7.5 Sampling Distribution of $\bar{X}$

* The sampling distribution of $\bar{x}$ is the probability distribution of all possible values of the sample mean $\bar{x}$.

• Now, we study the properites of the sampling distribution of $\bar{x}$ as other prob. distributions, interms of the following characteristics:
1. Expected value
2. standarad deviation
3. the shap of the distribution

to determine how close the sample mean $\bar{x}$ is to the population mean μ

1 Expected Value of $\bar{x}$

* The expected value of $\bar{x}$ equals to the mean of the population from which the sample is selected.

$$E(\bar{X}) = \mu \quad ---- (1)$$

* Hence, with a simple random sample, the expected value or mean of the sampling distribution of $\bar{x}$ is equal to the mean of the population.

* Recall that in Example*, the population mean $\mu = \$51,800$ and the expected value of $\bar{x}$ is $E(\bar{x}) = \$51,800$

Hence, $E(\bar{x}) = \mu = \$51,800$ and we say that the point estimator $\bar{x}$ is unbiased.

* Unbiased: A property of a point estimator that is present when the expected value of the point estimator is equal to the population parameter.

* Note that eq. (1) shows that $\bar{x}$ is unbiased estimator of the population mean μ.

2 Standard Deviation of $\bar{x}$

Let : N = Population size

n = sample size

σ = Population standard deviation

$\sigma_{\bar{x}}$ = standard deviation of the sampling distribution of $\bar{x}$.

* For finite Population : the standard deviation of $\bar{x}$ is

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{where}$$

$\sqrt{\frac{N-n}{N-1}}$ is the finite population correction factor.

* For infinite population "process" or when N is large and n is small

that is when $\frac{n}{N} \leq 0.05$, then the standard deviation of $\bar{x}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

This because the finite population correction factor becomes close to 1, and so has a little effect on the value of $\sigma_{\bar{x}}$. (so, we ignor it).

* Note that $\sigma_{\bar{x}}$ is also called the standard error of the point estimator $\bar{x}$

Hence, the standard error is used throughout statistical inference to refere to the standard deviation of a point estimator.

$\sigma_{\bar{p}}$ is the standard error of the poportion.

Recall Example* where the population standard deviation $\sigma = \$4000$ and $N = 2500$ managers, with simple random sample of size $n = 30$, we have $\frac{n}{N} = \frac{30}{2500} = 0.012 \leq 0.05$. Hence

the standard error of $\bar{x}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.3$

3 Shap (Form) of the sampling Distribution of $\bar{x}$

we consider two cases:

1 The population has a normal distribution:

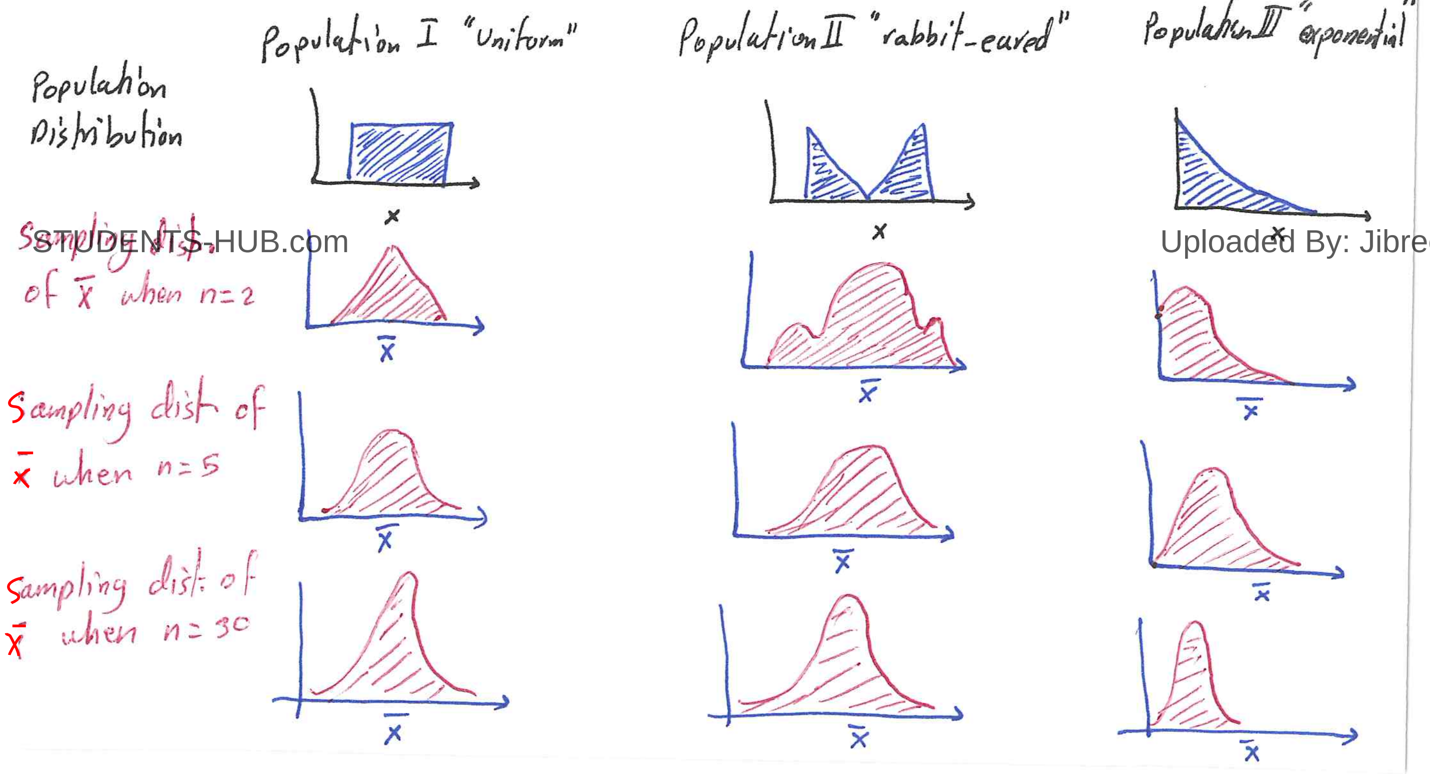
$\Rightarrow$ when the population has a normal distribution, the sampling distribution of $\bar{x}$ is normally distributed for any sample size.

2 The population does not have a normal distribution:

$\Rightarrow$ when the population, from which we select a simple random sample, does not have a normal distribution, we apply the Central limit Theorem to identify the shape of the sampling distribution of $\bar{x}$.

Central limit Theorem:

In selecting simple random samples of size n from a population, the sampling distribution of the sample mean $\bar{x}$ can be approximated by a normal distribution as the sample size becomes Large.



* Hence, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution if the sample size $n \geq 30$.

* If the population is highly skewed or If the population has an outliers, then the sampling distribution of $\bar{x}$ can be approximated by a normal distribution if the sample size $n \geq 50$.

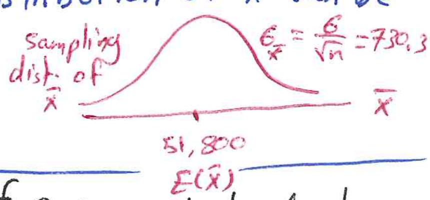
Recall Example* where $E(\bar{x}) = \$51,800$ and $\sigma_{\bar{x}} = 730.3$

$\Rightarrow$ we don't know if the population is normally distributed or not.

- So if the population has a normal distribution, then the sampling distribution of $\bar{x}$ is normally distributed.

- If the population does not have normal distribution, then
  1. the simple random sample of size 30 managers and
  2. the central limit Theorem

  enable us to conclude that the sampling distribution of $\bar{x}$ can be approximated by a normal distribution.



Example (Q18 page 276) A population has a mean of 200 and standard deviation of 50. A simple random sample of size 100 will be taken and the sample mean $\bar{x}$ will be used to estimat the population mean.

$\mu = 200$
$\sigma = 50$
$n = 100$

a) what is the expected value of $\bar{x}$? $E(\bar{x}) = \mu = 200$

b) what is the standard deviation of $\bar{x}$? $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = \frac{50}{10} = 5$

c) Show the sampling distribution of $\bar{x}$?

Since $n = 100 \geq 30$, it follows by the Central limit Theorem that the sampling distribution is approximated by a normal distribution with $E(\bar{x}) = 200$ and $\sigma_{\bar{x}} = 5$.

d) What does the sampling distribution of $\bar{x}$ show?

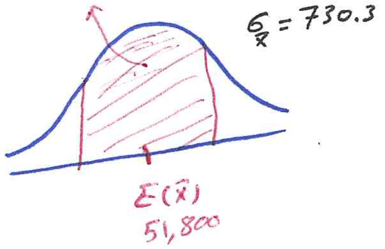
The probability distribution of $\bar{x}$. (see next page)

The sampling distribution of $\bar{x}$ provides probability information about how close the sample mean $\bar{x}$ to the population mean μ.

Recall Example* with $E(\bar{x}) = 51,800$ and $\sigma_{\bar{x}} = 730.3$

what is the Prop. that the sample mean will be within \$ 500 of the population mean?

$$P(51,800 - 500 \le \bar{x} \le 51,800 + 500) = P(51,300 \le \bar{x} \le 52,300)$$



$$z = \frac{\boxed{52,300} - 51,800}{730.30} = 0.68$$

$$z = \frac{\boxed{51,300} - E(\bar{x})}{\sigma_{\bar{x}}} = \frac{51,300 - 51,800}{730.3} = -0.68$$

$$P(51,300 \le x \le 52,300) = P(-0.68 \le z \le 0.68) = P(z \le 0.68) - P(z \le -0.68)$$
$$= 0.7517 - 0.2483 = 0.5034$$

• Hence, a simple random sample of size 30 has roughly 50/100 chance of providing a sample mean within \$ 500.

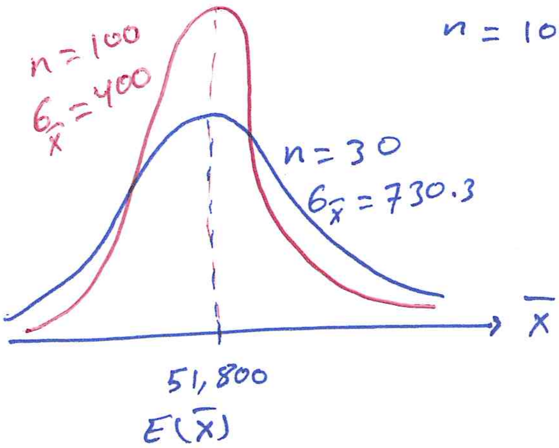
What is the relationship between the sample size n and the Sampling distribution of $\bar{x}$?

* Note that $E(\bar{x}) = \mu$ regardless of the sample size n.
* The standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ depends on the sample size n:

As n increases, the standard error $\sigma_{\bar{x}}$ decreases:

For example when $n = 30 \Rightarrow \sigma_{\bar{x}} = 730.3$ " in Example* "

$$n = 100 \Rightarrow \sigma_{\bar{x}} = \frac{4000}{\sqrt{100}} = \frac{4000}{10} = 400$$



Example (Q19 page 277) A population has a mean 200 and standard deviation 50. Suppose a simple random sample of size 100 is selected and $\bar{x}$ is used to estimate μ.

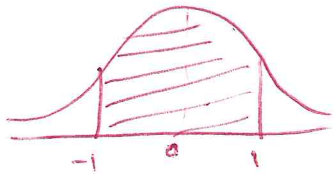
$\boxed{a}$ What is the prob. that the sample mean will be within ± 5 of the population mean? $E(\bar{x}) = 200$, $\sigma = 50$, $n = 100$

$P(200-5 \le \bar{x} \le 200+5) = P(195 \le \bar{x} \le 205)$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{10} = 5$

$z = \frac{\boxed{205} - E(\bar{x})}{\sigma_{\bar{x}}} = \frac{205-200}{5} = \frac{5}{5} = 1$

$z = \frac{\boxed{195} - E(\bar{x})}{\sigma_{\bar{x}}} = \frac{195-200}{5} = \frac{-5}{5} = -1$

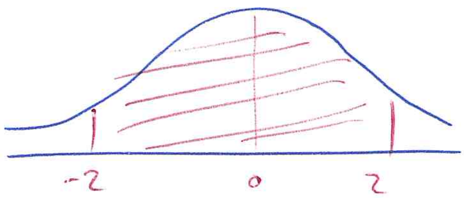


$P(195 \le \bar{x} \le 205) = P(-1 \le z \le 1) = P(z \le 1) - P(z \le -1) = 0.8413 - 0.1587$
$= 0.6826$

$\boxed{b}$ What is the prob. that the sample mean will be within ± 10 of the population mean?

$P(200-10 \le \bar{x} \le 200+10) = P(190 \le \bar{x} \le 210)$

$z = \frac{\boxed{210} - E(\bar{x})}{\sigma_{\bar{x}}} = \frac{210-200}{5} = \frac{10}{5} = 2$



$z = \frac{\boxed{190} - E(\bar{x})}{\sigma_{\bar{x}}} = \frac{190-200}{5} = \frac{-10}{5} = -2$

$P(190 \le \bar{x} \le 210) = P(-2 \le z \le 2) = P(z \le 2) - P(z \le -2) = 0.9772 - 0.0228$
$= 0.9544$

Example (Q21 page 277) Suppose a simple random sample of size 50 is selected from a population with $\sigma = 10$. Find the standard error if

$\boxed{a}$ The population is infinite: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{50}} = \frac{10}{7.07} = 1.4$

$\boxed{b}$ The population size is $N = 50{,}000$ since $\frac{n}{N} = 0.001 \le 0.05$, it follows that $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1.4$

$\boxed{c}$ The population size is $N = 5000$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1.4$
since $\frac{n}{N} = 0.01 < 0.05$

$\boxed{d}$ The population size is $N = 500$ $\sigma_{\bar{x}} = \sqrt{\frac{450}{499}}\,(1.4) = 0.9496\,(1.4) = 1.33$
since $\frac{n}{N} = 0.1 > 0.05$