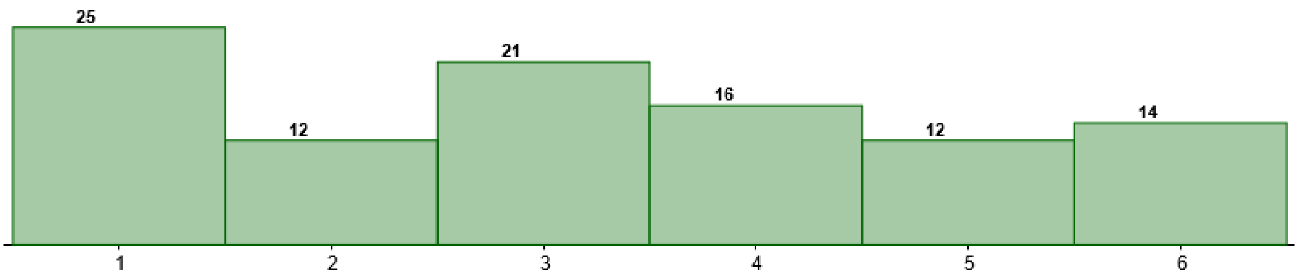
Chapter 3: Testing a Mean

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 suspect 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



Based on the data we collected, 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 - https://www.geogebra.org/m/UsoH4eNl

- 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 statistic from our data that estimates a parameter of interest!
  - With the dented coin, we were interested in testing whether the true proportion (π) of heads was still 0.50.
  - Our estimator was a sample proportion ($\hat{p}$), with our specific 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 numeric result!
  - While we could convert our numeric data into a binary grouping (e.g., rolling a 1 vs. not rolling a 1), maybe we'd like to preserve the numeric value.
  - We could take a sample mean ($\bar{x}$) as our estimator for the population mean (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…$\mu_0 = 3.5$
  - We took a sample of 100 dice rolls. Our sample estimate for the true mean is…$\bar{x} = 3.2$

- The Absolute Error of our estimate
  - Absolute Error is the distance between our specific estimate and the parameter
  - With the dented coin, the absolute error in our $\hat{p}$ if π really is 0.5 is…$|0.58 - 0.50| = 0.08$
  - With Mack's die, the absolute error in our $\bar{x}$ if μ really is 3.5 is…$|3.2 - 3.5| = 0.3$

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 Null Hypothesized Model!

Let's use the Art of Stat, sampling distribution sim to explore our question further: https://istats.shinyapps.io/SampDist_discrete/ - To get started…

- Select "Fair Die" as your Population Distribution
- Adjust the sample size to 100
- Draw a couple samples and notice what happens!

Let's complete our investigation!

Null Hypothesis: Mack's die should average to 3.5. Symbolically, we'd say…$\mu = 3.5$

- By running several thousand simulations, you can create a "Null Model" under this scenario.
- Under options, use the "Find probability" tool to estimate how often we would see a sample mean as low or lower than ours under the Null Model.

Do we have evidence against the Null Hypothesis? What do *you* think?

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 a statistic

- The **“Standard Error”** of a statistic can be described as Expected Error of that statistic as an estimator for its parameter.
- When testing a mean or proportion, the standard error is equivalent to the standard deviation of our Null Model
    - The “Standard Error of a Sample Mean” is abbreviated as… $\sigma_{\bar{x}}$ or $SE_{\bar{x}}$
    - The “Standard Error of a Sample Proportion” is abbreviated as… $\sigma_{\hat{p}}$ or $SE_{\hat{p}}$

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 estimate of 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?

We would expect it to be more accurate when drawing a larger sample

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.
  - The population distribution of heights from every adult female in the world is somewhat "Normally distributed" and centers somewhere around 63.7 inches.
- **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 "bumpy" version of the population!

- As sample size increases, your **sample** distribution will **converge in** shape to the population distribution.

DRAW SAMPLE DISTRIBUTIONS FOR 20, 200, AND 2000

- A **SAMPLING Distribution** is the general name for the distribution of a sample statistic. *A sampling distribution can act as a Null Model if we're doing a hypothesis test!*
  - In the dice example, we created a sampling distribution for $\bar{x}$ to see how $\bar{x}$ might vary for a particular sample size.
  - We could also create a sampling distribution for $\hat{p}$ for any of the previous proportions examples we've seen. For example. https://istats.shinyapps.io/SampDist_Prop/

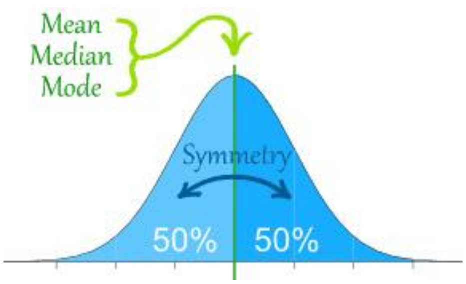
- As sample size increases, the **sampling** distribution of the sample mean will **converge in value** to the population mean.
  - This is known as the Law of Large Numbers, and a corollary to this law is that the variability in our estimator will decrease when using a larger and larger sample.
  - https://digitalfirst.bfwpub.com/stats_applet/stats_applet_11_largenums.html

DRAW SAMPLING DISTRIBUTIONS FOR 20, 200, AND 2000

Read on your own

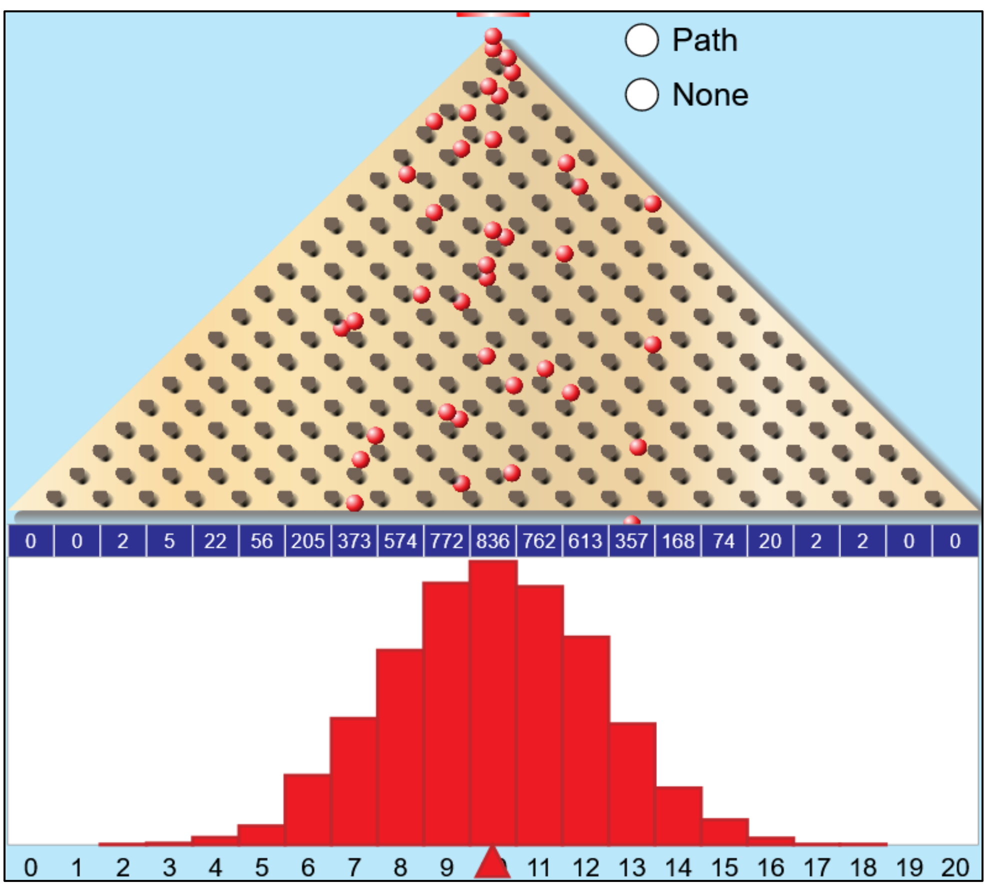
What does it mean to be “Normally distributed”?

- 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.
  - This symmetry also allows the balancing point (mean), the middle value (median), and the peak (mode) to be equal!
  - It slopes off gradually from the center as the probabilities approach 0 for larger discrepancies from the mean.

- 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 become decreasingly likely to occur (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—be careful not 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.

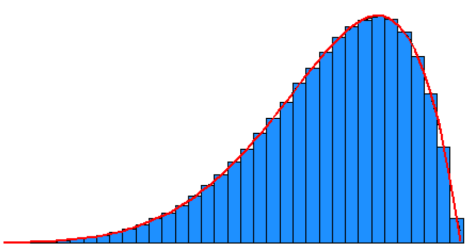
The U.S. Census has some general data on height. Using some posted data, I deduced that the heights of 18-year-old males is approximately "normally distributed" with a **mean** of **69.3 inches** and a **standard deviation** of **2.5 inches.**

How is this different from the loaded die investigation?

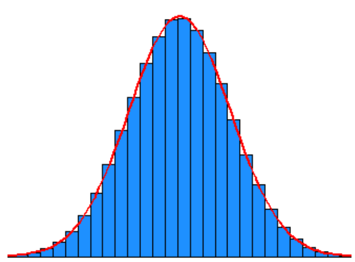
- When investigating Mack's die, the population distribution was a <u>discrete</u> variable, where we could model and sample from each possible value exactly
- In this investigation, height is a <u>continuous</u> variable, where we can't model the probability of being every possible height value. *What's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

Modeling a continuous population

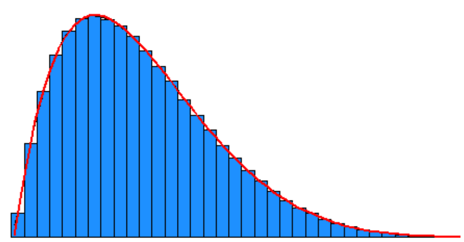
- There are many distribution shapes that continuous variables can take.
    - These can be modeled with mathematical functions!
    - Deriving these functions or understanding how they relate to a particular shape is beyond the scope of this course.
- If we can identify a variable as approximating a known distributional function (like a "normal distribution"), then we can sample from this function and create a Null model!



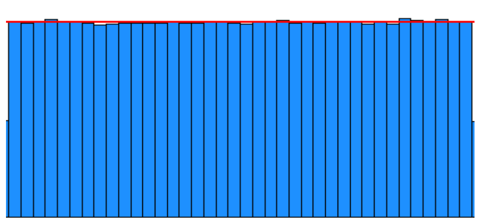
Left-Skewed Distribution



Normal Distribution



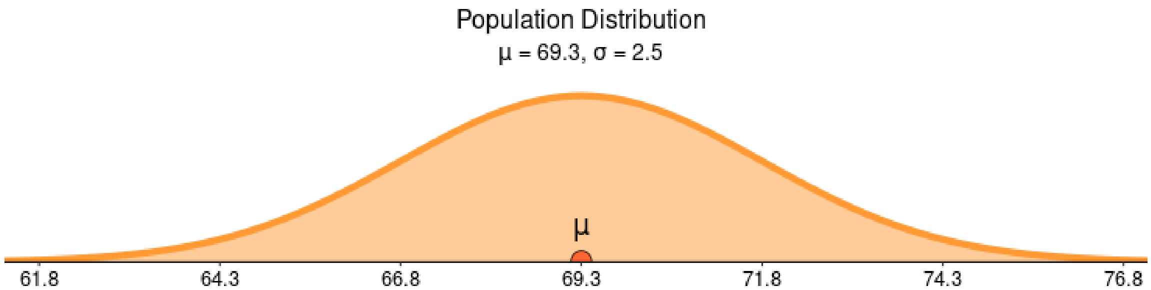
Right-Skewed Distribution

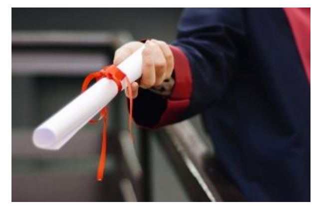


Uniform Distribution

Let's open the Sampling Distribution sim for **"Continuous"** Population Distribution to model the distribution for 18-year-old male heights - https://istats.shinyapps.io/sampdist_cont/

- 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? $\bar{x} = 70.5$

What is our Null hypothesized mean? $\mu_0 = 69.3$

If the null were true…what would be the absolute error of our sample mean? 70.5 – 69.3 = 1.2 inches

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?* Directional

How would we write our Null and Alternative Hypothesis for this investigation?

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

Make a conclusion. Should we reject the Null Hypothesis using a $\alpha = 0.02$ threshold of evidence?

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 population distribution.

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*

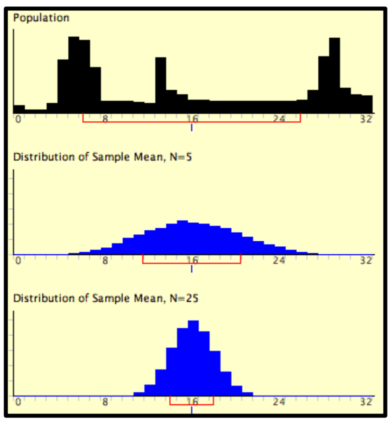
1) The distribution of $\bar{x}$ will center around the true population mean, μ. In other words, of $\bar{x}$ is an "unbiased" estimator of μ

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.

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

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 large enough.



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 these cases!

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

- Determining a sufficient n will depend on just how skewed the population is.
- Re-sampling with replacement from your sample can be used to determine if $\bar{x}$ might reasonably be normally distributed!

Practice: Using the Art of Stat simulator, which of the following population distributions might require a sample size larger than 30 in order 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 Additional Practice

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 for continuous variables sim: https://istats.shinyapps.io/sampdist_cont/ 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 α = 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 - https://www.youtube.com/watch?v=zeJD6dqJ5lo&t=498s

Chapter 3 Learning Goals

After this chapter, you should be able to…

- Identify $\hat{p}$ as a common estimator for π and $\bar{x}$ as a common estimator for μ
- Calculate the absolute error in a statistic as an estimate for a parameter
- Use a Null Model to interpret the likelihood of different sample means under a Null Hypothesis
- Understand and calculate the standard error of a sample statistic
    - Recognize it informally as the expected amount of error in a statistic when the parameter is unknown
    - Understand it conceptually as the standard deviation of all possible sample statistics we could generate for a particular sample size
    - Calculate it using the formula
- Distinguish between a population distribution, sample data distribution, and sampling distribution
- Notice various properties about sample and sampling distributions as n increases
    - Sample distribution will converge in shape to the population distribution
    - Sampling distribution will converge in value to the parameter it estimates—likewise, the standard error of a statistic will decrease
- Recognize visually when variables are normally distributed, and contrast this with uniform, left skewed, right skewed, and other asymmetric distribution shapes
- Remember characteristics about the sampling distribution for the sample mean
    - It will be centered around μ *($\bar{x}$ is an unbiased estimator of μ)*
    - The standard deviation of the sampling distribution for $\bar{x}$ is abbreviated as $\sigma_{\bar{x}}$ or $SE_{\bar{x}}$ and can be calculated using the formula provided in the notes.
    - The distribution will tend towards "normal," even when the population is not normally distributed *(see the Central Limit Theorem)*