Assignment 2

STAT3373

Isaac Shoebottom

Sept 25th, 2025

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.2
## v purrr     1.2.0
## -- 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
```

```
dd <- beaver2
```

Question 1

mu_0 = mean temperature when activ = 0

mu_1 = mean temperature when activ = 1

$$H_0 : \mu_0 = \mu_1, H_1 : \mu_0 \neq \mu_1$$

```
t.test(temp ~ activ, data = dd)
```

```
##
##  Welch Two Sample t-test
##
## data:  temp by activ
## t = -18.548, df = 80.852, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.8927106 -0.7197342
## sample estimates:
## mean in group 0 mean in group 1
##        37.09684        37.90306
```

Reject H_0, accept H_1, we conclude that mean temperatures differ by activity level

Now manually:

```r
summary_stats <- dd %>%
  group_by(activ) %>%
  summarise(
    n = n(),
    mean_temp = mean(temp),
    var_temp = var(temp)
  )

summary_stats
```

```
## # A tibble: 2 x 4
##   activ     n mean_temp var_temp
##   <dbl> <int>     <dbl>    <dbl>
## 1     0    38      37.1   0.0431
## 2     1    62      37.9   0.0468
```

Compute Standard Error and t stat

```r
x0 <- summary_stats$mean_temp[summary_stats$activ == 0]
x1 <- summary_stats$mean_temp[summary_stats$activ == 1]

s0 <- summary_stats$var_temp[summary_stats$activ == 0]
s1 <- summary_stats$var_temp[summary_stats$activ == 1]

n0 <- summary_stats$n[summary_stats$activ == 0]
n1 <- summary_stats$n[summary_stats$activ == 1]

SE <- sqrt(s0 / n0 + s1 / n1)

t_stat <- (x0 - x1) / SE
t_stat
```

```
## [1] -18.54789
```

Compute DF

```r
df <- (s0/n0 + s1/n1)^2 /
  ((s0/n0)^2/(n0-1) + (s1/n1)^2/(n1-1))

df
```

```
## [1] 80.85226
```

Compute p-value

```r
p_value <- 2 * pt(-abs(t_stat), df)
p_value
```

```
## [1] 7.269112e-31
```

This p value also matches the conclusion that t.test reaches, reject H_0, accept H_1. We conclude that mean temperatures differ by activity level

Question 2

```r
dd <- iris %>%
  filter(Species %in% c("setosa", "versicolor"))
```

mu_0 = mean Sepal.Length for setosa

mu_1 = mean Sepal.Length for versicolor

$$H_0 : \mu_0 = \mu_1, H_1 : \mu_0 \neq \mu_1$$

```
t.test(Sepal.Length ~ Species, data = dd)
```

```
##
##  Welch Two Sample t-test
##
## data:  Sepal.Length by Species
## t = -10.521, df = 86.538, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group setosa and group versicolor is not equ
## 95 percent confidence interval:
##  -1.1057074 -0.7542926
## sample estimates:
##     mean in group setosa mean in group versicolor
##                    5.006                    5.936
```

p-value < 0.05, reject H_0, accept H_1. This indicates a statistically significant difference in mean Sepal.Length between setosa and versicolor.