Estimation

Day 3

Do The Math

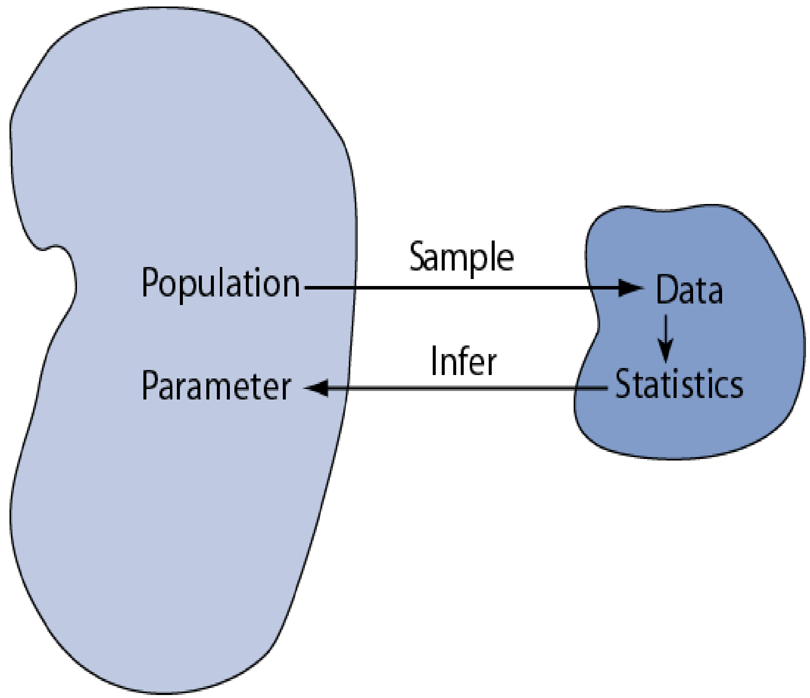
Chicago, IL
Bangalore, India
www.mu-sigma.com

2016

Contents

Statistical inference is the act of generalizing from a sample to a population with calculated degree of certainty

We want to learn about the population parameters…



…but we can only calculate *sample statistics*

$$\mu \longleftarrow \bar{x}$$

Statistical inference two steps – estimation and testing of hypothesis

Inference

Estimation

Testing of Hypothesis

Point Estimation

Interval Estimation

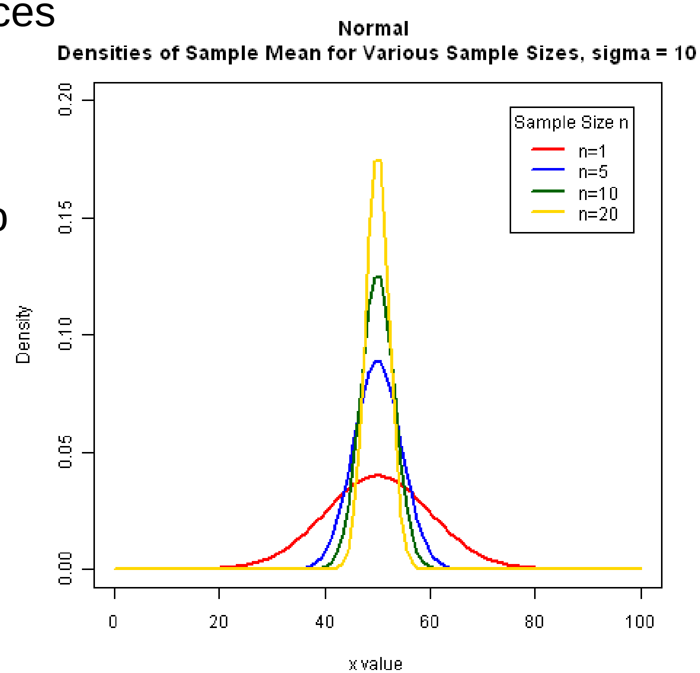
	Parameters	Statistics
Source	Population	Sample
Calculated?	No	Yes
Constants?	Yes	No
Examples	μ, σ, p	$\bar{x}, s, \hat{p}$

It is not enough just to estimate the parameter – one also needs to estimate the error to get a sense of certainty

- How precisely does a given sample mean (x-bar) reflect underlying population mean (μ)? How reliable are our inferences?
- What is the chance that I am observing this sample by fluke
- The standard error of the sample mean is the standard deviation of the sample mean's estimate of the population mean

Drawing inference from the samples

- To draw inference from the samples, we need to compute some appropriate numbers from the random sample
- To draw inference about the mean of the population (μ), we compute the sample mean $\bar{X}$
- The distribution of $\bar{X}$ is called the sampling distribution of the sample mean
- The sampling distribution is the basis of making inferences
- For instance if $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$
- As sample size increases the sample mean is subject to less and less variation



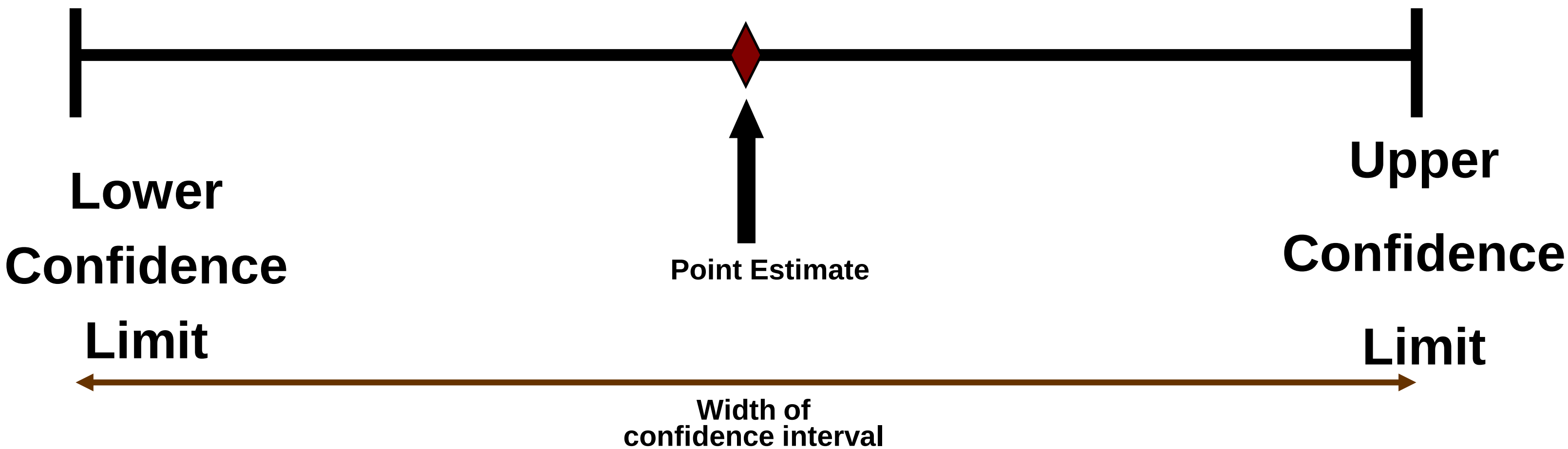
Estimate of population mean and its standard error

- The unknown population mean is μ is estimated by $\bar{x}$
- However estimate of population mean, is not always enough. We need to know how good this estimate is, i.e., on an average how much is my sample mean ($\bar{x}$) is going to vary
- An estimate of this standard deviation is given by $\frac{\sigma}{\sqrt{n}}$
- However, σ is not generally known, and it is estimated by the variance of the sample denoted by ***s***
- The expression, $\frac{s}{\sqrt{n}}$ is called the standard error of the estimate $\bar{x}$
- From the theory of sampling distributions, we have, if $\bar{x} \sim N(\mu, \sigma^2)$,

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$ and $$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$ (t – distribution with (n-1) degrees of freedom)

A confidence interval estimate provides more information about the population parameter than a point estimate

- A **point estimate** is just a single number – we are unaware of any kind uncertainty that may be associated with it
- A **confidence interval estimate** gives a range or an interval along with an associated **level of confidence**

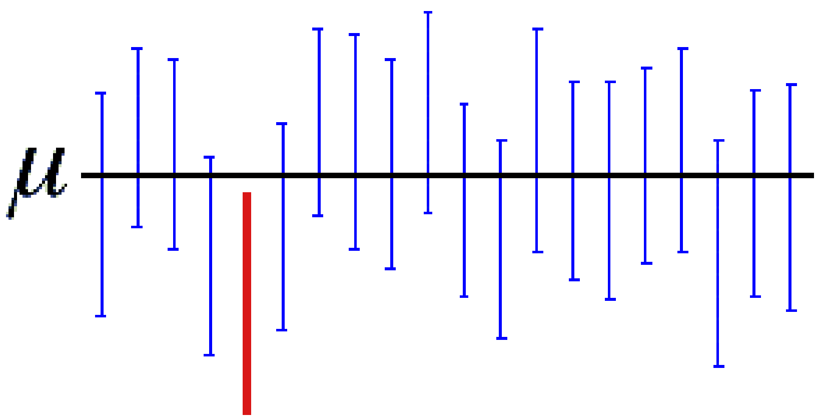


Mu Sigma Confidential

Factors that may affect confidence intervals

- Width of the confidence intervals are determined by three factors
  - The sample size ***n***
  - The variability in the population, usually **σ** or ***s***
  - The desired **level of confidence**

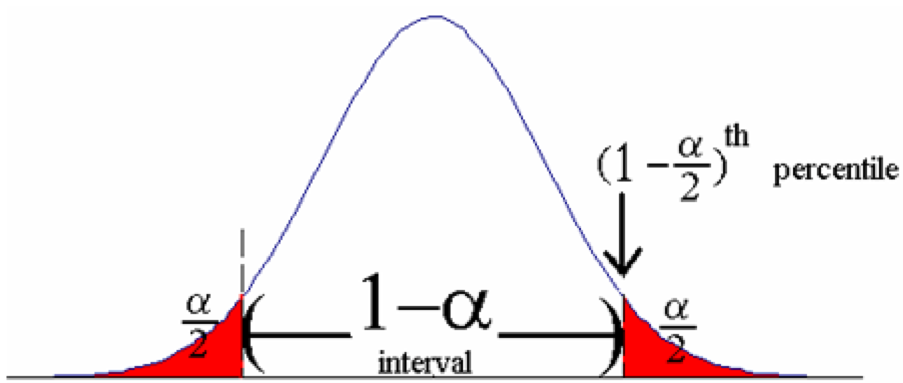
- For a 95% confidence interval about 95% of the similarly constructed intervals will contain the population parameter being estimated



A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population parameter.

Confidence Intervals for an unknown population mean and known variance

- The construction of a confidence interval for the population mean depends upon
  - The point estimate of the population
  - The level of confidence
  - The standard deviation of the sample mean $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
- Because the sample mean is normally distributed, we know that 95% of all sample means should lie within 1.96 standard deviations of the population mean and 2.5% will lie in each tail



Confidence interval is only a random interval as the end points of the interval are random variables

- A correct way to interpret confidence interval is – "if an infinite number of random samples are collected, and a $100(1 - \alpha)\%$ confidence interval for μ is computed from each sample, $100(1 - \alpha)\%$ of these intervals will contain the true value of μ"

- In practice we obtain only one random sample and calculate one confidence interval

- Since this interval either will or will not contain the true value of μ it is not reasonable to attach a probability level to this specific event

Confidence interval will vary based on the sample obtained

- A confidence interval estimate for μ is an interval of the form L ≤ μ ≤ U such that,

$$P\left(L \leq \mu \leq U\right) = 1 - \alpha$$

- Since Z follows a standard normal distribution, we have

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

- Manipulating the quantities inside the probability expression, we get

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence Intervals for an unknown population mean and unknown variance

- Confidence interval for mean where σ is unknown is given by

$$\overline{X} \pm t_{\frac{\alpha}{2},\, n-1} \frac{S}{\sqrt{n}}$$

- It follows Student's t-distribution with n-1 degrees of freedom

- The area in the tails of the t distribution is a little greater than the area in the tails of the standard normal distribution. This result is because we are using s as an estimate of which introduces more variability to the t statistic

- As the sample size increases, however, the t distribution approaches the standard normal distribution,

Confidence Interval for the Mean – Example using the t-distribution

- A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation of 0.09 inch. Construct a 95 percent confidence interval for the population mean. Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inches?

Given in the problem:

$$n = 10$$

$$\bar{x} = 0.32$$

$$s = 0.09$$

Compute the C.I. using the t-dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

Student's t-distribution Table

Compute the C.I.

using the t-dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,10-1} \frac{s}{\sqrt{n}}$$

$$= 0.32 \pm t_{.025,9} \frac{0.09}{\sqrt{10}}$$

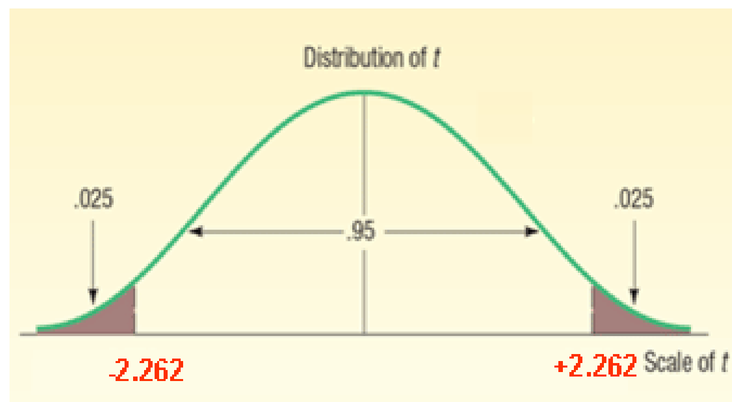
$$= 0.32 \pm 2.262 \frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude: the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169



Confidence Interval Estimates for the Mean – Example

- The manager of the Inlet Square Mall, near Ft. Myers, Florida, wants to estimate the mean amount spent per shopping visit by customers.

- A sample of 20 customers reveals the following amounts spent.

$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

Confidence Interval Estimates for the Mean – By Formula

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

$$= \bar{X} \pm t_{.05/2,20-1} \frac{s}{\sqrt{n}}$$

$$= 49.35 \pm t_{.025,19} \frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 2.093 \frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 4.22$$

The endpoints of the confidence interval are \$45.13 and \$53.57

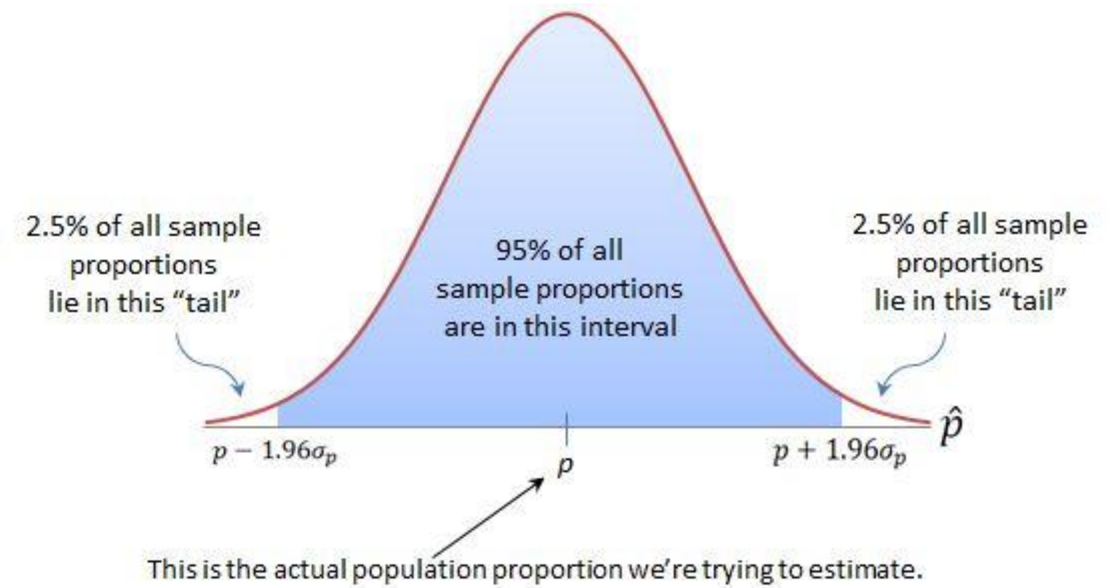
Conclude : It is reasonable that the population mean could be \$50.

The value of \$60 is not in the confidence interval. Hence, we conclude that the population mean is unlikely to be \$60.

Confidence Interval for a Population Proportion

- A confidence interval for a population proportion is constructed by taking the point estimate (pˆp^) plus and minus the margin of error. The margin of error is computed by multiplying a z multiplier by the standard error, SE(pˆ).

$$\hat{p} \pm z^* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$



Confidence Interval for a Population Proportion - Example

- The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. According to BBA union bylaws, at least three-fourths of the union membership must approve any merger. A random sample of 2,000 current BBA members reveals 1,600 plan to vote for the merger proposal. What is the estimate of the population proportion?

- Develop a 95 percent confidence interval for the population proportion. Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

First, compute the sample proportion:

$$p = \frac{x}{n} = \frac{1{,}600}{2000} = 0.80$$

Compute the 95% C.I.

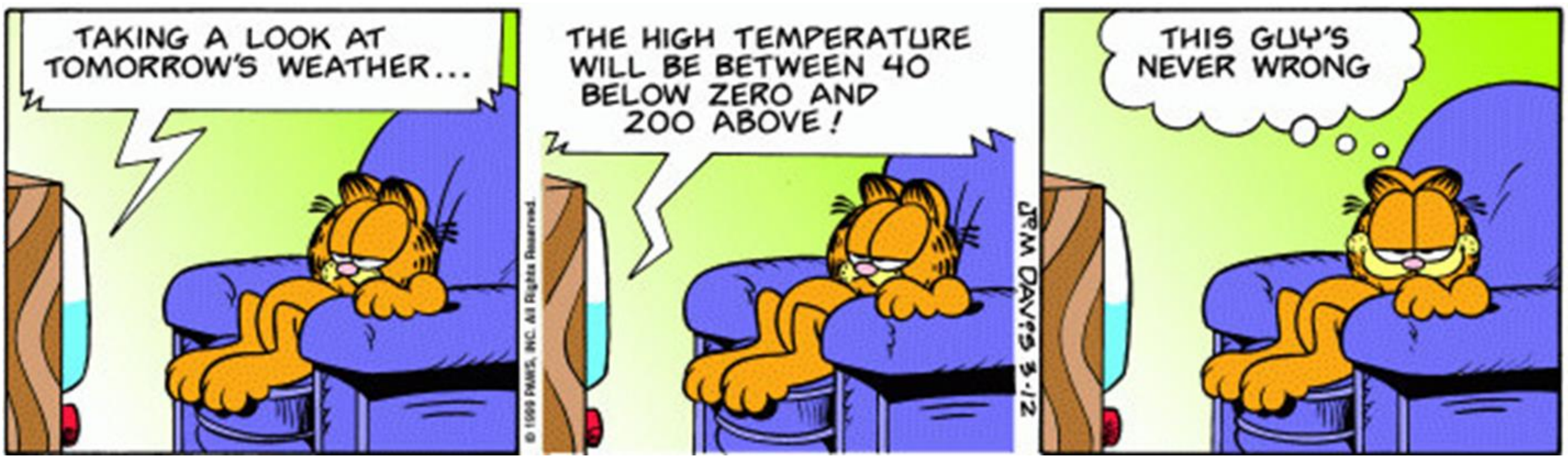
$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$

$$= (0.782, 0.818)$$

Conclude: The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

Confidence Interval



Confidence Intervals – What do they mean?

" I got the instructions from my Statistics Professor. He was 80% confident that the true location of the restaurant was in this neighborhood."