Chapter 8: Sampling Distributions

Section 8.1: Distribution of the Sample Mean

!!!,, HUGE theoretical!

SAMPLING DISTRIBUTION

Def The **sampling distribution** of a statistic is the distribution of all values of the statistic when all possible samples of the same size n are taken from the same population.

(typically represented as a probability distribution in the format of a table, histogram, or formula)

population mean: $\mu = \frac{1+2+3}{3} = 2$.

Ex1: Given three pool balls we will select two of the balls (with replacement) and find the average of their numbers.

$\bar{X}$ = mean of #s of two balls #S = 9

1 2 3 $S = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$

(a) Fill in the table to find $\bar{X}$ = the average of a sample of size two. (b) Fill in the table below using the data from (a).

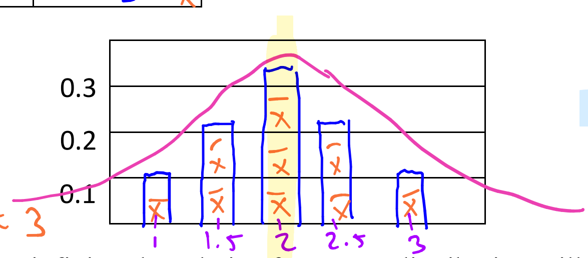
Outcome	Ball 1	Ball 2	Mean
1.	1	1	1 $=\bar{x}$
2.	1	2	1.5 $\bar{x}$
3.	1	3	2 $\bar{x}$
4.	2	1	1.5 $\bar{x}$
5.	2	2	2 $\bar{x}$
6.	2	3	2.5 $\bar{x}$
7.	3	1	2 $\bar{x}$
8.	3	2	2.5 $\bar{x}$
9.	3	3	3 $\bar{x}$

Mean	Frequency	Relative Frequency
1	1	1/9 = 0.111
1.5	2	2/9 = 0.222
2	3	3/9 = 0.333
2.5	2	2/9 = 0.222
3	1	1/9 = 0.111

$\Sigma f = 9$

(c) Draw the relative frequency distribution.

all samples of size n=2 from a population of N=3



Key approximately normal

As the number of samples approaches infinity, the relative frequency distribution will approach the sampling distribution.

$\mu = 2$

SAMPLING DISTRIBUTION OF THE MEAN

Def The **sampling distribution of the mean** is the distribution of sample means, with all samples having the same size n taken from the same population.

IMPORTANT NOTES

1. The sample means **target** the value of the population mean mean of sample mean ($\bar{x}$) = pop mean (μ)
(i.e. the mean of the sample means is equal to the population mean)

2. The distribution of the sample means tends to be a **normal distribution**.
(The distribution tends to become closer to a normal distribution as the sample size increases.)

Def An **unbiased estimator** is a statistic that targets the value of the corresponding population parameter in the sense that the sampling distribution of the statistic has a mean that is equal to the corresponding population parameter.

Good

Unbiased estimators: $\bar{x}$, $\hat{p}$, s^2 (sample mean, sample proportion, sample st.dev)

Biased estimators: Med, Range, s

Animation: http://onlinestatbook.com/stat_sim/sampling_dist/

SAMPLING DISTRIBUTION OF THE MEAN

Given

1. The random variable x has a distribution with mean μ and standard deviation σ.

   (the distribution may or may not be normal)

2. Simple random samples of size n are independent and selected from the population.

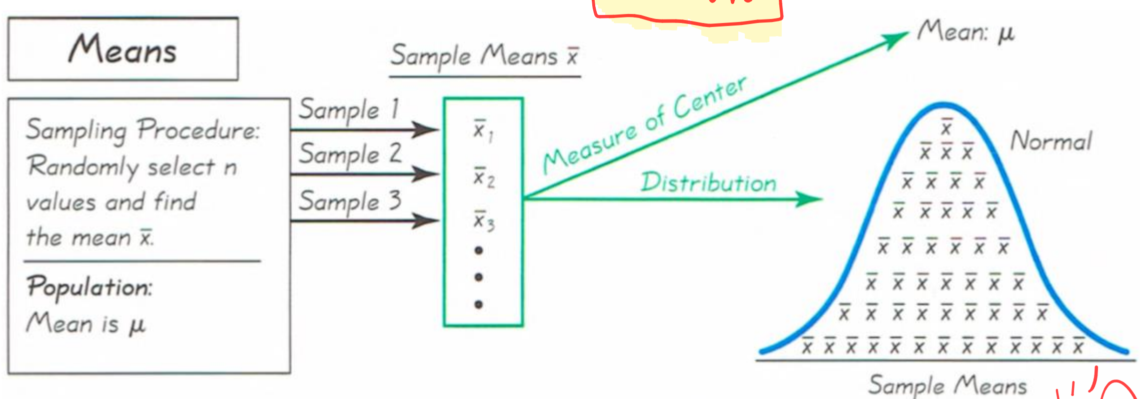
(SRS)

Conclusions

1. The distribution of all sample means $\bar{x}$ will approach a normal dist. as the sample size increases.
2. The mean of all sample means is the population mean: $\mu_{\bar{x}} = \mu$
3. The standard deviation of all sample means is given by $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

SAMPING DIST

as n increases, $\sigma_{\bar{x}}$ decreases!



$\bar{x}$ not x

Application of Result

4. The probability of a indep. SRS of size n having a (sample) mean between a and b is $P(a < \bar{x} < b)$.

We call $\sigma_{\bar{x}}$ the **standard error** of the **mean**.

Notes About Distributions

1. If the original distribution is normally distributed, then for any sample size n, the sample means will be normally distributed.
2. If the original population is not normally distributed then for samples of size $n \geq$ 30, the distribution of sample means can be approximated by a normal distribution.

IMPORTANT DISTINCTION: WHEN DO YOU USE A SAMPLING DISTRIBUTION?

We use the sampling distribution to compute probabilities of random samples of size $n > 1$.

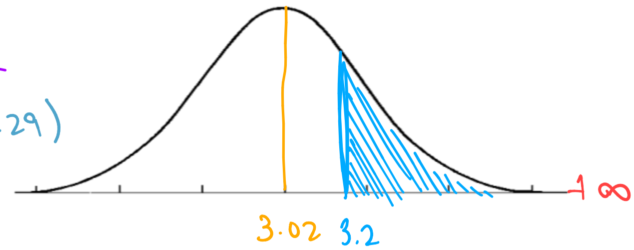
Ex2: The GPAs of all students enrolled at a large university have a normal distribution with a mean of 3.02 μ and a standard deviation of 0.29. σ

(a) Find the probability that one **randomly selected student** will have a GPA greater than 3.20.

→ same as previous section

$P(x > 3.20) = \text{normalcdf}(\overset{a}{3.2}, \overset{b}{1E99}, \overset{\mu}{3.02}, \overset{\sigma}{0.29})$

$= \boxed{0.267}$



(b) Find the probability that **25 randomly selected students** will have a mean GPA greater than 3.20.

$n = 25$ → use Sampling Dist ∇

$\mu_{\bar{x}} = \mu = 3.02$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.29}{\sqrt{25}}$

$P(\bar{x} > 3.20)$

$= \text{normalcdf}(\overset{a}{3.2}, \overset{b}{1E99}, \overset{\mu_{\bar{x}}}{3.02}, \overset{\sigma_{\bar{x}}}{\frac{0.29}{\sqrt{25}}})$

$= 9.56 \times 10^{-4} = \boxed{0.000956}$ "0"

different!

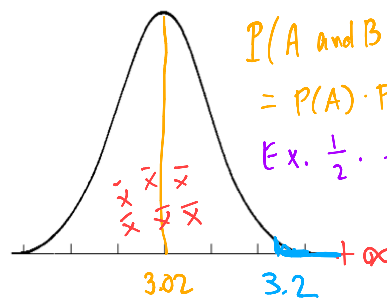
Reason? independent

$P(A \text{ and } B \text{ and } C \ldots)$

$= P(A) \cdot P(B) \cdot P(C) \cdots$

Ex. $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

$= 0.125$



(c) Find the probability that 10 randomly selected students will have a mean GPA between 2.90 and 3.10.

$\mu_{\bar{x}} = \mu = 3.02$

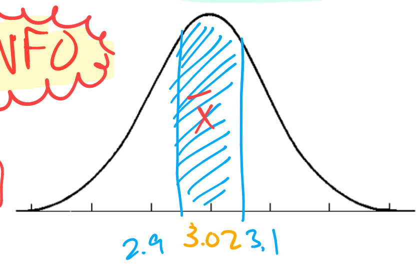
$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.29}{\sqrt{10}}}$

~~use normal~~? ~~n = 1~~

use Sampling? $n = 10$

{NEW SAMPLING INFO}

$P(2.9 < \bar{x} < 3.1) = \text{normalcdf}(\overset{a}{2.9}, \overset{b}{3.1}, \overset{\mu_{\bar{x}}}{3.02}, \overset{\sigma_{\bar{x}}}{\frac{0.29}{\sqrt{10}}}) = \boxed{0.713}$



Ex3: Go to the **Animation:** http://onlinestatbook.com/stat_sim/sampling_dist/

Create a crazy distribution that is very NON-normal. Write the *population* mean of your crazy distribution: $\mu =$ 13.06

Do 5 random samples, what is the mean? $n =$ 5, $\bar{x} =$ 11.26

Do another 5 random samples, what is the mean? $n =$ 10, $\bar{x} =$ 10.87

Do a total of 10 random samples, what is the mean? $n =$ ____, $\bar{x} =$ ______

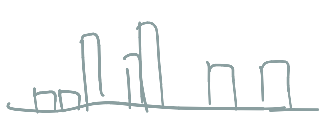
Do a total of 20 random samples, what is the mean? $n =$ 20, $\bar{x} =$ 12.26

Do a total of 40 random samples, what is the mean? $n =$ 40, $\bar{x} =$ 12.64

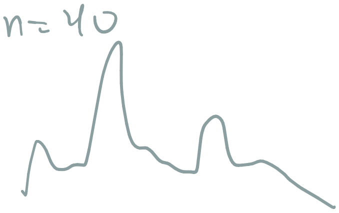
Do a total of 10,000 random samples, what is the mean? $n =$ 10k, $\bar{x} =$ 13.05

Sketch the resulting distribution of 20, 40, and 10,000 random sample. Overlay a normal distribution in your sketch.

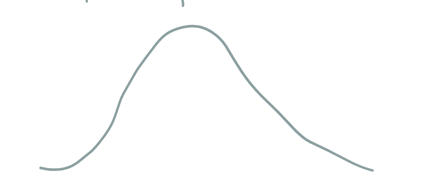
$n = 20$



$n = 40$



$n = 10{,}000$



Central Limit Theorem

Regardless of the shape of the underlying distribution X of the population,
the sampling distribution of $\bar{x}$ becomes *approximately* normal as the sample size, n, increases.

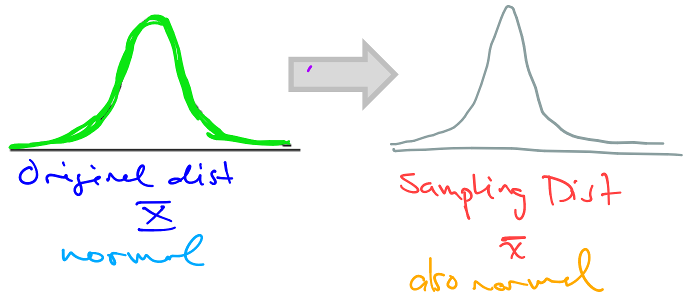
Ex4: Find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ for the given distributions.

(a) Given a normal distribution where $\mu = 10$, $\sigma = 3$ and $n = 9$.

Sampling

$\mu_{\bar{x}} = \mu = 10$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{9}} = \frac{3}{3} = 1$

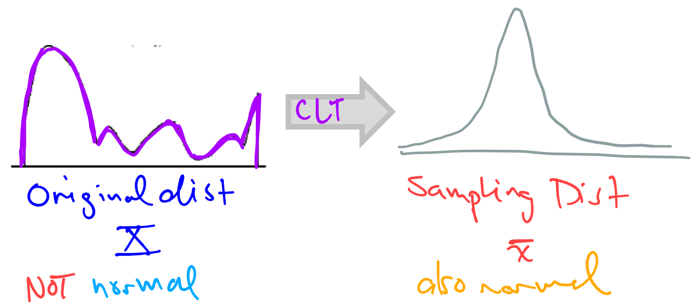


(b) Given a distribution with $\mu = 77$, $\sigma = 14$ and $n = 49$.

Sampling

$\mu_{\bar{x}} = \mu = 77$

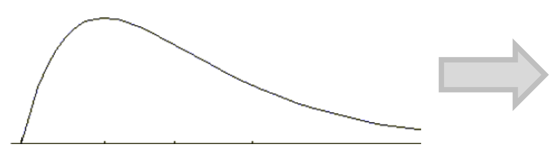
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = \frac{14}{7} = 2$



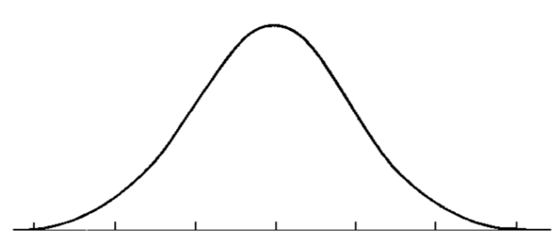
STEPS to Use Sampling Distribution for $\bar{x}$
Step 0: Are you selecting n people with n>1? If yes, continue:
1st: Check if $\bar{x}$ is Normal OR $n \geq 30$
2nd: Find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$
3rd: Use normalcdf for the rest…just don't forget to use information from *step 2!!*

Ex5: Let $x = CEO$ salaries (in thousands) where x is skewed right with $\mu = 139$, $\sigma = 45$.

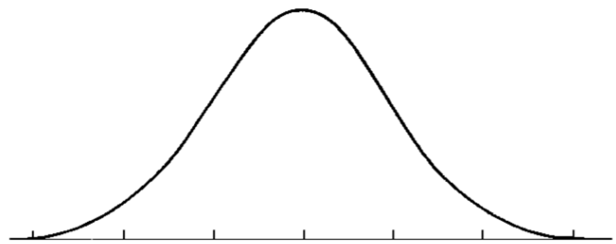
(a) If all possible random samples of 40 CEO salaries (in thousands) are taken, how would you describe the distribution of sample means? What would the standard deviation of the sample distribution be?



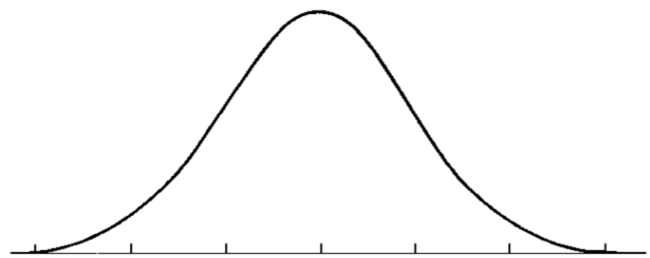
(b) What is the probability a sample of 40 CEOs make between 137 and 145 thousand dollars?



(c) Given a sample of 40 CEO's salaries, at what salary do the top 10% of CEO salaries begin at? $(M \rightarrow E)$



(d) What is the probability a sample of 40 CEO's makes less than 120 thousand dollars? Is this unusual?



WHY IS THE CLT SO POWERFUL?

In many situations knowing the shape of original distribution of X is impossible. By CLT, as long as we are willing to compute probabilities of samples of size 30 or more, *instead of a single individual*, we can use the normal distribution and compute the probabilities!