Class 16

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:
Recap Chapter 7.2
Lecture Chapter 7.3
Discussion: Chapters
Recap Chapter 7.2
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

Sample distribution of sample means (SDSM): If all possible random samples, each of size $n$, are taken from any population with mean $\mu$ and standard deviation $\sigma$, then the sampling distribution of sample means will have the following:

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

Furthermore, if the sampled population has a normal distribution, then the sampling distribution of $\bar{x}$ will also be normal for all samples of all sizes.

Discuss Later: What if the sampled population does not have a normal distribution?
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

The Central Limit Theorem: Assume that we have a population (arbitrary distribution) with mean \( \mu \) and standard deviation \( \sigma \).

If we take random samples of size \( n \) (with replacement), then for “large” \( n \), the distribution of the sample means, the \( \bar{x} \)‘s, is approximately normally distributed with

\[
\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

where in general \( n \geq 30 \) is sufficiently “large,” but can be as small as 15 or as big as 50 depending upon the shape of distribution!
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

According to SDSM, if we had a sample \( x_1, \ldots, x_n \) of size \( n = 1, 2, 3, 4, 5, 15, 30, \) and 50 from Uniform(0,200) or Normal(100,(57.7)^2)

<table>
<thead>
<tr>
<th>Sample Size, ( n )</th>
<th>Mean, ( \mu_{\bar{x}} )</th>
<th>SD, ( \sigma_{\bar{x}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>57.7350</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>40.8248</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>33.3333</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>28.8675</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>25.8199</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>14.9071</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>10.5409</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>8.1650</td>
</tr>
</tbody>
</table>

For Uniform \( a \) to \( b \)

\[
\mu = \frac{b - a}{2}
\]

\[
\sigma^2 = \frac{(b - a)^2}{12}
\]

Theoretical values from the SDSM

\[
\mu_{\bar{x}} = \mu
\]

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

I wrote a computer program to generate 1 million random observations $x_1, \ldots, x_{1000000}$ from the Uniform$(a=0, b=200)$ and also from Normal$(\mu=100, \sigma^2=(57.7)^2)$ distributions.

I showed you the histograms being formed!
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

Expanded program to generate $n$ million random observations

$$x_1, \ldots, x_{n \times 10^6}$$ from the Uniform($a=0, b=200$) and

also from the Normal($\mu=100, \sigma^2=(57.7)^2$) distributions,

for each of $n=1, 2, 3, 4, 5, 15, 30, \text{ and } 50$.

8 data sets of Uniform and Normal random observations
# 7: Sample Variability

## 7.2 The Sampling Distribution of Sample Means

Sample means and standard deviations from each of the $n$ million observations from the Uniform($a=0, b=200$) and Normal($\mu=100, \sigma^2=(57.7)^2$) distributions.

### i.e. $n=5$, $5 \times 10^6$

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>...</th>
<th>$x_{5000000}$</th>
</tr>
</thead>
</table>

Groups of $n=5$

| $x_1, x_2, x_3, x_4, x_5$ | $x_6, x_7, x_8, x_9, x_{10}$ | ... | $x_{4999996}, \ldots, x_{5000000}$ |

| $\bar{x}_1$ | $\bar{x}_2$ | ... | $\bar{x}_{10^6}$ |

Mean of groups

Histogram of $\bar{x}$'s

$8$ data sets of Uniform and Normal
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

Computed sample means and standard deviations from the one million means \( \bar{x}_1, \ldots, \bar{x}_{10^6} \):

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Mean ( \mu_{\bar{x}} )</th>
<th>Mean U ( \bar{x}_U )</th>
<th>Mean N ( \bar{x}_N )</th>
<th>SD ( \sigma_{\bar{x}} )</th>
<th>SD U ( s_{\bar{x}} )</th>
<th>SD N ( s_{\bar{x}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100.0642</td>
<td>100.0077</td>
<td>57.7350</td>
<td>57.7071</td>
<td>57.7888</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>99.9828</td>
<td>100.0037</td>
<td>40.8248</td>
<td>40.8418</td>
<td>40.8206</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>99.9909</td>
<td>99.9627</td>
<td>33.3333</td>
<td>33.3418</td>
<td>33.2984</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>99.9559</td>
<td>100.0642</td>
<td>28.8675</td>
<td>28.8946</td>
<td>28.8126</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>100.0074</td>
<td>100.0320</td>
<td>25.8199</td>
<td>25.7865</td>
<td>25.8397</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>100.0134</td>
<td>99.9517</td>
<td>14.9071</td>
<td>14.9035</td>
<td>14.8918</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>99.9934</td>
<td>99.9836</td>
<td>10.5409</td>
<td>10.5335</td>
<td>10.5352</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>99.9918</td>
<td>99.9890</td>
<td>8.1650</td>
<td>8.1605</td>
<td>8.1709</td>
</tr>
</tbody>
</table>
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 2 \quad 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ \overline{X} \]

\[ S_{\overline{X}} \]
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

\( n=3 \) \( \times 10^6 \) means

\( \mu = 100 \)
\( \sigma = 57.73 \)

Histogram of means from uniform

Histogram of means from normal

\( S_{\overline{X}} \)
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 4 \quad 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ S_{\bar{X}} \]
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\( n = 5 \quad 1 \times 10^6 \) means

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ S_{\bar{X}} \]

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 15 \quad 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \quad \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ \bar{X} \]

\[ S_{\bar{X}} \]
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 30 \quad 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ S_{\overline{X}} \]
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n = 50 \quad 1 \times 10^6 \text{ means} \]

\[ \mu = 100 \]
\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal

\[ s_{\overline{x}} \]

Rowe, D.B.
With a population mean $\mu$ and standard deviation $\sigma$.

Random samples of size $n$ with replacement, for “large” $n$, the distribution of the sample means quickly becomes normally distributed with

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

Generally $n \geq 30$ is sufficiently “large,” but can be as small as 15 or as big as 50 depending upon the shape of distribution!
7: Sample Variability

Questions?

Homework: Chapter 7 # 6, 21, 23, 29, 33, 35
Lecture Chapter 7.3
Chapter 7: Sample Variability (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Recall: When we take a sample of data \( x_1, \ldots, x_n \) from a population, then compute an estimate of a parameter it is called a sample statistic. i.e. \( \bar{x} \) for \( \mu \)

Sampling Distribution of a sample statistic: The distribution of values for a sample statistic obtained from repeated samples, all of the same size and all drawn from the same population.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

The Sampling Distribution of Sample Means

As the number of samples increases the empirical dist. turns into theoretical dist.

Figure from Johnson & Kuby, 2008.
3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: Scatter Diagram

Our data.

Height vs. Weight

$n=80$ responses

Find yourself. If not here then removed or you did not respond.
Assume that this is a population of data.

Does this population look normally distributed?

Put normal with same $\mu$ and $\sigma$. 

$N = 80$ values
If I’m interested in the mean $\mu_{\bar{x}}$ of a sample of size $n=15$, then by **SDSM** $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, in addition, by the **CLT** $\bar{x}$ is (hopefully) normally distributed.

I wrote a computer program to take a sample of $n=1,3,5,15$ from the population of $N$ heights with replacement $10^6$ times.
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_{\bar{x}} = 66.7$ and $\sigma_{\bar{x}} = \frac{4.5}{\sqrt{1}} = 4.5$.

$N=80$ values

$\mu = 66.7$

$\sigma = 4.5$
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, \( \mu_\bar{x} = 66.7 \) and \( \sigma_\bar{x} = \frac{4.5}{\sqrt{3}} = 3.2 \).

Histogram of the 1 million means

\( n=3 \)

Put normal with same \( \mu_\bar{x} \) and \( \sigma_\bar{x} \).

\( N=80 \) values

\( \mu = 66.7 \)

\( \sigma = 4.5 \)
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_x = 66.7$ and $\sigma_x = \frac{4.5}{\sqrt{5}} = 2.6$.

Histogram of the 1 million means

$N=80$ values

$\mu = 66.7$

$\sigma = 4.5$
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_x = 66.7$ and $\sigma_x = \frac{4.5}{\sqrt{15}} = 1.2$.

By CLT becomes normal

$N=80$ values

$\mu = 66.7$

$\sigma = 4.5$
Now that we believe that the mean $\bar{x}$ from a sample of $n=15$ is normally distributed with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, we can find probabilities.
7: Sample Variability
7.3 Application of the Sampling Distribution of Sample Means

To find these probabilities, we first convert to $z$ scores

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}, \quad c = \frac{a - \mu_{\bar{x}}}{\sigma_{\bar{x}}}, \quad d = \frac{b - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

and use the table in book.
7: Sample Variability
7.3 Application of the Sampling Distribution of Sample Means

Example:
What is probability that sample mean $\overline{x}$ from a random sample of $n=15$ heights is greater than 69" when $\mu = 66.7$ and $\sigma = 4.5$?
7: Sample Variability

Questions?

Homework: 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.

Before we discussed …
Discussion on Course

Part I: Descriptive Statistics

Chapter 1: Statistics
   Background material. Definitions.

Chapter 2: Descriptive Analysis and Presentation
   single variable data
      Graphs, Central Tendency, Dispersion, Position

Chapter 3: Descriptive Analysis and Presentation
   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
  Hypothesis testing

Chapter 9: Inferences Involving One Population
  Mean $\mu$ ($\sigma$ unknown), proportion $p$, variance $\sigma^2$

Chapter 10: Inferences Involving Two Populations
  Difference in means $\mu_1 - \mu_2$, proportions $p_1 - p_2$, variances $\sigma_1^2 / \sigma_2^2$

Part IV: More Inferential Statistics
Chapter 11: Applications of Chi-Square
  Chi-square statistics. …. We will discuss later.
Discussion on Course

Next Lecture will be on Chapter 8.

Chapter 8: Introduction to Statistical Inferences
Hypothesis testing