6 - Sampling Dist of Mean

October 20, 2025 5:42 PM

6 - Sampling Distribution of the Mean

We are about to establish one of the central links between *descriptive statistics* and *probability theory*. This will lead us to *inferential statistics*. The key idea is this:

The sample mean $\bar{X}$ is a random variable.

↳ and often $\bar{X}$ is normally distributed

Example Suppose we have a class of $N = 100$ students. Each student has a certain age, X. Assume that all the values of X are:

20	20	20	20	23	20	18	20	21	24
20	23	20	22	23	19	22	21	19	22
19	19	21	24	19	21	22	25	21	21
23	21	20	18	20	19	20	21	20	23
20	19	21	21	22	21	22	19	22	21
23	20	21	19	22	24	22	21	21	22
21	23	20	19	20	25	20	22	22	24
20	19	20	22	24	21	19	20	20	20
22	20	19	21	19	23	21	25	21	22
19	22	20	20	23	23	20	20	20	26

one sample

$\bar{X} = 21.0$

population data

from this population

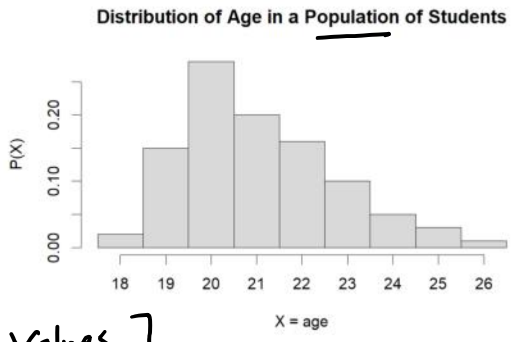
Using R, we find that the population mean and standard deviation are:

$N = 100$

$\mu = 21.04$

$\sigma = 1.673$

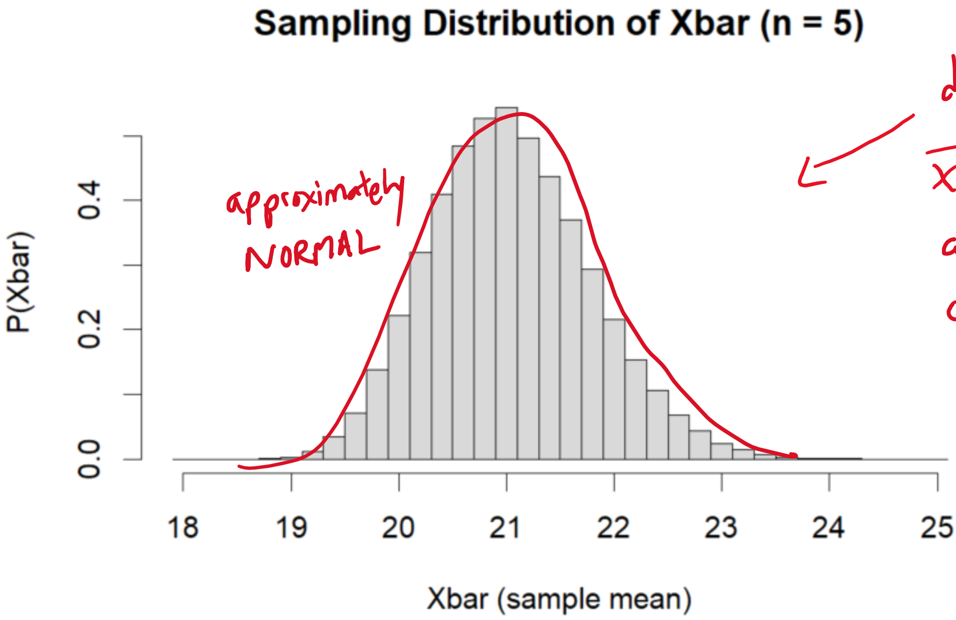
[Typically we would not know these parameter values.]



Suppose we randomly select a sample of $n = 5$ students.

It turns out that the sample mean $\bar{X}$ is a random variable whose distribution is/has:

- $\mu_{\bar{X}} = \mu = 21.04$
- $\sigma_{\bar{X}} = \sigma/\sqrt{n} = \dfrac{1.673}{\sqrt{5}} = 0.748$
- Shape of $\bar{X}$ distribution is approx. NORMAL



distribution of $\bar{X}$ values for a large number of possible samples

Definition The distribution of $\bar{X}$ is also called the *sampling distribution of the mean*.

sample

Observations:

- The values of $\bar{X}$ are now decimal values (like 20.8) instead of integer values like X.
- Therefore, there are more *possible* values of $\bar{X}$. (i.e., 19.8, 20.0, 20.2, ...)
- If we increased the sample size n, then $\bar{X}$ would become a *continuous* variable.
- It does not make much difference if we use *replacement* or *no replacement* as long as the population size N is much larger than the sample size n.

Mathematically we assume with replacement to have independence.

Summary of Symbols

X = the random variable (both sample and population)

N = population size
μ = population mean value of X
σ = population std dev. of X

pop. parameters (unknown usually)

n = sample size
$\bar{X}$ = sample mean of X
s = sample std dev of X

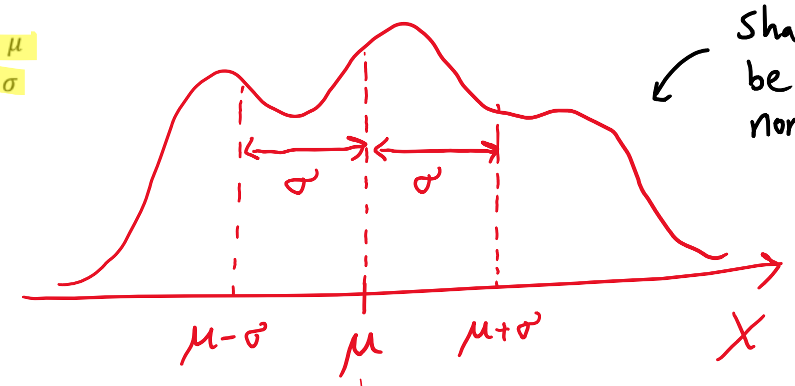
sample statistics (known but random)

Sampling Distribution of the Mean – In General

Suppose X is any random variable for individuals selected from a population of size N.
Then X has a probability distribution with parameters:

- mean μ
- std. dev. σ

Shape may be non-normal



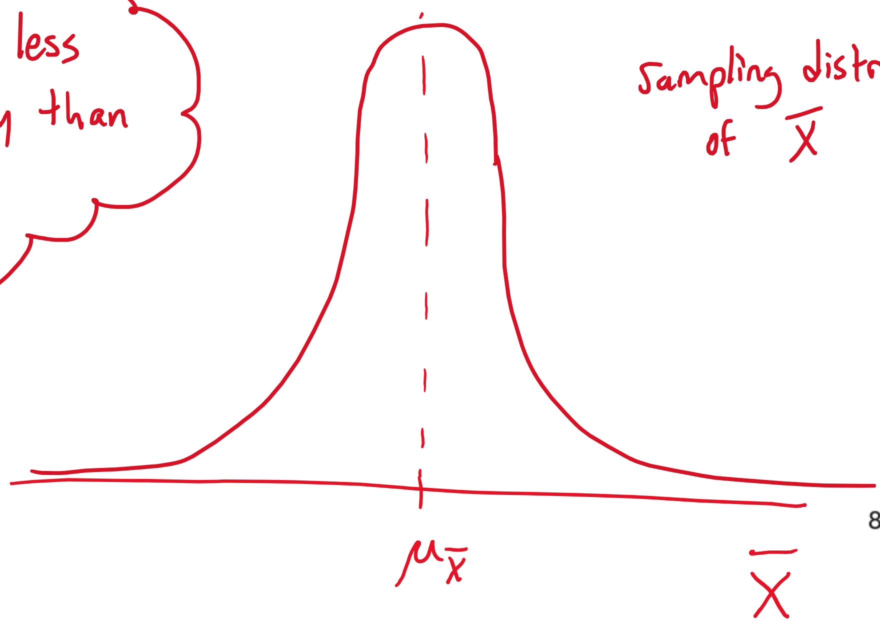
If samples of size n are randomly selected from the population, then the sample mean $\bar{X}$ is a random variable with its own probability distribution called the *sampling distribution of the mean*. The parameters of $\bar{X}$ are:

- mean $\mu_{\bar{X}} = \mu$
- std. dev. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ ← smaller than σ for $n > 1$

Think of $\mu_{\bar{X}}$ as "the average of the averages."

Think of $\sigma_{\bar{X}}$ as the typical error between $\bar{X}$ and μ. (the "standard error of the mean.")

$\bar{X}$ has less variability than X has



Sampling distribution of $\bar{X}$

Central Limit Theorem ($\bar{X}$ is approx NORMAL)

Using more advanced concepts from probability theory, it is possible to prove:

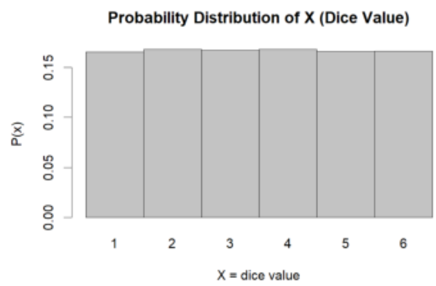
Central Limit Theorem Suppose a random variable X is defined on a population of size N. If a sample of n individuals are selected independently from the population, then the sampling distribution of $\bar{X}$ is:

- *normal* if X was normal
- *approximately normal* even if X was not normal ($n \geq 30$)
    - the approximation becomes better and better as $n \to \infty$

As a practical rule, if $\boxed{n \geq 30}$ then $\bar{X}$ follows a normal distribution (even if X does not).

Example (Dice Rolls) Suppose you roll $n = 1$ six-sided die. Let $X =$ the number that turns up. Then the probability distribution of X is given by:

x	$P(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6



The mean and standard deviation of X are:

$$\mu = \sum[x \cdot P(x)] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

$$\sigma^2 = \sum[(x - \mu)^2 \cdot P(x)] = (1 - 3.5)^2 \cdot \frac{1}{6} + (2 - 3.5)^2 \cdot \frac{1}{6} + (3 - 3.5)^2 \cdot \frac{1}{6} +$$

$$(4 - 3.5)^2 \cdot \frac{1}{6} + (5 - 3.5)^2 \cdot \frac{1}{6} + (6 - 3.5)^2 \cdot \frac{1}{6} = 2.91666 \ldots$$

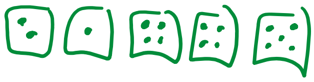
$$\sigma = \sqrt{2.91666} = \boxed{1.708}$$

What is the sampling distribution of $\bar{X}$ if we roll $n = 5$ dice?

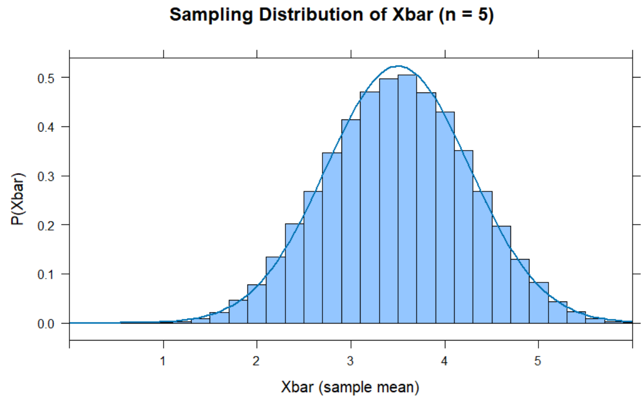
$n = 5$

$\mu_{\bar{x}} = \mu = 3.5$

$\sigma_{\bar{x}} = \sigma / \sqrt{n} = \frac{1.708}{\sqrt{5}} = 0.764$

e.g. 

$\bar{x} = 3.2$

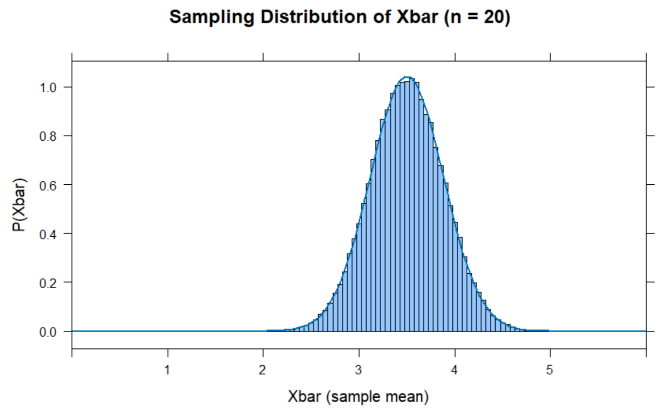


What is the sampling distribution of $\bar{X}$ if we roll $n = 20$ dice?

$n = 20$

$\mu_{\bar{x}} = \mu = 3.5$

$\sigma_{\bar{x}} = \sigma / \sqrt{n} = \frac{1.708}{\sqrt{20}} = 0.382$

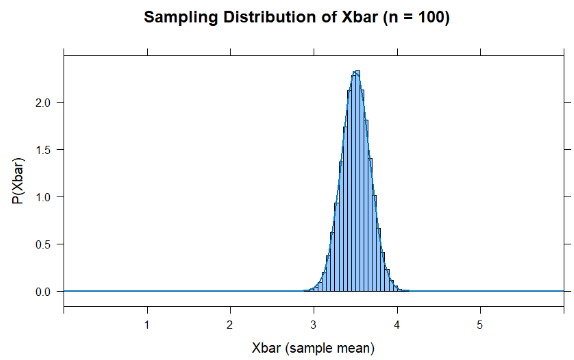


What is the sampling distribution of $\bar{X}$ if we roll $n = 100$ dice?

$n = 100$

$\mu_{\bar{x}} = \mu = 3.5$

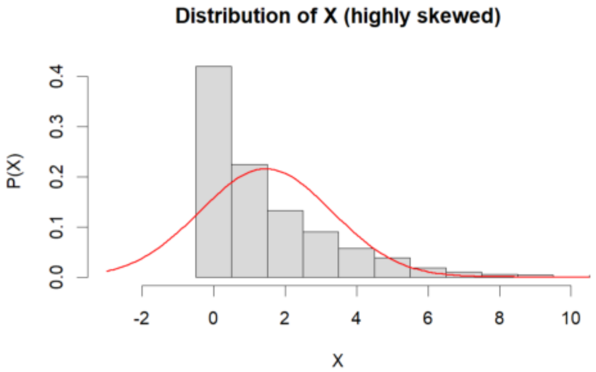
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{100}} = \frac{1.708}{10} = 0.1708$



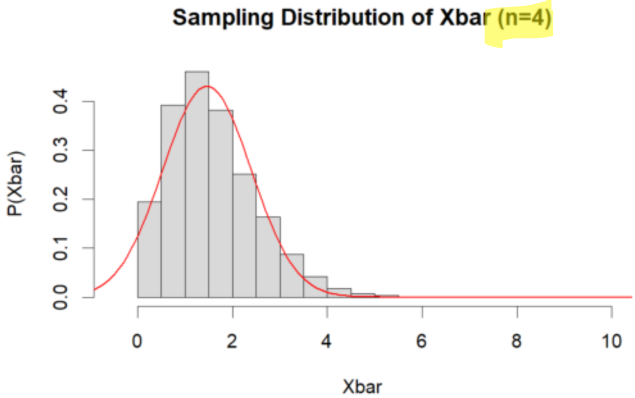
As the sample size, n, increases the distribution

- becomes narrower
- becomes closer to being normal

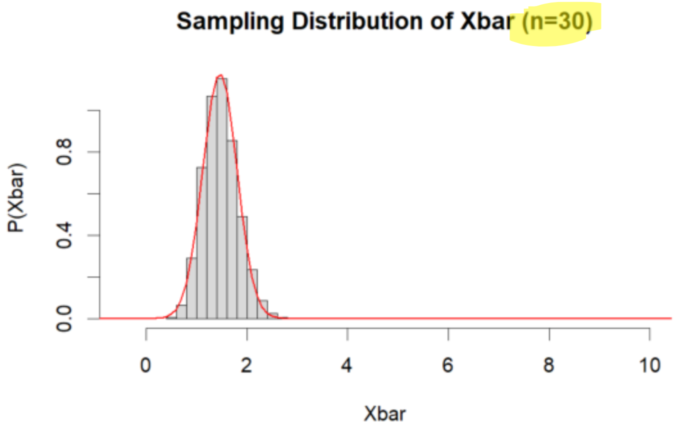
Example If the distribution of X is highly *skewed* then it requires a larger n before the sampling distribution of $\bar{X}$ becomes close to normal.



Exponential Distribution
$n=1$
$\bar{X}$ is Exponential (same as X)



$\bar{X}$ is in between being Exponential and being Normal



$\bar{X}$ is very close to being normally distributed

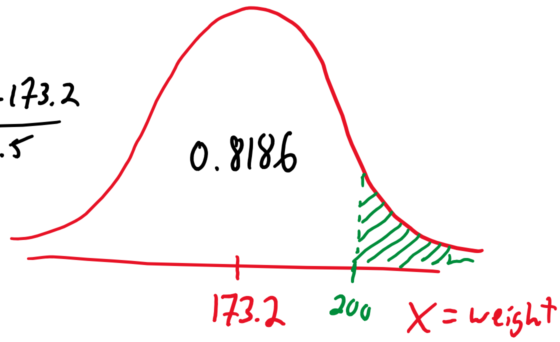
Example In human engineering and product design, it is important to consider the weights of people. Assume that the population of male BCIT students has normally distributed weights, with mean 173.2 lbs and a standard deviation of 29.5 lbs.

a. Find the probability that if a male student is randomly selected, his weight is greater than 200 lbs.

$\mu = 173.2$ lbs
$\sigma = 29.5$ lbs

$P(X > 200)$
$= P(Z > 0.908)$
$= 1 - 0.8186$
$= \boxed{0.1814}$

$Z = \frac{200 - 173.2}{29.5}$
$= 0.908$



b. An elevator has a maximum weight capacity of 7200 lbs. Find the probability that 36 randomly selected male students will exceed the elevator's weight capacity.

$n = 36$
$\mu_{\bar{X}} = \mu = 173.2$ lbs.
$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{29.5}{\sqrt{36}} = 4.917$ lbs

If $X_1 + X_2 + \dots + X_{36} > 7200$
then $\bar{X} > 7200/36 = 200$

$P(\text{Total} > 7200) = P(\bar{X} > 200)$
$= 1 - P(Z < 5.45)$
$= 0^+$

$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$
$= \frac{200 - 173.2}{4.917}$
$= 5.45$

Almost no chance that 36 male students are over 7200 lbs limit.

Example Assume that the population of adult body temperatures has a mean 37.0 °C. Also assume that the population standard deviation is 0.62 °C.

a. If a sample of $n = 108$ is randomly selected, find the probability of getting a mean of 36.8 °C or lower.

X is normal with $\mu = 37.0°C$ and $\sigma = 0.62°C$.

Since X is normal, $\bar{X}$ is also normal for any sample size n.

- $\mu_{\bar{x}} = \mu = 37.0°C$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.62°C}{\sqrt{108}} = 0.0597°C$

$P(\bar{X} < 36.8°C) = P(Z < -3.35) = 0.0004$ [Z-table]

$Z = \frac{36.8 - 37.0}{0.0597} = -3.35$

b. Suppose you take a sample of $n = 108$ randomly selected adults and find that the mean of their body temperatures is 36.8 °C. Would this be *unusually low* (in a statistical sense), given the data about the population of adult body temperatures? What does your result suggest?

The Z-score for $\bar{X} = 36.8°C$ was -3.35.

The is highly unusual $(Z < -2)$.

It suggests that the true population mean is actually less than $37.0°C$.

Finite Population Correction Factor

We have used the formula

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

under the assumption that individuals were sampled *independently* from the population, which requires either:

- the population is infinite or practically infinite ($n < 5\%$ of N), or
- sampling is *with replacement*

If sampling is done *without replacement* where $n \geq 5\%$ of N, then we need to adjust the formula above:

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

Not tested on Final Exam this term.

The expression $\sqrt{\frac{N-n}{N-1}}$ is called a *finite correction factor*.

Example Suppose we collect midterm scores for all 150 students in MATH 3042. We find the mean is 71% and the standard deviation is 6.5%. If 40 students are chosen without replacement, find the probability that their mean score is greater than 73.

Population: $N = 150$, $\mu = 71\%$, $\sigma = 6.5\%$ X = midterm score

Sample: $n = 40$ $\bar{X}$ = sample mean of midterm scores

$\mu_{\bar{X}} = \mu_X = 71\%$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{6.5\%}{\sqrt{40}} \cdot \sqrt{\frac{150-40}{150-1}} = 0.883\%$$

[would be $\frac{\sigma}{\sqrt{n}} = 1.028\%$]

$$P(\bar{X} > 73) = P(Z > 2.26) = 1 - 0.9881 = \boxed{0.0119}$$

$$Z = \frac{73-71}{0.883} = 2.2649 \approx 2.26$$