HW3

Zhuodiao Kuang

2023-10-09

```
library(MASS)
library(dplyr)
```

Problem 1

Some medical professionals claim that the average weight of American women is 171 pounds. The column lwt holds the mother's weight (in pounds) at last menstrual period, i.e. her pre-pregnancy weight. Use this column for the following questions.

a) Construct a 95% confidence interval of true mean weight of American women

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

So, the interval is

$$\bar{X} - \frac{s}{\sqrt{n}} t_{0.975} \leq \mu \leq \bar{X} + \frac{s}{\sqrt{n}} t_{0.975}$$

```
wt <- birthwt |> pull(lwt)
n <- length(wt)
df = n-1
t <- qt(0.975,df = df)
lq <- mean(wt) - sd(wt)/sqrt(n)*t
uq <- mean(wt) + sd(wt)/sqrt(n)*t
print(c(lq,uq))
```

```
[1] 125.4270 134.2027
```

So, the 95% confidence interval of true mean weight of American women is $[125.4270, 134.2027]$.

b) Interpret the confidence interval. We are 95% confident that the average weight of all American women in the population is between 125.4270 and 134.2027 pounds. This means that if we repeated the same sampling procedure many times and calculated a confidence interval for each sample, about 95% of these intervals would contain the true population mean weight of American women.

c) Comment on the validity of the statement above ("Some medical professionals claim that the average weight of American women is 171 pounds"). In other words, what can we say about this statement given our confidence interval from part a? Given our confidence interval of [125.4270 ,134.2027] for the true mean weight of American women, we can say that the statement that the average weight of American women is 171 pounds is very unlikely to be true. This is because 171 pounds is far outside the range of plausible values for the population mean weight of American women based on our sample data. If the statement were true, it would mean that our sample was not representative of the population or that there was a large sampling error or bias in our data collection process. Therefore, we have strong evidence to reject the claim that the average weight of American women is 171 pounds.

Problem 2

In this data set, we have a variable (smoke) indicating the smoking status of the mothers during pregnancy. Some doctors believe that smoking status is related to weight. Using the columns smoke and lwt, test this claim. (Note: a value of 1 indicates the mother is in the "smoking" group.)

a) Test for the equality of variances between the two groups. (Use a 5% significance level.) Testing the hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 vs H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{s_1^2}{s_2^2} \sim F_{n_1-1,n_2-1} under H_0$$

```
wt1 <- birthwt |> filter(smoke == 1) |> pull(lwt)
wt0 <- birthwt |> filter(smoke == 0) |> pull(lwt)
n1 <- length(wt1)
n0 <- length(wt0)
s1 <- var(wt1)
s0 <- var(wt0)
F = s1/s0

lq <-qf(0.025,n1-1,n0-1)
uq <-qf(0.975,n1-1,n0-1)

F;lq;uq
```

```
[1] 1.412636
```

```
[1] 0.6518345
```

```
[1] 1.50466
```

b) Given your answer from part a, what kind of hypothesis test will you perform? Given that two population variances are equal, we can test the hypothesis that both groups have the same mean(two-sided).

Testing the hypothesis:

$$H_0 : \mu_1 = \mu_2 vs H_1 : \mu_1 \neq \mu_2$$