Chapter 8: Sampling Distributions

El Mechry El Koudous

Fordham University

April 30, 2019

Overview

- In Chapters 6 and 7, we had access to the population mean and standard deviation, and we made probability statements about individual x values taken from the population.
- In this chapter, we will be taking many samples from a population, and each sample will have its own mean $\bar{x}$.
- Every time we take a sample from the population, we will get a different sample mean.
- We are going to regard the sample mean itself as a random variable $\bar{X} = \{\bar{x_1}, \bar{x_2}, \bar{x_3}, ..., \bar{x_n}\}$ with its own mean and standard deviation.
- The resulting probability distribution of the sample means is called the ***sampling distribution of the mean***.

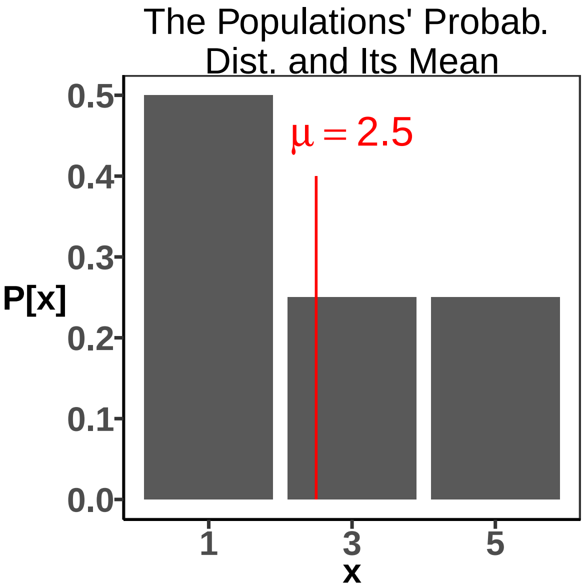
The Sampling Dist. of the Mean: Example

Population: Suppose our population has four people, and let x be number of bottles of Diet Pepsi in their refrigerator. Lets consider all possible simple random samples of $n = 2$.

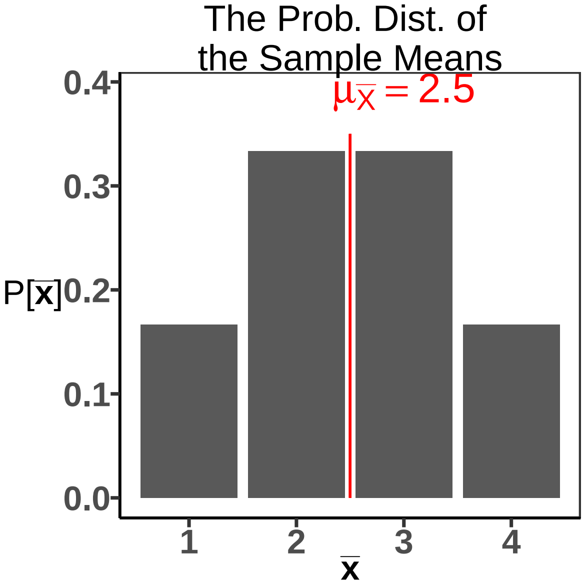
Person	x
Bill	1
Carl	1
Denise	3
Ed	5
	$\mu = 2.5$

Sample	Mean of this Sample	Probability of Selecting this Sample
Bill and Carl	$\bar{x} = 1$	1/6
Bill and Denise	$\bar{x} = 2$	1/6
Bill and Ed	$\bar{x} = 3$	1/6
Carl and Denise	$\bar{x} = 2$	1/6
Carl and Ed	$\bar{x} = 3$	1/6
Denise and Ed	$\bar{x} = 4$	1/6
	$\mu_{\bar{X}} = 2.5$	

The Sampling Dist. of the Mean: Example



The Sampling Dist. of the Mean: Example



Properties of the Sampling Dist. of the Mean

- The mean of the sampling distribution and its standard deviation can be calculated in the same way as the mean and standard deviation of a random variable.
- If the original population is normally distributed, the sampling distribution of the mean will also be normally distributed.
- The sampling distribution of the mean, denoted $\mathbb{E}[\bar{X}]$ or $\mu_{\bar{X}}$, will be related to the mean of the original population from which the samples were drawn μ_X
- The standard deviation of the sampling distribution of the mean is referred to as the standard error of the mean and is denoted $\sigma_{\bar{X}}$. It's also related to σ_X.

$\mathbb{E}[\bar{X}]$ and $\sigma_{\bar{X}}$

From the previous example $\bar{X} = \{1, 2, 3, 4\}$ with probabilities $P[\bar{X}_i] = \{1/6, 2/6, 2/6, 1/6\}$, so we can calculate $\mathbb{E}[\bar{X}]$ and $\sigma_{\bar{X}}$ the usual way.

$\bar{X}$	$P[\bar{X}]$	$P[\bar{X}_i] \times \bar{X}_i$	$P[\bar{X}_i] \times (\bar{X}_i - \mathbb{E}[\bar{X}])^2$
1	0.1667	0.1667	0.3750
2	0.3333	0.6667	0.0833
3	0.3333	1.0000	0.0833
4	0.1667	0.6667	0.3750
		$\mathbb{E}[\bar{X}] = 2.5$	$\sigma^2_{\bar{X}} = 0.9167$

The Sampling Dist. of the Mean: Example

Population: Suppose our population is five dogs, and let x be their weights. Lets consider all possible simple random samples of $n = 2$.

Dog	x
A	42
B	48
C	52
D	58
E	60

$\mu = 52$
$\sigma^2 = 43.2$

Sample	Weights	$\bar{X}$	$P[\bar{X}]$	$P[\bar{X}_i] \times \bar{X}_i$	...
A,B	42 , 48	45	1/10	4.5	4.9
A,C	42 , 52	47	1/10	4.7	5.2
A,D	42 , 58	50	1/10	5.0	0.4
A,E	42 , 60	51	1/10	5.1	0.1
B,C	48 , 52	50	1/10	5.0	0.4
B,D	48 , 58	53	1/10	5.3	0.1
B,E	48 , 60	54	1/10	5.4	0.4
C,D	52 , 58	55	1/10	5.5	0.9
C,E	52 , 60	56	1/10	5.6	1.6
D,E	58 , 60	59	1/10	5.9	0.9
				$\mu_{\bar{X}} = 52$	$\sigma^2 = 16.2$

Properties of the Sampling Dist. of the Mean

Central Limit Theorem: CLT

For large, simple random samples from a population that is ***not*** normally distributed, the sampling distribution of the mean will be approximately normal. As the sample size n is increased, the sampling distribution of the mean will more closely approach the normal distribution.

The central limit theorem is basic to the concept of statistical inference because it permits us to draw conclusions about the population based strictly on sample data, and without having any knowledge about the distribution of the underlying population.

Properties of the Sampling Dist. of the Mean

Suppose X represents a population with mean, μ_X, and standard deviation, σ_X. Also suppose we draw great many simple random samples of size n from this population, and let $\bar{X}$ be the sampling distribution of the mean, then:

- The mean of $\bar{X}$, usually denoted $\mu_{\bar{X}}$, is

$$\mu_{\bar{X}} = \mathbb{E}[\bar{X}] = \mu_X$$

- The standard deviation of $\bar{X}$, usually called the standard error and denoted $\sigma_{\bar{X}}$ is:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

- $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X}{\sqrt{n}})$ if the original population is normally distributed, or if the sample size $n \geq 30$ regardless of the distribution of the original population.

Sampling Dist.of the Mean: Role of $\sigma_{\bar{X}}$

Let $X \sim \mathcal{N}(120, 30)$. Suppose we take many samples from X with sizes: $n = 9$ and $n = 36$.What is the distribution of $\bar{X}$?
Answer:

(a) $n = 9$; $\mu_{\bar{X}} =$

Sampling Dist.of the Mean: Role of $\sigma_{\bar{X}}$

Let $X \sim \mathcal{N}(120, 30)$. Suppose we take many samples from X with sizes: $n = 9$ and $n = 36$.What is the distribution of $\bar{X}$?
Answer:

(a) $n = 9$; $\mu_{\bar{X}} = \mu_X = 120$; $\sigma_{\bar{X}} =$

Sampling Dist.of the Mean: Role of $\sigma_{\bar{X}}$

Let $X \sim \mathcal{N}(120, 30)$. Suppose we take many samples from X with sizes: $n = 9$ and $n = 36$.What is the distribution of $\bar{X}$?

Answer:

(a) $n = 9$; $\mu_{\bar{X}} = \mu_X = 120$; $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{30}{\sqrt{9}} = 10$;$\sigma_{\bar{X}} \sim \mathcal{N}(120, 10)$

(b) $n = 36$; $\mu_{\bar{X}} =$

Sampling Dist.of the Mean: Role of $\sigma_{\bar{X}}$

Let $X \sim \mathcal{N}(120, 30)$. Suppose we take many samples from X with sizes: $n = 9$ and $n = 36$.What is the distribution of $\bar{X}$?
Answer:

(a) $n = 9$; $\mu_{\bar{X}} = \mu_X = 120$; $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{30}{\sqrt{9}} = 10$;$\sigma_{\bar{X}} \sim \mathcal{N}(120, 10)$

(b) $n = 36$; $\mu_{\bar{X}} = \mu_X = 120$ and $\sigma_{\bar{X}} =$

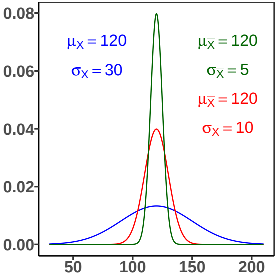
Sampling Dist.of the Mean: Role of $\sigma_{\bar{X}}$

Let $X \sim \mathcal{N}(120, 30)$. Suppose we take many samples from X with sizes: $n = 9$ and $n = 36$.What is the distribution of $\bar{X}$?

Answer:

(a) $n = 9$; $\mu_{\bar{X}} = \mu_X = 120$; $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{30}{\sqrt{9}} = 10$;$\sigma_{\bar{X}} \sim \mathcal{N}(120, 10)$

(b) $n = 36$; $\mu_{\bar{X}} = \mu_X = 120$ and $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{30}{\sqrt{36}} = 5$; $\sigma_{\bar{X}} \sim \mathcal{N}(120, 5)$



Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$P[\bar{X} > 128] =$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$P[\bar{X} > 128] = P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}]$$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$\begin{aligned} P[\bar{X} > 128] &= P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\ &= P[z > \frac{128 - 120}{\frac{30}{\sqrt{36}}}] \end{aligned}$$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$
\begin{aligned}
P[\bar{X} > 128] &= P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\
&= P[z > \frac{128 - 120}{\frac{30}{\sqrt{36}}}] \\
&= P[z > \frac{8}{5}]
\end{aligned}
$$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$P[\bar{X} > 128] = P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}]$$

$$= P[z > \frac{128 - 120}{\frac{30}{\sqrt{36}}}]$$

$$= P[z > \frac{8}{5}] = 1 - P[z < \frac{8}{5}]$$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$\begin{aligned} P[\bar{X} > 128] &= P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\ &= P[z > \frac{128 - 120}{\frac{30}{\sqrt{36}}}] \\ &= P[z > \frac{8}{5}] = 1 - P[z < \frac{8}{5}] = 1 - P[z < 1.6] \end{aligned}$$

Sampling Dist.of the Mean: Example 1

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 36 aircraft, what is the probability that the average flying time for the aircraft in the sample was at least 128 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{36}} = 5)$, and we are looking for $P[\bar{X} > 128]$. Recall that: $z_i = \frac{x_i - \mu}{\sigma}$ So:

$$
\begin{aligned}
P[\bar{X} > 128] &= P[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{128 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\
&= P[z > \frac{128 - 120}{\frac{30}{\sqrt{36}}}] \\
&= P[z > \frac{8}{5}] = 1 - P[z < \frac{8}{5}] = 1 - P[z < 1.6] = 1 - 0.9452 = 0.0548
\end{aligned}
$$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

$$P[111 < \bar{X} < 114] = P[\frac{111 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}]$$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

$$
\begin{aligned}
P[111 < \bar{X} < 114] &= P[\frac{111 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\
&= P[\frac{111 - 120}{\frac{30}{\sqrt{100}}} < z < \frac{114 - 120}{\frac{30}{\sqrt{100}}}]
\end{aligned}
$$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

$$\begin{aligned} P[111 < \bar{X} < 114] &= P[\frac{111 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\ &= P[\frac{111 - 120}{\frac{30}{\sqrt{100}}} < z < \frac{114 - 120}{\frac{30}{\sqrt{100}}}] \\ &= P[-3 < z < -2] \end{aligned}$$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

$$\begin{aligned} P[111 < \bar{X} < 114] &= P[\frac{111 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\ &= P[\frac{111 - 120}{\frac{30}{\sqrt{100}}} < z < \frac{114 - 120}{\frac{30}{\sqrt{100}}}] \\ &= P[-3 < z < -2] \\ &= P[z < -2] - P[z < -3] \end{aligned}$$

Sampling Dist.of the Mean: Example 2

Suppose flying time of aircraft is normally distributed with mean 120 and standard deviation 30. For a simple random sample of 100 aircraft, what is the probability that the average flying time for the aircraft in the sample was between 111 and 114 hours?
Answer: $X \sim \mathcal{N}(120, 30)$ so, $\bar{X} \sim \mathcal{N}(120, \frac{30}{\sqrt{100}} = 3)$, and we are looking for $P[111 < \bar{X} < 114]$. So:

$$\begin{aligned}
P[111 < \bar{X} < 114] &= P[\frac{111 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}] \\
&= P[\frac{111 - 120}{\frac{30}{\sqrt{100}}} < z < \frac{114 - 120}{\frac{30}{\sqrt{100}}}] \\
&= P[-3 < z < -2] \\
&= P[z < -2] - P[z < -3] = 0.0228 - 0.0013 = 0.0215
\end{aligned}$$

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim$

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] =$

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] = 0.0668$
- Between 78 and 82 pounds?

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] = 0.0668$
- Between 78 and 82 pounds?
  $P[78 < \bar{X} < 82] =$

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] = 0.0668$
- Between 78 and 82 pounds?
  $P[78 < \bar{X} < 82] = 0.4514$
- Above 83 pounds?

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] = 0.0668$
- Between 78 and 82 pounds?
  $P[78 < \bar{X} < 82] = 0.4514$
- Above 83 pounds?
  $P[\bar{X} > 83] =$

Sampling Dist.of the Mean: Example 3

Assume that a school district has 10,000 6^{th} graders. In this district, the average weight of a 6^{th} grader is 80 pounds, with a standard deviation of 20 pounds. Suppose you draw a random sample of 36 students. What is the probability that the average weight of a sampled student will be:

- Less than 75 pounds?
  $\bar{X} \sim \mathcal{N}(80, \frac{20}{\sqrt{36}} = 3.3333)$, so $P[\bar{X} < 75] = 0.0668$
- Between 78 and 82 pounds?
  $P[78 < \bar{X} < 82] = 0.4514$
- Above 83 pounds?
  $P[\bar{X} > 83] = 0.1841$

Sampling Dist.of the Mean: Example 4

ACT scores are normally distributed with mean of 20 and standard deviation of 5.

(a) What is the probability that a single student randomly selected will score 21 or higher?
$P[X > 21] =$

Sampling Dist.of the Mean: Example 4

ACT scores are normally distributed with mean of 20 and standard deviation of 5.

(a) What is the probability that a single student randomly selected will score 21 or higher?
$P[X > 21] = 0.4207$

(b) A random sample of 25 students is obtained. What is the probability that the mean score for these 25 students will be 21 or higher?
$\bar{X} \sim \mathcal{N}(20, \frac{5}{\sqrt{25}} = 1)$, so $P[\bar{X} > 21] =$

Sampling Dist.of the Mean: Example 4

ACT scores are normally distributed with mean of 20 and standard deviation of 5.

(a) What is the probability that a single student randomly selected will score 21 or higher?
$P[X > 21] = 0.4207$

(b) A random sample of 25 students is obtained. What is the probability that the mean score for these 25 students will be 21 or higher?
$\bar{X} \sim \mathcal{N}(20, \frac{5}{\sqrt{25}} = 1)$, so $P[\bar{X} > 21] = 0.1587$

(c) What if the normal distribution assumption was not given? How would your answer to part (b) change? Explain.

Sampling Dist.of the Mean: Example 4

ACT scores are normally distributed with mean of 20 and standard deviation of 5.

(a) What is the probability that a single student randomly selected will score 21 or higher?
$P[X > 21] = 0.4207$

(b) A random sample of 25 students is obtained. What is the probability that the mean score for these 25 students will be 21 or higher?
$\bar{X} \sim \mathcal{N}(20, \frac{5}{\sqrt{25}} = 1)$, so $P[\bar{X} > 21] = 0.1587$

(c) What if the normal distribution assumption was not given? How would your answer to part (b) change? Explain.
Since the original population is not normally distributed and the sample size is less than 30, we don't know the distribution of $\bar{X}$ and thus can't answer the question.

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?
$\bar{X} \sim \mathcal{N}(60, \frac{14}{\sqrt{49}} = 2)$, so

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?
$\bar{X} \sim \mathcal{N}(60, \frac{14}{\sqrt{49}} = 2)$, so $P[\bar{X} > 58] =$

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?
$\bar{X} \sim \mathcal{N}(60, \frac{14}{\sqrt{49}} = 2)$, so $P[\bar{X} > 58] = 0.8413$

(c) Can you determine the probability that a random sample of 49 bills will have a mean exactly equal to \$62?

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?
$\bar{X} \sim \mathcal{N}(60, \frac{14}{\sqrt{49}} = 2)$, so $P[\bar{X} > 58] = 0.8413$

(c) Can you determine the probability that a random sample of 49 bills will have a mean exactly equal to \$62?

Sampling Dist.of the Mean: Example 5

Restaurant bills at a given restaurant have a mean of \$60 and a standard deviation of \$14.

(a) Can you determine the probability that a given bill will be at least \$50?
We do not know the probability distribution of these bills, so we can't calculate this probability.

(b) Can you determine the probability that a random sample of 49 bills will have a mean greater than \$58?
$\bar{X} \sim \mathcal{N}(60, \frac{14}{\sqrt{49}} = 2)$, so $P[\bar{X} > 58] = 0.8413$

(c) Can you determine the probability that a random sample of 49 bills will have a mean exactly equal to \$62?
Since $\bar{X}$ is continuous, $P[\bar{X} = 62] = 0$

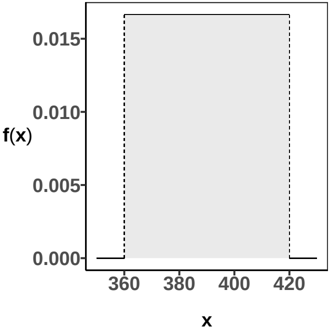
Sampling Dist.of the Mean: Example 6 I

The duration of direct flights from NYC to LA is uni formally distributed over the interval from 360 minutes to 420 minutes.

(a) Sketch the distribution of X. What's $\mathbb{E}[X]$ and σ_X

$$\mathbb{E}[X] = \frac{420 + 360}{2} = 390$$

$$\sigma_X = \sqrt{\frac{(420 - 360)^2}{12}} = 17.3205$$



Sampling Dist.of the Mean: Example 6 II

(b) Suppose we were to repeatedly take a random sample of size 100 from this distribution and compute the sample mean for each sample. What would the be the distribution of $\bar{X}$?
$\bar{X} \sim \mathcal{N}(\frac{420+360}{2} = 390, \frac{\sqrt{\frac{(420-360)^2}{12}}}{\sqrt{100}} = 1.7321)$

(c) What is the probability that a sample of 100 flights has a mean higher than 395 minutes of flight?
$P[\bar{X} > 395] = 0.0019$

Sampling Distribution of Proportions

The proportion of successes, x, in a sample consisting of n trials is:

$$\text{Sample proportion} = \hat{p} = \frac{\text{Number of successes}}{\text{Number of trials}} = \frac{x}{n}$$

When we take many samples form a population, each sample will have its own $\hat{p}$.

So, we will regard the sample proportion $\hat{p} = \{\hat{p}_1, \hat{p}_2, ..., \hat{p}_n\}$ as a random variable with its own mean and standard deviation.

The resulting probability distribution of the sample proportions is called ***the sampling distribution of Proportions.***

Properties of the Sampling Dist. of Proportions

Suppose we have a population with a proportion p. Also suppose we draw great many simple random samples of size n from this population, and let $\hat{p}$ be the sampling distribution of the proportions, then:

- The mean of $\hat{p}$, denoted $\mu_{\hat{p}}$ is

$$\mu_{\hat{p}} = \mathbb{E}[\hat{p}] = p$$

- The standard deviation of $\hat{p}$, usually called the standard error and denoted $\sigma_{\hat{p}}$ is:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

- According to the Central Limit Theorem, $\hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{pq}{n}}\right)$ if the sample size n is large enough such that $np > 5$, and $nq > 5$.

The Sampling Dist. of Proportions: Example 1

Suppose that 35% of people own a boat. Calculate the probability that, in a simple random sample of 50 people, more than 40% own a boat.

Answer: Here $p = 0.35$, $q = 0.65$, $n = 50$, and you are asked about $P[\hat{p} > 0.3]$. Since, $np = 17.5 > 5$ and $nq = 32.5 > 5$,

$\hat{p} \sim$

The Sampling Dist. of Proportions: Example 1

Suppose that 35% of people own a boat. Calculate the probability that, in a simple random sample of 50 people, more than 40% own a boat.

Answer: Here $p = 0.35$, $q = 0.65$, $n = 50$, and you are asked about $P[\hat{p} > 0.3]$. Since, $np = 17.5 > 5$ and $nq = 32.5 > 5$,
$\hat{p} \sim \mathcal{N}\left(0.35, \sqrt{\frac{0.35*0.65}{50}} = 0.0675\right)$

$$P[\hat{p} > 0.4] = 1 - P[\hat{p} < 0.4] = 1 - P\left[\frac{\hat{p} - \mu}{\sigma_{\hat{p}}} < \frac{0.4 - 0.35}{0.0675}\right]$$
$$= 1 - P[z < 0.7407] = 1 - 0.7706 = 0.2294$$

The Sampling Dist. of Proportions: Example 2

42% of 2016 female electorates voted for Trump. What is the probability that, in a simple random sample of 20 women, between 10 to 15 have voted for Trump?

Answer: Here $p = 0.42$, $q = 0.58$, $n = 20$, and you are asked about

The Sampling Dist. of Proportions: Example 2

42% of 2016 female electorates voted for Trump. What is the probability that, in a simple random sample of 20 women, between 10 to 15 have voted for Trump?

Answer: Here $p = 0.42$, $q = 0.58$, $n = 20$, and you are asked about $P[\frac{10}{20} < \hat{p} < \frac{15}{20}]$ or $P[0.5 < \hat{p} < 0.75]$. Since, $np = 8.4 > 5$ and $nq = 11.6 > 5$, $\hat{p} \sim$

The Sampling Dist. of Proportions: Example 2

42% of 2016 female electorates voted for Trump. What is the probability that, in a simple random sample of 20 women, between 10 to 15 have voted for Trump?

Answer: Here $p = 0.42$, $q = 0.58$, $n = 20$, and you are asked about $P[\frac{10}{20} < \hat{p} < \frac{15}{20}]$ or $P[0.5 < \hat{p} < 0.75]$. Since, $np = 8.4 > 5$ and $nq = 11.6 > 5$, $\hat{p} \sim \mathcal{N}\left(0.42, \sqrt{\frac{0.42*0.58}{20}} = 0.1104\right)$

$$\begin{aligned} P[0.5 < \hat{p} < 0.75] &= P\left[\frac{0.5 - 0.42}{0.1104} < \frac{\hat{p} - \mu}{\sigma_{\hat{p}}} < \frac{0.75 - 0.42}{0.1104}\right] \\ &= P[0.7246 < z < 2.9891] \\ &= 0.9986 - 0.7657 \\ &= 0.2329 \end{aligned}$$

The Sampling Dist. of Proportions: Example 3

A Gallup poll found that for those Americans who have lost weight, 31% believed the most effective strategy involved exercise. What is the probability that from a simple random sample of 300 Americans the sample proportion falls: above 32%? between 29% and 33% ? and below 30%?

Answer: Here $p = 0.31$, $q = 0.69$, $n = 300$. Since, $np > 5$ and $nq > 5$, $\hat{p}$

The Sampling Dist. of Proportions: Example 3

A Gallup poll found that for those Americans who have lost weight, 31% believed the most effective strategy involved exercise. What is the probability that from a simple random sample of 300 Americans the sample proportion falls: above 32%? between 29% and 33% ? and below 30%?

Answer: Here $p = 0.31$, $q = 0.69$, $n = 300$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.31, \sqrt{\frac{0.31*0.69}{300}} = 0.0267\right)$, so
$P[\hat{p} > 0.32] =$

The Sampling Dist. of Proportions: Example 3

A Gallup poll found that for those Americans who have lost weight, 31% believed the most effective strategy involved exercise. What is the probability that from a simple random sample of 300 Americans the sample proportion falls: above 32%? between 29% and 33% ? and below 30%?

Answer: Here $p = 0.31$, $q = 0.69$, $n = 300$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.31, \sqrt{\frac{0.31*0.69}{300}} = 0.0267\right)$, so $P[\hat{p} > 0.32] = 0.354$, $P[0.29 < \hat{p} < 0.33] =$

The Sampling Dist. of Proportions: Example 3

A Gallup poll found that for those Americans who have lost weight, 31% believed the most effective strategy involved exercise. What is the probability that from a simple random sample of 300 Americans the sample proportion falls: above 32%? between 29% and 33% ? and below 30%?

Answer: Here $p = 0.31$, $q = 0.69$, $n = 300$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.31, \sqrt{\frac{0.31*0.69}{300}} = 0.0267\right)$, so
$P[\hat{p} > 0.32] = 0.354$, $P[0.29 < \hat{p} < 0.33] = 0.5462$, and
$P[\hat{p} < 0.30] =$

The Sampling Dist. of Proportions: Example 3

A Gallup poll found that for those Americans who have lost weight, 31% believed the most effective strategy involved exercise. What is the probability that from a simple random sample of 300 Americans the sample proportion falls: above 32%? between 29% and 33% ? and below 30%?

Answer: Here $p = 0.31$, $q = 0.69$, $n = 300$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.31, \sqrt{\frac{0.31*0.69}{300}} = 0.0267\right)$, so
$P[\hat{p} > 0.32] = 0.354$, $P[0.29 < \hat{p} < 0.33] = 0.5462$, and
$P[\hat{p} < 0.30] = 0.354$

The Sampling Dist. of Proportions: Example 4

The proportion of college students who use marijuana is $p = 0.40$. What is the probability that the percentage of students who use marijuana in a sample of $n = 200$ is less than 38%?

Answer: Here $p = 0.40$, $q = 0.60$, $n = 200$. Since, $np > 5$ and $nq > 5$, $\hat{p}$

The Sampling Dist. of Proportions: Example 4

The proportion of college students who use marijuana is $p = 0.40$. What is the probability that the percentage of students who use marijuana in a sample of $n = 200$ is less than 38%?

Answer: Here $p = 0.40$, $q = 0.60$, $n = 200$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.4, \sqrt{\frac{0.4*0.6}{200}} = 0.0346\right)$, so
$P[\hat{p} < 0.38] = 0.2819$

The Sampling Dist. of Proportions: Example 5

In a typical class, about 80% of students receive a C or better. Out of a random sample of 100 students, what is the probability that less than 70% receive a C or better?

Answer: Here $p = 0.80$, $q = 0.20$, $n = 100$. Since, $np > 5$ and $nq > 5$, $\hat{p}$

The Sampling Dist. of Proportions: Example 5

In a typical class, about 80% of students receive a C or better. Out of a random sample of 100 students, what is the probability that less than 70% receive a C or better?

Answer: Here $p = 0.80$, $q = 0.20$, $n = 100$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.8, \sqrt{\frac{0.8*0.2}{100}} = 0.04\right)$, so
$P[\hat{p} < 0.70] =$

In a typical class, about 80% of students receive a C or better. Out of a random sample of 100 students, what is the probability that less than 70% receive a C or better?

Answer: Here $p = 0.80$, $q = 0.20$, $n = 100$. Since, $np > 5$ and $nq > 5$, $\hat{p} \sim \mathcal{N}\left(0.8, \sqrt{\frac{0.8*0.2}{100}} = 0.04\right)$, so
$P[\hat{p} < 0.70] = 0.0062$

Sampling Without Replacement and From a Finite Population

As the sample size is increased the standard errors get smaller and smaller. To correct the standard errors associated with large sample sizes, whenever $n > 0.05 \times N$, we multiply the standard errors by a population correction factor of $\sqrt{\frac{N-n}{N-1}}$. The standard errors then become:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \times \sqrt{\frac{N-n}{N-1}}$$

Sampling Without Replacement and From a Finite Population: Example

Of the 629 passenger vehicles imported by a small country, 117 were Volvos. A simple random sample of 300 passenger vehicles imported during that year is taken. What is the probability that at least 15% of the vehicles in this sample will be Volvos?

Answer: $p = 117/629 = 0.186$, $q = 0.814$, and $n = 300 > 0.05 * 629 = 31.45$. So:

$$\sigma_{\hat{p}} =$$

Sampling Without Replacement and From a Finite Population: Example

Of the 629 passenger vehicles imported by a small country, 117 were Volvos. A simple random sample of 300 passenger vehicles imported during that year is taken. What is the probability that at least 15% of the vehicles in this sample will be Volvos?

Answer: $p = 117/629 = 0.186$, $q = 0.814$, and $n = 300 > 0.05 * 629 = 31.45$. So:

$$\sigma_{\hat{p}} = \sqrt{\frac{0.186 * 0.814}{300}} \times \sqrt{\frac{629 - 300}{629 - 1}} =$$

Sampling Without Replacement and From a Finite Population: Example

Of the 629 passenger vehicles imported by a small country, 117 were Volvos. A simple random sample of 300 passenger vehicles imported during that year is taken. What is the probability that at least 15% of the vehicles in this sample will be Volvos?

Answer: $p = 117/629 = 0.186$, $q = 0.814$, and $n = 300 > 0.05 * 629 = 31.45$. So:

$$\sigma_{\hat{p}} = \sqrt{\frac{0.186 * 0.814}{300}} \times \sqrt{\frac{629 - 300}{629 - 1}} = 0.0225 * 0.7238 = 0.0163$$

$$P[\hat{p} > 0.15] = 1 - P[z < -2.21] = 0.9864$$