Chapter 8: Sampling Distributions

STAT 2601 – Business Statistics

Esam Mahdi

School of Mathematics and Statistics
Carleton University

January 12, 2026

Table of Contents

Review: Types of Samples

Why Sampling?

- Cost-effective and time-efficient
- Often more accurate than a census (less non-sampling error)
- Destructive testing requires sampling

Types of Samples:

1. **Probability Samples** (Random)
2. **Non-Probability Samples** (Not Random)

Probability Samples (Random Samples)

Probability samples are subsets of a population selected through random methods in which every member has a known, non-zero probability of inclusion, enabling researchers to draw statistically valid conclusions about the entire population.

1. **Simple Random Sample (SRS)**
   - Every sample of size n has equal chance
   - *Example: Random number generator to select student IDs*
2. **Stratified Random Sample**
   - Divide population into homogeneous strata
   - SRS from each stratum
   - *Example: Survey by gender (male, female)*
3. **Cluster Sample**
   - Divide population into clusters
   - Randomly select clusters
   - Sample all in chosen clusters
   - *Example: Randomly select classrooms, survey all students in those rooms*
4. **Systematic Random Sample**
   - Select every k-th element
   - Random start between 1 and k
   - *Example: Every 10th customer entering a store*

Non-Probability Samples

Warning

Results cannot be generalized to population with known precision.

1. **Convenience Sampling**
   - Easy to reach individuals
   - *Example: Surveying people in your dorm*
2. **Purposive/Judgmental Sampling**
   - Researcher selects based on judgment
   - *Example: Interviewing "typical" customers*
3. **Quota Sampling**
   - Ensure certain groups are represented
   - *Example: Survey 50 males and 50 females*
4. **Snowball Sampling**
   - Participants recruit other participants
   - *Example: Studying hard-to-reach populations*
5. **Self-Selection/Voluntary Sampling**
   - Individuals choose to participate
   - *Example: Online polls, call-in surveys*

Sampling Error

Definition

The difference between a sample statistic and the corresponding population parameter, arising from random chance in sample selection.

$$\text{Sampling Error} = \text{Sample Statistic} - \text{Population Parameter}$$

Example

- Population mean: $\mu = 65$
- Sample mean: $\bar{x} = 63.5$
- Sampling error: $\bar{x} - \mu = 63.5 - 65 = -1.5$

Important notes

- Sampling error is **not** a mistake
- It's inherent in random sampling
- Different samples yield different sampling errors

The Sampling Distribution of the Sample Mean: Example

Population: Uniform distribution with values: 2, 4, 6, 8
Population Parameters:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2+4+6+8}{4} = 5$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} = \sqrt{\frac{(2-5)^2+(4-5)^2+(6-5)^2+(8-5)^2}{4}} = \sqrt{5} \approx 2.236$$

Experiment: Draw all possible samples of size $n = 2$ (with replacement)

All Possible Samples of Size 2

Sample	Values	Sample Mean ($\bar{x}$)	Probability
1	(2,2)	2	1/16
2	(2,4)	3	1/16
3	(2,6)	4	1/16
4	(2,8)	5	1/16
5	(4,2)	3	1/16
6	(4,4)	4	1/16
7	(4,6)	5	1/16
8	(4,8)	6	1/16
9	(6,2)	4	1/16
10	(6,4)	5	1/16
11	(6,6)	6	1/16
12	(6,8)	7	1/16
13	(8,2)	5	1/16
14	(8,4)	6	1/16
15	(8,6)	7	1/16
16	(8,8)	8	1/16

Distribution of Sample Means

$\bar{x}$	Frequency	Probability
2	1	1/16 = 0.0625
3	2	2/16 = 0.1250
4	3	3/16 = 0.1875
5	4	4/16 = 0.2500
6	3	3/16 = 0.1875
7	2	2/16 = 0.1250
8	1	1/16 = 0.0625

Parameters of Sampling Distribution:

$$\mu_{\bar{x}} = \underline{\sum \bar{x} \cdot P(\bar{x})}$$

$$= 5.0$$

$$\sigma_{\bar{x}} = \sqrt{\sum (\bar{x} - \mu_{\bar{x}})^2 \cdot P(\bar{x})}$$

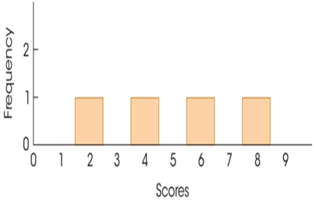
$$= \sqrt{2.5} \approx 1.581$$

Key Observations

- $\mu_{\bar{x}} = \mu = 5$ ✓
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.236}{\sqrt{2}} = 1.581$ ✓
- Distribution is symmetric and bell-shaped (even though population was uniform!)

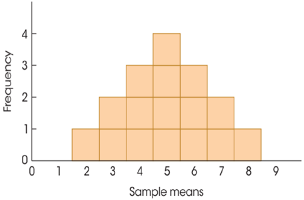
Visual Comparison

Population Distribution



Uniform, 4 values

Sampling Distribution $(n = 2)$



Symmetric, 7 values

Notice

- Sampling distribution has smaller spread
- Sampling distribution is more normal-like
- Center remains the same

Properties of Sampling Distribution of $\bar{X}$

Unbiasedness

The mean of all possible sample means equals the population mean:

$$\mu_{\bar{X}} = \mu$$

We say $\bar{X}$ is an **unbiased estimator** of μ.

Standard Error (Standard Deviation)

The **standard deviation** of the sampling distribution:

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

- Also called **standard error of the mean**
- Measures precision of $\bar{X}$ as an estimator
- Decreases as sample size increases

Assumptions and Conditions

For $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ to hold exactly:

1. **Random Sample:** Observations are independent
2. **Sample Size:** n is fixed
3. **Population:** Normally distributed (for small n)
4. **Sampling:** With replacement where the sample size, n, should be no more than 20% of the population (i.e., $n/N < 0.2$, where N is the population size)

Example

- Survey 50 Carleton students about study hours
- Population: All Carleton students (25,000+)
- Since $50 < 0.2 \times 25,000 = 5000$, we can use formula
- Standard error $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{50}}$

Central Limit Theorem (The Most Important Theorem!)

Central Limit Theorem

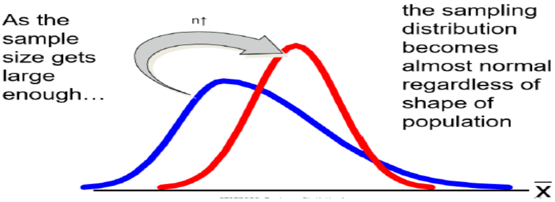
For a random sample of size n from **any** population with mean μ and standard deviation σ, when n is sufficiently large ($n \geq 30$ as a rule of thumb):

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu, \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\right)$$

or equivalently

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\right)$$

That is, the sampling distribution of $\bar{X}$ is approximately normal.



CLT: Visual Demonstration

Population	$n < 30$	$n \geq 30$
Any shape	Somewhat normal	Very normal!
μ, σ		

Key Points

- Works for **any** population shape
- Approximation improves as n increases
- If population is normal, $\bar{X}$ is normal for **any** n (small or large)
- Threshold number for large samples: $n \geq 30$

Probability Calculations for $\bar{X}$

Standardizing $\bar{X}$: Transform Normal to Standard Normal

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Example ($P(\bar{X} < a)$)

The heights of Carleton University students are normally distributed with mean $\mu = 170$ cm and standard deviation $\sigma = 10$ cm. Suppose a random sample of $n = 64$ students is selected. Determine the probability that the sample mean height is less than 168.5 cm.

Solution: We need to find $P(\bar{X} < 168.5)$, where $\bar{X} \sim N(\mu = 170, \sigma^2 = 100)$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{64}} = 1.25$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{168.5 - 170}{1.25} = -1.2$$

Thus,

$$P(\bar{X} < 168.5) = P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{168.5 - 170}{1.25}) = P(Z < -1.2) = 0.1151$$

Example ($P(\bar{X} > a)$)

Same parameters of the previous example. Find $P(\bar{X} > 172)$.

Solution:

$$P(\bar{X} > 172) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{172 - 170}{1.25}\right) = P(Z > 1.6)$$

Option 1

$$P(Z > 1.6) = 1 - P(Z < 1.6) = 1 - 0.9452 = 0.0548$$

Option 2

$$P(Z > 1.6) = P(Z < -1.6) = 0.0548$$

Example ($P(a < \bar{X} < b)$)

Find $P(169 < \bar{X} < 171.5)$.

Solution:

$$\begin{aligned} P(169 < \bar{X} < 171.5) &= P\left(\frac{169 - 170}{1.25} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{171.5 - 170}{1.25}\right) = P(-0.8 < Z < 1.2) \\ &= P(Z < 1.2) - P(Z < -0.8) \\ &= 0.8849 - 0.2119 \\ &= 0.6730 \end{aligned}$$

Real-Life Example

A coffee shop knows the average transaction is $\mu = \$8.5$ with $\sigma = \$2.2$. They take a random sample of 50 transactions on a busy Saturday.

1. **Find probability sample mean exceeds \$9.0:**

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.2}{\sqrt{50}} \approx 0.311$$

$$P(\bar{X} > 9.0) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{9.0 - 8.5}{0.311}\right) \approx P(Z > 1.61) = 0.0537$$

2. **Find probability sample mean between \$8.2 and \$8.8:**

$$\begin{aligned} P(8.20 < \bar{X} < 8.80) &= P\left(\frac{8.20 - 8.50}{0.311} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{8.80 - 8.50}{0.311}\right) \\ &\approx P(-0.96 < Z < 0.96) \\ &= P(Z < 0.96) - P(Z < -0.96) = 0.8315 - 0.1685 = 0.6629 \end{aligned}$$

Since the sample size is large ($n = 50$), the Central Limit Theorem (CLT) applies, so the sampling distribution of the mean is approximately normal even if the transaction on a busy Saturday (population) is not clearly stated as normally distributed.

Population and Sample Proportions

Population Proportion (p)

$$p = \frac{\text{Number of successes in population}}{\text{Population size}}$$

Sample Proportion ($\hat{p}$)

$$\hat{p} = \frac{X}{n} = \frac{\text{Number of successes in sample}}{\text{Sample size}}$$

where $X \sim \text{Binomial}(n, p)$

Example

- Population: All Carleton students (25,000)
- Success: Owns a laptop
- $p = 0.85$ (85% own laptops)
- Sample: $n = 100$ students, $X = 88$ own laptops
- $\hat{p} = 88/100 = 0.88$

Sampling Distribution of $\hat{p}$

Central Limit Theorem (CLT) for Proportions

For large n ($np \geq 5$ and $n(1-p) \geq 5$):

$$\hat{p} \sim N\left(\mu_{\hat{p}} = p, \sigma^2_{\hat{p}} = \frac{p(1-p)}{n}\right)$$

Standard Error for Proportion

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}, \quad \text{where} \quad q = 1 - p$$

Standardizing $\hat{p}$: Transform Normal to Standard Normal

$$\text{If } \hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right), \text{ then } Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Assumptions and Conditions

1. **Random Sample**
2. **Independence:** $n \leq 5\%$ of population
3. **Sample Size:** $np \geq 5$ and $n(1-p) \geq 5$

[

Real-Life Example: $P(\hat{p} < a)$] A survey suggests that 65% of Carleton students prefer online classes ($p = 0.65$). If we take a random sample of $n = 200$ students, what is the probability that the sample proportion $\hat{p}$ is less than 60%?

Solution:

1. **Check conditions:**

$$np = 200 \times 0.65 = 130 \geq 5 \quad \checkmark$$
$$n(1-p) = 200 \times 0.35 = 70 \geq 5 \quad \checkmark$$

   Conditions for normal approximation are satisfied.

2. **Calculate standard error:**

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.65 \times 0.35}{200}} = \sqrt{\frac{0.2275}{200}} = \sqrt{0.0011375} \approx 0.03373$$

3. **Standardize and Find probability:**

$$P(\hat{p} < 0.60) = P(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.60 - 0.65}{0.03373}) \approx P(Z < -1.48) = 0.0694$$

Interpretation: There is approximately a 6.9% chance that in a random sample of 200 Carleton students, fewer than 60% will prefer online classes, even though the true population proportion is 65%.

Example ($P(\hat{p} > a)$)

Same parameters of the previous example with $p = 0.65$ and $n = 200$, find $P(\hat{p} > 0.70)$.
Solution:

$$\sigma_{\hat{p}} = \sqrt{\frac{0.65 \times 0.35}{200}} \approx 0.03373$$

$$\begin{aligned}
P(\hat{p} > 0.70) &= P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.70 - 0.65}{0.03373}\right) \approx P(Z > 1.48) \\
&= 1 - P(Z \leq 1.48) \\
&= 1 - 0.9306 = 0.0694
\end{aligned}$$

Note: By symmetry of the normal distribution:

$$P(\hat{p} > 0.70) \approx P(Z > 1.48) = P(Z < -1.48) = P(\hat{p} < 0.60) \approx 0.0694$$

Example ($P(a < \hat{p} < b)$)

Same parameters of the previous example with $p = 0.65$ and $n = 200$, find $P(0.62 < \hat{p} < 0.68)$.

$$\begin{aligned}
P(0.62 < \hat{p} < 0.68) &= P\left(\frac{0.62 - 0.65}{0.0337} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.68 - 0.65}{0.0337}\right) \\
&\approx P(-0.89 < Z < 0.89) = P(Z < 0.89) - P(Z < -0.89) \\
&= 0.8133 - 0.1867 = 0.6266
\end{aligned}$$

Real-Life Example: Quality Control

A factory produces light bulbs. Historically, 5% are defective ($p = 0.05$). Quality control takes daily samples of 500 bulbs.

1. **Check conditions:**
$$np = 500(0.05) = 25 \geq 5 \quad \checkmark$$
$$n(1-p) = 500(0.95) = 475 \geq 5 \quad \checkmark$$

2. **Find probability sample proportion > 0.06:**
$$\sigma_{\hat{p}} = \sqrt{\frac{0.05(0.95)}{500}} = 0.00975$$
$$P(\hat{p} > 0.06) = P\left(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.06-0.05}{0.00975}\right) \approx P(Z > 1.03)$$
$$= 1 - P(Z < 1.03) = 1 - 0.8485 = 0.1515$$
$$\stackrel{\text{or}}{=} P(Z < -1.03) = 0.1515$$

3. **Interpretation:** About 15% of days will have defect rates over 6% purely by chance.

Summary

1. **Sampling Distributions** describe how statistics vary across samples
2. **Central Limit Theorem:** For large $n \geq 30$, $\bar{X}$ and $\hat{p}$ are approximately normal
3. **Standard Error** measures sampling variability:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

4. **Probability Calculations:** Standardize and use Z-table
5. $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{and} \quad Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$
6. **Assumptions Matter:** Check randomness, independence, sample size conditions

Practice Problems

1. Population: $\mu = 100$, $\sigma = 15$, $n = 36$. Find $P(95 < \bar{X} < 105)$.
2. Election poll: $p = 0.52$, $n = 400$. Find $P(\hat{p} < 0.50)$.
3. Explain why $\sigma_{\bar{X}}$ decreases as n increases.
4. Why is $n \geq 30$ a rule of thumb for CLT?
5. Suppose that 20% of university students regularly use AI tools (like ChatGPT or DeepSeek) to complete their assignments. Consider a random sample of 150 students and let $\hat{P}$ represent the proportion of these students who use AI tools for assignments.
    1. What is the sampling distribution of $\hat{P}$? Explain.
    2. Use your answer from part A to find the probability that more than 33 of these 150 students use AI tools for assignments.
6. Compare standard error formulas for means and proportions.