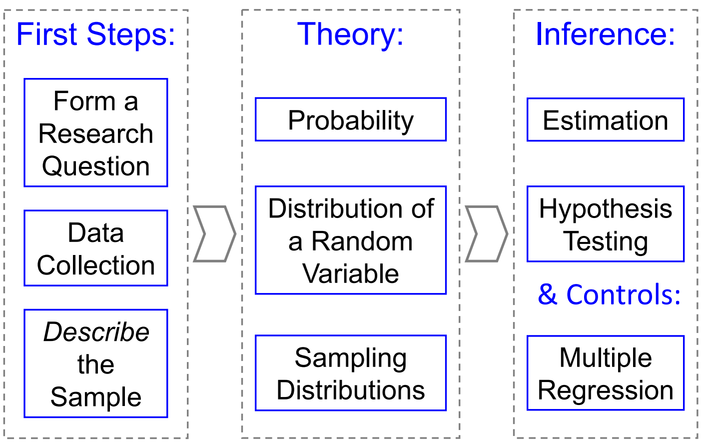
Confidence Interval Estimation: Single Proportion and Difference Between Proportions

Lecture 12

Reading: Chapter 11

ECO220Y: Overview



Birth Month Predicts ADHD Diagnosis

- Recall pre-reading: "The Link Between August Birthdays and A.D.H.D." *The New York Times*

  The rate of claims-based ADHD diagnosis among children in states with a September 1 cutoff was 85.1 per 10,000 children (309 cases among 36,319 children; 95% confidence interval [CI], 75.6 to 94.5) among those born in August and 63.6 per 10,000 children (225 cases among 35,353 children; 95% CI, 55.4 to 71.9) among those born in September, an absolute difference of 21.4 per 10,000 children (95% CI, 8.9 to 34.0); the corresponding difference in states without the September 1 cutoff was 8.9 per 10,000 children (95% CI, −14.9 to 20.8). [Layton et al. (2018), p. 2122]

Links to sources: https://www.nytimes.com/2018/11/28/opinion/august-birthdays-adhd.html and https://www.nejm.org/doi/full/10.1056/NEJMoa1806828

Estimation

- Estimator: Random variable based on sample statistics that is used to estimate a parameter
  - Point Estimator: Uses a single value
    - Ex: Infer population proportion is 0.0085
  - Interval Estimator: Uses a range of values and specifies the level of confidence
    - Ex: Infer 0.0076 and 0.0095 (0.0085 $\pm$ 0.0009) contains p with 95% confidence
    - As sampling error increases, width increases

Unbiasedness

- Unbiased estimator: Expected value equals the population parameter that it estimates
  - The sample mean is an unbiased estimator of the population mean: $E[\bar{X}] = \mu$
  - The sample proportion is an unbiased estimator of the population proportion: $E[\hat{P}] = p$
- Upward bias: E[estimator] > parameter
- Downward bias: E[estimator] < parameter

Law(s) of Large Numbers (LLN)

- Recall: $E[\bar{X}] = \mu$ and $SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$
  - SW (2011): Under general conditions, $\bar{X}$ will be near μ, with very high probability when n is large (i.e. $\bar{X}$ is a *consistent* estimator of μ)
- Similarly: $E[\hat{P}] = p$ and $SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$
- Use $\hat{P}$ to *make an inference* about p with interval estimation or hypothesis testing

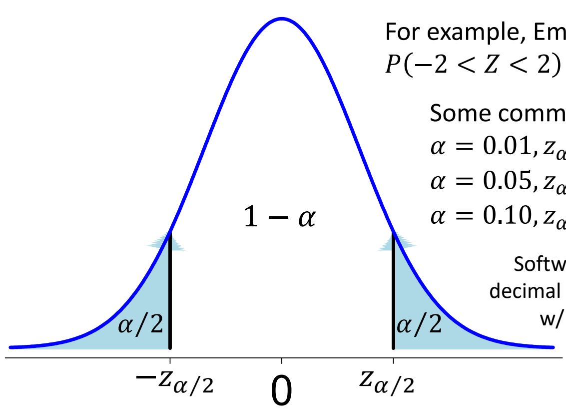
Confidence Interval (CI)

- CI Estimate = Point Estimate ± Margin of Error
  - Margin of Error (ME) = Measure related to desired confidence level * Measure of sampling error
    - Confidence level and sampling error affect *width* of CI
    - Confidence level: $(1-\alpha)$ where $0 < \alpha < 1$
      - For example, 0.95 means 95% confident, which is popular because it is a round number that sounds convincing
    - Significance level: α where $0 < \alpha < 1$
      - A 5% significance level (95% confidence) means $\alpha = 0.05$
    - Sampling distribution measures sampling error

Recall: Sampling Distribution of $\hat{P}$

- If random sampling & independence (10% condition): $\hat{P} = \frac{X}{n}$ where $X \sim B(n,p)$
  - $E[X] = np;\ V[X] = np(1-p)$
  - $E[\hat{P}] = p; V[\hat{P}] = p(1-p)/n$
  - The sampling distribution of $\hat{P}$ is approximatetly Normal if $p \pm 3 * \sqrt{p(1-p)/n}$ within (0,1)
  - But p unknown so check $\hat{P} \pm 3 * \sqrt{\hat{P}(1-\hat{P})/n}$
    - Alternate rule-of-thumb: $n\hat{P} \geq 10, n(1-\hat{P}) \geq 10$

To derive CI estimator of p start with $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$



For example, Empirical Rule: $P(-2 < Z < 2) = 0.954$

Some common α's:
$\alpha = 0.01, z_{\alpha/2} = 2.576$
$\alpha = 0.05, z_{\alpha/2} = 1.960$
$\alpha = 0.10, z_{\alpha/2} = 1.645$

Software gives third decimal place (approx. w/ Normal table)

Above example: $P(-1.4395 < Z < 1.4395) = 0.85$

Derive CI Estimator of p

- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P\left(-z_{\alpha/2} < \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$
- $P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$
- But $\sqrt{\frac{p(1-p)}{n}}$ unknown, so replace with $\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$
- Derivation presumes that the Normal approximation is reasonable, the 10% condition holds, and sample is random

10

Confidence Interval Estimator of p

- CI estimator of p: $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$
  - Standard Error (SE): $\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$
  - Margin of Error (ME): $z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$
  - Confidence Level: $(1-\alpha)$ where $0 < \alpha < 1$
    - For example, if $\alpha = 0.05$ (95% Confidence), then $z_{\alpha/2} = 1.96$
  - Lower Confidence Limit (LCL): $\hat{P} - ME$
  - Upper Confidence Limit (UCL): $\hat{P} + ME$

11

ADHD August: 95% CI Estimator

$$\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = \frac{309}{36{,}319} \pm 1.960\sqrt{\frac{\frac{309}{36{,}319} * \frac{36{,}010}{36{,}319}}{36{,}319}}$$

- $LCL = 0.00756 \; and \; UCL = 0.00945$
- Margin of error (ME) $= 0.00094$; 0.00851 ± 0.00094
- Standard error (SE) $= 0.00048$; $0.00851 \pm 1.960 * 0.00048$

We are 95% confident that among children born *in August* from 2007 through 2009 in any of the 18 U.S. states with a September 1st cutoff for kindergarten, the interval from 75.6 to 94.5 includes the *population* rate of claims-based ADHD diagnosis per 10,000 children. These are the *youngest* kindergarteners.

Does use of data from insurance claims affect interpretation?

12

Difference Between Proportions

- If $\hat{P}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$ & $\hat{P}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$ then $(\hat{P}_2 - \hat{P}_1)$ is Normal because it is a linear combination of independent Normal r.v.'s
  - $\mathrm{E}[\hat{P}_2 - \hat{P}_1] = E[\hat{P}_2] - E[\hat{P}_1] = p_2 - p_1$
  - $\mathrm{V}[\hat{P}_2 - \hat{P}_1] = V[\hat{P}_2] + V[\hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$
  - This tells the sampling distribution of the difference between two sample proportions

13

CI Estimator of $(p_2 - p_1)$

- Confidence Interval Estimator of $(p_2 - p_1)$:

$$(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$$

  - What is the point estimate?
  - Margin of error (ME)?
  - Standard error (SE) of the difference btwn proportions?
  - Assuming that both n_1 and n_2 are sufficiently large?
  - Must the 10% condition be met twice?

14

ADHD: August versus September

$$(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$$

$$\left(\frac{309}{36{,}319} - \frac{225}{35{,}353}\right) \pm 1.960\sqrt{\frac{\frac{309}{36{,}319} * \frac{36{,}010}{36{,}319}}{36{,}319} + \frac{\frac{225}{35{,}353} * \frac{35{,}128}{35{,}353}}{35{,}353}}$$

- Point estimate is 0.00214 with ME of 0.00126
- LCL is 0.00089 and UCL is 0.00340

We are 95% confident that the ADHD diagnosis rate per 10,000 children is from 8.9 to 34.0 *higher* for the youngest kindergarteners versus the oldest. The rate of 85.1 (August born) is considerably higher than 63.6 (September born). The natural randomness in birth month suggests that being younger may cause ADHD diagnoses. 15

Research on Charitable Giving: Karlan and List (2007)

Abstract: We conducted a natural field experiment to further our understanding of the economics of charity. Using direct mail solicitations to over 50,000 prior donors of a nonprofit organization, we tested the effectiveness of a matching grant on charitable giving. We find that the match offer increases both the revenue per solicitation and the response rate. Larger match ratios (i.e., \$3:\$1 and \$2:\$1) relative to a smaller match ratio (\$1:\$1) had no additional impact, however. The results provide avenues for future empirical and theoretical work on charitable giving, cost-benefit analysis, and the private provision of public goods.

Karlan, Dean, and John A. List. 2007. "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment." *American Economic Review*, 97(5): 1774 – 1793. https://www.aeaweb.org/articles.php?doi=10.1257/aer.97.5.1774

16

Excerpt from Table 2

	Control	**1:1 Ratio**	**2:1 Ratio**	**3:1 Ratio**
Imp. price of \$1 public good	1.00	0.50	0.33	0.25
Response Rate	0.018 (0.001)	0.021 (0.001)	0.023 (0.001)	0.023 (0.001)
Observations	16,687	11,133	11,134	11,129

What do column headings mean? Observations?

What does "implied price of \$1 public good" mean?

What does "response rate" mean? (See p. 347 of textbook)

In parentheses are the standard errors:

$s.e.[\hat{P}] = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$ and plugging in get $\sqrt{\frac{0.018(1-0.018)}{16{,}687}} = 0.001$

17

Cross-Tabulation of Raw Data

```
. tabulate ratio gave

           |          gave
     ratio |         0          1 |     Total
-----------+----------------------+----------
         0 |    16,389        298 |    16,687
         1 |    10,902        231 |    11,133
         2 |    10,882        252 |    11,134
         3 |    10,876        253 |    11,129
-----------+----------------------+----------
     Total |    49,049      1,034 |    50,083
```

You work with these data in DACM Module C.2

18

95% CI of Effect of 1:1 Match

- Control group: $\hat{P}_1 = \frac{X_1}{n_1} = \frac{298}{16{,}687} = 0.01786$
- 1:1 treatment: $\hat{P}_2 = \frac{X_2}{n_2} = \frac{231}{11{,}133} = 0.02075$

$$(0.02075 - 0.01786) \pm 1.96\sqrt{\frac{0.02075(0.97925)}{11{,}133} + \frac{0.01786\,(0.98214)}{16{,}687}}$$

$0.00289 \pm 1.96 * 0.00170$

0.00289 ± 0.00332

LCL $= -0.0004$

UCL $= 0.0062$

Infer causality?

We are 95% confident that offering people a 1 to 1 match will affect the percent choosing to donate by a small decrease of 0.04 percentage points to a considerable increase of 0.62 percentage points compared to no match.

95% CI of Effect of Any Match

- Control group: $\hat{P}_1 = \frac{X_1}{n_1} = \frac{298}{16{,}687} = 0.01786$
- All treatments: $\hat{P}_2 = \frac{X_2}{n_2} = \frac{736}{33{,}396} = 0.02204$

$$(0.02204 - 0.01786) \pm 1.96\sqrt{\frac{0.02204(0.97796)}{33{,}396} + \frac{0.01786\,(0.98214)}{16{,}687}}$$

$0.00418 \pm 1.96 * 0.00130$

0.00418 ± 0.00255

LCL $= 0.0016$

UCL $= 0.0067$

Infer causality?

We are 95% confident that offering people any match will increase the percent choosing to donate by 0.16 to 0.67 percentage points compared to no match.

	Control	1:1 Ratio	2:1 Ratio	3:1 Ratio
PANEL A: All States				
Response Rate	0.018 (0.001)	0.021 (0.001)	0.023 (0.001)	0.023 (0.001)
Observations	16,687	11,133	11,134	11,129
PANEL B: Blue States				
Response Rate	0.020 (0.001)	0.021 (0.002)	0.022 (0.002)	0.021 (0.002)
Observations	10,029	6,634	6,569	6,574
PANEL C: Red States				
Response Rate	0.015 (0.001)	0.021 (0.002)	0.024 (0.002)	0.026 (0.002)
Observations	6,648	4,490	4,557	4,547

Selecting the Sample Size

$$\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

- If target a ME ($\hat{P} \pm \tau$) with $(1-\alpha)$ confidence then solve for necessary n

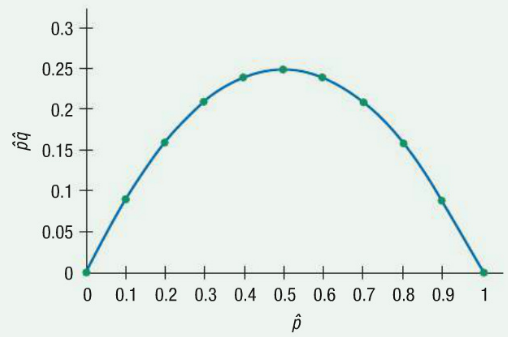
$$\tau = z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$$n = \left(\frac{z_{\alpha/2}}{\tau}\right)^2 \hat{P}(1-\hat{P})$$

- But $\hat{P}$ unknown *before* collect sample
  - Use $\hat{P}$ = 0.5: conservative, n surely big enough (see p. 367)
  - Use $\hat{P}$ = guess: efficient if sure p that above or below 0.5

What $\hat{p}$ Should We Use?

Often you'll have an estimate of the population proportion based on experience or perhaps on a previous study. If so, use that value as $\hat{p}$ in calculating what size sample you need. If not, the cautious approach is to use $\hat{p} = 0.5$. The graph below shows that $\hat{p} = 0.5$ gives the largest value of $\hat{p}\hat{q}$, and hence will determine the largest sample necessary regardless of the true proportion. It's the *worst-case* scenario.



How Many Torontonians to Sample?

- When CP24 hires Ipsos Reid, likely specify desired accuracy of an estimate for a news story (e.g. a margin of error no more than 3%)

$$n = \left(\frac{z_{\alpha/2}}{\tau}\right)^2 \hat{P}(1-\hat{P}) = \left(\frac{1.96}{0.03}\right)^2 0.5 * 0.5 = 1068$$

$$n = \left(\frac{z_{\alpha/2}}{\tau}\right)^2 \hat{P}(1-\hat{P}) = \left(\frac{1.96}{0.03}\right)^2 0.2 * 0.8 = 683$$

  - What if a higher confidence level is desired?
  - What if a smaller margin of error is desired?