Prob/Stat/Discrete **Name**________________________
Unit 6 Guided Notes

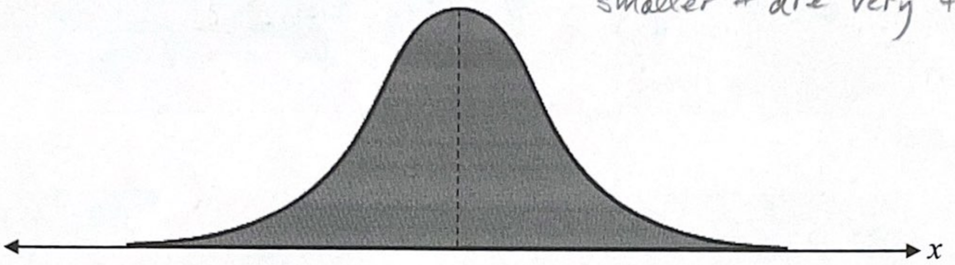
6.1: Introduction to Normal Distributions

Objectives

1. **Can you interpret graphs of normal probability distributions?**
2. **Can you create diagrams of a standard normal curve?**
3. **Can you find areas under the standard normal curve?**
4. **Can you use technology to find areas when given a *z*-score?**

What is normal? Human behavior and characteristics often follow a "normal" pattern, where a large number of people have similar behavior in the center, and a smaller number of people have unusual behavior.

- Industry & business : lifetime of TV, housing costs
- Animals, chemicals, physical properties often demonstrate normal behavior.
- Consider the heights of adult females... : lots have similar height in center
  smaller # are very tall or very short



Recall : Unit 5 was Discrete Prob. Distr

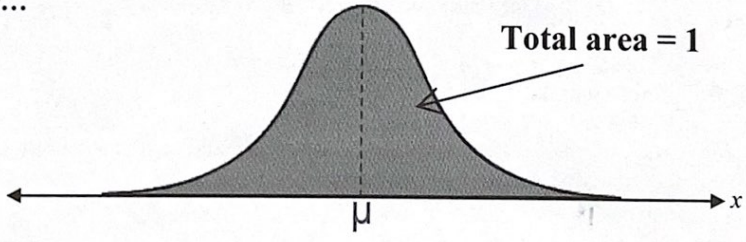
Normal Distribution

- A continuous probability distribution for a random variable, *x*.
- The most common continuous probability distribution in statistics.
- The graph of a normal distribution is called the normal curve.

Properties of Normal Distribution

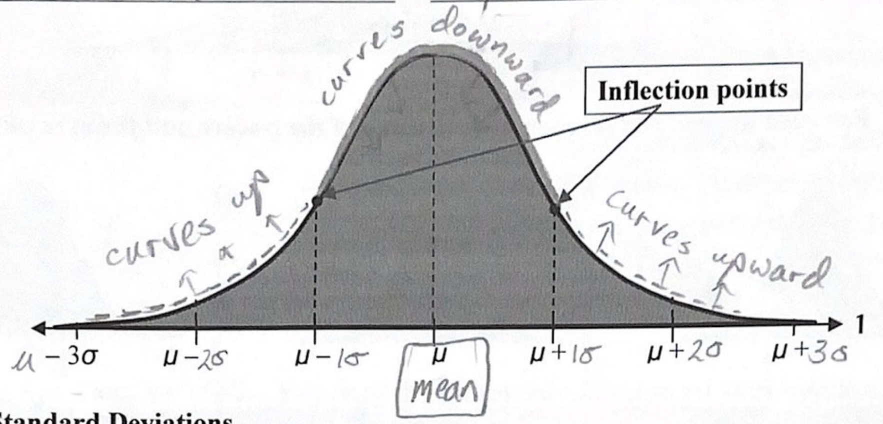
1. The mean, median, and mode are the SAME value.
2. The normal curve is bell-shaped and symmetric about the mean (μ).
3. The total area under the curve is equal to 1.
4. The normal curve approaches but never touches the *x*-axis as it extends farther and farther away from the mean.

more on next page…



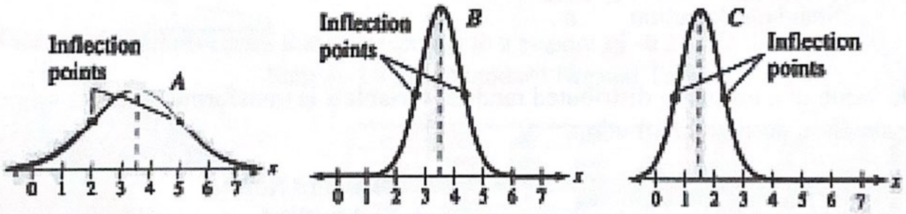
Properties of Normal Distributions, continued…

5. Between $\mu - \sigma$ and $\mu + \sigma$ (in the center of the curve), the graph curves downward. The graph then curves upward to the left of $\mu - \sigma$ and to the right of $\mu + \sigma$. The points at which the curve changes from curving upward to curving downward are called the inflection points.



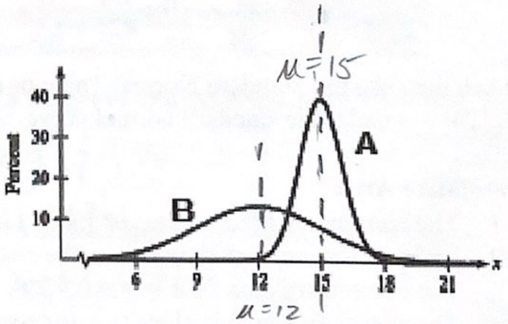
Means and Standard Deviations

- A normal distribution can have any mean and any positive standard deviation.
- The mean gives the location of the line of symmetry.
- The standard deviation describes the spread of the data.



$\mu = 3.5$	$\mu = 3.5$	$\mu = 1.5$
$\sigma = 1.5$	$\sigma = 0.7$	$\sigma = 0.7$

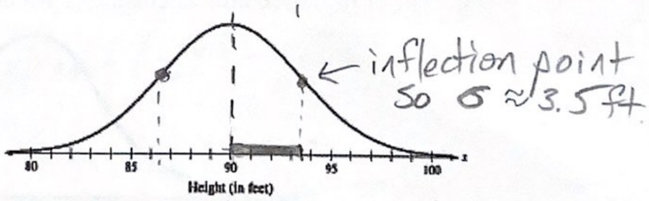
Examples:



1) Which curve has the larger mean? A ($\mu = 15$)

2) Which curve has the larger standard deviation?
B: more spread out

3) The heights of fully grown white oak trees are normally distributed. The curve represents the distribution. What is the mean height of a fully grown white oak tree? Estimate the standard deviation.

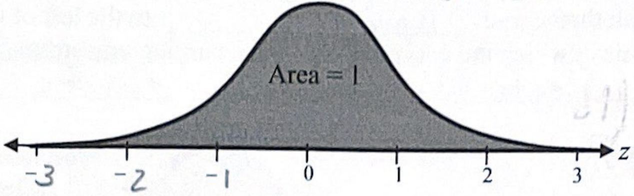


$3\sigma = 10$
So $\sigma \approx 3.3$ ft.

μ = mean height = $\boxed{90 \text{ feet}}$

σ = standard dev. ≈ $\boxed{3.3 \text{ ft}}$ or $\boxed{3.5 \text{ ft}}$

The Standard Normal Distribution: A normal distribution with a mean of 0 and a standard deviation of 1.

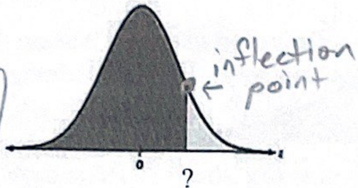


Example 4: For each normal curve, estimate the value of the *z*-score and the area of the shaded region.

a)

Recall:

$\boxed{68\text{-}95\text{-}99.7 \text{ Rule}}$



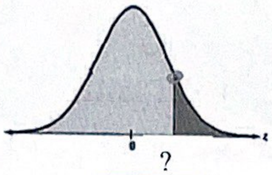
↑

$\sigma = 1$ So $\boxed{z = 1}$

Area = 50% + ½(68%)

= .50 + .34 = $\boxed{.84}$

b)

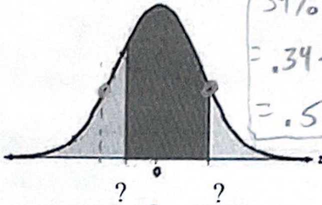


$\boxed{Z = 1}$

Area = 100% − 84%

= 1.00 − .84 = $\boxed{.16}$

c)



$\boxed{\begin{aligned}&34\% + \tfrac{1}{2}(34\%)\\&= .34 + .17\\&= .51\ \boxed{?}\end{aligned}}$

$\boxed{\text{Not really half...}}$

$z \approx -.5$ $Z = 1$

Area < 68%

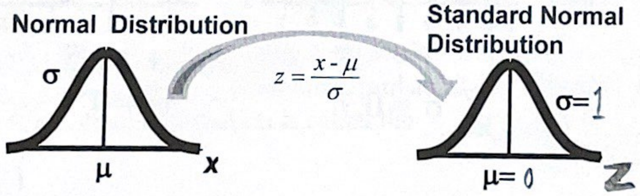
Area ≈ $\boxed{.55}$ (?)

Need more info!

Standardizing Values (*z* – scores): Any *x*-value can be transformed into a *z*-score by using the formula

$$z = \frac{\text{Value - Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

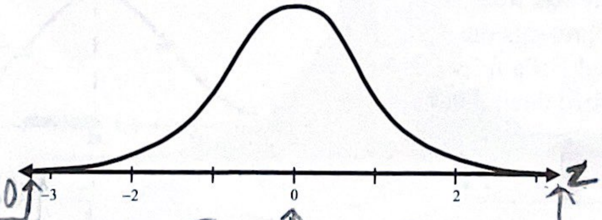
If each data value of a normally distributed random variable *x* is transformed into a *z*-score, the result will be the standard normal distribution.



We can then use the Standard Normal Table or a calculator to find the **cumulative area** (area to the left) under the standard normal curve.

Cumulative Area

1. The cumulative area (to the left) is close to 0 for *z*-scores close to $z = -3.49$.
2. The cumulative area increases as the *z*-scores increase.
3. The cumulative area for $z = 0$ is 0.5000. ⇒ 50% of the area
4. The cumulative area is close to 1 for *z*-scores close to $z = 3.49$.



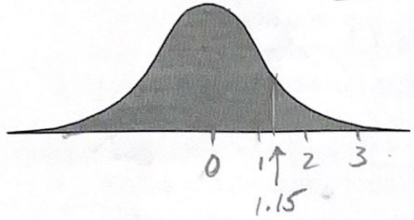
Area ≈ 0

Area ≈ .5

Area ≈ 1.0

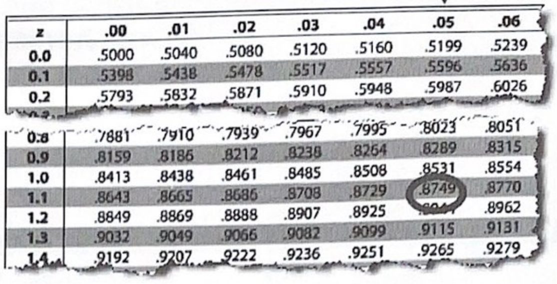
Example 5: Find the cumulative area that corresponds to a *z*-score of 1.15 for a normal curve.
($z < 1.15$)

Step 1: Draw a picture (**required!**)



Area ≈ 0.8746

Step 2: Use the **Standard Normal Table** (Table 4 in the foldout in your book.)



z	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	[illegible]	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279

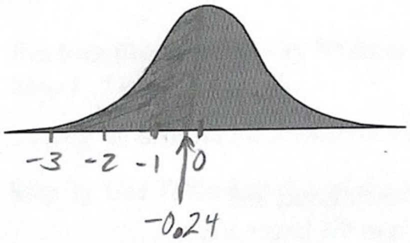
So, the cumulative area at $z = 1.15$ is 0,8749 __________.

Alternative method... on your TI-83…**You must draw a picture and record your keystrokes!**

Button	Comments
2nd, DISTR	
2: normalcdf(	*cdf* stands for Cumulative Distribution Function (gives values to the left)
-10,000,	Lower limit; use an extremely small value.
1.15)	The upper limit (the *z*-score when you are finding area to the LEFT)
ENTER	

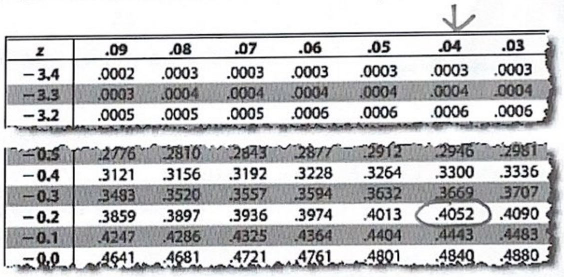
Example 6: Find the cumulative area that corresponds to a *z*-score of -0.24. ($z < -0.24$)

Step 1: Draw a picture.



Area ≈ ,4052

Step 2: Use the Standard Normal Table (Table 4.)



z	.09	.08	.07	.06	.05	.04	.03
−3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003
−3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004
−3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006
−0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981
−0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336
−0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707
−0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090
−0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483
−0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880

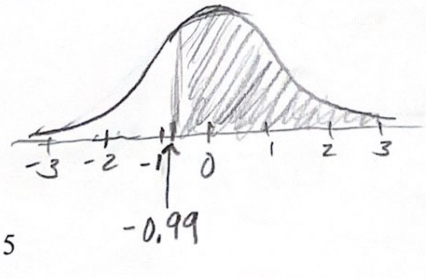
Alternative Method: Use your TI-83 **(You MUST draw a diagram and record your keystrokes for work.)**

normalcdf(-10,000, -0.24) = $\boxed{0.4052}$ ← ,405165...

Round up

Example 7: Find the area under the standard normal curve to the **right** of $z = -0.99$. ($z > -0.99$)

step 1: Picture



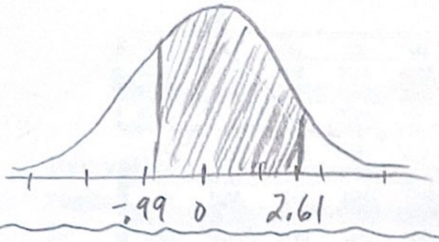
step 2: calculator

normalcdf(-.99, 10000) = $\boxed{0.8389}$

big number

Example 8: Find the area under the standard normal curve between $z = -0.99$ and $z = 2.61$.

Step 1:



Step 2:

normalcdf$(-0.99, 2.61)$

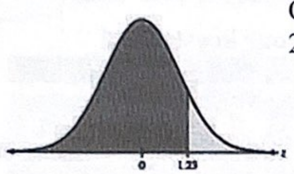
$= \boxed{0.8344}$

SUMMARY for finding area when the *z*-score is known:

→ **To find area to the LEFT of a *z*-score:**

1) Draw a picture.

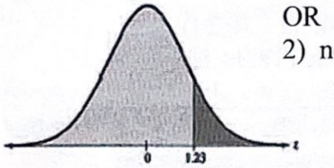
2) Use Table 4 (the value corresponds to the area to the left.)
OR
2) normalcdf (lower limit such as -1000, *z*-score)



→ **To find the area to the RIGHT of a *z*-score:**

1) Draw a picture

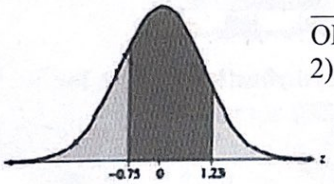
2) Use Table 4, and subtract this value from 1.
OR
2) normalcdf (*z*-score, up limit such as 1000)



→ **To find the area BETWEEN two *z*-scores:**

1) Draw a picture.

2) Use Table 4 to find the area to the left of each value, and subtract the smaller area from the larger area.
OR
2) normalcdf (smaller *z*-score, larger *z*-score)



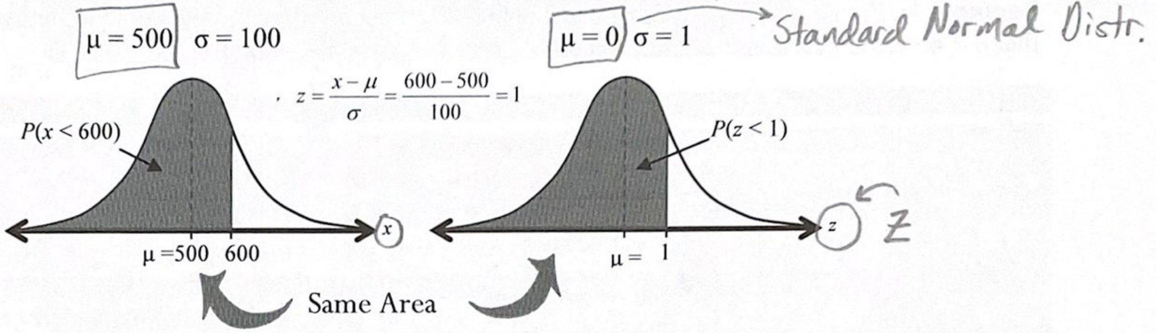
6.2 Notes: Normal Distributions: Finding Probabilities

Objectives:

5. Can you calculate z-scores?
6. Can you explain how area under a normal curve is related to probability?
7. Can you find probabilities for normally distributed variables?
8. Can you use technology to find probabilities?

Probability, Area, and Percent

- With a normal distribution (or any continuous distribution), the probability of getting a value greater or less than x is equal to the area of that shaded region.
- Area and Probability are both expressed as decimals. [Ex: 100% = 1.00]
- Move the decimal two times to the right to change this value to a percent.



$$P(x < 600) = P(z < 1)$$

Finding Probability with Normal Curves When Given an x Value:

Step 1) Draw a Diagram.

Step 2) Transform the x value to a z-score. $z = \frac{x-\mu}{\sigma}$

Step 3) Use Table 4 or the calculator to find the probability (area.)

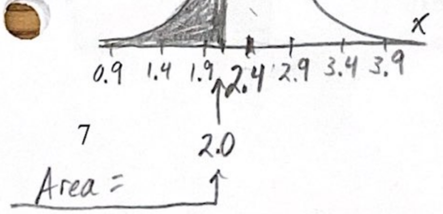
Random Variable : x = time before upgrading

Example 1: A survey indicates that people use their computers an average of 2.4 years before upgrading to a new machine. The standard deviation is 0.5 year. A computer owner is selected at random. Find the probability that he or she will use it for fewer than 2 years before upgrading. Assume that the variable x is normally distributed.

→ $\mu = 2.4$ yr
→ $\sigma = 0.5$

Find $P(x<2)$

Step 1 $P(x<2)$



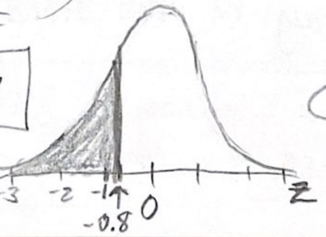
Step 2

$z = \frac{2-2.4}{0.5}$

$z = \frac{-0.4}{0.5}$

$\boxed{z = -0.8}$

$\boxed{P(z<-0.8)}$



Step 3

normalcdf (-10000, -0.8)

= 0.211855....

≈ $\boxed{0.2119}$

OR

21.19% of computer owners upgrade in less than 2 years.

Random variable: x = time in the store

$\mu = 45$

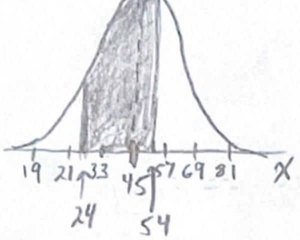
$\sigma = 12$

Example 2: A survey indicates that for each trip to the supermarket, a shopper spends an average of 45 minutes with a standard deviation of 12 minutes in the store. The length of time spent in the store is normally distributed and is represented by the variable *x*. A shopper enters the store. Find the probability that the shopper will be in the store for between 24 and 54 minutes. Find $P(24 < x < 54)$

$\Downarrow$

$P(-1.75 < z < 0.75)$

Step 1:



Step 2:

$z_1 = \frac{24-45}{12}$ $z_2 = \frac{54-45}{12}$

$z_1 = -1.75$ $z_2 = 0.75$

Step 3:

normalcdf$(-1.75, 0.75)$

$= \boxed{0.7333}$

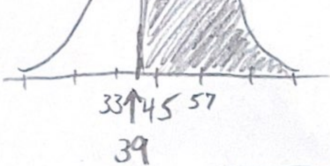
Interpret: 73.33% of shoppers will be in the store between 24 and 54 minutes

Example 3: Find the probability that the shopper will be in the store more than 39 minutes. (Recall that μ = 45 minutes and σ = 12 minutes.) Find $P(x > 39)$

$\Downarrow$

$P(z > -0.5)$

Step 1:



Step 2:

$z = \frac{39-45}{12}$

$\boxed{z = -0.5}$

Step 3:

normalcdf$(-0.5, 10000)$

$= \boxed{0.6915}$

69.15% of shoppers are expected to stay in the store over 39 min.

Example 4: If 200 shoppers enter the store, **how many** shoppers would you expect to be in the store more than 39 minutes? (*Note: this question is NOT asking for probability!*)

69.15% of shoppers

= 69.15% of 200

= .6915 × 200

percent of a number ⇒ Multiply!

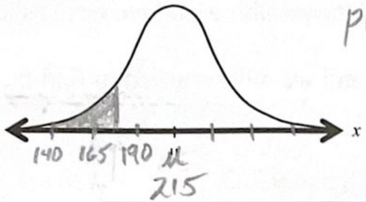
= 138.3 shoppers

You would expect about 138 shoppers (out of the 200) to stay in the store over 39 min.

Using Technology to Find Probability Without Standardizing:
Your TI-83 can automatically transform an x-value to a z-score (if you give μ & σ), which allows you to skip this step by hand.

Example 5: Assume that cholesterol levels of men in the United States are normally distributed, with a mean of 215 milligrams per deciliter and a standard deviation of 25 milligrams per deciliter. You randomly select a man from the United States. What is the probability that his cholesterol level is less than 175?

$P(x<175)$



$\text{normalcdf}(-10000, 175, 215, 25) = 0.0548$
low high μ σ
$= \boxed{5.48\%}$

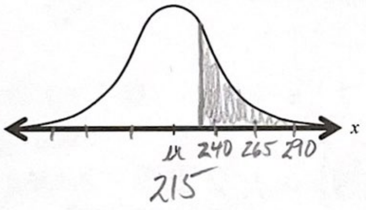
normalcdf (lower limit, upper limit, mean, standard deviation)

For this one, could we use "0" for low?

Button	Comments
2nd, DISTR	
2: normalcdf(	*cdf* stands for Cumulative Distribution Function (gives values to the left)
0,	Lower limit (the smallest possible cholesterol); use the lowest possible value when finding area to the LEFT
175,	Upper limit (the x-value when we are finding are to the LEFT)
215,	The mean
25)	The standard deviation
ENTER	

Example 6: Assume that cholesterol levels of men in the United States are normally distributed, with a mean of 215 milligrams per deciliter and a standard deviation of 25 milligrams per deciliter. You randomly select a man from the United States. What is the probability that his cholesterol level is more than 230?

$P(X>230)$



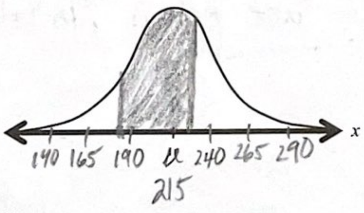
$\text{normalcdf}(230, 10000, 215, 25)$
low high μ σ

$= 0.2743$
$= \boxed{27.43\%}$

27.43% of men in the U.S. have cholesterol over 230

Example 7: Assume that cholesterol levels of men in the United States are normally distributed, with a mean of 215 milligrams per deciliter and a standard deviation of 25 milligrams per deciliter. You randomly select a man from the United States. What is the probability that his cholesterol level is between 185 and 225?

$P(185<x<225)$



$\text{normalcdf}(185, 225, 215, 25) = .5404$
$= \boxed{54.04\%}$

54.04% of men in U.S. have cholesterol between 185 and 225

Remember: Draw a diagram Write down your calculator entry

6.3 Notes: Finding *x* Values (Working BACKWARDS)

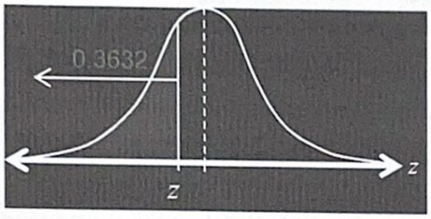
Objectives:

9. Can you find a *z*-score given the area under the normal curve?
10. Can you transform a *z*-score to an *x*-value?
11. Can you find a specific data value of a normal distribution when given the probability?

☆☆ **Working Backwards:** ☆☆ area ⇒ z-score

- ⊙ In section 6.2 we were given a normally distributed random variable *x* and we were asked to find a probability.
- ⊙ In this section, we will be given a probability or *z*-score and we will be asked to find the value of the random variable *x*.

Example 1: Find the *z*-score that corresponds to a cumulative area of 0.3632.



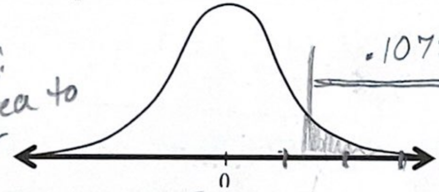
Try ↳

Button	Comments
2nd, DISTR	
3: invNorm(	This is the inverse function of finding areas, because we are working backwards.
0.3632)	The area to the left of the *z*-score you want to find.
ENTER	

$\text{invNorm}(0.3632) = -0.3499 \Longrightarrow \boxed{z \approx -0.35}$

To find a z-score when given area: invNorm(area to the LEFT of the *z*-score)
ALWAYS

Example 2: Find the *z*-score that has 10.75% of the distribution's area to its right.



.1075 to right

First: Find area to the LEFT!

$1 - 0.1075 = 0.8925$ to left

Then $\text{invNorm}(0.8925) = 1.2399$.

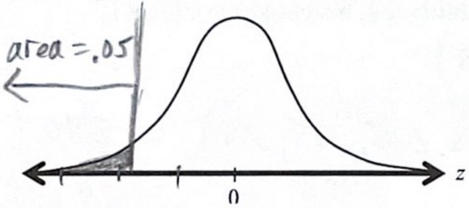
So $\boxed{z \approx 1.24}$

Example 3: Find the *z*-score that corresponds to *P*5. *What does that mean?*

↳ 5th percentile: 5% of all data values are below (to left)

⇓

area of .05 on left of what z-score?



area = .05

$\text{invNorm}(.05) = -1.6449$

$\boxed{z = -1.645}$

$\boxed{z\text{-score} \rightarrow x}$

Transforming a *z*-score to an *x* value: Working BACKWARDS! ☆ ☆ ☆

Reminder... the following formula is used to transform an *x*-value to a *z*-score:

$$\sigma \cdot z=\frac{x-\mu}{\cancel{\sigma}}\cdot\frac{\cancel{\sigma}}{1}$$

☆ Can we change this to derive a formula that isolates *x*?

Solve for x!

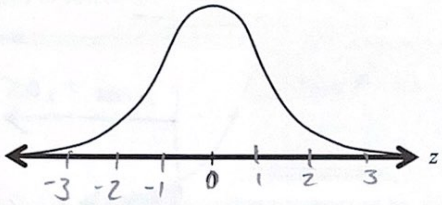
$$z\cdot\sigma = x-\mu$$

$$+\mu \qquad\qquad +\mu$$

$$\boxed{\mu + z\cdot\sigma = x}$$

$$\boxed{x=\mu+z\cdot\sigma}$$

Example 4: The speeds of vehicles along a stretch of highway are normally distributed, with a mean of 67 miles per hour and a standard deviation of 4 miles per hour. Find the speeds *x* corresponding to *z*-sores of 1.96, -2.33, and 0.



Given information:

Random Var: x = speeds of vehicles

$\mu = 67$ mph

$\sigma = 4$ mph

What are we trying to find? Speeds of cars with given z-scores

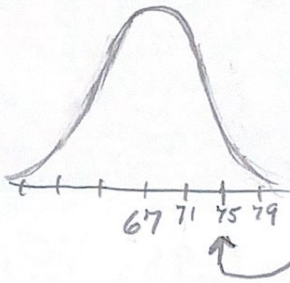
$$\boxed{x=\mu+z\cdot\sigma}$$

a) $z = 1.96$

above mean (pos.)

$x=67+1.96(4)$

$\boxed{x = 74.84 \text{ mph}}$



z = 1.96

almost 2 groups of 4 above the mean

b) $z = -2.33$

below Mean (neg.)

$x = 67+{}^{-}2.33(4)$

$\boxed{x = 57.68 \text{ mph}}$

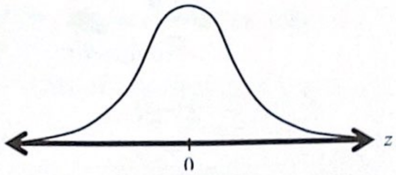
c) $z = 0$

↓ neither above or below! Right at the mean

$x=67+0(4)$

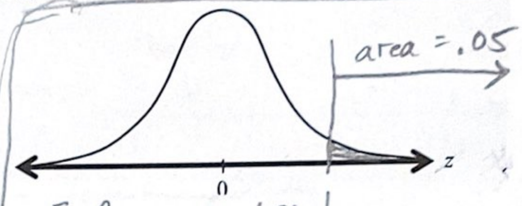
$\boxed{x = 67 \text{ mph}}$

Finding a Specific *x*-value when given Area:
We can work backwards from an area in a normal distribution to find an *x*-value (previously we found *z*-scores from areas.)



***x*-value** = invNorm (area to the LEFT, mean, standard deviation)

Example 5: Scores for a civil service exam are normally distributed, with a mean of 75 and a standard deviation of 6.5. To be eligible for civil service employment, you must score in the top 5%. What is the lowest score you can earn and still be eligible for employment?



Find area to left
$1 - .05 = \boxed{.95}$

Given information:
Random Var x = exam scores
$\mu = 75$
$\sigma = 6.5$

What are we trying to find? Score needed to be in top 5% (area = .05 to right)

Try

Button	Comments
2nd, DISTR	
3: invNorm(	This is the inverse function of finding areas, because we are working backwards.
0.95,	The area to the left of the *z*-score you want to find.
75,	The mean
6.5)	The standard deviation
ENTER	

$\text{invNorm}(\underset{\substack{\uparrow \\ \text{area to left}}}{0.95}, \underset{\substack{\uparrow \\ \mu}}{75}, \underset{\substack{\uparrow \\ \sigma}}{6.5}) = 85.69$ score at boundary

Assuming whole number scores, you would need to score an 86 (or higher)

6.4 Notes: Sampling Distributions and the Central Limit Theorem

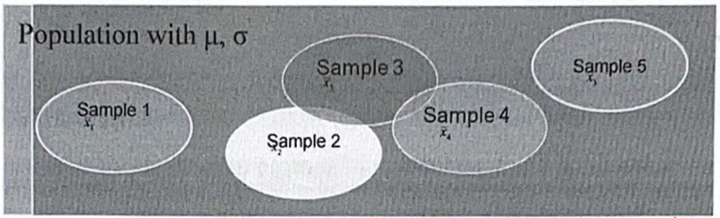
Objectives:

12. Can you find sampling distributions and verify their properties?
13. Can you find the standard error of the mean?
14. Can you interpret the Central Limit Theorem?

Sampling Distributions

If repeated samples of a population are taken, we can take the mean of EACH sample. If we use each mean as an entry, we create a **sampling distribution of the** sample **means..** $\bar{x}$

- Formed when samples of size *n* are repeatedly taken from a population.
- The mean is calculated from each sample, and then the means are collected to form a distribution. ⇒ Distribution of $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$

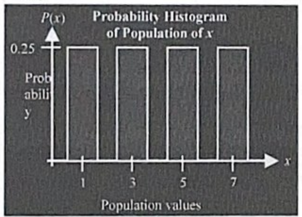


Note: Sample means can vary from each other & can vary from the population mean. This is expected and is called sampling error.

Work in groups to complete Objective #12. → p.28

Population Distribution (Probability)

$p(x)$

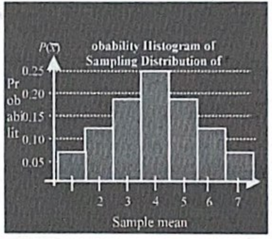


X

mean = 4

Sample Mean Distribution (Probability)

$p(\bar{x})$



$\bar{x}$

Notice that the sample mean distribution is symmetric, close to normal

Properties of Sampling Distributions of Sample Means

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

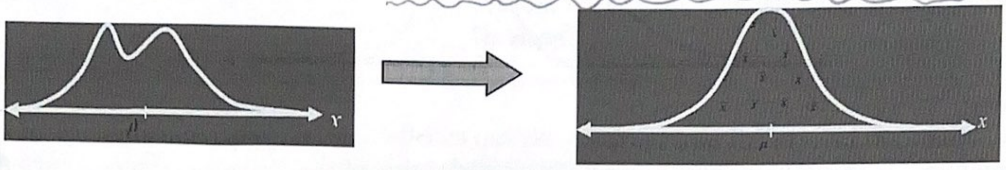
2. The standard deviation of the sample means, $\sigma_{\bar{x}}$, is equal to the population standard deviation, σ divided by the square root of the sample size, *n*.

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

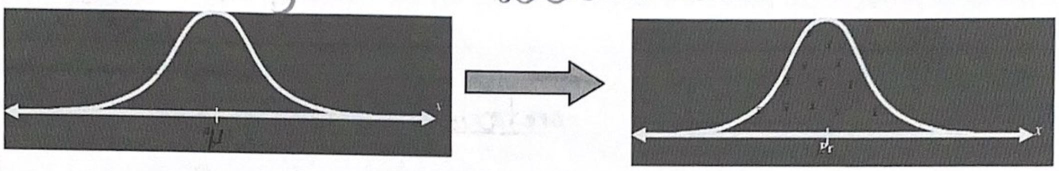
Note: Called the standard error of the mean.

The Central Limit Theorem

→ **1.** If samples of size $n \geq$ 30 are drawn from **any** population with mean = μ and standard deviation = σ, then the sampling distribution of the sample means approximates a normal distribution. The greater the sample size, the better the approximation.



→ 2. If the population itself is normally distributed, then the sampling distribution of the sample means is normally distribution for ***any*** sample size *n*.



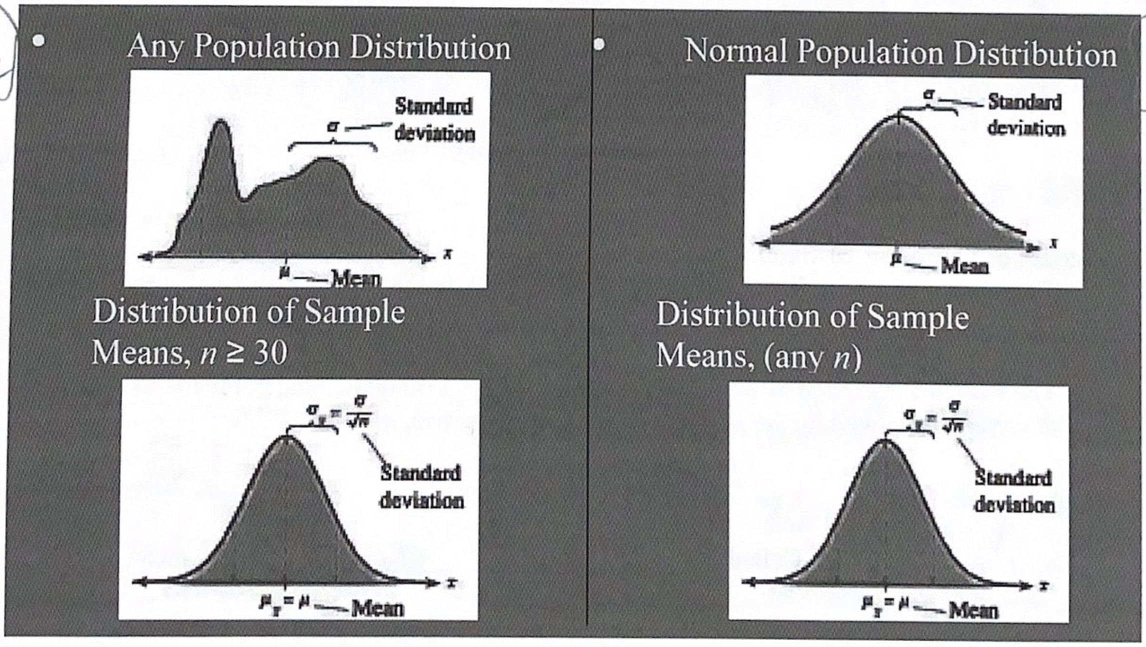
☆ 3. In either case, the sampling distribution of sample means has a mean equal to the population mean.

☆ 4. The sampling distribution of sample means has a standard deviation equal to the population standard deviation divided by the square root of *n*. This sample standard deviation is also called the standard error of the mean.

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

#1 above

#2 above



- Any Population Distribution

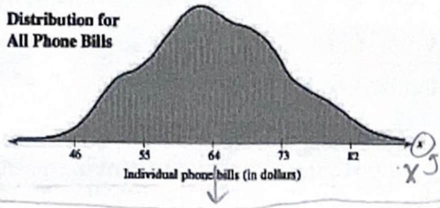
Distribution of Sample Means, $n \geq 30$

- Normal Population Distribution

Distribution of Sample Means, (any *n*)

$n = 36$ $\mu = 64$ $\sigma = 9$

Example 1: Phone bills for residents of a city have a mean of $64 and a standard deviation of $9. Random samples of 36 phone bills are drawn from this population and the mean of each sample is determined. Find the mean and standard error of the mean of the sampling distribution. Then sketch a graph of the sampling distribution of sample means.



59.50 ← 61 ← 62.50 ← 64 → 65.50 → 67 → 68.50
using $\sigma_{\bar{x}} = 1.50$

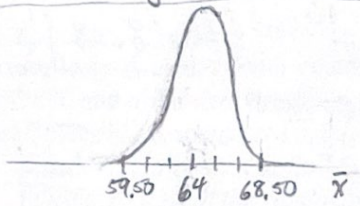
Given information:

$\mu = 64 \Rightarrow \mu_{\bar{x}} = \boxed{64}$ (SAME!)

$\sigma = 9$
$n = 36$ $\Rightarrow \boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}} = \frac{9}{\sqrt{36}}$

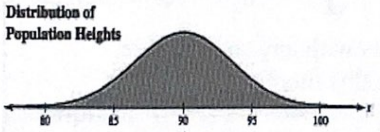
Standard error of the mean $= \frac{9}{6}$

$= \$1.50$



Interpretation: Since $\boxed{n \geq 30}$, by the Central Limit Theorem the sampling distribution can be approximated by a NORMAL distribution with $\mu_{\bar{x}} = 64$ & $\sigma_{\bar{x}} = 1.50$

Example 2: The heights of fully grown white oak trees are normally distributed, with a mean of 90 feet and standard deviation of 3.5 feet. Random samples of size 4 are drawn from this population, and the mean of each sample is determined. Find the mean and standard error of the mean of the sampling distribution. Then sketch a graph of the sampling distribution of sample means.



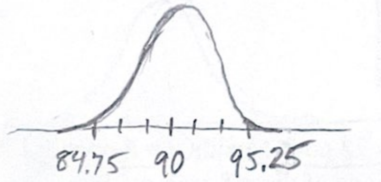
84.75 ← 86.5 ← 88.25 ← 90 → 91.75 → 93.50 → 95.25

Given information:

$\mu = 90$ ft $\Rightarrow \mu_{\bar{x}} = \boxed{90 \text{ ft.}}$

$\sigma = 3.5$ ft
$n = 4$ $\Rightarrow \sigma_{\bar{x}} = \frac{3.5}{\sqrt{4}} = \frac{3.5}{2} = \boxed{1.75 \text{ ft}}$

Standard error of the mean



Interpretation: Since the population is normally distributed, the sampling distr. of sample means is ALSO normal, with $\mu_{\bar{x}} = 90$ and $\sigma_{\bar{x}} = 1.75$

Probability and the Central Limit Theorem

Objective:

15. Can you apply the Central Limit Theorem to find the probability of a sample mean?

☆ **Finding z-scores with sampling distributions:** $z = \frac{\text{Value-Mean}}{\text{Standard Error}} = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ☆

Example 1: IQ scores have a population: normal distribution with a mean of 90 and a standard deviation of 11.. Nine students are randomly chosen from one high school.

↳ $n=9$

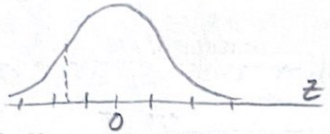
a) One student from the sample has an IQ of 120. What is the corresponding *z*-score?

$\mu=90 \Rightarrow \boxed{\mu_{\bar{x}}=90}$

$\sigma=11 \Rightarrow \boxed{\sigma_{\bar{x}} = 11/\sqrt{9} = 3.\bar{6}}$

$z = \frac{120-90}{3.\bar{6}} = \frac{30}{3.\bar{6}} = \boxed{8.18}$ Outlier!

b) Another student has an IQ of 84. What is the corresponding *z*-score?



$z = \frac{84-90}{3.\bar{6}} = \frac{-6}{3.\bar{6}} = \boxed{-1.64}$

☆☆ **Finding probabilities with sampling distributions:**

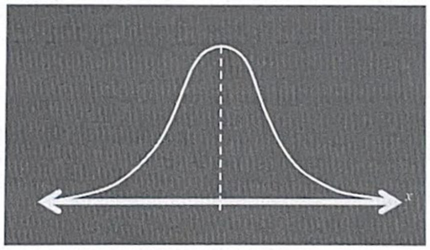
Step 1) Draw a picture.

Step 2) normalcdf (lower limit, upper limit, sample mean, standard error)

↳ gives area under that portion of the curve

↳ equals the probability of getting a value in that region

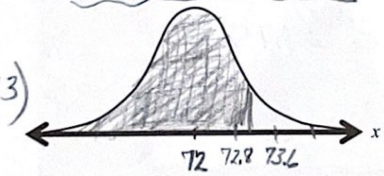
- NOTE: If the population is normal, we can do this with any sample size.
- If the population is NOT normal, we can only use this method for $n \geq 30$.



Example 2: The national mean weight for 8-year old males is 72 pounds with $\sigma = 5$ pounds. A sample is taken of 40 8-year old boys. Find the probability of one of these 8-year old boys weighing less than 73 pounds.

$n=40$

Find $P(\bar{x}<73)$



Random variable: X = weights of 8-year old boys

$\mu = 72 \Rightarrow \mu_{\bar{x}} = 72$

$\left.\begin{matrix}\sigma = 5 \\ n = 40\end{matrix}\right] \Rightarrow \sigma_{\bar{x}} = \frac{5}{\sqrt{40}} \approx 0.79$

normalcdf$(\underset{\text{low}}{-10000}, \underset{\text{high}}{73}, \underset{\mu_{\bar{x}}}{72}, \underset{\sigma_{\bar{x}}}{.79}) \approx .8972 = 89.7\%$

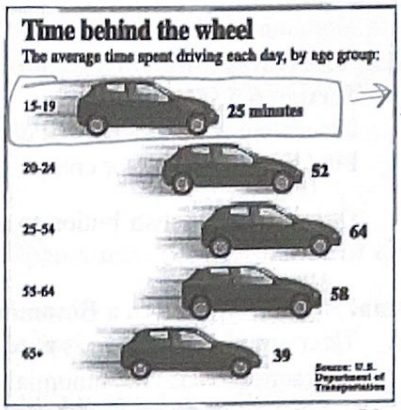
population: not necessarily normal, but $n \geq 30$, so Central Limit Thm applies

$\mu = 25$ min (from chart)

$\sigma = 1.5$

$\boxed{n=50}$

Example 3: The graph shows the length of time people spend driving each day. You randomly select 50 drivers age 15 to 19. What is the probability that the mean time they spend driving each day is between 24.7 and 25.5 minutes? Assume that σ = 1.5 minutes.

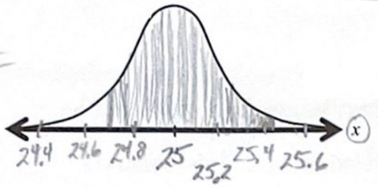


$\Rightarrow \mu = 25$ min.

$\mu_{\bar{x}} = 25$ min

$\sigma_{\bar{x}} = \frac{1.5}{\sqrt{50}} = .21$ min.

Step 1



Step 2 normalcdf(24.7, 25.5, 25, 0.21) = .9148 ≈ 91.5%

low, high, $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$

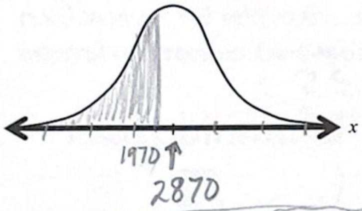
Interpret: 91.5% of samples of 50 drivers (age 15-19) will have mean daily time between 24.7 & 25.5

Example 4: A bank auditor claims that credit card balances are normally distributed, with a mean of $2870 and a standard deviation of $900. What is the probability that a randomly selected credit card holder has a credit card balance less than $2500? $P(x < 2500)$

$\mu = 2870$

$\sigma = 900$

Note: You are asked to find the probability associated with a certain value of the POPULATION of the random variable x (NOT from a sample). ⟹ old problem, don't need central limit theorem!



normalcdf(0, 2500, 2870, 900) = .3405 ≈ $\boxed{34\%}$

low, high, μ, σ

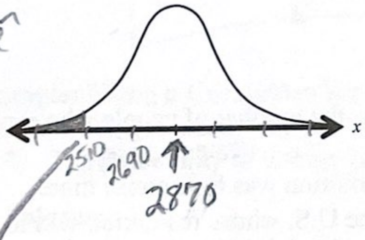
So 34% of the whole population have a balance under $2500.

Example 5: Using the same credit card balance information from Example 4, you randomly select 25 credit card holders. What is the probability that their mean credit card balance is less than $2500?

$\boxed{n = 25}$

Small sample OK since population is normal

Note: *You are asked to find the probability associated with a sample mean.* Do use central limit theorem



$\mu = 2870 \Rightarrow \mu_{\bar{x}} = 2870$

$\sigma = 900$, $n = 25$ $\rightarrow \sigma_{\bar{x}} = \frac{900}{\sqrt{25}} = 180$

normalcdf(0, 2500, 2870, 180) = 0.0199 ≈ $\boxed{2\%}$

low, high, $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$

Only 2% probability that a random sample will have mean under $2500

over 2 standard dev. from mean ⟹ unusual event ⟹ Could mean that the auditor's claim (Ex. 4) was incorrect

Section 6.5: Normal Approximations to Binomial Distributions

Recall from unit 5

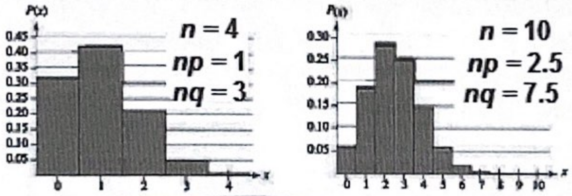
Section 6.5 Objectives:

Determine when the normal distribution can approximate the binomial distribution
- Find the correction for continuity
- Use the normal distribution to approximate binomial probabilities

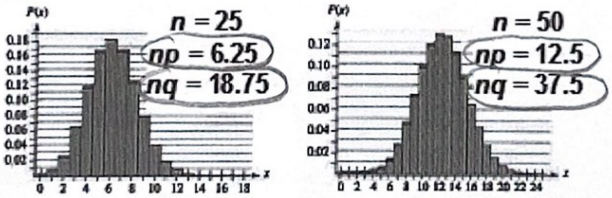
Normal Approximation to a Binomial
- The normal distribution is used to approximate the binomial distribution when it would be impractical to use the binomial distribution to find a probability.

Normal Approximation to a Binomial Distribution
- If $np \geq 5$ and $nq \geq 5$, then the binomial random variable x is approximately normally distributed with
  - mean $\mu = np$
  - standard deviation $\sigma = \sqrt{npq}$

Normal Approximation to a Binomial:
- Binomial distribution: $p = 0.25$



$np \not\geq 5$



Both $np \geq 5$ AND $nq \geq 5$

- As n increases the histogram approaches a normal curve.

Example: Approximating the Binomial:

Decide whether you can use the normal distribution to approximate x, the number of people who reply yes. If you can, find the mean and standard deviation.

1. Fifty-one percent of adults in the U.S. whose New Year's resolution was to exercise more achieved their resolution. You randomly select 65 adults in the U.S. whose resolution was to exercise more and ask each if he or she achieved that resolution.

Solution: You can use the normal approximation

$n = 65$, $p = 0.51$, $q = 0.49$

$np = 33.15 \geq 5$ ✓

$nq = 31.85 \geq 5$ ✓

- Mean: $\mu = \text{np} = 33.15$
- Standard Deviation:

$\sigma = \sqrt{npq} = \sqrt{65(.51)(.49)}$

≈ 4.03

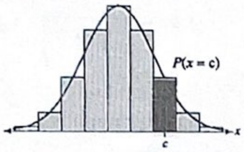
2. Fifteen percent of adults in the U.S. do not make New Year's resolutions. You randomly select 15 adults in the U.S. and ask each if he or she made a New Year's resolution.

Solution: You cannot use the normal approximation
$n = 15$, $p = 0.15$, $q = 0.85$
$np = 15(0.15) = 2.25 \leftarrow$ Fails here $2.25 \ngeq 5$
$nq = 15(0.85) = 12.75$

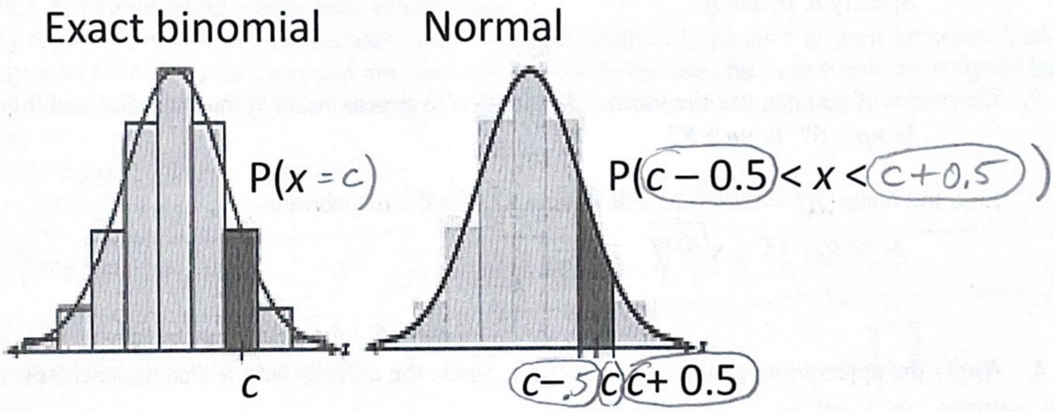
- Because $np < 5$, you cannot use the normal distribution to approximate the distribution of x.

Correction for Continuity

- The binomial distribution is discrete and can be represented by a probability histogram.
- To calculate *exact* binomial probabilities, the binomial formula is used for each value of x and the results are added.
- Geometrically this corresponds to adding the areas of bars in the probability histogram.



- When you use a *continuous* normal distribution to approximate a binomial probability, you need to move 0.5 unit to the left and right of the midpoint to include all possible x-values in the interval (**correction for continuity**).



Example: Using a Correction for Continuity:
Use a correction for continuity to convert the binomial intervals to a normal distribution interval.

1. The probability of getting between 270 and 310 successes, inclusive.

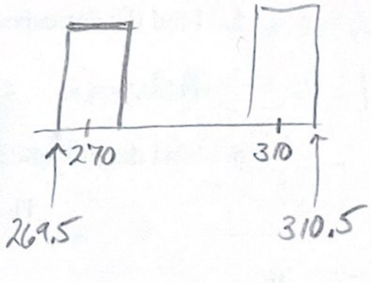
Solution:

- The discrete midpoint values are 270, 271, 272, ... 310

$270 \leq x \leq 310$

- The corresponding interval for the continuous normal distribution is

$269.5 < x < 310.5$



smallest

2. The probability of getting at least 158 successes.

Solution:

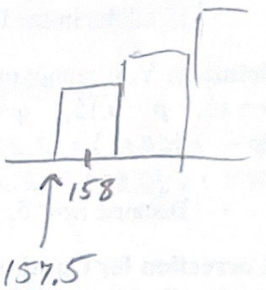
- The discrete midpoint values are

$158, 159, 160, \ldots$

$x \geq 158$

- The corresponding interval for the continuous normal distribution is

$x > 157.5$



3. The probability of getting less than 63 successes.

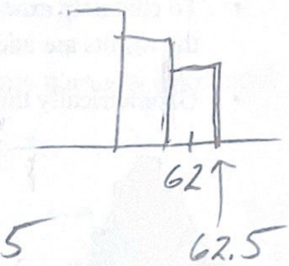
Solution:

- The discrete midpoint values are

$\ldots 60, 61, 62$ $x \leq 62$

- The corresponding interval for the continuous normal distribution is

$x < 62.5$



STEPS

Using the Normal Distribution to Approximate Binomial Probabilities

1. Verify that the binomial distribution applies.
   Specify *n, p*, and *q*.

2. Determine if you can use the normal distribution to approximate x, the binomial variable.
   Is $np \geq 5$? Is $nq \geq 5$?

3. Find the mean μ and standard deviation σ for the distribution.

$$\mu = np \quad \sigma = \sqrt{npq}$$

4. Apply the appropriate continuity correction. Shade the corresponding area under the normal curve.

   Add or subtract 0.5 from endpoints.

5. Find the corresponding *z*-score(s).

$$z = \frac{x - \mu}{\sigma}$$

6. Find the probability.

   Use the Standard Normal Table. OR normalcdf

Random Var: x counts people who say Yes, they do now exercise more, $X = 1, 2, 3, \dots, 65$

Binomial → achieved resolution → $p = .51$
→ didn't → $q = .49$
(2 possible results) with consistent probability

Example: Approximating a Binomial Probability

Fifty-one percent of adults in the U. S. whose New Year's resolution was to exercise more achieved their resolution. You randomly select 65 adults in the U. S. whose resolution was to exercise more and ask each if he or she achieved that resolution. What is the probability that fewer than forty of them respond yes? *(Source: Opinion Research Corporation)*

$n = 65$

Solution:

- Can use the normal approximation? (see the bottom of page 18) → Yes

$np = 65(.51) \geq 5$
$= 33.15 \geq 5$ ✓

$nq = 65(.49) \geq 5$
$= 31.85 \geq 5$ ✓

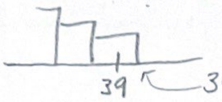
$\mu = n \cdot p$ $\sigma = \sqrt{npq}$

$\mu = 65 \cdot 0.51 = \boxed{33.15}$ $\sigma = \sqrt{65 \cdot 0.51 \cdot 0.49} \approx \boxed{4.03}$

Solution: Approximating a Binomial Probability

- Apply the continuity correction: Find $P(X < 40)$ or $P(X \leq 39)$

Fewer than 40 (…37, 38, 39) corresponds to the continuous normal distribution interval $X < 39.5$

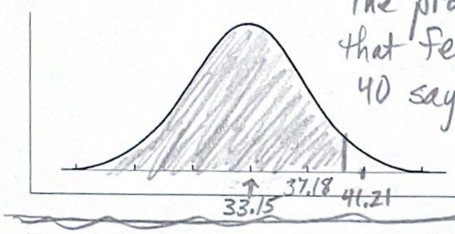


$\boxed{\text{normalcdf }(0, 39.5, 33.15, 4.03)}$

$= .9425$

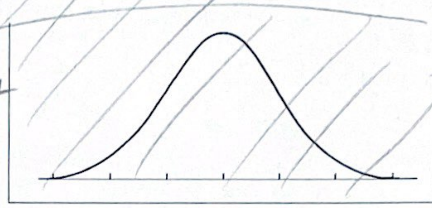
$\approx \boxed{94\%}$

Normal Distribution
$\mu = 33.15$ $\sigma = 4.03$



The probability that fewer than 40 say yes is about 94%

Standard Normal
$\mu = 0$ $\sigma = 1$



Example: Approximating a Binomial Probability

A survey reports that 86% of Internet users use Windows® Internet Explorer® as their browser. You randomly select 200 Internet users and ask each whether he or she uses Internet Explorer as his or her browser. What is the probability that exactly 176 will say yes? *(Source: OneStat.com)*

$p = .86$
$q = .14$
$n = 200$

Solution:

Can we use the normal approximation?

$np = 200(.86) = 172$ $nq = 200(.14) = 28$ ⇒ Yes, both are ≥ 5

$\mu = 200(.86) = \boxed{172}$ $\sigma = \sqrt{200(.86)(.14)} = 4.907 \approx \boxed{4.91}$

Solution: Approximating a Binomial Probability

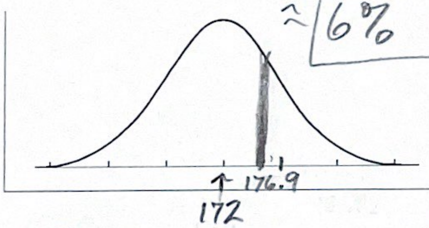
$P(X = 176)$

- Apply the continuity correction:

Exactly 176 corresponds to the continuous normal distribution interval $175.5 < x < 176.5$

normalcdf(175.5, 176.5, 172, 4.91)
$= 0.060065$
≈ 0.0601
$\approx \boxed{6\%}$

Normal Distribution
$\mu = 172$ $\sigma = 4.91$

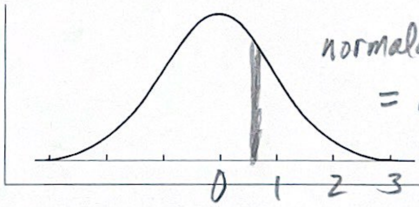


Standard Normal
$\mu = 0$ $\sigma = 1$

$z_1 = \frac{175.5 - 172}{4.91}$
$z_1 = 0.71$

$z_2 = \frac{176.5 - 172}{4.91}$
$= 0.92$

normalcdf(.71, .92)
$= 0.060065$
$\approx \boxed{6\%}$



Contrast with old method: binompdf(200, .86, 176) = 0.061169 ≈ $\boxed{6\%}$

n p x