Math 140
Introductory Statistics

Professor Bernardo Ábrego
Lecture 21
Sections 8.1

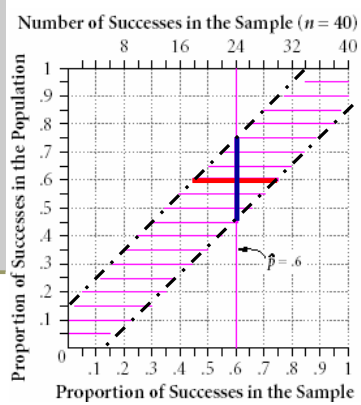
Discussion: Confidence Intervals

- D3. According to the 2000 U.S. Census, about 60% of Hispanics in the United States are of Mexican origin. Would it be reasonably likely in a survey of 40 randomly chosen Hispanics to find that 27 are of Mexican origin?

  Source: www.census.gov/prod/2001pubs/c2kbr01-3.pdf.
- D4. According to the 2000 U.S. Census, about 30% of people over age 85 are men. In a random sample of 40 people over age 85, would it be reasonably likely to get 60% who are men?

  Source: www.census.gov/prod/2001pubs/c2kbr01-10.pdf.
- D5. Suppose that in a random sample of 40 toddlers, 34 know what color Elmo is. What is the 95% confidence interval for the percentage of toddlers who know what color Elmo is?
- D6. Polls usually report a margin of error. Suppose a poll of 40 randomly selected statistics majors finds that 20 are female. The poll reports that 50% of statistics majors are female, with a margin of error of 15%. Use your completed chart to explain where the 15% came from.

Getting a Formula



- The endpoints of the horizontal segment (a reasonably likely interval) are:

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

- The horizontal segments are delimited by two curves that are almost parallel lines of slope 1.
- Thus the vertical segment has about the same length as the horizontal segment.
- Moreover, since the center of each horizontal segment is p, then the endpoints of the vertical segment are:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A Confidence Interval for a Proportion. (Any percent)

- A **confidence interval** for the proportion of successes p in the population is given by the formula

$$\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- n is the sample size, $\hat{p}$ is the proportion of successes in the sample. The value of z depends on how confident you want to be that $\hat{p}$ will be in the confidence interval. For a 95% confidence interval, use z = 1.96; for a 90% confidence interval, use z = 1.645; for a 99% confidence interval, use z = 2.576; and so on.
- This confidence interval is reasonably accurate when **three conditions** are met:
  - The sample was a simple random sample from a binomial population (every subject is either a success or a failure).
  - Both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10.
  - The size of the population is at least 10 times the size of the sample n.

A Confidence Interval and the Margin of Error

- A confidence interval for the proportion of successes p in the population is given by the formula

$$\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$z = \begin{cases} 1.645 & \text{for 90\% confidence} \\ 1.96 & \text{for 95\% confidence} \\ 2.576 & \text{for 99\% confidence} \end{cases}$$

- This confidence interval is reasonably accurate when three conditions are met:
  - The sample was a SRS from a binomial population (every subject is either a success or a failure).
  - $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.
  - The size of the population is at least 10 times the size of the sample n.

- The quantity

$$E = z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

is called **the margin of error**. It is half the length of the confidence interval.

Example: Safety Violations.

- Suppose you have a random sample of 40 buses from a large city and find that 24 have a safety violation. Find the 90% confidence interval for the proportion of all buses that have a safety violation.

Activity 8.3 (Simulated)

1. Generate a random sample of 40 numbers between 0 and 9.
2. Count the number of even digits in your sample of 40.
3. Construct a 95% confidence interval for the proportion of random digits that are even.
4. Repeat a 100 times and draw all of the intervals in the appropriate display (like Display 8.5 in p. 426)
5. What is the true proportion of all random digits that are even?
6. What percentage of the confidence intervals captured the true proportion? Is this what you expected? Explain.

- Examples:

1 6 9 6 3 0 9 1 3 1 2 8 3 5 6 0 0
0 6 7 8 3 1 4 9 6 5 6 9 6 7 5 3 2
6 8 1 4 1 2

Even: 20 $\hat{p} = \frac{20}{40} = 0.5$

95% Confidence Interval

$$0.5 \pm (1.96)\sqrt{\frac{0.5(1-0.5)}{40}} =$$

$$(0.34505, 0.65495)$$

- 8 9 4 4 2 6 6 8 7 4 9 6 3 4 8 8 2
7 4 4 2 0 3 6 1 6 0 5 2 0 9 8 2 7
2 1 7 2 5 8

Even: 27 $\hat{p} = \frac{27}{40} = 0.675$

95% Confidence Interval

$$0.675 \pm (1.96)\sqrt{\frac{0.675(1-0.675)}{40}} =$$

$$(0.52985, 0.82015)$$

Activity 8.3 (Simulated)

- Examples:

1 6 9 6 3 0 9 1 3 1 2 8 3 5 6 0 0
0 6 7 8 3 1 4 9 6 5 6 9 6 7 5 3 2
6 8 1 4 1 2

Even: 20 $\hat{p} = \frac{20}{40} = 0.5$

95% Confidence Interval

$$0.5 \pm (1.96)\sqrt{\frac{0.5(1-0.5)}{40}} =$$

$$(0.34505, 0.65495)$$

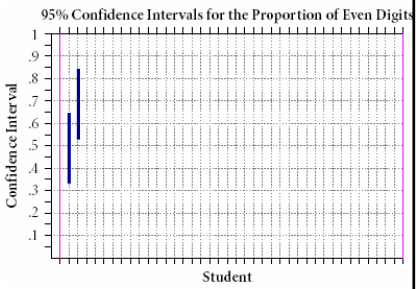
- 8 9 4 4 2 6 6 8 7 4 9 6 3 4 8 8 2
7 4 4 2 0 3 6 1 6 0 5 2 0 9 8 2 7
2 1 7 2 5 8

Even: 27 $\hat{p} = \frac{27}{40} = 0.675$

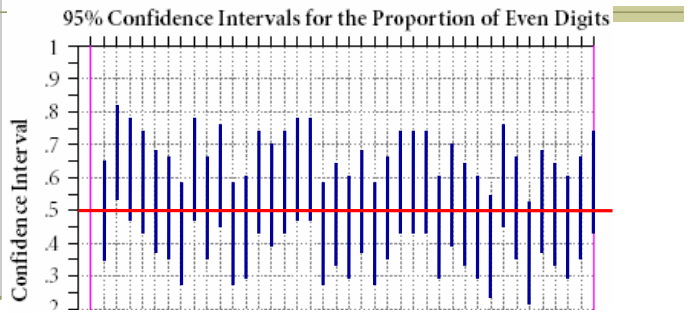
95% Confidence Interval

$$0.675 \pm (1.96)\sqrt{\frac{0.675(1-0.675)}{40}} =$$

$$(0.52985, 0.82015)$$



Activity 8.3 (Simulated)



How many intervals contain the true proportion of even numbers $p = 0.5$?

Answer: all except one, that is 38 out of 39. Or 97.43% of all of the intervals.

In general, if we randomly calculate a large number of 95% Confidence Intervals we should expect that about 95% of them will contain the true value of p.