Class 12

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Recap Chapter 7.2 – 7.3

Discussion of Chapters

Lecture Chapter 8.1 – 8.2

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Sample distribution of sample means (SDSM):
If random samples of size n, are taken from ANY population with mean μ and standard deviation σ, then the SDSM:

1. A mean $\mu_{\bar{x}}$ equal to μ
2. A standard deviation $\sigma_{\bar{x}}$ equal to $\dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem (CLT): The sampling distribution of sample means will more closely resemble the normal distribution as the sample size n increases.

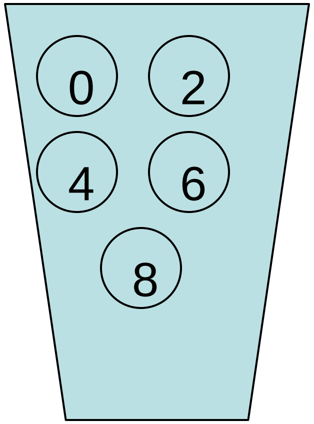
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

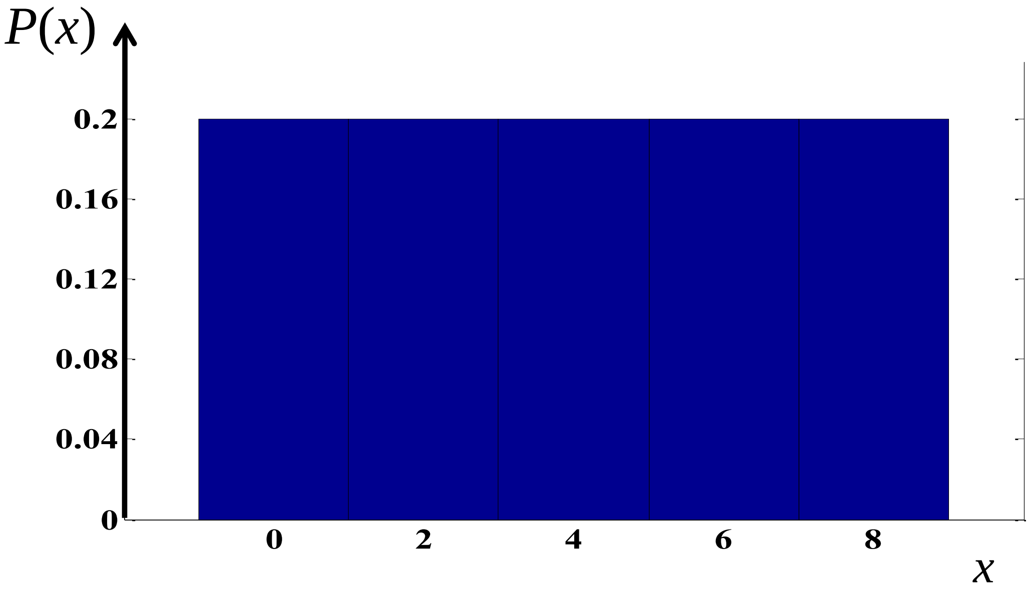
Example:

N=5 balls in bucket, select n=1 *with* replacement.

0, 2, 4, 6, 8.



x	$P(x)$
0	$1/5$
2	$1/5$
4	$1/5$
6	$1/5$
8	$1/5$



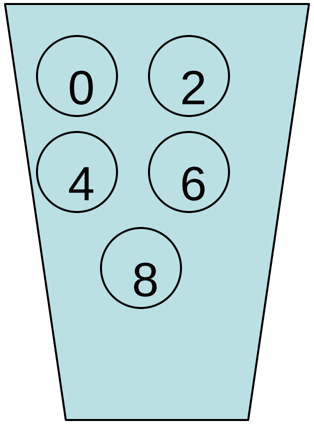
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

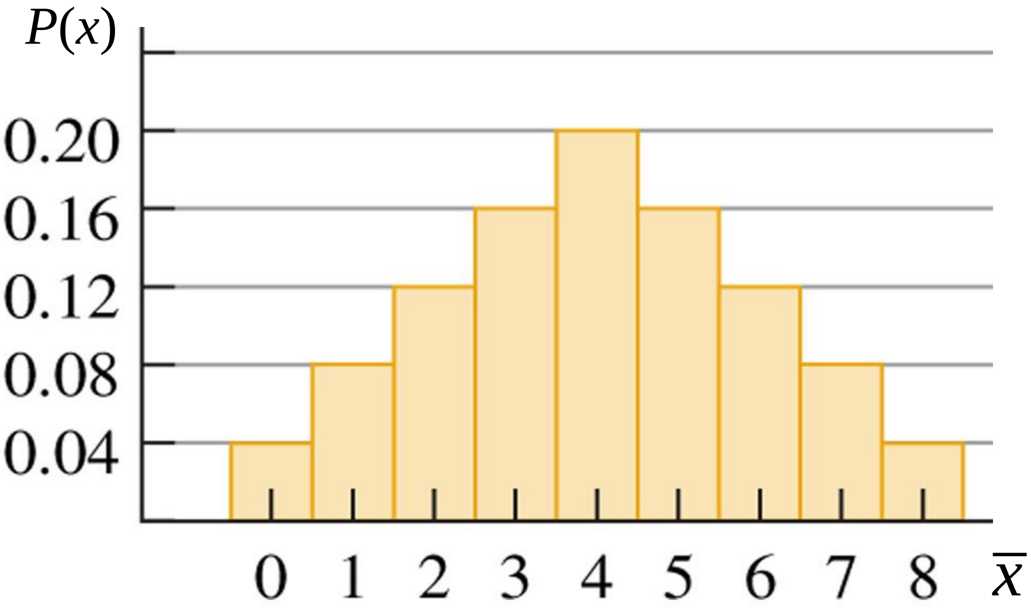
Example:

N=5 balls in bucket, select n=2 *with* replacement.

0, 2, 4, 6, 8.



$\bar{x}$	$P(\bar{x})$
0	1/25
1	2/25
2	3/25
3	4/25
4	5/25
5	4/25
6	3/25
7	2/25
8	1/25



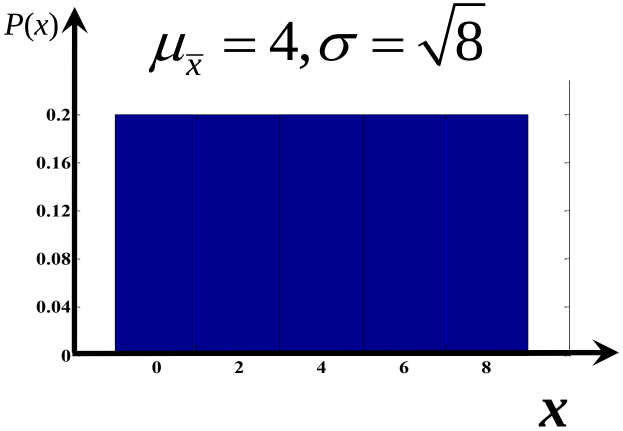
7: Sample Variability

1. A mean $\mu_{\bar{x}}$ equal to μ
2. A standard deviation $\sigma_{\bar{x}}$ equal to $\frac{\sigma}{\sqrt{n}}$

7.2 The Sampling Distribution of Sample Means

$\bar{x}$ from n=1 distribution

$$\mu_{\bar{x}} = 4, \sigma = \sqrt{8}$$



$\bar{x}$ from n=2 distribution

$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{2}$$

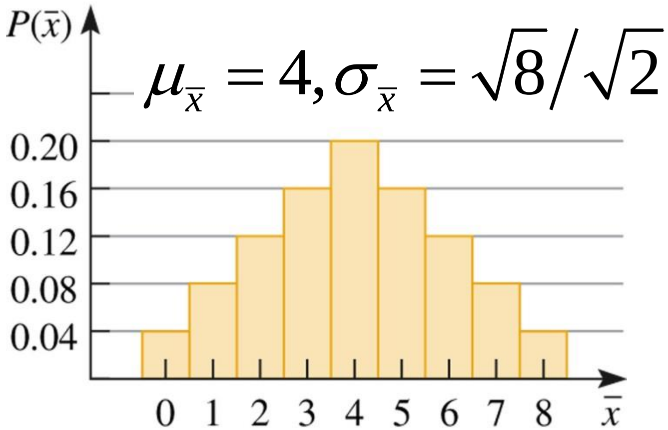
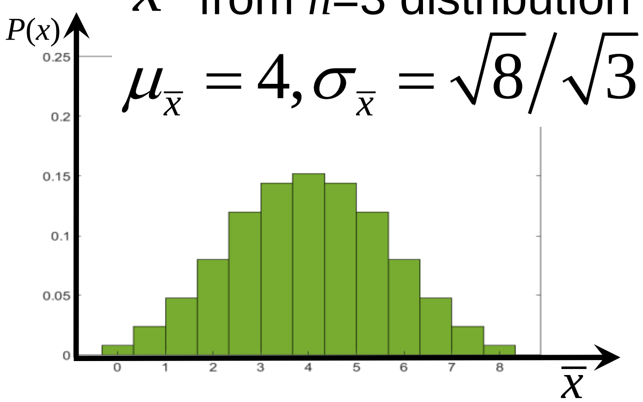


Figure from Johnson & Kuby, 2012.

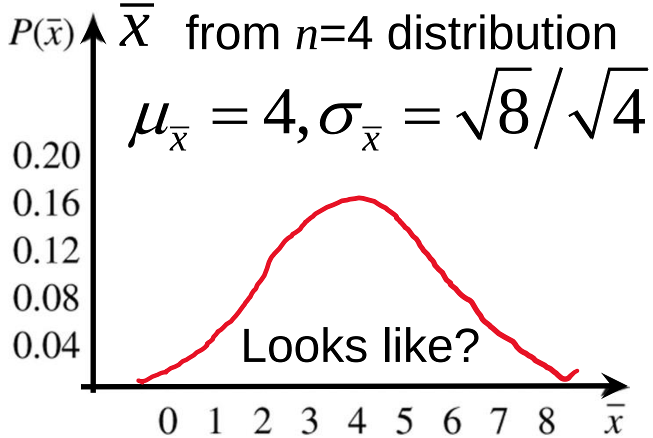
$\bar{x}$ from n=3 distribution

$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{3}$$



$\bar{x}$ from n=4 distribution

$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{4}$$



n large? →

$$\mu_{\bar{x}} = 4$$

$$\sigma_{\bar{x}} = \sqrt{8}/\sqrt{n}$$

Looks like?

7: Sample Variability

Questions?

Homework: Read Chapter 7.1-7.3

WebAssign

Chapter 7 # 6, 21, 23, 29, 33, 35

Discussion: Chapters

Discussion on Course

We're moving into a new phase of the course…

Part III on Inferential Statistics.

Parts I and II were all foundational material for

Part III.

Discussion on Course

Part I: Descriptive Statistics

Chapter 1: Statistics

Background material. Definitions.

Chapter 2: Descriptive Analysis and Presentation of single variable data

Graphs, Central Tendency, Dispersion, Position

Chapter 3: Descriptive Analysis and Presentation of bivariate data

Scatter plot, Correlation, Regression

Discussion on Course

Part II: Probability

Chapter 4: Probability

Conditional, Rules, Mutually Exclusive, Independent

Chapter 5: Probability Distributions (Discrete)

Random variables, Probability Distributions, Mean & Variance, Binomial Distribution with Mean & Variance

Chapter 6: Probability Distributions (Continuous)

Normal Distribution, Standard Normal, Applications, Notation

Chapter 7: Sample Variability

Sampling Distributions, SDSM, CLT

Discussion on Course

Part III: Inferential Statistics

Chapter 8: Introduction to Statistical Inferences

Confidence Intervals, Hypothesis testing

Chapter 9: Inferences Involving One Population

Mean μ (σ unknown), proportion p, variance σ^2

Chapter 10: Inferences Involving Two Populations

Difference in means μ_1-μ_2, proportions p_1-p_2, variances σ_1^2 / σ_2^2

Part IV: More Inferential Statistics

Chapter 11: Applications of Chi-Square

Chi-square statistics. …. We will discuss later.

Discussion on Course

Part IV: More Inferential Statistics

Chapter 11: Applications of Chi-Square

Hypothesis testing for Contingency Tables.

Chapter 12: Analysis of Variance

Hypothesis testing for differences in more than two means μ_1, μ_2, μ_3. ….

Chapter 13: Linear Correlation and Regression Analysis

Hypothesis testing on correlation coefficient ρ and slope β_1.

Chapter 14: Elements of Nonparametric Analysis

Distribution free hypothesis tests.

Lecture Chapter 8.1- 8.2

Chapter 8: Introduction to Statistical Inference

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

8: Introduction to Statistical Inference
8.1 The Nature of Estimation

The purpose of Statistical Inference is to use the info in a sample of data to increase knowledge of a population.

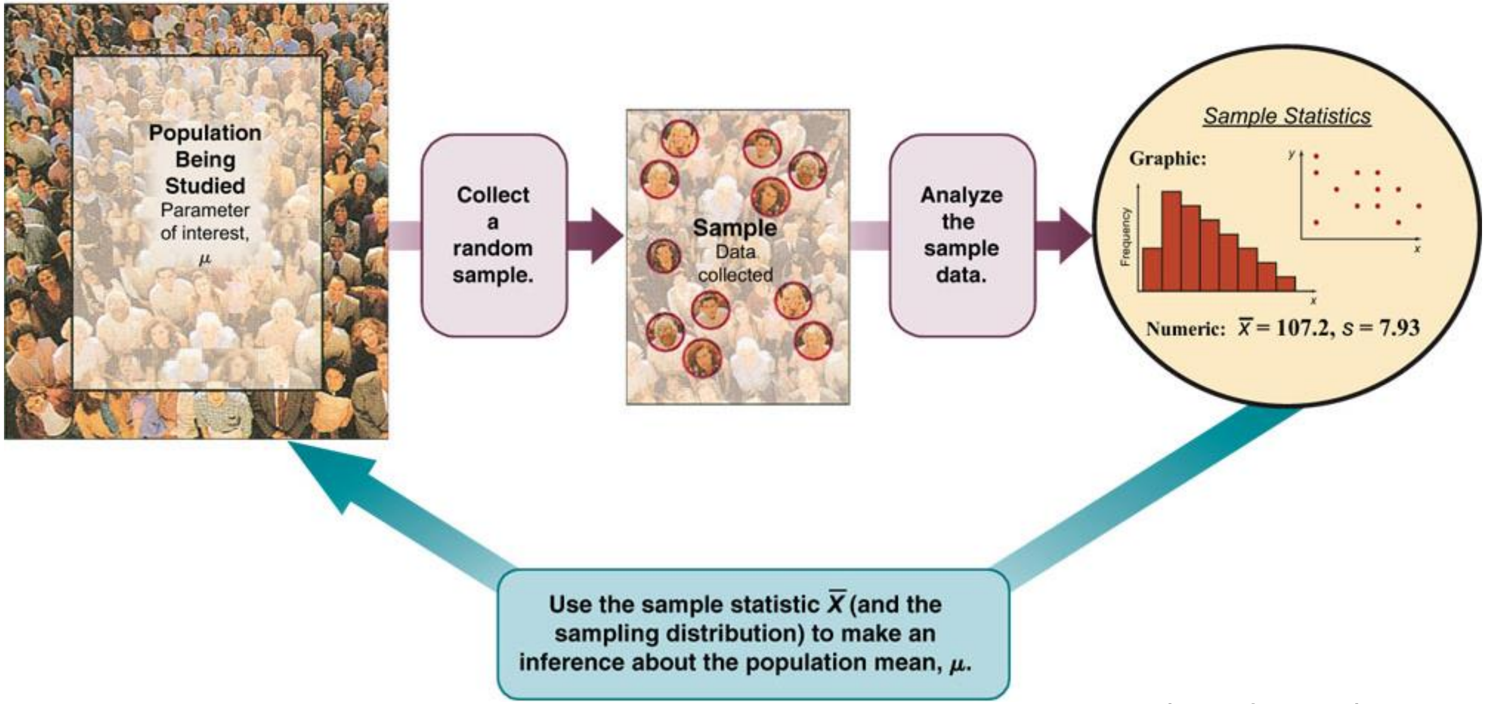


Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference
8.1 The Nature of Estimation

We discussed how if we compute a quantity from a population of data then it is called a parameter and if we estimate it from a sample of data then it is called a statistic.

Recall: Chapter 1 definitions.
Parameter: A numerical value summarizing all the data of an entire population.

Statistic: A numerical value summarizing the sample data.

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

More precisely, a single number to estimate a parameter is called a point estimate.

Point estimate for a parameter: A single number designed to estimate a quantitative parameter of a population, usually the value of the corresponding **sample statistic**.

i.e. $\bar{x}$ is a point estimate for μ

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

Interval estimate: An interval bounded by two values and used to estimate the value of a population parameter. The values that bound this interval are statistics calculated from the sample that is being used as the basis for estimation.

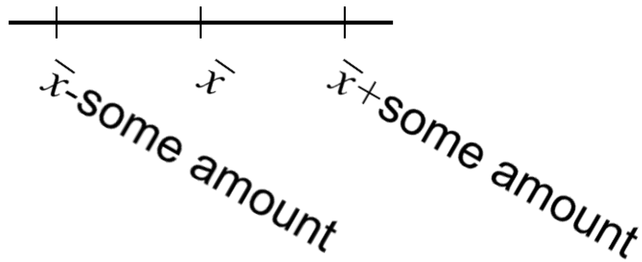
i.e. $\bar{x} \pm (\text{some amount})$ is an interval estimate for μ.

The interval estimate will be of the form

$\text{point estimate} \pm \text{some amount}$

8: Introduction to Statistical Inference

8.1 The Nature of Estimation



← Interval Estimate for μ.

Significance Level: Pre assigned probability of a parameter being outside our interval estimate, α.

$$P(\mu \text{ not in } \bar{x} \pm \text{some amount}) = \alpha$$

i.e. .05

$$1 - P(\bar{x} - \text{some amount} < \mu < \bar{x} + \text{some amount}) = \alpha$$

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

Take many samples and for each calculate interval estimate … then

Level of Confidence 1-α: The proportion of all interval estimates that include the parameter being estimated. i.e μ

$$P(\mu \text{ in } \bar{x} \pm \text{some amount}) = 1 - \alpha$$

$$P(\bar{x} - \text{some amount} < \mu < \bar{x} + \text{some amount}) = 1 - \alpha$$

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

Confidence Interval: An interval estimate with a specified level of confidence.

A range of values for the parameter with a level of confidence attached. (i.e. 95% confident)

point estimator $\pm$ some amount that depends on confidence level

$\bar{x} \pm$ some amount$(1-\alpha)$ $\quad$ $1-\alpha$

The general form for a confidence interval is

point estimate $\pm$ margin of error

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

The assumption for estimating mean μ using a known σ:
The sampling distribution of $\bar{x}$ has a normal distribution.

Recall from Chapter 6 that for the standard normal distribution,

$$P(-1.96 < z < 1.96) = 0.95$$

From the CLT in Chapter 7, we know that when n is "large," the sample mean $\bar{x}$ is approximately normally distributed with

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

What this implies is that $z=\dfrac{\overline{x}-\mu_{\overline{x}}}{\sigma_{\overline{x}}}$

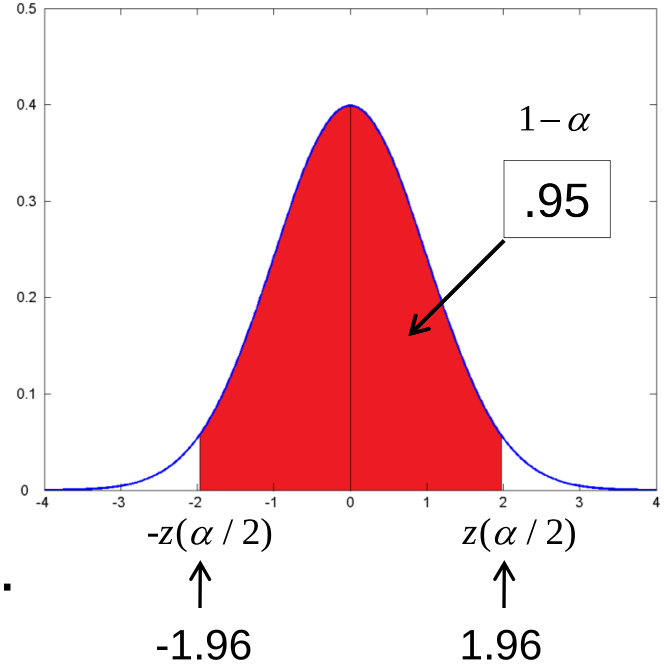
has an approximate standard normal distribution!

$$P(-1.96<z<1.96)=0.95$$

Or more generally,

$$P(-z(\alpha/2)<z<z(\alpha/2))=1-\alpha \quad .$$

$z(\alpha/2)$ called the confidence coefficient.



$z(a)$ is the value of z with an area a larger than it

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

$$P(-z(\alpha/2) < z < z(\alpha/2)) = 1-\alpha$$

With some algebra, we can see that …. (fill in)

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

$$P(-z(\alpha/2) < z < z(\alpha/2)) = 1-\alpha$$

With some algebra, we can see that

$$-z(\alpha/2) \quad < \quad z$$

$$-z(\alpha/2) \quad < \quad \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$-z(\alpha/2)\frac{\sigma}{\sqrt{n}} \quad < \quad \bar{x}-\mu$$

$$-z(\alpha/2)\frac{\sigma}{\sqrt{n}}-\bar{x} \quad < \quad -\mu$$

$$\bar{x}+z(\alpha/2)\frac{\sigma}{\sqrt{n}} \quad > \quad \mu$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

$$P(-z(\alpha/2) < z < z(\alpha/2)) = 1 - \alpha$$

With some algebra, we can see that

and

$$z(\alpha/2) > z$$

$$z(\alpha/2) > \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z(\alpha/2)\frac{\sigma}{\sqrt{n}} > \bar{x} - \mu$$

$$z(\alpha/2)\frac{\sigma}{\sqrt{n}} - \bar{x} > -\mu$$

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} < \mu$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Thus, a (1-α)×100% confidence interval for μ is

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

which if α=0.05, a 95% confidence interval for μ is

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \quad .$$

Confidence Interval for Mean:

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \quad \textbf{to} \quad \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}} \qquad \textbf{(8.1)}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

THE CONFIDENCE INTERVAL: A FIVE STEP PROCESS

Step 1 The Set-UP:

Step 2 Confidence Interval Criteria:

Step 3 The Sample Evidence:

Step 4 The Confidence Interval:

Step 5 The Results:

Your Book describes as a 5 step process.

Read this. Important.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Philosophically, μ is fixed and the interval varies.

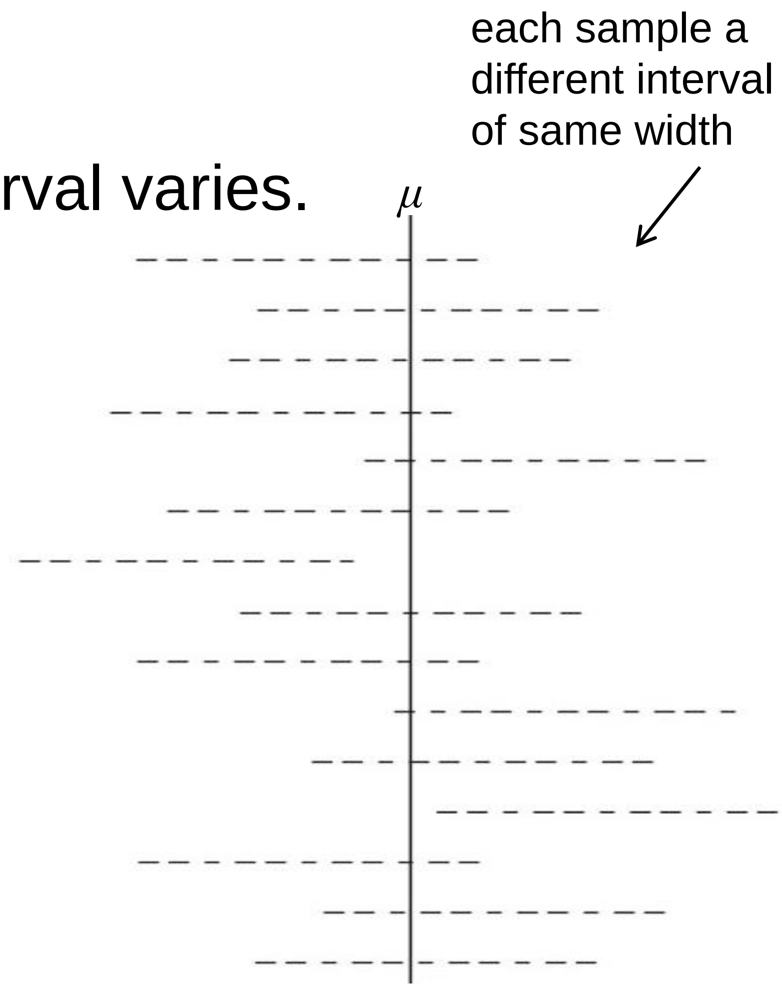


Figure from Johnson & Kuby, 2012.

If we take a sample of data, $x_1,\ldots,x_n$ and determine a confidence interval from it, we get.

$$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

If we had a different sample of data, $y_1,\ldots,y_n$ we would have determined a different confidence interval.

$$\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

We never truly know if our CI from our sample of data will

contain the true population mean μ.

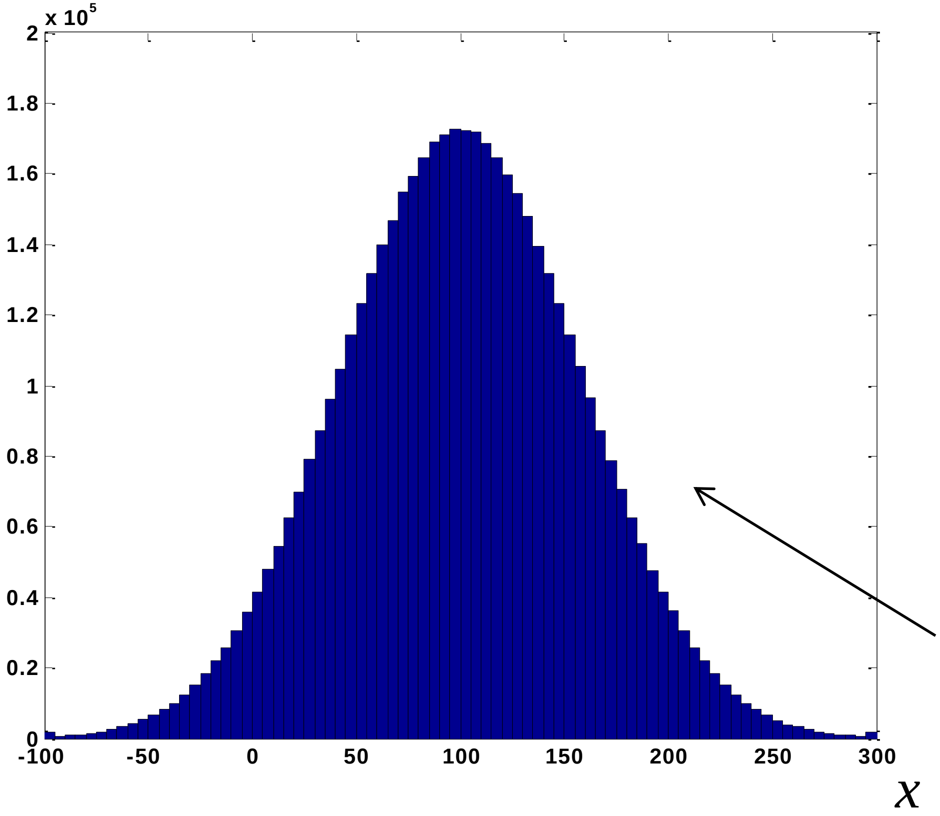
But we do know that there is a $(1-\alpha)\times 100\%$ chance

that a confidence interval from a sample of data will contain μ.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Example: Make 5×10^6 normal μ=100,σ=57.7 random values.



True	Random
$\mu = 100$	$\bar{x} = 100.0320$
$\sigma = 57.7350$	$s = 57.7623$

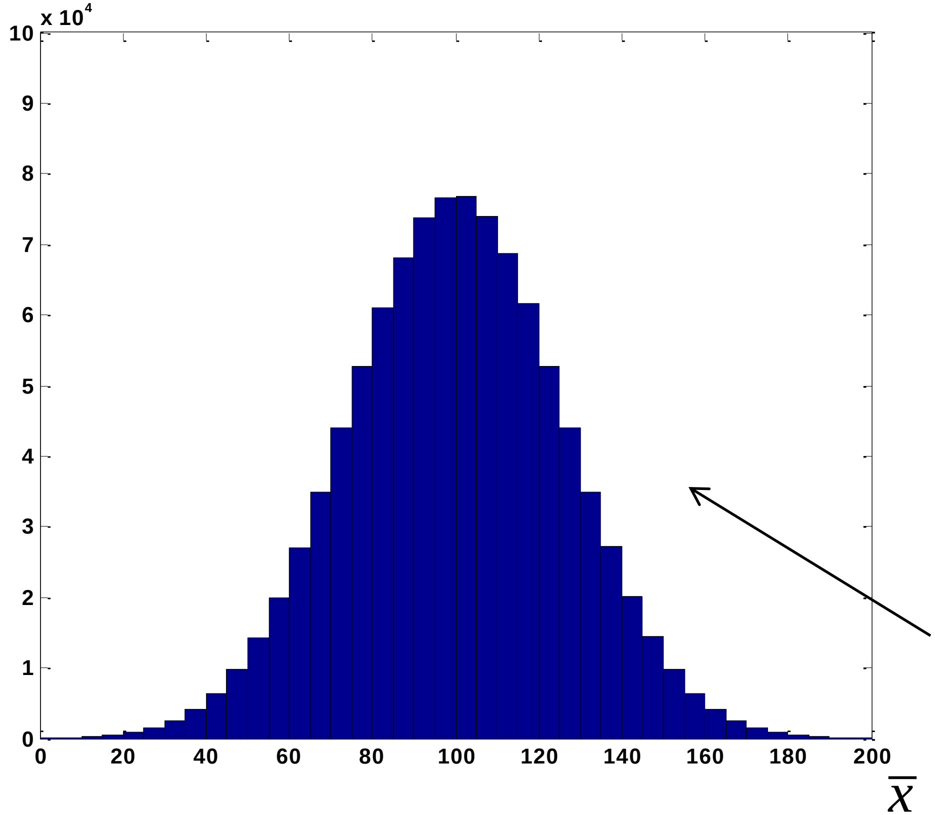
Sample mean and variance from the 5 million values.

Here is a histogram of the 5 million values.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Form 10^6 $\bar{x}$'s (means) from successive n=5 random numbers.



True	Random
$\mu_{\bar{x}} = 100$	$\bar{x}_{\bar{x}} = 100.0320$
$\sigma_{\bar{x}} = 25.8199$	$s_{\bar{x}} = 25.8397$

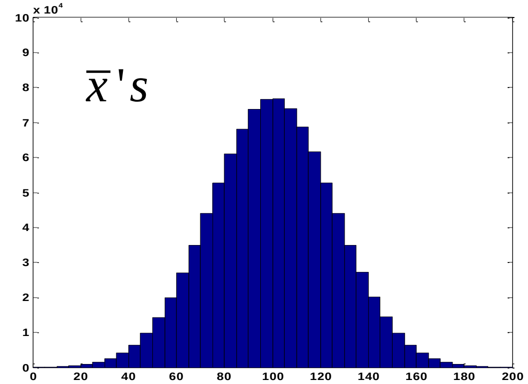
Sample mean and variance from the 1 million means.

$$\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \sigma / \sqrt{n}$$

Here is a histogram of the 1 million means.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)



Form 1 million U and L values from $\bar{x}$'s.

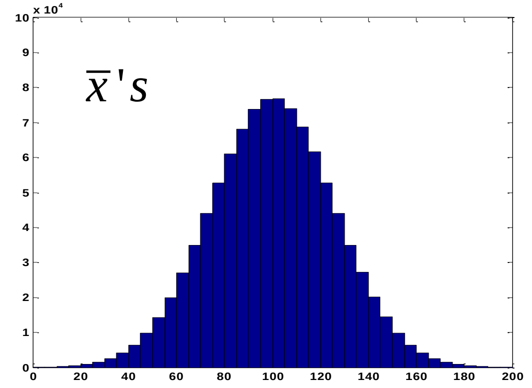
n=5 and $\sigma = 57.7$

$$U = \bar{x} + 1.96\sigma / \sqrt{n}$$

insert each $\bar{x}$

$$L = \bar{x} - 1.96\sigma / \sqrt{n}$$

8: Introduction to Statistical Inference
8.2 Estimation of Mean μ (σ Known)



Form 1 million U and L values from $\bar{x}$'s.

n=5 and $\sigma = 57.7$

$$U = \bar{x} + 1.96\sigma / \sqrt{n}$$

insert each $\bar{x}$

$$L = \bar{x} - 1.96\sigma / \sqrt{n}$$

$$U_\mu = \mu + 1.96\sigma / \sqrt{n}$$

insert true μ

$$L_\mu = \mu - 1.96\sigma / \sqrt{n}$$

Random, $\bar{x}$'s
$\bar{L}_{\bar{x}} = 49.4250$
$\bar{U}_{\bar{x}} = 150.6389$

We will also get 1 million L's and U's that we can use to make histograms.

8: Introduction to Statistical Inference

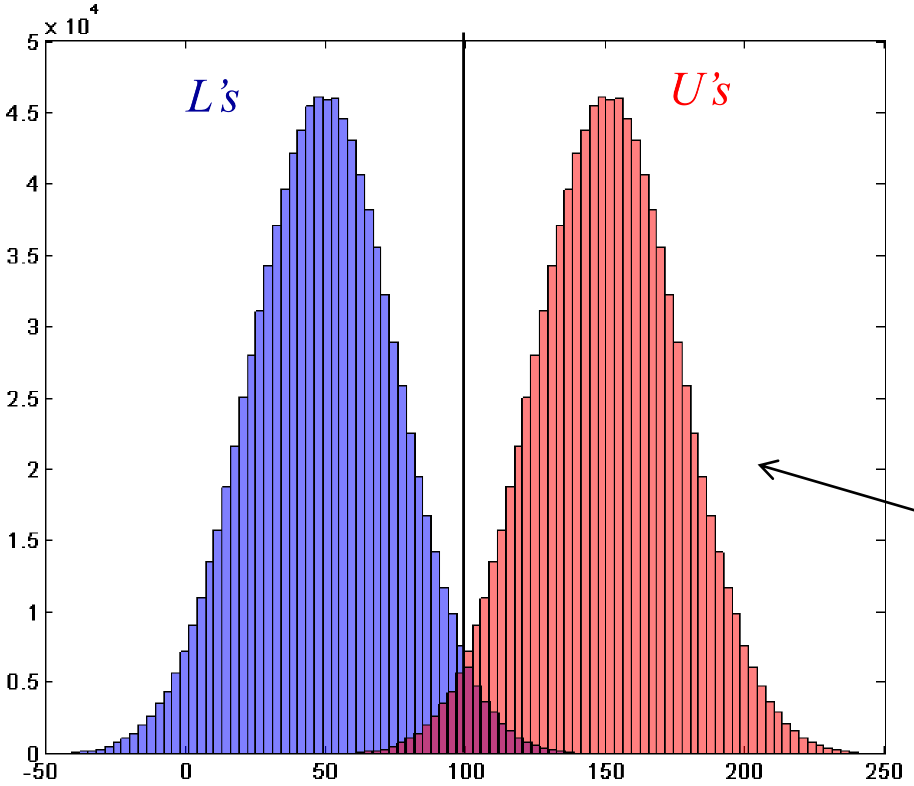
8.2 Estimation of Mean μ (σ Known)

$\mu = 100$
$\sigma = 57.7$

Form 1 million U and L values from $\bar{x}$'s.

$$U = \bar{x} + 1.96\sigma / \sqrt{n}$$

$$L = \bar{x} - 1.96\sigma / \sqrt{n}$$



True	Random
$L_\mu = 49.3930$	$L_{\bar{x}} = 49.4250$
$U_\mu = 150.6070$	$U_{\bar{x}} = 150.6389$

Histogram of the 1 million L's and U's.

8: Introduction to Statistical Inference

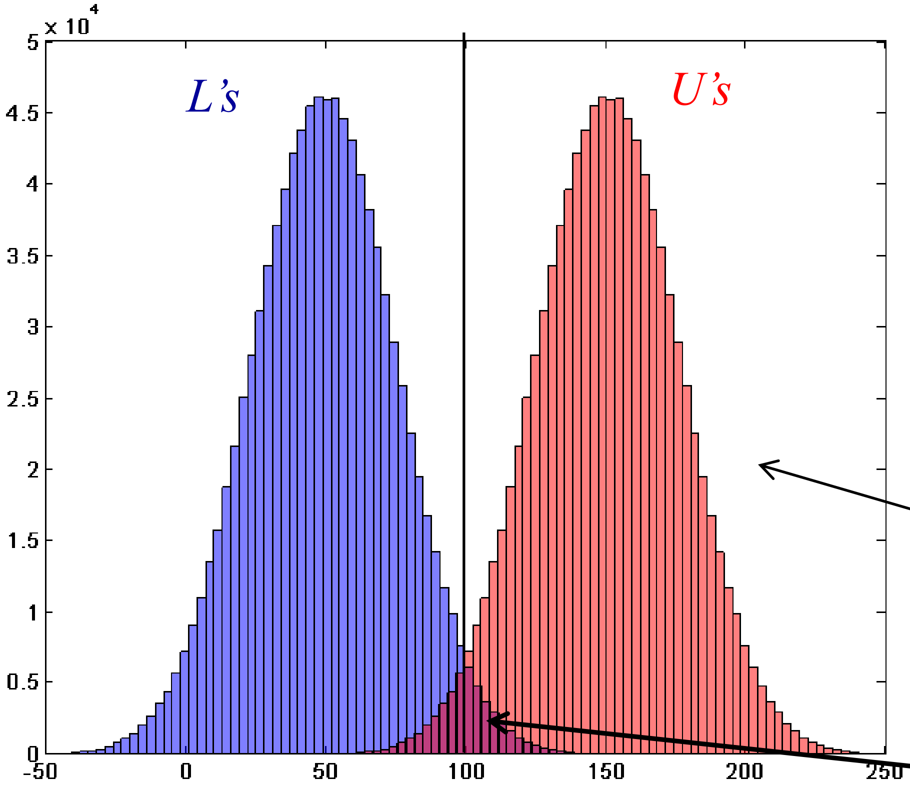
8.2 Estimation of Mean μ (σ Known)

$$\mu = 66.5$$
$$\sigma = 57.7$$

Form 1 million U and L values from $\bar{x}$'s.

$$U = \bar{x} + 1.96\sigma / \sqrt{n}$$

$$L = \bar{x} - 1.96\sigma / \sqrt{n}$$



True	Random
$L_\mu = 49.3930$	$L_{\bar{x}} = 49.4250$
$U_\mu = 150.6070$	$U_{\bar{x}} = 150.6389$

Histogram of the 1 million L's and U's.

2.5% of time upper less than 100
2.5% of time lower greater than 100
5% of time μ=100 not in interval

$$P(\mu \text{ not in } \bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}) = \alpha$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

69,69,67,66,71,74,75,70,67,73,64,65,68,67,65

Example:

A random sample of size n=15 student heights was taken from this class. Assume that we know that σ=4. Construct a 95% confidence Interval for μ.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

69,69,67,66,71,74,75,70,67,73,64,65,68,67,65

Example:

A random sample of size n=15 student heights was taken from this class. Assume that we know that σ=4. Construct a 95% confidence Interval for μ.

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

69,69,67,66,71,74,75,70,67,73,64,65,68,67,65

Example:

A random sample of size n=15 student heights was taken from this class. Assume that we know that σ=4. Construct a 95% confidence Interval for μ.

$$\bar{x} = 68.7 \qquad \sigma = 4$$
$$\alpha = .05$$

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

69,69,67,66,71,74,75,70,67,73,64,65,68,67,65

Example:

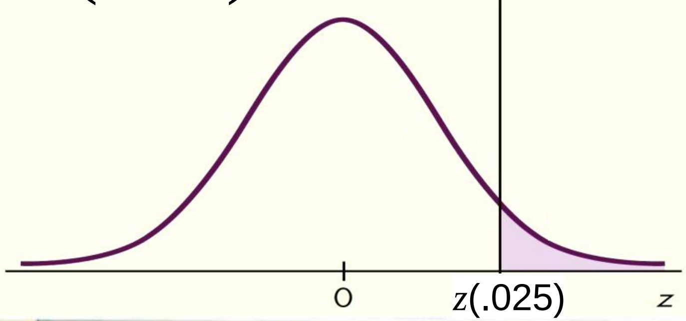
A random sample of size n=15 student heights was taken from this class. Assume that we know that σ=4. Construct a 95% confidence Interval for μ.

$$\bar{x} = 68.7 \qquad \sigma = 4$$

$$\alpha = .05$$

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

$$z(.025) = 1.96$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

69,69,67,66,71,74,75,70,67,73,64,65,68,67,65

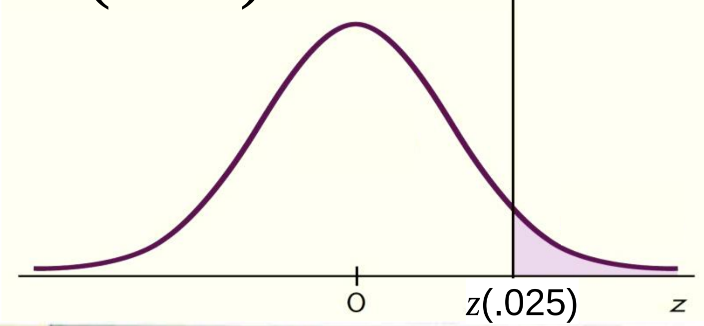
Example:

A random sample of size n=15 student heights was taken from this class. Assume that we know that σ=4. Construct a 95% confidence Interval for μ.

$\bar{x} = 68.7 \qquad \sigma = 4$

$\alpha = .05$

$z(.025) = 1.96$



$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

$$68.7 - 1.96\frac{4}{\sqrt{15}} \text{ to } 68.7 + 1.96\frac{4}{\sqrt{15}}$$

$$66.7 \text{ to } 70.7$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Sample Size Determination

Recall that our Confidence Interval was $\bar{x} \pm$ some amount

which was say $\bar{x} \pm z(\alpha/2)\dfrac{\sigma}{\sqrt{n}}$.

Maximum Error of Estimate

$$E = z(\alpha/2)\frac{\sigma}{\sqrt{n}} \qquad \textbf{(8.2)}$$

then we can rewrite as

Sample Size

$$n = \left(\frac{z(\alpha/2)\sigma}{E}\right)^2 \qquad \textbf{(8.3)}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

$$n = \left(\frac{z(\alpha/2)\sigma}{E} \right)^2$$

In this, $z(\alpha/2)$ is known with specification of α.

We can set an E and set σ (or get it from previous data) to obtain a minimum sample size n to achieve E.

Used a lot in Biological applications to determine how many subjects and most IRBs require an estimate of n.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Example:

Determine the sample size n needed to estimate the mean height in this class to within 1 inch with 95% confidence. Assume that we know that σ=4.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Example:

Determine the sample size n needed to estimate the mean height in this class to within 1 inch with 95% confidence. Assume that we know that $\sigma=4$.

$$n = \left(\frac{z(\alpha/2)\sigma}{E} \right)^2$$

$$E = 1$$

$$\alpha = .05$$

$$z(.025) = 1.96$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Example:

Determine the sample size n needed to estimate the mean height in this class to within 1 inch with 95% confidence. Assume that we know that σ=4.

$$E = 1$$

$$\alpha = .05$$

$$z(.025) = 1.96$$

$$n = \left(\frac{z(\alpha / 2)\sigma}{E} \right)^2$$

$$n = \left(\frac{1.96 * 4}{1} \right)^2$$

$$n = 61.46$$

$$n = 62$$

Chapter 8: Introduction to Statistical Inference

Questions?

Homework: Read Chapter 8.1-8.2
WebAssign
Chapter 8 # 5, 15, 22, 24, 35, 47