

Differences Between Groups

– Part 1

Evan L. Ray

November 20, 2017

Birth Weights (SDM4 22.39)

A study in the Journal of the American Medical Association reported a study examining the possible impact of air pollution caused by the September 11 attacks on New York's World Trade Center on the weight of babies. Researchers found that 16 out of 182 babies born to mothers who were exposed to heavy doses of soot and ash were classified as having low birth weight. 92 out of 2300 babies born in another New York City hospital whose mothers had not been near the site of the disaster were similarly classified.

- What is the research question?
- What is/are the population parameter(s) involved, and how are they related to the question at hand?

Setup

- Population parameters:
 - p_1 = proportion of babies born to mothers exposed to air pollution who have low birth weight
 - p_2 = proportion of babies born to mothers not exposed to air pollution who have low birth weight
- Research question:
 - Is $p_1 = p_2$?

Setup

- Population parameters:
 - p_1 = proportion of babies born to mothers exposed to air pollution who have low birth weight
 - p_2 = proportion of babies born to mothers not exposed to air pollution who have low birth weight
- Research question:
 - Is $p_1 = p_2$?
 - Is $p_1 - p_2 = 0$?

Setup

- Population parameters:
 - p_1 = proportion of babies born to mothers exposed to air pollution who have low birth weight
 - p_2 = proportion of babies born to mothers not exposed to air pollution who have low birth weight
- Research question:
 - Is $p_1 = p_2$?
 - Is $p_1 - p_2 = 0$? (Hypothesis test)
 - $H_0: p_1 - p_2 = 0$
 - $H_A: p_1 - p_2 \neq 0$
 - How big is the difference $p_1 - p_2 = 0$? (Confidence interval)

Sample Statistics

- Population parameters:
 - p_1 = proportion of babies born to mothers exposed to air pollution who have low birth weight
 - p_2 = proportion of babies born to mothers not exposed to air pollution who have low birth weight
- Sample statistics:
 - \hat{p}_1 = proportion of babies **in this sample** born to mothers exposed to air pollution who have low birth weight
 - \hat{p}_2 = proportion of babies **in this sample** born to mothers not exposed to air pollution who have low birth weight
 - $\hat{p}_1 - \hat{p}_2$ = differences in sample proportions

Sampling Distributions

- $\hat{p}_1 \sim \text{Normal} \left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}} \right)$
- $\hat{p}_2 \sim \text{Normal} \left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}} \right)$
- $\hat{p}_1 - \hat{p}_2 \sim \text{Normal} \left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$
- Check all usual assumptions for both \hat{p}_1 and \hat{p}_2 :
 - 2 outcomes
 - same probability of success for all people **within each group**
 - independence of people **within each group** (randomization, sample size < 10% of population size)
 - ≥ 10 successes and ≥ 10 failures in each group
- Also check independence across groups (no connection between observations in different groups)

Tests and Confidence Intervals in R

- 16 out of 182 babies born to mothers who were exposed to pollution had low birth weight. ($16/182 = 0.088$)
- 92 out of 2300 babies born to mothers who were **not** exposed to pollution had low birth weight. ($92/2300 = 0.04$)

```
prop.test(x = c(16, 92), n = c(182, 2300), conf.level = 0.95)
```

```
##  
## 2-sample test for equality of proportions with continuity  
## correction  
##  
## data: c(16, 92) out of c(182, 2300)  
## X-squared = 8.1867, df = 1, p-value = 0.00422  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## 0.00303607 0.09278811  
## sample estimates:  
##      prop 1      prop 2  
## 0.08791209 0.04000000
```


Conclusions

- The p-value was 0.0042.
- The data provide enough evidence to conclude that mothers exposed to heavy doses of air pollution give birth to babies with low birth weights more often than mothers who are not exposed to heavy doses of air pollution, at a significance level of $\alpha = 0.01$
- The confidence interval was [0.003, 0.093]
- We are 95% confident that the difference in the proportion of births where the baby has a low birth weight between mothers who were and were not exposed to heavy doses of air pollution is between 0.003 and 0.093.

1 sided Tests (but use 2 sided CI's)

- $H_0: p_1 = p_2$ (or $p_1 - p_2 = 0$)
- $H_A: p_1 > p_2$ (or $p_1 - p_2 > 0$)

```
prop.test(x = c(16, 92), n = c(182, 2300), conf.level = 0.95,  
          alternative = "greater")
```

```
##  
## 2-sample test for equality of proportions with continuity  
## correction  
##  
## data: c(16, 92) out of c(182, 2300)  
## X-squared = 8.1867, df = 1, p-value = 0.00211  
## alternative hypothesis: greater  
## 95 percent confidence interval:  
## 0.009774311 1.000000000  
## sample estimates:  
##      prop 1      prop 2  
## 0.08791209 0.04000000
```