Class 20

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 8.4 - 8.5

Lecture Chapter 8.6

Review Chapter 7
Recap Chapter 8.4 - 8.5
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.”

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”, $\mu \neq 69$”. 

Rowe, D.B.
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.”

<table>
<thead>
<tr>
<th></th>
<th>Party Great</th>
<th>Party a dud.</th>
</tr>
</thead>
<tbody>
<tr>
<td>We go.</td>
<td>Correct Decision</td>
<td>Type II Error</td>
</tr>
<tr>
<td>We do not go.</td>
<td>Type I Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

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 in error in judgment.
Example 2: Math 1700 Height

\[ H_0: \mu = 69'' \]

vs.

\[ H_a: \mu \neq 69'' \]

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

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

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

**Type II Probability ($\beta$):**
The probability of committing a type II error.

<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 Correct</td>
<td>Type II Error</td>
</tr>
<tr>
<td><strong>Reject $H_0$</strong></td>
<td>Decision $(1-\alpha)$</td>
<td>($\beta$)</td>
</tr>
<tr>
<td><strong>Reject $H_0$</strong></td>
<td>Type I Error</td>
<td>Type B Correct</td>
</tr>
<tr>
<td></td>
<td>$(\alpha)$</td>
<td>Decision $(1-\beta)$</td>
</tr>
</tbody>
</table>
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 (H₀) and alternative (Hₐ) hypotheses
H₀: μ = 69” vs. Hₐ: μ ≠ 69”

Step 2 The Hypothesis Test Criteria: Test statistic.
\[ z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \]
σ 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 # 13, 15, 21, 23, 25, 33, 45, 53, 55, 73, 77, 87, 89, 93, 103, 105, 107, 137, 143, and 149
Lecture Chapter 8.6
Chapter 8: Introduction to Statistical Inference (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
8: Introduction to Statistical Inference
8.6 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
8.6 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
8.6 Hypothesis Test of Mean ($\sigma$ Known): Classical Approach

THE CLASSICAL 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 (\geq)</td>
<td>Less than (&lt;)</td>
</tr>
<tr>
<td>2. Less than or equal to (\leq)</td>
<td>Greater than (&gt; )</td>
</tr>
<tr>
<td>3. Equal to (=)</td>
<td>Not equal to (\neq)</td>
</tr>
</tbody>
</table>

\[ H_0: \mu = 69" \quad \text{vs.} \quad H_a: \mu \neq 69" \]

Figure from Johnson & Kuby, 2008.
8: Introduction to Statistical Inference
8.6 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 $H_0$)
8: Introduction to Statistical Inference
8.6 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 $\rightarrow \mu_0 = 69$ 
Sample mean $\xrightarrow{\bar{x}} 70$
8: Introduction to Statistical Inference
8.6 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: Introduction to Statistical Inference
8.6 Hypothesis Test of Mean (σ 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 σ=4. Assume that n is “large” so that by the CLT, \( \bar{x} \) is normally distributed.

\[
\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{by SDSM}
\]
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)
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.
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 \( \mu \) 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
8.6 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
8.6 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

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

data indicates \( \mu < \mu_0 \) because \( \bar{x} \) is “a lot” smaller than \( \mu_0 \)
8: Introduction to Statistical Inference
8.6 Hypothesis Test of Mean (σ Known): Classical Approach

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 greater then

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

\( z(\alpha) \)

data indicates \( \mu > \mu_0 \)
because \( \bar{x} \) is “a lot” larger than \( \mu_0 \)
8: Introduction to Statistical Inference
8.6 Hypothesis Test of Mean (σ Known): Classical Approach

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

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

\(-z(\alpha / 2)\)

or if is greater than

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

\(z(\alpha / 2)\)

data indicates \( \mu \neq \mu_0 \), \( \bar{x} \) far from \( \mu_0 \)
8: Introduction to Statistical Inference
8.6 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.

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

$P(z > z(\alpha / 2)) = \alpha / 2$, $z(0.025) = 1.96$

since two sided test.

Figure from Johnson & Kuby, 2008.
8: Introduction to Statistical Inference
8.6 Hypothesis Test of Mean (σ Known): Classical Approach

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, 2008.
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

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

H₀: μ ≤ μ₀ vs. H₁: μ > μ₀

Generated 15×10⁶ normal data values from μ = 69" and σ=4". (Will repeat for μ = 72".)

Calculated 1×10⁶ means with n=15.
8: Introduction to Statistical Inference  
8.6 Hypothesis Test of Mean (σ Known): Classical Approach

\[ H_0: \mu \leq \mu_0 \text{ vs. } H_1: \mu > \mu_0 \]

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

Commit a Type I Error.

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

8.6 Hypothesis Test of Mean ($\sigma$ 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

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

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

\[ \alpha = .05 \]

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

Commit a Type II Error

\[ 1 - \beta \]

\[ \beta \]

\[ \mu_{critical} \text{ (same)} \]
8: Introduction to Statistical Inference

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

- Null Hypothesis: \(H_0: \mu \leq \mu_0\)
- Alternative Hypothesis: \(H_1: \mu > \mu_0\)

- Fail to Reject \(H_0\) when \(\mu = 69\)"
- Correct Decision \((1 - \alpha)\)
- Type II Error \((\beta)\)

- Reject \(H_0\) when \(\mu = 72\)"
- Type I Error \((\alpha)\)
- Correct Decision \((1 - \beta)\)

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

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

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

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

\[ z = \frac{\bar{x} - 69}{4 / \sqrt{15}} \]

\[ H_0: \mu \leq \mu_0 \text{ vs. } H_1: \mu > \mu_0 \]

Fail to Reject

\( z \)

Reject

H0 True
(\( \mu = 69'' \))

H0 False
(\( \mu = 72'' \))

Fail to Reject

Correct Decision
(1 - \( \alpha \))

Type II Error
(\( \beta \))

Reject

Type I Error
(\( \alpha \))

Correct Decision
(1 - \( \beta \))

Power of the test:

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

Discrimination ability.
Ability to detect difference.

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

8.6 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.

$H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$

$n=15 \times 10^6$

$\bar{x}'s$

$z_{critical}$
8: Introduction to Statistical Inference  

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

H$_0$: $\mu \leq \mu_0$ vs. H$_1$: $\mu > \mu_0$

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

8.6 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. $H_1: \mu > \mu_0$

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

$n=30$

$1 \times 10^6$

$\bar{x}'s$

$1-\alpha$

$1-\beta$
Chapter 8: Introduction to Statistical Inference

Questions?

Homework: Chapter 8 # 13, 15, 21, 23, 25, 33, 45 53, 55, 73, 77 87, 89, 93, 103, 105, 107 137, 143, and 149
Review Chapter 7
(Exam 5 Chapter)

Just the highlights!
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>

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

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

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

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

\[
\begin{align*}
\bar{x} &= 0, \text{ one time} & P(\bar{x} = 0) &= 1/25 \\
\bar{x} &= 1, \text{ two times} & P(\bar{x} = 1) &= 2/25 \\
\bar{x} &= 2, \text{ three times} & P(\bar{x} = 2) &= 3/25 \\
\bar{x} &= 3, \text{ four times} & P(\bar{x} = 3) &= 4/25 \\
\bar{x} &= 4, \text{ five times} & P(\bar{x} = 4) &= 5/25 \\
\bar{x} &= 5, \text{ four times} & P(\bar{x} = 5) &= 4/25 \\
\bar{x} &= 6, \text{ three times} & P(\bar{x} = 6) &= 3/25 \\
\bar{x} &= 7, \text{ two times} & P(\bar{x} = 7) &= 2/25 \\
\bar{x} &= 8, \text{ one time} & P(\bar{x} = 8) &= 1/25
\end{align*}
\]
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.

Note that intermediate values of 1, 3, 5, 7 are now possible.

Figure from Johnson & Kuby, 2008.
7: Sample Variability
7.3 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.
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
7.3 The Sampling Distribution of Sample Means

According to SDSM, if we had a sample $x_1, \ldots, x_n$ of size $n=1, 2, 3, 4, 5, 15, 30, \text{ and } 50$ 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>

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

Theoretical values from the SDSM

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
7: Sample Variability
7.3 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 $n=1, 2, 3, 4, 5, 15, 30, \text{ and } 50$.

8 data sets of Uniform and Normal random observations
### 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>i.e. ( n=5, \ 5 \times 10^6 )</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>( \overline{x}_1 )</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( x_6, x_7, x_8, x_9, x_{10} )</td>
<td>( \overline{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>( \overline{x}_{10^6} )</td>
</tr>
</tbody>
</table>

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

\[ \mu = 100 \]

\[ \sigma = 57.73 \]

Histogram of means from uniform

Histogram of means from normal
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

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

Histogram of means from normal

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

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\( n = 4 \quad 1 \times 10^6 \) means

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

Histogram of means from uniform

\( \bar{X} \rightarrow S_{\bar{X}} \)

Histogram of means from normal

\( \bar{X} \rightarrow S_{\bar{X}} \)

Rowe, D.B.
7: Sample Variability

7.3 The Sampling Distribution of Sample Means

\[ n = 5 \quad 1 \times 10^6 \text{ means} \]

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

Histogram of means from uniform

Histogram of means from normal

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

7.3 The Sampling Distribution of Sample Means

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

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

Histogram of means from uniform

Histogram of means from normal

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

Rowe, D.B.
7: Sample Variability

7.3 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

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!
Assume that this is a population of data.

Does this population look normally distributed?

$N=108$ values

$\mu = 67.0$

$\sigma = 3.7$
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 $n=15$ from the population of $N$ heights with replacement $10^6$ times.

$N=108$ values
7: Sample Variability

7.4 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_x = 67.0$ and $\sigma_x = \frac{3.7}{\sqrt{1}} = 3.7$.

N = 108 values

$\mu = 67.0$
$\sigma = 3.7$

Histogram of the 1 million means

$n=1$

Put normal with same $\mu_x$ and $\sigma_x$. 

Rowe, D.B.
7: Sample Variability
7.4 Application of the Sampling Distribution of Sample Means

By SDSM, $\mu_x = 67.0$ and $\sigma_x = \frac{3.7}{\sqrt{3}} = 2.1$.

N=108 values

$\mu = 67.0$

$\sigma = 3.7$
7: Sample Variability

7.4 Application of the Sampling Distribution of Sample Means

By SDSM, \( \mu_x = 67.0 \) and \( \sigma_x = \frac{3.7}{\sqrt{5}} = 1.7 \).

\[ N = 108 \text{ values} \]

\[ \mu = 67.0 \]

\[ \sigma = 3.7 \]
7: Sample Variability

7.4 Application of the Sampling Distribution of Sample Means

By SDSM, \( \mu_x = 67.0 \) and \( \sigma_x = \frac{3.7}{\sqrt{15}} = 1.0 \).

\( N=108 \) values

\( \mu = 67.0 \)

\( \sigma = 3.7 \)

By CLT becomes normal
Now that we believe $\overline{x}$ is normal with $\mu_{\overline{x}} = \mu$ and $\sigma_{\overline{x}} = \sigma / \sqrt{n}$, we can find probabilities by first converting to $z$ scores.

\begin{align*}
P(a < \overline{x} < b) & \quad \text{same area} \\
P(c < z < d) & \quad \text{same area} \\
Z = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} & \quad C = \frac{a - \mu_{\overline{x}}}{\sigma_{\overline{x}}} \quad D = \frac{b - \mu_{\overline{x}}}{\sigma_{\overline{x}}} \\
\end{align*}

and use the table in book.
7: Sample Variability
7.4 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 $\mu = 67.0$ and $\sigma = 3.7$?

We first convert to $z$ scores

$$z = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}}$$

where $d = \frac{b - \mu_x}{\sigma_{\bar{x}}} = \frac{69 - 67}{3.7 / \sqrt{15}} = 2.09$, then use the table in book.

$$P(2.09 < z) = .5 - P(0 < z < 2.09) = .5 - .4817 = .0183$$