Multilevel models

Part 2

03.02.22

[Click here for PDF of slides](#)

Announcements

- Tomorrow's lab: article evaluation for mini-project 02

- Looking ahead to the individual final project...

  - Choose a dataset with **correlated observations** and analyze it
  - More details after spring break

Learning goals

- Conduct univariate and bivariate EDA for multilevel models

- Write multilevel model , including assumptions about variance components, in by-level and composite forms

- Interpret the model parameters, fixed effects, and variance components

Data: Music performance anxiety

The data **musicdata.csv** come from the Sadler and Miller (2010) study of the emotional state of musicians before performances. The dataset contains information collected from 37 undergraduate music majors who completed the Positive Affect Negative Affect Schedule (PANAS), an instrument produces a measure of anxiety (negative affect) and a measure of happiness (positive affect). This analysis will focus on negative affect as a measure of performance anxiety.

The primary variables we'll use are

- **na**: negative affect score on PANAS (the response variable)

- **perform_type**: type of performance (Solo, Large Ensemble, Small Ensemble)

- **instrument**: type of instrument (Voice, Orchestral, Piano)

Look at data

id	diary	perform_type	na	gender	instrument
1	1	Solo	11	Female	voice
1	2	Large Ensemble	19	Female	voice
1	3	Large Ensemble	14	Female	voice
43	1	Solo	19	Female	voice
43	2	Solo	13	Female	voice
43	3	Small Ensemble	19	Female	voice

Draw the data structure, and add the Level One and Level Two observations and variables.

STA 310

6

Univariate exploratory data analysis
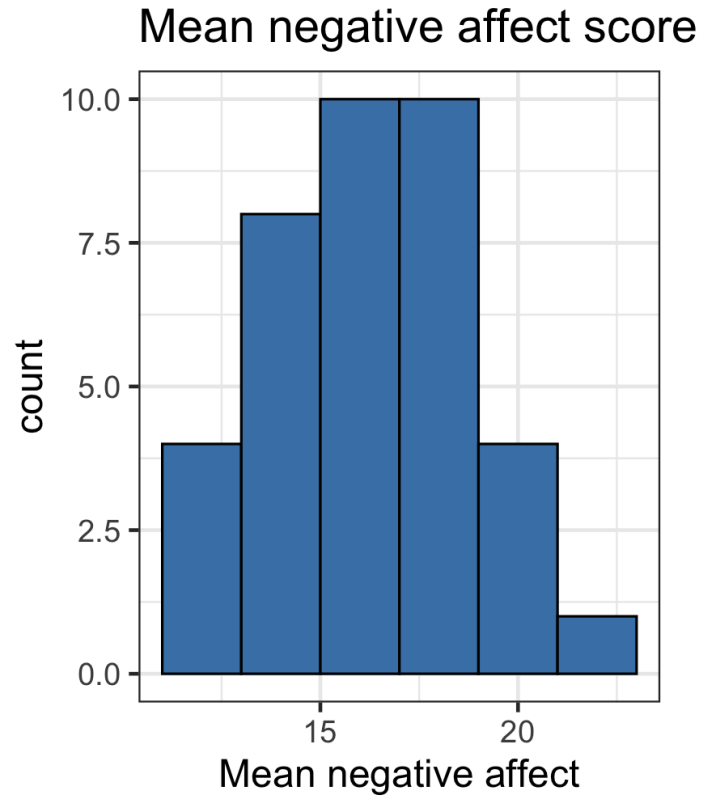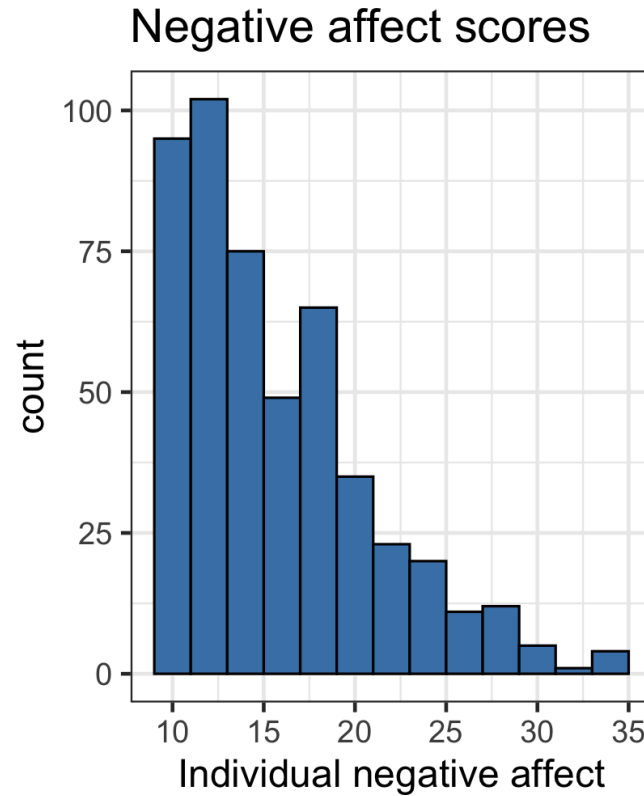
Level One variables

Two ways to approach univariate EDA (visualizations and summary statistics) for Level One variables:

- Use individual observations (i.e., treat observations as independent)

- Use aggregated values for each Level Two observation

Level Two variables

- Use a data set that contains one row per Level Two observation

Unviariate EDA

Bivariate exploratory data analysis

Goals

- Explore general association between the predictor and response variable
- Explore whether subjects at a given level of the predictor tend to have similar mean responses
- Explore whether variation in response differs at different levels of a predictor

There are two ways to visualize these associations:

- One plot of individual observations (i.e., treat observations as independent)

- Separate plots of responses vs. predictor for each Level Two observation (lattice plots)

Complete Part 1: Bivariate EDA in `lecture-16.Rmd`.

06:00

Fitting the model

Questions we want to answer

The goal is to understand variability in performance anxiety (**na**) based on performance-level and musician-level characteristics. Specifically:

> What is the association between performance type (large ensemble or not) and performance anxiety? Does the association differ based on instrument type (orchestral or not)?

We will fit the model in two parts:

1 Fit a separate model for each musician understand the association between performance type (Level One models).

2 Then fit a system of models to predict the fitted coefficients in the Level One models based on instrument type (Level Two models).

1. How many Level One models will we fit?
2. How many Level Two models will we fit?

[Click here](#) to submit your response.

1️⃣ Level One models

We'll start with a Level One model to understand the association between performance type and performance anxiety for the ith musician.

$$na_{ij} = a_i + b_i \ LargeEnsemble_{ij} + \epsilon_i, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Add the Level One model to the diagram.

For now, we will estimate a_i and b_i using least-squares regression.

Example Level One model

Below is partial data for observation #22

id	diary	perform_type	instrument	na
22	1	Solo	orchestral instrument	24
22	2	Large Ensemble	orchestral instrument	21
22	3	Large Ensemble	orchestral instrument	14
22	13	Large Ensemble	orchestral instrument	12
22	14	Large Ensemble	orchestral instrument	19
22	15	Solo	orchestral instrument	25

Level One model

```
music %>%
  filter(id == 22) %>%
  lm(na ~ large_ensemble, data = .) %>%
  tidy() %>%
  kable(digits = 3)
```
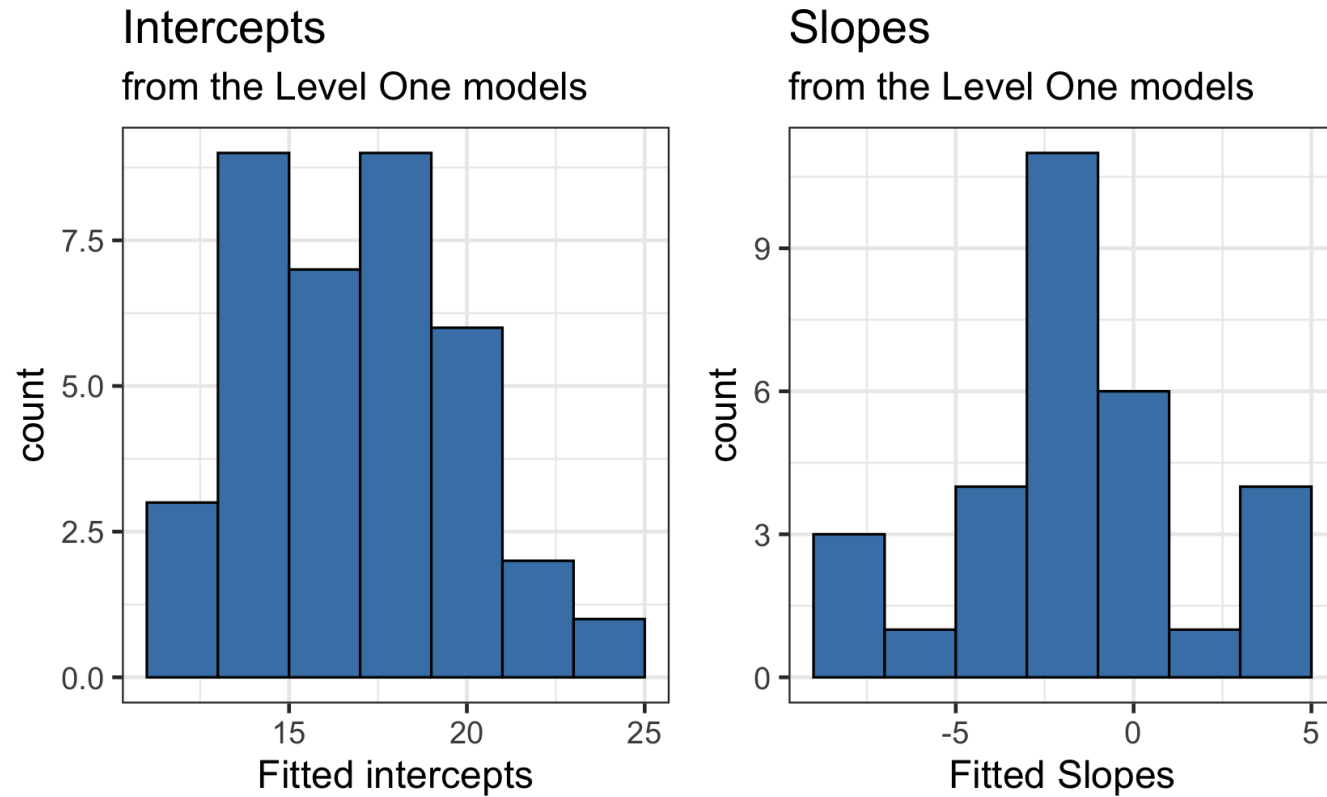
term	estimate	std.error	statistic	p.value
(Intercept)	24.500	1.96	12.503	0.000
large_ensemble1	-7.833	2.53	-3.097	0.009

Repeat for all 37 musicians.

Read the code in Part 2: Level One models in `lecture-16.Rmd`.

Suppose a student only played in small ensembles/ solos.

- What is the estimated intercept for this student?

- What is the estimated slope for this student?

04:00

STA 310

Recreated from BMLR Figure 8.9

Now let's consider if there is an association between the estimated slopes, estimated intercepts, and the type of instrument

Level Two models

The slope and intercept for the ith musician can be modeled as

$$a_i = \alpha_0 + \alpha_1\,Orchestra_i + u_i$$
$$b_i = \beta_0 + \beta_1\,Orchestra_i + v_i$$

Note the response variable in the Level Two models are not observed outcomes but the (fitted) slope and intercept from each musician

See Part 3: Level Two Models in `lecture-16.Rmd.`

How many observations used to estimate a? How many observations are used to estimate b? Click here to submit your response.

Estimated coefficients by instrument
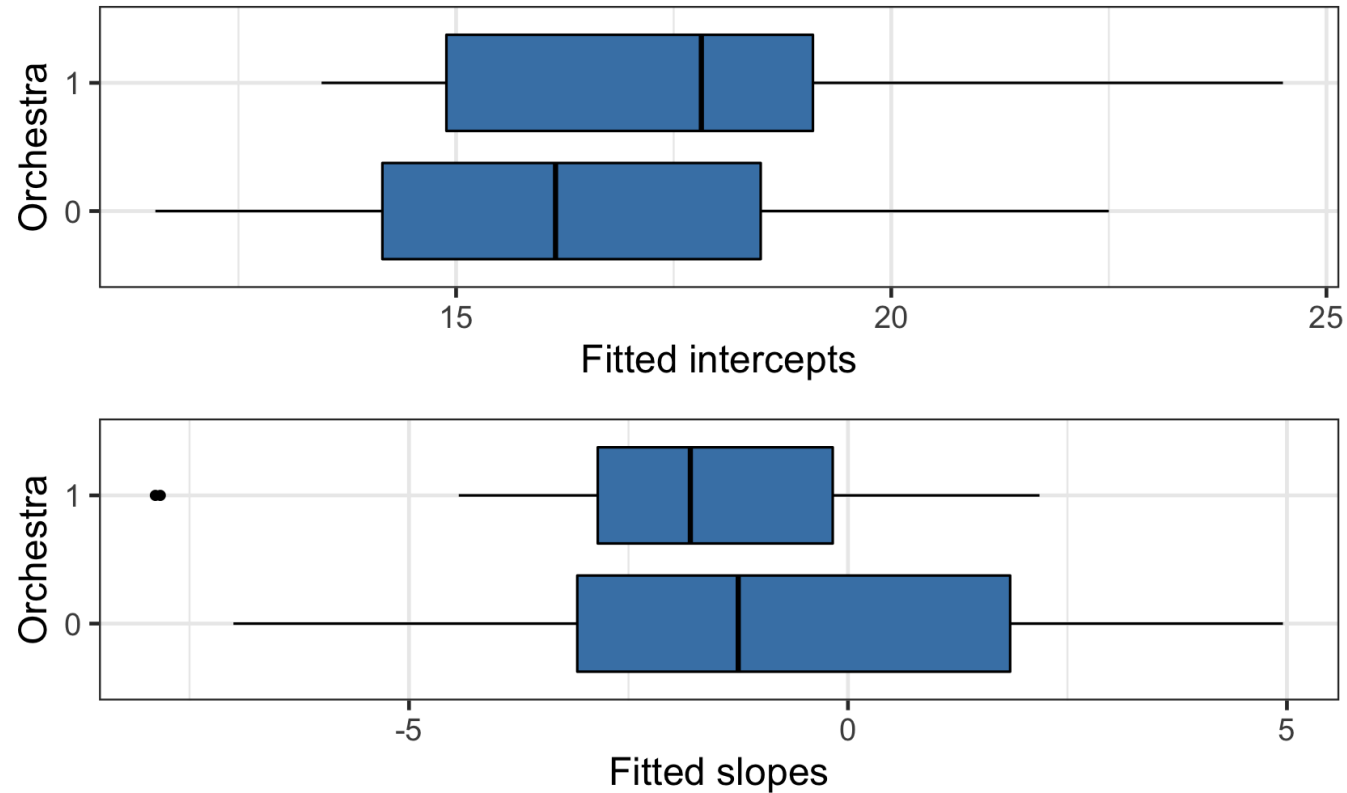
Level Two model

Model for intercepts

term	estimate	std.error	statistic	p.value
(Intercept)	16.283	0.671	24.249	0.000
orchestra1	1.411	0.991	1.424	0.163

Model for slopes

term	estimate	std.error	statistic	p.value
(Intercept)	-0.771	0.851	-0.906	0.373
orchestra1	-1.406	1.203	-1.168	0.253

Writing out the models

Level One

$$\hat{na}_{ij} = \hat{a}_i + \hat{b}_i \; LargeEnsemble_{ij}$$

for each musician.

Level Two

$$\hat{a}_i = 16.283 + 1.441 \; Orchestra_i$$
$$\hat{b}_i = -0.771 - 1.406 \; Orchestra_i$$

Composite model

$$\hat{na}_i = 16.283 + 1.441\, Orchestra_i - 0.771\, LargeEnsemble_{ij}$$
$$- 1.406\, Orchestra : LargeEnsemble_{ij}$$

- What is the predicted average performance anxiety before solos and small ensemble performances for vocalists and keyboardists? For those who place orchestral instruments?

- What is the predicted average performance anxiety before large ensemble performances for those who play orchestral instruments?

Disadvantages to this approach

⚠️ Weighs each musician the same regardless of number of diary entries

⚠️ Drops subjects who have missing values for slope (7 individuals who didn't play large ensemble performance)

⚠️ Does not share strength effectively across individuals

Plot the R^2 values calculated in Part 2: Level One Model of `lecture-16.Rmd`.

Unified approach to two-level modeling

Framework

Let Y_{ij} be the performance anxiety for the ith musician before performance j.

Level One

$$Y_{ij} = a_i + b_i \, LargeEnsemble + \epsilon_{ij}$$

Level Two

$$a_i = \alpha_0 + \alpha_1 \, Orchestra_i + u_i$$
$$b_i = \beta_0 + \beta_1 \, Orchestra_i + v_i$$

This approach uses likelihood-based methods (instead of least squares) to address the previously mentioned disadvantages

STA 310

Composite model

Plug in the equations for a_i and b_i to get the **composite model**

$$Y_{ij} = (\alpha_0 + \alpha_1\, Orchestra_i + \beta_0\, LargeEnsemble_{ij}$$
$$+ \beta_1\, Orchestra_i : LargeEnsemble_{ij})$$
$$+ (u_i + v_i\, LargeEnsemble_{ij} + \epsilon_{ij})$$

- The **fixed effects** to estimate are $\alpha_0, \alpha_1, \beta_0, \beta_1$
- The **random effects** are u_i, v_i, ϵ_{ij}

Note that we no longer need to estimate a_i and b_i directly as we did earlier. They conceptually connect the Level One and Level Two models.

Notation

- Greek letters denote the fixed effect model parameters to be estimated

  - e.g., $\alpha_0, \alpha_1, \beta_0, \beta_1$

- Roman letters denote the preliminary fixed effects at lower levels that will not be estimated directly.

  - e.g. a_i, b_i

- σ and ρ denote variance components that will be estimated

- ϵ_{ij}, u_i, v_i denote error terms

Error terms

- We generally assume that the error terms are normally distributed, e.g. error associated with each performance of a given musician is $\epsilon_{ij} \sim N(0, \sigma^2)$

- For the Level Two models, the errors are

  - u_i: deviation of musician i from the mean performance anxiety before solos and small ensembles after accounting for the instrument

  - v_i: deviance of musician i from the mean difference in performance anxiety between large ensembles and other performance types after accounting for instrument

- Need to account for fact that u_i and v_i are correlated for the ith musician

Fitted slopes and intercepts

r = -0.525
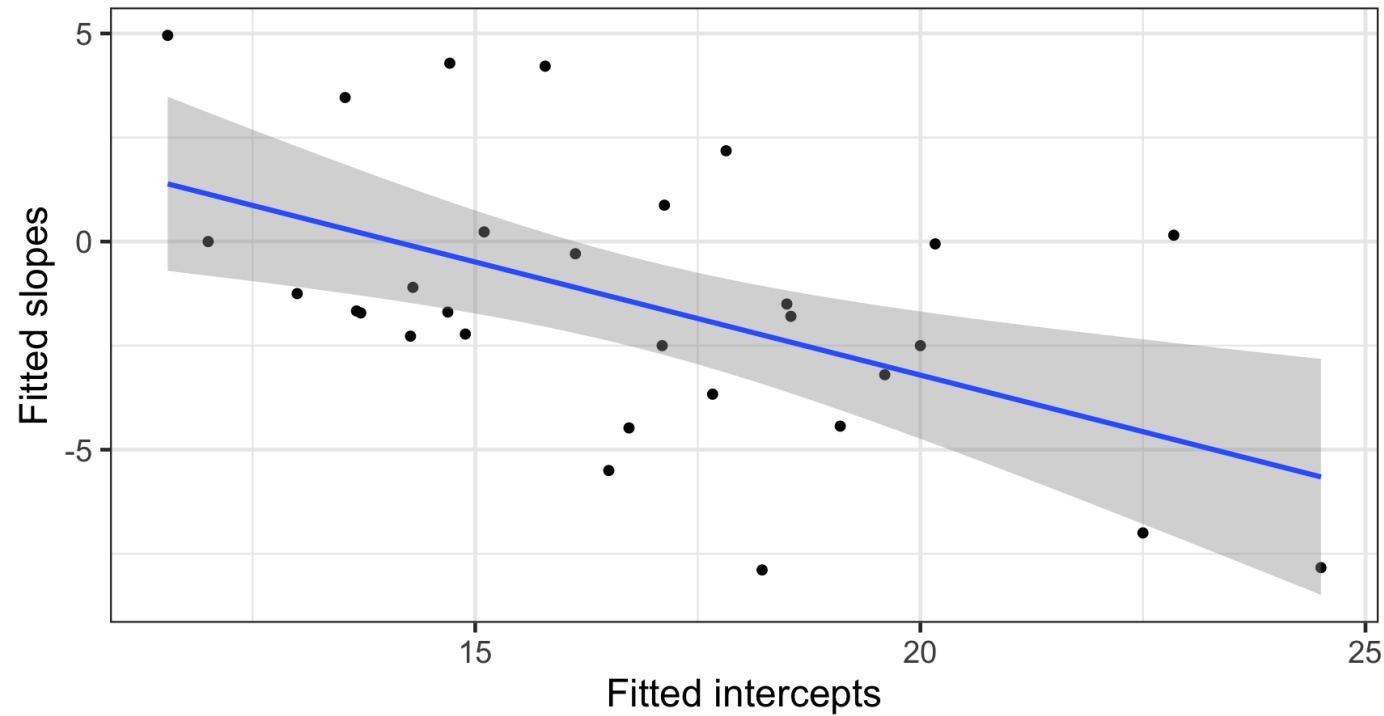
Recreated from Figure 8.11

Describe what we learn about the association between the slopes and intercepts based on this plot.

30

Distribution of Level Two errors

Use a **multivariate normal** distribution for the Level Two error terms

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \right)$$

where σ_u^2 and σ_v^2 are the variance of u_i's and v_i's respectively, and $\sigma_{uv} = \rho_{uv}\sigma_u\sigma_v$ is covariance between u_i and v_i

- What does it mean for $\rho_{uv} > 0$?

- What does it mean for $\rho_{uv} < 0$?

Visualizing multivariate normal distribution
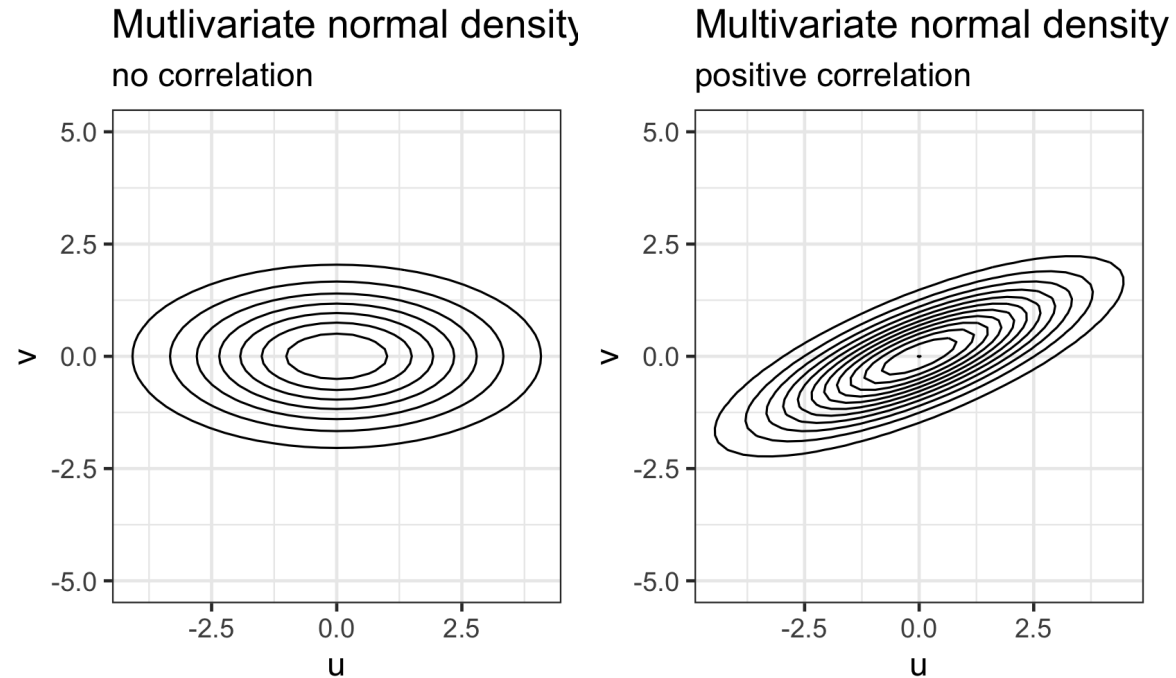
Recreated from Figure 8.12

Fit the model

Fit multilevel model using the `lmer` function from the lme4 package. Display results using hte `tidy()` function from the broom.mixed package.

```r
library(lme4)
library(broom.mixed)

music_model <- lmer(na ~ orchestra + large_ensemble +
                      orchestra:large_ensemble + (large_ensemble|id),
                   REML = TRUE, data = music)

tidy(music_model) %>% kable(digits = 3)
```

Fit the model

effect	group	term	estimate	std.error	statistic
fixed	NA	(Intercept)	15.930	0.641	24.833
fixed	NA	orchestra1	1.693	0.945	1.791
fixed	NA	large_ensemble1	-0.911	0.845	-1.077
fixed	NA	orchestra1:large_ensemble1	-1.424	1.099	-1.295
ran_pars	id	sd__(Intercept)	2.378	NA	NA
ran_pars	id	cor__(Intercept).large_ensemble1	-0.635	NA	NA
ran_pars	id	sd__large_ensemble1	0.672	NA	NA
ran_pars	Residual	sd__Observation	4.670	NA	NA

Acknowledgements

The content in the slides is from