model, ML';  
run;
```

## Question 2: SAS Output for LRT

Testing for no group x time interaction		
Structure	-2 Log Likelihood	Number of Parameters
Time, group only model	2064.3	5
Saturated model	1935.3	8
Difference	129.0	3

LRT yields  $G^2 = 129.0$  with 3 df ( $p < 0.0001$ ), so we reject the null hypothesis at  $\alpha = 0.05$  and conclude that the group x time interaction is needed.



## Question 2: SAS Code for Multivariate Wald

If we use the Multivariate Wald test, we can fit using REML in SAS and include the chisq option to get the chi-sq test statistic “for free”.

```
/* Multivariate Wald Test for no group x time interaction */  
proc mixed data=wtloss2 noclprint;  
  class id program monthcat t;  
  model weight = monthcat program program*monthcat/ chisq s;  
  repeated t/ type=cs subject=id r rcorr;  
  title 'CS Covariance, Wald Test for Interaction, REML';  
run;
```

## Question 2: SAS Output for Multivariate Wald

---

### Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
monthcat	3	174	873.62	291.21	<.0001	<.0001
program	1	58	23.89	23.89	<.0001	<.0001
program*monthcat	3	174	182.28	60.76	<.0001	<.0001

---

Multivariate Wald yields  $W^2 = 182.3$  with 3 df ( $p < 0.0001$ ), so we reject the null hypothesis at  $\alpha = 0.05$  and conclude that the group x time interaction is needed.

# Using Parametric Curves

- In many studies, the true underlying mean response process changes over time in a relatively smooth, monotonic pattern
- Fitting parsimonious models for the mean response in such cases will result in statistical tests of covariate effects with greater power than in analysis of response profiles
- These models also allow for an unbalanced design (mistimed measurements), as we can use the exact time that each measurement was taken on each subject

# Quadratic Time Trend

- We will first examine whether the means tend to change over time in a quadratic manner
- We treat month as continuous and use a model with month\*program interaction
- We use maximum likelihood (ML) with Likelihood Ratio tests to check model adequacy (not REML)
- We will test  $H_0$ : quadratic model vs.  $H_A$ : saturated model (most general mean model - these are nested)

Our Quadratic Model:  $E(Y_{ij}) = \beta_1 + \beta_2 \text{month}_{ij} + \beta_3 \text{month}_{ij}^2 + \beta_4 \text{prog}_{ij} + \beta_5 \text{month}_{ij} * \text{prog}_{ij} + \beta_6 \text{month}_{ij}^2 * \text{prog}_{ij}$

where  $\text{prog}_{ij} = 1$  if encouragement program and 0 otherwise.

# Quadratic Time Hypothesis Test: Fitting Models

First, we create a  $\text{month}^2$  variable, then fit the models.

```
/* Center the month variable at the midpoint */  
data wtloss3;  
    set wtloss2;  
    month_c = month - 4.5 ;  
run;  
/* Create quadratic month */  
data wtloss3;  
    set wtloss3;  
    month_csq = month_c*month_c;  
run;  
  
/* CS covariance, quadratic trend, ML */  
proc mixed data=wtloss3 method=ml noclprint;  
    class id program t;  
    model weight = month_c month_csq program  
                month_c*program month_csq*program/chisq s;  
    repeated t / type=cs subject=id r;  
    title 'Quadratic time trend, ML';  
run;
```

# Quadratic Time Hypothesis Test: Output/Results

Testing the Quadratic trend		
Structure	-2 Log Likelihood	Number of Parameters
Quadratic model	1935.7	6
Saturated model	1935.3	8
Difference	0.4	2

LRT yields  $G^2 = 0.4$  with 2 df ( $p = 0.819$ ), so we fail to reject the null hypothesis at  $\alpha = 0.05$  and conclude that the model with month as a quadratic effect seems to fit the data adequately.

## Linear vs. Quadratic Time

- Notice that though the quadratic model seems to fit the data adequately, the coefficients for  $\text{month}^2$  and  $\text{month}^2 * \text{program}$  are not statistically significant
- Perhaps a linear model would fit the data
- We can test this hypothesis with a LRT (using ML estimates) because the linear model is nested in the quadratic model
- $H_0$ : linear model vs.  $H_A$ : quadratic model

Our Linear Model:

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{month}_{ij} + \beta_3 \text{prog}_{ij} + \beta_4 \text{month}_{ij} * \text{prog}_{ij}$$

where  $\text{prog}_{ij} = 1$  if encouragement program and 0 otherwise.

# Linear Time: Fitting Model

```
/* CS covariance, linear trend, ML */  
proc mixed data=wtloss3 method=ml noclprint;  
  class id program t;  
  model weight = month_c program month_c*program/s;  
  repeated t/ type=cs subject=id r;  
  title 'Linear time trend, ML';  
run;
```

## Linear vs. Quadratic

Structure	-2 Log Likelihood	Number of Parameters
Linear model	1936.3	4
Quadratic model	1935.7	6
Difference	0.6	2

LRT yields  $G^2 = 0.6$  ( $p = 0.741$ ): we fail to reject the null hypothesis at  $\alpha = 0.05$  and conclude that the model with month as a linear effect fits the data adequately versus the model with month as a quadratic effect



# Our Final Model

- Our final model is the linear model:  
$$E(Y_{ij}) = \beta_1 + \beta_2 \text{month}_{ij} + \beta_3 \text{prog}_{ij} + \beta_4 \text{month}_{ij} * \text{prog}_{ij}$$
- We use a compound symmetry covariance structure
- We fit this model (using REML) to interpret all parameters
- We use our month variable starting at baseline

```
/* CS covariance , linear trend , REML */  
proc mixed data=wtloss2 noclprint;  
    class id program t;  
    model weight = month program month*program/s;  
    repeated t/ type=cs subject=id r;  
    title 'Final Model, Linear time trend , REML';  
run;
```

# Covariate Interpretation

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{month}_{ij} + \beta_3 \text{prog}_{ij} + \beta_4 \text{month}_{ij} * \text{prog}_{ij}$$

- For encouraged people:

$$\begin{aligned} E(Y_{ij}) &= \beta_1 + \beta_2 \text{month}_{ij} + \beta_3(1) + \beta_4 \text{month}_{ij} * (1) \\ &= (\beta_1 + \beta_3) + (\beta_2 + \beta_4) * \text{month}_{ij} \end{aligned}$$

- For unencouraged people:

$$\begin{aligned} E(Y_{ij}) &= \beta_1 + \beta_2 \text{month}_{ij} + \beta_3(0) + \beta_4 \text{month}_{ij} * (0) \\ &= \beta_1 + \beta_2 * \text{month}_{ij} \end{aligned}$$

- $\beta_3$  is the difference in baselines ( $\text{month}_{ij} = 0$ ) between the two groups
- $\beta_4$  is the difference in the linear effect of month between the two groups (quantifies how/if the two groups are not parallel)

# Final Model: SAS Output, Parameter Estimates

---

## Solution for Fixed Effects

Effect	program	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept		245.37 = beta1	4.1143	58	59.64	<.0001
month		-3.0210 = beta2	0.2795	178	-10.81	<.0001
program	1: encourage	-2.9948 = beta3	5.4655	58	-0.55	0.5858
program	2: none	0	.	.	.	.
month*program	1: encourage	-5.0833 = beta4	0.3712	178	-13.69	<.0001
month*program	2: none	0	.	.	.	.

---

## Covariate Interpretations cont.

- Subjects in the non-encouragement program lose lbs/month, on average.
- There are no baseline differences between the two groups (as expected: treatment was randomized)
- As each month passes, people in the encouragement program lose more lbs/month, on average, than those not in the encouragement program
- As each month passes, people in the encouragement program lose a total of lbs/month, on average
- We conclude that the both programs are beneficial for weight loss, but those in the encouragement program lose significantly more weight over time than those in the standard program.

## Covariate Interpretations cont.

- Subjects in the non-encouragement program lose 3.02 lbs/month, on average.
- There are no baseline differences between the two groups (as expected: treatment was randomized)
- As each month passes, people in the encouragement program lose 5.08 more lbs/month, on average, than those not in the encouragement program
- As each month passes, people in the encouragement program lose a total of 8.1 lbs/month, on average
- We conclude that the both programs are beneficial for weight loss, but those in the encouragement program lose significantly more weight over time than those in the standard program.

# Summary

We have discussed:

- Different ways of modeling the covariance structure
- How we can choose (formally with hypothesis tests, or informally with AIC) between different covariance structures
- How we can model the mean: with time as a categorical or continuous variable
- How we can choose between nested mean models with LRT (using ML estimation)
- How we interpret models and parameter estimates