M9 Problem Set: ANOVA for TAs

Pablo E. Gutiérrez-Fonseca

2024-03-26 13:40:22

We collected freshly laid frogspawn from a pond nestled in the Italian Alps and transported them back to our laboratory. Upon arrival, we partitioned them into 60 water containers. Specifically, 20 of these containers maintained a temperature of 13°C, while another 20 were set to 18°C. The remaining 20 containers were regulated at 25°C. By employing a substantial number of replicates, we bolster our confidence in attributing any anticipated disparities between groups to the variable under scrutiny, namely temperature.

1. What is the response variable?
2. What is the explanatory variable?
3. Perform a one-way ANOVA to assess your research question. It's important to note that Temperature should be recoded from a numerical variable to a factor (categorical variable) with three levels: "13", "18", and "25". These numbers represent the distinct categories of our explanatory variable, rather than actual count data. Therefore, we need to transform Temperature from numerical to a factor variable. Assume that your data are normally distributed. 3.1 Report the F value for this test. 3.2 Report the p-value for this test. 3.3 Are there differences in hatching time between frogs exposed to different temperatures? 3.4 What is the temperature at which the frogspawn takes the longest time to hatch?
4. Summarize: Write a concise one paragraph summary of this analysis. Remember that any summary should include the following:

a. Statement of the research hypothesis or study objectives.

b. Brief summary of methods (one sentence or less).

c. Statement of the statistical results (including type of test and shorthand: t(df)= obtained value, p = 0.xxx).

d. Description of any differences, if meaningful, along with an interpretation of why these results make sense (or don't make sense).

1. What is the response variable? Our response variable is Hatching_time.
2. What is the explanatory variable? Our explanatory variable is Temperature, with 3 levels: 13°C, 18°C and 25°C.

#1. Import libraries and load packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'tibble' was built under R version 4.3.1

## Warning: package 'tidyr' was built under R version 4.3.2

## Warning: package 'readr' was built under R version 4.3.2

## Warning: package 'purrr' was built under R version 4.3.1

## Warning: package 'dplyr' was built under R version 4.3.2

## Warning: package 'stringr' was built under R version 4.3.2

## Warning: package 'forcats' was built under R version 4.3.2

## Warning: package 'lubridate' was built under R version 4.3.2
```

```
library(dplyr)
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.1
```

#2. importing our data

```
frogs_messy_data <- read.csv("frogs_messy_data.csv")
```

```
frogs_tidy_data <- gather(frogs_messy_data, Temperature, Hatching_time, c(2:4)) %>%
                   # Hatching times (value) to be gathered by Temperature (key)
                   mutate(Temperature = parse_number(Temperature)) %>%
                   # To get rid of the non-numerical part
                   select("Hatching_time", "Temperature") %>%
                   # Keeping only the columns we need for the analysis
                   na.omit()
                   # To get rid of missing values (NAs)
# write.xlsx(frogs_tidy_data, file = "frogs_tidy_data.xlsx")
```

```
frogs_tidy_data$Temperature <- as.factor(as.character(frogs_tidy_data$Temperature))
# Makes temperature into factor variable

frogs_anova <- aov(Hatching_time ~ Temperature, data = frogs_tidy_data)
summary(frogs_anova)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Temperature  2 1020.9   510.5   385.9 <2e-16 ***
## Residuals   57   75.4     1.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References:
https://ourcodingclub.github.io/tutorials/anova/#anova