Homework 3

Mari Sanders

2024-10-12

Problem 1

a)

```
x <-
  birthwt %>%
  pull(lwt)

t.test(x, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 58.362, df = 188, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  125.4270 134.2027
## sample estimates:
## mean of x
##  129.8148
```

b)

We are 95% confident that the true average weight of American Women is between 125.4270 and 134.2027 pounds.

c)

Since 171 is not in our confidence interval, we can reject the hypothesis that the average weight of American women is 171 pounds.

Problem 2

a)

H_0 = The variance of the weight of non-smoking women is the same as the variance of smoking women.

H_a = The variance of the weight of non-smoking women is different than the variance of smoking women.

```r
not_smoke <-
  birthwt %>% select(lwt, smoke) %>% filter(smoke == 0) %>% select(-smoke) %>% pull(lwt)

smoke <-
  birthwt %>% select(lwt, smoke) %>% filter(smoke == 1) %>% select(-smoke) %>% pull(lwt)

var.test(not_smoke, smoke, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  F test to compare two variances
##
## data:  not_smoke and smoke
## F = 0.7079, num df = 114, denom df = 73, p-value = 0.09744
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4614313 1.0651436
## sample estimates:
## ratio of variances
##          0.7078964
```

b)

We should preform a two sample t test with variance, since we get that the p-value is 0.09 we fail to reject the null hypothesis that the variances are equal. This means that the variances are in fact equal.

c)

H_0 : The mean hypertension in smoking and non-smoking women is the same.

H_1 : The mean hypertension in smoking and non-smoking women is different.

```r
t.test(not_smoke, smoke, var.equal = TRUE, conf.level = 0.90)
```

```
##
##  Two Sample t-test
##
## data:  not_smoke and smoke
## t = 0.60473, df = 187, p-value = 0.5461
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -4.785414 10.306448
## sample estimates:
## mean of x mean of y
##  130.8957  128.1351
```

We get a p-value to be 0.5161. This means we fail to reject the null hypothesis that the mean hypertension in smkoking and nonsmoking women is the same.

Problem 3

a)

```r
x <- birthwt %>%
  filter(ht == 1) %>% select(ht) %>%
  sum()
n <- nrow(birthwt)
prop.test(x, n, conf.level = 0.99)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 142.31, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 99 percent confidence interval:
##  0.02926609 0.12892679
## sample estimates:
##          p
## 0.06349206
```

We are 99% confident that the true value of American women that suffer from hypertension is between 0.02926609 and 0.12892679.

Since the p-value is 0.0004 and the confidence interval does not include 0.2, we can reject the null hypothesis that the 20% of American women suffer from hypertension.

b) H_0 = The proportion of women who suffer from hypertension is 20%.

H_a = The proportion of women who suffer from hypertension is less than 20%.

```r
prop.test(x, n, conf.level = 0.9, alternative = "less")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 142.31, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is less than 0.5
## 90 percent confidence interval:
##  0.00000000 0.09324317
## sample estimates:
##          p
## 0.06349206
```

With a p-value less than 2.2e^-16, we can reject the null hypothesis that approximately 20% of American women suffer from hypertension. This means that less than 20% of American women suffer from hypertension.

Problem 4

H_0 : The proportion of uterine irritability for pregnant women who smoke is the same as the proportion of uterine irritability for pregnant women who don't smoke.

H_1 : The proportion of uterine irritability for pregnant women who smoke is different than the proportion of uterine irritability for pregnant women who don't smoke.

```r
non_smoke <-
  birthwt %>%
  filter(smoke == 0)  %>% filter(ui == 1) %>%
  select(ui) %>% pull() %>% length()

smoke <-
  birthwt %>%
  filter(smoke == 1) %>% filter(ui == 1) %>%
  select(ui) %>% pull() %>% length()

both <- c(non_smoke, smoke)
non_smoke_length <-
  birthwt %>%
  filter(smoke == 0) %>% nrow()
smoke_length <-
  birthwt %>%
  filter(smoke == 1) %>% nrow()
n <- c(non_smoke_length, smoke_length)

prop.test(both, n = n, conf.level = 0.99)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  both out of n
## X-squared = 0.41576, df = 1, p-value = 0.5191
## alternative hypothesis: two.sided
## 99 percent confidence interval:
##  -0.1960879  0.1056061
## sample estimates:
##    prop 1    prop 2
## 0.1304348 0.1756757
```

We fail to reject the hypothesis that

Problem 5

a)

An ANOVA test would be the most appropriate because `race` is has more than two groups and `bwt` is continuous.

b)

k > 2 populations of interest, independent samples, errors are normally distributed, variances are the same. These are all valid for our samples.

c) H_0 : Mean birth weight across all means are equal

H_1 : At least two of the mean birth weights are not equal.

```r
summary(aov(race~bwt, data = birthwt))
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## bwt           1   6.01   6.011   7.369 0.00726 **
## Residuals   187 152.54   0.816
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d)

```r
pairwise.t.test(birthwt$bwt,birthwt$race, p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  birthwt$bwt and birthwt$race
##
##   1     2
## 2 0.049 -
## 3 0.029 1.000
##
## P value adjustment method: bonferroni
```

Based on these results, we can see that there is a significant difference between white and black and between white and other race.