Distribution of the sample mean, \( \bar{y} \).

This lab looks at the sample mean \( \bar{y} \) by sampling from a population of 250 females who took an introductory statistics class. The characteristic we are interested in is the average or mean height of female students. We can look at the variation in a sample statistic by taking many samples (called repeated sampling) from the population and looking at the distribution of the sample statistics obtained.

1. Refer to the table titled “Heights of all females in the population.” This table contains a listing of the heights, in centimeters (cm), of the population members. Rather than list the names of the population members, this table numbers them by rows numbered (00, 01, \ldots, 09, 10, \ldots, 24) and columns numbered (0,1,2,\ldots, 9). For example, student 037 (Row 03 and Column 7) is 165 cm tall.

   a) Use the random numbers generated by JMP to select 4 random samples of size 10. Record the student numbers and heights on the answer sheet and calculate the sample mean for each sample.
   b) How different are the four sample means? What causes the differences?

This population has a population mean height \( \mu = 166.6 \text{ cm} \) and a population standard deviation \( \sigma = 7.6 \text{ cm} \). The distribution of heights is given below.

We are interested in the distribution of the sample mean for random samples of size \( n = 10 \).

c) Describe the shape of the theoretical distribution of the sample mean, \( \bar{y} \) for samples of size \( n = 10 \).

d) What is the mean of the theoretical distribution of the sample mean for samples of size \( n = 10 \)?

e) What is the standard deviation of the theoretical distribution of the sample mean for samples of size \( n = 10 \)?
Combine your 4 samples together to create a random sample of size \( n = 40 \).

f) Calculate the sample mean for this larger random sample.

g) Describe the shape of the theoretical distribution of the sample mean, \( \bar{y} \) for samples of size \( n = 40 \).

h) What is the mean of the theoretical distribution of the sample mean for samples of size \( n = 40 \)?

i) What is the standard deviation of the theoretical distribution of the sample mean for samples of size \( n = 40 \)?

j) Which should be a better estimate of the population mean, the sample mean for a random sample of size \( n = 10 \) or size \( n = 40 \)? Support your answer statistically.

2. Each member of the group should take one of the random samples and

a) Calculate the sample mean and sample standard deviation. Use your calculator to do these calculations.

b) Construct a 95% confidence interval for the population mean height of female students.

c) How many of your confidence intervals contain the population mean \( \mu = 166.6 \text{ cm} \)?

d) Put this on the white board. Once all the groups in lab have put their numbers of intervals that contain the population mean on the board, compute a relative frequency of capturing the population mean for your lab section. You can continue to work on other parts of this lab while waiting for all the groups to put their numbers on the board.

e) What should be the long run relative frequency of capturing the population mean?

f) Suppose you were to combine the four random samples of size \( n = 10 \) to create a random sample of size \( n = 40 \), would the confidence interval for the population mean height based on the sample of size \( n = 40 \) be wider, about the same width of narrower than a confidence interval for the population mean height based on the sample of size \( n = 10 \)? Explain your answer briefly. Note: You do not, and should not, compute the confidence interval for the sample of size \( n = 40 \).

3. Four random samples of size \( n = 10 \) were combined to create a random sample of size \( n = 40 \). For this larger random sample, the sample mean is \( \bar{y} = 164.45 \text{ cms} \) and the sample standard deviation is \( s = 10.213 \text{ cms} \). A normal quantile plot, outlier box plot and histogram for the sample data appear on the next page.

a) Comment on each of the plots and indicate how each supports or fails to support the condition that the distribution of female heights is nearly normal.

b) Construct a 95% confidence interval for the population mean height of female students.

c) Based on this interval could the population mean height of female students be 166.6 cms? Explain briefly.

d) Test the hypothesis that the population mean height of female students is 166.6 cms against the alternative that the population mean height is something different from 166.6. Go through the steps of a hypothesis test.
Distribution of a random sample of heights of 40 female students.
Stat 104 – Laboratory 11
Group Answer Sheet

Names of Group Members: ____________________, ____________________
____________________, ____________________

1. | Sample 1 | Sample 2 | Sample 3 | Sample 4 |
   | Number  | Height  | Number  | Height  |
   |         |         |         |         |
   |         |         |         |         |
   |         |         |         |         |
   |         |         |         |         |
   |         |         |         |         |
   | Sample Mean | Sample Mean | Sample Mean | Sample Mean |

b) How different are the four sample means? What causes the differences?

c) Describe the shape of the theoretical distribution of the sample mean, \( \bar{y} \), for samples of size \( n = 10 \).

d) What is the mean of the theoretical distribution of the sample mean for samples of size \( n = 10 \)?

e) What is the standard deviation of the theoretical distribution of the sample mean for samples of size \( n = 10 \)?
f) Calculate the sample mean for this larger random sample \( n = 40 \).

g) Describe the shape of the theoretical distribution of the sample mean, \( \bar{y} \) for samples of size \( n = 40 \).

h) What is the mean of the theoretical distribution of the sample mean for samples of size \( n = 40 \)?

i) What is the standard deviation of the theoretical distribution of the sample mean for samples of size \( n = 40 \)?

j) Which should be a better estimate of the population mean, the sample mean for a random sample of size \( n = 10 \) or size \( n = 40 \)? Support your answer statistically.

2. Each member of the group should take one of the random samples

<table>
<thead>
<tr>
<th></th>
<th>Sample Mean, ( \bar{y} )</th>
<th>Sample Std. Dev., ( s )</th>
<th>95% Confidence Interval for ( \mu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td>( \bar{y} - t*\left(\frac{s}{\sqrt{n}}\right) ), ( \bar{y} + t*\left(\frac{s}{\sqrt{n}}\right) )</td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c) How many of the intervals contain \( \mu = 166.6 \text{ cm} \)?

d) The relative frequency of capturing the population mean for your lab section.

e) What should be the long run relative frequency of capturing the population mean?
f) Suppose you were to combine the four random samples of size \( n = 10 \) to create a random sample of size \( n = 40 \), would the confidence interval for the population mean height based on the sample of size \( n = 40 \) be wider, about the same width of narrower than a confidence interval for the population mean height based on the sample of size \( n = 10 \)? Explain your answer briefly.

3. Four random samples of size \( n = 10 \) were combined to create a random sample of size \( n = 40 \). For this larger random sample, the sample mean is \( \bar{y} = 164.45 \) cms and the sample standard deviation is \( s = 10.213 \) cms. A normal quantile plot, outlier box plot and histogram for the sample data appear on the back of this sheet.

   a) Comment on each of the plots and indicate how each supports or fails to support the condition that the distribution of female heights is nearly normal.

      • Normal quantile plot

      • Outlier box plot

      • Histogram

   b) Construct a 95% confidence interval for the population mean height of female students.

   c) Based on this interval could the population mean height of female students be 166.6 cms? Explain briefly.
d) Test the hypothesis that the population mean height of female students is 166.6 cms against the alternative that the population mean height is something different from 166.6. Go through the steps of a hypothesis test.

Step 1: Check conditions.

Step 2: Set up hypotheses.

Step 3: Calculate test statistic and P-value.

Step 4: Use the P-value to reach a decision.

Step 5: State a conclusion within the context of the problem.