Chapter 6: Normal Probability Distributions

Section 6.1: The Standard Normal Distribution

Continuous Probability Distributions

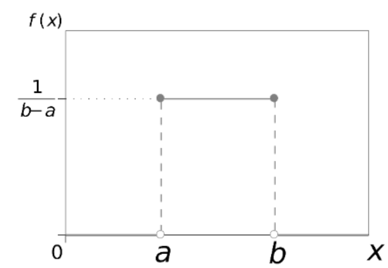
Def A ***density curve*** is the graph of a continuous probability distribution.

Requirements

1. The total area under the curve must equal 1. i.e. $\sum P(x) = 1$

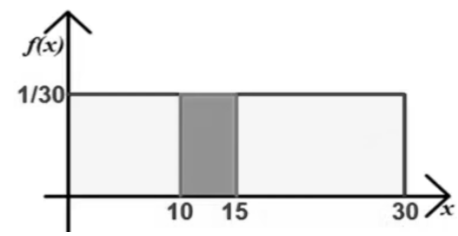
2. Every point on the curve must have a vertical height that is 0 or greater.

Uniform Probability Distribution



Ex: The bus to Union Station leaves every 30 minutes and is uniformly distributed. Find the probability that a randomly chosen person arriving at a random time will wait between 10 and 15 minutes?

$P(10 < x < 15) =$

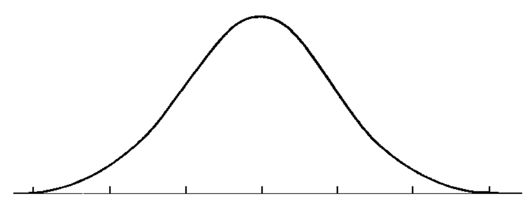


Normal Distributions

Def A continuous random variable has a normal distribution if its density curve is symmetric and bell-shaped.

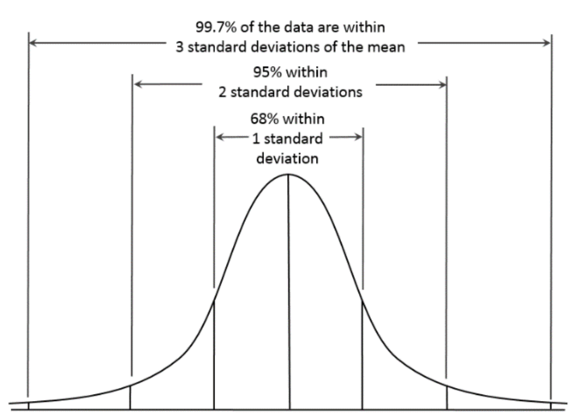
Specifically, the curve is given by: $y = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ *(Don't worry, we'll never use it.)*

Ex: The weights of all firefighters are normally distributed with a mean of 200 lbs and a standard deviation of 7 lbs. What's the probability that a randomly chosen firefighter weighs between 185 and 195 lbs?



Standard Normal Distribution

Def The **standard normal distribution** is a normal probability distribution with $\mu = 0$ and $\sigma = 1$.



Important Notes

1. The z-score is used on the horizontal axis.
2. The area of the region under the curve is equal to the associated probability of occurrence.

Two Ways to Find Area

1. Use Table A-2.

   Look up the area under the curve that lies to the left of z-score (may first need to convert data to z-score).

2. Use Graphing Calculator (TI-84 Plus)

   (a) $2^{nd} \Rightarrow$ VARS $\Rightarrow$ DISTR

   (b) normalcdf$\left(lower, upper, \mu, \sigma \right)$

Two Ways to Find Z-score

1. Use Table A-2.

   Look up the z-score associated with the area that lies to left.

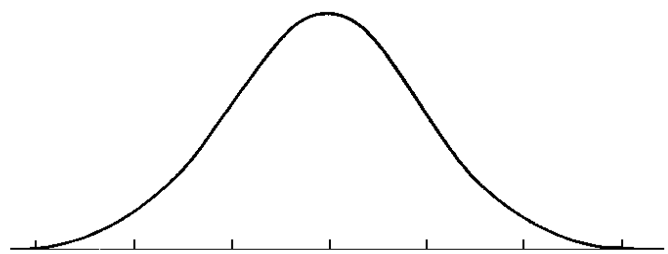
2. Use Graphing Calculator (TI-84 Plus)

   (a) $2^{nd} \Rightarrow$ VARS $\Rightarrow$ DISTR

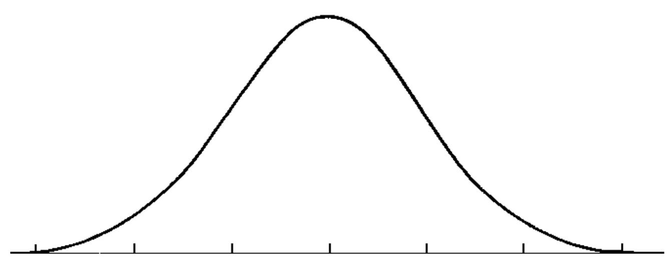
   (b) invNorm $\left(area, \mu, \sigma, Tail \right)$

Ex: Find the probability given the following z-scores for a standard normal distribution.

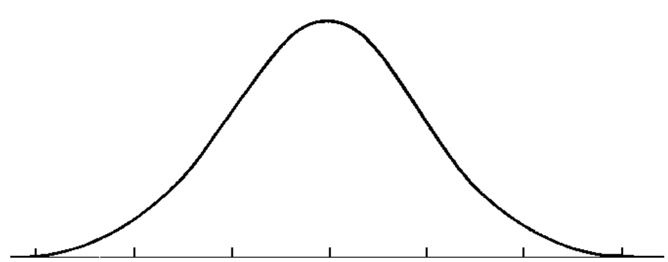
(a) $P(z < 1.35)$



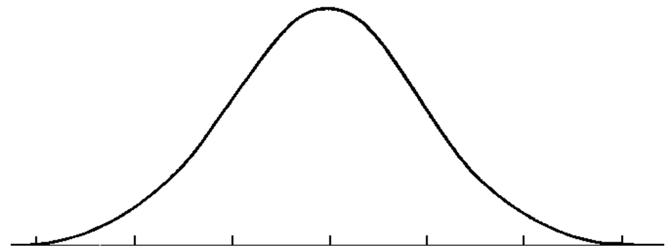
(b) $P(z > 0.68)$



(c) $P(-2.43 < z < 0.88)$

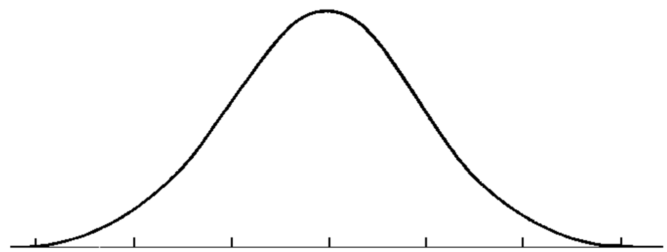


Ex: Find the z-score associated with the 15th percentile.



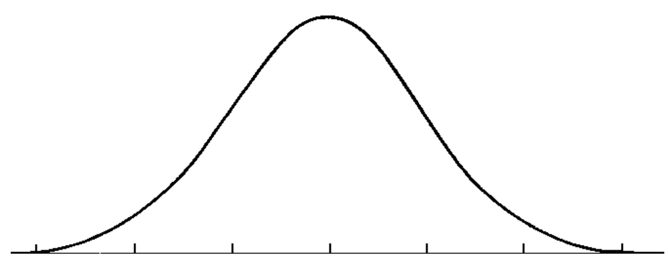
Ex: Find the probability given the following z-scores for a standard normal distribution.

(a) $P(z < -0.55)$



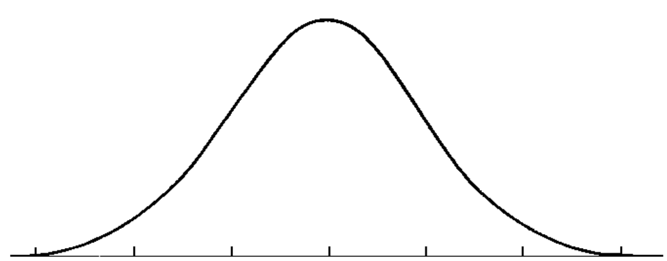
(b) Would $P(z \leq -0.55)$ differ from (a)?

Ex: Find the z-scores that separate the top 10% and bottom 10% of all values.



***Specific Notation**: z_α is the **critical value** that denotes a z-score with an area of α to its _______________ .

Ex: Find $z_{0.05}$



Section 6.2: Real Applications of Normal Distributions

Z Scores

Def A **z score** is the number of standard deviations that a given value x is above or below the mean.

FORMULA: Sample: $\boxed{z = \frac{x - \bar{x}}{s}}$ Population: $z = \frac{x - \mu}{\sigma}$

Also Given z-score find score (x)

$$\boxed{x = \bar{x} + z \cdot s}$$

Round-Off Rule: Round z scores to two decimal places.

Ex: Consider your height in inches. Calculate the standardized value (z-score) for your height given that in the United States the average height for women is 63.7 inches with a standard deviation of 2.7 inches and for men is 69.1 inches with a standard deviation of 2.9 inches. *Would you be considered tall for your gender?*

$z_{height} =$

x = varies by person

x = 68 inches (Jorge)

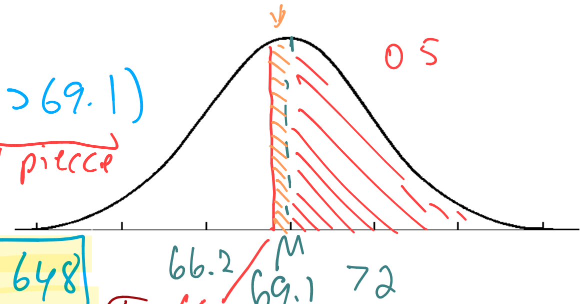
$z = \frac{68 - 69.1}{2.9} = \boxed{-0.38}$

x = height of men in US

What is the probability that someone of your gender is taller than you?

$P(x > 68) = \underbrace{P(68 < x < 69.1)}_{\text{orange piece}} + \underbrace{P(x > 69.1)}_{\text{red piece}}$

$= \text{normalcdf}(68, 69.1, 69.1, 2.9) + 0.5 = \boxed{0.648}$



Ex: The average for the statistics exam was 75 and the standard deviation was 8. Andrey was told by the instructor that he scored 1.5 standard deviations below the mean, and the scores were normally distributed.

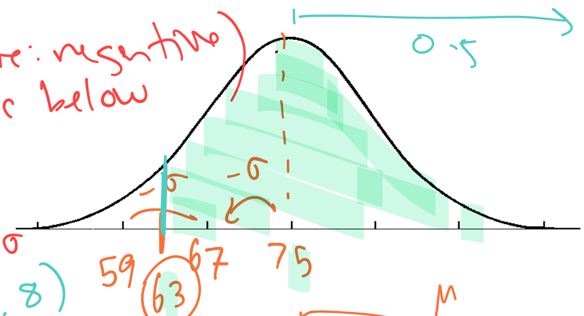
What is Andrey's score? What percentage of students scored higher?

Use $x = \bar{x} + zs$ $z = -1.5$ (note: negative b/c below)

$= 75 + (-1.5)(8)$

$= \boxed{63}$

$P(x > 63) = 0.5 + \text{normalcdf}(63, 75, 75, 8) = 0.933$ $\boxed{93.3\%}$



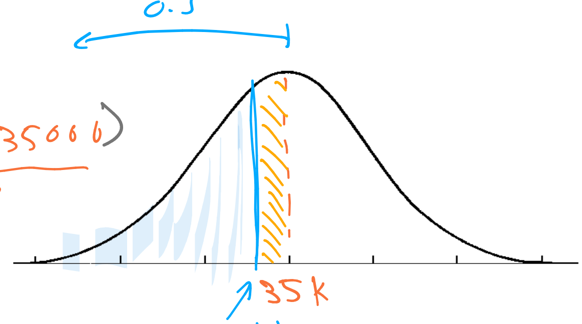
Ex: The life spans of a brand of automobile tires are normally distributed with a mean life span of 35,000 miles and a standard deviation of 2250 miles. The life span of a randomly selected tire is 34,000 miles. Find the z-score of this tire. Can you find the probability that a randomly selected automobile tire has a life-span less than or equal to 34,000 miles?

x = lifespan of tires in miles.

x = 34,000 mi

↳ z-score: $z = \frac{x - \bar{x}}{\sigma} = \frac{(34000 - 35000)}{2250}$

$z = \boxed{-0.44}$



$P(x \leq 34000) = 0.5 - P(34k < x < 35k)$

$= 0.5 - \text{normalcdf}(34000, 35000, 35000, 2250) = \boxed{0.328}$

When to use **Normalcdf**? → find prob	When to use **InvNorm**? → Find x-values (or z-score)
To find a probability when the <u>data</u> (μ, σ, x) are given. a, b	When the percent or area or probability is given and we are trying to find the <u>corresponding x-values.</u>

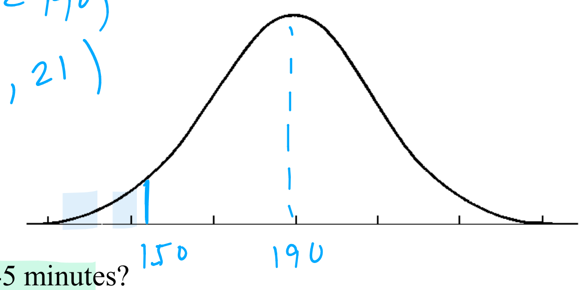
<u>Ex</u>: The completion times to run a road race are normally distributed with a mean of 190 minutes and a standard deviation of 21 minutes.

(a) What is the probability that a randomly selected runner will finish the race in less than 150 minutes?

$P(x<150) = P(x<190) - P(150<x<190)$

$= 0.5 - \text{normalcdf}(150, 190, 190, 21)$

$= \boxed{0.0284}$

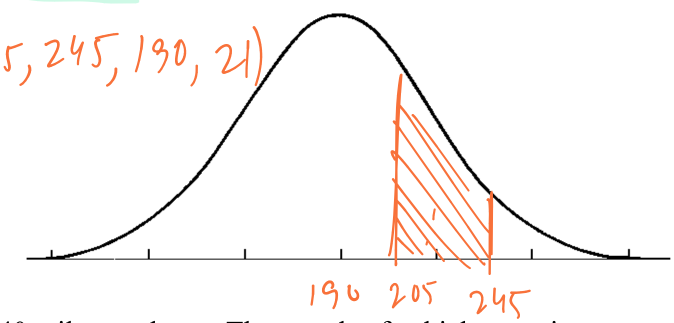


(b) What percentage of runners will finish the race between 205 and 245 minutes?

$P(205<x<245) = \text{normalcdf}(205, 245, 190, 21)$

$= 0.233$

$= \boxed{23.3\%}$



<u>Ex</u>: A construction zone on a highway has a posted speed limit of 40 miles per hour. The speeds of vehicles passing through this construction zone are normally distributed with a mean of 46 mph and a standard deviation of 4 mph.

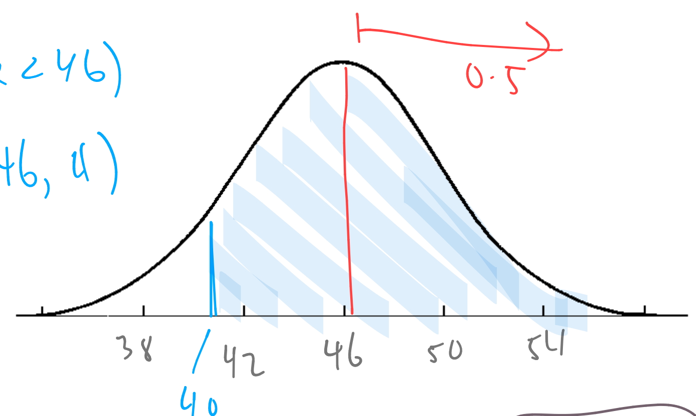
(a) What percentage of vehicles exceed the speed limit?

$P(x>40) = P(x>46) + P(40<x<46)$

$= 0.5 + \text{normalcdf}(40, 46, 46, 4)$

$= 0.933$

$\boxed{93.3\%}$



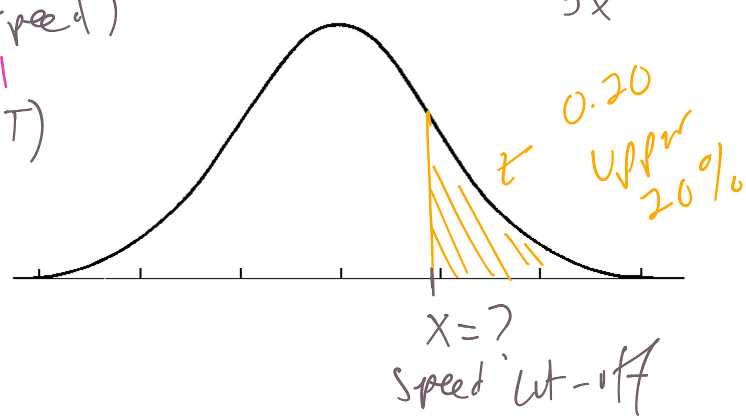
(b) If the police wish to ticket only those drivers whose speed falls in the upper 20th percentile, what is the minimum speed of a driver that will be ticketed?

$\le x$

Given $p = 0.2$, find x-value (speed)

$x = \text{invNorm}(\underset{p}{0.2}, \underset{\mu}{46}, \underset{\sigma}{4}, \underset{\text{tail}}{\text{RIGHT}})$

$= \boxed{49.4 \text{ mph}}$



Section 6.3: Sampling Distributions and Estimators

nuts! huge! for datasets.

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)

nuts! theoretical

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

1 2 3

all samples of 2 pool balls {11, 12, 13, ...}

$\bar{X}$ = mean # of 2 selected balls.

(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 $\bar{x}$

all total

Sample Distribution

NEW

Sampling Distribution

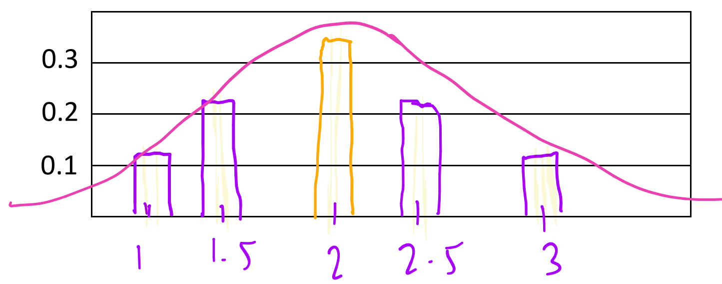
Mean	Frequency	Relative Frequency
1	1	$0.11\bar{1}$
1.5	2	$0.22\bar{2}$
2	3	$0.33\bar{3}$
2.5	2	$0.22\bar{2}$
3	1	$0.11\bar{1}$

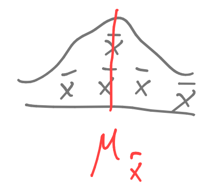
1/9
2/9
3/9
2/9
1/9

(c) Draw the relative frequency distribution.

- Looks approx normal
- Symmetric

Statistic: mean $\bar{x}$ (sample)





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

↑ theoretical

SAMPLING DISTRIBUTION OF THE MEAN

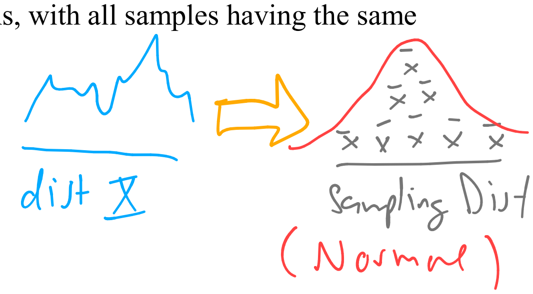
Ex above: pop N = 3, sample n = 2

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
   (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.

Facts to know

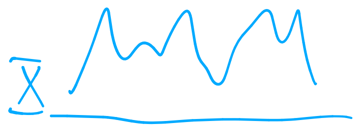
Unbiased estimators: $\hat{p}$, $\bar{x}$, and s^2 (Good!)

$\hat{p}$ = "p hat" = sample proportion

Biased estimators: Med, range, and s

GTQ!

Section 6.4: The Central Limit Theorem



CENTRAL LIMIT THEOREM

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 selected from the population. (SRS)

Conclusions

1. The distribution of sample means $\bar{x}$ will approach a normal dist. as the sample size increases. (n gets big!)
2. The mean of all sample means is the population mean (μ)

   $\mu_{\bar{x}}$

   $$\boxed{\mu_{\bar{x}} = \mu}$$ (always)

3. The standard deviation of all sample means is given by

   $\sigma_{\bar{x}}$ = "st. dev of sampling dist"

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

   Notice $\sigma_{\bar{x}}$ changes as n changes!

Animation: http://onlinestatbook.com/stat_sim/sampling_dist/ (Simulation app)

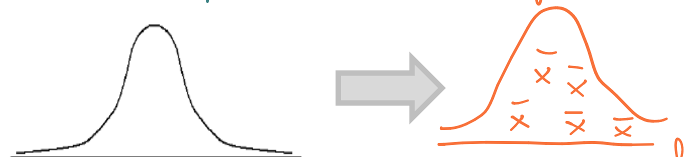
Notes About Distributions

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

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

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

OG X normal



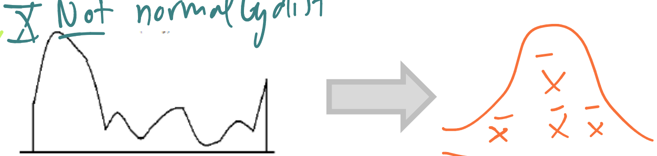
$\boxed{\mu_{\bar{x}} = \mu = 10}$ Sampling mean

notice $\cancel{9 > 30}$

$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3/\sqrt{9} = 3/3 = 1$ $\boxed{\sigma_{\bar{x}} = 1}$ Sampling st. dev.

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

X Not normally dist



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

$49 > 30$ ✓

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

STEPS to Use Sampling Distribution for $\bar{x}$
1st: Check if $\bar{x}$ is Normal (or is $n \geq 30$?) (check requirements)
2nd: Find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ (compute)
3rd: Use normalcdf for the rest…just don't forget to use information from step 2 (just like 6.1, 6.2)

can use any sample size!

Ex: 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.

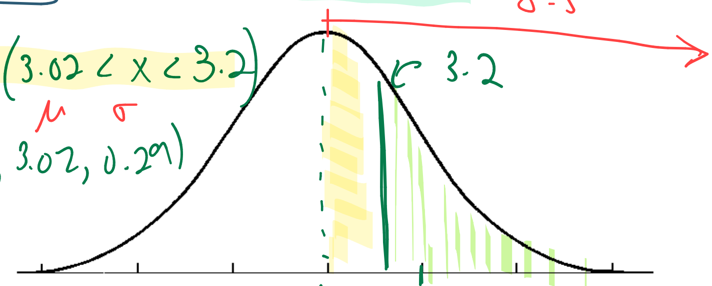
original way {6.1, 6.2

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

$$P(x > 3.20) = P(x > 3.02) - P(3.02 < x < 3.2)$$

$$= 0.5 - \text{normalcdf}(\overset{a}{3.02}, \overset{b}{3.2}, \overset{\mu}{3.02}, \overset{\sigma}{0.29})$$

$$= \boxed{0.267}$$



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

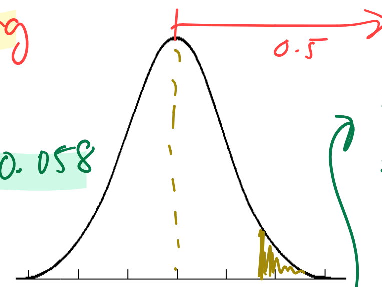
Note b/c using Sampling

$n = 25 \to$ use Sampling

- $\mu_{\bar{x}} = \mu = 3.02$
- $\sigma_{\bar{x}} = \sigma/\sqrt{n} = \dfrac{0.29}{\sqrt{25}} = 0.058$

$$P(\bar{x} > 3.2) = P(\bar{x} > 3.02) - P(3.02 < \bar{x} < 3.2)$$

$$= 0.5 - \text{normalcdf}(\overset{a}{3.02}, \overset{b}{3.2}, \overset{\mu_{\bar{x}}}{3.02}, \overset{\sigma_{\bar{x}}}{0.058})$$



$= 9.5646679E-4$

$= 9.5646679 \times 10^{-4}$

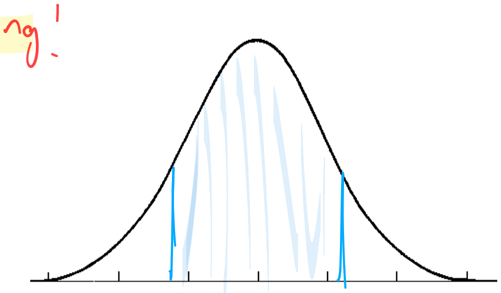
$= 0.0009564\ldots$

$= \boxed{0.000956}$

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

$n = 10 \to$ use sampling!

answer



$$P(2.90 < \bar{x} < 3.10) =$$

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

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

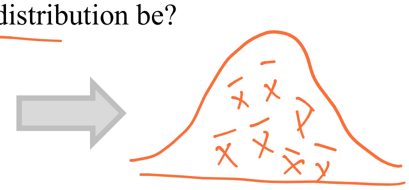
(a) If all possible random samples of 40 CEO salaries (n) (in thousands) are taken, how would you describe the distribution of sample means? (Q1) What would the standard deviation of the sample distribution be? (Q2)

Q1: ↳ becomes normal! by Central Limit Theorem

$n = 40 > 30$? ✓

not normal dist

↳ need $n > 30$



Q2: $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 45/\sqrt{40} = 7.115\ldots \approx \boxed{7.1 \text{ thousand dollars}}$

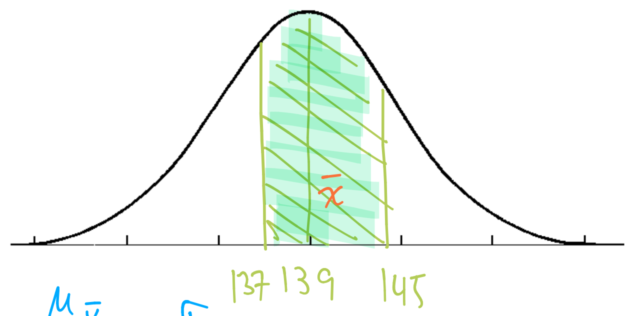
(b) What is the probability a sample of 40 CEOs make between 137 and 145 thousand dollars?

↳ sampling

orig dist $\begin{cases} \mu = 139 \\ \sigma = 45 \end{cases}$

Sampling: $\mu_{\bar{x}} = \mu = 139$

$\sigma_{\bar{x}} = \sigma/\sqrt{40} = 45/\sqrt{40}$

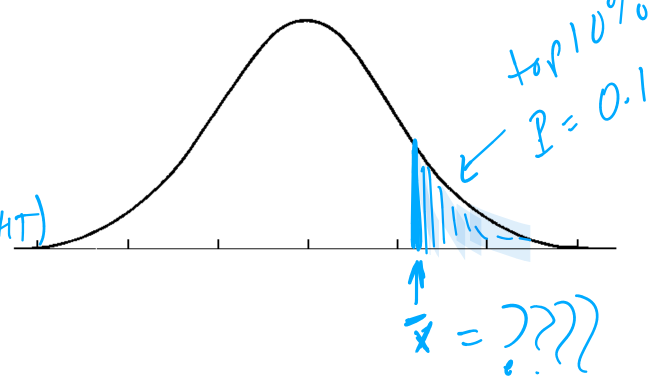


$$P(137 < \bar{x} < 145) = \text{normalcdf}(\overset{a}{137}, \overset{b}{145}, \overset{\mu_{\bar{x}}}{139}, \overset{\sigma_{\bar{x}}}{45/\sqrt{40}})$$

$$= \boxed{0.411}$$

(c) Given a sample of $\boxed{\text{40 CEO's salaries}}$, at what salary does the top 10% of CEO salaries begin at?

↳ use sampling



$$\bar{x} = \text{invNorm}(p, \mu_{\bar{x}}, \sigma_{\bar{x}}, \text{tail})$$

$$= \text{invNorm}(0.1, 139, 45/\sqrt{40}, \text{RIGHT})$$

$$= 148.118....$$

$\bar{x} = 148.1$

(M→E) "with a sample of 40 CEO salaries, the top 10% of their salaries begins at 148.1 thousand dollars."

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