Chapter 18
Sampling Distribution Models

Data without concepts are blind, concepts without data are empty.

Immanuel Kant, (1724-1804)

“Data! Data! Data!” He cried impatiently. “I can't make bricks without clay!”

Sherlock Holmes

The rest of the course depends on the concepts in this chapter!!

Statistics vs Guesswork

Population parameter:

Sample statistic:

Definition (sampling distribution)
The sampling distribution of a sample statistic calculated from a sample of $n$ measurements is the probability distribution of values taken by the statistic in all possible samples of size $n$ from the same population.

Note: in the above definition the probability distribution is based on all possible samples of size $n$ that can be selected (with replacement, order counts) from the population. In some cases the sampling distribution can be determined exactly; in other cases it must be approximated by using the computer to draw some of the possible samples of size $n$ and drawing a histogram.

A Sampling Distribution Model for the Sample Proportion $\hat{p}$

EXAMPLE If a coin is fair the probability of a head on any toss of the coin is $p = 0.5$.

Imagine tossing this fair coin 5 times and calculating the proportion $\hat{p}$ of the 5 tosses that result in heads (note that $\hat{p} = \frac{x}{n}$, where $x$ is the number of heads in 5 tosses).

OBJECTIVE: Calculate the sampling distribution of $\hat{p}$, the proportion of heads in 5 tosses of a fair coin.

NOTE: the sampling distribution of $\hat{p}$ is the probability distribution of $\hat{p}$, where $\hat{p}$ is the proportion of heads in 5 tosses of a fair coin.

STEP 1:
The possible values of $\hat{p}$ are: $\frac{0}{5} = 0$, $\frac{1}{5} = .20$, $\frac{2}{5} = .40$, $\frac{3}{5} = .60$, $\frac{4}{5} = .80$, $\frac{5}{5} = 1$. 
STEP 2:
Use the binomial probabilities given below for n=5, p = 0.5, to fill in the probabilities for the values of \( \hat{p} \) in the table below the binomial probabilities.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.03125</td>
</tr>
<tr>
<td>1</td>
<td>0.15625</td>
</tr>
<tr>
<td>2</td>
<td>0.3125</td>
</tr>
<tr>
<td>3</td>
<td>0.3125</td>
</tr>
<tr>
<td>4</td>
<td>0.15625</td>
</tr>
<tr>
<td>5</td>
<td>0.03125</td>
</tr>
</tbody>
</table>

This table showing all the possible values of \( \hat{p} \) and their corresponding probabilities is the sampling distribution of \( \hat{p} \), the proportion of heads, in 5 tosses of a fair coin.

Use the probability distribution of \( \hat{p} \) to answer the following questions:

Question 1:
What is \( \hat{p} \)?

Question 2:
What is \( SD(\hat{p}) \), the standard deviation of \( \hat{p} \)? (compare your answer to \( \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(0.5)}{5}} = \frac{0.5}{\sqrt{5}} \approx 0.2236 \))

EXAMPLE Class Activity: sampling from the bins.
Expected Value and Standard Deviation of the Sampling Distribution of $\hat{p}$

**Properties of the sampling distribution of $\hat{p}$:**

1. $E(\hat{p}) = p$ (the expected value of the sampling dist. of $\hat{p} = \text{the “success” probability } p \text{ of the sampled population}$)

2. $SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$, where $p$ is the success probability of the “sampled” population and $n$ is the sample size.

⇒ This is why we use the sample proportion $\hat{p}$ to estimate the unknown value of the success probability $p$!!

**WHAT ABOUT THE SHAPE OF THE SAMPLING DISTRIBUTION OF $\hat{p}$?**

**In Addition:** The sampling distribution of $\hat{p}$ can be approximated by a normal model when the sample size $n$ is large enough. $n$ large enough means $np \geq 10$ and $n(1-p) \geq 10$.

**EXAMPLE** The frequency of color blindness (dyschromatopsia) in the Caucasian American male population is about 8%. We can't actually select all possible random samples of, say, $n = 1000$, Caucasian American males but IMAGINE the results of all possible simple random samples of $n = 1000$ Caucasian American males.

What would a histogram of the proportions of color blind males in each of those samples look like? Where do you expect the center of that histogram to be? We know that the center will be at the true proportion of the population (about .08) and we can call that $p$.

We don't have to just imagine. We can simulate. We want to simulate a bunch of those random samples of 1000 that we can't actually select. To do that, we'll simulate many independent random samples of size 1000, keeping the same probability $p = .08$ of success.

Below is a histogram of the proportions of Caucasian American males for 2000 independent samples, each of size $n = 1000$, when the true proportion is $p = .08$.

It should be no surprise that we don't get the same proportion for each sample that we select, even though the underlying true value $p = .08$ is the same for the population from which we've drawn the samples. Each $\hat{p}$ comes from a different independent sample.
Does it surprise you that the histogram is unimodal? Symmetric? That it is centered at $p = .08$? 

*It's an amazing and fortunate fact that a normal model is just the right one for the histogram of sample proportions.*

---

**The sampling distribution model for a proportion**

Provided that the sampled values are independent and the sample size $n$ is large enough, the sampling distribution of $\hat{p}$ is modeled by a normal distribution with $E(\hat{p}) = p$ and standard deviation $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$, where $q = 1 - p$.

That is, $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$

$n$ large enough means $np \geq 10$ and $nq \geq 10$

---

**EXAMPLE** (binge drinking by college students)

According to a study by the Harvard School of Public Health (H. Wechsler, G. W. Dowdall, A. Davenport, and W. DeJong, “Binge Drinking on Campus: Results of a National Study”) 44% of college students engage in binge drinking (for men, binge drinking is defined as having 5 or more drinks in a row, and for women as having 4 or more drinks in a row). A professor surveyed 244 students at her college and 36% of them admitted to binge drinking in the past week. Assume the 44% given in the study and compute the probability that in a sample of 244 students 36% or less have engaged in binge drinking.

**SOLUTION**

Let $\hat{p}$ be the proportion in a sample of 244 that engage in binge drinking.

We want to compute $P(\hat{p} \leq .36)$.

$$E(\hat{p}) = p = .44, \quad SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.44*.56}{244}} = .032$$

Since $np = 244*.44 = 107.36$ and $nq = 244*.56 = 136.64$ are both $\geq 10$, we can model the distribution of $\hat{p}$ with a normal distribution, so

$$\hat{p} \sim N(.44, .032)$$

and

$$P(\hat{p} \leq .36) = P\left(\frac{\hat{p} - .44}{.032} \leq \frac{.36 - .44}{.032}\right) = P(z \leq -2.5) = .5 - .4938 = .0062$$

END OF EXAMPLE
EXAMPLE Professor Stickler has a large statistics class of over 300 students. He asked them the ages of their cars and obtained the following probability distribution:

<table>
<thead>
<tr>
<th>Age (x)</th>
<th>Probability (p(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\frac{1}{11}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{11}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{2}{11}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{2}{11}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{2}{11}$</td>
</tr>
<tr>
<td>7</td>
<td>$\frac{3}{11}$</td>
</tr>
<tr>
<td>8</td>
<td>$\frac{3}{11}$</td>
</tr>
</tbody>
</table>

Questions:
1) Find the sampling distribution of the sample mean $\bar{x}$ for samples of size $n = 2$.
2) What is the probability of obtaining a sample mean between 4 and 6 years, inclusive, that is, what is $P(4 \leq \bar{x} \leq 6)$?

SOLUTION There are $7^2 = 49$ possible sample of size $n = 2$. Each of the 49 possible samples is listed below with the corresponding sample mean for that particular sample and the probability of each sample.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\bar{x}$</th>
<th>Probability</th>
<th>Sample</th>
<th>$\bar{x}$</th>
<th>Probability</th>
<th>Sample</th>
<th>$\bar{x}$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 2</td>
<td>2</td>
<td>$\frac{1}{196}$</td>
<td>6, 8</td>
<td>7</td>
<td>$\frac{6}{196}$</td>
<td>5, 3</td>
<td>4</td>
<td>$\frac{2}{196}$</td>
</tr>
<tr>
<td>2, 4</td>
<td>3</td>
<td>$\frac{2}{196}$</td>
<td>6, 5</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
<td>5, 7</td>
<td>6</td>
<td>$\frac{6}{196}$</td>
</tr>
<tr>
<td>2, 6</td>
<td>4</td>
<td>$\frac{2}{196}$</td>
<td>6, 3</td>
<td>4.5</td>
<td>$\frac{4}{196}$</td>
<td>3, 2</td>
<td>2.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>2, 8</td>
<td>5</td>
<td>$\frac{3}{196}$</td>
<td>6, 7</td>
<td>6.5</td>
<td>$\frac{2}{196}$</td>
<td>3, 4</td>
<td>3.5</td>
<td>$\frac{2}{196}$</td>
</tr>
<tr>
<td>2, 5</td>
<td>3.5</td>
<td>$\frac{2}{196}$</td>
<td>8, 2</td>
<td>5</td>
<td>$\frac{2}{196}$</td>
<td>3, 6</td>
<td>4.5</td>
<td>$\frac{2}{196}$</td>
</tr>
<tr>
<td>2, 3</td>
<td>2.5</td>
<td>$\frac{1}{196}$</td>
<td>8, 4</td>
<td>6</td>
<td>$\frac{1}{196}$</td>
<td>3, 8</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>2, 7</td>
<td>4.5</td>
<td>$\frac{3}{196}$</td>
<td>8, 6</td>
<td>7</td>
<td>$\frac{3}{196}$</td>
<td>3, 5</td>
<td>4</td>
<td>$\frac{2}{196}$</td>
</tr>
<tr>
<td>4, 2</td>
<td>3</td>
<td>$\frac{2}{196}$</td>
<td>8, 8</td>
<td>8</td>
<td>$\frac{2}{196}$</td>
<td>3, 3</td>
<td>3</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 4</td>
<td>4</td>
<td>$\frac{1}{196}$</td>
<td>8, 5</td>
<td>6.5</td>
<td>$\frac{1}{196}$</td>
<td>3, 7</td>
<td>5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 6</td>
<td>5</td>
<td>$\frac{1}{196}$</td>
<td>8, 3</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 2</td>
<td>4.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 8</td>
<td>6</td>
<td>$\frac{1}{196}$</td>
<td>8, 7</td>
<td>7.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 4</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 5</td>
<td>4.5</td>
<td>$\frac{1}{196}$</td>
<td>5, 2</td>
<td>3.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 6</td>
<td>6.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 3</td>
<td>3.5</td>
<td>$\frac{1}{196}$</td>
<td>5, 4</td>
<td>4.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 8</td>
<td>7.5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>4, 7</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
<td>5, 6</td>
<td>5.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 5</td>
<td>6</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>6, 4</td>
<td>4</td>
<td>$\frac{2}{196}$</td>
<td>5, 8</td>
<td>6.5</td>
<td>$\frac{1}{196}$</td>
<td>7, 3</td>
<td>5</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>6, 6</td>
<td>6</td>
<td>$\frac{1}{196}$</td>
<td>5, 5</td>
<td>5</td>
<td>$\frac{1}{196}$</td>
<td>7, 7</td>
<td>7</td>
<td>$\frac{1}{196}$</td>
</tr>
</tbody>
</table>

Question 1:
From the table above we see that $\bar{x}$ can assume the values 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8.
$\bar{x} = 2$ occurs in only one sample (2,22), so $P(\bar{x} = 2) = \frac{1}{11} \times \frac{1}{11} = \frac{1}{121}$; similarly, $\bar{x} = 2.5$ occurs in 2 samples: (2, 3) and (3,2); therefore, $P(\bar{x} = 2.5) = \frac{1}{11} + \frac{1}{11} = \frac{2}{116}$. The complete probability distribution for $\bar{x}$ is shown in the table below:

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>Probability (p(\bar{x}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\frac{1}{196}$</td>
</tr>
<tr>
<td>2.5</td>
<td>$\frac{2}{196}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{3}{196}$</td>
</tr>
<tr>
<td>3.5</td>
<td>$\frac{4}{196}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{5}{196}$</td>
</tr>
<tr>
<td>4.5</td>
<td>$\frac{6}{196}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{7}{196}$</td>
</tr>
<tr>
<td>5.5</td>
<td>$\frac{8}{196}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{9}{196}$</td>
</tr>
<tr>
<td>6.5</td>
<td>$\frac{10}{196}$</td>
</tr>
<tr>
<td>7</td>
<td>$\frac{11}{196}$</td>
</tr>
<tr>
<td>7.5</td>
<td>$\frac{12}{196}$</td>
</tr>
<tr>
<td>8</td>
<td>$\frac{13}{196}$</td>
</tr>
</tbody>
</table>

This is the sampling distribution of $\bar{x}$ because it specifies the probability associated with each possible value of $\bar{x}$.

Question 2:
from the sampling distribution for the sample mean in the table above,
$P(4 \leq \bar{x} \leq 6) = p(4) + p(4.5) + p(5) + p(5.5) + p(6) = \frac{12}{196} + \frac{18}{196} + \frac{24}{196} + \frac{26}{196} + \frac{28}{196} = \frac{108}{196}$. 

Expected Value and Standard Deviation of the Sampling Distribution of \( \bar{X} \)

**EXAMPLE**  Recall the example on p. 36 where we found the sampling distribution of \( \pi \) from samples of size \( n = 2 \) from the population with probability distribution

\[
\begin{array}{c|cccccccc}
 x \text{(Age)} & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
p(x) & \frac{1}{11} & \frac{1}{11} & \frac{1}{11} & \frac{2}{11} & \frac{2}{11} & \frac{3}{11} & \frac{1}{11} \\
\end{array}
\]

Note that

\[
\mu = E(x) = \frac{1}{11}(2) + \frac{1}{11}(3) + \frac{2}{11}(4) + \frac{2}{11}(5) + \frac{2}{11}(6) + \frac{3}{11}(7) + \frac{1}{11}(8) = 5.714
\]

and

\[
\sigma^2 = Var(x) = E[(x - 5.714)^2] = (2 - 5.714)^2(\frac{1}{11}) + (3 - 5.714)^2(\frac{1}{11}) + ... + (8 - 5.714)^2(\frac{1}{11}) = 3.4898
\]

In the example we found the sampling distribution of \( \pi \) to be

\[
\begin{array}{c|cccccccc}
 \pi & 2 & 2.5 & 3 & 3.5 & 4 & 4.5 & 5 & 5.5 & 6 & 6.5 & 7 & 7.5 & 8 \\
p(\pi) & \frac{1}{196} & \frac{2}{196} & \frac{5}{196} & \frac{8}{196} & \frac{12}{196} & \frac{18}{196} & \frac{24}{196} & \frac{24}{196} & \frac{28}{196} & \frac{24}{196} & \frac{21}{196} & \frac{18}{196} & \frac{9}{196} \\
\end{array}
\]

\[
E(\pi) = \mu = 5.714 \quad \text{and} \quad Var(\pi) = \frac{\sigma^2}{n} = 1.7449
\]

**Properties of the sampling distribution of \( \bar{X} \):**

1. \( E(\bar{X}) = \mu \) (the mean of the sampling distribution of \( \bar{X} \) is equal to the mean \( \mu \) of the population from which the sample is selected)

2. \( SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} \), where \( \sigma \) is the standard deviation of the population from which the sample is selected and \( n \) is the sample size.

\( \Rightarrow \) This is why we use the sample mean \( \bar{X} \) to estimate the unknown value of the population mean \( \mu \)!!

**WHAT ABOUT THE SHAPE OF THE SAMPLING DISTRIBUTION OF \( \bar{X} \)?**

**In Addition:** The sampling distribution of \( \bar{X} \) is approximately normal for large sample sizes (the Central Limit Theorem)
"The World is Normal" Theorem

The Central Limit Theorem

**IMPORTANT**

Sampling from normal populations

- If the population is normally distributed, then the sampling distribution of $\bar{x}$ is normally distributed for any sample size $n$.

Sampling from normal populations that are NOT normal

What can we say about the shape of the sampling distribution of $x$ when the population from which the sample is selected is not normal?

**The Central Limit Theorem** (for the sample mean $\bar{x}$)

The mean $\mu$ of a random sample has a sampling distribution whose shape can be approximated by a Normal model. The larger the sample, the better the approximation will be.

**The importance of the Central Limit Theorem:**

When we select simple random samples of size $n$, the sample means we find will vary from sample to sample. We can model the distribution of these sample means with a probability model that is

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

This normal distribution is the sampling distribution model for the sample mean $\bar{x}$.

**Note:** for the purpose of applying the central limit theorem for the sample mean $\bar{x}$, we will consider a sample size to be large when $n \geq 30$. 

Histogram of 260 sample mean salaries ($n = 50$)
IMPORTANT REMARK: it is crucial to realize that, in practice, we select only one sample at any one time and therefore observe only one of the possible values of the sample mean.

The Central Limit Theorem (for the sample proportion $\hat{p}$)
If a random sample of $n$ observations is selected from a population (any population), and $x$ “successes” are observed, then when $n$ is sufficiently large, the sampling distribution of the sample proportion $\hat{p} = \frac{x}{n}$ can be approximated by a normal model.

The importance of the Central Limit Theorem:
When we select simple random samples of size $n$, the sample proportions $\hat{p} = \frac{x}{n}$ that we obtain will vary from sample to sample. We can model the distribution of these sample proportions with a probability model that is

$$N \left( p, \sqrt{\frac{p(1-p)}{n}} \right).$$

This normal distribution is the sampling distribution model for the sample proportion $\hat{p}$.

Note: for the purpose of applying the central limit theorem for the sample proportion $\hat{p}$, we will consider a sample size $n$ to be large when $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

EXAMPLE: A random sample of $n = 64$ observations is drawn from a population with mean $\mu = 15$ and $\sigma = 4$.

a. Give the mean $E(\bar{x})$ and the standard deviation $SD(\bar{x})$ of the sampling distribution of $\bar{x}$.

b. Describe the shape of the sampling distribution model for $\bar{x}$. Does your answer depend on the sample size?

c. Calculate the standard normal $z$-score corresponding to a value of $\bar{x} = 15.5$.

ANSWER:

a. $E(\bar{x}) = \mu = 15; SD(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{8} = .5$.

b. The shape of the sampling distribution of $\bar{x}$ is approximately normal (by the central limit theorem) with mean $E(\bar{x}) = 15$ and standard deviation $SD(\bar{x}) = .5$. The answer depends on the sample size since $SD(\bar{x}) = \sigma/\sqrt{n}$.

c. $z = \frac{\bar{x} - E(\bar{x})}{SD(\bar{x})} = \frac{15.5 - 15}{.5} = 1$. □
Distribution of $\pi$ for various sample sizes from various populations

A uniform probability distribution

A skewed probability distribution
EXAMPLE: (Last digit of telephone numbers)

Population probability distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>p(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.1</td>
</tr>
<tr>
<td>1</td>
<td>.1</td>
</tr>
<tr>
<td>2</td>
<td>.1</td>
</tr>
<tr>
<td>3</td>
<td>.1</td>
</tr>
<tr>
<td>4</td>
<td>.1</td>
</tr>
<tr>
<td>5</td>
<td>.1</td>
</tr>
<tr>
<td>6</td>
<td>.1</td>
</tr>
<tr>
<td>7</td>
<td>.1</td>
</tr>
<tr>
<td>8</td>
<td>.1</td>
</tr>
<tr>
<td>9</td>
<td>.1</td>
</tr>
</tbody>
</table>

Population parameters: $\mu = 4.5, \sigma = 2.872$;
$n = 10, n = 20, n = 30$; 100 random samples for each value of the sample size $n$.

$n = 10, \frac{\sigma}{\sqrt{n}} = \frac{2.872}{\sqrt{10}} = .91$

$n = 20, \frac{\sigma}{\sqrt{n}} = \frac{2.872}{\sqrt{20}} = .64$

$n = 30, \frac{\sigma}{\sqrt{n}} = \frac{2.872}{\sqrt{30}} = .52$

Mean of 100 means: 4.509 s = .707 min = 2.8 max = 6.05

Mean of 100 means: 4.547 s = 1.003 min = 2.2 max = 7

Mean of 100 means: 4.525 s = .583 min = 3.17 max = 5.73
EXAMPLE: Consider a population that contains values of $x$ equal to 00, 01, 02, …, 97, 98, 99. Assume that the values of $x$ are equally likely. For sample sizes $n = 20, 30, 40, 50$, generate 100 random samples and calculate $\bar{x}$ for each sample. For each sample size construct a relative frequency histogram of the 100 values of $\bar{x}$. Note the decrease in the standard deviation of the $\bar{x}$'s as the sample size $n$ increases.

Population probability distribution:

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>…</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x)$</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>…</td>
<td>.01</td>
</tr>
</tbody>
</table>

Population parameters: $\mu = 49.5; \sigma = 28.87$

\[ n = 20, \quad \frac{\sigma}{\sqrt{n}} = \frac{28.87}{\sqrt{20}} = 6.46 \]

\[ n = 30, \quad \frac{\sigma}{\sqrt{n}} = \frac{28.87}{\sqrt{30}} = 5.27 \]

\[ n = 40, \quad \frac{\sigma}{\sqrt{n}} = \frac{28.87}{\sqrt{40}} = 4.56 \]

\[ n = 50, \quad \frac{\sigma}{\sqrt{n}} = \frac{28.87}{\sqrt{50}} = 4.08 \]
EXAMPLES

EXAMPLE 2:
The probability distribution of 6-month incomes of junior account executives \( X \) has mean $20,000 and standard deviation $5,000.

a) A single executive’s income is $20,000. Can it be said that this executive’s income exceeds 50% of all account executive incomes?

b) 64 account executives (that is, \( n = 64 \)) are randomly selected. What is the probability that the sample mean exceeds $20,500?

c) Why do we need to use the Central Limit Theorem to solve part b)?

ANSWER:
a) Distribution of \( X \)?

b) \( \mu = 20,000, \sigma = 5,000, n = 64 \). \( E(\bar{X}) = SD(\bar{X}) = \) By the Central Limit Theorem, \( \bar{X} \sim \)

\[ P(\bar{X} > 20,500) = \]

EXAMPLE 3:
A sample of size \( n=16 \) is drawn from a normally distributed population \( X \) with mean \( \mu = 20 \) and \( \sigma = 8 \).

a) What is the probability that the sample mean is at least 24?

b) What is the probability that the sample mean is between 16 and 24, inclusive?

c) Do we need the Central Limit Theorem to solve a) or b)? Why or why not?

ANSWER: \( X \sim N(20, 8); \ \bar{X} \sim ? \)

a)

b)

c)

EXAMPLE 4:
The life \( X \) of a particular brand flashlight battery is normally distributed with mean 20 hours and standard deviation 10 hours. The manufacturer has a money-back guarantee that the average battery life in a case of 24 batteries exceeds 16 hours. What is the probability that a 24-battery case satisfies the guarantee?

ANSWER: \( \mu = 20 \text{ hours}, \sigma = 10 \text{ hours} \)

EXAMPLE 5:
Cans of salmon are supposed to have a net weight of 6 oz. The cannery owner claims that the net weight is a random variable with mean \( \mu = 6.05 \) oz. and stand. dev. \( \sigma = 0.18 \) oz. As an inspector for the U.S. Food and Drug Administration, your job is to make sure that the cans are filled with an adequate amount of salmon. You select a random sample of 36 cans from the cannery's production line.

a) Find the probability that the mean weight of the sample of 36 cans is less than 5.97 oz.
b) Based on your answer in part a), what would you conclude about the owner’s claim that the mean fill is 
\( \mu = 6.05 \text{ oz.} \)?

EXAMPLE 6:
A business school claims that 1 year after graduation the average weekly income of its graduates is $600.

a) Suppose the distribution of weekly income has a standard deviation of $100. What is the probability that 25 randomly selected graduates have an average weekly income of less than $550?

b) If a random sample of 25 graduates actually had an average weekly income of $550, what would you conclude about the validity of the claim that the average weekly income is $600?

ANSWER: 
\( X = \text{weekly income} \quad \mu = 600, \quad \sigma = 100 \quad X \sim ? \)

EXAMPLE 7:
12% of students at NCSU are left-handed. What is the probability that in a sample of 50 students, the sample proportion that are left-handed is less than 11%?

ANSWER: 
\( E(\hat{p}) = p = .12; SD(\hat{p}) = \sqrt{\frac{.12 \times .88}{50}} = .046 \); by the CLT the distribution of \( \hat{p} \) can be modeled with a normal distribution, so \( \hat{p} \sim N(.12, .046) \).
\( P(\hat{p} < .11) = P(\frac{\hat{p} - .12}{.046} < \frac{.11 - .12}{.046}) = P(z < - .22) = .4129. \)