

ECN 3972 – Lab 2

Start by doing three things:

1. Import HTWT1.xls into R. (Again note that this is an Excel file.) You should see 20 observations of three variables: observation number, height (inches above 5') which is labeled as "X," and weight (in pounds) which is labeled as "Y."
2. Type `library(mosaic)`
3. Type `library(psych)`
4. Get descriptive statistics for the data by typing `describe(HTWT1)`

The sample mean for weight is 169.4 and the sample standard deviation is 16.33.

Confidence Intervals & Hypothesis Tests in R

Confidence Intervals:

We can create confidence intervals and do hypothesis testing in R. Let's start with confidence intervals. We can build a confidence interval for the mean of weight using our sample. To do so, let's assign some variables for the sample mean, standard deviation, and sample size (from the results above):

```
xbar <- 169.4  
s <- 16.33  
N <- 20
```

Let's say we want a 95% confidence interval:

```
error <- qt(0.975, df=N-1)*s/(sqrt(N))  
lefttail <- xbar - error  
righttail <- xbar + error  
lefttail  
  
## [1] 161.7573  
  
righttail  
  
## [1] 177.0427
```

So, our 95% confidence interval for the population mean (μ) of weight is ($161.7573 \leq \mu \leq 177.0427$).

Note that we used 0.975 above in the calculation because confidence intervals are always two-sided.

We can use a simple command as well to get the confidence interval:

```
t.test(HTWT1$Y)
```

The results are:

One Sample t-test

data: HTWT1\$Y

t = 46.401, df = 19, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

161.7588 177.0412

sample estimates:

mean of x

169.4

Note: I did not specify a confidence level; the default is 0.95. But we could get the 99% confidence interval by entering:

```
t.test(HTWT1$Y, conf.level = 0.99)
```

The results are:

One Sample t-test

data: HTWT1\$Y

t = 46.401, df = 19, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

99 percent confidence interval:

158.9553 179.8447

sample estimates:

mean of x

169.4

Hypothesis Tests – Two sided

Now, let's try hypothesis testing on the weight variable. Suppose I wanted to test:

$H_0: \mu = 175$

$H_1: \mu \neq 175$

This is a two-sided test again. Let's do this at the $\alpha = 0.05$ level of significance. We need to assign the hypothesized value of μ . We will call it μ_0 :

```
mu0 <- 175
```

Now, we need to compute the critical value at $\alpha = 0.05$ level of significance:

```
alpha <- 0.05
critt <- qt(1-alpha/2, df =N-1)
c(-critt, critt)

## [1] -2.093024  2.093024
```

So, our decision rule will be to reject H_0 if our calculated t-value < -2.093024 or if our calculated t-value > 2.093024 . So, let's calculate our t-value:

```
calct <- (xbar - mu0)/(s/sqrt(N))
calct

## [1] -1.533617
```

Our calculated t-value is only -1.533617 so we cannot reject H_0 and we must conclude that the average weight may be equal to 175.

We could also get the calculated t value from the t.test command we used above. But, we have to enter in the hypothesized value of μ :

```
t.test(HTWT1$Y, mu=175)
```

The result:

One Sample t-test

data: HTWT1\$Y

t = -1.5339, df = 19, p-value = 0.1415

alternative hypothesis: true mean is not equal to 175

95 percent confidence interval:

161.7588 177.0412

sample estimates:

mean of x

169.4

The difference is due to rounding errors in the calculations.

Hypothesis Tests – One-tailed

Now, let's try a one-sided test. Suppose we want to know if average height is less than 10 inches and use a significance level of $\alpha = 0.01$. We can set up our hypotheses:

$H_0: \mu_{ht} \geq 10$

$H_1: \mu_{ht} < 10$

We need to assign values. (Note that I'm changing variable names so we don't overwrite what we have so far.)

```
xbarht = 10.35  
sht = 2.21  
muht0 = 10
```

Now, let's get the critical t-value. Remember, this is a one-sided test:

```
alpha2 = 0.01  
critt2 = qt(1-alpha2, df=N-1)  
-critt2  
  
## [1] -2.539483
```

Why did we request “-critt2” instead of “critt2”? Because the alternative hypothesis is that $\mu_{ht} < 10$. This means that we will be in the left side of the distribution (less than zero).

So, the critical t-value is -2.539483. This means that we will reject H_0 if our calculated t-value is less than -2.539483. Let's calculate our t:

```
calct2 <- (xbarht-muht0)/(sht/sqrt(N))  
calct2  
  
## [1] 0.7082568
```

Because 0.7082568 is NOT < -2.539483, we cannot reject H_0 and conclude that average height may be greater than or equal to 10.

Again, we can use the t.test function to get our calculated t. We have to tell the program that we are doing a one-sided test with an alternative hypothesis of less than (the default is a two-sided test):

```
t.test(HTWT1$X, mu=10, alternative = "less")
```

Results:

One Sample t-test

data: HTWT1\$X

t = 0.7092, df = 19, p-value = 0.7566

alternative hypothesis: true mean is less than 10

95 percent confidence interval:

-Inf 11.20335

sample estimates:

mean of x

10.35

One more?

Here's another one-sided test for weight:

$H_0: \mu_{wt2} \leq 162$

$H_1: \mu_{wt2} > 162$

Enter the data:

```
muwt0 <- 162
```

Let's again test at the $\alpha = 0.05$ level of significance. Let's gather our critical t-value:

```
alpha3 <- 0.05
critt3 <- qt(1-alpha3, df=N-1)
critt3

## [1] 1.729133
```

First, why did I ask for "critt3" and not "-critt3"? Because the alternative hypothesis is GREATER THAN! This means we are in the upper tail.

So, our decision rule is to Reject H_0 if our calculated t-value > 1.729133 .

Let's calculate our t:

```
calct3 <- (xbar- muwt0)/(s/sqrt(N))
calct3

## [1] 2.026565
```

We reject H_0 because $2.026565 > 1.729133$.

And again, using t.test:

```
t.test(HTWT1$Y, mu=162, alternative = "greater")
```

Results:

One Sample t-test

data: HTWT1\$Y

t = 2.0269, df = 19, p-value = 0.02847

alternative hypothesis: true mean is greater than 162

95 percent confidence interval:

163.0873 Inf

sample estimates:

mean of x

169.4