Math 170S
Lecture Notes Section 7.3 *†
Confidence intervals for proportions

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

*Version date: Saturday 31st October, 2020, 23:23.
†This notes is based on Hanbaek Lyu’s and Liza Rebrova’s notes from the previous quarter, and I would like to thank them for their generosity. “Nanos gigantum humeris insidentes (I am but a dwarf standing on the shoulders of giants)”.
1 Confidence intervals for proportions: Example

In a certain political campaign, one candidate conducted a poll for which 185 out of 351 voters favor this candidate. This candidate then calculates the percentage of people who voted for them, which is

\[ \frac{185}{351} \approx 0.527. \]

This is higher than 50%, and the candidate now feels confident about winning. Is the candidate’s confidence justified?
2 CI for proportions: Settings

• **Object:** $Y$ is a **Bernoulli** random variable with **unknown parameter** $p$.

• **Input:**
  
  – Random samples $y_1, \ldots, y_n$ for $Y$. Note that each $y_i$ is either 0 or 1.
  
  – Confidence constant $1 - \alpha$

• **Output:** The value $\varepsilon$ that allows us to say

  “$p$ is contained in the interval $[\bar{y} - \varepsilon, \bar{y} + \varepsilon]$ with confidence (approximately) $1 - \alpha$."

The interval $[\bar{y} - \varepsilon, \bar{y} + \varepsilon]$ is the **confidence interval** for $p$. This interval is centered at $\bar{y}$, and the length of the interval is $2\varepsilon$. 
3 The value $\varepsilon$

Theorem 1. In this case, the value $\varepsilon$ is given by

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\bar{y}(1 - \bar{y})}{n}},$$

where $z_{\alpha/2}$ is the real number such that

$$P[N(0, 1) \geq z_{\alpha/2}] = \alpha/2.$$ 

The value $z_{\alpha/2}$ can be computed from the Table V in Appendix B.
4 Answer for Example

Suppose that confidence constant $1 - \alpha$ is 0.95. Then

$$\bar{y} = \frac{185}{351} = 0.527;$$

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\bar{y}(1 - \bar{y})}{n}} = (1.96) \sqrt{\frac{(0.527)(0.473)}{351}} \approx 0.052.$$

The 95% confidence interval is equal to

$$[(0.527) - 0.052, (0.527) + 0.052] = [0.475, 0.579].$$

Hence there is some possibility that $p$ is less than 0.5, so the candidate should be very careful when campaigning as we **cannot say that we are 95% sure that they will win the election.**
Remark 2. Note that textbook has two other formulas ((7.3.4) and (7.3.5) in the textbook) for estimating $\varepsilon$. We will only focus on the formula provided in the lecture notes for now.
5 Settings: One-sided intervals

• **Object:** $Y$ is a **Bernoulli** random variables with **unknown parameter** $p$.

• **Input:**
  
  − Random samples $y_1, \ldots, y_n$ for $Y$. Note that each $y_i$ is either 0 or 1.
  
  − Confidence constant $1 - \alpha$

• **Output:** The value $\varepsilon$ that allows us to say

  “$p$ is contained in the interval $[0, \bar{y} + \varepsilon]$ with confidence $1-\alpha$,”

  or

  “$p$ is contained in the interval $[\bar{y} - \varepsilon, 1]$ with confidence $1-\alpha$.”
6 Value of $\varepsilon$: One-sided

Theorem 3. In this case, the value $\varepsilon$ is given by

$$\varepsilon = z_\alpha \sqrt{\frac{y(1-y)}{n}},$$

where $z_\alpha$ is the real number such that

$$P[N(0,1) \geq z_\alpha] = \alpha.$$

The value $z_\alpha$ can be computed from the Table V in Appendix B.
7 Answer for Example: One-sided

We again take the confidence constant $1 - \alpha$ to be 0.95. In an election, we only care if the candidate gets more than 50% of the vote, so we want

“$p$ is contained in the interval $[\bar{y} - \varepsilon, 1]$ with confidence (approximately) $1 - \alpha$,“

Then we have

$$\bar{y} = \frac{185}{351} = 0.527;$$

$$\varepsilon = z_\alpha \sqrt{\frac{\bar{y}(1 - \bar{y})}{n}} = (1.645) \sqrt{\frac{(0.527)(0.473)}{351}} \approx 0.044.$$
The confidence interval is then equal to

\[
[0.527 - 0.044, 1] = [0.483, 1].
\]

Again we see that \( p \) can be less than 0.5 in this confidence interval, so we cannot say that we are 95% sure that they will win the election.
Let’s try to calculate the winning probability of our candidate, i.e., we want to find $\alpha$ so that

“$p$ is contained in the interval $[0.5, 1]$ with confidence (approximately) $1-\alpha$.”

This implies that (BT)

$$\bar{y} - \varepsilon = 0.5$$

$$z_\alpha \sqrt{\frac{\bar{y}(1-\bar{y})}{n}} = \bar{y} - 0.5$$

$$z_\alpha = \frac{(\bar{y} - 0.5) \sqrt{n}}{\sqrt{\bar{y}(1 - \bar{y})}}$$

$$z_\alpha = \frac{(0.527 - 0.5) \sqrt{351}}{\sqrt{(0.527)(1 - 0.527)}}$$

$$z_\alpha \approx 1.01.$$ 

By Table V, $1 - \alpha$ is approximately 0.8438. So the candidate’s chance of winning is 84.38%.
Example: Comparing two proportions

Two disinfectants were tested for their ability to remove coronavirus from a dry surface.

- The first detergent is successful on 63 out of 91 trials;
- The second detergent is successful on 42 out of 79 trials.

Can we say that one detergent is stronger than the other confidently?
9 Two proportions: Settings

- **Object:** $Y_1, Y_2$ are independent Bernoulli random variables with unknown parameter $p_1, p_2$.

- **Input:**
  - Sample mean $\bar{y}_1$ for $Y_1$ from $n_1$ many samples, and sample mean $\bar{y}_2$ for $Y_2$ from $n_2$ many samples;
  - Confidence constant $1 - \alpha$.

- **Output:** The value $\varepsilon$ that allows us to say

  \[
  \text{“} p_1 - p_2 \text{ is contained in the interval } \left[ (\bar{y}_1 - \bar{y}_2) - \varepsilon, (\bar{y}_1 - \bar{y}_2) + \varepsilon \right] \text{ with confidence (approximately) } 1 - \alpha. \text{”}
  \]
10 The value $\varepsilon$: Two proportions

Theorem 4. In this case, the value $\varepsilon$ is given by

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\bar{y}_1(1 - \bar{y}_1)}{n_1} + \frac{\bar{y}_2(1 - \bar{y}_2)}{n_2}},$$

where $z_{\alpha/2}$ is the real number such that

$$P[N(0,1) \geq z_{\alpha/2}] = \alpha/2.$$

The value $z_{\alpha/2}$ can be computed from the Table V in Appendix B.
11 Answer for Example: Two proportions

Suppose that confidence constant $1 - \alpha$ is 0.9. Then (BT)

$$\bar{y}_1 - \bar{y}_2 = \frac{63}{91} - \frac{42}{79} = 0.16;$$

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{\bar{y}_1(1 - \bar{y}_1)}{n_1} + \frac{\bar{y}_2(1 - \bar{y}_2)}{n_2}}$$

$$= (1.645) \sqrt{\frac{(0.692)(0.308)}{91} + \frac{(0.532)(0.468)}{79}} \approx 0.12192.$$

The confidence interval is then equal to

$$[(0.16) - 0.12192, (0.16) + 0.12192] = [0.03808, 0.28192].$$

Because the interval lies entirely to the right of zero, we can say that the first disinfectant is better than the second one with at least 90% confidence.