Class 22: Foundations for inference II

November 13, 2017

Variability in estimates

Pew Research Survey

Young, Underemployed and Optimistic

Coming of Age, Slowly, in a Tough Economy

Young adults hit hard by the recession. A plurality of the public (41%) believes young adults, rather than middle-aged or older adults, are having the toughest time in today's economy. An analysis of government economic data suggests that this perception is correct. The recent indicators on the nation's labor market show a decline in the

Tough economic times altering young adults' daily lives, long-term plans. While negative trends in the labor market have been felt most acutely by the youngest workers, many adults in their late 20s and early 30s have also felt the impact of the weak economy. Among all 18- to 34-year-olds, fully half (49%) say they have taken a job they didn't want just to pay the bills, with 24% saying they have taken an unpaid job to gain work experience. And more than one-third (35%) say that, as a result of the poor economy, they have gone back to school. Their personal lives have also been affected: 31% have postponed either getting married or having a baby (22% say they have postponed having a baby and 20% have put off getting married). One-in-four (24%) say they have moved back in with their parents after living on their own.

http://pewresearch.org/pubs/2191/young-adults-workers-labor-market-pay-careers-advancement-recession

Margin of error

The general public survey is based on telephone interviews conducted Dec. 6-19, 2011, with a nationally representative sample of 2,048 adults ages 18 and older living in the continental United States, including an oversample of 346 adults ages 18 to 34. A total of 769 interviews were completed with respondents contacted by landline telephone and 1,279 with those contacted on their cellular phone. Data are weighted to produce a final sample that is representative of the general population of adults in the continental United States. Survey interviews were conducted under the direction of Princeton Survey Research Associates International, in English and Spanish. Margin of sampling error is plus or minus 2.9 percentage points for results based on the total sample and 4.4 percentage points for adults ages 18-34 at the 95% confidence level.

- 41% $\pm$ 2.9%: We are 95% confident that 38.1% to 43.9% of the public believe young adults, rather than middle-aged or older adults, are having the toughest time in today's economy.
- 49% $\pm$ 4.4%: We are 95% confident that 44.6% to 53.4% of 18–34 years olds have taken a job they didn't want just to pay the bills.

Parameter estimation

- We are often interested in **population parameters**.
- Since complete populations are difficult (or impossible) to collect data on, we use **sample statistics** as **point estimates** for the unknown population parameters of interest.
- Sample statistics vary from sample to sample.
- Quantifying how sample statistics vary provides a way to estimate the **margin of error** associated with our point estimate.
- But before we get to quantifying the variability among samples, let's try to understand how and why point estimates vary from sample to sample.

Suppose we randomly sample 1,000 adults from each state in the US. Would you expect the sample means of their heights to be the same, somewhat different, or very different?

Parameter estimation

- We are often interested in **population parameters**.
- Since complete populations are difficult (or impossible) to collect data on, we use **sample statistics** as **point estimates** for the unknown population parameters of interest.
- Sample statistics vary from sample to sample.
- Quantifying how sample statistics vary provides a way to estimate the **margin of error** associated with our point estimate.
- But before we get to quantifying the variability among samples, let's try to understand how and why point estimates vary from sample to sample.

Suppose we randomly sample 1,000 adults from each state in the US. Would you expect the sample means of their heights to be the same, somewhat different, or very different?

Not the same, but only somewhat different.

Mythbusters Statistics

An experiment conducted by the *MythBusters*, a science entertainment TV program that aired on the Discovery Channel, tested if a person can be subconsciously influenced into yawning if another person near them yawns.

- 50 people were randomly assigned to two groups: 34 to a group where a person near them yawned (treatment group) and 16 to a group where there wasn't a person yawning near them (control group).
- The results of the experiment are in a file "yawn.csv" posted on the course website: http://fall17.cds101.com/pages/datasets/

Mythbusters Statistics

Open the dataset and either in RStudio or on a piece of paper, fill out a contingency table like the one below:

	Treatment	**Control**	**Total**
Yawn			
Not Yawn			
Total			

Mythbusters Statistics

1. What is the null hypothesis?

Mythbusters Statistics

1. What is the null hypothesis?
2. What is the alternative hypothesis?

Mythbusters Statistics

1. What is the null hypothesis?
2. What is the alternative hypothesis?
3. What value of α should we use?

Mythbusters Statistics

1. What is the null hypothesis?
2. What is the alternative hypothesis?
3. What value of α should we use?
4. Should we use a one-sided or two-sided hypothesis test?

Mythbusters Statistics

1. What is the null hypothesis?
2. What is the alternative hypothesis?
3. What value of α should we use?
4. Should we use a one-sided or two-sided hypothesis test?
5. What quantities do we need to subtract to find the observed difference between the yawning rates under the two different groups?

Hypothesis test with `infer`

Conduct a hypothesis test with `infer`. Is the *Mythbusters* result statistically significant?

Confidence interval

How to compute the confidence interval for this experiment.