Notes: MS 204 Chapter 3 two population proportion

Warm up

1. A pollster makes a 95 percent confidence interval for the proportion of college students that binge drink, finding it to be (46%, 54%). What is her point estimate and the margin of error?

2. How many people did the pollster survey?

Overview

- Inference for the difference of two population proportions

An example

Scientists predict that global warming may have big effects on the polar regions within the next 100 years. One of the possible effects is that the northern ice cap may completely melt. Would this bother you a great deal if it actually happened?

The General Social Survey (GSS) asks the same question, below are the distributions of responses from the 2010 GSS as well as from a group of introductory statistics students at Duke University.

```
library(tidyverse)
library(mosaic)
## response
icecap <- c(rep("Bothered", 523), rep("Not bothered", 262))

## explanatory
sample.group <- c(rep("GSS", 454), rep("Duke", 69), rep("GSS", 226), rep("Duke", 36))
tally(icecap ~ sample.group)
```

```
##                sample.group
## icecap          Duke GSS
##   Bothered        69 454
##   Not bothered    36 226
```

Inference for a difference in two population proportions

parameter

point estimate

population

Inference

Central limit theorem for difference in two proportions

Assumptions?

Confidence interval

```r
prop.test(icecap ~ sample.group, alternative = "two.sided")
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  tally(icecap ~ sample.group)
## X-squared = 0.010255, df = 1, p-value = 0.9193
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.1134510  0.0924426
## sample estimates:
##    prop 1    prop 2
## 0.6571429 0.6676471
```

Hypothesis test

Do these data provide convincing evidence that there's a difference in the fraction of all Duke students and the fraction of all Americans who care a great deal about the ice cap?

pooled proportion

Mythbusters yawning example:

parameter

point estimate

population

```
seeded <- c(rep(0, 12), rep(1, 24), rep(0, 4), rep(1, 10))
yawned <- c(rep(0, 36), rep(1, 14))
yawners.df <- data.frame(seeded, yawned)
prop.test( ~ yawned | seeded, data = yawners.df)
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  tally(yawned ~ seeded)
## X-squared = 3.853e-31, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2616754  0.3499107
## sample estimates:
##    prop 1    prop 2
## 0.7500000 0.7058824
```