Introduction to Multivariate Models: Modeling Multivariate Outcomes with Mixed Model Repeated Measures Analyses

Applied Multilevel Models for Cross-Sectional Data
Lecture 11

ICPSR Summer Workshop
University of Colorado Boulder
Covered this Section

- General linear models in matrices for...
  - One (univariate) conditionally normal outcome
  - Estimated using maximum likelihood in PROC MIXED

- Expanding linear models from univariate outcomes to multivariate outcomes:
  - Multiple variable analyses, simultaneously
  - All outcomes are then assumed to be conditionally multivariate normally distributed

- Models for covariances
Example Data

- A health researcher is interested in examining the impact of dietary habits and exercise on pulse rate

- A sample of 18 participants is collected
  - Diet factor (BETWEEN SUBJECTS):
    - Nine are vegetarians
    - Nine are omnivores
  - Exercise factor (BETWEEN SUBJECTS) with random assignment:
    - Aerobic stair climbing
    - Racquetball
    - Weight training
  - Three pulse rates (WITHIN SUBJECTS):
    - After warm-up
    - After jogging
    - After running

- We will consider the three pulse rate observations to be our first step into multivariate analyses
### Original Data: Wide Format

- The data:

<table>
<thead>
<tr>
<th>Exercise Type</th>
<th>Pulse After Warmup</th>
<th>Pulse After Jogging</th>
<th>Pulse After Running</th>
<th>Diet Type</th>
<th>personID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>112</td>
<td>166</td>
<td>215</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>111</td>
<td>166</td>
<td>225</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>89</td>
<td>132</td>
<td>189</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>95</td>
<td>134</td>
<td>186</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>66</td>
<td>109</td>
<td>150</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>69</td>
<td>119</td>
<td>177</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>125</td>
<td>177</td>
<td>241</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>85</td>
<td>117</td>
<td>186</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>97</td>
<td>137</td>
<td>185</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>93</td>
<td>151</td>
<td>217</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>77</td>
<td>122</td>
<td>178</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>78</td>
<td>119</td>
<td>173</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>81</td>
<td>134</td>
<td>205</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>88</td>
<td>133</td>
<td>180</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>88</td>
<td>157</td>
<td>224</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>58</td>
<td>99</td>
<td>131</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>85</td>
<td>132</td>
<td>186</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>78</td>
<td>110</td>
<td>164</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.5</td>
<td>264.36111</td>
</tr>
<tr>
<td>134.11111</td>
<td>315</td>
</tr>
<tr>
<td>189.55556</td>
<td>373.72222</td>
</tr>
</tbody>
</table>
Comparing Univariate and Multivariate Normal Distributions

- The univariate normal distribution:

\[
f(x_p) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]
\]

- The univariate normal, rewritten with a little algebra:

\[
f(x_p) = \frac{1}{(2\pi)^{\frac{1}{2}}|\sigma^2|^{\frac{1}{2}}} \exp \left[ -\frac{(x - \mu)(\sigma^2)^{-1}(x - \mu)}{2} \right]
\]

- The multivariate normal distribution

\[
f(x_p) = \frac{1}{V} \frac{1}{(2\pi)^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}}} \exp \left[ -\frac{(x_p^T - \mu)^T \Sigma^{-1} (x_p^T - \mu)}{2} \right]
\]

- When \( V = 1 \) (one variable), the MVN is a univariate normal distribution
Multivariate Normal Notation

- Standard notation for the multivariate normal distribution of \( v \) variables is \( N_v(\mu, \Sigma) \)
  - Our example would use a trivariate normal: \( N_3(\mu, \Sigma) \)

- **In data:**
  - The multivariate normal distribution serves as the basis for most every statistical technique commonly used in the social and educational sciences
    - General linear models (ANOVA, regression, MANOVA)
    - General linear mixed models (HLM/multilevel models)
    - Factor and structural equation models (EFA, CFA, SEM, path models)
    - Multiple imputation for missing data

  - Simply put, the world of commonly used statistics revolves around the multivariate normal distribution
    - Understanding it is the key to understanding many statistical methods including longitudinal analyses
Bivariate Normal Plot #1

\[ \mu = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \Sigma = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]
Bivariate Normal Plot #2 (Multivariate Normal)

$$\mu = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 \end{bmatrix} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$$

Density Surface (3D)
Density Surface (2D): Contour Plot
LINEAR MODELS WITH MATRICES
General Linear Models in Matrices

- Matrix expression of the GLM is important in that many descriptions of multivariate statistical models use matrix form
  - This is the starting point for learning the “language of multivariate”

- In this section, we will use an empty model with a single outcome
  - Pulse 3: Pulse after running
Linear Models with Matrices

- The basic linear model for observation \( p \) (of \( N \)), as modeled by \( k \) predictor variables (some of which may be interactions):
  \[
y_p = \beta_0 + \beta_1 X_{p1} + \cdots + \beta_k X_{pk} + e_p
\]

- The equation above, for a single univariate outcome \( y_p \) can be expressed more compactly by a set of matrices:
  \[
y_p = x_p \beta + e_p
\]

- \( y_p \) is of size \((1 \times 1)\) – a scalar (univariate/single outcome)
- \( x_p \) is of size \((1 \times (1 + k))\) – the 1 before the + \( k \) is for the intercept
- \( \beta \) is of size \(((1+k) \times 1)\)
- \( e_p \) is of size \((N \times 1)\) – one outcome means one error per person \( p \)
Unpacking the Equation

For any person $p$:

$$y_p = \beta_0 + \beta_1 x_{p1} + \cdots + \beta_k x_{pk} + e_p$$
Assumed Distributions

• The conditional distribution of $y_p$ has a normal distribution:
  - Mean is the predicted value of $y_p$ (conditional mean)
  - Error variance is the variance of $y_p$ (conditional variance)
  \[ f(y_p|x_p) \sim N_1(x_p\beta, \sigma_e^2) \]

• Because we have only one dependent variable we have a univariate normal distribution
  - Mean is determined by (model for the mean):
    • Independent variables
    • Linear model coefficients in $\beta$
  - Variance is determined only by $\sigma_e^2$ (model for the variance)

• This is why checking only the dependent variable for normality isn’t a good idea
  - Conditional distribution of $Y$ given $X$ is normal
  - No assumptions about $X$
The Normal Distribution as a Likelihood Function

- How ML estimation works with conditionally normal outcomes in GLMs is that each person contributes a portion to the total sample log likelihood:

- First, we find the (not-log) likelihood of a single observation
  \[
  L(\sigma_e^2) = \frac{1}{(2\pi)^{1/2} |\sigma_e^2|^{1/2}} \exp \left( -\frac{1}{2} (y_p - \hat{y}_p)(\sigma_e^2)^{-1}(y_p - \hat{y}_p) \right)
  \]

- From that we get the log-likelihood for that same single observation
  \[
  \log L(\sigma_e^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log|\sigma_e^2| - \frac{1}{2} (y_p - \hat{y}_p)(\sigma_e^2)^{-1}(y_p - \hat{y}_p)
  \]

- \( \hat{y}_p = x_p \beta \), is the **conditional mean** of \( y_p \) (model for the means)
- \( \sigma_e^2 \) is the error variance (or the residual variance), the conditional variance of \( y_p \) (the model for the variances)
How PROC MIXED Finds Estimates

- For a given value of $\sigma_e^2$, there is an equation that provides the fixed effects (model for the means) in $\beta$

$$\beta = (X^T(\sigma_e^2)^{-1}X)^{-1}X^T(\sigma_e^2)^{-1}y$$

- $X$ is a matrix for all $N$ people with all $k$ predictors (size $N \times k$)
- $y$ is a column vector with all persons outcomes (size $N \times 1$)
- $\sigma_e^2$ is the value of the error variance that is currently being evaluated

- For each iteration, PROC MIXED
  1. Finds $\sigma_e^2$, then uses it to find $\beta$
  2. Then uses $\beta$ to find $\hat{y}_p$ for all people
  3. Then evaluates the log likelihood
Empty Model in SAS For Pulse After Running

• Syntax:

```sas
*GLM with ML for one DV (can use wide data as only one DV is present):;
PROC MIXED DATA=WORK.DIETWIDE METHOD=ML COVTEST NOPROFILE ITDETAILS IC NAMELEN=50;
MODEL PULSE3 = /S;
REPEATED / R RCORR;
RUN;
```

- Although we only have one outcome, the REPEATED line is used to demonstrate how SAS handles the error variances
- We will shortly use the REPEATED line when we have multiple outcomes

• Output:

<table>
<thead>
<tr>
<th>CovP1: $\sigma^2_e$ for that iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Iteration History</strong></td>
</tr>
<tr>
<td>CovP1</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>1.0000</td>
</tr>
<tr>
<td>727.25</td>
</tr>
</tbody>
</table>

Convergence criteria met.
More SAS Output

- Parsing the relevant SAS output gives us:

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Residual</td>
</tr>
</tbody>
</table>

- As this is the empty model, this is equal to the variance of Y

<table>
<thead>
<tr>
<th>Solution for Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
</tbody>
</table>
Putting Output Into Matrices and Likelihoods

- For this analysis, the matrix of all predictors for all people:
  \[ X = 1_{(18 \times 1)} \]

- So, the intercept came from:
  \[
  \beta = (X^T (\sigma_e^2)^{-1}X)^{-1}X^T (\sigma_e^2)^{-1}y
  = \left(1^T_{(1 \times 18)}(\sigma_e^2)^{-1}1_{(18 \times 1)}\right)^{-1}1^T_{(1 \times 18)}(\sigma_e^2)^{-1}y_{(18 \times 1)}
  = \left(\frac{N}{\sigma_e^2}\right)^{-1} \sum_{p=1}^{18} \frac{y_p}{\sigma_e^2} = \frac{\sigma_e^2}{N} \left(\frac{1}{\sigma_e^2}\right) \sum_{p=1}^{18} y_p = \bar{y}
  \]

- And...the log likelihood for a person would be:
  \[
  \log L(\sigma_e^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log|\sigma_e^2| - \frac{1}{2} (y_p - \hat{y}_p)(\sigma_e^2)^{-1}(y_p - \hat{y}_p)
  = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(727.25) - \frac{1}{2} (y_p - 189.56)(727.25)^{-1}(y_p - 189.56)
  \]
MULTIVARIATE MODELS
From Univariate to Multivariate

• The first set of slides covered how linear models work when we have a conditional univariate normal outcome
  ➢ In the study, however, there were three outcomes

• We wish to model all three outcomes simultaneously
  ➢ Simultaneous modeling allows for:
    • Determining differences across outcomes in addition to differences within outcomes (i.e., as created by predictors)
    • Providing a mechanism to model and simultaneously test:
      – Within subjects factors (pulse rate across intensity levels)
      – Between subjects factors (diet, exercise)
      – Interactions of within and between subjects factors

• Our mechanism for studying the multivariate relationships will be to treat all three outcomes as being part of a (eventually conditional) multivariate normal distribution
Multivariate Setup for Data: Stacked (Long) Format

<table>
<thead>
<tr>
<th>personID</th>
<th>dEXERCISE_ASC</th>
<th>dEXERCISE_R</th>
<th>dEXERCISE_WT</th>
<th>dDIET_M</th>
<th>dDIET_V</th>
<th>intensity</th>
<th>pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Warm Up</td>
<td>112</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>105</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Running</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Pulse: Jogging</td>
<td>95</td>
</tr>
</tbody>
</table>

**Intensity**: Variable that denotes which pulse observation is given on that row of data.

*CONVERTING DATA TO STACKED FORM FOR PROC MIXED:*

```
DATA WORK.dietstack;
  SET WORK.dietwide;

  FORMAT intensity intensities.; *ADDING A FORMAT STATEMENT FOR INTENSITY VARIABLE;*

  *FIRST OUTCOME: PULSE 1 (AFTER WARM UP);*
  pulse = pulse1;
  intensity = 1;
  dINTENSITY_S = 1; dINTENSITY_J = 0; dINTENSITY_R = 0; *DUMMY CODED VARIABLES FOR ANALYSIS;
  OUTPUT; *OUTPUT MAKES THE LINE OF DATA GET WRITTEN TO THE NEW DATA SET;

  *SECOND OUTCOME: PULSE 2 (AFTER JOGGING);*
  pulse = pulse2;
  intensity = 2;
  dINTENSITY_S = 0; dINTENSITY_J = 1; dINTENSITY_R = 0;
  OUTPUT;

  *THIRD OUTCOME: PULSE 3 (AFTER RUNNING);*
  pulse = pulse3;
  intensity = 3;
  dINTENSITY_S = 0; dINTENSITY_J = 0; dINTENSITY_R = 1;
  OUTPUT;
```

RUN;
Why Stacked Data?

• Stacked data seem a bit counter-intuitive if you are used to repeated measures types of experiments
  ➢ Most repeated measures analysis programs take wide-format data

• In short, stacked data allow for a more concise method of matching IVs to DVs, making it easy to:
  ➢ Specify if some IVs are different across observations (important in longitudinal research)

  ➢ Keep more data in a maximum likelihood-based analysis if one or more outcomes are missing (see lecture on missing data later in October)
    • HINT: the MVN for a person uses a smaller covariance matrix
    • Use the rows you observe
Multivariate Empty Model

• What a multivariate empty model will give us is very similar to the univariate empty model:
  ➢ The mean for each variable (we can think of this as a mean vector)
    • Three means in our analysis – this will equal our mean vector for this analysis
    • Model for the means now is for a mean vector
  ➢ An estimate of the variance for each variable
    • Three variances in our analysis
  ➢ An estimate of the covariance for each pair variables
    • Three covariances in our analysis – all of these will be equal to our covariance matrix for this analysis
    • The model for the variance is now a model for the covariance matrix

• The trick, in syntax, is to figure out how to get access to all parts

• The trick, in multivariate modeling, is to get an appropriate* model for the covariance matrix so you can believe your model for the means
  ➢ *Appropriate = best fitting and most parsimonious
SAS Syntax for Multivariate Empty Model

TITLE "EMPTY MULTIVARIATE MODEL, VC ERROR: (PREDICTORS ARE INDICATORS OF WHICH VARIABLE)";
PROC MIXED DATA=WORK.dietstack METHOD=ML COVTEST NOPROFILE ITDETAILS IC NAMELEN=50;
MODEL pulse = dINTENSITY_W dINTENSITY_R / S DDFM=KENWARDROGER;
REPEATED / SUBJECT=personID TYPE=VC R RCORR;
RUN;

- Pulse still shows up to the left of the equals sign (because pulse is one column now)

- To the right of the equals sign we now need predictors that will allow us to get a mean estimate for each variable
  - Without predictors here, we would only get one term (an intercept)
    \[ y_p = \beta_0 + \beta_1 dWARMUP_p + \beta_2 dRUNNING_p + e_p \]

- With your knowledge of linear models, what does:
  \[ \beta_0 = ? \] Predicted pulse for jogging (W=0, R=0)
  \[ \beta_1 = ? \] Difference in pulse between jogging and warmup
  \[ \beta_2 = ? \] Difference in pulse between jogging and running
SAS Syntax for Multivariate Empty Model

TITLE "EMPTY MULTIVARIATE MODEL, VC ERROR: (PREDICTORS ARE INDICATORS OF WHICH VARIABLE)";
PROC MIXED DATA=WORK.dietstack METHOD=ML COVTEST NOPROFILE ITDETAILS IC NAMELEN=50;
MODEL pulse = dINTENSITY_W dINTENSITY_R / S DDFM=KENWARDROGER;
REPEATED / SUBJECT=personID TYPE=VC R RCORR;
RUN;

- SUBJECT = personID: Indicates that observations with the same personID are all from the same subject/person (and as such get put into a single multivariate normal distribution)

- TYPE = VC: The type line gives access to the model for the covariance matrix
  - VC stands for variance components
    - The default, estimates $\sigma^2 I$ (or one single residual variance that is shared/the same for each outcome)
    - Does not estimate any residual covariances between outcomes: assumes residuals are independent

  - Many types exist in SAS
    (http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/statug_mixed_sect020.htm#statug.mixed.repeatedstmt_type) – to be discussed shortly

- R: SAS’ notation for the residual covariance matrix (the letter prints the matrix)
- RCORR: the correlation matrix version of the R covariance matrix (the word prints the matrix)
Helpful SAS Output Information

- SUBJECTS – should equal your sample size

- Covariance parameters – number of parameters estimated for the covariance matrix (1 = our variance)

- Max Obs Per Subject – should equal your max per subject

- If any of these are off, the model is specified incorrectly in syntax.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Parameters</td>
<td>1</td>
</tr>
<tr>
<td>Columns in X</td>
<td>3</td>
</tr>
<tr>
<td>Columns in Z</td>
<td>0</td>
</tr>
<tr>
<td>Subjects</td>
<td>18</td>
</tr>
<tr>
<td>Max Obs Per Subject</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations Read</td>
<td>54</td>
</tr>
<tr>
<td>Number of Observations Used</td>
<td>54</td>
</tr>
<tr>
<td>Number of Observations Not Used</td>
<td>0</td>
</tr>
</tbody>
</table>

Lecture 11: Introduction to Multivariate Models
Multivariate Output from PROC MIXED: Fixed Effects

Solution for Fixed Effects

| Effect        | Estimate | Standard Error | DF | t Value | Pr > |t| |
|---------------|----------|----------------|----|---------|------|---|
| Intercept     | 134.11   | 5.1611         | 54 | 25.99   | <.0001|
| dINTENSITY_W  | -46.6111 | 7.2988         | 54 | -6.39   | <.0001|
| dINTENSITY_R  | 55.4444  | 7.2988         | 54 | 7.60    | <.0001|

\[ \mathbf{y}_p = \begin{bmatrix} \text{Pulse}_{1p} \\ \text{Pulse}_{2p} \\ \text{Pulse}_{3p} \end{bmatrix}; \mathbf{X}_p = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}; \mathbf{\beta} = \begin{bmatrix} 134.11 \\ -46.61 \\ 55.44 \end{bmatrix} \]

- Therefore, for any given observation, the predicted (mean vector):

\[ \hat{\mathbf{y}}_p = \mathbf{X}_p \mathbf{\beta} = \begin{bmatrix} 134.11 - 46.61 \\ 134.11 \\ 134.11 + 55.44 \end{bmatrix} = \begin{bmatrix} 87.50 \\ 134.11 \\ 189.56 \end{bmatrix} \]

- These are the means for each outcome from the means vector
Multivariate Output from PROC MIXED: Variances

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z Value</th>
<th>Pr &gt; Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>personID</td>
<td>479.46</td>
<td>92.2717</td>
<td>5.20</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Estimated R Matrix for Subject 1

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>479.46</td>
<td>479.46</td>
<td>479.46</td>
</tr>
<tr>
<td>2</td>
<td>479.46</td>
<td>479.46</td>
<td>479.46</td>
</tr>
</tbody>
</table>

Estimated R Correlation Matrix for Subject 1

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

\[
\mathbf{R} = \sigma^2 \mathbf{I}_{(3 \times 3)} = 479.46 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 479.46 & 0 & 0 \\ 0 & 479.46 & 0 \\ 0 & 0 & 479.46 \end{bmatrix}
\]

Note: look at how large the standard error is – variances are very hard to estimate (and covariances even harder)…large samples needed.
Putting Output Into Distributional Terms

- For multivariate data today we are assuming that the multivariate distribution of all outcomes is multivariate normal, conditional on the IVs (although we don’t always have to assume conditional MVN)

\[ f(y_p | X_p) \sim N_V (X_p \beta, V_p) \]

Where:
- \( y_p \) is a \( V \times 1 \) vector of outcomes for person \( p \)
- \( X_p \) is a \( V \times (k + 1) \) matrix of \( k \) predictors for person \( p \)
- \( \beta \) is a \( (k + 1) \times 1 \) vector of fixed effects
- \( X_p \beta = \hat{y}_p \) is the predicted conditional mean vector of \( y_p \)
- \( V_p \) is the residual covariance matrix (SAS notation) for person \( p \)

- Here \( V_p = R \) for all people
  - More complicated models bring about more terms into how \( V_p \) is formed
  - In 2-level multilevel models and structural equation models
    \[ V_p = Z_p G Z_p^T + R_p \]
Visualizing the Log-Likelihood for Our Example

- From our example we found: $\beta = \begin{bmatrix} 134.11 \\ -46.61 \\ 55.44 \end{bmatrix}$; $R = \begin{bmatrix} 479.46 & 0 & 0 \\ 0 & 479.46 & 0 \\ 0 & 0 & 479.46 \end{bmatrix}$; $\hat{y}_p = \begin{bmatrix} 87.50 \\ 134.11 \\ 189.56 \end{bmatrix}$

- The conditional MVN for a person is:

$$f(y_p | X_p) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(y_p - \hat{y}_p)^T R^{-1} (y_p - \hat{y}_p)}{2} \right)$$

- And...the log likelihood for a person would be:

$$\log L(R) = -\frac{3}{2} \log(2\pi) - \frac{1}{2} \log|R| - \frac{1}{2} (y_p - \hat{y}_p)^T R^{-1} (y_p - \hat{y}_p)$$

$$= -\frac{3}{2} \log(2\pi) - \frac{1}{2} \log(110,219,172)$$

$$-\frac{1}{2} \begin{bmatrix} \text{Pulse1}_p \\ \text{Pulse2}_p \\ \text{Pulse3}_p \end{bmatrix}^T \begin{bmatrix} 0.002 & 0 & 0 \\ 0 & 0.002 & 0 \\ 0 & 0 & 0.002 \end{bmatrix} \begin{bmatrix} \text{Pulse1}_p \\ \text{Pulse2}_p \\ \text{Pulse3}_p \end{bmatrix}$$

Lecture 11: Introduction to Multivariate Models
MODELS FOR (THE VARIANCES) COVARIANCE MATRICES
Modeling Covariances

• The estimated covariance matrix in our example analysis was:

\[
R = \begin{bmatrix}
479.46 & 0 & 0 \\
0 & 479.46 & 0 \\
0 & 0 & 479.46 \\
\end{bmatrix}
\]

• From the beginning, however, we found the sample covariance matrix to be:

\[
S = \begin{bmatrix}
264.36 & 315.00 & 373.72 \\
315.00 & 446.75 & 536.49 \\
373.72 & 539.49 & 727.25 \\
\end{bmatrix}
\]

• Deciding on the right model for the covariance matrix is a balance between power and model fit:
  - More parameters = (possibly) better fit + less statistical power

• NOTE: Model fit ≠ Effect size
  - We are not explaining anything by finding a good fitting model
  - Model fit is necessary, but not sufficient
A (Possibly) Better Model for the Covariance Matrix

• An UNSTRUCTURED covariance matrix is one where every term is a model parameter and is estimated:

\[
R = \begin{bmatrix}
\sigma_{e_1}^2 & \sigma_{e_1,e_2} & \sigma_{e_1,e_3} \\
\sigma_{e_1,e_2} & \sigma_{e_2}^2 & \sigma_{e_2,e_3} \\
\sigma_{e_1,e_3} & \sigma_{e_2,e_3} & \sigma_{e_3}^2 \\
\end{bmatrix}
\]

• As this is an empty model, what would you expect the estimates to be?
# The Unstructured Model Estimates: Covariance Parameters

## Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z Value</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>personID</td>
<td>264.36</td>
<td>88.1204</td>
<td>3.00</td>
<td>0.0013</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>personID</td>
<td>315.00</td>
<td>109.88</td>
<td>2.87</td>
<td>0.0041</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>personID</td>
<td>446.77</td>
<td>148.92</td>
<td>3.00</td>
<td>0.0013</td>
</tr>
<tr>
<td>UN(3,1)</td>
<td>personID</td>
<td>373.72</td>
<td>135.79</td>
<td>2.75</td>
<td>0.0059</td>
</tr>
<tr>
<td>UN(3,2)</td>
<td>personID</td>
<td>539.49</td>
<td>184.99</td>
<td>2.92</td>
<td>0.0035</td>
</tr>
<tr>
<td>UN(3,3)</td>
<td>personID</td>
<td>727.25</td>
<td>242.42</td>
<td>3.00</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

## Estimated R Matrix for Subject 1

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>264.36</td>
<td>315.00</td>
<td>373.72</td>
</tr>
<tr>
<td>2</td>
<td>315.00</td>
<td>446.77</td>
<td>539.49</td>
</tr>
<tr>
<td>3</td>
<td>373.72</td>
<td>539.49</td>
<td>727.25</td>
</tr>
</tbody>
</table>

## Estimated R Correlation Matrix for Subject 1

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>0.9166</td>
<td>0.8523</td>
</tr>
<tr>
<td>2</td>
<td>0.9166</td>
<td>1.0000</td>
<td>0.9465</td>
</tr>
<tr>
<td>3</td>
<td>0.8523</td>
<td>0.9465</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Lecture 11: Introduction to Multivariate Models
Comparing Covariances

• The unstructured model provided a new estimated $R$ covariance matrix:

\[
R_{UN} = \begin{bmatrix}
264.36 & 315.00 & 373.72 \\
315.00 & 446.75 & 536.49 \\
373.72 & 539.49 & 727.25
\end{bmatrix}
\]

• The estimated covariance matrix in our example analysis was:

\[
R_{VC} = \begin{bmatrix}
479.46 & 0 & 0 \\
0 & 479.46 & 0 \\
0 & 0 & 479.46
\end{bmatrix}
\]

• From the beginning, however, we found the sample covariance matrix to be:

\[
S = \begin{bmatrix}
264.36 & 315.00 & 373.72 \\
315.00 & 446.75 & 536.49 \\
373.72 & 539.49 & 727.25
\end{bmatrix}
\]

• So, which model is correct: VC or UN?
  
  ➢ Good news: VC is nested within UN so we can use a likelihood ratio test
Model Comparison

• We will compare the fit of the VC model to the UN model using a likelihood ratio test

\[ H_0: \mathbf{R} = \sigma_e^2 \mathbf{I} \text{ (3 fixed effects + 1 variance = 4 parameters)} \]

\[ H_A: \mathbf{R} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1,e_2} & \sigma_{e_1,e_3} \\ \sigma_{e_1,e_2} & \sigma_{e_2}^2 & \sigma_{e_2,e_3} \\ \sigma_{e_1,e_3} & \sigma_{e_2,e_3} & \sigma_{e_3}^2 \end{bmatrix} \text{ (3 fixed effects + 6 var/cov = 9 parameters)} \]

• In SAS PROC MIXED, this is done for us automatically:

<table>
<thead>
<tr>
<th>Null Model Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

• Therefore, we find that the UN model fits better.
Why the Right Covariance Matrix Matters: Standard Errors and Inferences Made from Fixed Effects

- Fixed effects from the UN model:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>134.11</td>
<td>4.9820</td>
<td>18</td>
<td>26.92</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dINTENSITY_W</td>
<td>-46.6111</td>
<td>2.1230</td>
<td>18</td>
<td>-21.96</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dINTENSITY_R</td>
<td>55.4444</td>
<td>2.2976</td>
<td>18</td>
<td>24.13</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Fixed effects from the VC model:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>134.11</td>
<td>5.1611</td>
<td>54</td>
<td>25.99</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dINTENSITY_W</td>
<td>-46.6111</td>
<td>7.2988</td>
<td>54</td>
<td>-6.39</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dINTENSITY_R</td>
<td>55.4444</td>
<td>7.2988</td>
<td>54</td>
<td>7.60</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What Happens with Differing Models for the Covariances

• The different models for the covariances generally don’t change the model for the means (the fixed effects) much
  ➢ Exceptions: unbalanced data

• The standard errors for the fixed effects are derived from the R matrix that was estimated:
  \[ V(\beta) = (X^T R^{-1} X)^{-1} \]
  ➢ Putting the wrong R matrix in the model will lead to the wrong SEs
  ➢ The wrong SEs will end up giving you inaccurate p-values
  ➢ Inaccurate p-values will lead to the wrong inferences

• Therefore, the main part of a multivariate model is to determine the appropriate model for the variances
  ➢ This is why we have Repeated Measures (one type of R matrix) and MANOVA (an unstructured R matrix)
  ➢ Unless sample size isn’t an issue, the model selected should the model that fits best with the least number of parameters

• Maximum likelihood has made many types of covariance matrices possible
MULTIPLE MODELS FOR COVARIANCE MATRICES
A Multivariate Modeling Demonstration

• To demonstrate the process of finding the best fitting/most parsimonious covariance matrix, we will estimate five models
  1. Variance Components
  2. Variance Components with Heterogeneous Variances
  3. Compound Symmetry
  4. Compound Symmetry with Heterogeneous Variances
  5. Unstructured

• The unstructured model from the previous slides will be the best one can do – but the question remains as to whether any simpler forms would be approximately correct but have fewer parameters

• The choice of a covariance matrix is typically aided by the types of outcomes:
  - Time sensitive? Auto regressive/Toeplitz
  - Same outcome after multiple trials? Unstructured/Compound symmetry
  - Are outcomes region or geography specific? Spatial models
Type 1: Variance Components

- **R matrix form:**
  \[
  R = \sigma^2 I = \begin{bmatrix}
  \sigma_e^2 & 0 & 0 \\
  0 & \sigma_e^2 & 0 \\
  0 & 0 & \sigma_e^2
  \end{bmatrix}
  \]

- **Estimated R matrix:**
  \[
  R = \begin{bmatrix}
  479.46 & 0 & 0 \\
  0 & 479.46 & 0 \\
  0 & 0 & 479.46
  \end{bmatrix}
  \]

- **Model Fit Statistics:**
  - -2 Log L = 486.6
  - Parameters = 4 (3 fixed effects + 1 variances)
Type 2: Heterogeneous Variances/Zero Covariances
TYPE = UN(1) in PROC MIXED

- **R matrix form:**
  \[
  R = \begin{bmatrix}
  \sigma_{e_1}^2 & 0 & 0 \\
  0 & \sigma_{e_2}^2 & 0 \\
  0 & 0 & \sigma_{e_3}^2
  \end{bmatrix}
  \]

- **Estimated R matrix:**
  \[
  R = \begin{bmatrix}
  264.36 & 0 & 0 \\
  0 & 446.77 & 0 \\
  0 & 0 & 727.25
  \end{bmatrix}
  \]

- **Model Fit Statistics:**
  - -2 Log L = 482.1
  - Parameters = 6 (3 fixed effects + 3 variances)

- **Model comparison:**
  - LRT compared with TYPE=VC: \(-2LL = 4.49, df = 2, p = .106\)
  - VC is preferred to this model
Type 3: Compound Symmetry [TYPE = CS in PROC MIXED]

- **R matrix form:**
  
  $$
  R = \begin{bmatrix}
  \sigma_e^2 + \sigma & \sigma & \sigma \\
  \sigma & \sigma^2 & \sigma \\
  \sigma & \sigma & \sigma_e^2 + \sigma
  \end{bmatrix}
  $$

- **Estimated R matrix:**
  
  $$
  R = \begin{bmatrix}
  479.46 & 409.41 & 409.41 \\
  409.41 & 479.46 & 409.41 \\
  409.41 & 409.41 & 479.46
  \end{bmatrix}
  $$

- **Model Fit Statistics:**
  
  - $-2 \text{ Log L} = 435.3$
  - Parameters = 5 (3 fixed effects + 1 variances + 1 covariance)

- **Model comparison:**
  
  - LRT compared with TYPE=VC: $-2LL = 51.31, df = 1, p < .0001$
  - CS is preferred to VC (so we now use CS as null model)

Btw, this is univariate repeated measures ANOVA.
Type 4: Compound Symmetry/Heterogeneous Variances

[TYPE = CSH in PROC MIXED]

• R matrix form:

\[
R = \begin{bmatrix}
\sigma_{e_1}^2 & \sigma_{e_1} \sigma_{e_2} \rho & \sigma_{e_1} \sigma_{e_3} \rho \\
\sigma_{e_1} \sigma_{e_2} \rho & \sigma_{e_2}^2 & \sigma_{e_2} \sigma_{e_3} \rho \\
\sigma_{e_1} \sigma_{e_3} \rho & \sigma_{e_2} \sigma_{e_3} \rho & \sigma_{e_3}^2
\end{bmatrix}
\]

• Estimated R matrix:

\[
R = \begin{bmatrix}
217.16 & 310.26 & 403.39 \\
310.26 & 433.06 & 509.79 \\
403.39 & 509.79 & 732.06
\end{bmatrix}
\]

• Model Fit Statistics:
  - \(-2 \log L = 415.8\)
  - Parameters = 7 (3 fixed effects + 3 variances + 1 covariance)

• Model comparison:
  - LRT compared with TYPE=CS: \(-2LL = 19.5\), \(df = 2\), \(p < .0001\)
  - CSH is preferred to CS (so we now use CSH as null model)
Type 5: Unstructured Covariance Matrix [TYPE = UN in PROC MIXED]

- **R matrix form:**
  \[
  R = \begin{bmatrix}
  \sigma_{e_1}^2 & \sigma_{e_1,e_2} & \sigma_{e_1,e_3} \\
  \sigma_{e_1,e_2} & \sigma_{e_2}^2 & \sigma_{e_2,e_3} \\
  \sigma_{e_1,e_3} & \sigma_{e_2,e_3} & \sigma_{e_3}^2 
  \end{bmatrix}
  \]

- **Estimated R matrix:**
  \[
  R = \begin{bmatrix}
  264.36 & 315.00 & 373.72 \\
  315.00 & 446.75 & 536.49 \\
  373.72 & 539.49 & 727.25 
  \end{bmatrix}
  \]

- **Model Fit Statistics:**
  - $-2 \log L = 408.1$
  - Parameters = 9 (3 fixed effects + 3 variances + 3 covariances)

- **Model comparison:**
  - LRT compared with TYPE=CSH: $-2LL = 7.7, df = 2, p < .021$
  - UN is preferred to CSH – and UN is the winner!

Btw, this is Multivariate ANOVA – or MANOVA
Wrapping Up

- This section was our first step into multivariate modeling where multiple outcomes were modeled using a conditional multivariate normal distribution