

## Chapter 3: Testing a Mean

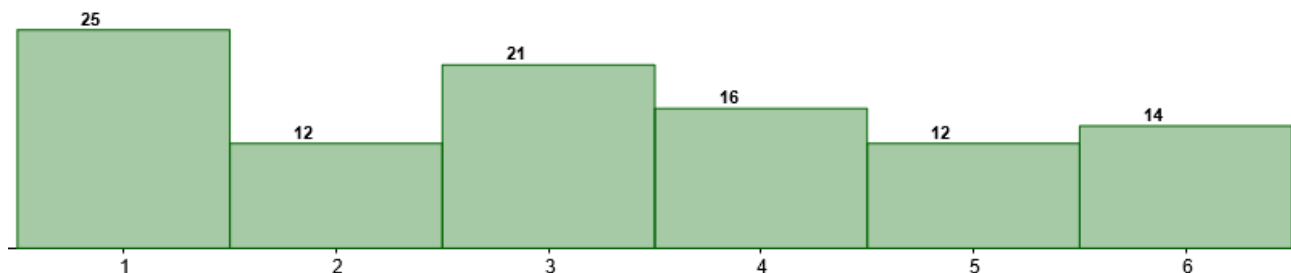


**Investigation:** Mack runs a casino game where players roll a single, 6-sided die as part of their play. In general, **players win more money** when they get **higher** dice rolls and they **lose more money** with **lower** dice rolls.

You recently found out that Mack was fired from a previous casino for **loading a die** to make players **lose** more money, and then pocketing the extra money—you wonder if he might try the same deception here! When Mack goes on break, you decide to test his die by rolling it 100 times and recording your results. You record the following:

1	2	3	4	5	6
25	12	21	16	12	14

Table 1. Mack's die frequencies



Based on a visual inspection of this data, do you think we have evidence that Mack has loaded the die to make players lose more money? What is your thought process to making that determination?

GeoGebra [Dice Roll Simulation](#) (or google search “Geogebra dice roll simulation”)

- Set dice to 1.
- Try rolling 100 dice and observing the results.

What could we measure or record with each simulation of 100 dice rolls that would help us determine if our own sample result could reasonably be generated from a fair die?

### Identifying an Estimator

- An “Estimator” would be a \_\_\_\_\_ from our data that estimates a \_\_\_\_\_ of interest!
  - One choice for our estimator is the \_\_\_\_\_ ( $\hat{p}$ ) of 1’s we might see from 100 rolls.
  - This is estimating the true proportion ( $\pi$ ) of 1’s we would see if rolling this die continuously.
- But maybe the die isn’t specifically rigged to produce 1’s, but rigged to simply favor lower values or avoid higher values. Maybe we want to consider the \_\_\_\_\_ nature of this variable!
  - Under that condition, we might choose a \_\_\_\_\_ ( $\bar{x}$ ) from our 100 dice rolls would serve as our estimator for  $\mu$ .
  - $\mu$  would represent the mean across \_\_\_\_\_ rolls we could make with this die.
- To determine if our results are unusual, we might start with the **Absolute Error** of our estimate.
  - Absolute Error is the distance between our specific estimate and the parameter.
  - If using  $\hat{p}$  as our estimator, our absolute error (assuming a fair die) is...
  - If using  $\bar{x}$  as our estimator, our absolute error is...

But we can’t use absolute error alone to answer our question—we need to determine how often we would observe an absolute error **at least this large** by using an appropriate \_\_\_\_\_.

Let’s use the [sampling distribution discrete population](#) from the Art of Stat web apps page (choose the sampling distribution for discrete population).

- Select “Fair Die” as your Population Distribution
- Adjust the sample size to 100

### Let’s complete our investigation using $\bar{x}$ as our estimator!

Null Hypothesis: Mack’s die is \_\_\_\_\_. should average to \_\_\_\_\_.  
Symbolically, we’d say...



- By running several thousand simulations, you can create a “Null Model” under this scenario.
- Let’s use the “Find probability” tool (bottom left) to estimate how often we would see a sample mean as low or lower than ours according to this Null Model.

How would we interpret this probability (p-value) in context?

If you were trying to decide whether to fire Mack, do you think this is convincing enough?

### Exploring the Null Model

Use the “Summary Statistics” checkbox to explore some characteristics about our Null Model.

What was the lowest sample mean observed in your simulation? \_\_\_\_\_

What was the highest sample mean observed in your simulation? \_\_\_\_\_

What was the standard deviation of the sample means in this simulation? \_\_\_\_\_

The Standard Error of an estimator

- Since we don’t know the parameter when we do inference, we can’t just find the absolute error of our estimate. Instead, we must *estimate* that error.
- The **Standard Error** of a statistic represents the \_\_\_\_\_ when using it to estimate it’s parameter. It reflects our uncertainty in \_\_\_\_\_.
  - The “Standard Error of a Sample Mean” is abbreviated as...
  - The “Standard Error of a Sample Proportion” is abbreviated as...



### Standard Deviation vs. Standard Error

- A standard error is kind of like a special kind of standard deviation. It’s the standard deviation of a set of values when those values happen to be possible \_\_\_\_\_!
- So the standard deviation being reported from our simulation will be the SE for our sample mean!

When taking a sample of 100 dice rolls, we should expect to see an error of \_\_\_\_\_ on average when using our sample mean as an estimator for the true mean.

**Challenge Question:** We take a sample of 20 dice rolls. We *expect* that sample mean to be \_\_\_\_\_ accurate compared to a sample mean generated from 100 rolls. Therefore, the standard error of our sample mean for the n = 20 sample will have a(n) \_\_\_\_\_ standard error than the sample mean from n = 100.

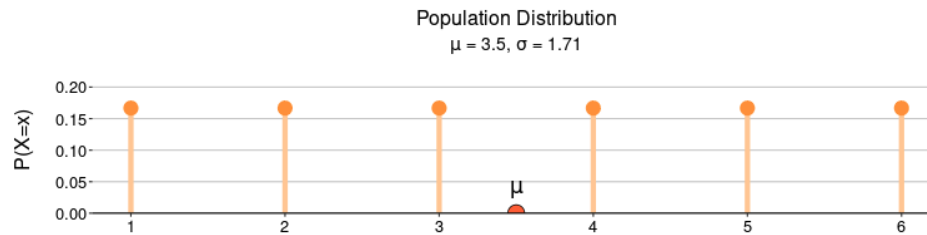
- |                  |            |
|------------------|------------|
| 1. more          | 1. bigger  |
| 2. less          | 2. smaller |
| 3. about equally | 3. equal   |

Let’s explore the standard error of our sample mean when we change the sample size. Re-generate a null model for the average dice roll for each sample size (10,000 simulations is fine). For each one, record the minimum sample mean, maximum sample mean, and standard error for the sample mean.

Size 20	Size 500
Minimum:	Minimum:
Maximum:	Maximum:
Standard Error:	Standard Error:

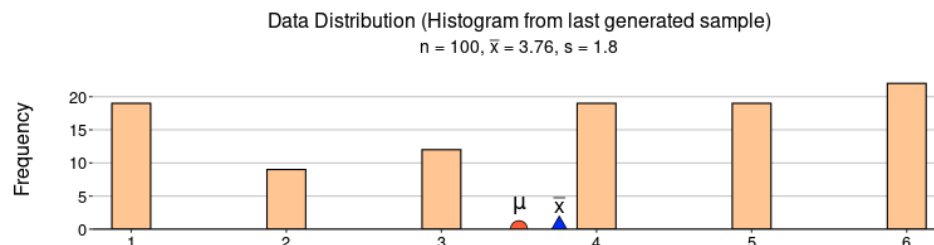
## Distributions and Convergence Properties

- **A Population Distribution** represents the true underlying distribution of a particular variable.
  - The population distribution for rolling a single die is *uniformly* distributed across 1, 2, 3, 4, 5, 6.



- *But variable distributions come in all shapes—so be aware of that!*

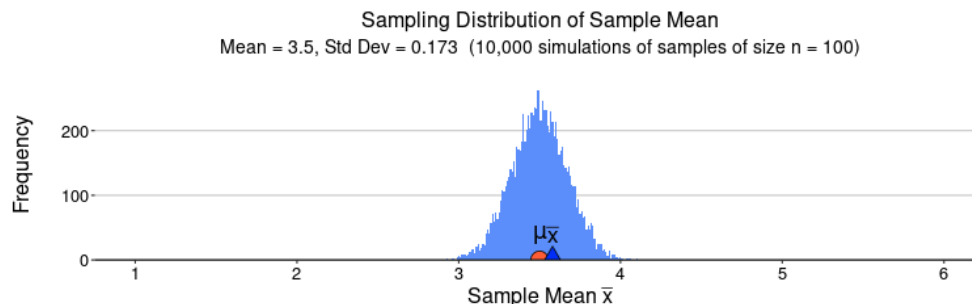
- **A Sample Data Distribution** is the distribution of measurements collected from our *sample*.
  - A sample is an incomplete picture of the population. It represents the shape of the limited data we have collected so far.
  - A sample distribution will look like a \_\_\_\_\_ version of the population!



- As sample size increases, your **sample** distribution will **converge in** \_\_\_\_\_ to the \_\_\_\_\_.

- Now imagine taking many samples of size  $n$ . Just as each sample will vary a bit, we can imagine there being a distribution of possible statistics we could get! We call this “theoretical” distributional idea a **SAMPLING Distribution**.

- In the dice example, we created a sampling distribution for  $\bar{x}$  when assuming a fair die to see how  $\bar{x}$  might vary for a particular sample size.



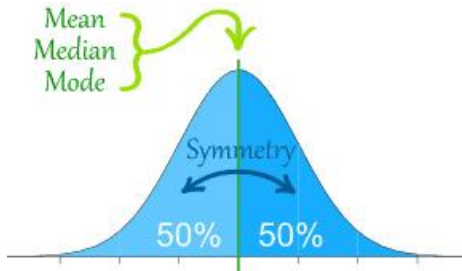
- As sample size increases, the **sampling** distribution of the sample mean will **converge in value** to the \_\_\_\_\_.
- This is known as the Law of Large Numbers. A corollary to this law is that the variability in our estimator will \_\_\_\_\_ when using a larger and larger sample.



## Read on your own

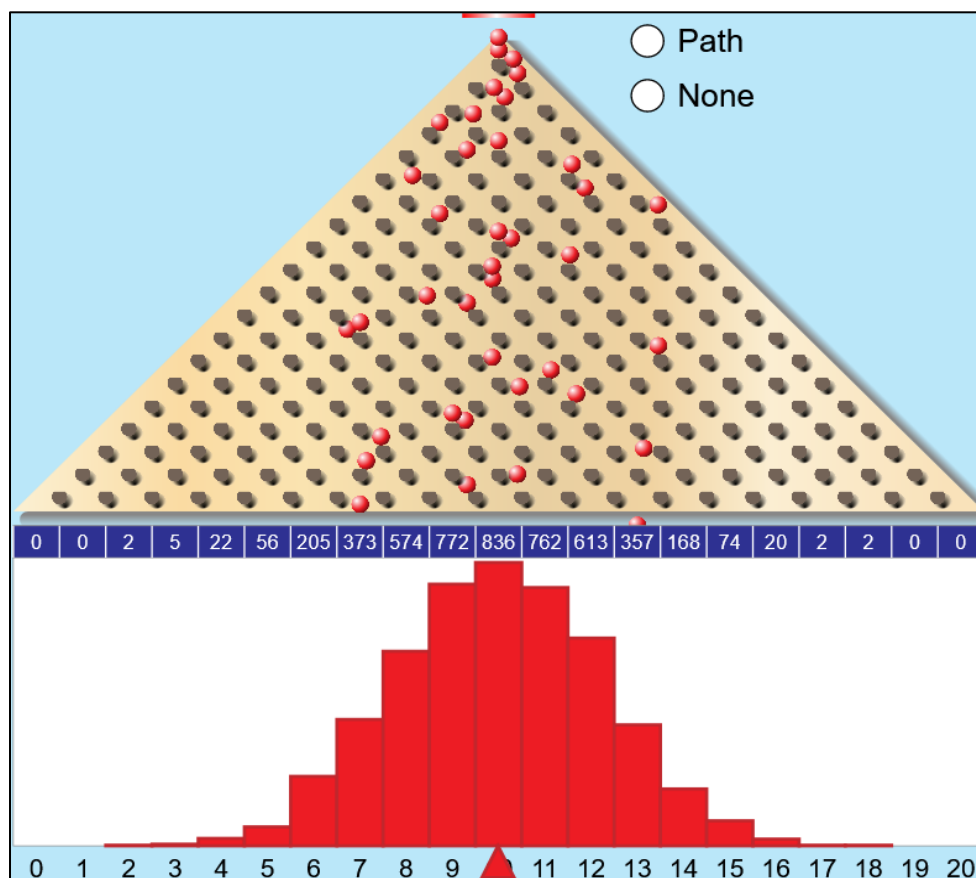
### What does it mean to be “Normally distributed”?

- A normally distributed variable has an identifiable bell-curve shape and shows up in a lot of places!



It is symmetric about the center, meaning that 50% of data will be below the center and 50% is above.

- The **normal distribution** may be thought of as the **distribution of random errors**. The idea is that some variables will cluster around a center point, with random chance effects making discrepancies farther from the center less and less likely.
- In what situations do we see normal distributions? *Think about the landing places of plinko balls in plinko probability.*
  - There is a most likely value that observations will gravitate to (a clear center).
  - Random variation is equally likely to increase or decrease the value (symmetric).
  - Discrepancies farther from the center are less likely (non-uniform).



Many biological measurements largely determined by genetics (heights, lengths), quality control measurements (the weight of a mass-distributed item, the time it takes to complete a process), and other common variables are normally distributed. But many are not—**don't assume everything is “normal.”**

### Reflection Questions

---

**3.1.** When trying to estimate a population proportion ( $\pi$ ) with a sample of data, what might we commonly use as an estimator? What might we commonly use as an estimator for a population mean ( $\mu$ )?

**3.2.** How did we statistically test whether Mack's die might reasonably be generating fair results? If we had set  $\alpha = 0.01$  as our threshold for firing, what would we have decided?

**3.3.** What is the difference between the **absolute error** of an *estimate* and the **standard error** of an *estimator*?

**3.4.** What is the difference between a *sample data* distribution and a *sampling* distribution? Which of these would we use to determine if our sample statistic is unusually far from expectation?

---

**Investigation:** Research has noted that taller people tend to have more success—especially among males. Might this be true at the high school level as well? One way to examine this is to test whether male valedictorians and salutatorians are taller than average.

To investigate this, I asked my students last year whether they were the valedictorian of their class. If restricting to those who identified their gender as “man,” there are **15** students with an average height of **71.47** inches.

As a benchmark for comparison, let's look at the heights of all men who took my class last year. Their heights were approximately “normally distributed” with a **mean of 70.00 inches** and a **standard deviation of 2.87 inches**.

In your own words, what is the research question we are answering here?

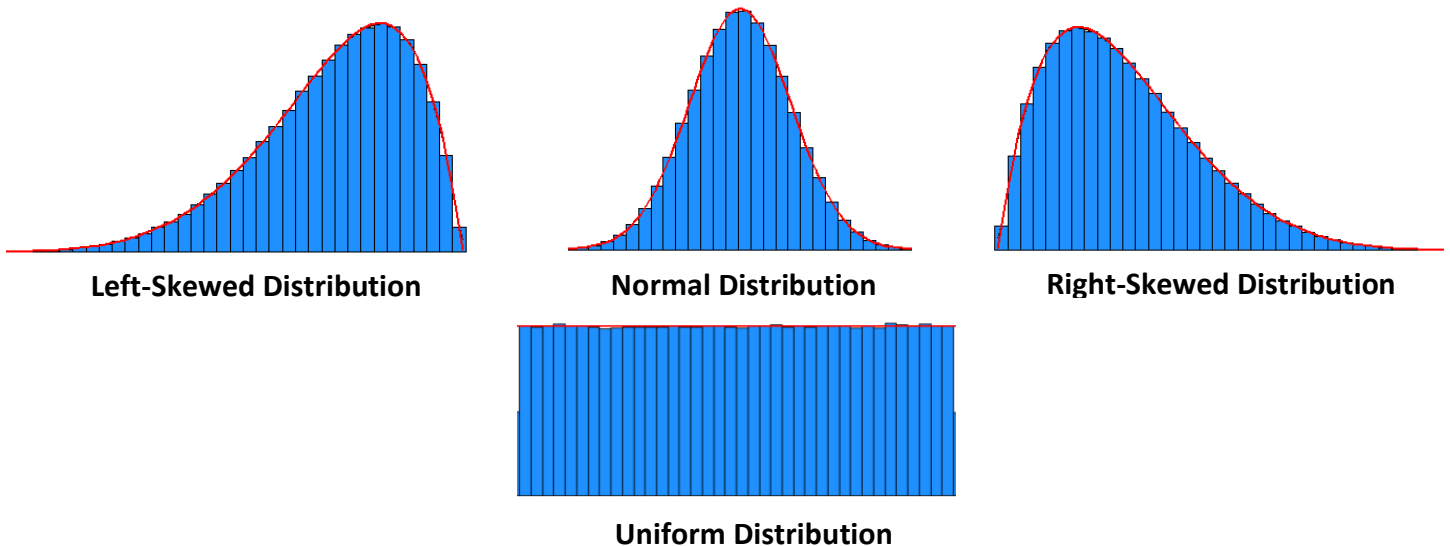


**How is this different from the loaded die investigation?**

- When investigating Mack’s die, the population distribution was a \_\_\_\_\_ variable, where we could model and sample from **each** possible value exactly.
- In this investigation, height is a \_\_\_\_\_ variable, where we can’t model the probability of being every possible height value. *What’s the probability of being 66.5786... inches?*

**Modeling a continuous population**

- Now that our variable (male valedictorian heights) is continuous, we must identify the distributional shape we are sampling from.
- If we can identify a variable as approximating a known distributional function (like a “normal distribution”), then we can sample from this function and create a Null model!



The unknown parameter in our investigation (\_\_\_) is the average height of all \_\_\_\_\_ who might ever take this class.

What is our sample estimate for that unknown parameter?

What is our Null hypothesized value for that parameter?

If the null were true...what would be the absolute error of our sample mean?

Is this a Directional or Non-directional investigation? *Would our results only matter (or support our theory) if they are specifically lower or specifically higher than Expectation? Or would a discrepancy in either direction be equally noteworthy?*

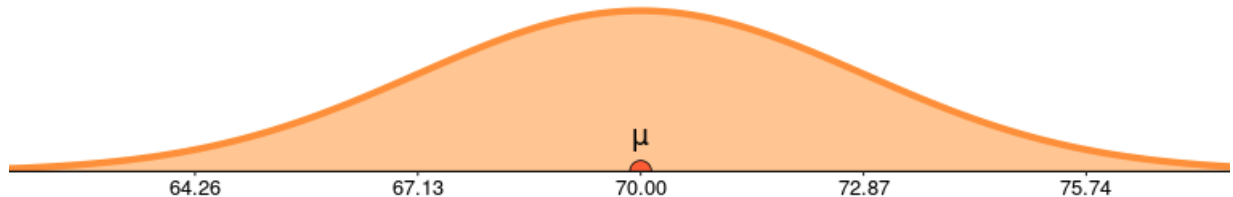
How would we symbolically write our Null and Alternative Hypotheses (*the two possible answers to our research question*) for this investigation?



## Chapter 3: Testing a Mean

Let's open the [Sampling Distribution Continuous Population](#) applet from the Art of Stat Web apps page to model the distribution of male heights in our class generally.

- Choose "bell-shaped."
- Enter a custom mean and standard deviation to match our null hypothesized parameters
- Set sample size to \_\_\_\_.
- 



Simulate a sampling distribution for  $\bar{x}$  (null model) and draw it over top of the population distribution above. Let's shade the p-value region.

Using the “Find Probability” feature, what is your simulated p-value (right tail region)?

**Example:** The Art of Stat applet has some pre-loaded data involving the Yellow Taxi Ride distances in New York City on Dec. 31 2016.

***Before we open the data...***...think about what shape you would expect this distribution to take. Would you expect it to be normally distributed? Or some other shape? Draw your prediction below!



Then, we'll check the applet to see what it actually looks like!

Population Distribution (your guess)	Population Distribution (actual)



## Chapter 3: Testing a Mean

Let's take a random sample of 30 ride distances from this population and take note of our sample mean. As expected, the sample distribution resembles a bumpy version of the \_\_\_\_\_.

If we continued taking samples of size 30 and plotting the average distance of each sample, what shape would the **sampling** distribution take?

Sampling Distribution for n = 30 (your guess)	Sampling Distribution for n = 30 (actual)

Try adjusting n to a larger value like 100. Does the sampling distribution shape change?

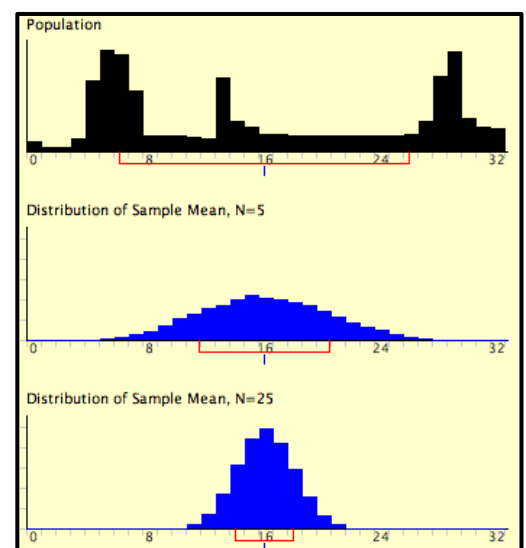
### Properties of the Sampling Distribution for *Sample Means*

- 1) The distribution of  $\bar{x}$  will center around the true population mean,  $\mu$ . In other words, of  $\bar{x}$  is an \_\_\_\_\_ estimator of  $\mu$ .
- 2) The standard deviation of the distribution of  $\bar{x}$  (the "Standard Error of  $\bar{x}$ ") can also be derived by calculating the **population** standard deviation of your variable divided by the square root of the sample size used to generate that sample mean.

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- 3) **Central Limit Theorem:** Even if a variable is **not** normally distributed, the distribution of  $\bar{x}$  will **become normally distributed** if the sample size generating those sample means is \_\_\_\_\_.

The distribution of sample means will begin to look more like a **distribution of random errors!**



**When will the sampling distribution for  $\bar{x}$  be normally distributed?**

- If the population you are sampling from is *already* **approximately normally distributed**, then the distribution of  $\bar{x}$  will be normally distributed for **any** sample size.
- If the population is **not normally distributed**, but is **approximately symmetric** (e.g., uniformly distributed), then the distribution of  $\bar{x}$  should be normally distributed, even with small sample sizes.  **$n \geq 10$**  is likely a good benchmark!
- If the population has *some noticeable* **asymmetry**, but no significant skewness, then  **$n \geq 30$**  is a relatively safe benchmark (sometimes less is ok too!)
- If the population has a **significant skewness**, then you may need a much larger sample size. In our class, we'll use  **$n \geq 100$**  as a generic benchmark, though sometimes that isn't enough!

---

**Reflection Questions**

**3.5.** Exam scores is a moderately left skewed variable. If we were to take a sample of size 100 exam scores and plot those *individual* observations, what general shape would we expect that distribution to be?

**3.6.** Let's again take this same left skewed variable. If I were to *repeatedly* take samples of 100 exam scores and plot the *sample means* I get each time, what general shape would we expect that distribution to take?  
*Hint: this is related to the Central Limit Theorem.*

**3.7.** Let's say this exam score variable has a mean of 85 and a standard deviation of 11. If I take a random sample of 100 exam scores, what is the "expected" error in my sample mean here?

**3.8.** As our sample size approaches the population size (or infinity), the standard error of our estimator is converging to what value? And likewise, the sampling distribution for that estimator will do what?

**Chapter 3 Additional Practice (Videos available in the Ch 3 module on Canvas!)**

---

**Practice:** Among students attending the University of Illinois, the mean ACT score was 30.1 with a standard deviation of 2.8. If we took a random sample of 50 students and calculated the average ACT score for these 50 students, how much error would we *expect* in this sample mean?

At “Purdont” University, the mean ACT score is 29.5 with a standard deviation of 2.8. If we took a random sample of 50 students from Purdont and calculated *their* average ACT score, how would that compare to the expected accuracy of our sample mean from the University of Illinois?

- A. The expected amount of error in the Purdont sample mean is **smaller** than the Illinois mean.
- B. The expected amount of error in the Purdont sample mean is **equal to** that of the Illinois mean.
- C. The expected amount of error in the Purdont sample mean is **bigger** than the Illinois mean.

Now, imagine that we took a sample of **100** students from the University of Illinois instead of 50. Which statement best describes how this new sample mean should compare to the sample mean generated from 50 students?

- A. The new sample mean would definitely be closer to the true mean of 30.1.
- B. We would expect the new sample mean to be closer to 30.1, but it’s possible it may not be.
- C. The new sample mean would definitely be farther from the true mean of 30.1.
- D. We would expect the new sample mean to be farther from 30.1, but it’s possible it may not be.

**Practice:** Let’s continue with [the applet](#) we have been using and look at the **age of FIFA19 soccer players** example under the “real population data” drop-down list.

According to the distribution, FIFA19 has a mean age of 25.10 and a standard deviation of 4.67.

While FIFA19 is comprised of players from countries all over the world, someone wants to know whether players from a smaller, professional league tend to be younger on average (perhaps because they have more up-and-coming stars trying to break through).

Of the 172 players in the small league, the average age is 24.72

Let’s consider these 172 current players as a representative sample of all players that might ever play for this smaller league.



### Chapter 3: Testing a Mean

What is the null hypothesized mean for the average age of players in the smaller league?

Symbolically write the null and alternative hypotheses for this investigation (*think first—is this a directional or non directional investigation?*)

Now, let's use the applet to create a null model and test out our hypothesis!

- Choose a sample size of \_\_\_\_\_
- Simulate a Null Model on the sim and roughly copy it below.

Notice the standard deviation of your sampling distribution. Compare it to the result you get when calculating the standard error of  $\bar{x}$ .  $SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Using the “Find Probability” feature, what is your simulated p-value? *Note that the sim always finds the area to the left, so subtract this value from 1 to get the right tail area!*

If you're curious to explore the functional representation of the normal distribution and the existence of the Central Limit Theorem further, check out this video from [3Blue1Brown](#)

