Comparing the means of two groups

Rida Kanatbekovna Azhigulova, R lab 5, BIO/MTH-245 - Applied Statistics for the Life Sciences

2025-02-24

Contents

Introduction

This is an R Lab 5 to compare the means of two groups. To solve the problems (#1, #2 and #3) I used bumpus, countries and leg shaving csv provided by Professor.

Analysis of Bumpus Data

```
# Load necessary libraries
library(ggplot2)
library(dplyr)
library(car)  # For Levene's test
```

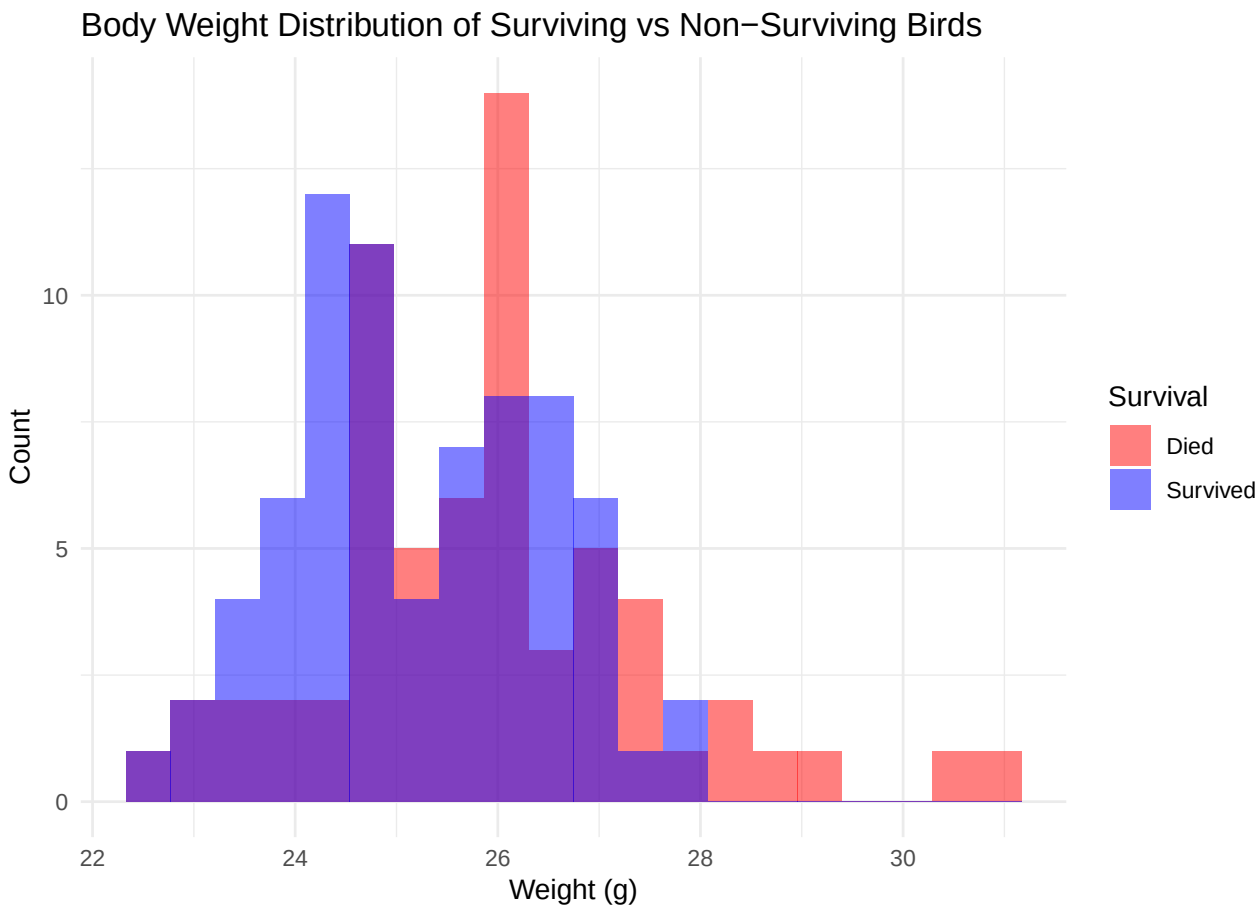
```
# Load the dataset
bumpus <- read.csv("./bumpus.csv", stringsAsFactors = FALSE)
```

```
# Summary statistics by survival status
bumpus %>%
  group_by(survival) %>%
  summarise(mean_weight = mean(weight_g, na.rm = TRUE),
            sd_weight = sd(weight_g, na.rm = TRUE),
            n = n())
```

```
## # A tibble: 2 x 4
##   survival mean_weight sd_weight     n
##   <chr>          <dbl>     <dbl> <int>
## 1 died            25.9      1.63    64
## 2 survived        25.2      1.26    72
```

```
# Plot histogram of body weight distribution
ggplot(bumpus, aes(x = weight_g, fill = as.factor(survival))) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 20) +
  scale_fill_manual(values = c("red", "blue"), labels = c("Died", "Survived")) +
  labs(title = "Body Weight Distribution of Surviving vs Non-Surviving Birds",
       x = "Weight (g)",
       y = "Count",
       fill = "Survival") +
```

```
  theme_minimal()
```



Interpretation: The histogram compares the body weight distributions of surviving (blue) and non-surviving (red) English house sparrows. The distributions show some overlap but also key differences. Surviving birds tend to have a higher concentration of weights around 24-26 grams, whereas non-surviving birds exhibit a slightly broader distribution with a peak near 26 grams and more individuals in the higher weight range (above 28 grams). The shape suggests that while weight might have influenced survival, other factors could also be at play. The distributions appear approximately normal, though some skewness is present. To confirm differences in means and variance, statistical tests like Welch's t-test and Levene's test should be performed.

```
# Perform and print t-test for weight by survival status
t_test_result <- t.test(weight_g ~ survival, data = bumpus, var.equal = FALSE)
print(t_test_result)
```

```
##
##  Welch Two Sample t-test
##
## data:  weight_g by survival
## t = 2.5161, df = 117.88, p-value = 0.01321
## alternative hypothesis: true difference in means between group died and group survived is not equal t
## 95 percent confidence interval:
##  0.1351376 1.1339597
## sample estimates:
##     mean in group died mean in group survived
##               25.86094               25.22639
```

```r
# Levene's test for equality of variances
leveneTest(weight_g ~ as.factor(survival), data = bumpus)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   1  0.8256 0.3652
##       134
```

Analysis of Countries Data

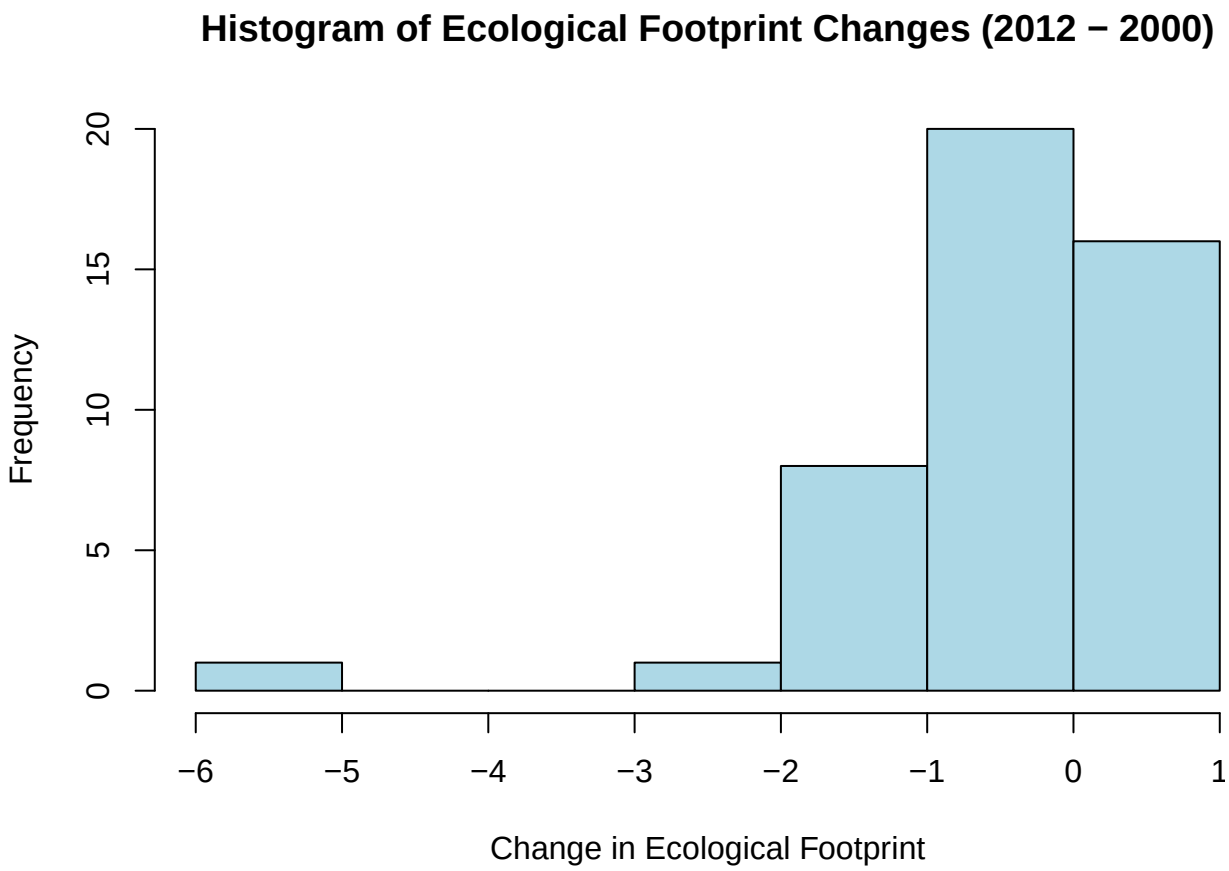
```r
# Load dataset for countries
countries <- read.csv("countries.csv")
```

```r
# Create a new column for differences
countries$difference <- countries$ecological_footprint_2012 - countries$ecological_footprint_2000

# Summary statistics of the difference
summary(countries$difference)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  -5.240  -0.545  -0.190  -0.417   0.265   0.670     150
```

```r
# Plot histogram of ecological footprint changes
hist(countries$difference,
     main = "Histogram of Ecological Footprint Changes (2012 - 2000)",
     xlab = "Change in Ecological Footprint",
     col = "lightblue",
     border = "black")
```



Interpretation: The histogram displays the distribution of changes in ecological footprint between 2000 and 2012. Most countries show small decreases in their ecological footprint, with the majority of values concentrated around -1 and 0. A few countries exhibit larger reductions (e.g., -5 or -6), but there are no substantial increases. This suggests that, overall, ecological footprints have either remained stable or slightly declined over time. A paired t-test would help determine whether the observed changes are statistically significant.

```r
# Perform and print paired t-test for ecological footprint data
print(t.test(countries$ecological_footprint_2012,
             countries$ecological_footprint_2000,
             paired = TRUE))
```

```
##
##  Paired t-test
##
## data:  countries$ecological_footprint_2012 and countries$ecological_footprint_2000
## t = -2.7553, df = 45, p-value = 0.008434
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.7217469 -0.1121662
## sample estimates:
## mean difference
##      -0.4169565
```

Analysis of Leg Shaving Data

```r
# Load dataset for leg shaving experiment
leg_shaving <- read.csv("leg shaving.csv")
```

```r
# Check the structure of the leg shaving dataset
str(leg_shaving)
```

```
## 'data.frame':    5 obs. of  5 variables:
##  $ person                    : chr  "Z" "F" "M" "S" ...
##  $ hair_width_change_control : int  -3 0 10 10 -11
##  $ hair_width_change_test    : int  -1 -8 9 -3 -35
##  $ hair_growth_change_control: num  -0.02 -0.01 -0.02 0.01 0.06
##  $ hair_growth_change_test   : num  -0.02 -0.03 0.02 -0.03 0.04
```

```r
# Summary statistics for leg shaving dataset
summary(leg_shaving)
```

```
##     person          hair_width_change_control hair_width_change_test
##  Length:5           Min.   :-11.0             Min.   :-35.0
##  Class :character   1st Qu.: -3.0             1st Qu.: -8.0
##  Mode  :character   Median :  0.0             Median : -3.0
##                     Mean   :  1.2             Mean   : -7.6
##                     3rd Qu.: 10.0             3rd Qu.: -1.0
##                     Max.   : 10.0             Max.   :  9.0
##  hair_growth_change_control hair_growth_change_test
##  Min.   :-0.020             Min.   :-0.030
##  1st Qu.:-0.020             1st Qu.:-0.030
##  Median :-0.010             Median :-0.020
##  Mean   : 0.004             Mean   :-0.004
##  3rd Qu.: 0.010             3rd Qu.: 0.020
##  Max.   : 0.060             Max.   : 0.040
```

```r
# Perform and print paired t-test for hair width changes (shaved vs. unshaved legs)
# Perform paired t-test
ttest_result <- t.test(leg_shaving$hair_width_change_test,
                       leg_shaving$hair_width_change_control,
                       paired = TRUE)

# Tidy the t-test results using broom
ttest_tidy <- tidy(ttest_result)

# Print the results in a nicely formatted table
knitr::kable(ttest_tidy, caption = "Paired t-test Results for Hair Width Changes")
```

Table 1: Paired t-test Results for Hair Width Changes

estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
-8.8	-1.904958	0.1294963	4	-21.62585	4.025854	Paired t-test	two.sided

Interpretation: The p-value suggests no significant difference in hair thickness due to shaving. Confidence interval includes zero, reinforcing the lack of significant effect.