**Stat 301 – Day 11**

**Confidence intervals and Sampling from Finite Population**

|  |
| --- |
| **Last Time:** Normal approximation to the binomial → another way to get confidence intervals* **Goal of a confidence interval:** To estimate the parameter of interest from the sample statistic with a set of plausible values for the parameter
* One proportion *z*-confidence interval (aka Wald interval)
	+ $\hat{p}$ + z\*$\sqrt{\hat{p}\left(1-\hat{p}\right)/n} $ *estimate + margin of error*
	+ Validity conditions: Binomial process, At least 10 successes and at least 10 failures in the sample
* Interpretation of confidence interval: I’m 95% confident that <<parameter>> is between <<lower>> and <<upper>>
* Interpretation of confidence level: This method works (the parameter is inside the sample interval) 95% of the time (lots of random samples from sample process) in the long run
* Advantages of Wald interval over Exact Binomial interval
 |

**So what do we do if we have a small sample size and validity conditions are not met?**

The *Adjusted Wald procedure* uses a different estimate and a different standard error. The main idea is the sample proportion is pulled away from 0 and 1 by “shrinking” it towards 0.50:

 $\tilde{p}=\frac{X+.5z^{2}}{n+z^{2}}, \tilde{n}=n+z^{2}$

Suppose we want 95% confidence, these values are approximately?

**Key Idea:** This Plus Four Method (95% confidence) has been shown to have a coverage rate much closer to the desired confidence level, even with small sample sizes, and to not be as wide as Binomial confidence intervals.

**Example 1:** From Investigation 1.11

(a) Calculate a 95% confidence interval for the probability of death in a heart transplant operation at St. George’s hospital based on 8 deaths in the last 10 operations

 (i) One proportion *z* interval

 (ii) Adjusted Wald/Plus Four method

**Example 2:**

* Open the **Sampling Words** [applet](http://www.rossmanchance.com/applets/OneSample.html?population=gettysburg)
* Use pull-down menu to change **Variable** to “Short”

(a) Do you know the value of the population parameter?

* Check **Show Sampling Options**
* Set **Sample size** to 5
* Press **Draw Samples**

(b) Do you know the value of the sample proportion?

* Press **Draw Samples** 9 more times.

(c) Are you starting to see a pattern to the distribution of sample proportions?

* Change the **Number of samples** to 990 (for 1000 total) and **press Draw Samples**.

(d) Does this appear to be an unbiased sampling method? How are you deciding?

(e) Does the distribution of sample proportions appear to be well-modelled by a normal distribution? (Are you surprised?)

(f) What are the mean and SD values of the distribution of sample proportions?

(g) Predict what will happen to the distribution of sample proportions (shape, center, variability) if we change the sample size to 10.

* Select the **Fixed** radio button
* Select 1000 samples

(h) Were your predictions correct?

(i) Repeat (g) and (h) for samples of 20 words.

|  |
| --- |
| **Key Result** (p. 102)**:** The Central Limit Theorem for a sample proportion states that when drawing random samples from a *large but finite population* with a large enough sample size, then the sampling distribution of the sample proportion $\hat{p}$ will be well modeled by a normal distribution with mean equal to $π$, the population proportion of successes, and standard deviation equal to$SD\left(\hat{p}\right)=\sqrt{π(1-π)/n}$.We consider the sample size large enough if n× $π$ > 10 and n×(1 – $π$) > 10.We consider the population size (*N*) large if it is more than 20 times the size of the sample (*n*). |

**Example 3:** Investigation 1.16 (p. 110)