Class 20

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 8.3 - 8.4

Lecture Chapter 8.5

Review Chapter 7
Recap Chapter 8.3 - 8.4
Example 1: Friend’s Party.

$H_0$: ”The Party will be a great time”

vs.

$H_a$: “The party will be a dud.”

Example 2: Math 1700 Students Height

$H_0$: The mean height of Math 1700 students is 69”, $\mu = 69$.

vs.

$H_a$: The mean height of Math 1700 students is not 69”, $\mu \neq 69$. 
8: Introduction to Statistical Inference
8.4 The Nature of Hypothesis Testing

Example 1: Friend’s Party

\( H_0 \): ”The Party will be a great time”

vs.

\( H_a \): “The party will be a dud.”

If do not go to party and it’s great, we made an error in judgment.

If go to party and it’s a dud, we made an error in judgment.
8: Introduction to Statistical Inference
8.4 The Nature of Hypothesis Testing

**Example 2: Math 1700 Height**

\( H_0: \mu = 69” \)

vs.

\( H_a: \mu \neq 69” \)

<table>
<thead>
<tr>
<th></th>
<th>( \mu = 69 )</th>
<th>( \mu \neq 69 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to reject ( H_0. )</td>
<td>Correct Decision</td>
<td>Type II Error</td>
</tr>
<tr>
<td>Reject ( H_0. )</td>
<td>Type I Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

If we reject \( H_0 \) and it is true, we made in error in judgment.

If we do not reject \( H_0 \) and it is false, we have made an error in judgment.

Rowe, D.B.
8: Introduction to Statistical Inference
8.4 The Nature of Hypothesis Testing

Type I Error: …true null hypothesis $H_0$ is rejected.

**Level of Significance ($\alpha$):** The probability of committing a type I error. (Sometimes $\alpha$ is called the false positive rate.)

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Do Not</strong></td>
<td>Type A</td>
<td>Type II</td>
</tr>
<tr>
<td><strong>Reject $H_0$</strong></td>
<td>Correct Decision</td>
<td>Error ($\beta$)</td>
</tr>
<tr>
<td></td>
<td>$(1-\alpha)$</td>
<td></td>
</tr>
</tbody>
</table>

Type II Error: … favor … null hypothesis that is actually false.

Type II Probability ($\beta$): The probability of committing a type II error.
8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

We need to determine a measure that will quantify what we should believe.

**Test Statistic:** A random variable whose value is calculated from the sample data and is used in making the decision “reject $H_0$: or “fail to reject $H_0$.”

**Example:** Friend’s Party  
Fraction of parties that were good.

**Example:** Math 1700 Heights  
Sample mean height.
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean (σ Known): Probability Approach

Step 1 The Set-Up: Null ($H_0$) and alternative ($H_a$) hypotheses

$H_0$: $\mu = 69$” vs. $H_a$: $\mu \neq 69”$

Step 2 The Hypothesis Test Criteria: Test statistic.

$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

$\sigma$ known, $n$ is “large” so by CLT $\bar{x}$ is normal $z^*$ is normal

Step 3 The Sample Evidence: Calculate test statistic.

$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$

$n=15$, $\bar{x} = 67.2$, $\sigma = 4$

Step 4 The Probability Distribution:

$P(z > |z^*|) = p$ – value $\rightarrow 0.0819$

Step 5 The Results:

$p$ – value $\leq \alpha$, reject $H_0$, $p$ – value $> \alpha$ fail to reject $H_0$

$\alpha = 0.05$
Chapter 8: Introduction to Statistical Inference

Questions?

Homework: Chapter 8 # 5, 15, 19, 22, 23, 24, 25, 35, 47, 51, 57, 59, 81, 87, 91, 93, 97, 105, 106, 107, 109, 119, 139, 140, 145, 149, 157, 159
Lecture Chapter 8.5
Chapter 8: Introduction to Statistical Inference (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean \( \mu \) (\( \sigma \) Known):
A Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS
Step 1 The Set-Up:

Step 2 The Hypothesis Test Criteria:

Step 3 The Sample Evidence:

Step 4 The Probability Distribution:

Step 5 The Results:
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS
Step 1 The Set-Up:
   a. Describe the population parameter of interest.

   The population parameter of interest is the mean $\mu$, the height of Math 1700 students.
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Probability Approach

THE PROBABILITY-VALUE HYPOTHESIS TEST: 5 STEPS

Step 1 The Set-Up:
   b. State the null hypothesis ($H_0$) and the alternative hypotheses ($H_a$).

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Greater than or equal to (≥)</td>
<td>Less than (&lt;)</td>
</tr>
<tr>
<td>2. Less than or equal to (≤)</td>
<td>Greater than (&gt; )</td>
</tr>
<tr>
<td>3. Equal to (=)</td>
<td>Not equal to (≠)</td>
</tr>
</tbody>
</table>

$H_0: \mu = 69''$ vs. $H_a: \mu \neq 69''$

Figure from Johnson & Kuby, 2012.
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

Scenario:
True $\mu$ not likely
$\mu_0=69$.

Reject $H_0$: $\mu=\mu_0$
(do not believe $H_0$)

Let's say we set a cut-off mean
$\mu_{critical}=71$

Hypothesized mean $\mu_0=69$

Sample mean $\bar{x}=72$
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

Scenario:
True $\mu$ likely $\mu_0 = 69$.

Fail to reject $H_0: \mu = \mu_0$ (not enough evidence not to believe $H_0$)

let's say we set a cut-off mean $\mu_{critical} = 71$

hypothesized mean $\mu_0 = 69$

sample mean $\bar{x} = 70$
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

We need a “better” (objective) way to set a “cut-off “ value or “cut-off” values for which we would either believe $H_0$ or for which we would not have enough evidence not believe $H_0$.

We need to use the normal distribution and probabilities.
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 2 The Hypothesis Test Criteria:

a. Check the assumptions.

Assume we know from past experience that $\sigma=4$.
Assume that $n$ is “large” so that by the CLT
$\bar{x}$ is normally distributed.

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

by SDSM
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 2 The Hypothesis Test Criteria:

b. Identify the probability distribution and the test statistic to be used.

The standard normal distribution is to be used because $\bar{x}$ is expected to have a normal distribution.

Test Statistic for Mean:

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

where $\sigma$ is assumed to be known (8.4)
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean (σ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 2 The Hypothesis Test Criteria:
c. Determine the level of significance, \( \alpha \).

After much consideration, we assign a tolerable probability of a Type I error to be \( \alpha = 0.05 \).

Type I Error: When a true null hypothesis \( H_0 \) is rejected.
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean (σ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 3 The Sample Evidence:

a. Collect a sample of information.
   Take a random sample from the population with mean μ that being questioned.

b. Calculate the value of the test statistic.

\[
z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74
\]

Assuming \( n = 15 \) and 67.2 is sample mean. With known \( \sigma = 4 \).
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean (σ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 4 The Probability Distribution:
   a. Determine the critical region and critical value(s).

Critical Region: The set of values for the test statistic that will cause us to reject the null hypothesis.

Critical value(s): The “first” or “boundary” value(s) of the critical region(s).
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean (σ Known): Classical Approach

There are three possible hypothesis pairs for the mean.

\[ H_0: \mu \geq \mu_0 \text{ vs. } H_a: \mu < \mu_0 \]

Reject \( H_0 \) if less than

\[ z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < -z(\alpha) \]

data indicates \( \mu < \mu_0 \) because \( \bar{x} \) is “a lot” smaller than \( \mu_0 \)
There are three possible hypothesis pairs for the mean.

\[ H_0: \mu \leq \mu_0 \text{ vs. } H_a: \mu > \mu_0 \]

Reject \( H_0 \) if \( z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \) is greater than \( z(\alpha) \)

Data indicates \( \mu > \mu_0 \) because \( \bar{x} \) is "a lot" larger than \( \mu_0 \)
There are three possible hypothesis pairs for the mean.

\[ H_0: \mu = \mu_0 \text{ vs. } H_a: \mu \neq \mu_0 \]

Reject \( H_0 \) if less than \(-z(\alpha / 2)\)

or if is greater than \( z(\alpha / 2)\)

\[
z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}
\]

data indicates \( \mu \neq \mu_0, \bar{x} \text{ far from } \mu_0 \)
8: Introduction to Statistical Inference
8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

THE CLASSICAL HYPOTHESIS TEST: 5 STEPS
Step 4 The Probability Distribution:
   a. Determine the critical region and critical value(s).
   b. Determine whether or not the calculated test statistic is in the critical region.

$H_0: \mu = 69" \text{ vs. } H_a: \mu \neq 69"$

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

$z(.025) = 1.96$

since two sided test.
THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 5 The Results:
   a. State the decision about $H_0$.
      Need a decision rule.

Decision rule:
   a. If the test statistic falls within the critical region, then the decision must be reject $H_0$.
   b. If the test statistic is not in the critical region, then the decision must be fail to reject $H_0$. 
THE CLASSICAL HYPOTHESIS TEST: 5 STEPS

Step 5 The Results:

b. State the conclusion about \( H_a \).

With \( \alpha = 0.05 \), there is not sufficient evidence to reject \( H_0 \).

Fail to reject \( H_0 \).

Figure from Johnson & Kuby, 2012.
Let’s examine the hypothesis test

\[ H_0: \mu \leq 69'' \text{ vs. } H_a: \mu > 69'' \]

with \( \alpha=0.05 \) for the heights of Math 1700 students.

Generate random data values.
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

Generated $15 \times 10^6$

normal data values

from $\mu = 69''$ and $\sigma = 4''$.

Calculated

$1 \times 10^6$ means with $n=15$.

(Will repeat for $\mu = 72''$.)

Rowe, D.B.
8. Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

\[ H_0: \mu \leq \mu_0 \text{ vs. } H_a: \mu > \mu_0 \]

\[ n=15 \]
\[ 1 \times 10^6 \]
\[ \bar{x}'s \]

\[ \alpha=.05 \]

When the true mean \( \mu = 69'' \), we reject \( H_0 \) \( \alpha \) fraction of the time.

Commit a Type I Error.

Given \( \alpha \), we want \( \mu_{critical} \).
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

Instead of \( \mu_{\text{critical}} \) we find critical \( z \), \( z_{\text{critical}} = z(\alpha) \).

Do this by assuming that \( H_0: \mu = 69 \) is true, then calculate

\[
z = \frac{\bar{x} - 69}{4 / \sqrt{15}}
\]
8: Introduction to Statistical Inference  

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

\[ H_0: \mu \leq 69 \text{ vs. } H_a: \mu > 69 \]

\[ \alpha = 0.05 \]

When the true mean \( \mu = 72 \)”, we do not reject \( H_0 \) \( \beta \) fraction of the time.

Commit a Type II Error

When the true mean \( \mu = 72 \)”, we do not reject \( H_0 \) \( \beta \) fraction of the time.

Commit a Type II Error
8: Introduction to Statistical Inference  

8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

Hypothesis: $H_0: \mu \leq \mu_0$ vs. $H_a: \mu > \mu_0$

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ True ($\mu=69''$)</th>
<th>$H_0$ False ($\mu=72''$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to Reject $H_0$</td>
<td>Correct Decision (1-$\alpha$)</td>
<td>Type II Error ($\beta$)</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>Type I Error ($\alpha$)</td>
<td>Correct Decision (1-$\beta$)</td>
</tr>
</tbody>
</table>

Power of the test: 
$$1 - \beta = P(\text{Reject } H_0 | H_0 \text{ False})$$

Discrimination ability. Ability to detect difference.
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

\[ z_s = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \]

- \( n = 15 \)
- \( 1 \times 10^6 \)
- \( \bar{x}'s \)

- Discrimination ability.
- Ability to detect difference.

\[ z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \]

\[ 1 - \beta = P(\text{Reject } H_0 | H_0 \text{ False}) \]

Rowe, D.B.
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

We want our $\alpha$, Prob of Type I Error to be small.

So why not just decrease $\alpha$?

Decreasing $\alpha$ increases $\beta$.

And vice versa.
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

$H_0: \mu \leq \mu_0$ vs. $H_a: \mu > \mu_0$

What is the solution?

Increase $n$.

Figure shows $n$ increased to $n = 30$ from $n = 15$.

Note $\alpha$ and $\beta$ both smaller with larger $n$. 

Rowe, D.B.
8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

What is the solution?

Increase $n$.

Figure shows $n$ increased to $n = 30$ from $n = 15$.

Note $\alpha$ and $\beta$ both smaller with larger $n$.

$H_0: \mu \leq \mu_0$ vs. $H_a: \mu > \mu_0$

$z = \frac{\bar{x} - 69}{4 / \sqrt{30}}$

$n=30$

$1 \times 10^6$

$\bar{x}'s$

$1 - \alpha$

$1 - \beta$

Rowe, D.B.
Chapter 8: Introduction to Statistical Inference

Questions?

Homework: Chapter 8 # 5, 15, 19,
22, 23, 24, 25, 35, 47, 51,
57, 59, 81,
87, 91, 93, 97, 105, 106, 107,
109, 119,
139, 140, 145, 149, 157, 159
Review Chapter 7
(Exam 5 Chapter)

Just the highlights!
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

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 $\bar{x}$. 

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

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

\[
\begin{array}{c|c|c}
 x & P(x) & P(x) \\
 0 & 1/5 & 0.20 \\
 2 & 1/5 & 0.16 \\
 4 & 1/5 & 0.12 \\
 6 & 1/5 & 0.08 \\
 8 & 1/5 & 0.04 \\
\end{array}
\]

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

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

$\sigma = \sqrt{8} = 2\sqrt{2}$
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

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

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

25 possible samples
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

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

\[
\begin{array}{cccccc}
0 & 1 & 2 & 3 & 4 & \text{Prob. of each samples} \\
1 & 2 & 3 & 4 & 5 & \text{mean } = 1/25 = 0.04 \\
2 & 3 & 4 & 5 & 6 & \\
3 & 4 & 5 & 6 & 7 & \\
4 & 5 & 6 & 7 & 8 & \\
\end{array}
\]

\[\overline{x} = 0, \text{ one time} \]
\[P(\overline{x} = 0) = 1/25\]

\[\overline{x} = 1, \text{ two times} \]
\[P(\overline{x} = 1) = 2/25\]

\[\overline{x} = 2, \text{ three times} \]
\[P(\overline{x} = 2) = 3/25\]

\[\overline{x} = 3, \text{ four times} \]
\[P(\overline{x} = 3) = 4/25\]

\[\overline{x} = 4, \text{ five times} \]
\[P(\overline{x} = 4) = 5/25\]

\[\overline{x} = 5, \text{ four times} \]
\[P(\overline{x} = 5) = 4/25\]

\[\overline{x} = 6, \text{ three times} \]
\[P(\overline{x} = 6) = 3/25\]

\[\overline{x} = 7, \text{ two times} \]
\[P(\overline{x} = 7) = 2/25\]

\[\overline{x} = 8, \text{ one time} \]
\[P(\overline{x} = 8) = 1/25\]
7: Sample Variability
7.2 The Sampling Distribution of Sample Means

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

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

Represent this distribution function with a histogram.

Figure from Johnson & Kuby, 2012.

note that intermediate values of 1, 3, 5, 7 are now possible.
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 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
7.2 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(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

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
7.2 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.

<table>
<thead>
<tr>
<th>i.e. ( n=5, \ 5\times10^6 )</th>
<th>Groups of ( n=5 )</th>
<th>Mean of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( x_1, x_2, x_3, x_4, x_5 )</td>
<td>( \bar{x}_1 )</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( x_6, x_7, x_8, x_9, x_{10} )</td>
<td>( \bar{x}_2 )</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>( x_{5000000} )</td>
<td>( x_{4999996}, \ldots, x_{5000000} )</td>
<td>( \bar{x}_{10^6} )</td>
</tr>
</tbody>
</table>

8 data sets of Uniform and Normal

Histogram of \( \bar{x} \)'s
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

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

Histogram of means from uniform

Histogram of means from normal

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

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

\[ n=2 \quad 1\times10^6 \text{ means } \overline{x} \]

Histogram of means from uniform

Histogram of means from normal

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

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

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

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

Histogram of means from uniform

Histogram of means from normal

same classes

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

scale same

limited range

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

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

$\mu = 100 \quad \sigma = 57.73$

Histogram of means from uniform

Histogram of means from normal

scale same

$S_{\bar{x}}$

same classes

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

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

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

Histogram of means from uniform

Histogram of means from normal

scale same

$S_{\bar{x}}$

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal

same classes

scale same

Rowe, D.B.
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

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

Histogram of means from uniform

Histogram of means from normal

$\mu = 100$

$\sigma = 57.73$

Rowe, D.B.
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!
Assume that this is a population of data. Does this population look normally distributed?

**Height**

- \( \mu = 66.7 \)
- \( \sigma = 3.9 \)

Put normal with same \( \mu \) and \( \sigma \).
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 \( \sigma_{\bar{x}} = \frac{3.9}{\sqrt{1}} = 3.9 \).

\[ \mu = 66.7 \]
\[ \sigma = 3.9 \]

N=32 values

Rowe, D.B.
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, \( \mu_x = 66.7 \) and \( \sigma_x = \frac{3.9}{\sqrt{3}} = 2.3 \).

Histogram of the 1 million means

\( n=3 \)

\( \mu = 66.7 \)
\( \sigma = 3.9 \)

Put normal with same \( \mu_x \) and \( \sigma_x \).
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_\bar{x} = 66.7$ and $\sigma_\bar{x} = \frac{3.9}{\sqrt{5}} = 1.7$.

$N=32$ values

$\mu = 66.7$

$\sigma = 3.9$

Histogram of the 1 million means $n=5$

Put normal with same $\mu_\bar{x}$ and $\sigma_\bar{x}$. 
7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_x = 66.7$ and $\sigma_x = \frac{3.9}{\sqrt{15}} = 1.0$.

Histogram of the 1 million means

$N=32$ values

$\mu = 66.7$

$\sigma = 3.9$

By CLT becomes normal

Put normal with same $\mu_x$ and $\sigma_x$.
7: Sample Variability
7.3 Application of the Sampling Distribution of Sample Means

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.

$$P(a < \bar{x} < b)$$

$$P(b < \bar{x})$$

$$P(\bar{x} < a)$$
7: Sample Variability
7.3 Application of the Sampling Distribution of Sample Means

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

$$P(a < x < b)$$

$$P(c < z < d)$$

$$z = \frac{x - \mu_x}{\sigma_x}$$

$$P(b < x) = P(d < z)$$

$$P(x < a) = P(z < c)$$

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 $\bar{x}$ from a random sample of $n=15$ heights is greater than 69” when $\mu = 66.7$ and $\sigma = 3.9$?

\[
P(69 < \bar{x})
\]

we first convert to $z$ scores

\[
z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}
\]

where $d = \frac{b - \mu}{\sigma_{\bar{x}}} = \frac{69 - 66.7}{3.9 / \sqrt{15}} = 2.28$, then use the table in book.

\[
1 - P(z < 2.28) = 1 - 0.9987 = 0.0012
\]
7: Sample Variability

Questions?

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