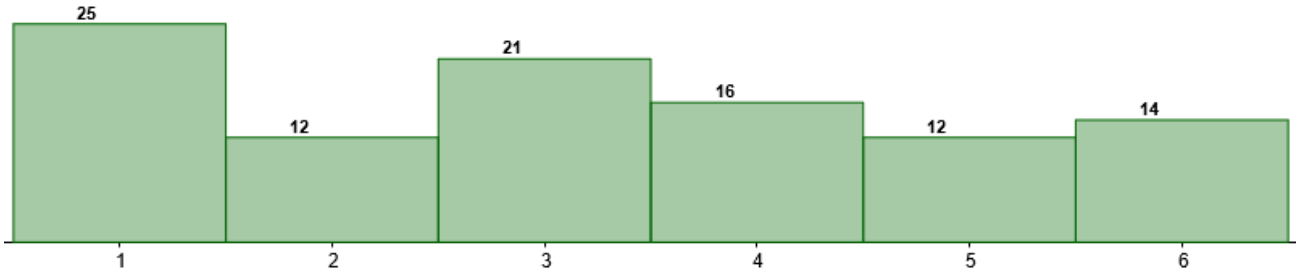
Chapter 3: Testing a Mean

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 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? Or is this still a reasonable result to see by random chance?

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/record with each simulation to 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!
  - For example, we examined the true proportion (π) of heads when flipping a dented coin, where the null hypothesis said that the true heads rate was _____.
  - Our estimator was a _____ ($\hat{p}$) from 100 tosses, with our sample estimate being of $\hat{p} = 0.58$.

- While each flip of the coin produces a binary result (heads or tails), rolling a die produces a_____ result! We could convert to a binary grouping (e.g., rolling a 1 or not), but maybe we should preserve the numeric value.
  - μ would represent the mean dice roll across ____ rolls we could make with this die.
  - The _____ ($\bar{x}$) from our 100 dice rolls would serve as our estimator for μ.

- Our Investigation
  - We're testing whether the die is fair or not. In this case, our Null hypothesis will be that all dice rolls we could ever make with this die should average to…
  - We took a sample of 100 dice rolls. Our sample estimate for the true mean is…

- The **Absolute Error** of our estimate
  - Absolute Error is the _____ between our specific estimate and the parameter
  - With the dented coin, if we assumed π really is 0.5, then absolute error is…
  - With Mack's die, if μ really is 3.5, then the absolute error in our $\bar{x}$ 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 under the 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!

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? How low should it be to fire Mack?

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 an estimator represents the _____ of that estimator.
- When testing a mean or proportion, the standard error is equivalent to the standard deviation of our _____.
    - The "Standard Error of a Sample Mean" is abbreviated as…
    - The "Standard Error of a Sample Proportion" is abbreviated as…

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: Consider if we were to take a sample of 200 dice rolls, then find the average roll from our sample of 200. Would we *expect* that sample mean to be more accurate, less accurate, or about equally accurate as a sample mean generated from 100 dice rolls?

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 200**	**Size 2000**
Minimum:	Minimum:	Minimum:
Maximum:	Maximum:	Maximum:
Standard Error:	Standard Error:	Standard Error:

Distributions and Convergence Properties

- **A Population Distribution** represents the entire distribution of a particular variable.
  - The population distribution for rolling a single die is *uniformly* distributed across 1, 2, 3, 4, 5, 6.
  - Another example might be the weights of all babies born at full term. This distribution is more symmetrically distributed with a mean right around 7 lbs.
- **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 _____ _____.

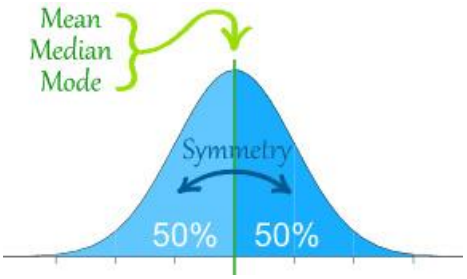
- If I take a sample of size n, and you take a sample of size n, we'll have different data and slightly different statistics. Imagine repeating this process many times—there's a distribution of possible statistics we could get! We call this "theoretical" distributional idea a **SAMPLING Distribution.**
  - 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.
  - We could also create a sampling distribution for $\hat{p}$ in a similar manner using a different sim.

- 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"?

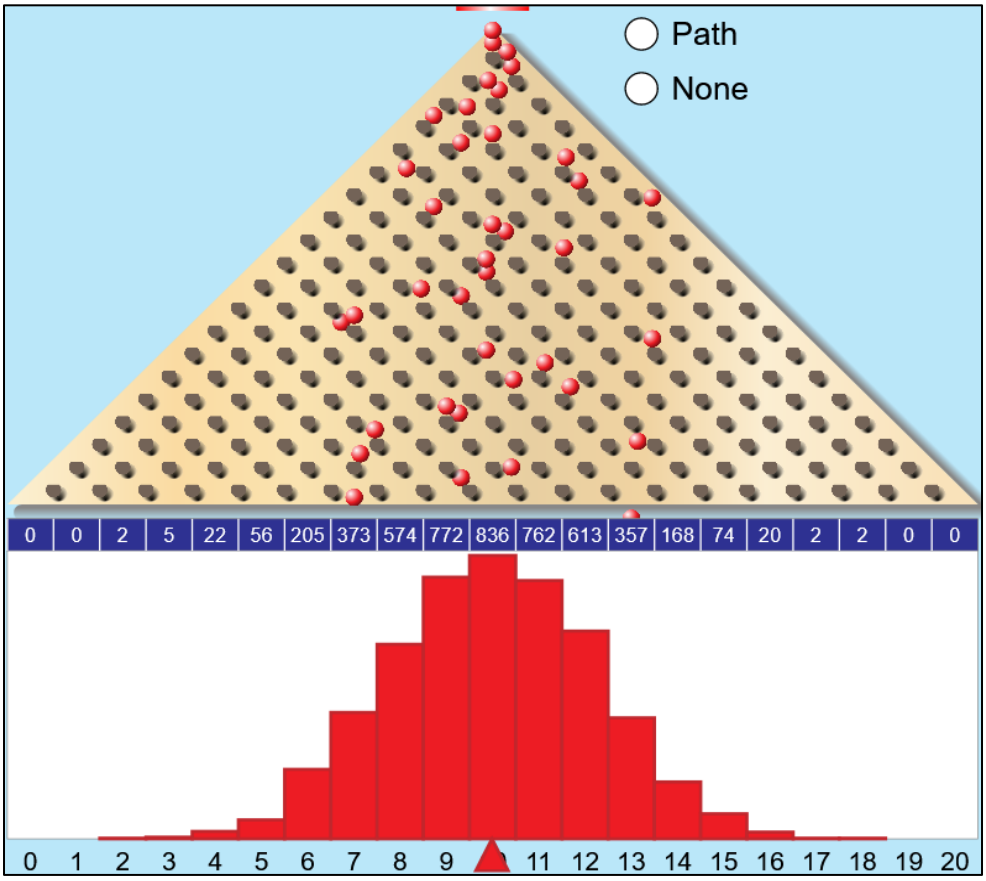
- A normally distributed variable has an identifiable bell-curve shape and shows up in a lot of places!



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."**

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 are taller than average.

Let's say we contacted a representative sample of high schools. We find that 24 of them had a male valedictorian this year, and the average height of those 24 students was 70.5 inches.

The U.S. Census has some general data on height. Using some posted data, I found that the heights of 18-year-old males is approximately "normally distributed" with a **mean** of **69.3 inches** and a **standard deviation** of **2.5 inches.**

What is the population we hope to generalize to?

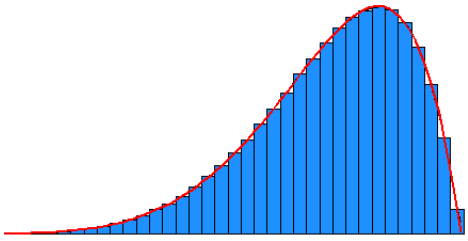
What is our research question?

How is this different from the loaded die investigation?

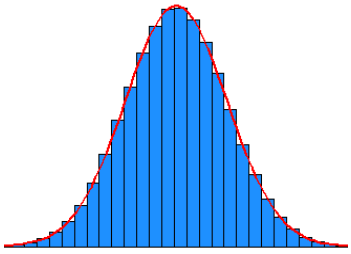
- When investigating Mack's die, the population distribution was a _____ variable, where we could model and sample from **each** possible value exactly.
- In this investigation, height is a _____ variable, where we can't model the probability of being every possible height value. *What's the probability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

Modeling a continuous population

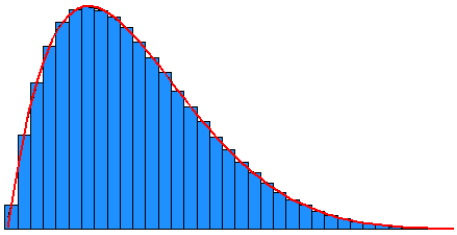
- Now that our variable (male valedictorian heights) is continuous, we have to identify the distributional shape we are sampling from.
- If we can identify a variable as approximating a known distributional function (like a "normal distribution"), then we can sample from this function and create a Null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



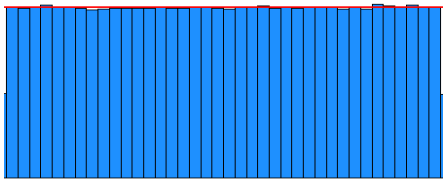
Left-Skewed Distribution



Normal Distribution



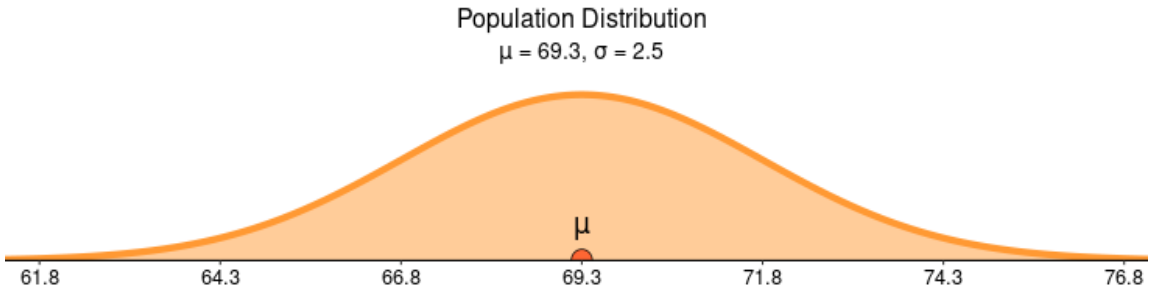
Right-Skewed Distribution



Uniform Distribution

Let's open the **Sampling Distribution Continuous Population** simulation from the Art of Stat Web apps page to model the distribution for 18-year-old male heights

- Choose "bell-shaped."
- Enter a custom mean and standard deviation to match the U.S. Census data.
- Set sample size to 24.



Completing our investigation

What is our sample estimate for the true mean?

What is our Null hypothesized mean?

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?

Simulate a Null Model. Using the "Find Probability" feature, what is your simulated p-value?

Let's say we have set our allowable Type I error to $\alpha = 0.02$. What should we conclude?

Example: The Art of Stat sim has some pre-loaded data involving the Airbnb prices for all New York City listings in 2019.

Before opening the sim…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 open up that example and draw the actual population distribution on the right.

Population Distribution (your guess)	**Population Distribution (actual)**

Let's take a random sample of 30 listings from this population and take note of our sample mean. As expected, we should note that the sample distribution resembles a bumpy version of the _____.

If we continued taking samples of size 30 and plotting the average price 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, μ. In other words, of $\bar{x}$ is an _____ estimator of μ.

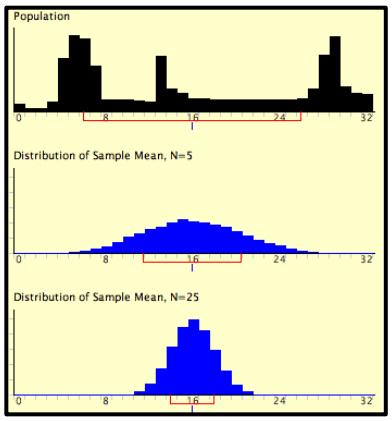
2) The standard deviation of the distribution of $\bar{x}$ (the "Standard Error of $\bar{x}$") can also be derived by calculating the ***population*** standard deviation divided by the square root of the sample size used to generate that sample mean.

$$SE_{\bar{x}} \text{ (also symbolized as } \sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}}$$

3) (a) If the population distribution is normally distributed, then the distribution of $\bar{x}$ is also normally distributed, regardless of the sample size n

3 (b) **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**!



If the population is not normal but *doesn't* have a large skew (long tail), then the distribution of $\bar{x}$ will typically be normally distributed at fairly small sample sizes.

- In our class, we'll use **n > 30** as a generic benchmark for cases with mild asymmetry, but not highly skewed.

If the population is **highly skewed** and has a **long tail**, then n might need to be _____ for $\bar{x}$ to be normally distributed.

- Determining a sufficient n will depend on just how skewed the population is (which could be estimated with a "resampling" approach in software).
- In our class, we'll use **n > 100** as a generic benchmark for variables that are highly skewed (though there are variables for which even n = 100 won't be enough!

If Time: Using the Art of Stat simulator, which of the following population distributions might require a sample size larger than 30 for the sampling distributions to be approximately normally distributed? If larger than 30, how large does n need to be?

World Life Expectancy: _____

Delay of Flights Arriving in ATL: _____

Bimodal Distribution: _____

Chapter 3 Reflection Questions

When trying to estimate a population proportion (π) 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 (μ)?

How did we statistically test whether Mack's die might reasonably be generating fair results? If we had set $\alpha = 0.05$ as our threshold for firing, what would we have done?

What is the difference between the **absolute error** of an *estimate* and the **standard error** of an *estimator*? For which calculation would we need to know (or assume) the parameter value?

What is the difference between a *sample data* distribution and a *sampling* distribution? Which one do we generally think about as a more "theoretical" idea?

Let's say that some variable is right skewed. If we were to take a sample of size 100 and plot those individual observations, what general shape would we expect that distribution to be?

Let's again take this same right skewed variable. If I were to repeatedly take samples of size 100 and plot the sample means I get each time, what general shape would we expect that distribution to take?

In your own words, how did we statistically test whether Mack's die seemed biased or not? *Hint, it involved a process similar to what the last question describes.*

What are the two different ways we learned to symbolize the standard error of a sample mean?

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 (if you need it!)

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: For this sim, use the "NYC Taxi rides" data from the Art of Stat sim.

Consider a NYC taxi company that wonders whether the average distance for rides is *different* on rainy days. The database in the simulation shows that trips tend to be an average of 2.44 miles across all weather events.

They collect some additional ride length data for 66 trips that occur while it's rainy and find the average ride length to be 2.72 miles.

What is the null hypothesized mean in this investigation?

What is the taxi company's estimate for the true mean in this investigation?

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

Open the Sampling Distribution Continuous Population sim and navigate to the "NYC Taxi Rides" dataset.

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

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!*

Make a conclusion. Should we reject the Null Hypothesis using a $\alpha = 0.05$ threshold of evidence?

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