Monday, Oct 4

Confidence Intervals for μ and p

Recall that the "recipe" for a confidence interval is

$$\overbrace{\text{point estimate} \pm \underbrace{\text{standard score} \times \text{standard error}}_{\text{margin of error}}}^{\text{confidence interval}}.$$

The confidence interval for μ is

$$\bar{x} \pm t\frac{s}{\sqrt{n}},$$

where t is determined by the desired *confidence level* and the *degrees of freedom* $(n-1)$.

The confidence interval for p is

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

where z is determined by the desired *confidence level*.

Note: These assume that we are (a) sampling with replacement, (b) sampling from an infinite population of units, or (c) that we are sampling without replacement but that N is much larger than n so that $1 - n/N \approx 1$ and so we can ignore it.

Example: A molecular biologist estimated that the probability of success of a polymerase chain reaction (PCR) under certain conditions is 0.7, with a margin of error of 0.1. What is the *confidence interval* for the probability of success of the PCR?

Example: Conservation researchers conducted a survey to estimate the density of larkspur in a region. They did this by dividing the region into transects and then selecting some of the transects at random to observe. Their *estimate* was 0.76 larkspur per square meter. They also reported that the *standard error* was 0.11 larkspur per square meter. What is the margin of error and confidence interval for the density of larkspur in the region?

Example: Fisheries researchers estimated the total number of Northern Pike in a lake using a mark and recapture survey (more on that later). They reported a *confidence interval* of 28000 to 30000 pike. What is the *point estimate* and the *margin of error* for the number of Northern Pike in the lake?

Choosing a Sample Size for Estimating p with $\hat{p}$

How do we *choose* n for estimating p with $\hat{p}$? Consider that

$$m = z\sqrt{\frac{p(1-p)}{n}} \quad \Leftrightarrow \quad n = \frac{z^2 p(1-p)}{m^2}.$$

Then how do we choose p? Two approaches:

1. A *good guess* of p based on experience and/or expertise.
2. An *upper bound* on the sample size by using $p = 0.5$ because

$$n = \frac{z^2 p(1-p)}{m^2} \leq \frac{z^2 0.5(1-0.5)}{m^2}.$$

Example: Recall the survey of Hobbits with foot lice. In that survey it was found that in a sample of 100 Hobbits, 25 had foot lice. How large of a sample size would we need to get a margin of error of 0.01 to estimate the proportion of Hobbits with foot lice if we use a confidence level of 95%?

Example: The *sensitivity* of a medical test is the probability of a positive result when the condition (e.g., disease) is present. Rapid strep tests have a sensitivity of about 0.95. A new rapid strep test has been developed but its sensitivity is unknown. How many tests would be necessary to estimate its sensitivity with a margin of error of 0.02 if we use a confidence level of 95%?

Choosing a Sample Size for Estimating μ with $\bar{x}$

How do we choose n for estimating μ with $\bar{x}$? Consider that

$$m = z\frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad n = \frac{z^2\sigma^2}{m^2}$$

Then how do we choose σ? A good guess based on experience and/or expertise.

Example: Recall the twin study comparing left hippocampus volume between twins unaffected and affected by schizophrenia. How many twin pairs would we need to get the margin of error for estimating μ to 0.05 cubic cm, assuming a confidence level of 95%?

Example: A survey of $n = 31$ black cherry trees yielded a point estimate of μ of 30.2 cubic feet. The sample standard deviation was 16.4 cubic feet, so the margin of error was about 6.1 cubic feet, assuming a confidence level of 95%. What n would be necessary for a new survey that would produce a margin of error of (approximately) 2 cubic feet?