Lecture 21: Difference of two proportions

Chapter 6.2

Question for today

How do we infer about a difference in proportions $p_1 - p_2$?

Surveys

The way a question is phrased in survey can influence a person's response. Ex on p.281: the Pew Research Center conducted a survey with the following question:

Surveys

The way a question is phrased in survey can influence a person's response. Ex on p.281: the Pew Research Center conducted a survey with the following question:

By 2014 all Americans will be required to have health insurance. X while Y. Do you approve of disapprove of this policy?

Surveys

The way a question is phrased in survey can influence a person's response. Ex on p.281: the Pew Research Center conducted a survey with the following question:

By 2014 all Americans will be required to have health insurance. X while Y. Do you approve of disapprove of this policy?

where X and Y were randomly ordered between

- People who do not buy insurance will pay a penalty
- People who cannot afford it will receive financial help from the government

Surveys

The way a question is phrased in survey can influence a person's response. Ex on p.281: the Pew Research Center conducted a survey with the following question:

By 2014 all Americans will be required to have health insurance. X while Y. Do you approve of disapprove of this policy?

where X and Y were randomly ordered between

- People who do not buy insurance will pay a penalty
- People who cannot afford it will receive financial help from the government

Let's infer about the difference in proportion of people who approve.

Example from Text

	Sample size n_i	Approve (%)	Disapprove (%)	Other (%)
people who do not buy it will pay a penalty given first	771	47	49	3
people who cannot afford it will receive financial help from the gov't given first	732	34	63	3

Example from Text

Lump in Other and Disapprove into "Don't Approve"

	Sample size n_i	Approve (%)	Don't Approve (%)
people who do not buy it will pay a penalty given first	771	47	53
people who cannot afford it will receive financial help from the gov't given first	732	34	66

Example from Text

Lump in Other and Disapprove into "Don't Approve"

	Sample size n_i	Approve (%)	Don't Approve (%)
people who do not buy it will pay a penalty given first	771	47	53
people who cannot afford it will receive financial help from the gov't given first	732	34	66

So $\widehat{p}_1 - \widehat{p}_2 = 0.47 - 0.34 > 0$: people are more likely to support Obamacare in the first scenario.

Conditions for Using Normal Model

What p_1 & p_2?

Confidence Intervals

What is a 90% confidence interval for the difference in proportions?

Interpretation

Interpretation

More support Obamacare if stated as follows:

People who do not buy it will pay a penalty while people who cannot afford it will receive financial help from the government.

Hypothesis Tests

Exercise 6.31 on Page 319

A 2010 survey asked 827 randomly sample voters in California "How do you feel about drilling for oil and natural gas off the coast of California?"

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Don't Know	104	131
Total	438	389

Exercise 6.31 on Page 319

A 2010 survey asked 827 randomly sample voters in California "How do you feel about drilling for oil and natural gas off the coast of California?"

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Don't Know	104	131
Total	438	389

Test at the $\alpha = 0.10$ significance level if the proportion of college graduates who support off-shore drilling is different than that of non-college graduates.

Jury Selection

Preview of next lecture: In many trials a big issue is the racial makeup of the jury.

Jury Selection

Preview of next lecture: In many trials a big issue is the racial makeup of the jury.

Question: is there a way to figure out if there is a racial bias in jury selection?

Jury Selection

Say we have a juror pool (registered voters) where the racial breakdown is:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%

Jury Selection

If we pick $n = 100$ jurors at random (i.e. unbiasedly), we expect the breakdown of counts to be:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	72	7	12	9	$n = 100$

Jury Selection

Say we observe the following counts:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	0	0	100	0	$n = 100$

Jury Selection

Say we observe the following counts:

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	0	0	100	0	$n = 100$

Fairly obvious bias in juror selection!

Jury Selection

But what about the following? Is there a bias? i.e. a non-random mechanism at play?

Race	White	Black	Hispanic	Other	Total
Registered Voters	72%	7%	12%	9%	100%
Representation	75	6	11	8	$n = 100$

Next Two Lectures

Chi-square tests are used to compare expected counts with observed counts.

Next Two Lectures

Chi-square tests are used to compare expected counts with observed counts.

Two tests we'll see:

- Goodness-of-fit tests: for frequency tables
- Tests for independence: for contingency/two-way tables