Week 10 & 11: Hypothesis Testing

Example 1

On page 527 of the OpenStax textbook, the National Institute of Standards and Technology provides data on conductivity properties of material. Following are measurements for 11 randomly selected pieces of a particular type of glass.

```
x <- c(1.11, 1.07, 1.11, 1.07, 1.12, 1.08, 0.98, 0.98, 1.02, 0.95, 0.95)
```

Is there convincing evidence that the average conductivity of this type of glass is greater tan one? Use a significance level of 0.05.

We find sample mean and standard deviation below.

```
mean(x)
```

```
## [1] 1.04
```

```
sd(x)
```

```
## [1] 0.06587868
```

The test statistics is the following.

$$t = \frac{1.04 - 1}{\frac{0.0659}{\sqrt{11}}} = 2.01$$

We use the pt function to find the p-value. The degrees of freedom is 51.

```
pt(2.01, 10)
```

```
## [1] 0.9639127
```

```
1-pt(2.01, 10)
```

```
## [1] 0.03608734
```

Because the p-value (p=0.036) is less than our alpha value, we reject the null hypothesis. The data support the claim that the mean conductivity level is greater than one.

It is possible to use the pre-defined function to reach the same conclusion.

```
t.test(x, mu = 1, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  x
## t = 2.0138, df = 10, p-value = 0.03586
## alternative hypothesis: true mean is greater than 1
## 95 percent confidence interval:
##  1.003999      Inf
## sample estimates:
## mean of x
##      1.04
```

Example 2

A researcher measured the body temperatures of a sample of randomly selected adults. We read the data file from Professor Nick Horton's website.

```r
dat <- read.csv(url("https://nhorton.people.amherst.edu/is5/data/Normal_temperature.csv"))
head(dat)
```

```
##   Temp
## 1 97.3
## 2 97.4
## 3 97.4
## 4 97.4
## 5 97.4
## 6 97.5
```

The sample mean and standard deviation are the following, along with the sample size.

```r
mean(dat$Temp)
```

```
## [1] 98.28462
```

```r
sd(dat$Temp)
```

```
## [1] 0.6823789
```

```r
length(dat$Temp)
```

```
## [1] 52
```

The conventional wisdom says that "normal" body temperature is 98.6°F. Based on this value, the test statistic is

$$t = \frac{98.285 - 98.6}{\frac{0.6824}{\sqrt{52}}} = -3.34$$

The p-value is below.

```r
pt(-3.34, 51)
```

```
## [1] 0.0007861981
```

```r
2*pt(-3.34, 51)
```

```
## [1] 0.001572396
```

For a two-tailed test, the p-value (p=0.0016) is less than a typical α value (5%). We reject the null hypothesis of 98.6; strong evidence suggests that the mean is not 98.6.

Again we can use the t.test function for the entire procedure.

```r
t.test(dat$Temp, mu = 98.6)
```

```
## 
##  One Sample t-test
## 
## data:  dat$Temp
## t = -3.3329, df = 51, p-value = 0.001606
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  98.09464 98.47459
## sample estimates:
## mean of x 
##  98.28462
```

Follow Up

On page 544 of the OpenStax textbook, Homework problem 85, the mean work week for engineers in a start-up company is believed to be 60 hours. A newly hired engineer hopes that it's shorter. She asked 10 engineers for the lengths of their mean work weeks, and have the following: 70, 45, 55, 60, 65, 55, 55, 60, 50, 55. Should she count on the mean work week to be shorter than 60 hours?