# Homework 3

Yuki Joyama

2023-10-23

## Problem 1

Let n = 189, $\bar{X}$ (sample mean) = 129.81, $s$ (sample standard deviation) = 30.58

a) Because we do not know the true population variance $\sigma^2$, 95% confidence interval of true mean weight can be obtained by the following:

$$\bar{X} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$$

Thus, 95% confidence interval is (125.43, 134.2)

b) We are 95% confident that the true mean weight of American women will be between (125.43, 134.2).

c) The statement is not valid because 171 pounds is not included in the 95% confidence interval.

## Problem 2

a) Let $s_1$ and $s_2$ be the sample standard deviation from non-smoking group and smoking group accordingly. And let $n_1$ and $n_2$ be the number of observations in each group.
$\alpha = 0.05$
$H_0$: the variances in two groups are equal
$H_1$: the variances in two groups are not equal

The test statistic is
$F = \frac{s_1^2}{s_2^2} = 0.71 \sim F_{n_1-1,n_2-1}$

Now, $F_{n_1-1,n_2-1,1-\alpha/2} = 1.53 >$ F
Thus, we fail to reject the null hypothesis and accept $H_0$.

b) The answer from part a indicates that the variances between two groups are equal at 5% significance level. Therefore, I will perform two-sample independent t-test for equal variances.

c) $\alpha = 0.10$ Let $\mu_1$ and $\mu_2$ be the population mean of the two groups. ($\bar{X_1}, \bar{X_2}$: sample mean)
$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 \neq \mu_2$

The pooled estimate of the standard deviation $s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = 30.63$

The test statistic is
$t = \frac{\bar{X_1}-\bar{X_2}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 0.6 \sim t_{n_1+n_2-2}$

$t_{n_1+n_2-2,1-\alpha/2} = 1.65 >$ t so we fail to reject the null hypothesis and accept $H_0$.
At 10% significance level, we can say that there is no statistically significant difference in the mother's weight at last menstrual period between smoking and non-smoking groups.

# Problem 3

a) $\alpha = 0.01$
Let p be the proportion of pregnant American women suffer from hypertension. ($p_0 = 0.20$)
I will perform one-sample test for binomial proportion.

99% confidence interval for one population proportion can be calculated by:

$$\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Thus, 99% CI is (0.02, 0.11)

Because 0.20 is not included in the 99% CI, we can conclude that we are 99% confident that our data does not support this claim.

b) $\alpha = 0.1$

I will perform one-sample binomial test for proportions.
$H_0$: $p = p_0$
$H_1$: $p < p_0$

Test statistic is $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ = -4.69 ~ N(0, 1)

$z_\alpha$ = -1.28 $> z$
Therefore, we reject the null hypothesis and accept $H_1$. We can conclude that at 10% significant level, the true proportion is indeed less than the claimed 20%.

# Problem 4

$\alpha = 0.01$
Let $p_1$ and $p_2$ be the proportion of pregnant American women having uterine irritability in smoking and non-smoking groups.

I will perform two-sample test for binomial proportion.
$H_0$: $p = p_0$
$H_1$: $p \neq p_0$

Now, $\hat{p_1} = 0.18$, $\hat{p_2} = 0.13$, $n_1 = 74$, $n_2 = 115$

$\hat{p} = \frac{n_1\hat{p_1}+n_2\hat{p_2}}{n_1+n_2} = 0.15$, $\hat{q} = 1 - \hat{p} = 0.85$

Test statistic is $z = \frac{\hat{p_1}-\hat{p_2}}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1}+\frac{1}{n_2})}}$ = 0.85 ~ N(0,1)

$z_{1-\alpha/2} = 2.58 >$ z
Thus, we fail to reject the null hypothesis.

We can conclude that the proportions of uterine irritability in the pregnant women are not different in smoking and non-smoking group at 1% significance level.

# Problem 5

a) There is more than two levels in race variable so one-way ANOVA would be appropriate in this case.

b)

- There are k populations of interest (k > 2)
  This is met because k = 3.
- The samples are drawn independently from the underlying populations
  This is met.
- Homoscedasticity

```
##
##  Bartlett test of homogeneity of variances
##
## data:  bwt by race
## Bartlett's K-squared = 0.65952, df = 2, p-value = 0.7191
```

Given the result from Bartlett test, (p-value > 0.05), we can say that the three race groups have equal variances.

- The distributions of the error terms are normal

```
##
##  Shapiro-Wilk normality test
##
## data:  aov_residuals
## W = 0.9894, p-value = 0.1738
```

Shapiro-Wilk test reveals that p-value > 0.05, so we can say that the error terms are normally distributed.

c) $\alpha = 0.05$
   Let $\mu_1$, $\mu_2$, $\mu_3$ be the mean birth weight in each race levels.
   $H_0$: $\mu_1 = \mu_2 = \mu_3$
   $H_1$: at least two of the birth weight means in each race differ

```
##              Df   Sum Sq Mean Sq F value  Pr(>F)
## race          1  3790184 3790184   7.369 0.00726 **
## Residuals   187 96179472  514329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Given the result from one-way ANOVA test, because p-value < 0.05, we reject the null hypothesis and accept the alternative hypothesis.
At 5% confidence level, we can say that at least two of the birth weight means in each race category differ.

d)

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

The adjusted p-value for the mean difference in birth weight between race 1 and 2 is 0.049; race 1 and 3 is 0.029; race 2 and 3 is 1.000.
This indicates that at 5% significance level, means of birth weight differ between race 1 and 2, and race 1 and 3.