Chapter 5 Normal Probability Distributions

STA 2023 SECTION 5.1, 5.2, and 5.3 Introduction to Normal Distributions and the Standard Normal Distributions

Gaussian dist.

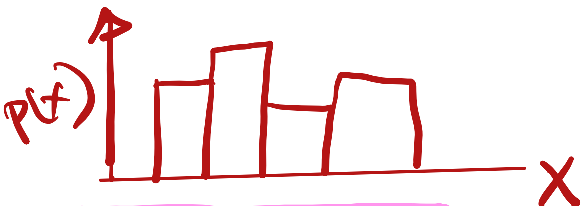
CTS prob dist.

Learning Outcomes:

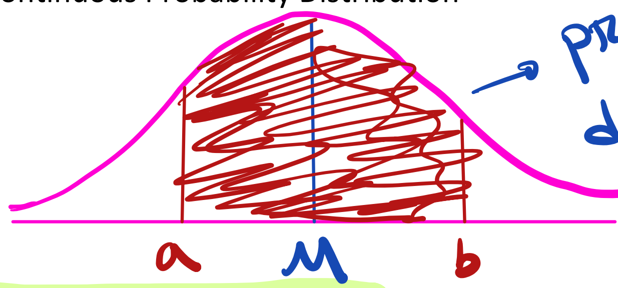
1) Interpret graphs of normal probability distributions
2) Find areas under the standard normal curve
3) Find a *z*-score given the area under the normal curve
4) Find a specific data value of a normal distribution given the probability

Review : prob histogram

Discrete Probability Distribution –



Continuous Probability Distribution –



prob density curve

<u>Probability Density Curve</u>

- Represents the probability distribution of a **continuous** random variable
- The area under the curve is equal to 1 or 100%
- The area under the curve between 2 values has 2 interpretations:
  - The area represents the proportion of the values that fall in this interval
  - The area represents the probability that a particular value falls in this interval

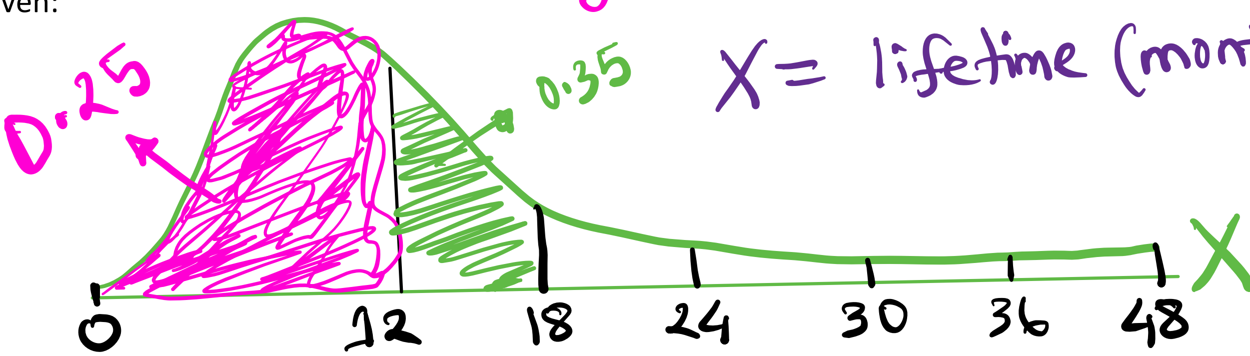
The following probability density curve represents the lifetime, in months, of a certain brand of laptop battery.

Area = Probability = Proportion = % of values

CTS

Given:

X = lifetime (month)



a. Find the proportion of batteries with lifetimes between 12 and 18 months.

proportion of batteries = 35% = 0.35

b. Find the proportion of batteries with lifetimes less than 18 months.

$P(X < 18) = 0.25 + 0.35 = 0.60$

c. What is the probability that a randomly chosen battery lasts more than 18 months?

$P(X > 18) = 1 - 0.60 = 0.40$

Note: For any continuous random variable, the area under a specific value of x is equal to zero, so $P(a < x < b)$ *is equivalent to* $P(a \leq x \leq b)$.

CTS dist $\Rightarrow P(a < X < b) = P(a \leq X \leq b)$

The Normal Distribution

- a special type of probability density curve
- has properties:
  - the mean is located at the center of the bell-shaped curve
  - the mean, median and mode are all equal
  - the graph is symmetric about the mean
  - the horizontal axis represents the continuous random variable
  - the curve never touches the horizontal axis
  - the AREA under the curve is equal to 1 or 100%
  - the AREA under the curve can represent a proportion of the population of values OR a probability of a particular value falling in an interval

location-scale family of dist

μ = location parameter = mean

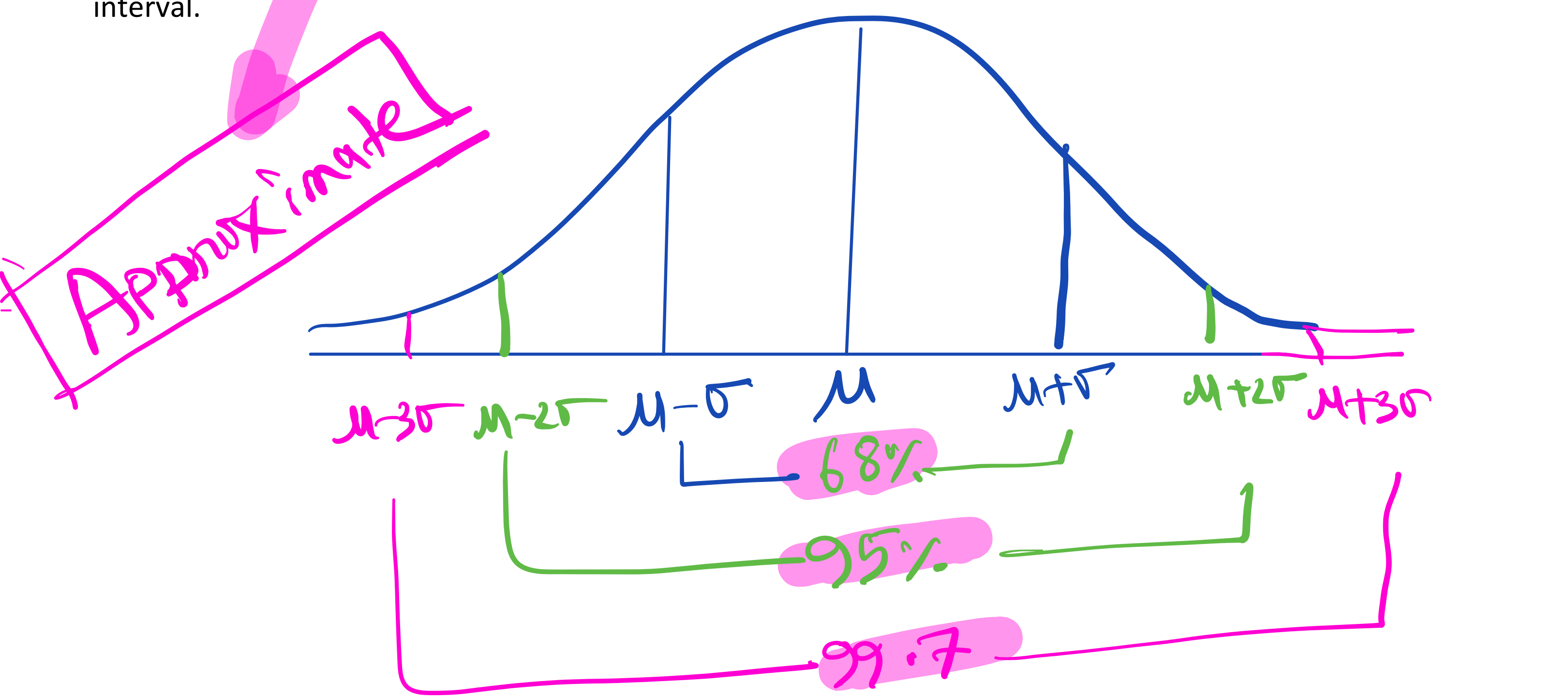
σ = scale parameter = st. dev.

The Standard Normal Distribution

- the population mean is $\boxed{\mu = 0}$ and the population standard deviation is $\boxed{\sigma = 1}$.
- The horizontal axis represents the continuous random variable z .

$Z \sim N(0, 1)$ | X = non-Standard Normal dist $X \sim N(\mu, \sigma)$

Recall: The Empirical Rule – the percent represents the portion of the population values that fall in this range or interval , but can also represent the probability that a given $z - score$ falls in this range or interval.



R.V. $X \sim N(\mu, \sigma)$

μ, and σ will be given

R.V. $Z \sim N(0, 1)$

* normalcdf(LB, UB, mean, St.dev.)

LB → start, UB → end

To find Area = prob^y = proportion = % of values

* invNorm(AL, mean, St.dev.)

To find Z-score or percentile value for a "given" prob^y

$1E99 = \infty$

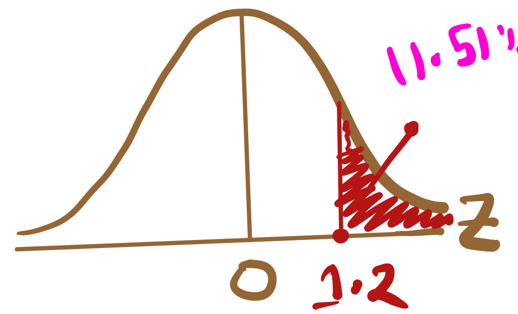
→ 2nd comma

Find the area under the curve for the following:

a. $z > 1.2$

$= \text{normalcdf}(1.2, 1E99, 0, 1)$

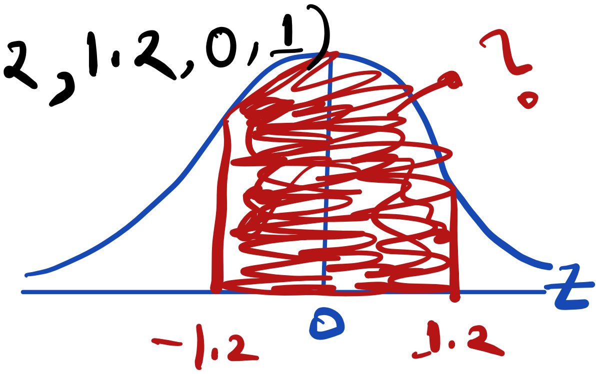
$\boxed{= 0.1151}$



b. $-1.2 < z < 1.2$

$P(-1.2 < Z < 1.2) = \text{normalcdf}(-1.2, 1.2, 0, 1)$

$= \boxed{0.7699}$



c. $z < -1.2$

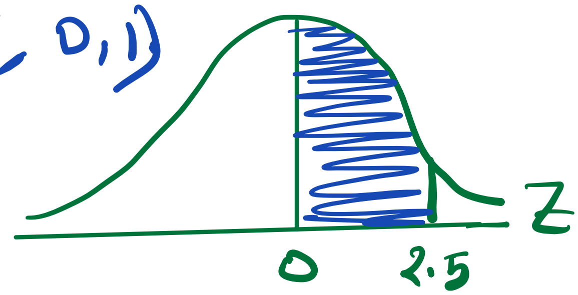
$P(Z < -1.2) = \text{normalcdf}(-1E99, -1.2, 0, 1)$

$\boxed{= 0.1151}$

d. $0 \leq z \leq 2.5$

$P(0 \leq Z \leq 2.5) = \text{normalcdf}(0, 2.5, 0, 1)$

$= 0.4938$

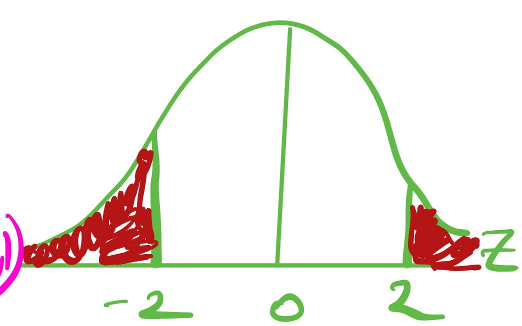


e. *Area that lies more than 2 standard deviations from the mean*

$P(Z > 2$ or $P\ Z < -2)$

$= P(Z > 2) + P(Z < -2)$

$= 2P(Z > 2 = 2 \cdot \text{normalcdf}(2, 1E99, 0, 1)$



$= 0.0455$

Alternative

$P(Z > 2 \text{ or } Z < -2)$

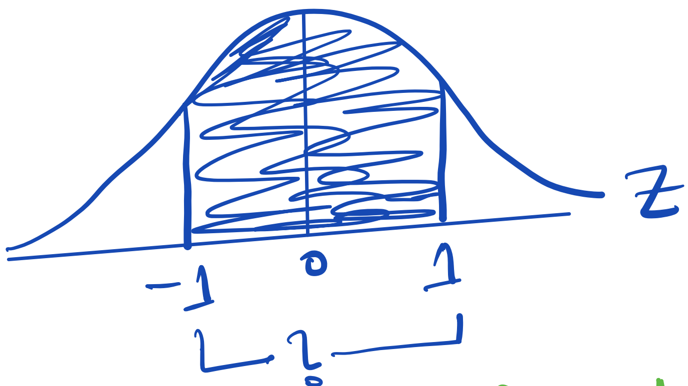
$= 1 - P(-2 < Z < 2)$

$= 0.0455$

Empirical Rule

within 1 st.dev.

68%



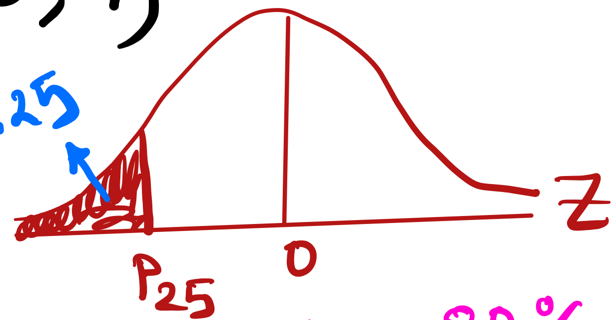
normalcdf $(-1, 1, 0, 1) = 0.6827$

2nd VARS → [3] invNorm( )

Finding a $z-score$ from a known area or probability: invNorm( )

① P_{25} = invNorm(.25, 0, 1) of Z

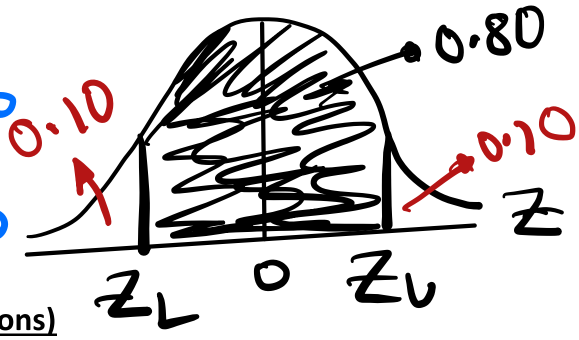
= −0.6745



② Find Z-score responsible for middle 80% area.

Z_L = invNorm(0.10, 0, 1) = −1.2816

Z_U = invNorm(0.90, 0, 1) = 1.2816



Applications of Normal Distributions (Nonstandard Normal Distributions)

Standard Normal Distribution:

$Z \sim N(0,1)$

Nonstandard Normal Distribution:

$X \sim N(\mu, \sigma)$

*We can use normal distributions to represent a percent of the population OR a probability that an $x-value$ falls in an interval.

The heights of men have a normal distribution with a mean of 69.2 inches and standard deviation of 2.8 inches.

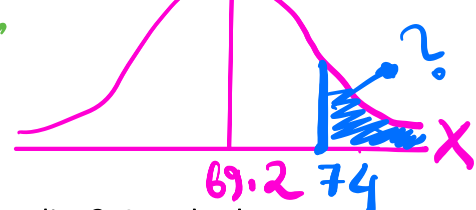
$\mu = 69.2$, $\sigma = 2.8$

a. What is the probability that a randomly chosen man is taller than 74 inches?

R.V. X = Heights of men $\sim N(69.2, 2.8)$

$P(X > 74)$ = normalcdf(74, 1E99, 69.2, 2.8)

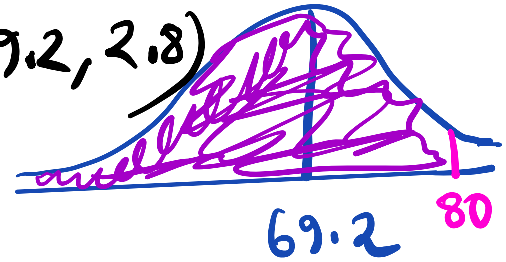
$= 0.0432 = 4.32\%$



b. What percentage of men will fit through a standard doorway without bending? Standard doorways are set at 6’8” or 80 inches.

$P(X < 80) = \text{normalcdf}(-1E99, 80, 69.2, 2.8)$

$= 0.9999$



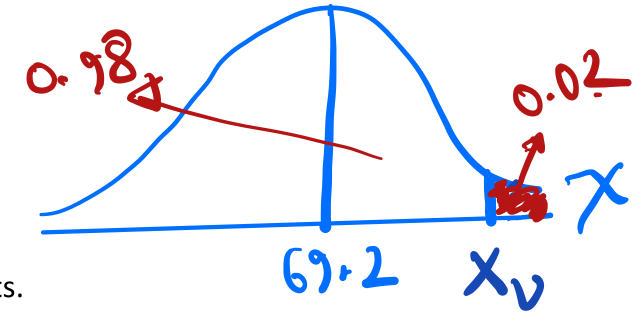
c. Find the 25[th] percentile for the heights of men.

$P_{25} = 67.31$ inch $= \text{invNorm}(\quad)$

d. If *Tall Clubs International* wants to include only the tallest 2% of men as members, what is the height requirement to join?

$P_{98} = \text{invNorm}(.98, 69.2, 2.8)$

$= 74.95$

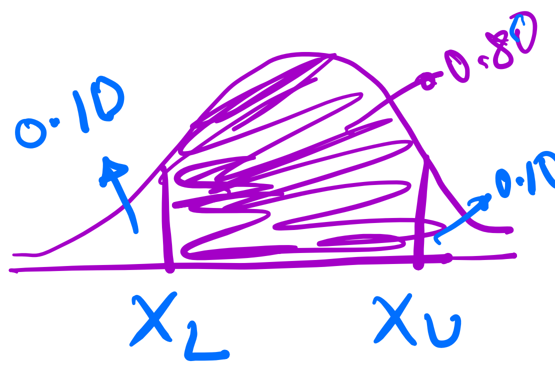


e. Find the range of values for the middle 80% of men’s heights.

HW

$X_L =$

$X_U =$



Notes:

Any normal distribution can be transformed into the standard normal distribution by transforming each **x-value** to its **z-score**. This means we can find the **area** under the curve of any normal distribution using the standard normal distribution using the same method discussed in section 5.1.

$$X \sim N(\mu, \sigma) \Rightarrow \frac{X - \mu}{\sigma} = Z \sim N(0,1)$$

$X \sim N(22, 9)$

Example 1: A national study found that college students with jobs worked an average of 22 hours per week. The standard deviation is 9 hours. Assume that the lengths of time college students work are normally distributed and are represented by the variable *x*. *(Adapted from Sallie Mae/Ipsos Public Affairs)*

a) If a college student with a job that is selected at random, what is the probability that they work more than 28 hours a week? Round to three decimal places.

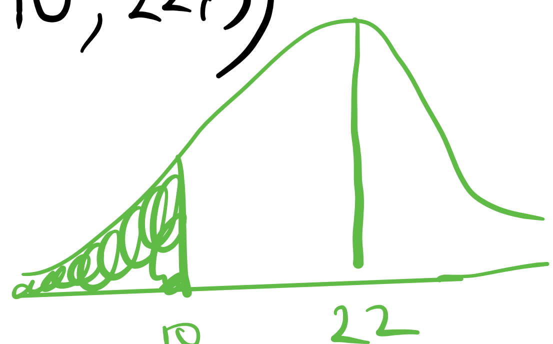
$P(X > 28) = \text{normalcdf}(28, 1E99, 22, 9)$

$= 0.2525$

b) A college student with a job is selected at random. Find the probability that the student works for less than 10 hours per week. Round to three decimal places.

$P(X < 10) = \text{normalcdf}(-1E99, 10, 22, 9)$

$= 0.0912$



c) What is the probability that a student selected at random will work between 18 and 24 hours per week? Round to three decimal places.

$P(18 < X < 24)$

$= \text{normalcdf}(18, 24, 22, 9) = 0.2596$

Example 2: The time it takes to receive a food delivery is normally distributed with a mean of 30 minutes and a standard deviation of 6 minutes.

$X \sim N(30, 6)$

a) What percentage of delivery times will be less than 20 minutes? Round to four decimal places.

$P(X < 20) = \text{normalcdf}(-1E99, 20, 30, 6)$

$= 0.0478$

b) What percentage of delivery times will be greater than 35 minutes? Round to four decimal places.

HW

c) What percentage of delivery times will take between 28 and 38 minutes? Round to four decimal places.

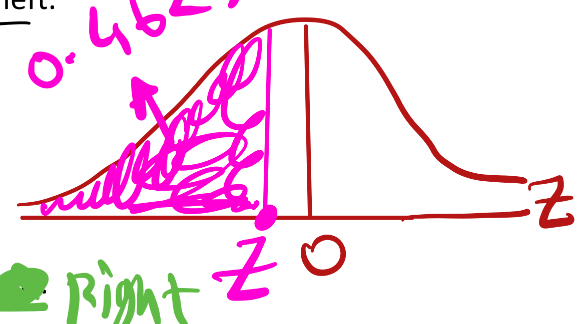
HW

$\boxed{AL = 1 - AR}$

Notes:
Finding the value for x or z-score that corresponds to an area under the normal curve:

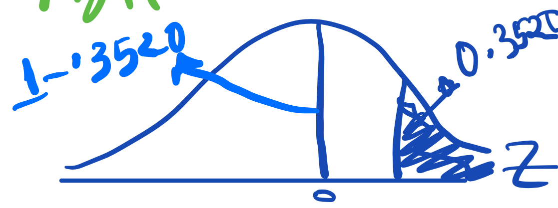
Example 3: Find a *z*-score(s) for the given area. Round to two decimal places.

a) Find the *z*-score that has 46.27 % of the distribution's area to its left.



$Z = \text{invNorm}(0.4627, 0, 1)$
$= -0.0936$

b) Find the z-score that has 35.20% of the distribution's area to its left. Right



$Z = \text{invNorm}(0.648, 0, 1)$

= 0.3799

CLT

STA 2023 SECTION 5.4 SAMPLING DISTRIBUTIONS AND THE CENTRAL LIMIT THEOREM NOTES

Learning Outcomes:

1) Find sampling distributions and verify their properties
2) Interpret the Central Limit Theorem
3) Apply the Central Limit Theorem to find the probability of a sample mean

Recall: A **population** is set of similar items or events which is of interest for some question or experiment. A **parameter** is a numerical description of a **population characteristic**. A **sample** is a subset of a population. A **statistic** is a numerical description of a **sample characteristic**.

A Sampling distribution is the probability distribution of a sample statistic that is formed when random samples of size *n* are repeatedly taken from a population and the same sample statistic is calculated from each sample.

If the sample statistic of a sampling distribution is the sample mean, then the distribution is a **sampling distribution of sample means.** Every sampling statistic has a sampling distribution.

Consider a population with mean μ, and standard deviation σ. If you take random samples of size n from that population and then compute the **mean** of each sample, the means of each sample together form a ____.

$$\boxed{\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Properties of Sampling Distributions of Sample Means:

1) The mean of the sample means, denoted $\mu_{\bar{x}}$, is equal to the population mean μ.

2) The standard deviation of the of the sample means, denoted $\sigma_{\bar{x}}$,, is called the Standard error and is found using the following formula: $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

The Central Limit Theorem is an important result in inferential statistics.

The Central Limit Theorem: If x has a distribution (may or may not be normal) with *mean of* μ *and standard deviation of* σ, and if simple random samples all of the same size n are taken from the population in such a way that all possible samples of size n have the same chance of being selected, then:

- If the original population (x) has a normal distribution, then the distribution of sample means $(\bar{x})$ is also normally distributed for *any sample size* n.

Normal produce Normal, we don't care

about sample size

- If the original population (x) **does not** have a normal distribution, then $\boxed{\textbf{sample size } n > 30}$ is needed to approximate the normal distribution. The approximation gets better as n gets larger.

- The standard deviation of the sample means $(\sigma_{\bar{x}})$ is smaller than the standard deviation of the population (σ) and is often called the *standard error of the means*.

$\sigma_{\bar{x}} < \sigma$

- Standard error of the means: $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$ = Standard devition of sample mean

The Central Limit Theorem allows you to treat a sampling distribution of sample means like a normal distribution (when the sample size is 30 or more or the population is itself normally distributed).

Since any normal distribution can be transformed into a standard normal distribution, you can find the area under a sampling distribution of sample means curve using the same steps discussed in section 5.1

1. If $X \sim$ Normal dist. then $\bar{x} \sim$ Normal dist

or

2. If $X \nsim$ Normal we must have $n > 30$

$\bar{x} \overset{CLT}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Example 1: A study analyzed the sleep habits of college students. The study found that the mean sleep time was 6.9 hours, with a standard deviation of 1.5 hours. Random samples of 100 sleep times are drawn from this population, and the mean of each sample is determined to create a sampling distribution of sample means.

a) Is the sampling distribution of sample means from samples of size 100 normally distributed? Why or why not?

b) Find the probability that the mean sleep time of a sample of size 100 will be larger than 7 hours.