09_homework_lmm_anova

Bill Perry

```r
library(janitor)
```

```
Attaching package: 'janitor'
```

```
The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(readxl)
library(lme4)
```

```
Loading required package: Matrix
```

```r
library(lmerTest)
```

```
Attaching package: 'lmerTest'
```

```
The following object is masked from 'package:lme4':

    lmer
```

```
The following object is masked from 'package:stats':

    step
```

```r
library(broom.mixed)
library(performance)
library(sjPlot)
library(skimr)
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.2     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
```

```
── Conflicts ──────────────────────────────────── tidyverse_conflicts() ──
✖ tidyr::expand() masks Matrix::expand()
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
✖ tidyr::pack()   masks Matrix::pack()
✖ tidyr::unpack() masks Matrix::unpack()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```
theme_set(theme_light())
```

Assignment Overview

This homework assignment analyzes crayfish growth data from Sargent and Lodge (2014) to examine differences in growth rates between native and invasive populations of rusty crayfish (*Orconectes rusticus*) using mixed effects models with lake as a random effect.

Learning Objectives

By completing this assignment, you will be able to:

1. **Understand mixed effects models and random effects**
2. **Perform exploratory data analysis for hierarchical data**
3. **Conduct linear mixed effects analysis using lme4**
4. **Test assumptions for mixed effects models**
5. **Interpret fixed and random effects**
6. **Calculate intraclass correlation coefficients**
7. **Create publication-quality figures**
8. **Write scientific methods and results sections**

Data Description

The dataset contains growth measurements from a common garden experiment where young-of-year (YOY) rusty crayfish from native (Ohio) and invasive (Wisconsin) populations were grown in enclosures in three northern Wisconsin lakes during summer 2011. The hierarchical structure includes individual crayfish nested within lakes.

Key variables:

- - range: Population origin (Native vs Invasive)
- - fixed effect - lake: Lake location (Big, High, Papoose)
- - random effect - growth_per_day: Daily growth rate (mm/day)
- - response variable - initial_length: Starting length (mm)
- - final_length: Ending length (mm)
- - days: Duration of experiment

Part 1: Data Loading and Preparation

1.1 Load and Clean the Data

```
cray_df <- read_csv("data/sargent_lodge_crayfish.csv")
```

```
Rows: 84 Columns: 9
── Column specification ───────────────────────────────────────────
```

```
Delimiter: ","
chr (4): range, lake, collection_loc, mat_id
dbl (5): initial_length, final_length, days, growth_per_day, avg_temp

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Part 2: Statistical Analysis Setup

2.1 Analysis Type and Model

Type of Analysis: Linear Mixed Effects Model (Hierarchical/Nested Design)

Model Equation: Growth Rate_ij = β_0 + β_1(Range_i) + u_j + ε_ij

Where:

- - Growth Rate_ij = daily growth rate for individual i in lake j
- - β_0 = fixed intercept (overall mean)
- - β_1 = fixed effect of range (Native vs Invasive)
- - u_j ~ $N(0, \sigma^2\_lake)$ = random effect of lake j
- - ε_ij ~ $N(0, \sigma^2\_error)$ = residual error for individual i in lake j

Hypotheses:

- *Fixed Effect*
  - *- Range:*
    - - H_0: $\beta_1 = 0$ (no difference in growth between ranges)
    - - H_1: $\beta_1 \neq 0$ (difference exists between ranges)
- *Random Effect*
  - *- Lake:* - Accounts for correlation within lakes and lake-to-lake variability

Variables:

- - Response: growth_per_day (continuous, mm/day)
- - Fixed Effect: range (categorical, 2 levels: Native, Invasive)
- - Random Effect: lake (categorical, 3 levels: Big, High, Papoose)
- - Level 1: Individual crayfish (n = 84)
- - Level 2: Lakes (n = 3)

Biological Rationale: Individual crayfish within the same lake are likely to be more similar to each other than to crayfish in different lakes due to shared environmental conditions. The mixed effects model accounts for this hierarchical structure while testing for range differences.

Part 3: Exploratory Data Analysis

3.1 Summary Statistics

3.2 Exploratory Visualizations

Part 4: Mixed Effects Model Analysis

4.1 Fit the Mixed Effects Model

Part 5: Assumption Testing

5.1 Check Mixed Effects Model Assumptions

```
# check_model_result <- check_model(growth_mixed_model)
# plot(check_model_result)
```

5.2 Formal Assumption Tests

```
# shapiro_residuals_result <- shapiro.test(residuals(growth_mixed_model))
# shapiro_residuals_result
#
# shapiro_random_result <- shapiro.test(random_effects_pred_df$random_intercept)
# shapiro_random_result
```

```
# check_homogeneity_result <- check_homogeneity(growth_mixed_model)
# check_homogeneity_result
```

Part 7: Publication Figure

7.1 Create Publication-Quality Figure

Submission Guidelines

What to turn in -

1. a quarto markdown file and dataframe if you modified the original. All of the code should be able to run with what you turn in. **(2 points)**

2. a self-contained html file showing the code and output **(2 points)**

3. annotations in the quarto file that shows or tells what is being done in the r code chunks describing what you are trying to do - credit will be given even if it does not work as long as you detail what you are doing. As we start to move into more statistics you will be expected to interpret the results. **(2 points)**

Points

- summary stats - 2 point
- assumptions and hypotheses - 3 points
- exploratory graphs - 2 point
- interpretation - 4 points
- Final figure - 1 point
- Results - 2 points