Chapter 7 Homework

Due 2-23-2018

Problem 7.7

b) Estimate the variation in reading scores using a 99% Confidence Interval

Which Confidence Interval?

Since we are trying to estimate the variation in the reading scores with a Confidence Interval, we should use $100(1-\alpha)\%$ CI for σ^2 which involves χ^2.

Assumptions:

1) The population distribution is normal.

Confidence Interval Equation:

The Confidence Interval for Variance is as follows:

$$C.I. = \left(\frac{(n-1)\,s^2}{\chi^2_U}, \sigma^2, \frac{(n-1)\,s^2}{\chi^2_L} \right)$$
$$C.I. = \left(\frac{(150-1)\,9.537^2}{147.2568}, \sigma^2, \frac{(150-1)\,9.537^2}{117.098} \right)$$
$$C.I. \approx \left(92.0311, \sigma^2, 115.7338 \right)$$

(1)

Conclusion/Interpretation:

We are 95% confidence that the Variance in the large public school system is between 92.0311 and 115.7338 points.

c) Do the data indicate that the standard deviation in reading scores is greater than 9, the standard deviation for all students taking the exam the previous yea? Use $\alpha = 0.01$ in reaching your conclusion

Hypotheses:

With this test wanting to compare the standard deviation of theses students scores to the previous year's statewide results, we will use a Statistical Test for σ^2 and compare it to the Statewide Variance from last year of 81.

$H_0 : \sigma^2 \leq \sigma_0^2 = 81$
$H_1 : \sigma^2 > \sigma_0^2 = 81$

Assumptions:

1) We will still be assuming the population distribution is normal.

Test Statistic:

In order to find out the results of our statistical test, we will find the value of χ^2 to determine if we can reject the null hypothesis.

$$\chi_0^2 = \frac{(n-1)\, s^2}{\sigma_0^2}$$

$$\chi_0^2 = \frac{(149)\left(9.537^2\right)}{9^2}$$

$$\chi_0^2 = 167.3111$$

(2)

Rejection Region:

Since we have a one tail greater than test, if $\chi_0^2 \geq \chi_U^2$, then we can reject the null hypothesis. Using the qchisq function in R to get the upper bound, we have a $\chi_U^2 = 192.073$

P-Value:

In order to get the P-value for chi-square value, we will utilize the R function pchisq. Subtracting the value returned from 1 gives us a p-value of 0.1449.

Conclusion/Interpretation:

Since our $\chi_0^2 < \chi_U^2$ and out p-value $> \alpha$, we do not have extreme enough results to reject the Null Hypothesis that the standard deviation of our group is less than or equal to the statewide performance last year.

Problem 7.16

a) 95% C.I. for $\frac{\sigma_1}{\sigma_2}$

Assume the Computer Group as Group 1 and Conventional Group as Group 2. Since we are estimating a ratio of variances at the 0.95 Confidence Level, we will use the following equation:

$$C.I. = \left(\sqrt{\frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}},\ \sqrt{\frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1}} \right)$$

$$C.I. = \left(\sqrt{\frac{53.77^2}{36.94^2}(1.58)},\ \sqrt{\frac{53.77^2}{36.94^2}\left(\frac{1}{1.58}\right)} \right)$$

$$C.I. \approx (1.1580, 1.830)$$

(3)

Conclusion/Interpretation:

We are 95% confident the true ratio of standard deviation's between the Computer test taking group and the Conventional test taking group lie between 0.0944 and 1.8005.

b) Hypothesis test for $\sigma_1 = \sigma_2$; $\alpha = 0.05$

Hypotheses:

Since we want to check for equality amongst our standard deviations, we will use the statistical test for σ_1^2 and σ_2^2. Group 1 will still be Computers, and Group 2 will be Conventional.
$H_0 : \sigma_{Comp} = \sigma_{Con}$
$H_1 : \sigma_{Comp} \neq \sigma_{Con}$

Assumptions:

  1) Samples come from normally distributed populations.

Test Statistic:

$$F_0 = \frac{53.77^2}{36.94^2}$$
$$F_0 = 2.1188$$

<div align="right">(4)</div>

Rejection Region:

  We are looking for both $F_{1-\frac{\alpha}{2}}$ and $F_{\frac{\alpha}{2}}$ to determine if the H_0 is in line for rejection. Thus, by using the table, $F_{0.025} = \frac{1}{1.53} = 0.6536$ and $F_{0.975} = 1.53$. These will be the two values we will compare our 2.1188 to.

P-Value:

  In order to arrive at the p-value for 2.1188 with df_1 and df_2 both being 90, we will utilize the pf function in R to get the appropriate p-value for our F statistic. The function yields a cumulative probability to the left of 0.9997. Thus, our p-value is 0.0003

Conclusion/Interpretation:

  Since our F statistic of 2.1188 is greater than our high end F Boundary of 1.53, there is sufficient evidence to reject the hypothesis that the population standard deviations are equal. We have extreme enough evidence to appear that the population standard deviations are different between the group of students who took their exam on the computer and the group that took their test in a conventional method. Using the p-value, we have evidence to point in the same direction. This is because it is less than our $\frac{\alpha}{2}$ of 0.025.

c)95% C.I. for $\mu_1 - \mu_2$

Which Confidence Interval Should We Use?:

  Due to the fact that we are dealing with unequal variances or unequal standard deviations, we are going to follow down the path of using Satterwaite's approximation of $\mu_{Comp} - \mu_{Conv}$.

Assumptions:

  1) Independent Samples for both groups. 2) We are assuming unequal variances hold from the problem b)

Confidence Interval Equation:

  In order to complete Satterwhaitte's approximation, we must find our degrees of freedom. They are found by using this equation:

$$df = \frac{\left(\frac{s_{Comp}^2}{n_{Comp}} + \frac{s_{Conv}^2}{n_{Conv}}\right)}{\frac{\left(\frac{s_{Comp}^2}{n_{Comp}}\right)^2}{n_{Comp}-1} + \frac{\left(\frac{s_{Conv}^2}{n_{Conv}}\right)^2}{n_{Conv}-1}}$$

$$df = \frac{\left(\frac{53.77^2}{91} + \frac{36.94^2}{91}\right)}{\frac{\left(\frac{53.77^2}{91}\right)^2}{90} + \frac{\left(\frac{36.94^2}{91}\right)^2}{90}}$$

$$df = \frac{31.7716 + 14.9952}{11.2159 + 2.4984}$$

$$df = \frac{46.7668}{13.7143}$$

$$df = 3.4101$$

$$df = 3$$

Now that we have our degrees of freedom, we can find the C.I. for the difference in the two means.

$$C.I. = (y_{\bar{Comp}} - y_{\bar{Conv}}) \pm t_{\frac{\alpha}{2}, df}\sqrt{\frac{s_{Comp}^2}{n_{Comp}} + \frac{s_{Conv}^2}{n_{Conv}}}$$

$$C.I. = (484.45 - 487.38) \pm 3.182\sqrt{\frac{53.77^2}{91} + \frac{36.94^2}{91}}$$

$$C.I. = -2.93 \pm 3.182\sqrt{46.7668}$$

$$C.I. = -2.93 \pm 90.1468$$

$$C.I. \approx (-93.0768, 87.2168)$$

$$(5)$$

Conclusion/Interpretation:

We are 95% confident that the difference in the mean SAT scores of the Computer group and Conventional test taking group is between -93.0768 and 87.2168.

d) Hypothesis test for $\mu_1 = \mu_2$; $\alpha = 0.05$

Hypotheses:

Since we are checking the means of the two groups to see if they're equal, we will use Satterwaite's approximation to find the differences since the variances are still unequal from part b).
$H_0 : \mu_{Comp} - \mu_{Conv} = 0$
$H_1 : \mu_{Comp} - \mu_{Conv} \neq 0$

Assumptions:

1) Independent Samples
2) Unequal Variances

Test Statistic:

$$t_0 = \frac{(y_{C\bar{o}mp} - y_{C\bar{o}nv}) - D_0}{\sqrt{\frac{s_{Comp}^2}{n_{Comp}} + \frac{s_{Conv}^2}{n_{Conv}}}}$$

$$t_0 = \frac{(484.45 - 487.38) - 0}{\sqrt{\frac{53.77^2}{91} + \frac{36.94^2}{91}}}$$

$$t_0 = \frac{-2.93}{\sqrt{46.7668}}$$

$$t_0 = \frac{-2.93}{6.8386}$$

$$t_0 = -0.4284$$

Rejection Region:

In order to find the rejection boundary of our problem, we must check our a $t_{\frac{0.025}{2},3}$. The 3 came from the Satterwaite's df approximation above. That t value = 3.182.

P-Value:

The p-value on a t score of -0.4284 is 0.3486. This was derived from the pt function in R.

Conclusion/Interpretation:

Since our the absolute value of our Test Statistic is less than the t-value of our rejection bound of 3.182, there's not sufficient evident to reject the hypothesis that the two means are equal. There does not seem to be evidence to suggest that the two group's mean scores are different.

Problem 7.18

a) Perform a BFL test using R; you must present results in usual hypothesis test format; $\alpha = 0.05$

Hypotheses:

Since there are multiple groups, we will be using the BFL test for multiple variances.
$H_0 : \sigma_W^2 = \sigma_R^2 = \sigma_Z^2$
H_1: at least one of these variances are different

Assumptions:

1) The assumptions for ANOVA on our transformed response variable hold.

Test Statistic:

We will turn to R's lawstat package and the lavene function set around the median to get our test statistic. In this case, our test statistic is F = 1.4392.

Rejection Region:

To find the boundary of our Rejection Region, we will look up our F statistic at α, t - 1, N - t. That $F_{0.05,2,21}$ statistic is 3.47.

P-value:

Just like we got our test stat from R's 'lavene' function centered around the median, we will also get the p-value from there as well. The p-value is 0.2596.

Conclusion/Interpretation:

Since our p-value is larger than 0.05 and our test statistic is less than the rejection region, we do not have extreme enough data to reject the hypothesis that the variances are equal. It does not appear that at least one of the variances is different.

Chapter 7 Homework

Kyle Ligon

February 17, 2018

Problem 7.7 b)

```
#T_U
qchisq((1-0.05)/2, 149)
```

```
## [1] 147.2568
```

```
#T_L
qchisq(0.05/2, 149)
```

```
## [1] 117.098
```

```
#Lower End
(149)*(9.537^2)/(147.2568)
```

```
## [1] 92.03107
```

```
#Upper End
(149)*(9.537^2)/(117.098)
```

```
## [1] 115.7338
```

Problem 7.7 c)

```
#chisquare value
(149*9.537^2)/(9^2)
```

```
## [1] 167.3111
```

```
#upper val in Chi Sq test
qchisq(0.99, 149)
```

```
## [1] 192.073
```

```
#probability in the right tail of our chi-sq dist
1-pchisq(167.3111, 149)
```

```
## [1] 0.1449358
```

Problem 7.16 a) and b)

```
f <- (53.77^2)/(36.94^2)
sqrt((1.58*f))
```

```
## [1] 1.829665
```

```
sqrt((1/1.58)*f)
```

```
## [1] 1.158016
```

```
# part b test stat
f
```

```
## [1] 2.118782
```

```
# part b p-value
1- pf(2.1188, 90, 90)
```

```
## [1] 0.0002228897
```

7.16 c) Satterwhaite's Approximation

```
kk <- (53.77^2)/(91)
kk
```

```
## [1] 31.77157
```

```
jj <- (36.94^2)/(91)
jj
```

```
## [1] 14.9952
```

```
kk^2/90
```

```
## [1] 11.21592
```

```
jj^2/90
```

```
## [1] 2.498402
```

```
31.7716+14.9952
```

```
## [1] 46.7668
```

```
11.2159+2.4984
```

```
## [1] 13.7143
```

```
46.7668/13.7143
```

```
## [1] 3.410076
```

7.16 d) p-value

```
pt(-0.4284, df = 3)
```

```
## [1] 0.3486281
```

7.18 BFL test for Wild Ranch and Zoo group

```
library(lawstat)
library(reshape2)
library(tidyverse)
wild <- c(114.7, 128.9, 111.5, 116.4, 134.5, 126.7, 120.6, 129.59)
ranch <- c(120.4, 91.0, 119.6, 119.4, 150.0, 169.7, 100.9, 76.1)
zoo <- c(103.1, 90.7, 129.5, 75.8, 182.5, 76.8, 87.3, 77.3)
weight_m <- matrix(c(wild, ranch, zoo), nrow = 8, ncol = 3, byrow = F)
weight <- data.frame(weight_m)
colnames(weight) <- c("wild", "ranch", "zoo")
weight <- melt(weight)
levene.test(y = weight$value, group = weight$variable, location = "median")
```

```
##
##  modified robust Brown-Forsythe Levene-type test based on the
##  absolute deviations from the median
##
## data:  weight$value
## Test Statistic = 1.4392, p-value = 0.2596
```