8.1 Distribution of the Sample Mean

Definitions:

Sampling Distribution – a probability distribution of a statistic, such as the sample mean $\bar{x}$ or the sample proportion $\hat{p}$, for all possible values of the statistic computed from a sample size of $n$.

Sampling Distribution of the Sample Mean – a sampling distribution of the mean $\bar{x}$ of samples, all of size $n$, taken from a population with mean $\mu$.

Example: Suppose it is known that the mean score on an intelligence quotient test is 100 with a standard deviation of 15. Since this is a population, we have the parameters $\mu = 100$ and $\sigma = 15$. Suppose we obtain 1000 samples of size $n = 9$ and calculate the mean of each sample. Then we will have the data set $\{ \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{1000} \}$. [See Fig. 1 on p. 368.]

Shape. If we construct a frequency distribution of this set using a class width of 1, the histogram will be bell-shaped as shown in Fig. 2b on p. 369. The distribution is called a sampling distribution of the sample mean. If the population from which the samples are taken is normally distributed, then the sampling distribution will be normally distributed. That is, if the curve for $X$ is normal, then the curve for $\bar{x}$ is normal.

Center. If we were to calculate the mean of the data set, we would obtain a number close to the mean of the population $\mu$. We call this value the mean of the sampling distribution of $\bar{x}$, and give it the symbol $\mu_\bar{x}$, where

$$\mu_\bar{x} = \mu \approx \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_{1000}}{1000}$$

Spread. If we were to calculate the standard deviation of the data set, we would obtain a number close to the standard deviation of the population divided by the square root of the sample size, to wit, $\sigma / \sqrt{n}$. We call this value the standard de-
viation of the sampling distribution of \( \bar{x} \), or more tersely, the standard error of the mean, and give it the symbol \( \sigma_x \), where

\[
\sigma_x = \frac{\sigma}{\sqrt{n}} \approx \frac{1}{\sqrt{n}} \sum_{i=1}^{100} (\frac{\bar{x}_i - \bar{x}}{100})^2
\]

[See Fig. 4 on p. 370.] Thus, as \( n \) increases the standard deviation decreases, and the sampling curve begins to get narrow, because we are sampling more of the population and therefore expect the sampling distribution of \( \bar{x} \) to collapse on the one value \( \mu \). So as \( n \to N \), \( \bar{x} \to \mu \).

**Mean of the Sampling Distribution of \( \bar{x} \) –** for a simple random sample of size \( n \) drawn from a population with mean \( \mu \), the sampling distribution of \( \bar{x} \) will have mean

\[
\mu_x = \mu
\]

**Standard Deviation of the Sampling Distribution of \( \bar{x} \) –** also called standard error of the mean, for a simple random sample of size \( n \) drawn from a population with standard deviation \( \sigma \), the sampling distribution of \( \bar{x} \) will have standard deviation

\[
\sigma_x = \frac{\sigma}{\sqrt{n}}
\]

**Example:** To cut the standard error of the mean in half, by what factor must the sample size be increased?

**Solution:** The population standard deviation \( \sigma \) is constant. We want to choose a new sample size \( n_2 \) greater than the old sample size \( n_1 \), so that the new standard error \( \sigma_{x,2} \) is one half that of the old standard error \( \sigma_{x,1} \). So we have

\[
\sigma = \sigma_{x,2} \sqrt{n_2} = \sigma_{x,1} \sqrt{n_1} \quad \Rightarrow \quad \frac{\sigma_{x,2}}{\sigma_{x,1}} = \sqrt{\frac{n_1}{n_2}} = \frac{1}{2}
\]

\[
\Rightarrow \quad \frac{n_1}{n_2} = 4 \quad \Rightarrow \quad n_2 = 4n_1
\]
We would have to increase the sample size by a factor of 4 to reduce the standard error by a factor of one half.

**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. [See Figs. 8 – 10.] We will assume the sampling distribution of \( \bar{x} \) to be normal if \( n \geq 30 \).

**Shape of the Sampling Distribution of \( \bar{x} \)** – normal if population is normal or \( n \geq 30 \).

*Example:* 8.1.19, 25, 29

**Exercises:**

2. Determine the mean and standard deviation of the sampling distribution of the sample mean (9 – 12).
3. Understand the sampling distribution (13 – 14, 37).
4. Determine the probability of a sampling distribution of the sample mean (15 – 18).
5. Describe the distribution of the sample mean for samples taken from a population that is normal (19, 21).
6. Describe the distribution of the sample mean for samples taken from a population that is not normal (25, 27, 29).
8.2 Distribution of the Sample Proportion

Definitions:

**Sample Proportion** – for a random sample of size \( n \) in which there is a binomial response and \( x \) is the number of successes, the proportion of the sample indicated by \( \hat{p} \) is the ratio

\[
\hat{p} = \frac{x}{n}
\]

Note that it is always the case that \( 0 \leq \hat{p} \leq 1 \).

**Example**: In a town of 500 households, 220 have a dog. What is the proportion of dog owners?

**Solution**: We have \( n = 500 \) and \( x = 220 \), so

\[
\hat{p} = \frac{220}{500} = 0.44
\]

**Sampling Distribution of the Proportion** – a sampling distribution of the proportion \( \hat{p} \) of samples, all of size \( n \), taken from a population with proportion \( p \).

**Example**: In a Gallup poll, 76% of Americans believe that the state of moral values in the US is getting worse. We wish to verify this. First we choose 2000 samples of size \( n = 10 \). We calculate the sample proportion \( \hat{p} \) of each sample and construct a frequency distribution using a class width of 0.1. The resulting histogram is shown in Fig. 13(a) on p. 381, plotted alongside a normal curve centered at the population proportion \( p = 0.76 \). Note that the data is skewed left. We calculate a mean of 0.76 and a standard deviation of 0.136.

We repeat the experiment for \( n = 20 \). This time we construct a histogram using a class width of 0.04 and plot it alongside the same normal curve centered at \( p = 0.76 \), as shown in Fig. 13(b) on p. 381. The histogram is still skewed left but not as much as for the case \( n = 10 \). We calculate a mean of 0.76 and a standard deviation of 0.096.

Finally, we repeat the experiment for \( n = 60 \), and construct a histogram using a class width of
0.035. The resulting histogram is seen in Fig. 13(c), which is bell-shaped and fits the normal curve fairly well. We calculate a mean of 0.76 and a standard deviation of 0.054.

These three cases confirm that standard deviation varies as $1/\sqrt{n}$, since we have

$$0.136\sqrt{10} \approx 0.096\sqrt{20} \approx 0.054\sqrt{60}$$

This is consistent with the sampling distribution for the sample mean, where we found that $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. So likewise, $\sigma_{\hat{p}}$ is proportional to $1/\sqrt{n}$.

Recall that for the binomial distribution in Chapter 6, we had $\mu_X = np$ and $\sigma_X = \sqrt{np(1-p)}$. Since we are now talking about a proportion $x/n$, dividing the right-hand side of both equations by $n$ gives the mean and standard deviation of $\hat{p}$.

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

This is in fact the center and spread of the sampling distribution of $\hat{p}$. As for the shape, the sampling distribution of $\hat{p}$ is approximately normal if

$$np(1-p) \geq 10$$

$$n \leq 0.05N$$

The last condition indicates that we cannot sample more than 5% of the population. This must be true in order to satisfy the independence assumption. The success or failure of a trial – selecting someone with a particular characteristic – cannot depend on a previous result. Therefore, we require our sample size to be small relative to the population.

The characteristics of the sampling distributions for sample mean and sample proportion are summarized in the following table.
Chapter VIII  Sampling Distributions  8-6

<table>
<thead>
<tr>
<th>Sampling Distribution</th>
<th>$\bar{x}$</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape</strong></td>
<td>normal if: population normal or $n \geq 30$</td>
<td>normal if: $np(1-p) \geq 10$ and $n \leq 0.05N$</td>
</tr>
<tr>
<td><strong>Center</strong></td>
<td>$\mu_x = \mu$</td>
<td>$\mu_{\hat{p}} = p$</td>
</tr>
<tr>
<td><strong>Spread</strong></td>
<td>$\sigma_x = \frac{\sigma}{\sqrt{n}}$</td>
<td>$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$</td>
</tr>
</tbody>
</table>

*Example:* 8.2.7, 11, 17, 23

*Exercises:*

2. Describe the sampling distribution of a sample proportion (7 – 10).
4. Determine the sample size for a normal sampling distribution (23 – 24).