

# Lab 9 – Hypothesis Testing

2024-11-11

## Contents

```
t.test(mtcars$mpg, conf.level = 0.9)

##
## One Sample t-test
##
## data: mtcars$mpg
## t = 18.9, df = 31, p-value <0.0000000000000002
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  18.284 21.897
## sample estimates:
## mean of x
##    20.091
```

```
t.test(x = mtcars$mpg,
       mu = 18,
       conf.level = 0.9)

##
## One Sample t-test
##
## data: mtcars$mpg
## t = 1.96, df = 31, p-value = 0.059
## alternative hypothesis: true mean is not equal to 18
## 90 percent confidence interval:
##  18.284 21.897
## sample estimates:
## mean of x
##    20.091
```

In this example, note that our point estimate did not change, nor did our 90% confidence interval. What did change, however, was the calculated t-statistic as well as the p-value.

**Question 1:** Explain why both the t-statistic and the p-value changed in our second call of `t.test()` while the point estimate and the confidence interval stayed the same.

---

The t-statistic and p-value both changed because we specified  $\mu_0 = 18$  in the second call. Both of these values are related to our null hypothesis; changing these will change them as well

---

**Question 2:** Using the subsets of the data below, find and report 95% confidence intervals for total enrollment at public and private colleges.

```
library(dplyr)
college <- read.csv("https://collinn.github.io/data/college2019.csv")

col_priv <- filter(college, Type == "Private")
col_pub <- filter(college, Type == "Public")
```

---

```
## Not expected to know how to extract this, but you can get the
## information from the t.test function OR by computing it directly
## with Estimate +/- MOE
t.test(col_priv$Enrollment)[["conf.int"]]
```

```
## [1] 2443.4 2997.3
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(col_pub$Enrollment)[["conf.int"]]
```

```
## [1] 10388 12262
## attr(,"conf.level")
## [1] 0.95
```

---

**Question 3:** Below is code to recreate the two subsets of the hawk data containing only Red-tailed hawks

```
## RT Hawk data
hawks <- read.csv("https://collinn.github.io/data/hawks.csv")
hawks <- filter(hawks, Species == "RT")

## Subset of RT Hawk data
set.seed(89)
idx <- sample(seq_len(nrow(hawks)), size = 20)
hawks2 <- hawks[idx, ]
```

- **Part A:** Find the mean body weight for Red-tailed hawks in both the `hawks` and `hawks2` datasets. How do they compare?
- **Part B:** Using the average weight of the Cooper's hawks, perform a t-test on each dataset to test the hypothesis that  $\mu_0 = 1050$ . If you were testing at the  $\alpha = 0.05$  level, which would you conclude?
- **Part C:** Explain why the conclusions you came to were different in Part C.

---

```
## Part A
mean(hawks$Weight)
```

```
## [1] 1094.7
```

```
mean(hawks2$Weight)
```

```
## [1] 1094.7
```

```
## Part B
t.test(hawks$Weight, mu = 1050) # would reject null
```

```
##
## One Sample t-test
##
```

```
## data: hawks$Weight
## t = 5.63, df = 568, p-value = 0.000000028
## alternative hypothesis: true mean is not equal to 1050
## 95 percent confidence interval:
## 1079.1 1110.3
## sample estimates:
## mean of x
## 1094.7
```

```
t.test(hawks2$Weight, mu = 1050) # would not reject null
```

```
##
## One Sample t-test
##
## data: hawks2$Weight
## t = 1.13, df = 19, p-value = 0.27
## alternative hypothesis: true mean is not equal to 1050
## 95 percent confidence interval:
## 1011.7 1177.7
## sample estimates:
## mean of x
## 1094.7
```

```
## Part C
# Conclusions different because different sample size (n = 19 v n = 568)
```