Hypothesis testing involves a few stages

- Identify the null and alternative hypotheses
- Compute the observed value of the test statistic
- Compute the p-value – the probability of seeing the observed value or worse under H_0.

For a test of proportion or a test of mean when the data is normal there are functions to do this for you. These are `prop.test` and `t.test`

For example. Suppose a car insurance specialist sees the following claim values from a shop

```
1200, 1500, 2500, 1700, 1800, 3600, 800, 1700
```

If most shops have a mean of \$1500 does this shop have a mean higher than \$1500?

The hypotheses would be

$$H_0 : \mu = 1500, \quad H_A : \mu > 1500.$$

The test statistic would be

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

And larger value of the observed are in the direction of the alternative. So the p-value could be found with

```
> x = c(1200, 1500, 2500, 1700, 1800, 3600, 800, 1700)
> t = (mean(x) - 1500)/(sd(x)/sqrt(length(x)))
> t
[1] 1.151
> pt(t,length(x)-1, lower.tail=F)
[1] 0.1438
```

This is not statistically significant.

`R` has a built in function to do this. It needs the data, the null hypothesis specified by the value for μ and a specification of the alternative, in this case `"greater"`.

```
> x = c(1200, 1500, 2500, 1700, 1800, 3600, 800, 1700)
> mean(x)
[1] 1850
> t.test(x, mu=1500, alt="greater")

        One Sample t-test

data:  x
t = 1.151, df = 7, p-value = 0.1438
alternative hypothesis: true mean is greater than 1500
95 percent confidence interval:
 1274  Inf
```

```
sample estimates:
mean of x
     1850
```

Notice we get the same thing, and a confidence interval to boot.

0.1 I forgot to check something about the hypotheses. What was it? Check that it is okay and explain why.

For proportions, we have a similar function, `prop.test`. Suppose, that 50% support the president in Iraq last week, but 56% do the next week in a sample of size 500. Does this indicate a difference?

Here the null and alternative would be

$$H_0 : p = .5, \quad H_A : p > .5$$

and the test statistic would be

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

Larger values support the alternative so the p-value can be found with

```
>  p = .5; n = 500
> t = (.56-.5)/sqrt(p*(1-p)/n)
> t
[1] 2.683
> pnorm(t,lower.tail=FALSE)
[1] 0.003645
```

The p-value is small and indicates that 3 times in 1000 is how likely a value this large or larger is when the true proportion is .5.

Again, this can be done more easily with a built-in function `prop.test`

```
> prop.test(.56*500,500, p =.5,alt="greater")

        1-sample proportions test with continuity correction

data:  0.56 * 500 out of 500, null probability 0.5
X-squared = 6.962, df = 1, p-value = 0.004163
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5223 1.0000
sample estimates:
   p
0.56
```

0.2 The p-value above is 0.004163 and is not the same as we got before. This is due to a continuity correction used by `prop.test`. Repeat the above, only using the extra argument `correct=FALSE`. Do you get the "correct" answer now?

0.3 Historically about 50% of students fail or withdraw form MTH 020. After changes in college policy, this rate was 45% for 1000 students. Does this indicate a change in the rate or is the difference explainable by sampling variation? Explain.

0.4 The response times for a set of subjects who are talking on a cell phone are

```
0.70 0.61 0.59 0.65 0.40 0.44 0.61 0.50 0.48 0.54
```

Suppose, it is known that those not talking on the cell phone will respond in 0.30 on average. Does this data indicate that the response time is greater when talking on a cell phone? Explain.

0.5 In trying to model the amount of cell phone usage per month, a student tracks the lengths of 15 cell phone calls. They are

```
14  2  1  2  1 17 17  2  1  1  6  1  1  2  2
```

Do a t-test to determine if the mean is more than 4. Explain.

0.6 In the last example, is the assumption of normality valid for the data? If yes, explain why you think so. If no, you can do a non-paramteric test using `wilcox.test`. First take a transform of the data so that it is symmetric. Then use `wilcox.test`, Finally reverse your transform.

0.0.1 The wilcox test

The `wilcox.test` uses the fact that for a symmetric distribution the signed-rank statistic is known.

This statistic is found by ranking all the data in absolute value from smallest to largest and then adding those ranks corresponding to the positive values of x. If this is a "large" number then there appears to be a bigger part of the distribution to the right of the symmetry point, if it is a "small" value, then the same only to the left.

What is "large" and "small" depends on the distribution. This only depends on n under the null assumption of symmetry. So we can simulate the distribution and see what large and small are.

To find the signed-rank statistic for a sample we can do

```
> x = rnorm(10)                  # a sample from symmetric median=0
> x
 [1] -0.16746 -1.71749 -1.79411 -0.66055 -0.66740  0.07278
 [7]  0.31649  0.29657 -0.78128 -0.07997
> abs(x)
 [1] 0.16746 1.71749 1.79411 0.66055 0.66740 0.07278 0.31649
 [8] 0.29657 0.78128 0.07997
> rank(abs(x))
```

```
 [1]  3  9 10  6  7  1  5  4  8  2
> x > 0
 [1] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE
> sum(rank(abs(x))[x>0])
[1] 10
```

Really only one line is important the `sum(rank(abs(x))[x>0])`

To simulate the distribution when $n = 10$ we can then do

```
> res = c()
> for(i in 1:250) {
+ x = rnorm(10)
+ res[i] = sum(rank(abs(x))[x>0])
+ }
> quantile(res,c(.05,.95))
   5%   95%
11.00 43.55
> hist(res)
```

So if our data is the first 10 values of the phone data

```
14  2  1  2  1 17 17  2  1  1
```

We can find the signed-rank statistic as follows

```
> x =  c(14,2,1,2,1,17,17,2,1,1)
> x = x-4                        # center
> sum(rank(abs(x))[x>0])
[1] 27
> sum(res > 27)/length(res)
[1] 0.52
```

So this p-value is not small.

The `wilcox.test` does essentially this

```
> wilcox.test(x,mu = 0,alt="greater")

        Wilcoxon signed rank test with continuity correction

data:  x
V = 27, p-value = 0.541
alternative hypothesis: true mu is greater than 0
```

0.7 Do the above with $n = 15$. Simulate the sampling distribution of the signed-rank statistic and test the full phone dataset. Compare with the wilcox.test results.

0.8 Simulate the signed-rank statistic for $n = 50$. What is the shape of the sampling distribution? Guess that the mean is. (It involves a power of n.)