

07_homework_one_way_anova

Bill Perry

```
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() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(janitor)
```

```
Attaching package: 'janitor'

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

  chisq.test, fisher.test
```

```
library(readxl)
library(car)
```

```
Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:dplyr':

  recode

The following object is masked from 'package:purrr':

  some
```

```
library(broom)
library(ggplot2)

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 one-way ANOVA.

Learning Objectives

By completing this assignment, you will be able to:

1. Understand one-way ANOVA concepts and applications
2. Perform exploratory data analysis for group comparisons
3. Conduct one-way ANOVA analysis
4. Test statistical assumptions for ANOVA
5. Interpret ANOVA results and effect sizes
6. Create publication-quality figures
7. 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 northern Wisconsin lakes during summer 2011.

Key variables:

- - range: Population origin (Native vs Invasive)
 - - predictor variable - growth_per_day: Daily growth rate (mm/day)
 - - response variable - lake: Lake location (Big, High, Papoose)
 - - 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

```
# df <- df %>%
#   mutate(
#     range = factor(range, levels = c("Native", "Invasive")),
#     lake = factor(lake, levels = c("Big", "High", "Papoose"))
#   )
#
# head(cray_df)
```

Part 2: Statistical Analysis Setup

2.1 Analysis Type and Model

We are going to do a one way ANOVA on invasive and native growth rates. I am interested in seeing this overall... Is it the right test? Whats wrong?

Type of Analysis: One-way ANOVA

Model Equation: Growth Rate = $X + X + X$

Where:

- - Response variable
- - μ
- - Range
- - ε

Hypotheses:

- - H_0 :
- - H_1 :

Variables:

- - Response: growth_per_day (continuous, mm/day)
- - Factor: range (categorical, 2 levels: Native, Invasive)

Biological Rationale: We expect invasive populations to

Part 3: Exploratory Data Analysis

3.1 Summary Statistics

```
# cray_df %>%  
#   group_by(range) %>%  
#   skim()
```

3.2 Exploratory Visualizations

Part 4: One-Way ANOVA Analysis

4.1 Fit the ANOVA Model

Note using a Type 3 Sums of Squares shows a slight difference due to the unbalanced design

4.2 Effect Size Calculation

Eta-squared represents the **proportion of total variance explained** by the factor (range).

- Formula: $\eta^2 = SS_{\text{between}} / SS_{\text{total}}$
- Range: 0 to 1
- Interpretation: If $\eta^2 = 0.21$, then 21% of the variance in growth rate is explained by population range

Omega-squared is a **less biased estimate** of effect size than eta-squared.

- Formula: $\omega^2 = (SS_{\text{between}} - df_{\text{between}} \times MS_{\text{within}}) / (SS_{\text{total}} + MS_{\text{within}})$
- Range: 0 to 1 (but can be slightly negative)
- More conservative than η^2 because it adjusts for bias in small samples

Why Calculate Both?

- **Eta-squared (η^2)**: Easier to calculate and interpret, but slightly **overestimates** effect size
- **Omega-squared (ω^2)**: More accurate, **unbiased estimate** of population effect size

Effect Size Interpretation Guidelines:

Effect Size	η^2 / ω^2	Interpretation
Small	0.01	1% of variance explained
Medium	0.06	6% of variance explained
Large	0.14	14% of variance explained

Example Output Interpretation:

If your results show:

```
eta_squared omega_squared
1          0.21          0.20
```

This means:

- **21%** of the variance in crayfish growth rate is explained by population range (η^2)
- **20%** is the unbiased estimate of variance explained (ω^2)
- This represents a **large effect size** (much larger than 0.14)
- Population range is a strong predictor of growth rate

Bottom line: Both metrics tell you how much of the differences in crayfish growth can be attributed to whether they're from native vs. invasive populations, with omega-squared being the more conservative (and accurate) estimate.

Part 5: Assumption Testing

5.1 Check ANOVA Assumptions

The easy way

```
# # Create diagnostic plots
# par(mfrow = c(2, 2))
# plot(growth_anova_model)
# par(mfrow = c(1, 1))
```

5.2 Formal Assumption Tests

```
# shapiro_test_result <- shapiro.test(residuals(model))
# shapiro_test_result
```

```
# levene_test_result <- leveneTest(response ~ factor, data = cradfy_df)
# levene_test_result
```

Part 6: Publication Figure

6.1 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