STATISTICS
TEXAS A&M UNIVERSITY

Handout 11

Repeated Measures Data Analysis

Objectives of Repeated Measures Analyses

- We will discuss basic issues in repeated measures analysis.
- We will list the advantages of the mixed model approach.
- We will produce group profile plot for a repeated measures data.

Basic Issues in Repeated Measures Data

- *Repeated measures* refers to multiple measurements on the same experimental unit (or subject).
- Measures taken on the same subject tend to be more similar than measures taken on different subjects.
- Measures made close in time on the same subject tend to be more highly correlated than measures made far apart in time.
- The analysis of repeated measures data accounts for the presence of correlation between the observations obtained on the same subject and for possible non constant variances.

Repeated Measures Data

- *Repeated measures* are made over time or space. Treatments are applied to EUs in a random fashion and measurements are taken at each of the several times or spaces.

- Independence and homogeneity are not valid assumptions.
  - Since time or space points are not randomly assigned to subjects, random errors are not reasonable to assume.
  - Homogeneity is not valid because:
    - **A:** Two measures on the same subject are likely to be closer to each other than two measures on the different subjects
    - **B:** Two measures made close in time or space on the same subject are likely to be more highly correlated than two measures made apart in time or space.

- In a split plot experiment, levels of subplot treatment are randomly assigned to subplots units within wholeplot units resulting in equal correlation between all pairs of measurements in the same wholeplot unit.

- The analysis of repeated measures data accounts for correlations between observations obtained on the same subject and also some possible heterogeneous variances among observations obtained on the same subject

Traditional Approaches

- Univariate ANOVA using PROC GLM
  - Treats repeated measures as split-plot in time
  - Assumes equal correlations regardless of distance in time. Commonly used structure is Compound Symmetry (CS)
  - Risks a higher type I error rate for fixed effect tests
  - This approach accommodates (**A**, previous slide) but not (**B**, previous slide) and called Split-plot in time.
- Multivariate ANOVA using PROC GLM
  - Requires data points are the same across subjects
  - Uses the complete case analysis. Meaning: if one repeated measure is missing from that subject, all data for that subject is eliminated
  - Avoids modeling the covariance structures and therefore the tests might be less powerful and less efficient for fixed factors. Meaning: uses UN covariance structure with MME.

Advantages of Mixed Model Analyses

- Mixed model analyses enable a flexible approach to modeling covariance structures (accounts for within-subject correlations).
- Mixed model analyses handle unbalanced data with unequally spaced time points within and across subjects.
- The generalized least squares (GLS) method is generally superior to the ordinary least squares (OLS) method, assuming an appropriate covariance structure.
- The presence of missing data does not pose the major problems for analysis that can occur with a traditional analysis, such as multivariate ANOVA approach.

Asthma Example

fev1uniExample.sas

- A pharmaceutical company wants to examine the effects of three drugs on the respiratory ability of asthma patients.
- The three drugs are labeled, **a, c**, and **p**. Drug **a** is a standard drug used to treat asthma. Drug **c** is a potential competitor that was developed by the pharmaceutical company. Drug **p** is a placebo.
- Each of the three drugs is randomly assigned to 24 patients.
- A total of 72 patients are included in the study.
- The assigned drug is administered to each patient. Then a standard measure of respiratory ability called **fev1** (forced exhaled volume in one second) is measured hourly for eight hours following treatment. Also, a **baseline fev1** is measured immediately before administering the drugs.

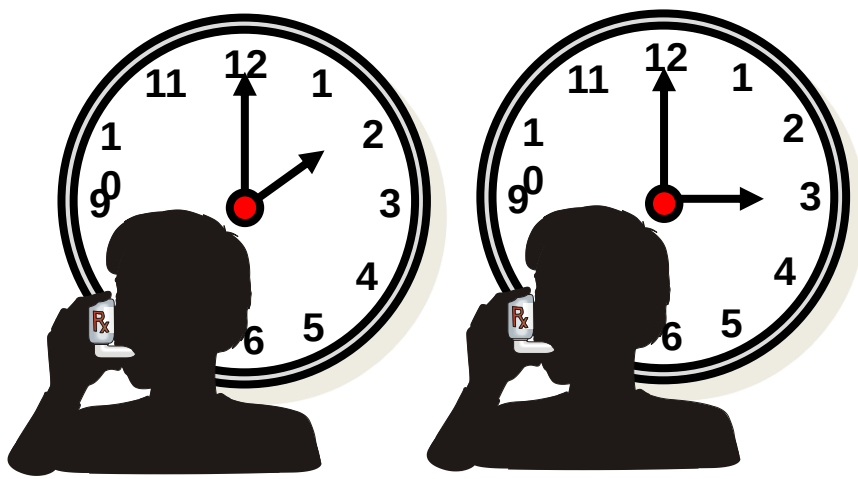
Asthma Example

DRUGS

- New (c)
- Standard (a)
- Placebo (p)

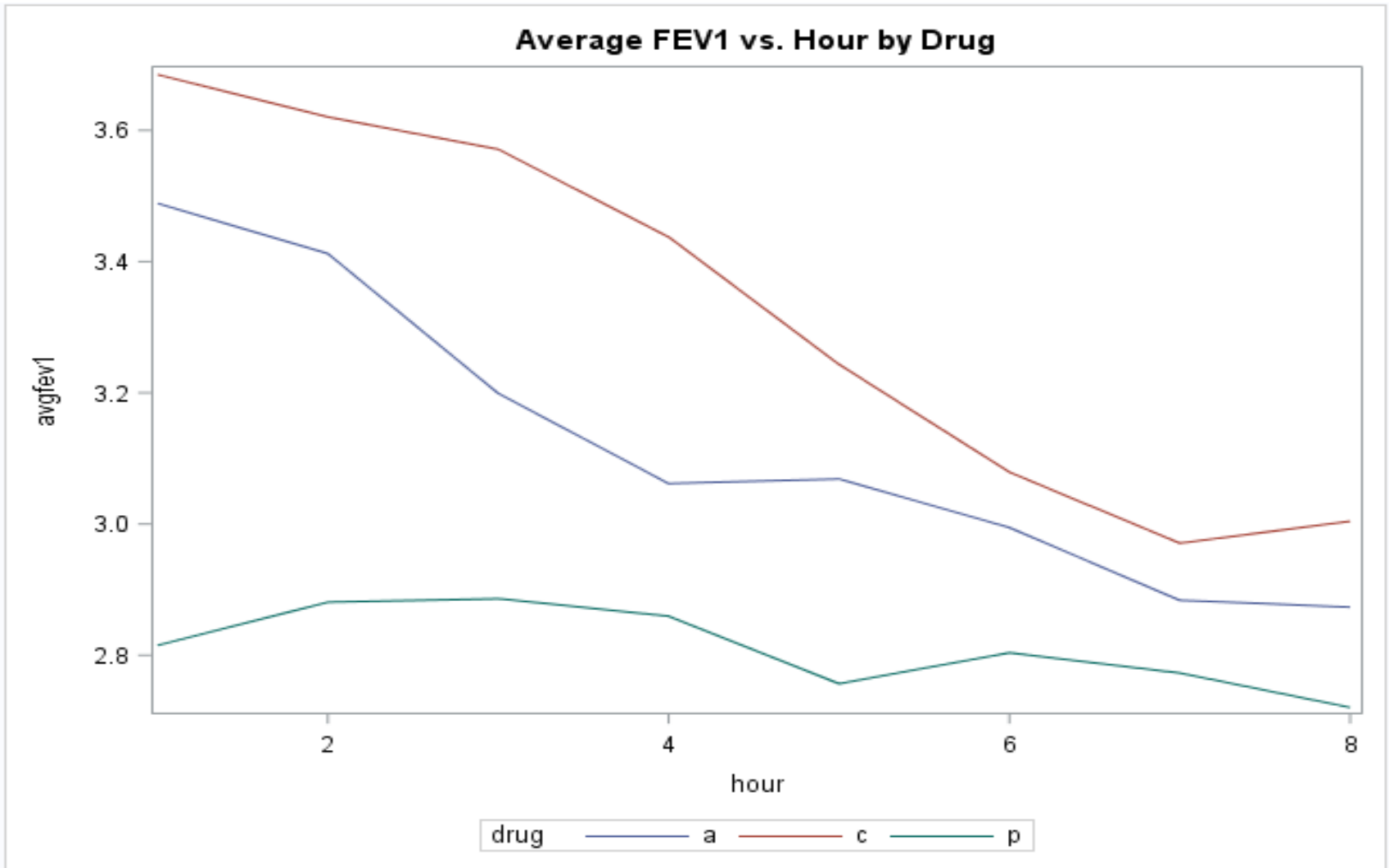
y – **fev1**
x – **basefev1**

HOURS



PATIENTS

Producing Profile Plots for the Three Drugs



Handout11_fev1uniExample.sas

Question

Which of the following is **false**?

a. To use PROC MIXED, the repeated measures must be arranged in the univariate data structure.
b. The repeated measurements provide redundant information for the subject and you should analyze one or the average of the measurements.
c. For exploratory data analysis of repeated measures that you might produce profile plots for individuals or groups to examine the trend over time.

Mixed Model Analyses of Repeated Measures Data

Objectives

- We will list the four-step procedure for mixed model analyses, especially how to choose the appropriate covariance structures for repeated measures.

- We will analyze repeated measures data using the MIXED procedure.

Four-Step Procedure for Mixed Model Analyses

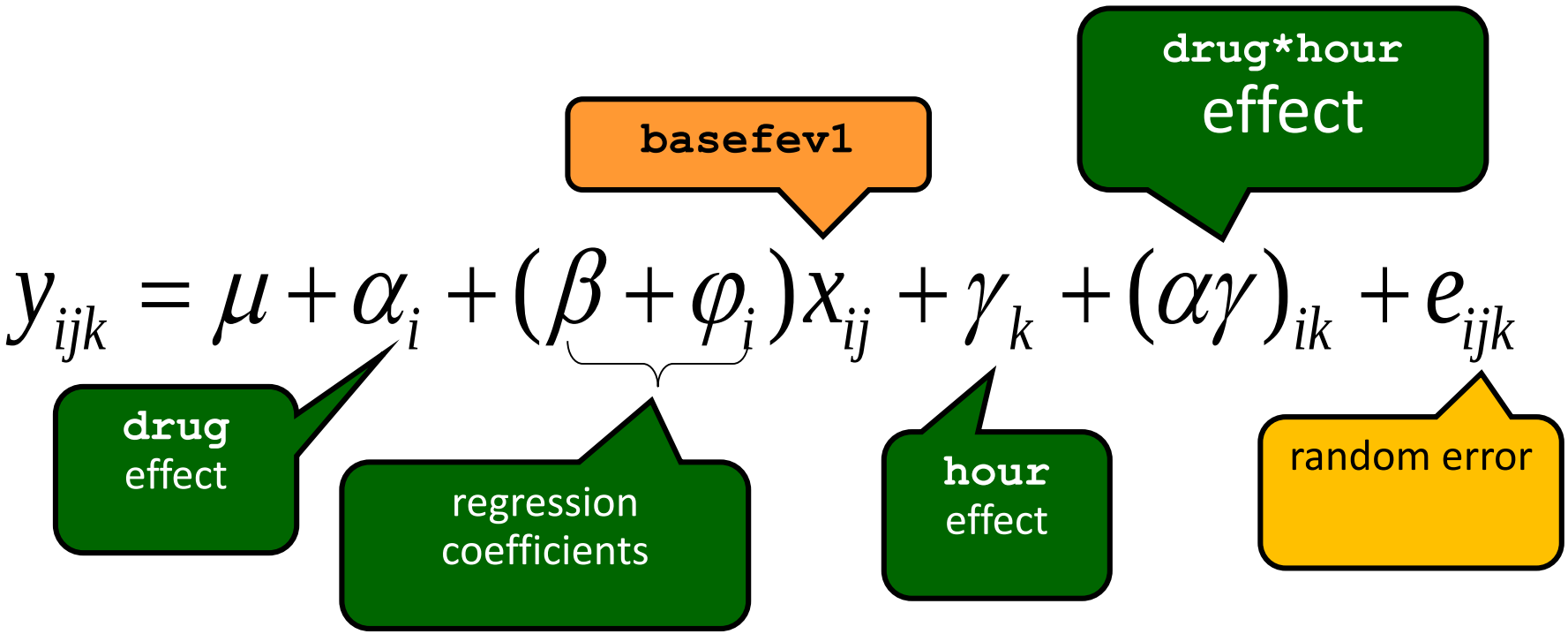
1. Model the mean structure, usually by specification of the fixed effects.

2. Specify the covariance structures for between-subject and within-subject effects.

3. Fit the mean model, accounting for the covariance structure using GLS. This step might include making the mean model more parsimonious.

4. Make statistical inferences based on the results of Step 3. This step might also include making the mean model more parsimonious.

Step 1: Model the Mean Structure



$$e_{ijk} \sim N(0, \mathbf{R})$$

y_{ijk} : the fev1 at kth hour on the jth patient assigned to ith drug

x_{ij} : the basefev1 for the jth patient taking ith drug

The mean model is E(y_{ijk})

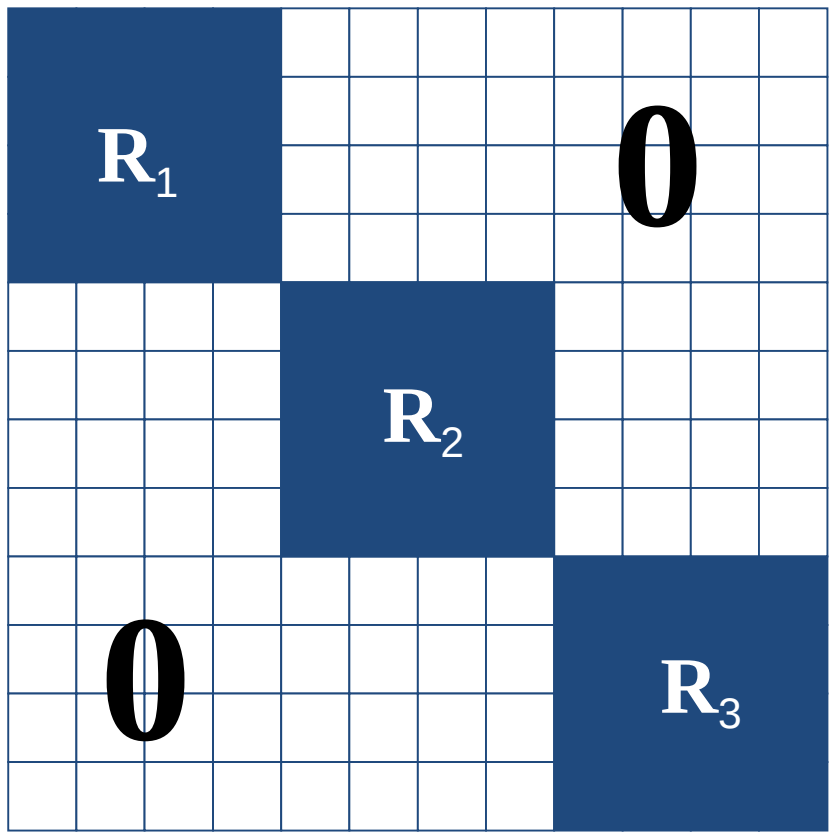
Step 2: Specify the Covariance Structure

What is the covariance structure for the within-subject measurements **R**?

- The repeated measures are assumed to be correlated within subjects and independent across subjects.
- The SUBJECT= option in the REPEATED statement in the MIXED procedure defines a block diagonal matrix **R**.
- The TYPE= option in the REPEATED statement specifies the covariance structure in each block.

What topic appeared earlier where SUBJECT= and TYPE= options in the RANDOM statement is presented?

Block Diagonal Covariance Matrix

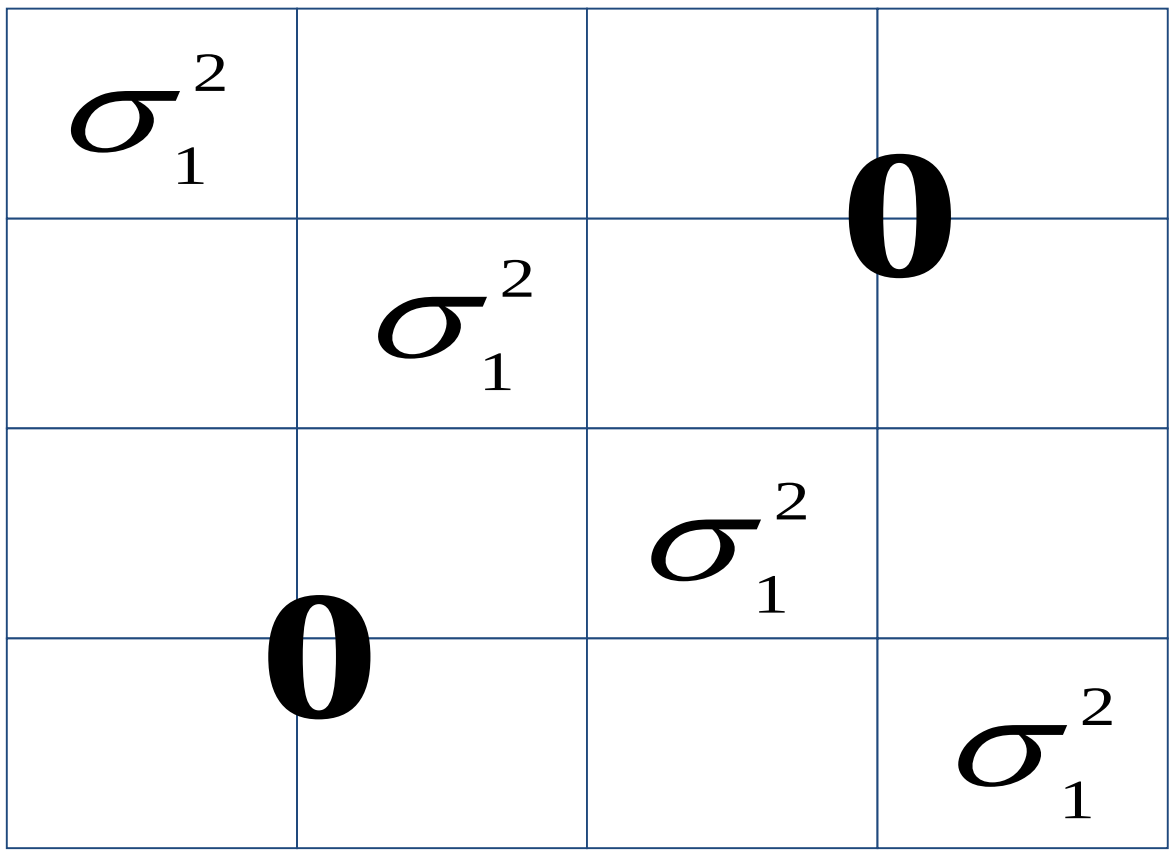


4 repeated measures on each of the 3 subjects so 12 observations and 12x12 covariance matrix.

V1, V2, V3 must have the same structure but can have different parameter estimates.

Each block (subject) can take on a variety of the covariance structures

Covariance Matrix in Each Block — Variance Component (VC) (Default)



Independent and equal variance structure which is not reasonable for repeated measures.

Covariance Matrix in Each Block — Unstructured Covariance

σ_1^2	σ_{12}	σ_{13}	σ_{14}
	σ_2^2	σ_{23}	σ_{24}
		σ_3^2	σ_{34}
			σ_4^2

Most complex structure. Multivariate ANOVA assumes and it is more complicated than is necessary

Covariance Matrix in Each Block — Compound Symmetry

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho & \rho \\ & 1.0 & \rho & \rho \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix} = \begin{bmatrix} \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 \\ & & \sigma_b^2+\sigma_e^2 & \sigma_b^2 \\ & & & \sigma_b^2+\sigma_e^2 \end{bmatrix}$$

The simplest correlation model. It assumes that the correlation is constant regardless of the distance between time points. This may be a valid assumption if repeated measures are not obtained over time. PROC MIXED provide the covariance instead of ρ.

When you specify

```
random patient(drug);
```

$$\mathbf{G} = \begin{bmatrix} \sigma_p^2 & & & \\ & \sigma_p^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} \sigma_e^2 & & & \\ & \sigma_e^2 & & \\ & & \ddots & \\ & & & \sigma_e^2 \end{bmatrix}$$

$$\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R} = \begin{bmatrix} \sigma_p^2 + \sigma_e^2 & \sigma_p^2 & \sigma_p^2 & \sigma_p^2 \\ & \sigma_p^2 + \sigma_e^2 & \sigma_p^2 & \sigma_p^2 \\ & & \ddots & \sigma_p^2 \\ & & & \sigma_p^2 + \sigma_e^2 \end{bmatrix}$$ (for one patient)

Covariance Matrix in Each Block — First-order Autoregressive AR(1)

$$\sigma^2 \begin{pmatrix} 1.0 & \rho & \rho^2 & \rho^3 \\ & 1.0 & \rho & \rho^2 \\ & & 1.0 & \rho \\ & & & 1.0 \end{pmatrix}$$

For unequally spaced time points, use spatial covariance structures instead of this.

The correlation between adjacent observations is ρ. The correlations for any pairs of observations 2 (3,4,….,d) units apart is ρ^2 (ρ^3, ρ^4,…, ρ^d).

Covariance Matrix in Each Block — Toeplitz

$$\sigma^2 \begin{bmatrix} 1.0 & \rho_1 & \rho_2 & \rho_3 \\ & 1.0 & \rho_1 & \rho_2 \\ & & 1.0 & \rho_1 \\ & & & 1.0 \end{bmatrix}$$

For unequally spaced time points, use spatial covariance structures instead of this.

Similar but more general than AR(1). The correlation is not a power function of ρ anymore but it is a number between 0 and 1. PROC MIXED produces $\sigma^2\rho_i$. RCORR Option can be used to obtain the ρ_i.

Account for Unequal Spacing?

TYPE=	Time points are		
	Equally spaced, same across subjects	Unequally spaced, same across subjects	Different across subjects
VC	n/a	n/a	n/a
UN	Yes	Yes	No
CS	Yes	Yes	Yes
AR(1)	Yes	No	No
TOEP	Yes	No	No

Covariance Matrix in Each Block — Spatial Power

$$\sigma^2 \begin{bmatrix} 1.0 & \rho^{|t_1-t_2|} & \rho^{|t_1-t_3|} & \rho^{|t_1-t_4|} \\ & 1.0 & \rho^{|t_2-t_3|} & \rho^{|t_2-t_4|} \\ & & 1.0 & \rho^{|t_3-t_4|} \\ & & & 1.0 \end{bmatrix}$$

Allows unequal spacing of the repeated measures. This is the direct Generalization of the AR(1).

Covariance Matrix in Each Block — Spatial Exponential

$$\sigma^2 \begin{bmatrix} 1.0 & e^{-\frac{d_{12}}{\rho}} & e^{-\frac{d_{13}}{\rho}} & e^{-\frac{d_{14}}{\rho}} \\ & 1.0 & e^{-\frac{d_{23}}{\rho}} & e^{-\frac{d_{24}}{\rho}} \\ & & 1.0 & e^{-\frac{d_{34}}{\rho}} \\ & & & 1.0 \end{bmatrix}$$

d_{ij} is the distance in time between time i and time j.

Account for Unequal Spacing?

TYPE=	Time points are		
	Equally spaced, same across subjects	Unequally spaced, same across subjects	Different across subjects
VC	n/a	n/a	n/a
UN	Yes	Yes	No
CS	Yes	Yes	Yes
AR(1)	Yes	No	No
TOEP	Yes	No	No
Spatial	Yes	Yes	Yes

Selecting an Appropriate Covariance Model

- You need to know your data (for example, how the time points are spaced) before choosing candidate correlation structures.

- If you ignore important correlations by using a model that is too simple, you risk increasing the Type I error rates for your fixed effects tests.

- If your model is too complex, you sacrifice power and efficiency for your fixed effects tests.

How to Select a Covariance Model

Use your subject matter knowledge and information from the data collection process and previous relevant studies.

- Independent correlation structure is not appropriate for repeated measures when time points are only hours apart.
- AR(1) or TOEP are not appropriate if time points are unequally spaced.

Consider of the following statistical tools to select appropriate covariance model:

- graphical tools to examine the patterns of correlations over time.
- information criteria to compare relative fit of various covariance structures.
- likelihood ratio tests to compare nested variance structures.

Visualizing the Correlation Structure

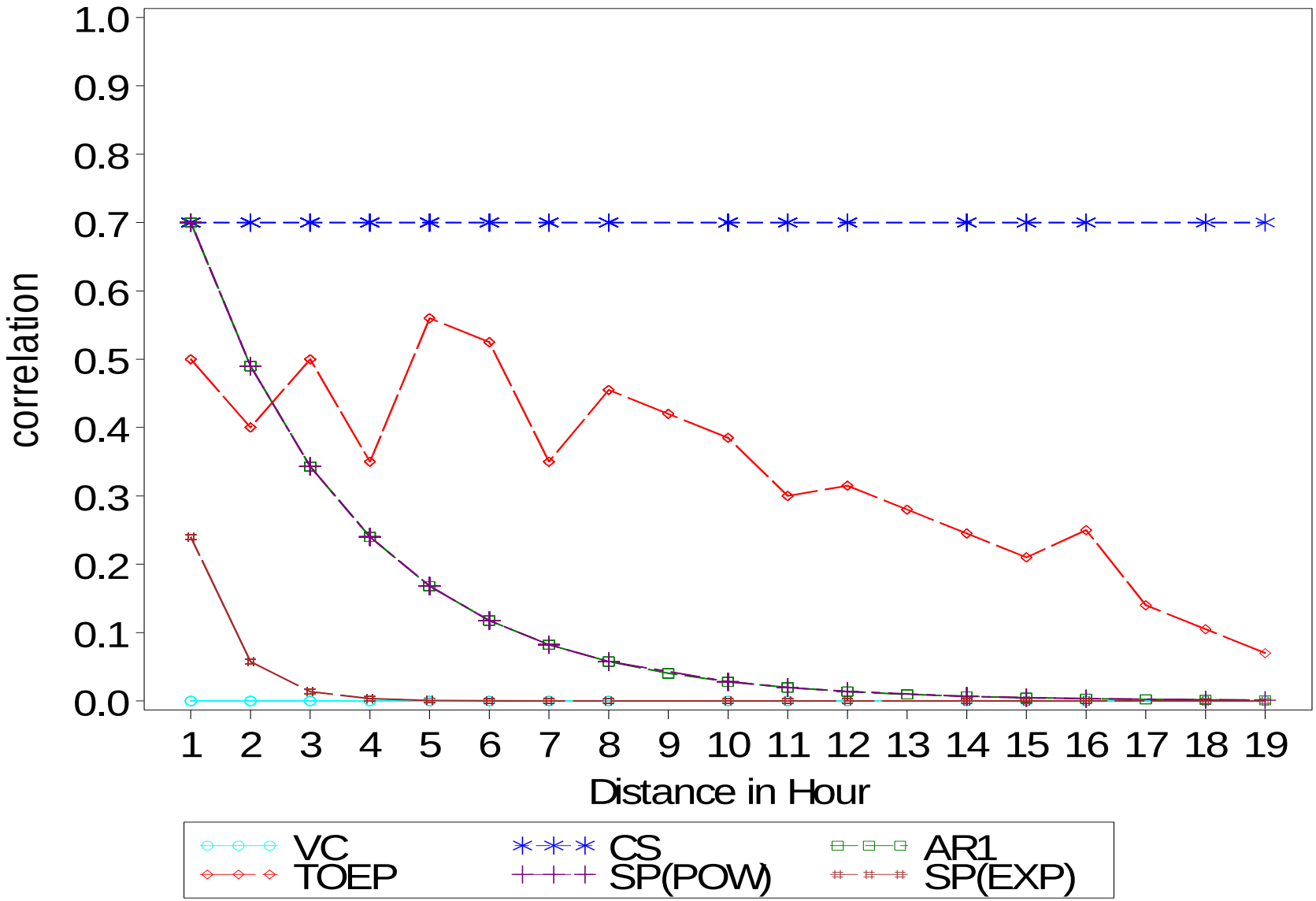
Estimated R Matrix for patient(drug) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.2321	0.2210	0.2162	0.2091	0.1789	0.1680	0.1328	0.1731
2	0.2210	0.2629	0.2372	0.2469	0.2226	0.1845	0.1602	0.1992
3	0.2162	0.2372	0.2586	0.2566	0.2224	0.1951	0.1729	0.2090
4	0.2091	0.2469	0.2566	0.3033	0.2434	0.2083	0.1948	0.2300
5	0.1789	0.2226	0.2224	0.2434	0.2879	0.2351	0.2069	0.2502
6	0.1680	0.1845	0.1951	0.2083	0.2351	0.2620	0.2187	0.2491
7	0.1328	0.1602	0.1729	0.1948	0.2069	0.2187	0.2748	0.2375
8	0.1731	0.1992	0.2090	0.2300	0.2502	0.2491	0.2375	0.3035

Estimated R Correlation Matrix for patient(drug) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.0000	0.8947	0.8828	0.7884	0.6921	0.6814	0.5260	0.6524
2	0.8947	1.0000	0.9098	0.8742	0.8092	0.7028	0.5959	0.7051
3	0.8828	0.9098	1.0000	0.9163	0.8153	0.7496	0.6487	0.7459
4	0.7884	0.8742	0.9163	1.0000	0.8239	0.7389	0.6747	0.7582
5	0.6921	0.8092	0.8153	0.8239	1.0000	0.8562	0.7358	0.8465
6	0.6814	0.7028	0.7496	0.7389	0.8562	1.0000	0.8152	0.8833
7	0.5260	0.5959	0.6487	0.6747	0.7358	0.8152	1.0000	0.8224
8	0.6524	0.7051	0.7459	0.7582	0.8465	0.8833	0.8224	1.0000

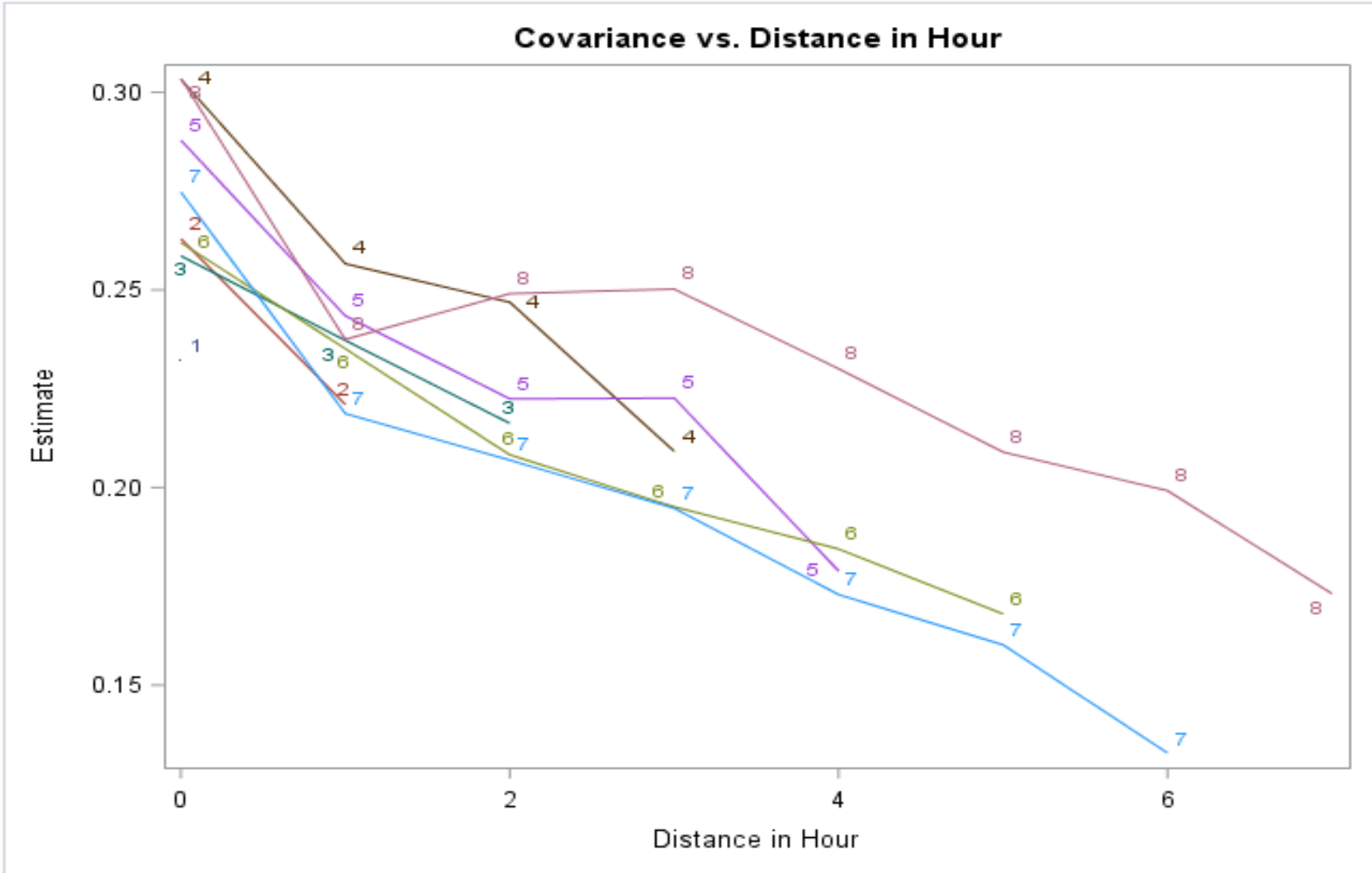
Based on the graphs of covariance / correlation patterns, what are your choices of covariance structures for this data?

Handout11_fev1uniExample.sas

Example Graphs of Correlation Structures



Covariance over Time



Handout11_fev1uniExample.sas

Observations

Covariance versus Distance in Hour

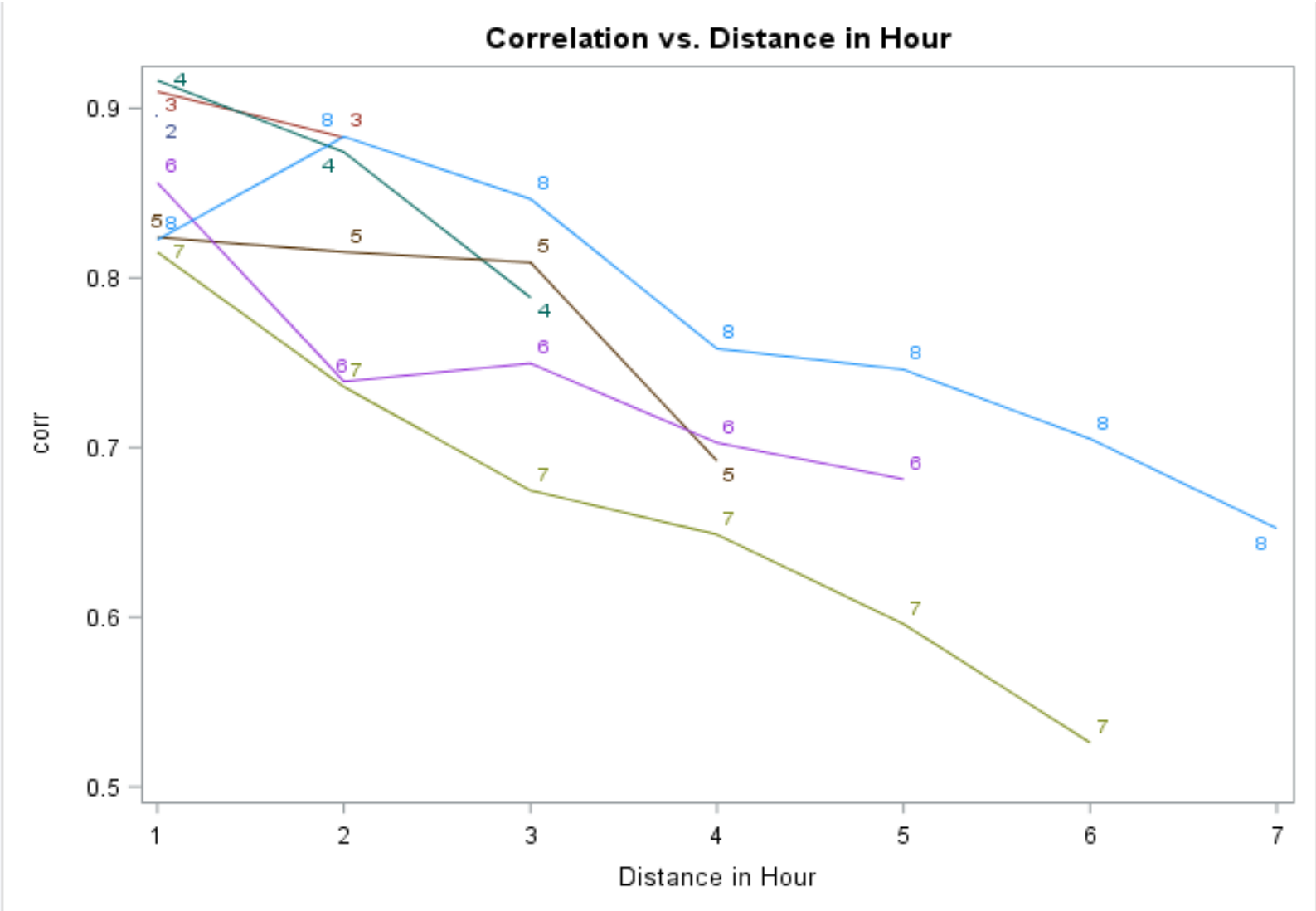
- Distance between pairs of observations increases, covariance tends to decrease.
- The pattern of decreasing covariance with distance is approximately the same for all reference times, indicated by the number on each line
  - The covariance among adjacent observations (distance=1) is consistently between 0.21 and 0.26 and does not appear to depend on the value of the ending time point
  - The covariance at distance 2 is a little smaller and does not seem to depend on the ending time point either.

- The values plotted at distance 0 are the variances among the observations at each of the 8 hours and values range from 0.23 to 0.31. Since the range is not big, there does not see to be an increasing or decreasing trend in the variances with time. Meaning: constant variance over time is reasonable.

As a result AR(1) or TOEPLITZ covariance models may be appropriate.

Handout11_fev1uniExample.sas

Correlation over Time



Handout11_fev1uniExample.sas

Observations

Correlation versus Distance in Hour

- The correlation decreases from approximately 0.81-0.91 at distance 1 to approximately 0.65 at a larger distance. A flat correlation plot over distance would suggest Compound Symmetry. This non-flat plot for the shorter distance indicates a within-subject correlation model for which a correlation is a decreasing function of distance.

- Unlike the covariance plot, you cannot use the correlation plot to diagnose possible heterogeneous variance over time because the correlation is always 1 at distance 0, which is not shown on the plot

Handout11_fev1uniExample.sas

Information Criteria

- Akaike Information Criteria (AIC)

  tends to favor more complex model

  AIC=2k-2log(maximized likelihood), k is the number of parameters

- Finite-sample corrected Akaike Information Criteria (AICc)

  reduces the bias of AIC for small samples

  AICc=AIC+$\frac{2k(k+1)}{n-k-1}$, n is the sample size

- Schwarz's Bayesian Information Criteria (BIC)

  tends to favor less complex model

  BIC=log(n)k-2log(maximized likelihood)

Information Criteria for Model Covariance Structure

Obs	Descr	AR1	Toep	UN
1	-2 Res Log Likelihood	276.1	229.0	150.4
2	AIC (Smaller is Better)	280.1	245.0	222.4
3	AICC (Smaller is Better)	280.1	245.3	227.6
4	BIC (Smaller is Better)	284.7	263.2	304.4

You should obtain consistent model comparison results using different information criteria. If that is not the case, do you need to re-specify your model.

- ❍ no
- ❍ yes

Handout11_fev1uniExample.sas

Construct a Likelihood Ratio Test

1. Fit two models with nested covariance structures.

2. Compute the difference of the –2 Res log likelihood values between the two models.

3. Compute the difference in the number of parameters estimated from the two models.

4. Compare the result from Step 2 with a χ^2 distribution with the degrees of freedom obtained from Step 3.

Likelihood Ratio Test

$\lambda(x) = \dfrac{L(\hat{\theta}_0 \mid x)}{L(\hat{\theta} \mid x)}$ where $\hat{\theta}$ is the MLE of θ and $\hat{\theta}_0$ is the MLE of θ under the restricted parameter space specified in the null model. If the null is specified as H_0: $\theta = \theta_0$ then the numerator is the likelihood function at $\theta = \theta_0$.

$$\log\{\lambda(x)\} = \log\{L(\hat{\theta}_0 \mid x)\} - \log\{L(\hat{\theta} \mid x)\}$$

$$\Rightarrow -2\log\{\lambda(x)\} = -2\log\{L(\hat{\theta}_0 \mid x)\} - [-2\log\{L(\hat{\theta} \mid x)\}]$$

$$= -2\log(reduced) - [-2\log(full)]$$

It can be shown that $-2\log\{\lambda(x)\}$ follows a chi-square distribution with r degrees of freedom, with r being the difference in the number of parameters estimated from the two models.

Likelihood Ratio Tests

Likelihood Ratio Tests
Toep vs UN, AR1 vs UN, and AR1 vs Toep

ChiToepUn	dfToepUn	pToepUn	ChiAR1Un	dfAR1Un	pAR1Un	ChiAR1Toep	dfAR1Toep	pAR1Toep
78.5850	28	.000001082	125.679	34	1.9271E-12	47.0939	6	1.7922E-8

H_0: TOEPLITZ versus H_a: UN
H_0: AR(1) versus H_a: UN
H_0: AR(1) versus H_a: TOEPLITZ

Based on the graphs, fit statistics tables from PROC MIXED, and the likelihood ratio test result, which covariance structure would you choose?

a. AR(1)
b. TOEP
c. UN
d. I have no idea
e. None of the above

Handout11_fev1uniExample.sas

Step 3: Fit the Mean Model Accounting for the Covariance Structure

- Use the appropriate covariance structure to fit the mean model using the MIXED procedure.

- You might need to reduce the mean model.

Model Development Guidelines

1. Begin with a fairly complete mean model.

   Evaluate the interactions and eliminate one at a time

2. Use REML to choose the appropriate covariance structure based on AICc or BIC.

3. Use ML to choose the appropriate mean model based on AICc or BIC.

4. Fit the final model using REML.

Fit Statistics from ML and REML

- Differences in the model fit statistics under REML reflect differences in the covariance parameter estimates.
- Differences in the model fit statistics under ML reflect differences in all the parameter estimates.
- When comparing different mean models, differences under ML are a better reflection of the importance of the fixed effects than differences under REML.

Fit the Mean Model Accounting for the Covariance Structure

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
TOEP(2)	patient(drug)	0.2314
TOEP(3)	patient(drug)	0.2191
TOEP(4)	patient(drug)	0.2099
TOEP(5)	patient(drug)	0.1937
TOEP(6)	patient(drug)	0.1858
TOEP(7)	patient(drug)	0.1724
TOEP(8)	patient(drug)	0.1612
Residual		0.2694

Fit Statistics	
-2 Res Log Likelihood	229.0
AIC (Smaller is Better)	245.0
AICC (Smaller is Better)	245.3
BIC (Smaller is Better)	263.2

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	685.97	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	66	0.41	0.6667
basefev1	1	65.9	74.53	<.0001
basefev1*drug	2	65.9	0.51	0.6026
hour	7	186	13.50	<.0001
drug*hour	14	249	3.73	<.0001

Handout11_fev1uniExample.sas

Comparison

When comparing different covariance models, you should keep the mean models the same, and use which of the following estimation methods, and then compare fit statistics from different models? Smaller values indicate better models.

a. ML
b. REML
c. ML or REML

When comparing different mean models by examining the fit statistics, you should keep the covariance model the same and use which of the following estimation methods?

a. ML
b. REML
c. ML or REML

Comparison

Fit Statistics	
-2 Log Likelihood	129.7
AIC (Smaller is Better)	199.7
AICC (Smaller is Better)	204.4
BIC (Smaller is Better)	279.4

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	710.07	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	72	0.44	0.6428
basefev1	1	71.8	81.31	<.0001
basefev1*drug	2	71.8	0.56	0.5757
hour	7	194	14.09	<.0001
drug*hour	14	261	3.89	<.0001

Fit Statistics	
-2 Log Likelihood	130.8
AIC (Smaller is Better)	196.8
AICC (Smaller is Better)	201.0
BIC (Smaller is Better)	272.0

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	717.09	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	75.5	7.72	0.0009
basefev1	1	71.8	80.15	<.0001
hour	7	194	14.05	<.0001
drug*hour	14	261	3.89	<.0001

Obs	Descr	Full	Reduced
1	-2 Log Likelihood	129.7	130.8
2	AIC (Smaller is Better)	199.7	196.8
3	AICC (Smaller is Better)	204.4	201.0
4	BIC (Smaller is Better)	279.4	272.0

Fitting the Mean Model with REML

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
TOEP(2)	patient(drug)	0.2284
TOEP(3)	patient(drug)	0.2162
TOEP(4)	patient(drug)	0.2070
TOEP(5)	patient(drug)	0.1908
TOEP(6)	patient(drug)	0.1826
TOEP(7)	patient(drug)	0.1691
TOEP(8)	patient(drug)	0.1579
Residual		0.2664

Fit Statistics	
-2 Res Log Likelihood	227.9
AIC (Smaller is Better)	243.9
AICC (Smaller is Better)	244.2
BIC (Smaller is Better)	262.1

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	689.03	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	71.3	7.30	0.0013
basefev1	1	67.8	75.69	<.0001
hour	7	186	13.45	<.0001
drug*hour	14	249	3.72	<.0001

Step 4: Make Statistical Inferences

Based on the results of Step 3, you might want to

- compare means for **`drug*hour`**
- fit a regression model treating **`hour`** as a continuous variable.

Make Statistical Inferences

Comparison of Means

- Perform all pairwise comparisons.
- Test the effects of **`drug`** at a given **`hour`**.
- Define contrasts of any other meaningful comparisons.

Comparisons using Regression

You can use a regression technique to see how the **`fev1`** means change over time for different treatments.

STATISTICS
TEXAS A&M UNIVERSITY

Comparison of Means

Effect	drug	hour	_drug	_hour	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
drug*hour	a	1	a	2	0.07667	0.05624	401	1.36	0.1736	Tukey-Kramer	0.9996
drug*hour	a	1	a	3	0.2896	0.06470	366	4.48	<.0001	Tukey-Kramer	0.0024
drug*hour	a	1	a	4	0.4271	0.07037	285	6.07	<.0001	Tukey-Kramer	<.0001
drug*hour	a	1	a	5	0.4200	0.07940	226	5.29	<.0001	Tukey-Kramer	<.0001
drug*hour	a	1	a	6	0.4942	0.08355	153	5.91	<.0001	Tukey-Kramer	<.0001
drug*hour	a	1	a	7	0.6050	0.09005	112	6.72	<.0001	Tukey-Kramer	<.0001
drug*hour	a	1	a	8	0.6154	0.09508	62.3	6.47	<.0001	Tukey-Kramer	<.0001
drug*hour	a	1	c	1	-0.2180	0.1490	109	-1.46	0.1463	Tukey-Kramer	0.9986
drug*hour	a	1	c	2	-0.1535	0.1490	109	-1.03	0.3054	Tukey-Kramer	1.0000
drug*hour	a	1	c	3	-0.1043	0.1490	109	-0.70	0.4855	Tukey-Kramer	1.0000
drug*hour	a	1	c	4	0.02945	0.1490	109	0.20	0.8437	Tukey-Kramer	1.0000
drug*hour	a	1	c	5	0.2236	0.1490	109	1.50	0.1363	Tukey-Kramer	0.9980
drug*hour	a	1	c	6	0.3878	0.1490	109	2.60	0.0105	Tukey-Kramer	0.5972
drug*hour	a	1	c	7	0.4961	0.1490	109	3.33	0.0012	Tukey-Kramer	0.1508
drug*hour	a	1	c	8	0.4628	0.1490	109	3.11	0.0024	Tukey-Kramer	0.2525
drug*hour	a	1	p	1	0.6449	0.1490	109	4.33	<.0001	Tukey-Kramer	0.0070
drug*hour	a	1	p	2	0.5791	0.1490	109	3.89	0.0002	Tukey-Kramer	0.0312
drug*hour	a	1	p	3	0.5737	0.1490	109	3.85	0.0002	Tukey-Kramer	0.0350
drug*hour	a	1	p	4	0.6003	0.1490	109	4.03	0.0001	Tukey-Kramer	0.0197
drug*hour	a	1	p	5	0.7033	0.1490	109	4.72	<.0001	Tukey-Kramer	0.0016
drug*hour	a	1	p	6	0.6562	0.1490	109	4.40	<.0001	Tukey-Kramer	0.0053
drug*hour	a	1	p	7	0.6870	0.1490	109	4.61	<.0001	Tukey-Kramer	0.0025
drug*hour	a	1	p	8	0.7395	0.1490	109	4.96	<.0001	Tukey-Kramer	0.0006
drug*hour	a	2	a	3	0.2129	0.05624	401	3.79	0.0002	Tukey-Kramer	0.0333
drug*hour	a	2	a	4	0.3504	0.06470	366	5.42	<.0001	Tukey-Kramer	<.0001
drug*hour	a	2	a	5	0.3433	0.07037	285	4.88	<.0001	Tukey-Kramer	0.0004
drug*hour	a	2	a	6	0.4175	0.07940	226	5.26	<.0001	Tukey-Kramer	<.0001
drug*hour	a	2	a	7	0.5283	0.08355	153	6.32	<.0001	Tukey-Kramer	<.0001
drug*hour	a	2	a	8	0.5388	0.09005	112	5.98	<.0001	Tukey-Kramer	<.0001
drug*hour	a	2	c	1	-0.2947	0.1490	109	-1.98	0.0505	Tukey-Kramer	0.9456

Tests of Effect Slices

Effect	hour	Num DF	Den DF	F Value	Pr > F
drug*hour	1	2	109	18.14	<.0001
drug*hour	2	2	109	12.64	<.0001
drug*hour	3	2	109	10.44	<.0001
drug*hour	4	2	109	7.72	0.0007
drug*hour	5	2	109	5.24	0.0067
drug*hour	6	2	109	1.65	0.1976
drug*hour	7	2	109	0.83	0.4404
drug*hour	8	2	109	1.73	0.1820

Estimate (hr1-hr8) for *a* versus (hr1-hr8) for *c*

drug	hour								
	1	2	3	4	5	6	7	8	
a	1							-1	0
c	-1							1	0
p									
	0							0	0

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
(hr1-hr8) for a vs (hr1-hr8) for c	-0.06542	0.1345	62.3	-0.49	0.6283

The Regression Model

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
TOEP(2)	patient(drug)	0.2295
TOEP(3)	patient(drug)	0.2153
TOEP(4)	patient(drug)	0.2059
TOEP(5)	patient(drug)	0.1904
TOEP(6)	patient(drug)	0.1829
TOEP(7)	patient(drug)	0.1706
TOEP(8)	patient(drug)	0.1595
Residual		0.2682

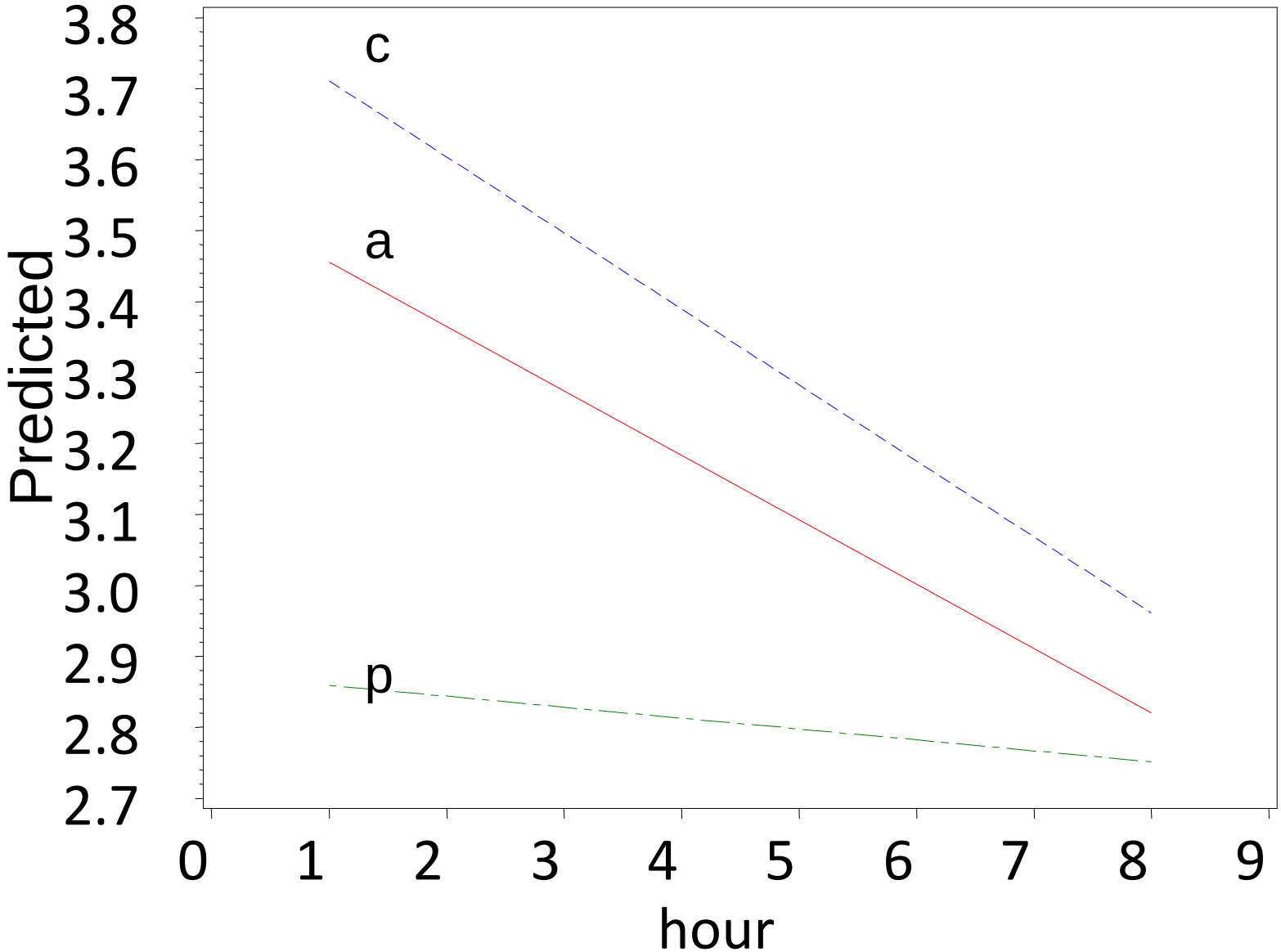
Fit Statistics	
-2 Res Log Likelihood	196.6
AIC (Smaller is Better)	212.6
AICC (Smaller is Better)	212.8
BIC (Smaller is Better)	230.8

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
7	695.60	<.0001

Solution for Fixed Effects						
Effect	drug	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.5225	0.2911	73.8	1.80	0.0767
basefev1		0.8877	0.1021	67.8	8.69	<.0001
drug	a	0.6715	0.1569	116	4.28	<.0001
drug	c	0.9438	0.1568	116	6.02	<.0001
drug	p	0	.	.	.	.
hour		-0.01535	0.01329	66	-1.16	0.2520
hour*drug	a	-0.07532	0.01879	66	-4.01	0.0002
hour*drug	c	-0.09180	0.01879	66	-4.89	<.0001
hour*drug	p	0	.	.	.	.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
basefev1	1	67.8	75.57	<.0001
drug	2	116	19.19	<.0001
hour	1	66	85.82	<.0001
hour*drug	2	66	13.57	<.0001

The Regression Model



What Else Could Have Been Done?

- Use the GROUP= option to fit a heterogeneous variance model.
- Use the RANDOM statement in addition to the REPEATED statement account for additional between-subject variations.
- Fit a random coefficient model and look at the EBLUPs.

The GROUP= Option

GROUP=**drug**

drug *a* patients

σ_a^2	σ_{1a}	σ_{2a}	σ_{3a}
	σ_a^2	σ_{1a}	σ_{2a}
		σ_a^2	σ_{1a}
			σ_a^2

drug *c* patients

σ_c^2	σ_{1c}	σ_{2c}	σ_{3c}
	σ_c^2	σ_{1c}	σ_{2c}
		σ_c^2	σ_{1c}
			σ_c^2

The GROUP= Option

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
Variance	patient(drug)	drug a	0.2838
TOEP(2)	patient(drug)	drug a	0.2457
TOEP(3)	patient(drug)	drug a	0.2330
TOEP(4)	patient(drug)	drug a	0.2263
TOEP(5)	patient(drug)	drug a	0.2303
TOEP(6)	patient(drug)	drug a	0.2376
TOEP(7)	patient(drug)	drug a	0.2276
TOEP(8)	patient(drug)	drug a	0.2316
Variance	patient(drug)	drug c	0.2778
TOEP(2)	patient(drug)	drug c	0.2430
TOEP(3)	patient(drug)	drug c	0.2299
TOEP(4)	patient(drug)	drug c	0.2134
TOEP(5)	patient(drug)	drug c	0.1859
TOEP(6)	patient(drug)	drug c	0.1619
TOEP(7)	patient(drug)	drug c	0.1479
TOEP(8)	patient(drug)	drug c	0.1555
Variance	patient(drug)	drug p	0.2399
TOEP(2)	patient(drug)	drug p	0.1959
TOEP(3)	patient(drug)	drug p	0.1822
TOEP(4)	patient(drug)	drug p	0.1767
TOEP(5)	patient(drug)	drug p	0.1531
TOEP(6)	patient(drug)	drug p	0.1481
TOEP(7)	patient(drug)	drug p	0.1159
TOEP(8)	patient(drug)	drug p	0.06273

Fit Statistics	
-2 Res Log Likelihood	195.6
AIC (Smaller is Better)	243.6
AICC (Smaller is Better)	245.9
BIC (Smaller is Better)	298.2

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
23	719.40	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	42.5	0.38	0.6864
basefev1	1	63.4	77.71	<.0001
basefev1*drug	2	42.2	0.39	0.6787
hour	7	195	12.94	<.0001
drug*hour	14	170	3.42	<.0001

RANDOM in Addition to REPEATED Statement

$$Var(y) = \mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$$

Models with the REPEATED statement only

- The covariance structure for the data (**V**) is completely determined by the covariance structure for the residuals (**R**)

Models with both RANDOM and REPEATED statements

- The covariance structure for the data (**V**) is indirectly modeled through the **G** matrix, plus the **R** matrix
- When **R** and **ZGZ'+R** result in the same structure, having both statements in the model is redundant.
- In general, be careful not to over parameterize your model with the RANDOM in addition to the REPEATED statement

RANDOM in Addition to REPEATED Statement

REPEATED / subject=patient(drug) type=cs;

V matrix follows the Compound Symmetry (CS)

Add

RANDOM patient(drug);

V matrix statement still follows the CS so having both in the model is redundant

If you change into

REPEATED / subject=patient(drug) type=AR(1);

RANDOM patient(drug);

V no longer follows the CS.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
patient(drug)		0.1874
AR(1)	patient(drug)	0.5400
Residual		0.08307

Fit Statistics	
-2 Res Log Likelihood	248.0
AIC (Smaller is Better)	254.0
AICC (Smaller is Better)	254.0
BIC (Smaller is Better)	260.8

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
drug	2	66	0.44	0.6442
basefev1	1	66	74.90	<.0001
basefev1*drug	2	66	0.59	0.5567
hour	7	395	17.05	<.0001
drug*hour	14	424	3.93	<.0001

Random Coefficient Models

Random coefficient models

- are alternative ways to indirectly model the covariance for repeated measures data
- model the variations among the subjects in terms of regression coefficients (intercept and slopes)
- have a **G** matrix (block-diagonal) and a **R** matrix ($\sigma^2\mathbf{I}_n$ by default), and the resulting **V** matrix enables the correlations within subject to change over time
- are usually models of choice when data was not obtained at fixed time points and within-subject correlations cannot be modeled adequately by a specified covariance structure.

Random Coefficient Models

Estimated G Matrix					
Row	Effect	drug	patient	Col1	Col2
1	Intercept	a	201	0.2639	-0.01434
2	hour	a	201	-0.01434	0.003648

Estimated V Matrix for patient(drug) 201 a								
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.2818	0.2282	0.2175	0.2068	0.1962	0.1855	0.1748	0.1641
2	0.2282	0.2640	0.2141	0.2071	0.2001	0.1930	0.1860	0.1789
3	0.2175	0.2141	0.2536	0.2074	0.2040	0.2006	0.1972	0.1938
4	0.2068	0.2071	0.2074	0.2505	0.2079	0.2081	0.2084	0.2086
5	0.1962	0.2001	0.2040	0.2079	0.2546	0.2157	0.2196	0.2235
6	0.1855	0.1930	0.2006	0.2081	0.2157	0.2661	0.2308	0.2383
7	0.1748	0.1860	0.1972	0.2084	0.2196	0.2308	0.2848	0.2532
8	0.1641	0.1789	0.1938	0.2086	0.2235	0.2383	0.2532	0.3109

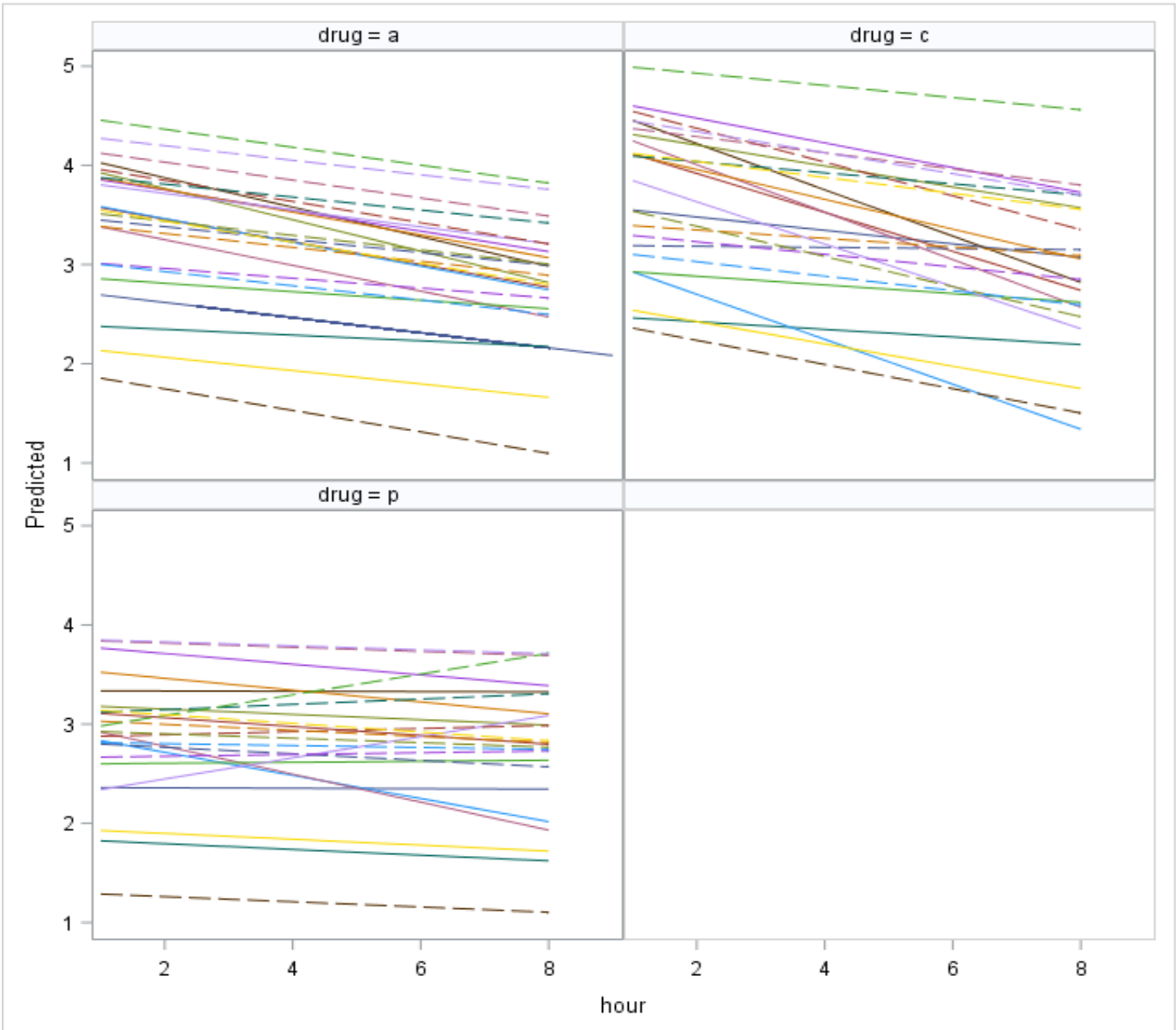
Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	patient(drug)	0.2639
UN(2,1)	patient(drug)	-0.01434
UN(2,2)	patient(drug)	0.003648
Residual		0.04285

Fit Statistics	
-2 Res Log Likelihood	219.4
AIC (Smaller is Better)	227.4
AICC (Smaller is Better)	227.5
BIC (Smaller is Better)	236.6

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
3	672.72	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
basefev1	1	68	74.96	<.0001
drug	2	68.4	18.88	<.0001
hour	1	69	85.76	<.0001
hour*drug	2	69	12.96	<.0001

Making Predictions in Random Coefficients Model



Predicted versus hour grouped by patients

Handout11_ fev1uniExample.sas

Fitting a Random Coefficient Model with a REPEATED Statement

Correlated errors

Estimated R Matrix for patient(drug) 201 a

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.05431	0.01693	0.005277	0.001645	0.000513	0.000160	0.000050	0.000016
2	0.01693	0.05431	0.01693	0.005277	0.001645	0.000513	0.000160	0.000050
3	0.005277	0.01693	0.05431	0.01693	0.005277	0.001645	0.000513	0.000160
4	0.001645	0.005277	0.01693	0.05431	0.01693	0.005277	0.001645	0.000513
5	0.000513	0.001645	0.005277	0.01693	0.05431	0.01693	0.005277	0.001645
6	0.000160	0.000513	0.001645	0.005277	0.01693	0.05431	0.01693	0.005277
7	0.000050	0.000160	0.000513	0.001645	0.005277	0.01693	0.05431	0.01693
8	0.000016	0.000050	0.000160	0.000513	0.001645	0.005277	0.01693	0.05431

Estimated V Matrix for patient(drug) 201 a

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	0.2650	0.2219	0.2045	0.1952	0.1883	0.1822	0.1764	0.1706
2	0.2219	0.2561	0.2155	0.2006	0.1937	0.1894	0.1858	0.1824
3	0.2045	0.2155	0.2521	0.2140	0.2016	0.1972	0.1954	0.1943
4	0.1952	0.2006	0.2140	0.2531	0.2175	0.2076	0.2058	0.2064
5	0.1883	0.1937	0.2016	0.2175	0.2592	0.2260	0.2187	0.2193
6	0.1822	0.1894	0.1972	0.2076	0.2260	0.2702	0.2396	0.2347
7	0.1764	0.1858	0.1954	0.2058	0.2187	0.2396	0.2862	0.2581
8	0.1706	0.1824	0.1943	0.2064	0.2193	0.2347	0.2581	0.3072

Estimated G Matrix

Row	Effect	drug	patient	Col1	Col2
1	Intercept	a	201	0.2247	-0.00822
2	hour	a	201	-0.00822	0.002497

Handout11_fev1uniExample.sas

Question

Which of the following is **true**?

a. The GROUP= option allows different covariance parameter estimates for different subgroups among subjects.
b. Random coefficient models are alternative ways of indirectly modeling the correlations among the repeated measures.
c. When using both the RANDOM and REPEATED statement to analyze a repeated measures data, you need to be careful not to over-specify the model.
d. All of the above.
e. None of the above.