STA 291
Spring 2009

1

LECTURE 16
TUESDAY, 31 March

Central Limit Theorem

- Thanks to the CLT …
- We know $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is approximately standard normal (for sufficiently large *n*, even if the original distribution is discrete, or skewed).
- Ditto $\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$

Example

- The scores on the Psychomotor Development Index (PDI) have mean 100 and standard deviation 15. A random sample of 36 infants is chosen and their index measured. What is the probability the *sample mean* is below 90?

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{90 - 100}{15/\sqrt{36}} = -4$$

- If we *knew* the scores were normally distributed and we randomly selected a single infant, how often would a *single* measurement be below 90?

$$z = \frac{X - \mu}{\sigma} = \frac{90 - 100}{15} = -0.67$$

- **Statistical Inference: Estimation**
  - Inferential statistical methods provide predictions about characteristics of a population, based on information in a sample from that population
  - For quantitative variables, we usually estimate the population mean (for example, mean household income)
  - For qualitative variables, we usually estimate population proportions (for example, proportion of people voting for candidate A)

Two Types of Estimators

- Point Estimate
  - A single number that is the best guess for the parameter
  - For example, the sample mean is usually a good guess for the population mean
- Interval Estimate
  - A range of numbers around the point estimate
  - To give an idea about the precision of the estimator
  - For example, “the proportion of people voting for A is between 67% and 73%”

Point Estimator

- A point estimator of a parameter is a (sample) statistic that predicts the value of that parameter
- A good estimator is
  - ***unbiased***: Centered around the true parameter
  - ***consistent***: Gets closer to the true parameter as the sample size gets larger
  - ***efficient***: Has a standard error that is as small as possible

Unbiased

- Already have *two* examples of unbiased estimators—
- Expected Value of the $\overline{X}$'s: μ—that makes $\overline{X}$ an unbiased estimator of μ.
- Expected Value of the $\hat{p}$'s: p—that makes $\hat{p}$ an unbiased estimator of p.
- Third example: $s^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X_i}\right)^2$

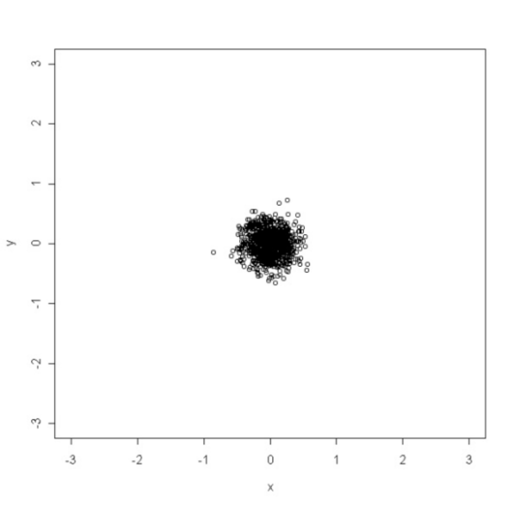
Efficiency

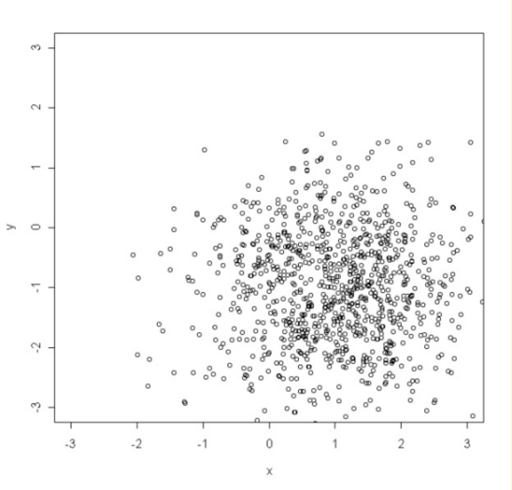
- An estimator is *efficient* if its standard error is small compared to other estimators

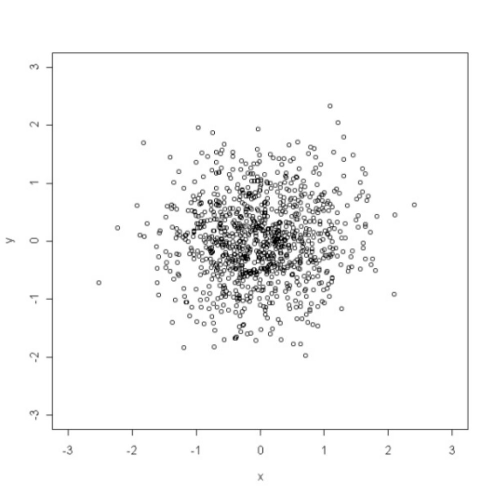
- Such an estimator has high precision

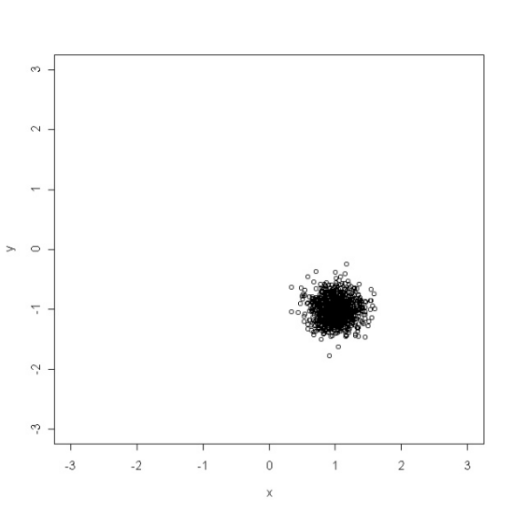
- A good estimator has ***small standard error*** and ***small bias*** (or no bias at all)

Bias versus Efficiency









Confidence Interval

- An inferential statement about a parameter should always provide the probable accuracy of the estimate
- How close is the estimate likely to fall to the true parameter value?
- Within 1 unit? 2 units? 10 units?
- This can be determined using the sampling distribution of the estimator/ sample statistic
- In particular, we need the standard error to make a statement about accuracy of the estimator

Confidence Interval—Example

- With sample size n = 64, then with 95% probability, the sample mean falls between

$$\mu - 1.96\frac{\sigma}{\sqrt{64}} = \mu - 0.245\sigma \quad \& \quad \mu + 1.96\frac{\sigma}{\sqrt{64}} = \mu + 0.245\sigma$$

Where μ = population mean and

σ = population standard deviation

Confidence Interval

- A confidence interval for a parameter is a range of numbers within which the true parameter likely falls

- The probability that the confidence interval contains the true parameter is called the *confidence coefficient*

- The confidence coefficient is a chosen number close to 1, usually 0.95 or 0.99

Confidence Intervals

- The sampling distribution of the sample mean $\bar{X}$ has mean μ and standard error $\dfrac{\sigma}{\sqrt{n}}$

- If n is large enough, then the sampling distribution of $\bar{X}$ is approximately normal/bell-shaped (Central Limit Theorem)

Confidence Intervals

- To calculate the confidence interval, we use the Central Limit Theorem
- Therefore, we need sample sizes of at least, say, $n = 30$
- Also, we need a z–score that is determined by the confidence coefficient
- If we choose 0.95, say, then $z = 1.96$

Confidence Intervals

- With 95% probability, the *sample mean* falls in the interval

$$\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}$$

- Whenever the sample mean falls within 1.96 standard errors from the population mean, the following interval contains the *population mean*

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

Attendance Question #17

Write your name and section number on your index card.

Today's question (Choose one):