Announcements

- One Handout Today
- Previous Lecture's handout up front
- Unclaimed exams & homeworks (1 & 2) are up front
- Did you accidently pick up an exam that wasn't yours? Please return it to the front table, no questions asked.

Chapter 19: The Margin of Error in the Media and the Confidence Interval in Statistics (p. 357)

- "Margin of Error" as it is called in the media and "Confidence Interval" as it is called your book (p. 356-357) are closely related and they are both indicators of the "strength" or "believability" or "accuracy" of a statistic They are expressions of confidence in what conclusions you might draw from a survey result.
- A "margin of error" as reported in the popular press generally implies a confidence interval with a confidence level of 95%.

Calculating the "Margin of Error" from a sample percentage

- See the WSJ handout for the Harris Poll (telephone)
- Let's treat "vote for Bush" as a "success" (p) all other responses as a "failure" (q, AKA 1-p). The article goes on to note that the margin of error is plus or minus 3% and 1092 persons nationwide were deemed likely to vote.
- The Harris Poll Indicated that 49% of likely voters would vote for Bush.

Method

- Let's then **substitute** our percentages from the sample ($\hat{p}$) for our population percentages p (we will pretend that our parameters are actually unknown even though today we now know what happened).
- We can now calculate the **standard error** of $\hat{p}$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{.49 * .51}{1092}} = .0151$$

- As long as n is large, and large is

$$n\hat{p} \geq 10 \text{ or } n(1-\hat{p}) \geq 10$$

- Then the sampling distribution is approximately normal with mean p and standard error SE $_{\hat{p}}$

- **Using our table of standard normal curve areas, we can find a value Z* such that a CENTRAL AREA of .95 (or 95%) falls between +Z* and –Z* (and this means .025 in each tail area). The Z* which would satisfy this is +1.96 and –1.96**
- For any normal distribution then, approximately 95% of the values are within ±1.96 standard deviations of the mean. Since for large samples the sampling distribution of a sampling proportion $\hat{p}$ is approximately normal with a mean p and $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$
- The confidence interval has the form $\hat{p} \pm Z^{*} SE(\hat{p})$
- If we multiply the SE of .0151 by 1.96 (this is Z*) and then report "49% + or – 3 percentage points" we have constructed a 95% confidence interval.

- And what it says is that we expect 95% of all samples of this size (1,092) to generate a statistic $\hat{p}$ that is within ±3% of the true population proportion, p.
- It implies that while 95% of all possible samples of this size should generate statistics that fall within 3% of the "truth" (population parameter) unfortunately, 5% (100%-95%) of the time, just by chance, the samples will generate statistics that fall outside of this interval.
- The formula given in your text for a 95% confidence interval is: (p. 359) $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$
- This can be used as long as long as n is large
- Note that when you substitute $\hat{p}$ for p, your book calls the standard deviation "the standard error of a statistic" (p. 355)

Understanding Confidence

- A CONFIDENCE INTERVAL then is a range of values (i.e. values derived from sample information) that we think covers or contains the true (unknown) parameter.
- The Harris Poll says 49% of those surveyed would vote for Bush if the election were held tomorrow with a margin of error of 3%. This suggests a range around the sample statistic of 49% is 46% to 52%. This interval of 46% to 52% is supposed to "cover" or "contain" the true percentage among American voters.
- This plus or minus 3% is effectively the same as plus or minus 2 Standard Errors (1.96 to be exact) and this is the way the media typically expresses results from polls. What they are saying is that they are "95% confident that the interval 46% to 52% covers or contains the true percentage of all voters who would cast a vote for Bush if the election were held right now.

Interpretation

- IF WE COULD REPEAT THIS PROCEDURE 100 TIMES, 95 OF THOSE TIMES (or 95%) OF THOSE TIMES, WE WOULD HAVE SELECTED A SAMPLE SUCH THAT THE CONFIDENCE INTERVAL GENERATED CONTAINS OR COVERS THE TRUE POPULATION PARAMETER.
- 5 OF THOSE TIMES, THE INTERVAL WE GIVE YOU DOES NOT! WE HOPE THAT THIS IS ONE OF THE GOOD 95 AND NOT ONE OF THE BAD 5.
- The level of confidence (e.g. 68%, 90%, 95%, 99%) have corresponding Z- values (i.e. 1, 1.64 or 1.65, 1.96, 2.58). Typically, the popular media uses 95% or a Z=1.96 to construct the confidence interval.

Properties of Confidence (p.360-361)

- You can never been 100% confident. There is always the chance that you could have a very bad sample and are nowhere near the true population parameter.
- Other things equal, as confidence decreases, the interval grows narrower and it is more precise but less certain, as confidence level increases the interval grows wider and is less precise but more certain to contain the parameter.
- Other things equal, as your sample size increases, your interval grows narrower and it is more precise. As your sample size decreases, the interval is wider and less precise.
- Other things equal, if your standard error increases, the interval becomes wider and less precise. As your standard error decreases, the interval becomes narrower and more precise.