SECTION 8.1 – Distribution of the Sample Mean

Sampling Distribution

The sampling distribution of a statistic is a probability distribution for all possible values of the statistic computed from a sample of size *n*.

Sampling Distribution of the Sample Mean

The **sampling distribution of the sample mean** $\bar{x}$ is the probability distribution of all possible values of the random variable *x* computed from a sample of size *n* from a population with mean μ and standard deviation σ.

The Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

Suppose that a simple random sample of size *n* is drawn from a population with mean μ and standard deviation σ.

- The sampling distribution of $\bar{x}$ has mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. The standard deviation of the sampling distribution of $\bar{x}$, $\sigma_{\bar{x}}$, is called the **standard error of the mean**.

The Shape of the Sampling Distribution of $\bar{x}$ If *X* Is Normal

If a random variable *X* is normally distributed, the distribution of the sample mean, $\bar{x}$, is normally distributed.

The Central Limit Theorem

Regardless of the shape of the underlying population, the sampling distribution of $\bar{x}$ becomes approximately normal as the sample size, *n*, increases.

Note:
The more skewed the distribution of the population is, the larger the sample size needed to invoke the Central Limit Theorem. We will err on the side of caution and say that, if the distribution of the population is unknown or not normal, then the distribution of the sample mean is approximately normal provided that the sample size is greater than or equal to 30.

The z-Score of a Sample Mean

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

Shape, center, and spread of the population	**Shape, center, and spread for the Distribution of the Sample Mean**
Population is **normal** with mean μ and standard deviation σ and **large** sample size ($n \geq 30$).	Regardless of sample size *n*, the shape of the sample distribution is **normal** with $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.
Population is **normal** with mean μ and standard deviation σ and **small** sample size ($n < 30$).	Regardless of sample size *n*, the shape of the sample distribution is **normal** with $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.
Population is **NOT normal** with mean μ and standard deviation σ and **large** sample size ($n \geq 30$).	The Central Limit Theorem is applied here. Since, the shape of the distribution is not normal, yet the sample size is large, the shape of the sample distribution is **approximately normal** with $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.
Population is **NOT normal** with mean μ and standard deviation σ and **small** sample size ($n < 30$).	Here we cannot conclude anything. We would have to use normal probability plots to assess any normality, if it exists. (Section 7.3).

SECTION 8.1 – Distribution of the Sample Mean

☺ **Exercises**:

1) The following table provides the proposed starting five players and their positions and heights for the 2018-2019 Los Angeles Lakers.

Player	Position	Height (in.)
Kentavious Caldwell-Pope (C)	Shooting Guard	77
Lonzo Ball (B)	Point Guard	78
LeBron James (J)	Small Forward	80
Brandon Ingram (I)	Power Forward	81
JaVale McGee (M)	Center	84

a) Compute the population mean, μ .

b) List all possible samples with size $n = 2$. There should be ${}_5C_2 = 10$ samples. (Use the first two columns in the table below).

c) Construct a sampling distribution for the mean by listing the sample means and their corresponding probabilities. (Use the third and fourth columns in the table below).

Samples	Heights	$\bar{x}$	$P(\bar{x})$

d) Construct a dotplot for the sampling distribution.

e) Compute the mean of the sampling distribution.

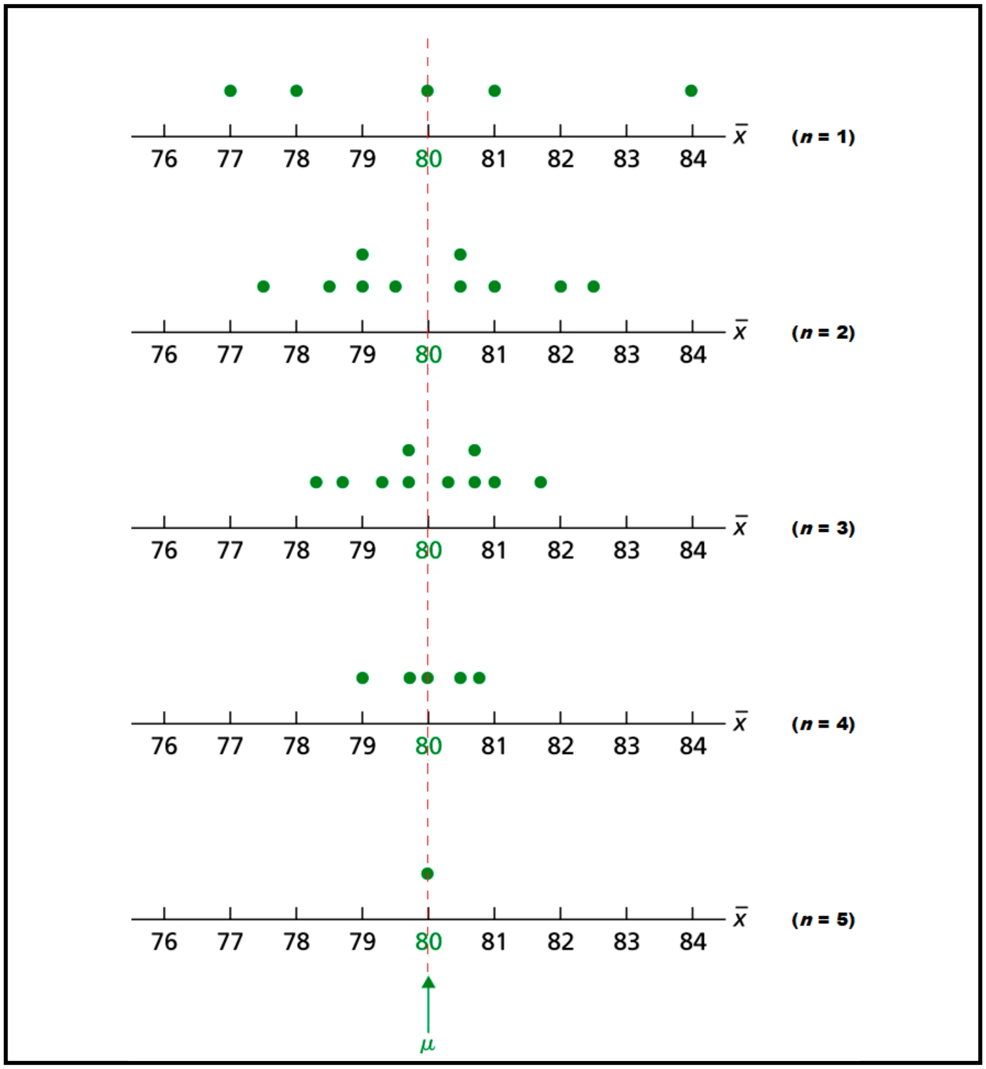
f) Compute the probability that the sample mean is within 1 inch of the population mean height.

g) Compute the probability that the sample mean is within 2 inches of the population mean height.

SECTION 8.1 – Distribution of the Sample Mean

☺ **Exercise (continued)**:

Dotplots for the sampling distributions of the sample mean for the heights of the five starting players for samples of sizes 1, 2, 3, 4, and 5.



Sample size and sampling error illustrations for the heights of the basketball players.

Sample size n	Number of possible samples	Number within 1" of μ	% within 1" of μ	Number within 0.5" of μ	% within 0.5" of μ
1	5	2	40%	1	20%
2	10	6	60%	3	30%
3	10	7	70%	3	30%
4	5	5	100%	3	60%
5	1	1	100%	1	100%

As stated earlier, the larger the sample size, the smaller the sampling error tends to be in estimating a population mean, μ, by a sample mean, $\bar{x}$.

SECTION 8.1 – Distribution of the Sample Mean

☺ **Exercises**:

2) A simple random sample of size $n = 36$ is obtained from a population with $\mu = 64$, $\sigma = 18$.

a) Describe the sampling distribution of $\overline{x}$.

b) What is $P(\overline{x} < 62.6)$?

c) What is $P(\overline{x} > 68.7)$?

d) What is $P(59.8 < \overline{x} < 65.9)$?

3) A simple random sample of size $n = 20$ is obtained from a population with $\mu = 64$, $\sigma = 17$.

a) What must be true regarding the distribution of the population in order to use the normal model to compute probabilities involving the sample mean? Assuming that this condition is true, describe the sampling distribution of $\overline{x}$.

b) Assuming the requirements described in part **(a)** are satisfied, determine $P(\overline{x} < 67.3)$.

c) Assuming the requirements described in part **(a)** are satisfied, determine $P(\overline{x} > 65.2)$.

SECTION 8.1 – Distribution of the Sample Mean

☺ **Exercises**:

4) **Old Faithful**. The most famous geyser in the world, Old Faithful in Yellowstone National Park, has a mean time between eruptions of 85 minutes. If the interval of time between eruptions is normally distributed with standard deviation 21.25 minutes, answer the following questions:
Source: www.unmuseum.org

a) What is the probability that a randomly selected time interval between eruptions is longer than 95 minutes?

b) What is the probability that a random sample of 20 time intervals between eruptions has a mean longer than 95 minutes?

c) What is the probability that a random sample of 30 time intervals between eruptions has a mean longer than 95 minutes?

d) What effect does increasing the sample size have on the probability? Provide an explanation for this result.

e) What might you conclude if a random sample of 30 time intervals between eruptions has a mean longer than 95 minutes?

f) On a certain day, suppose there are 22 time intervals for Old Faithful. Treating these 22 eruptions as a random sample, the likelihood the mean length of time between eruptions exceeds _______ minutes is 0.20.