Lsn 15 - MA206Y

Clark

Admin

The folks at MythBusters investigated whether yawning is contageous. Fifty people attending a local flea market were recuited to participate. Subjects were ushered, one at a time, into one of thre rooms by co-host Kari. She yawned as she ushered subjects into two of the rooms, and for the other room she did not yawn. The researchers decided in advance, with a random mechanism, which subjects went to which room. As time passed, the researchers watched to see which subjects yawned.

What was the point of having two groups here?

Is this an observational study or a randomized experiment? How are you deciding?

Identify the explanatory and response variables in this study. Classify them as categorical or quantitative.

In words, what is our null hypothesis and which is our alternative. **The null hypothesis is typically a statement of no effect**. State the null and alternative hypothesis in terms of association between the explanatory and response variables in this study

There are two long-run probabilities (parameters) in this study. What are they?

Write the null and alternative hypothesis in terms of the two long-run probabilities using appropriate symbols.

The researchers found that 11 of 34 subjects who had been given a yawn seed actually yawned themselves, comapred with 3 of 16 who had not been given a yawn seed. A table:

Proportion who yawned in "yawn seed group":

Proportion who yawned in control group:

Difference in conditional proportions:

This is your **statistic**: $\hat{p}_1 - \hat{p}_2$

```r
yawn.dat=read.table("http://www.isi-stats.com/isi/data/chap5/Yawning.txt",header=T)
yawn.dat %>% group_by(YawnSeed,Response)%>%summarize(count=n())
```

```
## # A tibble: 4 x 3
## # Groups:   YawnSeed [2]
##   YawnSeed Response count
##   <fct>    <fct>    <int>
## 1 Control  NoYawn      13
## 2 Control  Yawn         3
## 3 Seeded   NoYawn      23
## 4 Seeded   Yawn        11
```

The key to our simulation analysis is to assume that if yawning is NOT contageious, then the 14 yawners would have yawned regardless of whether or not theyhad seen the yawn seed.

Card experiment

In our simulation what was the proportion who yawned in the treatment group?

What was the proportion who yawned in the control group?

Let's go to the applet

Simulate 1000 shuffles. Is the null distribution for the difference in proportions centered at zero? Explan

Is the observed value out in the tail of the null distribution?

Let's calculate a P-value. What's our conclusion?

What we just went through was a hypothesis test for two sample proportion. As opposed to what we previously had been dong, we now have two groups that we are comparing. Our groups are defined by our explanatory variable and we are examining whether the explanatory variable impacts the response variable.

Here our parameter is:

Our null and alternative hypothesis are something of the form:

To estimate our parameter we will use a statistic, in this case we use:

We can simulate the distribution of the statistic under H_0 using the two proportion applet, or we can use a theory based test.

Though this is beyond the scope of the course, there's actually an exact way to calculate this. R.A. Fisher, the father of modern statistics, came up with a way to count all the possible simulations. He had to do this since computers weren't invented. But the original story was a lady claimed to be able to tell whether the tea or the milk was added first to a cup. Fisher proposed giving her eight cups of team, four of each variety, in random order. He then found the exact probability that she was guessing. Based off this, there is a test called Fisher's exact test:

```
Yawns<-
matrix(c(11, 3, 23, 13),
       nrow = 2)
fisher.test(Yawns, alternative = "greater")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  Yawns
## p-value = 0.2583
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.5226512       Inf
## sample estimates:
## odds ratio
##   2.044169
```

The key point here is that in 1935, Fisher **wanted** to be able to do a simulation and couldn't, so we had to use counting techniques. We now can, if we desire, simulate.

However, as we can see in our plot, we do have hope for a theory based test. What does the $\hat{p}_1 - \hat{p}_2$ curve look like?

Is this surprising?

Let's work 5.1.8-5.1.11

5.2.2