PS: Indpendent and Dependent t-test for TAs

Pablo E. Gutiérrez-Fonseca

2023-08-11

R practice.

Install packages.

```
library(car)
```

```
## Loading required package: carData
```

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Load the water pollution data into R.

```
##   Birds    Species Gender  Treatment value
## 1     1 Hylocichla      M Pre_weight    46
## 2     2 Hylocichla      M Pre_weight    44
## 3     3 Hylocichla      M Pre_weight    48
## 4     4 Hylocichla      M Pre_weight    42
## 5     5 Hylocichla      M Pre_weight    51
## 6     6 Hylocichla      M Pre_weight    51
```

Make a subgroup of Weight before migration (Pre_weight)

```r
df_bird_Pre_weight <- df_bird_all %>%
  filter(Treatment == 'Pre_weight')
```

```r
summary(df_bird_Pre_weight)
```

```
##      Birds          Species             Gender           Treatment
##  Min.   : 1.00   Length:28          Length:28          Length:28
##  1st Qu.: 7.75   Class :character   Class :character   Class :character
##  Median :14.50   Mode  :character   Mode  :character   Mode  :character
```

```
##  Mean   :14.50
##  3rd Qu.:21.25
##  Max.   :28.00
##      value
##  Min.   :41.00
##  1st Qu.:46.75
##  Median :49.00
##  Mean   :48.68
##  3rd Qu.:52.00
##  Max.   :55.00
```

Part I.
2. Analyze data distribution. Is your pre-weight data normally distributed?

```
shapiro.test(df_bird_Pre_weight$value)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_bird_Pre_weight$value
## W = 0.95557, p-value = 0.2725
```

3. Test to see if you meet the assumption of equal variance for this test. Based on the results of this test for unequal variance, which test should you use?

```
# Perform Levene's test
result <- leveneTest(value~Gender, data=df_bird_Pre_weight)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
print(result)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.0061 0.9381
##       26
```

```
p_value <- result$Pr[1]
# Significance level
alpha <- 0.05

# Interpret the result
if (p_value < alpha) {
  cat("Reject null hypothesis: Variances are significantly different.\n")
} else {
  cat("Fail to reject null hypothesis: Variances are not significantly different.\n")
}
```

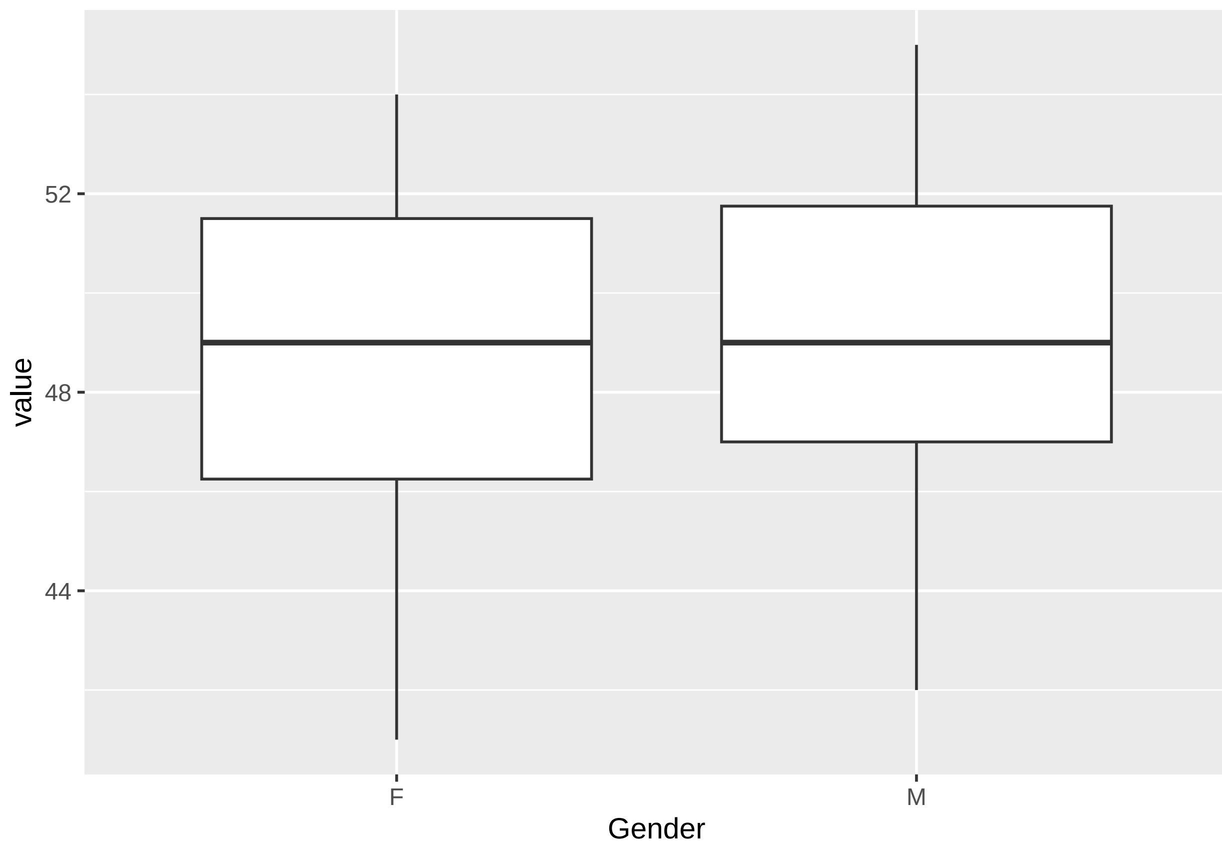
```
## Fail to reject null hypothesis: Variances are not significantly different.
```

4. Run this analysis. What is the test statistic (obtained) value for the test you selected?
Use alternative to determine your hypothesis. Default= "two.sided".
alternative = c("two.sided", "less", "greater"),

```
# Perform t-test
t_test_result <- t.test(value~Gender,alternative = "two.sided",
                        var.equal=TRUE, data=df_bird_Pre_weight)
print(t_test_result)
```

```
##
##  Two Sample t-test
##
## data:  value by Gender
## t = -0.78654, df = 26, p-value = 0.4387
## alternative hypothesis: true difference in means between group F and group M is not equal to 0
## 95 percent confidence interval:
##  -4.387692  1.959120
## sample estimates:
## mean in group F mean in group M
##        48.07143        49.28571
```

```
# Create a boxplot plot using ggplot2
ggplot(df_bird_Pre_weight, aes(x = Gender, y = value)) +
  geom_boxplot()  # Add jitter for better visualization
```



Part II.
Analyze data distribution. Is your pre-weight data normally distributed?

```
shapiro.test(df_bird_Pre_weight$value)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_bird_Pre_weight$value
## W = 0.95557, p-value = 0.2725
```

Test to see if you meet the assumption of equal variance for this test. Based on the results of this test for unequal variance, which test should you use?

```
# Perform Levene's test
result <- leveneTest(value~Species, data=df_bird_Pre_weight)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
print(result)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  1.2128 0.2809
##       26
```

```r
p_value <- result$Pr[1]
# Significance level
alpha <- 0.05

# Interpret the result
if (p_value < alpha) {
  cat("Reject null hypothesis: Variances are significantly different.\n")
} else {
  cat("Fail to reject null hypothesis: Variances are not significantly different.\n")
}
```

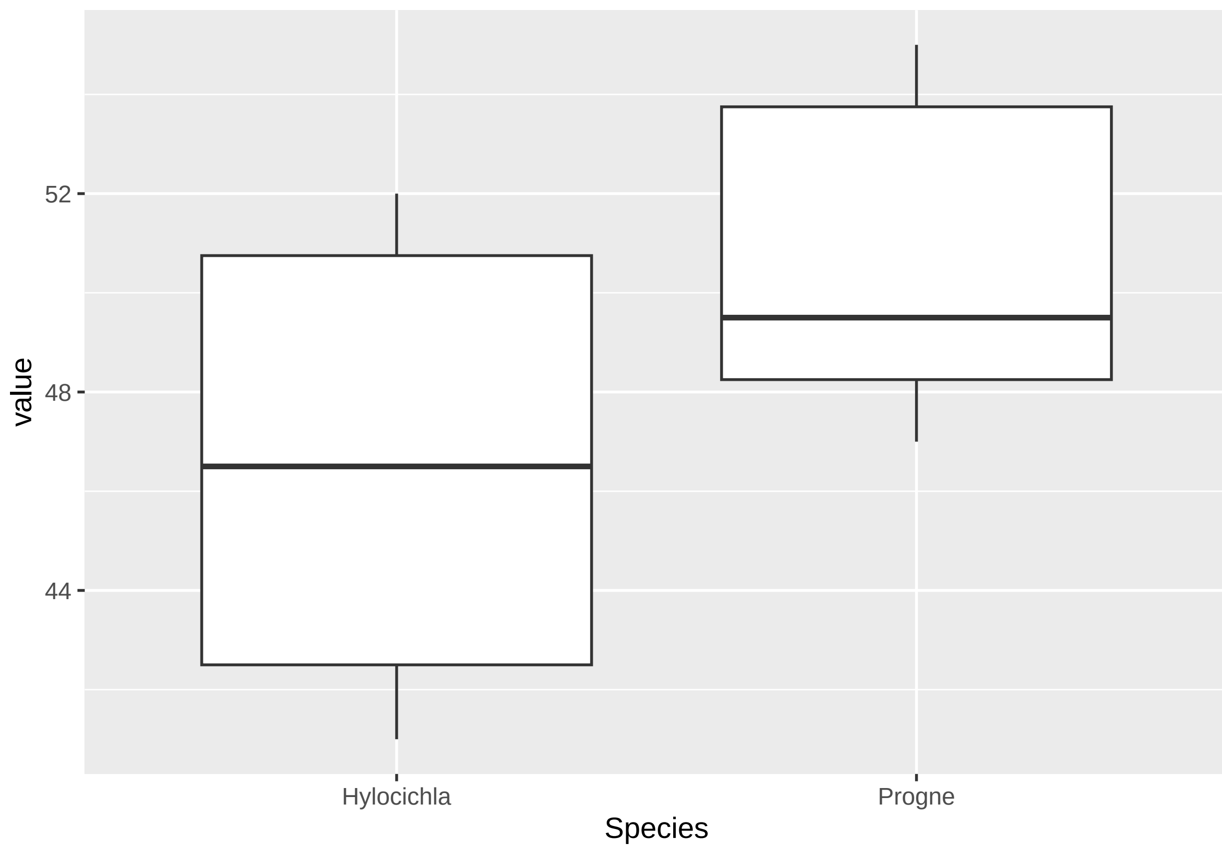
```
## Fail to reject null hypothesis: Variances are not significantly different.
```

8. Run this analysis. What is the test statistic (obtained) value for the test you selected?
Use alternative to determine your hypothesis. Default= "two.sided".
alternative = c("two.sided", "less", "greater"),

```r
# Perform t-test
t_test_result <- t.test(value~Species, alternative = "two.sided",
                        var.equal=TRUE, data=df_bird_Pre_weight)
print(t_test_result)
```

```
##
##  Two Sample t-test
##
## data:  value by Species
## t = -2.891, df = 26, p-value = 0.007655
## alternative hypothesis: true difference in means between group Hylocichla and group Progne is not equ
## 95 percent confidence interval:
##  -6.721783 -1.135360
## sample estimates:
## mean in group Hylocichla     mean in group Progne
##                 46.71429                 50.64286
```

```r
# Create a boxplot plot using ggplot2
ggplot(df_bird_Pre_weight, aes(x = Species, y = value)) +
  geom_boxplot()  # Add jitter for better visualization
```



Part III. 12. Analyze data distribution. Is your pre-weight data normally distributed?

```
shapiro.test(df_bird_all$value)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_bird_all$value
## W = 0.97528, p-value = 0.3028
```

Test to see if you meet the assumption of equal variance for this test. Based on the results of this test for unequal variance, which test should you use?

```
# Perform Levene's test
result_paired_t_test <- leveneTest(value~Treatment, data=df_bird_all)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
print(result_paired_t_test)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.0035 0.9529
##       54
```

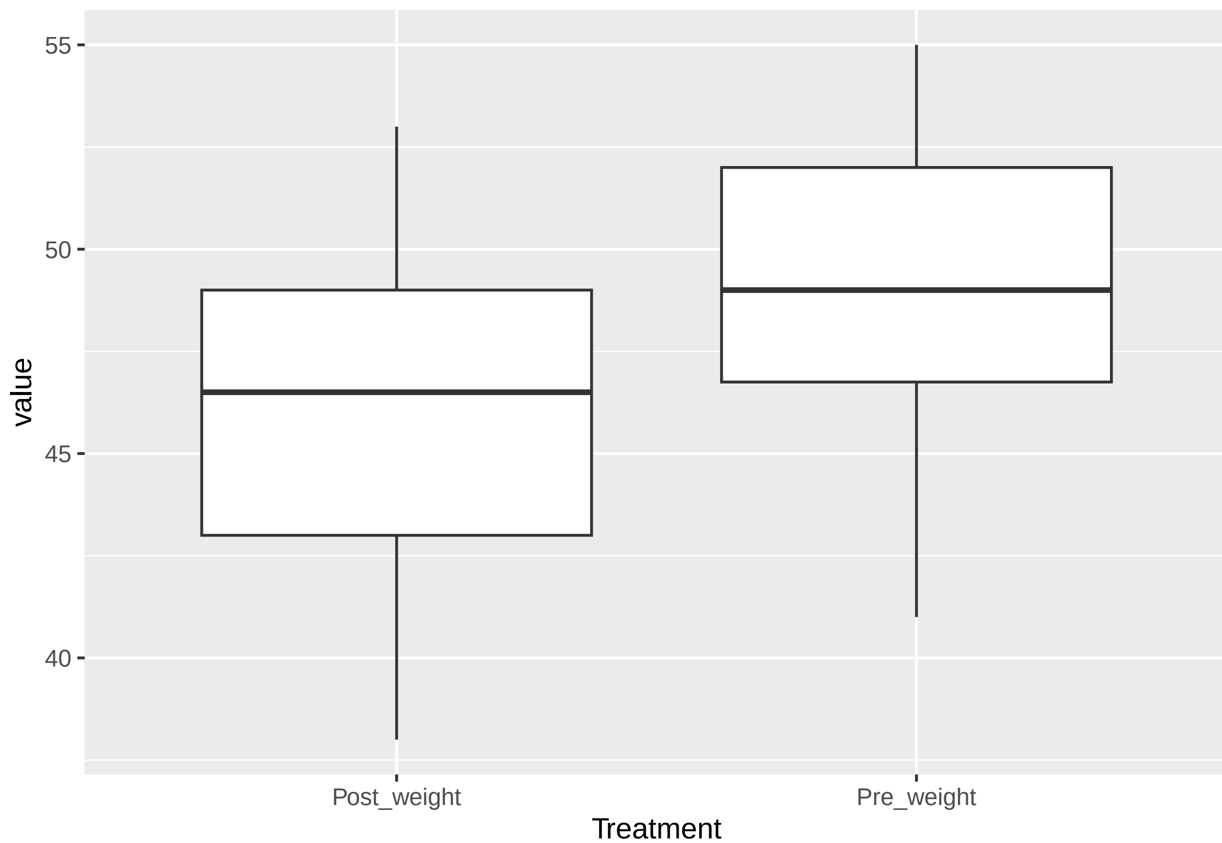
```r
p_value <- result_paired_t_test$Pr[1]
# Significance level
alpha <- 0.05

# Interpret the result
if (p_value < alpha) {
  cat("Reject null hypothesis: Variances are significantly different.\n")
} else {
  cat("Fail to reject null hypothesis: Variances are not significantly different.\n")
}
```

```
## Fail to reject null hypothesis: Variances are not significantly different.
```

```r
t_paired_result <- t.test(value ~ Treatment, paired = TRUE,
                          data = df_bird_all)
print(t_paired_result)
```

```
##
##  Paired t-test
##
## data:  value by Treatment
## t = -3.4184, df = 27, p-value = 0.002013
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.400616 -1.099384
## sample estimates:
## mean of the differences
##                   -2.75
```

```r
# Create a boxplot plot using ggplot2
ggplot(df_bird_all, aes(x = Treatment, y = value)) +
  geom_boxplot()  # Add jitter for better visualization
```



Part IV. From the data.frame make two groups, make two groups (females and males)

```
df_bird_female <- df_bird_all %>%
  filter(Gender == 'F')
```

```
df_bird_male <- df_bird_all %>%
  filter(Gender == 'M')
```

Assumptions for Females

```
shapiro.test(df_bird_female$value)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_bird_female$value
## W = 0.95617, p-value = 0.282
```

```
leveneTest(value~Treatment, data=df_bird_female)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##       Df F value Pr(>F)
## group  1  0.0053 0.9424
##       26
```

```r
t_paired_female <- t.test(value ~ Treatment, paired = TRUE,
                          alternative= 'less',
                          data = df_bird_female)
print(t_paired_female)
```

```
##
##  Paired t-test
##
## data:  value by Treatment
## t = -1.9868, df = 13, p-value = 0.03422
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.1164104
## sample estimates:
## mean of the differences
##               -1.071429
```

Assumptions for Males

```r
shapiro.test(df_bird_male$value)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df_bird_male$value
## W = 0.97278, p-value = 0.6567
```

```r
leveneTest(value~Treatment, data=df_bird_male)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1   0.334 0.5683
##       26
```

```r
t_paired_male <- t.test(value ~ Treatment, paired = TRUE,
                        alternative= 'less',
                        data = df_bird_male)
print(t_paired_male)
```

```
##
##  Paired t-test
##
## data:  value by Treatment
```

```
## t = -3.1601, df = 13, p-value = 0.003762
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -1.946804
## sample estimates:
## mean of the differences
##               -4.428571
```