Online Review Course of
Undergraduate Probability and Statistics

Review Lecture 11
Confidence Intervals

Chris A. Mack
Adjunct Associate Professor

Course Website: www.lithoguru.com/scientist/statistics/review.html

Sampling Distribution of the Mean

- Let $X_1, X_2, \ldots, X_n$ be iid random variables from an infinite population with mean μ and finite variance σ^2.

$$\bar{X} = \frac{1}{n}\sum_{i=1,n} X_i \qquad E(\bar{X}) = \mu \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Central Limit Theorem:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \; as \; n \to \infty$$

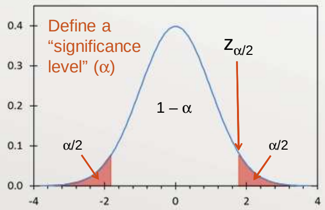
Sampling Distribution of the Mean

- We say that $\bar{X}$ is an estimator for the population mean, and that $\bar{x}$ is a point estimate of the population mean
  - We want our estimator to be unbiased (or low bias)
  - We want our estimator to have a small variance (standard error)
  - An efficient estimator is one that has lower bias and/or smaller variance than another estimator
- Our estimator $\bar{X}$ is unbiased provided our sample is random (each X_i is independent), and it is the most efficient
  - For a symmetric distribution, the median is an unbiased estimator, but is less efficient than the sample mean

Errors in our Estimate of the Mean

- How large might $\bar{X} - \mu$ get?
  - We know by the central limit theorem that for a moderately large sample $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

α	$z_{\alpha/2}$
0.10	1.65
0.05	1.96
0.02	2.33
0.01	2.58



Creating a Confidence Interval

- Let's look at the two boundary points, $\pm z_{\alpha/2}$

$$\pm z_{\alpha/2} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow \mu = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- The probability that the true mean will fall within this range is 1 – α
- Generally, we approximate σ with s

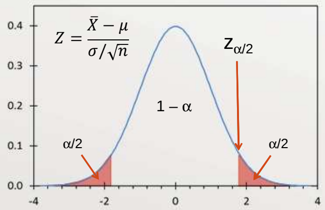
Interval Estimate of the Mean

- Confidence interval (CI):

Large Sample Confidence Interval — Margin of Error

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$z_{\alpha/2}$	α
1.65	0.10
1.96	0.05
2.33	0.02
2.58	0.01



Confidence Intervals

- Pick the value of α you want
- Do you have a large enough sample so that the sampling interval can be considered normal?
  - If yes, use the $z_{\alpha/2}$ value
  - If no, but underlying distribution is normal, use $t_{\alpha/2}$ value
- Create the confidence interval
  - We are (1-α)100% confident that the true population mean is captured by our interval
  - If we ran this experiment 100 times, we expect that our confidence intervals would capture the true mean (1-α)100 of those times
- Every statistic has a confidence interval (not just the mean)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Another Statistic: Proportion

- Proportion examples:
  - What percentage of UT students smoke?
  - How many people prefer brand X over brand Y?
  - What fraction of the molecules have reacted?
- With two options, the population will follow a binomial distribution
  - p = population proportion (probability) of "success"
  - q = 1 – p = proportion (probability) of "failure"

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Another Statistic: Proportion

- Proportion estimator
  - Sample size = n, X_i = Bernoulli RV
  - $X = \sum_{i=1,n} X_i$ = binomial RV

Recall from Lecture 8

$$\hat{p} = \frac{X}{n} \qquad E(\hat{p}) = \frac{E(X)}{n} = \frac{np}{n} = p$$

$$var(\hat{p}) = \frac{var(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} = \frac{pq}{n}$$

Standard Error $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Proportion Confidence Intervals

- The binomial distribution can be well approximated with a normal distribution whenever $np > 10$ and $nq > 10$
- When this is the case, use $z_{\alpha/2}$
  - Margin of error = $z_{\alpha/2} SE(\hat{p})$
- Example: 95% CI (α = 0.05)
  - 95% CI = $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Review #11: What have we learned?

- Explain estimator, point estimate, and interval estimate
- Under what conditions can we use the normal approximation for the sampling distribution of the mean?
- Know how to generate a confidence interval for any statistic
- What is the standard error for a proportion estimate?
- Under what conditions can we use the normal approximation for the sampling distribution of the proportion?