Notes: MS 204 Chapter 3

Overview

- Inference for a single population proportion
- Inference for the difference of two population proportions

An example

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

a. All 1000 get the drug
b. 500 get the drug, 500 don't

The General Social Survey asked the same question, and we observed the following results.

```
library(tidyverse)
library(mosaic)
drug.example <- c(rep("All 1000 get the drug", 99), rep("500 get the drug, 500 don't", 571))
tally( ~ drug.example)
```

```
## drug.example
## 500 get the drug, 500 don't       All 1000 get the drug
##                         571                          99
```

Inference for a single population proportion

parameter

point estimate

population

Inference

Central limit theorem for proportions

Assumptions?

Hypothesis test

Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

Confidence interval

Code

```
prop.test(~ drug.example, p = 0.8, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  NULL$drug.example  [with success = 500 get the drug, 500 don't]
## X-squared = 11.103, df = 1, p-value = 0.0008618
## alternative hypothesis: true p is not equal to 0.8
## 95 percent confidence interval:
##  0.8225647 0.8777859
## sample estimates:
##         p
## 0.8522388
```

```
prop.test(~ drug.example, p = 0.83, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  NULL$drug.example  [with success = 500 get the drug, 500 don't]
## X-squared = 2.1934, df = 1, p-value = 0.1386
## alternative hypothesis: true p is not equal to 0.83
## 95 percent confidence interval:
##  0.8225647 0.8777859
## sample estimates:
##         p
## 0.8522388
```

Patriots example:

```
patriots <- c(rep("Won coin toss", 19), rep("Lost coin toss", 6))
prop.test(~ patriots, p = 0.5, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  NULL$patriots  [with success = Lost coin toss]
## X-squared = 5.76, df = 1, p-value = 0.0164
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1015806 0.4552084
## sample estimates:
##    p
## 0.24
```

```
prop.test(~ patriots, p = 0.5, alternative = "two.sided", success = "Won coin toss")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  NULL$patriots  [with success = Won coin toss]
## X-squared = 5.76, df = 1, p-value = 0.0164
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.5447916 0.8984194
## sample estimates:
##    p
## 0.76
```

Choosing a sample size

How many people should we sample in order to cut the margin of error with a 95 percent interval down to 1 percent?