

## Lab 7

Goal: To gain experience with hypothesis tests for a proportion

### Part 1 – Running a Hypothesis Test in Rcmdr

The goal of this part of the lab is to become familiar with using Rcmdr to run some hypothesis tests.

Let's revisit the first example from class. We were interested in testing whether the proportion of women at Amherst College is higher than the national reported proportion of 50.7% (US Census Bureau, 2009). Treating our class as a random sample of Amherst College students (which may or may not be reasonable), we found a test statistic value of  $\hat{p} = \frac{17}{28} = 0.6071$ .

We wish to test:

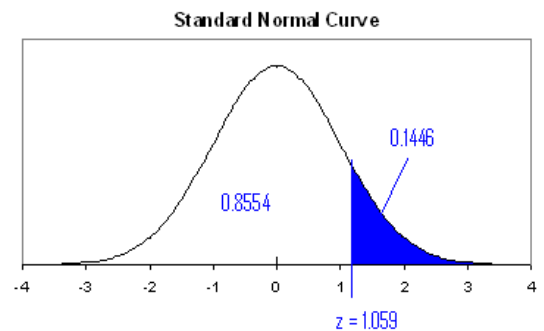
$$H_0: p = 0.507$$

$$H_A: p > 0.507,$$

where  $p$  is the true proportion of women at Amherst College.

We calculated a test statistic of:  $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = 1.059$ .

Using Normal theory, we found the  $p$ -value to be 0.1446.



Rcmdr can also do hypothesis tests for us, but it needs the raw data. Let's read in the classroom data we've used before. Note that it is still available on the course CMS site.

To open it in **Rcmdr**, follow the following steps:

1. Click Rcmdr Window, and select **Data > Import data > from Excel, Access, or dBase data set....**
2. In the window that opens, enter a **name** you would like to call your data set, e.g. *Class*. Click OK.
3. Find the data set file from wherever you saved it, click **Open**. In the window that opens, select the spreadsheet with the "cleaned" data and then click OK.

Now, go to **Statistics → Proportions → Single-sample proportion test**

- Highlight the GENDER variable
- Select the ">" alternative hypothesis and set the null hypothesis value to 0.507
- Leave the "Type of Test" set to the default, which is the Normal Approximation

There's a lot of numbers that might not make sense, but you should see that the  $p$ -value matches what we got by hand!

**Q1]** Compute a 95% confidence interval for the true proportion of Amherst College women.

```
> .Table
Gender
Female Male
  17    11

> prop.test(rbind(.Table), alternative='two.sided', p=.507, conf.level=.95,
+ correct=FALSE)
```

1-sample proportions test without continuity correction

```
data: rbind(.Table), null probability 0.507
X-squared = 1.1234, df = 1, p-value = 0.2892
alternative hypothesis: true p is not equal to 0.507
95 percent confidence interval:
 0.4240904 0.7643431
sample estimates:
      p
0.6071429
```

Based on the Rcmdr output above, the 95% confidence interval is (0.424 , 0.764).

Now, look at your computed confidence interval and determine whether it includes 0.507.

Yes, our computed confidence interval does include 0.507.

**Q2]** Using Rcmdr, test the hypotheses:

$$H_0: p = 0.507$$
$$H_A: p \neq 0.507,$$

What do you conclude? Use an  $\alpha$ -level of 0.05 here to determine significance.

Based on the output above, the  $p$ -value is 0.6071. We do not have enough evidence to reject the null hypothesis. We cannot claim that the proportion of women at Amherst College differs from that of the United States.

There is actually an equivalence here between confidence intervals and two-sided hypothesis tests. Rejecting the null hypothesis here using an  $\alpha$ -level of 0.05 is actually equivalent to

computing a 95% confidence interval and finding that the null hypothesis value is outside of the interval.

**Q3]** Test whether the majority of students at Amherst College are liberal. Use our class data, and assume the class is a representative sample of students at the school. State the hypotheses and make your conclusion.

I did this problem using Rcmdr. You have to recode the Politics variable into liberal and nonliberal.

We are testing the following hypotheses:

$$H_0 : p = 0.50$$

$$H_A : p > 0.50$$

The resulting  $p$ -value is 0.607, the same as the earlier problem! That's just a coincidence. We have no evidence that a majority of Amherst College students are liberal.

```
> .Table
```

```
Rpolitics
```

```
  Liberal Non-liberal
```

```
      17      11
```

```
> prop.test(rbind(.Table), alternative='greater', p=.5, conf.level=.95,  
+ correct=FALSE)
```

1-sample proportions test without continuity correction

```
data: rbind(.Table), null probability 0.5
```

```
X-squared = 1.2857, df = 1, p-value = 0.1284
```

```
alternative hypothesis: true p is greater than 0.5
```

```
95 percent confidence interval:
```

```
 0.452424 1.000000
```

```
sample estimates:
```

```
 p
```

```
0.6071429
```

## Part 2 – Assumptions and Errors

I haven't yet laid out the assumptions needed to do a one-proportion Z-test, so here they are:

- We have a simple random sample from the population
- Sample individuals are independent
- We are sampling less than 10% of the total population
- We are not too close to 0 or 1
  - $n \cdot p_0 \geq 10$
  - $n \cdot q_0 \geq 10$

**Q4]** Discuss each of the assumptions above to see if they are satisfied by our example in Question 2 of Part 1.

Notice that even after we make a decision about our hypotheses, we could have made a mistake.

- We could reject the null hypothesis when it is actually true – Type I error
- We could fail to reject the null hypothesis when it is actually false – Type II error

More formally, these are both conditional probabilities.

Type I Error =  $P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$

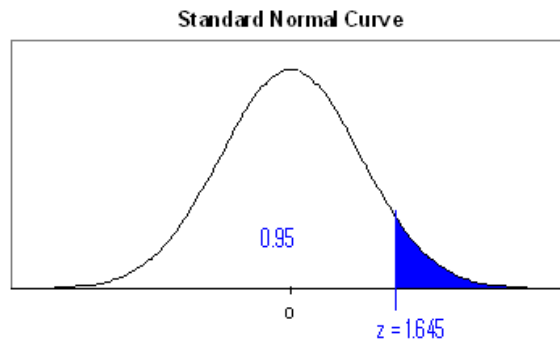
Type II Error =  $P(\text{do not reject } H_0 | H_A \text{ is true}) = \beta$

Power =  $P(\text{reject } H_0 | H_A \text{ is true}) = 1 - \beta$

The *power* of a test is a measure of its ability to correctly reject the null hypothesis. To calculate it, we need to have a specific value for the parameter in the alternative hypothesis.

An ideal hypothesis test would have a low  $\alpha$  and a high power. However, these quantities often oppose each other.

Let's return to our example from earlier. With an  $\alpha = 0.05$  and a one-sided test with a ">" alternative, we reject the null hypothesis for any Z statistics greater than 1.645. This corresponds to rejecting the null hypothesis whenever we get a sample estimate of 0.6626.



### Calculating Power

The power of the test needs a specific alternative hypothesis value to be calculated. Suppose instead that the alternative is that the proportion is 0.600.

$$H_0: p = 0.507$$

$$H_A: p = 0.600,$$

Now

To find this probability, we use a normal distribution but have a mean of 0.6 and a standard deviation of  $\sqrt{\frac{0.6(1-0.6)}{28}} = 0.09258$ .

$$P(\hat{p} > 0.6626 | p = 0.6) = P\left(Z > \frac{0.6626 - 0.6}{0.09258}\right) = P(Z > 0.6762) = 1 - .7517 = 0.2483$$

The power is pretty low!

**Q5]** Calculate the power again, but using an  $\alpha = 0.10$  cutoff. What happens to the power? As the Type I error ( $\alpha$ ) increases from 0.05 to 0.10, what happens to the type II error?

With an  $\alpha = 0.10$ , I will reject the null hypothesis if I get a Z statistic value that is greater than 1.28. This corresponds to a sample proportion of  $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \Rightarrow 1.28 =$

$$\frac{\hat{p} - 0.507}{\sqrt{\frac{(0.507)(0.493)}{28}}} \Rightarrow \hat{p} = 0.6280. \text{ The power is}$$

$$P(\hat{p} > 0.6280 | p = 0.6) = P\left(Z > \frac{0.6280 - 0.6}{0.09258}\right) = P(Z > 0.3024) = 1 - 0.6179 = 0.3821.$$

The power has increased and the probability of a Type II error decreases.

**Q6]** Calculate the power again with an  $\alpha = 0.05$  cutoff, but with a sample size of  $n = 100$ . What happens to the power? What happens to the type II error?

With an  $\alpha = 0.05$ , I will reject the null hypothesis if I get a  $Z$  statistic value that is greater than 1.645. This corresponds to a sample proportion of  $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \Rightarrow 1.645 =$

$$\frac{\hat{p} - 0.507}{\sqrt{\frac{(0.507)(0.493)}{100}}} \Rightarrow \hat{p} = 0.5892. \text{ The power is}$$

$$P(\hat{p} > 0.5892 | p = 0.6) = P\left(Z > \frac{0.5892 - 0.6}{0.04899}\right) = P(Z > -0.2205) = 1 - 0.4129 = 0.5871.$$

The power has increased and the probability of a Type II error decreases.

The point: The only way to keep both types of errors low is to increase the sample size.