M7 Problem Set: Independent t-test for TAs

Pablo E. Gutierrez Fonseca

2024-05-06 16:56:00

We have a data set from a former student's biology thesis (I LOVE it when students share their data!).

butterflies_pesticides.xlsx reports the mean wing length for 80 butterflies (40 of which were treated with pesticides during the larval stage (labeled Group 2) and 40 that were not (control group labeled 1))

Our task will be to examine how pesticides impact butterfly development. Let's start with the Butterfly Pesticides Data.

Assume that we want to test to see if there is a difference between butterfly wing length for those treated with pesticides (group 2) and those not treated with pesticides (group 1).

Use an alpha level of 0.05

1. Report the p-value for the goodness of fit test for the wing length data.
2. Is your wing length data normally distributed?
3. Conduct the appropriate test and report the obtained value for this test.
4. Report the appropriate p-value for this test.
5. Is there a significant difference between the pesticide treated and control group wing length?
6. Is this difference meaningful? How do you know?
7. 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. Import libraries and load packages

```r
library(tidyverse)
library(lessR)
library(readxl)
library(BSDA)
```

#2. importing our data

```r
butterflies <- read_excel("butterflies_pesticides.xlsx")
```

1. Report the p-value for the goodness of fit test for the wing length data

```r
shapiro.test(butterflies$`Wing Length`)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  butterflies$‘Wing Length‘
## W = 0.97134, p-value = 0.06913
```

2. Is your wing length data normally distributed?
   Yes
3. Conduct the appropriate test and report the obtained value for this test.
4. Report the appropriate p-value for this test.
5. Is there a significant difference between the pesticide treated and control group wing length?

```r
# 6. Test for equal variance
var.test(`Wing Length` ~ Group, data=butterflies)
```

```
##
##  F test to compare two variances
##
## data:  Wing Length by Group
## F = 0.40274, num df = 39, denom df = 39, p-value = 0.005512
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2130078 0.7614644
## sample estimates:
## ratio of variances
##           0.402738
```

```r
# we would conclude that the variances are not equal between the groups of butterflies.
```

```r
t.test(`Wing Length` ~ Group, data=butterflies,
       alternative = c("two.sided"),
       mu = 0,
       paired = FALSE,
       var.equal = FALSE,
       conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  Wing Length by Group
## t = 3.9832, df = 66.029, p-value = 0.0001721
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  19.29467 58.07713
## sample estimates:
## mean in group 1 mean in group 2
##        632.1756        593.4897
```

6. Is this difference meaningful? How do you know?

```r
mean_A <- mean(butterflies$`Wing Length`[butterflies$Group == 1])
mean_B <- mean(butterflies$`Wing Length`[butterflies$Group == 2])

sd_after <- sd(butterflies$`Wing Length`)

# Calculate Cohen's d
abs(mean_A - mean_B)/sd_after
```

```
## [1] 0.8170949
```