Class 15

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 7.1 - 7.3

Lecture Chapter 7.3 continued

TA go over Exam 3.
Recap Chapter 7.1 - 7.3
7: Sample Variability
7.2 Sampling Distributions

When we take a random sample $x_1, \ldots, x_n$ from a population, one of the things that we do is compute the sample mean $\bar{x}$. The value of $\bar{x}$ is not $\mu$. Each time we take a random sample of size $n$ (with replacement), we get a different set of values $x_1, \ldots, x_n$ and a different value for $\bar{x}$. 
7: Sample Variability
7.2 Sampling Distributions

$N=5$ balls in bucket, select $n=1$ with replacement.
Population data values: 0, 2, 4, 6, 8.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/5</td>
</tr>
<tr>
<td>2</td>
<td>1/5</td>
</tr>
<tr>
<td>4</td>
<td>1/5</td>
</tr>
<tr>
<td>6</td>
<td>1/5</td>
</tr>
<tr>
<td>8</td>
<td>1/5</td>
</tr>
</tbody>
</table>

$\mu = \sum_{i=1}^{n} [x_i P(x_i)] = 4$

$\sigma^2 = \sum_{i=1}^{n} [(x_i - \mu)^2 P(x_i)] = 8$

$\sigma = \sqrt{8} = 2\sqrt{2}$

Rowe, D.B.
7: Sample Variability
7.2 Sampling Distributions

\( N=5 \) balls in bucket, select \( n=2 \) with replacement.

Population data values: 0, 2, 4, 6, 8.

25 possible samples:

- (0,0) (2,0) (4,0) (6,0) (8,0)
- (0,2) (2,2) (4,2) (6,2) (8,2)
- (0,4) (2,4) (4,4) (6,4) (8,4)
- (0,6) (2,6) (4,6) (6,6) (8,6)
- (0,8) (2,8) (4,8) (6,8) (8,8)
7: Sample Variability
7.2 Sampling Distributions

Example: \( N=5 \), values: 0, 2, 4, 6, 8, \( n=2 \) (with replacement).

Prob. of each samples mean = \( 1/25 = 0.04 \)

\begin{align*}
\bar{x} = 0, \text{ one time} & \quad P(\bar{x} = 0) = 1/25 \\
\bar{x} = 1, \text{ two times} & \quad P(\bar{x} = 1) = 2/25 \\
\bar{x} = 2, \text{ three times} & \quad P(\bar{x} = 2) = 3/25 \\
\bar{x} = 3, \text{ four times} & \quad P(\bar{x} = 3) = 4/25 \\
\bar{x} = 4, \text{ five times} & \quad P(\bar{x} = 4) = 5/25 \\
\bar{x} = 5, \text{ four times} & \quad P(\bar{x} = 5) = 4/25 \\
\bar{x} = 6, \text{ three times} & \quad P(\bar{x} = 6) = 3/25 \\
\bar{x} = 7, \text{ two times} & \quad P(\bar{x} = 7) = 2/25 \\
\bar{x} = 8, \text{ one time} & \quad P(\bar{x} = 8) = 1/25 \\
\end{align*}
7: Sample Variability
7.2 Sampling Distributions

Example: \( N=5 \), values: 0, 2, 4, 6, 8, \( n=2 \) (with replacement).

\[
\begin{align*}
P(\bar{x} = 0) &= 1 / 25 \\
P(\bar{x} = 1) &= 2 / 25 \\
P(\bar{x} = 2) &= 3 / 25 \\
P(\bar{x} = 3) &= 4 / 25 \\
P(\bar{x} = 4) &= 5 / 25 \\
P(\bar{x} = 5) &= 4 / 25 \\
P(\bar{x} = 6) &= 3 / 25 \\
P(\bar{x} = 7) &= 2 / 25 \\
P(\bar{x} = 8) &= 1 / 25 
\end{align*}
\]

Represent this distribution function with a histogram.

Figure from Johnson & Kuby, 2008.

Note that intermediate values of 1, 3, 5, 7 are now possible.
Example: $N=5$, values: 0, 2, 4, 6, 8, $n=1$ or 2 (with replacement).
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.
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

Example: \( N=5 \), values: 0, 2, 4, 6, 8, \( n=2 \) (with replacement).

\[
\mu_{\bar{x}} = \sum_{\bar{x}} \bar{x}P(\bar{x})
\]

\[
\]

\[
\mu_{\bar{x}} = 4 \quad \text{Exactly as from formula!}
\]

\[
\mu_{\bar{x}} = \frac{1}{N} \sum_{i=1}^{N} x_i = 4
\]

\[
P(\bar{x} = 0) = 1/25 \quad P(\bar{x} = 1) = 2/25 \quad P(\bar{x} = 2) = 3/25 \quad P(\bar{x} = 3) = 4/25 \quad P(\bar{x} = 4) = 5/25 \quad P(\bar{x} = 5) = 4/25 \quad P(\bar{x} = 6) = 3/25 \quad P(\bar{x} = 7) = 2/25 \quad P(\bar{x} = 8) = 1/25
\]
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

**Example:** \( N=5 \), values: 0, 2, 4, 6, 8, \( n=2 \) (with replacement).

\[
\sigma_{\bar{x}}^2 = \sum_{\bar{x}} (\bar{x} - \mu_{\bar{x}})^2 P(\bar{x})
\]

\[
\]

\[
\sigma_{\bar{x}}^2 = 4
\]

\[
\sigma_{\bar{x}} = 2
\]

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2\sqrt{2}}{\sqrt{2}} = 2
\]

\[
P(\bar{x} = 0) = 1/25
\]

\[
P(\bar{x} = 1) = 2/25
\]

\[
P(\bar{x} = 2) = 3/25
\]

\[
P(\bar{x} = 3) = 4/25
\]

\[
P(\bar{x} = 4) = 5/25
\]

\[
P(\bar{x} = 5) = 4/25
\]

\[
P(\bar{x} = 6) = 3/25
\]

\[
P(\bar{x} = 7) = 2/25
\]

\[
P(\bar{x} = 8) = 1/25
\]

Exactly as from formula!
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

The CLT: 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

Questions?

Homework: Chapter 7 # 1, 6, 21, 23, 29, 35
Lecture Chapter 7.3 continued
Chapter 7: Sample Variability (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

Aside:

For a discrete distribution \( P(x) \), we found the mean \( \mu \) and variance \( \sigma^2 \) from:

\[
\mu = \sum_x xP(x) \quad \text{and} \quad \sigma^2 = \sum_x (x - \mu)^2 P(x) .
\]

AKA Don’t need to remember.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

Aside:

For a discrete distribution $P(x)$, we found the mean $\mu$ and variance $\sigma^2$ from

$$\mu = \sum_x xP(x) \quad \text{and} \quad \sigma^2 = \sum_x (x - \mu)^2 P(x).$$

For a continuous distribution $f(x)$, we find the mean $\mu$ and variance $\sigma^2$ from

$$\mu = \int_x x f(x) dx \quad \text{and} \quad \sigma^2 = \int_x (x - \mu)^2 f(x) dx.$$
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

Aside:

Uniform\(\left( a, b \right)\)

\[
f(x) = \frac{1}{b - a}
\]

\[
\mu = \int_{a}^{b} xf(x)dx = \frac{b - a}{2}
\]

\[
\sigma^2 = \int_{a}^{b} (x - \mu)^2 f(x)dx = \frac{(b - a)^2}{12}
\]

Normal\(\left( \mu, \sigma^2 \right)\)

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x - \mu}{\sigma})^2}
\]

\[
\mu = \int_{-\infty}^{\infty} xf(x)dx = \mu
\]

\[
\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx = \sigma^2
\]
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

**Example:** The random observation \( x \) comes from either a Uniform(0,200) population or Normal population.

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

**Uniform\((a=0,b=200)\)**

\[ f(x) = \frac{1}{b-a} \]

\[
\mu = 100 \\
\sigma = 57.7
\]

**Normal\((\mu=100,\sigma^2 = 200^2/12)\).**

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma}(x-\mu)^2} \]

\[
\mu = 100 \\
\sigma = 57.7
\]

Generate data \( x_1, \ldots, x_n \), for \( n=1, 2, 3, 4, 5, 15, 30, \) and 50. Calculate \( \bar{x} \), and repeat one million times.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

According to SDSM, if we had a sample $x_1, ..., x_n$ of size $n=1, 2, 3, 4, 5, 15, 30, \text{ and } 50$ from Uniform($a=0, b=200$), then

<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>

Theoretical values from the SDSM

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

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\mu = \frac{b-a}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

According to SDSM, if we had a sample \( x_1, ..., x_n \) of size \( n = 1, 2, 3, 4, 5, 15, 30, \) and 50 from Normal(\( \mu = 100, \sigma^2 = (57.7)^2 \)), then

<table>
<thead>
<tr>
<th>Sample Size, ( n )</th>
<th>Mean, ( \mu_x )</th>
<th>SD, ( \sigma_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>

Theoretical values from the SDSM

\[
\mu_x = \mu \\
\sigma_x = \frac{\sigma}{\sqrt{n}}
\]
7: Sample Variability
7.3 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.

Show histograms!
7: Sample Variability

7.3 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 \( n=1, 2, 3, 4, 5, 15, 30, \) and 50.

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

7.3 The Sampling Distribution of Sample Means

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

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

observations are uniform

observations are normal

Rowe, D.B.
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

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

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

observations are uniform

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

observations are normal

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

observations are uniform

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

observations are normal
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

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

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

observations are uniform

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

observations are normal

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\[ n=5 \quad 5 \times 10^6 \text{ observations} \]

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

observations are uniform

observations are normal

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

observations are uniform

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

observations are normal

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\[
\begin{align*}
\mu &= 100 \\
\sigma &= 57.73
\end{align*}
\]

- Observations are uniform
- Observations are normal

Rowe, D.B.
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

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

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

- Observations are uniform
- Observations are normal

Rowe, D.B.
7: Sample Variability
7.3 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.

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( \ldots )</th>
<th>( x_{5000000} )</th>
</tr>
</thead>
</table>

Groups of \( n=5 \)

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

Mean of groups

| \( \bar{x}_1 \) | \( \bar{x}_2 \) | \( \ldots \) | \( \bar{x}_{10^6} \) |

8 data sets of Uniform and Normal

Histogram of \( \bar{x} \)'s
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

Computed sample means and standard deviations from the one million means $\bar{x}_1, ..., \bar{x}_{10^6}$

<table>
<thead>
<tr>
<th>Sample Size $n$</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

7.3 The Sampling Distribution of Sample Means

$n = 1 \times 10^6$ means $\bar{x}$

$\mu = 100$

$\sigma = 57.73$

Histogram of means from uniform

Histogram of means from normal

$S_{\bar{x}}$

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\[ n=3 \quad 1 \times 10^6 \text{ means } \bar{x} \]

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

Histogram of means from uniform

Histogram of means from normal

\[ S_{\bar{x}} \]

\[ \bar{X} \]

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

$n=5 \quad 1 \times 10^6$ means $\bar{x}$

$\mu = 100$

$\sigma = 57.73$

Histogram of means from uniform

$\bar{X}$

$S_{\bar{X}}$

Histogram of means from normal

$\bar{X}$

$S_{\bar{X}}$

Rowe, D.B.
### 7: Sample Variability

#### 7.3 The Sampling Distribution of Sample Means

- **$n=15$**
- **$1 \times 10^6$ means $\bar{x}$**

- **$\mu = 100$**
- **$\sigma = 57.73$**

Histogram of means from uniform

Histogram of means from normal

*Rowe, D.B.*
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\( n = 30 \quad 1 \times 10^6 \text{ means } \overline{x} \)

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

Histogram of means from uniform

Histogram of means from normal
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal

Rowe, D.B.
7: Sample Variability
7.3 The Sampling Distribution of Sample Means

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 # 1, 6, 21, 23, 29, 35
TA go over Exam 3.