STA13-B
Elementary Statistics
Fall 2007

Lecture 21

Instructor: Katie Pollard

Announcements

- Midterm solutions are posted
- Turn in Homework 6 by section (B01-B07)
- Homework 7 is due next Thursday 11/29

  (10 problems on these last 2 assignments)

Chapter 9

- CI widths
- Bound on error of estimation
- Sample size formula

Estimating CIs

- Confidence intervals are of the form

  estimate ± (critical value)(standard error)
- Standard error=estimated population standard deviation divided by the square root of *n*.
- Distribution to use for the critical value:

  π and μ with σ known : N(0,1)

  μ with σ unknown: $t(df=n-1)$
- What determines CI width?

Comparing CI Widths

- For a fixed confidence level α, the $100(1-\alpha)\%$ CI (for μ or π) will decrease in width when:
  - sample size *n* increases
  - the population distribution is narrower (small σ)

  These both decrease the standard error.
- For a fixed sample size *n*, the CI will decrease in width when the confidence level α decreases (*i.e.* $z_{\alpha/2}$ smaller in magnitude).

Coverage Probability

- It is tempting, but incorrect, to say that the probability that a particular 95% CI contains the population characteristic (*e.g.* π or μ) is equal to 0.95.
- The correct interpretation is that ~95% of the time the random interval given by the CI will contain the true value of the parameter.
- A particular CI either contains the population characteristic or it does not.

Rules of Thumb

- Always need to have a random sample of size $n \leq 10\%$ of population
- For π (z critical value)

$$np \geq 10 \text{ AND } n(1-p) \geq 10$$

- For μ when σ is known (z critical value) or μ when σ is unknown (t critical value)

$n \geq 30$ OR data approximately normal

To test for normality, make a boxplot.

Bound on Error of Estimation

- When the sampling distribution of a test statistic is approximately normal, then the bound on the error of estimation for a 100(1-α)% CI is B = ($z_{\alpha/2}$)(standard error).
- B is half the total width of the CI.
- We have 100(1-α)% confidence that the estimate is within B of the population parameter of interest.

Sample Size Formula

- We can use a target level for B to determine a reasonable sample size *n* for a study.
- To estimate the population parameter to within a set amount B with 100(1-α)% confidence, the required sample size is

$$n = \left(\frac{(z_{\alpha/2})(\text{sd})}{\text{B}} \right)^2$$

where sd is the population standard deviation.

Sample Size Computation

Example: What is the sample size needed to estimate the proportion of breast cancer patients that go into remission after treatment to within 5% with 99% confidence if the true proportion were $\pi = 0.75$?

$\alpha = 0.01 \rightarrow z_{\alpha/2} = z_{0.005} = -2.58$, $B = 0.05$

$sd = \sqrt{\pi(1 - \pi)} = \sqrt{(0.75)(0.25)} = 0.38$

$n = ((-2.58)(0.38)/0.05)^2 = 384.5 \approx 385$

Practical Tips

- The population standard deviation is typically not known in advance.
- It is common practice to estimate it from a preliminary study or to use a conservative guess (*i.e*. one that gives a larger n):
  - For the population proportion, using $\pi = 0.5$ gives the maximum possible sd: sd=0.25.
  - For the sample mean, using sd=range/4 is a good guess when the distribution is not too skewed.

Components You Input into the Sample Size Formulas

- B is the bound on error of estimation. This means how close want to be to the true value.
- The confidence level is the same critical value used in the CI. This is from N(0,1), so assumes π and σ are known.
- Unknown parameters: use conservative values
  - $\pi=0.5$ for proportion
  - $\sigma=\text{range}/4$ for mean

Sample Size Computation

Example: What is the sample size needed to estimate the proportion of breast cancer patients that go into remission after treatment to within 5% with 99% confidence if the true proportion is unknown?

$\alpha = 0.01 \rightarrow z_{\alpha/2} = z_{0.005} = -2.58$, B = 0.05

Use $\sigma = \sqrt{\pi(1-\pi)} = \sqrt{(0.5)(0.5)} = 0.5$

$n = ((-2.58)(0.5)/0.05)^2 = 665.6 > 385$

Example on Board