Examining a proportion with R

In this little lab, I'll show you how to grab some actual data over the web and perform a hypothesis test in a couple of ways. Note that there's a Discourse assignment along these very lines. If you work through the examples on this sheet, you should have no problem with the Discourse assignment.

Grabbing random data

We've played with my random data generator before, though I've made a few changes. Here's some code to grab some random data and look at the first few rows.

```
df = read.csv("https://marksmath.org/cgi-bin/random_data.csv?username=mark")
head(df)
```

The output should look something like so:

first_name	last_name	age	gender	height	weight	income	smoke100	exerany	handedness
George	Howerton	51	male	70.98	222.96	216670	Y	Y	R
Rae	Cherry	53	female	64.33	127.77	9488	Y	Y	R
Tamara	Gore	24	female	65.35	126.19	4869	N	Y	L
Michael	Longwell	22	male	73.98	170.72	438	N	N	R

There's actually 100 rows of data and that's just the first few.

A question

Note the last column named "handedness". The value can be either L or R, depending on whether the individual is left or right handed. According to Wikipedia, approximately 12% of the population is left handed. Let's explore the question of whether this data supports the claim that claim or not.

A confidence interval

We begin with a 95% confidence interval for the number of people who are left handed according to this data. First, we'll need to know how the data splits up. R provides an easy way to do this:

```
table(df$handedness)
```

For me, 8 of the 100 folks are left handed so that $\hat{p} = 0.08$. Thus, my standard error can be computed by:

```
p = 0.08
se = sqrt(p*(1-p)/100)
```

Now, I can compute and display my 95% confidence interval as follows:

```
c(p-2*se, p+2*se)
```

A hypothesis test

Supposedly, about 12% of the population is left handed. My value for $\hat{p}$ was less than that, though.

Let's perform a p-test. To do so, we compute the probability of getting measured value of $\hat{p} = 0.08$ under the assumption that the actual value is $p_0 = 0.12$. I guess we can do this like so:

```
phat = 0.08
p0 = 0.12
se0 = sqrt(p0*(1-p0)/100)
pnorm(phat,p0,se0)
```

If this p-value is greater than than 0.05, we cannot reject the hypothesis that 12% of the population is left handed. Otherwise, we can reject the null.

An automatic approach

R provides some really high level functionality that does a lot of this for you. See if you can find the p-value here:

```
prop.test(8,100, p = 0.12, alternative = "less", correct = F)
```

Of course, I'm using the value that I found of 8. Your value is probably a bit different but the p-value here should agree with the p-value that you computed above.