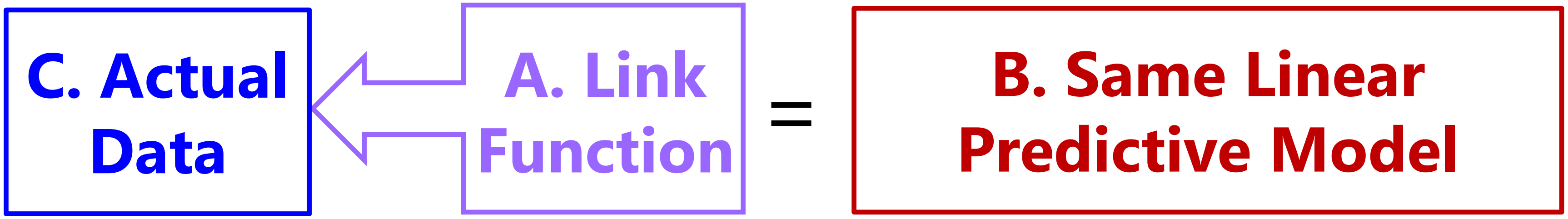
Introduction to Multivariate General Linear Models

- Topics:
    - Taxonomy of multivariate dependency: balanced or unbalanced
    - Multivariate models for balanced outcomes in univariate software
    - $\boldsymbol{R}$ matrix choices for residual variance and covariance
    - Fixed effects parameterization choices

3 Parts of Generalized Linear Models



C. Actual Data ← **A. Link Function** = **B. Same Linear Predictive Model**

A. **Link Function**: Transformation of conditional mean to keep *predicted outcomes* within the bounds of the outcome

B. **Same Linear Predictor**: How the model linearly predicts the *link-transformed* conditional mean of the outcome
   - Btw, I call this as the "**model for the means**" more generally

C. **Conditional Distribution**: How the outcome residuals could be distributed given the possible values of the outcome

- **Now we need to consider how the model needs to adapt when residuals are correlated → capture "dependency"**
  - Btw, I call this as the "**model for the variance**" more generally

Types of Multivariate Dependency

- **Dependency arises whenever multiple outcomes are collected from the same sampling unit**, for example:
  - A single outcome across repeated occasions or under multiple conditions, or multiple outcomes from the same person (“repeated measures” data)
  - Multiple persons from the same pair (“dyadic” data)
  - Multiple persons from the same group (“clustered” data)
- A **key distinction** in guiding modeling options is whether the sampling design is “**balanced**”—**is structured the same** for every sampling unit
  - **Balanced:** all persons have the same *potential* occasions, conditions, or outcomes (where *potential* allows missingness) from a common set
  - **Unbalanced:** no common set (e.g., observed occasions differ across persons, number of persons within a group differs across groups)
- We will not cover unbalanced outcomes in this class—they will be covered instead in classes focused on multilevel models (*aka*, mixed-effects models, hierarchical linear models) involving random intercepts and slopes

Estimating (Balanced) Multivariate Models

- Multivariate models can be estimated by **“tricking” univariate software** for general(*ized*) linear models (e.g., SAS MIXED, STATA MIXED) if each **variable is either a predictor OR an outcome**, not both, such as when:
  - You want to examine **mean differences** across the outcomes (e.g., over time or across conditions, as in traditional Repeated Measures ANOVA)
  - You want to test **differences in the effects of predictors** across outcomes (i.e., as in traditional MANOVA)
  - In this case we can build correlations (directly or indirectly) into the model between outcomes from the same person

- Multivariate models will need to be estimated in **“truly” multivariate software** (i.e., as path analysis models or structural equation models) if some **variables are both predictors and outcomes**, such as in mediation
  - e.g., X → M → Y, in which M is both an outcome of X and a predictor of Y
  - This involves regressions instead of correlations between outcomes

- For both types of analyses we will use **likelihood estimation instead of least squares**, so that cases with missing outcomes are not removed from the model (for what happens with missing predictors, stay tuned)

Back to General Linear Models…

- Regardless of software, **multivariate relations among outcomes from the same sampling unit** can be specified in one of two ways:
  - **Directly** → is only possible for models with normal residuals (GLM)
    - Linear predictor will only include fixed effects, like usual, because residual dependency is captured directly via residual covariances
  - **Indirectly** → is the only option using true likelihood estimation for non-normal outcomes (i.e., generali*zed* linear models)
    - Add random intercepts to the linear predictor that capture residual dependency (so the usual conditional distributions can still be used)

- To understand the difference, we first need to describe models for <u>independent observations</u> using new vocabulary—fun with matrices!
  - Let's start with this general linear model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i}) + \boldsymbol{e_i}$
    - In this "scalar" notation, the assumed independence is hidden…
    - What follows is the "**direct**" way of including relations among outcomes (we will see the "**indirect**" way at work in generali*zed* linear models)

Example: $N = 6$ persons, $n = 1$ outcome

- This GLM as scalar: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i}) + \boldsymbol{e_i}$ with binary x_i = 0 or 1
- This GLM using matrices with $k = 2$ fixed effects: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}$

$$\boldsymbol{Y} \qquad \boldsymbol{X} \qquad \boldsymbol{\beta} \qquad \boldsymbol{E}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

$$= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} \beta_0 1 + \beta_1 x_1 \\ \beta_0 1 + \beta_1 x_2 \\ \beta_0 1 + \beta_1 x_3 \\ \beta_0 1 + \beta_1 x_4 \\ \beta_0 1 + \beta_1 x_5 \\ \beta_0 1 + \beta_1 x_6 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

$\boldsymbol{Y} = N * 1$ outcome vector

$\boldsymbol{X} = N * k$ "design" matrix for predictors that have fixed effects

$\boldsymbol{\beta} = k * 1$ fixed effects vector

$\boldsymbol{E} = N * 1$ residual vector

where (by default) $\boldsymbol{e_i} \sim \mathrm{N}(0, \sigma_e^2)$

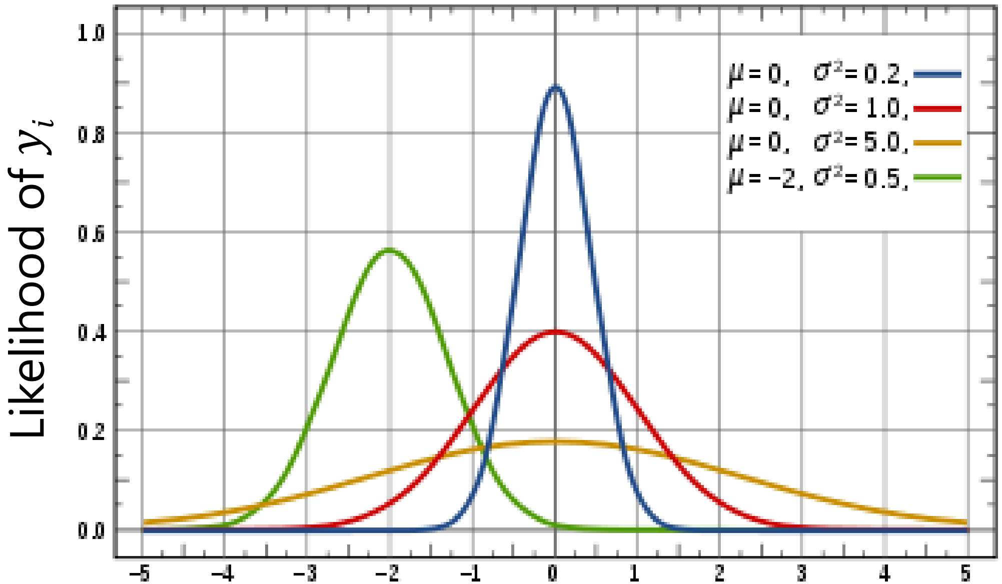
Across all 6 persons, the **combined residual variance-covariance** matrix is "**VC**":

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

Off-diagonal 0 values → independent residuals

σ_e^2 → all persons share common residual variance

Review: Univariate Normal PDF



Univariate Normal PDF:

$$f(y_i)=\frac{1}{\sqrt{2\pi\sigma_e^2}}*\exp\left[-\frac{1}{2}*\frac{\left(y_i-\hat{y}_i\right)^2}{\sigma_e^2}\right]$$

Sum over persons of log of $f(y_i)$ =
Model Log-Likelihood → Model Fit

- This PDF tells us how **likely** (i.e., **tall**) any value of $\boldsymbol{y_i}$ is given two things:
  - Conditional mean $\boldsymbol{\hat{y}_i}$
  - Residual variance $\boldsymbol{\sigma_e^2}$
- We can see this work using the NORMDIST function in excel!
  - Easiest for **empty** model:

$$\boldsymbol{y_i = \beta_0 + e_i}$$

- We can check our math via software using ML!

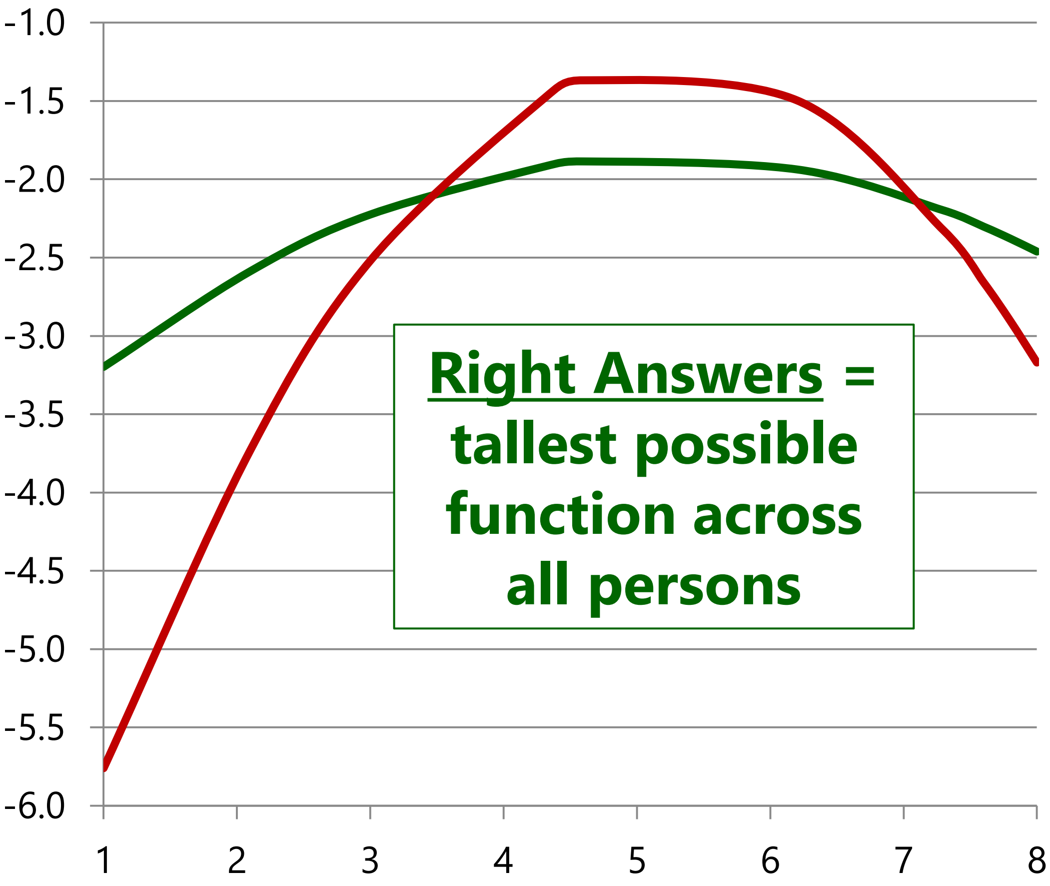
Image borrowed from: https://en.wikipedia.org/wiki/Normal_distribution

Univariate ML via Excel “NORMDIST”

Key idea: Normal Distribution formula → data height

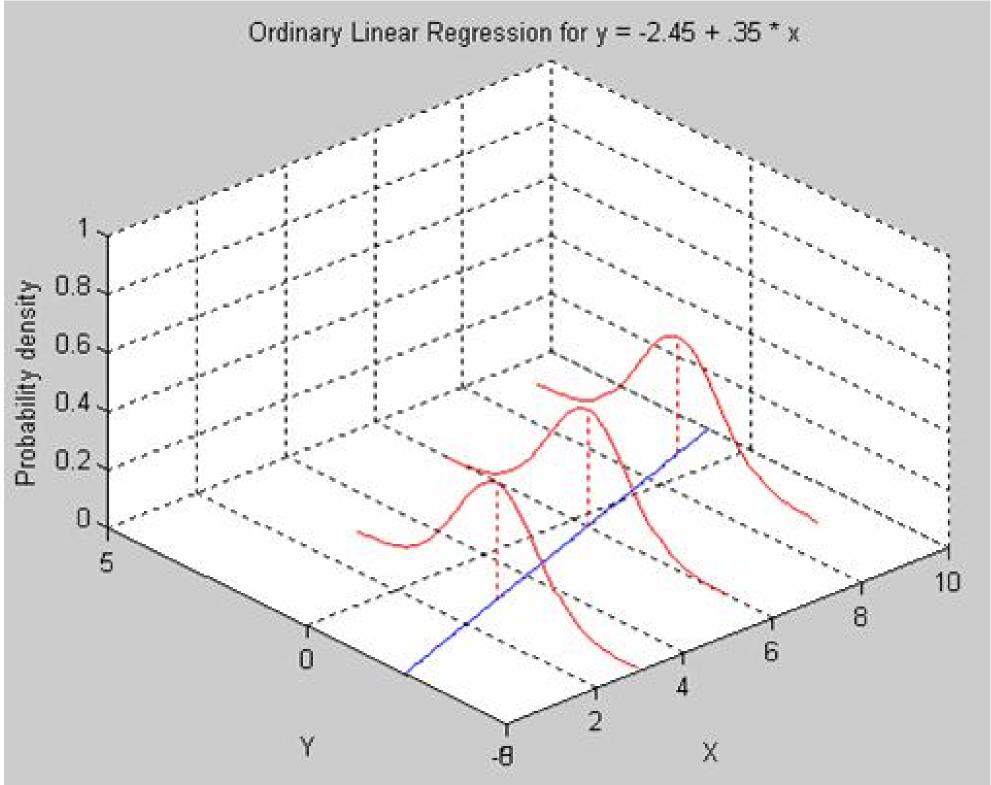
	Right	Wrong
Mean	5.19	5.24
Variance	6.56	2.00

Outcome	**Right Log(Height)**	**Wrong Log(Height)**
1.0	-3.20	-5.76
2.1	-2.59	-3.73
3.0	-2.22	-2.52
4.3	-1.92	-1.49
4.6	-1.89	-1.37
6.2	-1.94	-1.50
7.3	-2.20	-2.33
7.6	-2.30	-2.66
7.8	-2.38	-2.90
8.0	-2.46	-3.17
SUM = Model LL = taller is better		
	-23.09	-27.42



Right Answers = tallest possible function across all persons

Review: Conditional Univariate Normal



Univariate Normal PDF:

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2}\right]$$

- This function applies for any value of x_i, such as in GLM:
  - Fixed effects (intercept, predictor slopes) create a conditional mean for each person, $\hat{y}_i$
  - We assume the same residual variance σ_e^2 holds for all values of $\hat{y}_i$

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$$e_i = y_i - \hat{y}_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Left image borrowed from: http://www.omidrouhani.com/research/logisticregression/html/logisticregression.htm

From Univariate to Multivariate…

- What if the 6 observations were **2 outcomes (e.g., T1 and T2 occasions here) each from 3 persons** instead? We need a new model that builds in per-person residual dependency (and maybe different residual variances across outcomes as well, stay tuned)
- If the outcomes are in separate columns, then **to use univariate software we need to "stack" the separate outcomes** into a single column (i.e., go from a "wide" to "long" data structure) like this:

"Wide" Structure:
1 row per <u>person</u>

Person	**yT1**	**yT2**
1	y_{1t1}	y_{1t2}
2	y_{2t1}	y_{2t2}
3	y_{3t1}	y_{3t2}

"Long" Structure:
1 row per <u>outcome</u>

Person	**Time**	**x**	**y**
1	1	0	y_{1t1}
1	2	1	y_{1t2}
2	1	0	y_{2t1}
2	2	1	y_{2t2}
3	1	0	y_{3t1}
3	2	1	y_{3t2}

In the "long" structure we add a time column to index which occasion is in each row

We also make a centered version of the time index, x, to include as a predictor

Multivariate: $N = 3$ persons, $n = 2$ outcomes

- Multiv. GLM: $\boldsymbol{y}_{it} = \boldsymbol{\beta}_{00} + \boldsymbol{\beta}_{01}(\boldsymbol{x}_{it}) + \boldsymbol{e}_{it}$ with binary x_i = 0 or 1 requires a **per-person model** in matrices: $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{E}_i$

$$\begin{array}{ccccccc} \boldsymbol{Y}_i & & \boldsymbol{X}_i & \boldsymbol{\beta} & & \boldsymbol{E}_i \\ \begin{bmatrix} y_{t1} \\ y_{t2} \end{bmatrix} & = & \begin{bmatrix} 1 & x_{t1} \\ 1 & x_{t2} \end{bmatrix} & \begin{bmatrix} \beta_{00} \\ \beta_{01} \end{bmatrix} & + & \begin{bmatrix} e_{t1} \\ e_{t2} \end{bmatrix} \end{array}$$

where $\boldsymbol{e}_{it} \sim \mathrm{MVN}(0, \boldsymbol{R})$, and **2*2** $\boldsymbol{R}$ is chosen to predict **3 unique terms**:

MVN = multivariate normal

$$\boldsymbol{R} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{bmatrix}$$

(same matrix pattern would be repeated including whichever **complete rows** each person has)

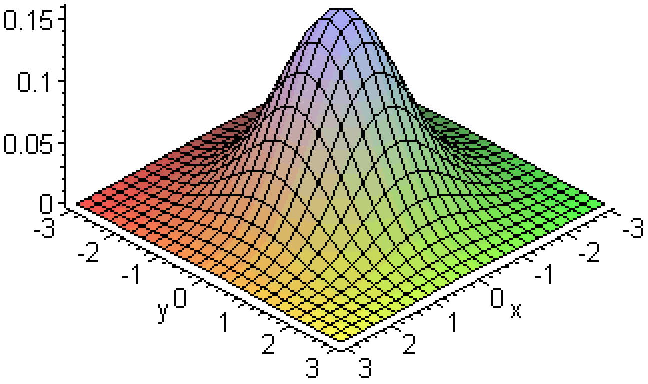
Across all 3 persons, the **combined residual variance-covariance** matrix now has a "**block diagonal**" structure with three 2*2 per-person pockets of variances (σ^2) and covariance (σ):

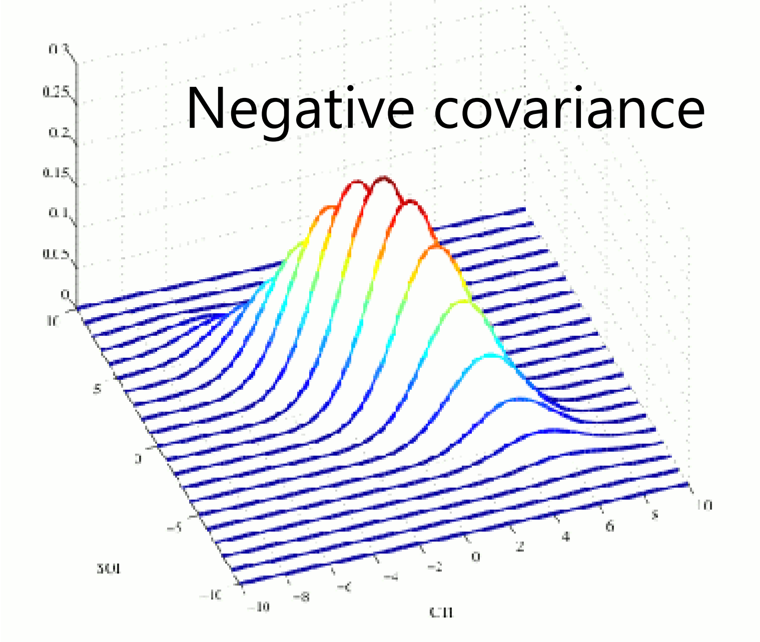
$$\begin{bmatrix} \boxed{\begin{matrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{matrix}} \end{bmatrix}$$

Off-block-diagonal 0 values → no residual covariances *across persons*
Same symbols → all persons share common residual variances and covariance

Welcome to Multivariate Normal!

- Same principle as univariate normal, but **LL is calculated for each person's SET of outcomes** (then LL is summed over persons)
- Model parameters to be found include parameters that predict **EACH outcome's residual variance and their residual covariances**
- So each outcome's likelihood height has its own dimension, but the joint shape depends on the covariance between outcomes:


No covariance


Negative covariance

Multivariate Normal for $\boldsymbol{Y}_i$ (height for all n outcomes for person i)

$$\text{Univariate Normal PDF: } f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

$$\text{Multivariate Normal PDF: } f(\mathbf{Y}_i) = \left(2\pi\right)^{-n/2} * \left|\mathbf{R}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} * \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)^T \left(\mathbf{R}_i\right)^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right]$$

- In MVN, the single $\hat{\boldsymbol{y}}_i$ per person is replaced with $\boldsymbol{X}_i\boldsymbol{\beta}$ from the model
- The model also gives $\boldsymbol{R}_i$ → the predicted residual variance-covariance matrix across outcomes (built using the same terms for each person here)
- Uses $|\boldsymbol{R}_i|$ = determinant of $\boldsymbol{R}_i$ = summary of *non-redundant* info
  - Reflects sum of variances across outcomes controlling for covariances
- $(\boldsymbol{R}_i)^{-1}$ → matrix inverse → like dividing (so can’t be 0 or negative)
  - $(\boldsymbol{R}_i)^{-1}$ must be “positive definite”, which in practice means no 0 residual variances and no out-of-bounds residual correlations between outcomes
  - Otherwise, SAS uses “generalized inverse” → questionable results

What about Missing Data?

- In univariate software using likelihood estimation: **each <u>row</u> must be complete** (otherwise, software drops it from model)
  - So whole people are not removed if they are missing some outcomes
  - (Untestable) assumption is then "**missing at random**" which means random after taking into account the person's other rows—said differently, the shape of each person's likelihood function would stay the same given complete or incomplete cases
- Whole people will be removed if they are **missing a person-level predictor**, which then implies the predictor is **missing completely at random** (what it sounds like)
  - Two solutions: **multiple imputation**, or **treat the predictor as an outcome** in a "truly" multivariate model—both require making distributional assumptions for the predictor with missing values… stay tuned

"Long" Structure:
1 row per <u>outcome</u>

Person	**Time**	**x**	**y**
1	1	0	y_{1t1}
1	2	1	y_{1t2}
2	1	0	y_{2t1}
2	2	1	y_{2t2}
3	1	0	y_{3t1}
3	2	1	y_{3t2}

For now, we will pre-select our sample for complete observations to keep the sample (and all model fit statistics) comparable across models.

Multivariate: $N = 2$ persons, 3 outcomes

- Multiv. GLM: $\boldsymbol{y_{it}} = \boldsymbol{\beta_{00}} + \boldsymbol{\beta_{01}}(\boldsymbol{x1_{it}}) + \boldsymbol{\beta_{02}}(\boldsymbol{x2_{it}}) + \boldsymbol{e_{it}}$ with 2 binary predictors **per-person model**: $\boldsymbol{Y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{E_i}$

$$\begin{array}{ccccccc} \boldsymbol{Y_i} & & \boldsymbol{X_i} & \boldsymbol{\beta} & & \boldsymbol{E_i} \\ \begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix} & = & \begin{bmatrix} 1 & x1_{t1} & x2_{t1} \\ 1 & x1_{t2} & x2_{t2} \\ 1 & x1_{t3} & x2_{t3} \end{bmatrix} & \begin{bmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{02} \end{bmatrix} & + & \begin{bmatrix} e_{t1} \\ e_{t2} \\ e_{t3} \end{bmatrix} \end{array}$$

where $\boldsymbol{e_{it}} \sim \text{MVN}(0, \boldsymbol{R})$, and **3*3 $\boldsymbol{R}$** is chosen to predict **6 unique terms**:

$$\boldsymbol{R} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

(same matrix pattern would be repeated including whichever **complete rows** each person has)

Across both persons, the **combined residual variance-covariance** matrix now has a "**block diagonal**" structure with two 3*3 per-person pockets of variances (σ^2) and covariances (σ):

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & 0 & 0 & 0 \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & 0 & 0 & 0 \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ 0 & 0 & 0 & \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ 0 & 0 & 0 & \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

Off-block-diagonal 0 values → no residual covariances *across persons*
Same symbols → all persons share common residual variances and covariances

What should the $\boldsymbol{R}$ Matrix Look Like?

- **Goal: predict all unique variances and covariances in $\boldsymbol{R}$**
  - The "**direct**" way of doing so uses only different $\boldsymbol{R}$ patterns ("R-side" models, as opposed to "G-side" models, stay tuned)

- **Next are 3 "direct" choices for unordered multiple outcomes** (btw, there are more choices for outcomes ordered in time or space)
  - SAS MIXED: `REPEATED DVindex /TYPE=?? SUBJECT=PersonID R RCORR;`
  - SAS GLIMMIX: `RANDOM DVindex /TYPE=?? SUBJECT=PersonID RESIDUAL;`
  - Stata MIXED: Goes into option ***`residuals(??, t(DVindex))`***
    - Not possible in STATA GLM or MEGLM (as far as I know)

- **The 3 choices for $\boldsymbol{R}$ patterns we will use differ in 2 respects**:
  - Is residual variance (σ^2) the same across outcomes?
    - <u>If so</u>, then are residual <u>covariances</u> (σ) are also the same across outcome pairs (remember: covariance is unstandardized correlation)
    - <u>If not</u>, might residual <u>correlations</u> (r) still be the same across outcome pairs (because covariances will differ if variances differ)

Option 1 for $\boldsymbol{R}$ Patterns: The Answer Key

- **Option 1 is an "unstructured" $\boldsymbol{R}$ matrix:** all variances and covariances are estimated separately (i.e., **non-constant, outcome-specific dependency**)
  - UN is a *description*, not a prediction, so it will fit best (i.e., as tallest LL)
  - Requires parameters = $\frac{n*(n+1)}{2}$ for n outcomes (so is hard to estimate past 5ish outcomes in smaller samples)
  - Left: $\boldsymbol{R}_{\mathbf{cov}}$ is a covariance matrix; $\boldsymbol{R}_{\mathbf{cor}}$ is a correlation matrix

$$\mathbf{R}_{\text{cov}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

$$\mathbf{R}_{\text{cor}} = \begin{bmatrix} 1 & \frac{\sigma_{12}}{\sigma_1\sigma_2} & \frac{\sigma_{13}}{\sigma_1\sigma_3} \\ \frac{\sigma_{12}}{\sigma_1\sigma_2} & 1 & \frac{\sigma_{23}}{\sigma_2\sigma_3} \\ \frac{\sigma_{13}}{\sigma_1\sigma_3} & \frac{\sigma_{23}}{\sigma_2\sigma_3} & 1 \end{bmatrix}$$

- Btw, **an unstructured $\boldsymbol{R}$ matrix** is also known as a "multivariate (MANOVA) model" or the "multivariate approach" to repeated measures (RM) ANOVA
  - Why the difference? When people say "RM ANOVA" or "MANOVA" they are often referring to the use of **least squares** instead of **likelihood estimation**
    - **Same model**, but **least squares uses only complete outcomes per person** (so outcomes are assumed missing completely at random), but likelihood estimation uses all possible outcomes (assumed missing at random instead)

Two More Choices for Patterns of *R*

- **Option 2 is a "compound symmetry heterogeneous"** ***R*** matrix: separate variances, but covariances are created using a common correlation ("CSH"):
  - Uses $n+1$ **total parameters**; after correcting for different outcome variances, all outcome residuals are correlated to the same extent (i.e., constant dependency)

$$\mathbf{R}_{\text{cov}} = \begin{bmatrix} \sigma_1^2 & \text{CSH}\sigma_1\sigma_2 & \text{CSH}\sigma_1\sigma_3 \\ \text{CSH}\sigma_2\sigma_1 & \sigma_2^2 & \text{CSH}\sigma_2\sigma_3 \\ \text{CSH}\sigma_3\sigma_1 & \text{CSH}\sigma_3\sigma_2 & \sigma_3^2 \end{bmatrix} \quad \mathbf{R}_{\text{cor}} = \begin{bmatrix} 1 & \text{CSH} & \text{CSH} \\ \text{CSH} & 1 & \text{CSH} \\ \text{CSH} & \text{CSH} & 1 \end{bmatrix}$$

CSH is still not available within STATA MIXED ☹

- **Option 3 is a "compound symmetry"** ***R*** matrix: equal variances and equal covariances (so only **2 parameters** no matter how many outcomes)
  - All **dependency is constant across outcomes** and is caused by person mean diffs: the "CS" parameter

$$\mathbf{R}_{\text{cov}} = \begin{bmatrix} \text{CS}+\sigma_e^2 & \text{CS} & \text{CS} \\ \text{CS} & \text{CS}+\sigma_e^2 & \text{CS} \\ \text{CS} & \text{CS} & \text{CS}+\sigma_e^2 \end{bmatrix}$$

$$\mathbf{R}_{\text{cor}} = \begin{bmatrix} 1 & \frac{\text{CS}}{\text{CS}+\sigma_e^2} & \frac{\text{CS}}{\text{CS}+\sigma_e^2} \\ \frac{\text{CS}}{\text{CS}+\sigma_e^2} & 1 & \frac{\text{CS}}{\text{CS}+\sigma_e^2} \\ \frac{\text{CS}}{\text{CS}+\sigma_e^2} & \frac{\text{CS}}{\text{CS}+\sigma_e^2} & 1 \end{bmatrix}$$

  - Also known as the "univariate" approach to RM ANOVA (if using least squares) and a "random intercept only" model (in the "indirect" way of capturing dependency)

How to Choose among $\boldsymbol{R}$ Matrices

- Use **likelihood ratio tests (LRT)**: treat difference in −2LL as regular χ^2 with DF = # parameters different (also, smallest AIC and BIC win)
  - VC (equal variances, no covariances) is the default and is nested in all others
    - CS fit better than VC? There's covariance (dependency) across outcomes
  - CS is nested in CSH, which are both nested in UN (= the data)
    - CSH fit better than CS? Then variances need to differ by outcome
    - UN fit better than CSH? Then correlations need to differ across outcome pairs

- Goal: find a **simpler** model that **fits not worse than UN**
  - UN will always fit best by −2LL because it is trying to recreate the complete data results (assuming missing at random)
  - Why not just use UN always? If may not always be estimable, and using a simpler model that fits not worse can lead to greater power (because more unnecessary parameters → less power)

- Btw, $\boldsymbol{R}$ matrix residual variances and covariances can also be allowed to differ across groups (see example 4a); test if that helps with LRTs
  - And in univariate models, residual variance can differ by predictors, too!

Assessing Relative Model Fit, In General

- **Model for the Means (linear predictor of fixed effects)** → which fixed effects of predictors are included in the model
  - Because fixed effects of predictors are unbounded, you can always use **univariate or multivariate Wald tests** to see if they contribute to the model (with denominator DF depending on software availability)
  - Could use LRTs, but only for models estimated with maximum likelihood (not residual maximum likelihood, a better choice for normal residuals)

- **Model for the Variance** → what the pattern of variance and covariance of residuals from the same sampling unit should be
  - DOES require assessment of relative model fit using LRTs: Because variances cannot be negative, you cannot use Wald test *p*-values (i.e., that show up in MIXED output next to the variance estimate)
  - Conditional distributions can only be compared using LRTs (usually with a mixture χ^2) or information criteria (AIC, BIC) if they are nested
    - e.g., Poisson and Negative Binomial differ by "stretchy k"; binomial and beta-binomial differ by "stretchy ϕ"; zero-inflation models add an intercept in another submodel that predicts the logit of being an extra 0

Why Use Multivariate Models?

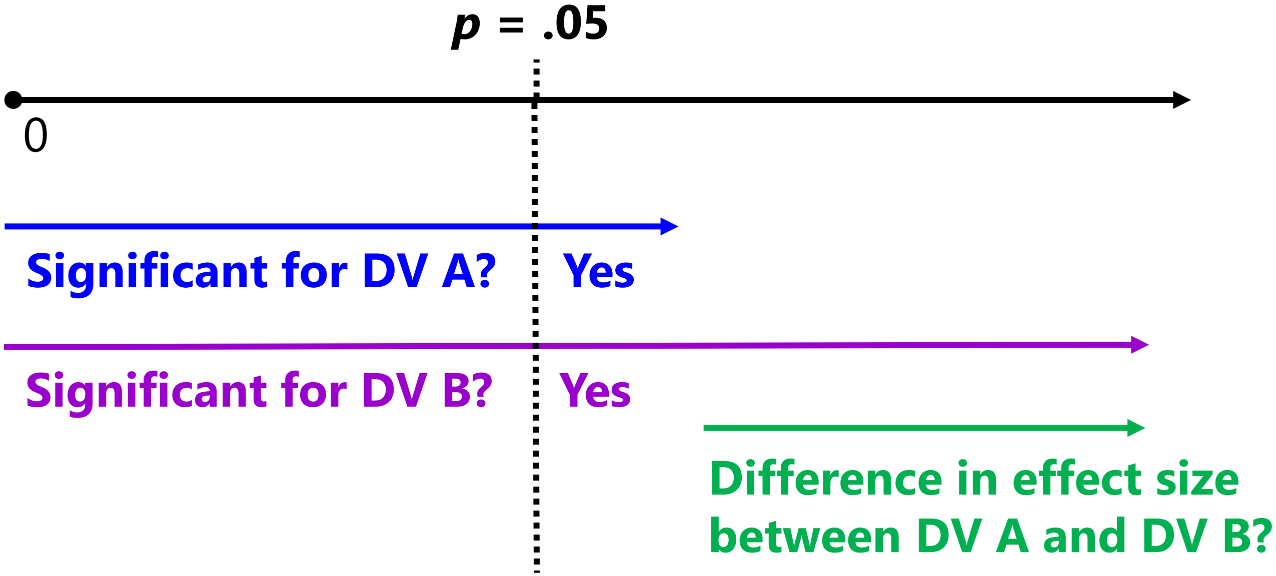
- **Repeated measures** designs (e.g., sampling over occasions or conditions) readily lend themselves to multivariate models
  - Goal is to examine **mean differences** across the per-person outcomes (i.e., as in traditional RM ANOVA but using likelihood for missing data)
  - Usual "**general intercept**" fixed effects strategy is likely most useful

- Multivariate models are also the optimal way to predict multiple outcomes (DVs)—simultaneously rather than in separate models
  - Examine **differences in predictor effects** across outcomes (and then constrain effects that are similar in size to be equal for greater power)
    - Outcomes should be transformed to common scale (such as by z-scoring them) if not *similarly interpretable* already (e.g., such as variants of same scale)
    - If so, a "**DV-specific intercepts**" fixed effects strategy may be more useful
  - Predicting each outcome is a better alternative to predicting difference scores or to "controlling for time 1" ANCOVA ("residualized change", bleh)

Differences in Effect Size across DVs

Scenario 1: Fixed effect is significant for both DVs:

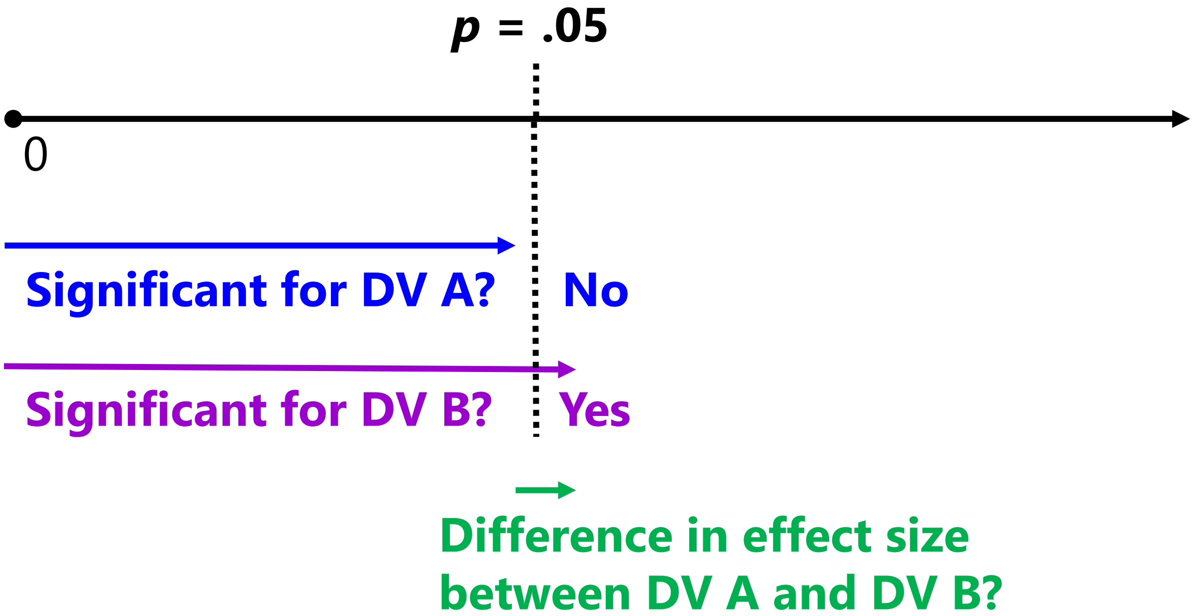
Absolute Value of Effect Size



Just because a predictor is **significant for both DVs** does not mean it has the **same magnitude** of relationship across DVs!

Differences in Effect Size across DVs

Scenario 2: Fixed effect is significant for DV B only:



Absolute Value of Effect Size

***p* = .05**

0

Significant for DV A? **No**

Significant for DV B? **Yes**

Difference in effect size between DV A and DV B?

Also, just because a predictor is **non-significant for one DV but significant for another DV** does not mean it has **different magnitudes** of relationships across DVs!

Fixed Effects Parameterizations

- Here are 2 ways of fitting the same model predicting y_{id} (a stacked outcome including DV = A, B, or C indexed by a categorical variable DV) from a general intercept (i.e., a column of 1s), a person-level predictor x_i, and 3 dummy-coded contrasts: *dvA*, *dvB*, and *dvC*:
  - If **DV=A**, then dvA=**1**, dvB=0, dvC=0
  - If **DV=B**, then dvA=0, dvB=**1**, dvC=0
  - If **DV=C**, then dvA=0, dvB=0, dvC=**1**

1. "**General intercept**": provides fixed effects for a reference DV and DV differences in fixed effects relative to the reference DV
   - So *fixed effects for non-reference DV* are found as linear combinations
2. "**DV-specific intercepts**": provides effects separately by DV
   - So *DV differences in fixed effects* are found as linear combinations
   - This is always how path models are parameterized (stay tuned)

DV as a Categorical Predictor

- DV-specific dummy codes can be replaced by a categorical DV predictor
  - SAS: put in on the **CLASS** statement; STATA: use i. prefix for predictor
- For a predictor with C categories, the program automatically then creates $\boldsymbol{C}$ **new internal variables**, for example "DV" with $C = 3$:

DV	_IsA	_IsB	_IsC
A	1	0	0
B	0	1	0
C	0	0	1

Default reference category by program:

- **SAS** (and SPSS): **highest/last** is reference
- STATA: **lowest/first** is default reference
  - Can easily change reference category, e.g., last = ref → ib(last).DV

- It then figures out how many of these internal variables are needed to create C means—if using an intercept, then it's $C - 1$; without the intercept, is C
- It enters them until it hits that criterion—the one left out is your reference; if all C internal variables are included, then each is a custom intercept
- When referring the categorical predictor (e.g., ESTIMATE, LINCOM, CONTRAST), you must tell it what to do with EACH of these internal variables [e.g., 1 0 0]

"General Intercept" Parameterization

- **Empty Model:** $\boldsymbol{\hat{y}_{id} = \beta_{00} + \beta_{01}(dvA_{id}) + \beta_{02}(dvB_{id})}$
  - β_{00} = intercept for DV=C (i.e., when dvA=0 and dvB=0)
  - β_{01} = mean difference for DV=C vs. DV=A
  - β_{02} = mean difference for DV=C vs. DV=B
  - $\beta_{02} - \beta_{01}$ = mean difference for DV=A vs. DV=B
  - $\beta_{00} + \beta_{01}$ = intercept for DV=A
  - $\beta_{00} + \beta_{02}$ = intercept for DV=B

With the general intercept (=1 for all), the dvA and dvB variables act like typical group differences (just between outcomes instead of groups).

- **Add $\boldsymbol{x_i}$ predictor:** $\boldsymbol{+ \beta_{10}(x_i) + \beta_{11}(dvA_{id})(x_i) + \beta_{12}(dvB_{id})(x_i)}$
  - β_{10} = x_i slope for DV=C (i.e., when x*dvA=0 and x*dvB=0)
  - β_{11} = difference in x_i slope for DV=C vs. DV=A
  - β_{12} = difference in x_i slope for DV=C vs. DV=B
  - $\beta_{12} - \beta_{11}$ = difference in x_i slope for DV=A vs. DV=B
  - $\beta_{10} + \beta_{11}$ = x_i slope for DV=A
  - $\beta_{10} + \beta_{12}$ = x_i slope for DV=B

“General Intercept” Parameterization

- Including a general intercept is convenient when differences between occasions or conditions are of most interest (i.e., then the difference scores are captured directly by fixed effects)
- Here are 2 equivalent versions of this model in SAS MIXED:

```
CLASS PersonID DV;
MODEL y = dvA dvB x dvA*x dvB*x
              / SOLUTION DDFM=Satterthwaite;
REPEATED DV / R RCORR TYPE=?? SUBJECT=PersonID;
```

```
CLASS PersonID DV;
MODEL y = DV x DV*x
              / SOLUTION DDFM=Satterthwaite;
REPEATED DV / R RCORR TYPE=?? SUBJECT=PersonID;
```

Note SAS REPEATED and STATA RESIDUALS stay the same across syntax variants

- Here are 2 equivalent versions of this model in STATA MIXED:

```
mixed y c.dvA c.dvB c.x c.dvA#c.x c.dvB#c.x,  ///
        variance reml dfmethod(satterthwaite) ///
        residuals(??,t(DV))
```

```
mixed y ib(last).DV c.x ib(last).DV#c.x,      ///
        variance reml dfmethod(satterthwaite) ///
        residuals(??,t(DV))
```

I used ib(last) to make C the reference DV (as in SAS)

"DV-Specific Intercept" Parameterization

- **Empty Model:** $\hat{y}_{id} = \underline{\boldsymbol{\beta_{00}(dvC_{id})}} + \boldsymbol{\beta_{01}(dvA_{id})} + \boldsymbol{\beta_{02}(dvB_{id})}$
  - β_{00} = intercept for DV=C (i.e., when dvA=0 and dvB=0)
  - β_{01} = intercept for DV=A (i.e., when dvB=0 and dvC=0)
  - β_{02} = intercept for DV=B (i.e., when dvA=0 and dvC=0)
  - $\beta_{01} - \beta_{00}$ = mean difference for DV=C vs. DV=A
  - $\beta_{02} - \beta_{00}$ = mean difference for DV=C vs. DV=B
  - $\beta_{02} - \beta_{01}$ = mean difference for DV=A vs. DV=B

Without the general intercept (=1 for all), the dvA, dvB, and dvC variables act like "switches" that turn on the fixed effects for its DV.

- **Add $\boldsymbol{x_i}$ predictor:** $+ \underline{\boldsymbol{\beta_{10}(x_i)(dvC_{di})}} + \boldsymbol{\beta_{11}(x_i)(dvA_{di})} + \boldsymbol{\beta_{12}(x_i)(dvB_{di})}$
  - β_{10} = x_i slope for DV=C (i.e., when x*dvA=0 and x*dvB=0)
  - β_{11} = x_i slope for DV=A (i.e., when x*dvB=0 and x*dvC=0)
  - β_{12} = x_i slope for DV=B (i.e., when x*dvA=0 and x*dvC=0)
  - $\beta_{11} - \beta_{10}$ = difference in x_i slope for DV=C vs. DV=A
  - $\beta_{12} - \beta_{10}$ = difference in x_i slope for DV=C vs. DV=B
  - $\beta_{12} - \beta_{11}$ = difference in x_i slope for DV=A vs. DV=B

“DV-Specific Intercept” Parameterization

- Removing the general intercept is convenient when fixed effects per DV are of most interest or when not all DVs get all fixed effects
- Here are 2 equivalent versions of this model in SAS MIXED:

```
CLASS PersonID DV;
MODEL y = dvC dvA dvB dvC*x dvA*x dvB*x
            / NOINT SOLUTION DDFM=Satterthwaite;
REPEATED DV / R RCORR TYPE=?? SUBJECT=PersonID;
```

```
CLASS PersonID DV;
MODEL y = DV DV*x
            / NOINT SOLUTION DDFM=Satterthwaite;
REPEATED DV / R RCORR TYPE=?? SUBJECT=PersonID;
```

Note SAS REPEATED and STATA RESIDUALS stay the same; in STATA I used ib(last) to make C the reference DV (same as in SAS)

- Here are 2 equivalent versions of this model in STATA MIXED:

```
mixed y c.dvC c.dvA c.dvB c.DVc#c.x c.dvA#c.x c.dvB#c.x, ///
        noconstant variance reml dfmethod(satterthwaite) ///
        residuals(??,t(DV))
```

```
mixed y ib(last).DV ib(last).DV#c.x,                      ///
        noconstant variance reml dfmethod(satterthwaite) ///
        residuals(??,t(DV))
```

Specifying Fixed Effects: Caveats

- Btw, to constrain the x_i slope to be equal across DVs, remove its DV-interaction terms—just enter x_i as a main effect (either version)
- You *can* mix-and-match parameterizations, but do so *carefully*!
  - e.g., here is how to retain separate effects of $x1_i$ for DVs A, B, and C, but have an effect of $x2_i$ only for DV=A and DV=B
  - **General intercept version:**

$$\begin{aligned}\hat{y}_{id} &= \beta_{00} + \beta_{01}(dvA_{id}) + \beta_{02}(dvB_{id}) \\ &\quad + \beta_{10}(x1_i) + \beta_{11}(dvA_{id})(x1_i) + \beta_{12}(dvB_{id})(x1_i) \\ &\quad + \beta_{21}(dvA_{id})(x2_i) + \beta_{22}(dvB_{id})(x2_i)\end{aligned}$$

  - **DV-specific intercept version:**

$$\begin{aligned}\hat{y}_{id} &= \beta_{00}(dvC_{id}) + \beta_{01}(dvA_{id}) + \beta_{02}(dvB_{id}) \\ &\quad + \beta_{10}(dvC_{id})(x1_i) + \beta_{11}(dvA_{id})(x1_i) + \beta_{12}(dvB_{id})(x1_i) \\ &\quad + \beta_{21}(dvA_{id})(x2_i) + \beta_{22}(dvB_{id})(x2_i)\end{aligned}$$

  - Either way, β_{21} and β_{22} give the effect of $x2_i$ for DV=A and DV=B (but it *looks* logically inconsistent in the general intercept version)

Wrapping Up…

- So far the generalized linear models we have examined have been **univariate models**—their LLs assume **persons are independent**

- But analyzing data in which each sampling unit has more than one outcome requires **multivariate models** instead
  - We need to add model terms that capture **dependency**, this semester for balanced designs (i.e., all persons have the same *potential* outcomes)
  - For plausibly normal outcomes, **dependency can be modeled directly**: we can allow same or different residual variances and covariances across outcomes (in a **person-specific R matrix** of type UN, CSH, or CS)
  - We have to use **likelihood ratio tests** ($-2\Delta LL$ as χ^2) to compare nested models to decide which fits least worse to protect our fixed effect SEs

- For convenience, fixed effects can be specified in 2 different ways
  - Single general intercept → DV terms reflect DV **differences**
  - Multiple DV-specific intercepts → DV terms are **switches** for own effects