

# Lab2

2024-09-06

STAT 5000LAB #2

FALL 2024 DUE TUE SEP 10TH NAME: SAM OLSON

**Directions:** Complete the exercises below. When you are finished, turn in any required files online in Canvas, then check-in with the Lab TA for dismissal.

## Introduction to t-Tests in R

Refer to the `fuel_economy.csv` data file posted in Canvas. This data set has information about an observational study of automobiles driven in Canada, including the following two columns:

**Cylinders:** category variable with two levels - 4 or 6

**Consumption:** numeric response variable with the fuel consumption in miles per gallon (mpg)

Researchers are interested in exploring whether there is a difference in the average fuel consumption of vehicles with engines built using differing numbers of cylinders. The code to conduct a two-sample t-test in R is explained below. The full R program is provided in the file `fuel_economy_Lab2.R` posted on Canvas.

- First, load in the data using the *Import Dataset* tool in R Studio. Be sure to change the variable type on the Cylinders column to “factor” and enter “4, 6” as the levels.

```
library(readr)
fuel <- read_csv("fuel_economy.csv",
                 col_types=cols(Cylinders=col_factor(levels=c("4", "6"))))
View(fuel)
```

- Next, compute the corresponding summary statistics within in group.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v purrr      1.0.2
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
sum_stats = fuel |>
  group_by(Cylinders) |>
  summarize(
    Y_n = n(),
    Y_mean = mean(Consumption.mpg),
    Y_sd = sd(Consumption.mpg)
  )
sum_stats
```

```
## # A tibble: 2 x 4
##   Cylinders   Y_n Y_mean Y_sd
##   <fct>     <int> <dbl> <dbl>
## 1 4         2111   34.1  6.67
## 2 6         1041   24.5  2.55
```

- Then, use the `t.test()` function to conduct a test for the difference in mean fuel consumption between 4 and 6 cylinder vehicles. Indicate the response variable name before the `~` and the category variable name after, use the `data` option to provide the name of the dataset, and use the `var.equal` option set to "TRUE" to indicate the population variances are assumed equal.

```
HT = t.test(Consumption.mpg~Cylinders, data=fuel, var.equal=TRUE)
HT
```

```
##
## Two Sample t-test
##
## data: Consumption.mpg by Cylinders
## t = 44.664, df = 3150, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 4 and group 6 is not equal to 0
## 95 percent confidence interval:
##  9.148005 9.988067
## sample estimates:
## mean in group 4 mean in group 6
##      34.08100      24.51297
```

You can see what pieces of information are stored in the HT variable using the `names()` function. You can access these pieces of information using the `$` operator, e.g.

```
names(HT)
```

```
## [1] "statistic" "parameter" "p.value"    "conf.int"  "estimate"
## [6] "null.value" "stderr"     "alternative" "method"    "data.name"
```

```
HT$null.value
```

```
## difference in means between group 4 and group 6
##                                     0
```

### Assignment

1. State the hypotheses for the two-sided test.

2. From the output, find/compute the difference in the two sample means.

3. From the output, find/compute the estimate of the pooled standard deviation.

4. From the output, find/compute the test statistic for the hypothesis test.

5. From the output, find/compute the degrees of freedom for the test.

6. From the output, find/compute the  $p$ -value for the two-sided hypothesis test.

7. Interpret the results of the two-sided test in the context of the research question.



8. By default, R conducts the two-sided hypothesis test. You can change this by adding the parameter `alternative=greater` or `alternative=less` inside the `t.test()` function. Provide a research question corresponding to either the "greater" or "less" one-sided test.

**Total:** 25 points **# correct:** %: