Decoding Normal Distribution: Understanding Data and Risk

Contents

Title: Decoding Normal Distribution: Understanding Data and Risk

Overview

This lesson introduces students to the concept of normal distribution in statistics. It covers how to draw accurate conclusions and assess risk using the empirical rule, understanding the distribution of data within standard deviations from the mean.

Learning Standards

Aligns with MA19.A2.23: Use mathematical and statistical reasoning about normal distributions to draw conclusions and assess risk; focus on informal arguments.

Outcomes

- Comprehend how to draw conclusions and assess risks using knowledge of normal distribution.
- Understand that nearly all data falls within three standard deviations of the mean.
- Grasp the empirical rule: 68% of data within one standard deviation, 95% within two, and 99.7% within three standard deviations from the mean.

Duration

Approximately 80 minutes

INTRODUCE. Gauge students' understanding of normal distribution.	20 min.
EDUCATE. Explain normal distribution and empirical rule.	30 min.
INSPIRE. Apply normal distribution to relevant contexts.	20 min.
ASSESSMENT. QuantHub scavenger hunts	10 min.

Materials

Learning Resources:

- https://www.quanthub.com/exploring-common-statistical-distributions/ (Appendix A)
- https://www.quanthub.com/exploring-distributions-in-data-analysis/ (Appendix B)
- https://www.quanthub.com/understanding-the-role-of-distribution-in-statistical-analysis/ (Appendix C)

Activity Scripts for Teachers (Appendices D-E)
Independent Study Dataset (Appendix F)
Central Tendency Calculation Cheat Sheet (Appendix G)
Central Tendency Calculation Worksheet (Appendix H)

Background Knowledge

Basic Statistics Concepts:

- **Mean** (Average): The sum of all data points divided by the number of points. It's the central value of a data set.
- **Median**: The middle value in a sorted list of numbers.
- **Mode**: The most frequently occurring value in a data set.
- **Range**: The difference between the highest and lowest values in a data set.
- **Variance**: A measure of how much each number in a set differs from the mean.
- **Standard Deviation**: A measure that is used to quantify the amount of variation or dispersion in a set of data values. It's the square root of the variance.

Concepts to Emphasize

- **Relevance of the Mean and Standard Deviation**: Explain how these two measurements are essential in understanding the distribution of data in a normal distribution.
- **Application of the Empirical Rule**: Discuss how the empirical rule can be used to draw conclusions and assess risks in real-world scenarios. For instance, understanding how grades are distributed in a class or how height varies in a population.
- **Normal Distribution in Everyday Life**: Provide examples of normal distribution seen in everyday life, like test scores, heights, or measurement errors.

Introduce Activity (20 minutes)

The activities in the Introduce phase are designed to help students make connections between past and present learning experiences, expose prior conceptions, and organize thinking toward the essential questions and learning outcomes of the learning sequence.

Teacher Objective: The teacher should facilitate discussions, encourage participation from all students, and gently guide the conversation towards understanding variability and distribution in everyday life. Choose one or more of the below strategies.

Strategy: Interactive Questioning (5-10min)

“Let's start by discussing what you already know about variation. Can anyone give me an example of variation in our everyday lives? How about something related to the heights of people in our class?”

Students might respond with the following examples:

- "I think variation in height is when some people in our class are really tall, like Sarah and John, while others, like Emma and David, are shorter."
- "It's like how some students in our class are almost as tall as the teacher, and some are much shorter. That's variation in height."

Strategy: Group Brainstorming. (10 min)

Ask students to form small groups and brainstorm why most heights cluster around a certain value. Encourage them to think about variability in heights.

You can ask the following probing questions to get students considering different aspects:

- Biological Factors:
    - What genetic factors might contribute to people having similar heights?
- Variability and Distribution:
    - Why do you think some people are much taller or shorter than average? What could cause these variations?

Use the following strategy to bridge the discussion to the Educate Activity.

Strategy: Introducing the Concept. (10 min)

Use individual or group responses to segue into the concept of normal distribution, explaining that it's a way to understand how data like heights is distributed. (Example script in Appendix D.)

Educate Activity (30 minutes)

This activity helps Explain the topic to students in two parts. First, students are asked to share their initial models and explanations from experiences in the Introduce phase. Second, resources and information are provided to support student learning and introduce scientific or technological concepts. Students use these resources and information, as well as ideas of other students, to construct pr revise their evidence-based models and explanations.

Teacher Objective: The teacher should ensure that the explanation is clear and accessible, using visual aids effectively and checking for understanding regularly. Use the https://www.quanthub.com/exploring-common-statistical-distributions/ to introduce the concept of normal distribution. Explain the empirical rule, using visual aids like histograms and density plots from the document. Example script in Appendix E.

Perform the following strategies (Independent learners can read the script in Appendix E):

Strategy: Visual Introduction. (10 min)

Example Script: "Let's delve deeper into the characteristics of a normal distribution curve. This curve is symmetric around its center, representing the average score. In a normally distributed set, the mean, median, and mode are identical. Observe the bell shape of the curve, indicating how scores are spread around the average."

Appendix E provides the following visuals:

1. Diagram illustrating the mean, median, mode, and symmetry in the context of the test scores.
2. Density plot showing the characteristics of normal distribution with respect to the test scores.

Example Guiding Question: "Why do you think the mean, median, and mode are the same in a normal distribution of test scores?"

Strategy: Conceptual Explanation. (10 min)

Example Script: "Let's examine the empirical rule, also known as the 68-95-99.7 rule, in the context of these reading scores. This rule indicates how scores are spread around the mean: about 68% within one standard deviation, 95% within two, and 99.7% within three standard deviations."

Appendix E provides the following visuals:

1. **Visual Example: Diagram showing the empirical rule applied to the test score distribution]**

Check for understanding: "Based on this rule, how much data would you expect to fall within two standard deviations of the mean in our test score example?"

Strategy: Interactive Q&A. (10 min) Example Script: "I encourage everyone to ask questions or share your thoughts about the empirical rule. How do you think this rule helps in understanding the distribution of data, like test scores, in real-world scenarios?” Example responses: 1. "This rule is important because it helps us understand the performance of a large group of students. For instance, it can indicate how many students are performing within the expected range." 2. "This rule can also help identify outliers in data. Outliers are unusual scores that don't fit the general trend. Spotting outliers can be crucial in educational settings to provide additional support or challenge to those students."	**Independent Learning Strategy: Written Reflection. (10 min)** Consider the empirical rule as a powerful tool to analyze data distribution, like test scores. Reflect on how this rule could help you understand the performance spread within a large group. Can you think of ways it might reveal the majority who fall within the expected range? Write a 50-word max. response.

Inspire Activity (20 minutes)

These activities provide time for students to apply their understanding of concepts and skills. They might apply their understanding to similar phenomena or problems. This activity is meant to provide students an opportunity to reflect on the lesson and examine essential questions.

Teacher Objective: The teacher should provide guidance on data collection and analysis, facilitate group discussions, and ensure that each group understands and correctly applies the empirical rule.

Perform the following strategies:

Strategy	Independent Learner Strategy
Strategy: Real-World Data Collection. Divide the class into small groups of 4-5 students. Each group will choose a real-world variable that they believe follows a normal distribution that they can collect data on while in class. Examples include: • Hand Span: The distance from the thumb to the little finger when the hand is fully stretched. This can vary widely and is easy to measure with rulers or tape measures. • Arm Span: Measuring the arm span (from fingertip to fingertip with arms stretched out) of each student. It's interesting to compare this with height, as they are often similar in length. • Pulse Rate: Students can measure their pulse rate over a minute under resting conditions. This biological variable can sometimes follow a normal distribution. Students will collect and record this data themselves (e.g., measuring heights).	**Independent Learner Strategy: Real-World Data Collection.** Learner should engage with the data set provided in Appendix F. The learner should add their information to the final cell.

Strategy: Data Analysis and Visualization.

Using the collected data, each group or individual will create a histogram to visualize the distribution.

Students will then calculate the mean and standard deviation of their data set. (Worksheet for calculations attached)

Strategy: Application of the Empirical Rule.

Groups or individual students will apply the empirical rule to their data, determining what percentage of their data falls within one, two, and three standard deviations from the mean.

They will annotate their histogram to show these intervals.

Strategy: Presentation and Discussion.

Example script: "After seeing these presentations, let's discuss as a class. How does the empirical rule help us understand the variability in real-world data? What are some limitations we might encounter?" Facilitate a discussion with students.

Example responses include:

- "I think the empirical rule is really useful because it tells us how data is spread out. It's like a guideline for understanding averages and extremes.
- "One limitation I can think of is what if the data isn't normally distributed? Like if we were looking at income, it's not usually a perfect bell curve, so the rule might not work well."

Assessment (10 minutes)

QuantHub Scavenger Hunt for

- https://www.quanthub.com/exploring-common-statistical-distributions/
- https://www.quanthub.com/understanding-the-role-of-distribution-in-statistical-analysis/
- https://www.quanthub.com/exploring-distributions-in-data-analysis/

Acceleration and Intervention Strategies

- For advanced students: Provide more complex datasets or challenge them to apply the concept to assess risks in real-life scenarios.
- For students needing support: Offer step-by-step guided practice with simpler datasets and additional one-on-one explanation.

Specific Example/Context: Buying a Car

Using normal distribution to assess the financial risk associated with the potential resale value of the car. Here's how this can be applied:

1. **Researching Resale Values**: Start by researching the historical resale values of the car model you are interested in. Gather data on how much the car model has been sold for in the past few years after being used for a similar duration and mileage.
2. **Analyzing the Data with Normal Distribution**: Notice that the resale values roughly follow a normal distribution. Most of the cars have been sold for a price around the average, with fewer cars being sold for significantly higher or lower prices.
3. **Calculating Mean and Standard Deviation**: Calculate the mean (average resale price) and the standard deviation (a measure of how spread out the resale prices are) of the data. This gives an idea of the expected resale price of the car and the variability in the prices.
4. **Assessing Financial Risk**: Using the standard deviation, you can assess the financial risk of your investment in the car. A higher standard deviation means there's a higher risk and uncertainty about the resale value of the car. A lower standard deviation suggests more predictability and lower risk.
5. **Making an Informed Decision**: Use this information to decide whether to buy the car. If the standard deviation is high, and you are risk-averse or have a tight budget, you might reconsider or look for a car model with a more predictable resale value. If you are willing to take the risk for a potentially higher resale value, you might proceed with the purchase.
6. **Applying the Empirical Rule**: Use the empirical rule (68-95-99.7 rule) to estimate the probability of the car's resale value falling within certain price ranges. This helps understand the likelihood of getting a good return on your investment when you decide to sell the car.

By applying the concepts of normal distribution to the scenario of buying a car, you can make a more informed decision, understanding the financial risks involved in terms of the potential variability in the car's resale value. This example demonstrates how mathematical concepts like the normal distribution can be applied in practical, real-world financial decisions.

Statistics can seem like a complex subject, but at its heart, it's about understanding how different things are spread out or distributed. Let's break down some of the most common types of statistical distributions to make them more understandable.

Normal Distribution (Gaussian Distribution)

What It Is: Imagine a graph that's bell-shaped and perfectly symmetrical. That's a normal distribution used for things that vary around an average.

Key Feature: This distribution follows the 68-95-99.7 rule – about 68% of values are within 1 standard deviation from the mean, 95% within 2 standard deviations, and 99.7% within 3 standard deviations. In a normal distribution, this rule allows us to predict the likelihood of a data point falling within a certain range. This predictability is crucial in many fields, such as quality control, finance, and social sciences, where understanding the variability and predictability of data is essential.

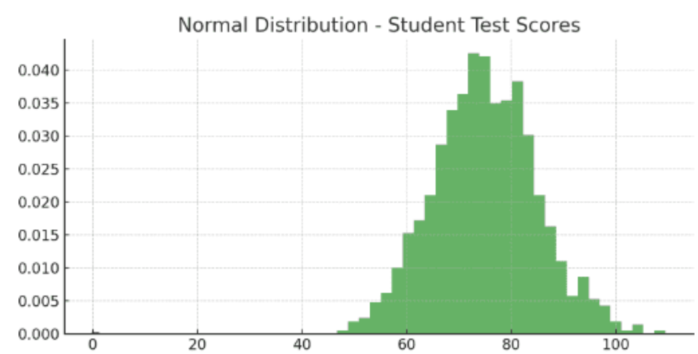
When to Use: Great for continuous data that clusters around a mean. It's common in natural and social sciences, especially with large samples.

Example: Imagine a high school where we are looking at students' scores on a standardized math test. After compiling all the scores, we find that they form a normal distribution. The average (mean) score is 75 out of 100, with a standard deviation of 10 points.

According to the Empirical Rule:

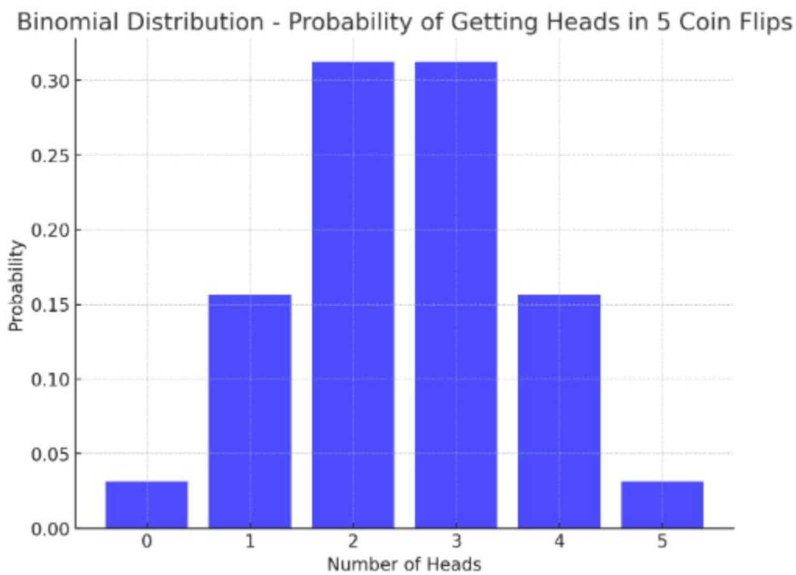
- About 68% of the students scored between 65 (75 – 10) and 85 (75 + 10).
- Around 95% scored between 55 (75 – 2×10) and 95 (75 + 2×10).
- Nearly 99.7% scored between 45 (75 – 3×10) and 105 (75 + 3×10).

This means most students scored close to the average, with fewer students obtaining very high or very low scores.



Binomial Distribution
What It Is: This distribution counts the number of successes in a fixed number of trials. Each trial has only two possible outcomes – success or failure.
Key Feature: Useful when you're dealing with two outcomes like yes/no, pass/fail, or win/lose.
When to Use: Perfect for situations with two outcomes like pass/fail. It's used when you want to know the probability of a certain number of successes in a series of independent trials.
Example: Consider a game where a player f1ips a coin 5 times, and we count how many times the coin lands on heads (success). If the coin is fair, the probability of getting heads in a single f1ip is 0.5 (50%).

If we define X as the number of heads (successes) in 5 f1ips, X follows a binomial distribution.
The probability of getting exactly 3 heads, for example, can be calculated using the binomial formula.

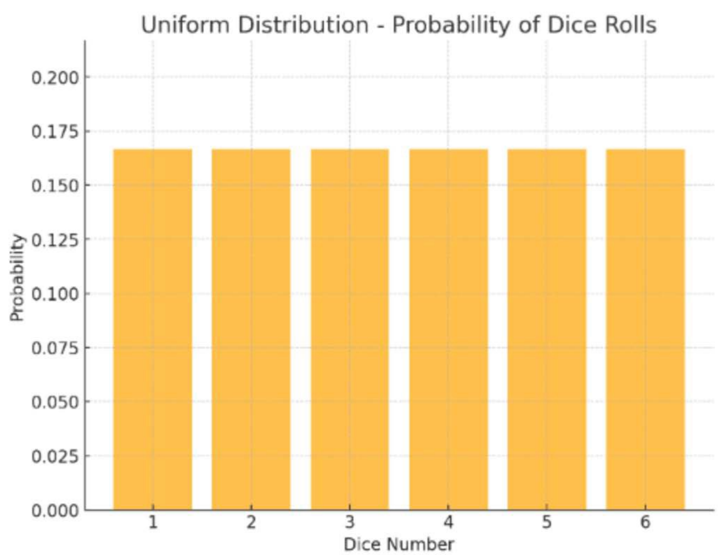


Number of Heads (Successes)	Probability
0	0.03125
1	0.15625
2	0.3125
3	0.3125
4	0.15625
5	0.03125

Uniform Distribution

What It Is: All outcomes are equally likely to happen. Think of a perfectly f1at line on a graph. Example: Rolling a fair die. The outcomes 1, 2, 3, 4, 5, and 6 are all equally likely.
Key Feature: Every value between a minimum and maximum is just as likely as any other.
When to Use: If every outcome in a scenario is equally likely, like in simulations or games of chance, this is your go-to distribution. Example: Let's say we have a perfectly fair six-sided die. Each side, numbered 1 through 6, has an equal probability of landing face up.

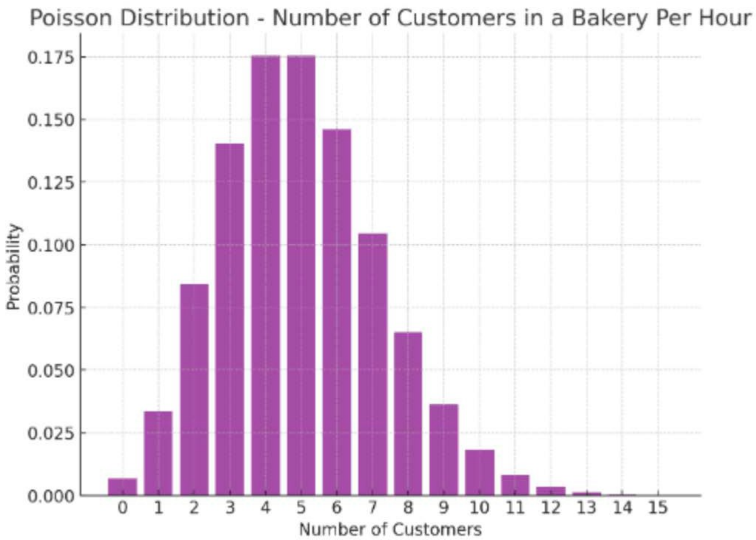
- In this case, the probability of rolling any specific number (1, 2, 3, 4, 5, or 6) is exactly the same, 1/6.
- The uniform distribution here is discrete since there are a finite number of equally likely outcomes.



Poisson Distribution

What It Is: Used to count the number of times something happens over a set period or area.
Key Feature: Good for counting events, especially when they're spread out randomly over time or space.
When to Use: When you're counting the number of events in a set time or space and these events are independent of each other, like the number of emails you get in an hour.
Example: A small bakery observes the number of customers arriving during a particular hour every day. Let's say they record an average of 5 customers per hour.

If we let X be the number of customers arriving in an hour, X can be modeled by a Poisson distribution with a rate (λ) of 5 customers per hour.
Using this model, the bakery can calculate probabilities for different numbers of customers arriving in an hour, like the probability of exactly 8 customers arriving.



Exponential Distribution

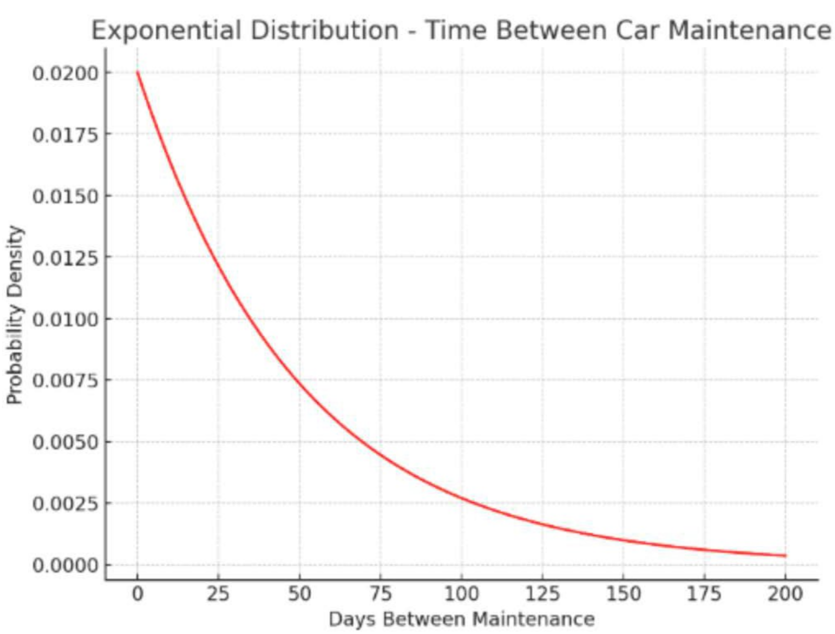
What It Is: It's all about the time between events, especially when those events happen at a constant rate but independently of each other. Key Feature: Often used to model waiting times or lifespans.

When to Use: Ideal for modeling the time until an event happens, like how long before a machine breaks down. It's widely used in survival analysis and reliability engineering.

Example: A car rental service is analyzing the time between car maintenance appointments. They find out that, on average, a car needs maintenance every 50 days.

Here, the time until a car needs maintenance follows an exponential distribution.

With a mean time of 50 days, the service can estimate the probability of a car needing maintenance within a certain time frame, like within the next 30 days.



**The behavior of the exponential distribution might initially seem counterintuitive when thinking about the probability of events like car maintenance. However, it's important to understand that the exponential distribution models the time until the next event (e.g., the need for maintenance) occurs, and it assumes that events happen independently of each other at a constant average rate.*

Understanding these distributions can be super helpful in various fields, from science to economics. Each distribution has its unique characteristics and applications, making them valuable tools in your statistical toolbox. So, next time you're faced with a data set, think about which distribution might best describe it!

****Fun Fact** Poisson Distributions and Calvary Getting Kicked in the Head**

A Surprising Application: An interesting application of the Poisson distribution, often mentioned in historical contexts, relates to the number of Prussian cavalry officers killed by horse kicks each year.

Historical Data: The distribution was used to model this seemingly random and rare event. Historical data from the Prussian army showed the number of such incidents over a period of years, and it was found that these fatalities followed a pattern that could be described by the Poisson distribution.

Significance: This example became famous because it was one of the early instances where a mathematical model (the Poisson distribution) was used to describe a real-world phenomenon that involved discrete events occurring at a low, constant rate in a fixed interval of time or space.

Appendix B: Exploring Distributions in Data Analysis

Data analysis is like detective work, where you search for clues hidden in numbers and patterns. One of the first steps in this exciting journey is exploring distributions. Let's dive into what this means and how it can be done, especially if you're a high school student at an 8th-grade reading level.

What Does It Mean to Explore Distributions?

Starting with frequency tables

Before you get into the nitty-gritty of data distribution, start with something simple: frequency tables. Imagine we're looking at a soccer league and analyzing the number of goals scored by players in a season. A frequency table will show the number of goals scored by players. This is a great way to get a first look at your data and spot any interesting trends or unusual patterns.

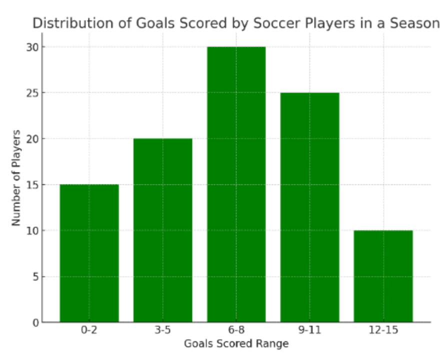
Collect and organize data

Next, gather all the data you need. This could mean doing surveys, setting up experiments, or just observing something and taking notes. Once you've got your data, put it in order so it's easier to work with. You can use a spreadsheet or special software to help with this.

Goals Scored Range	Number of Players
0-2	15
3-5	20
6-8	30
9-11	25
12-15	10

Visualize the data

A picture is worth a thousand numbers! Create graphs like histograms, box plots, or density plots to see what your data looks like. These graphs can tell you a lot about your data, like where most of the values are bunched up (the central tendency), how spread out they are (the spread), and if they lean more one way than another (the skewness).



Calculate descriptive statistics

This is where you crunch some numbers to get a better understanding of your data. Find out the average (mean), the middle value (median), and the most common value (mode). Also, look at how spread out your data is by calculating the range, interquartile range, standard deviation, and variance. These calculations can tell you a lot about the shape and characteristics of your data.

- Mean (Average) Goals Scored: The average number of goals scored by a player in the season is approximately 6.9.
- Variance: The variance of goals scored is about 13.67. This number gives us an idea of how much the number of goals scored by each player varies from the average.
- Standard Deviation: The standard deviation is approximately 3.70. This value tells us, on average, how much individual goal-scoring performances differ from the mean number of goals scored.

Identify the type of distribution

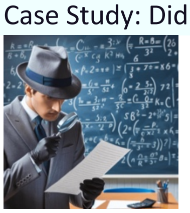
Now, combine what you've seen in your graphs with the numbers you've calculated. This can help you figure out what kind of distribution your data might have. For example, if your data looks like a bell curve and is evenly spread out, it might be a normal distribution. If it's about counting successes in a set number of tries, like getting heads in coin flips, it might be a binomial distribution.

Conduct further analysis

Once you know what type of distribution you're dealing with, you can do more detailed analysis. This might include testing hypotheses (like checking if one player's average score is really different from another's), building models to predict future trends, or other statistical methods that fit your data's distribution. The kind of analysis you do depends on what you want to find out from your study.

Exploring distributions is a crucial step in understanding what your data is trying to tell you. It's like piecing together a puzzle – first, you lay out all your pieces (collect and organize data), then you start to see how they fit together (visualize and calculate statistics), and finally, you get a clear picture of what it all means (identify the distribution and conduct further analysis). With these steps, you're well on your way to becoming a data analysis pro!

Case Study: Did the Math Test Get Leaked?



A high school teacher, Mr. Anderson, is investigating potential cheating in a recent exam. He suspects that some students may have had access to the exam questions beforehand. To assess this, he employs statistical analysis, focusing on the distribution and patterns of the students' answers.

Mr. Anderson collects the answer sheets of all students. He notes the frequency of each answer choice for each question, organizing this data into frequency tables. He pays particular attention to questions where the majority of students chose the same answer, especially if it's an uncommon or difficult question.

To better understand answer patterns, Mr. Anderson creates visualizations. He uses bar charts to represent the frequency of chosen answers for each question. He also employs heat maps to visualize patterns, such as clusters of students who have identical or nearly identical sets of answers, which could suggest copying or shared information.

Mr. Anderson calculates descriptive statistics for the answers, such as the mean, median, and mode for each question. He also looks at the variance and standard deviation to understand the spread and variability in the answers. He compares these statistics with the expected performance based on previous exams and class performance, looking for anomalies such as unusually high scores on typically hard questions.

He analyzes the distribution of the answers. Under normal circumstances, he expects a certain degree of variability in student responses, especially on more difficult questions. If he finds that the distribution of answers is unusually narrow (i.e., many students selecting the same answers), particularly on questions that typically have a wide spread of responses, it may suggest that the students had prior access to the exam questions.

Appendix C: Understanding the Role of Distribution in Statistical Analysis

In the world of statistics, “distribution” is a term that often pops up. But what does it mean, and why is it so important, especially when we’re trying to make sense of data? Let’s dive into the world of distributions and see how they help us in statistical analysis.

What is a Distribution in Statistics?

A distribution is like a summary of all the possible values or ranges of values in a dataset and how frequently they appear. Think of it as a way to organize and understand your data.

How Does Distribution Help?

A glimpse into data patterns

Distributions give us a snapshot of our data. They reveal patterns, showing us things like the average value (central tendency), how spread out the values are (variability), and whether more values are on one side of the average than the other (skewness).

A basis for predictions

Distributions are the backbone of statistical models. They help us make educated guesses or predictions about larger groups (populations) based on a smaller sample of data.

Identifying distribution types

Knowing whether your data follows a normal distribution, a binomial distribution, or another type can guide us in choosing the right statistical tests and understanding our data better.

Understanding probability and outcomes

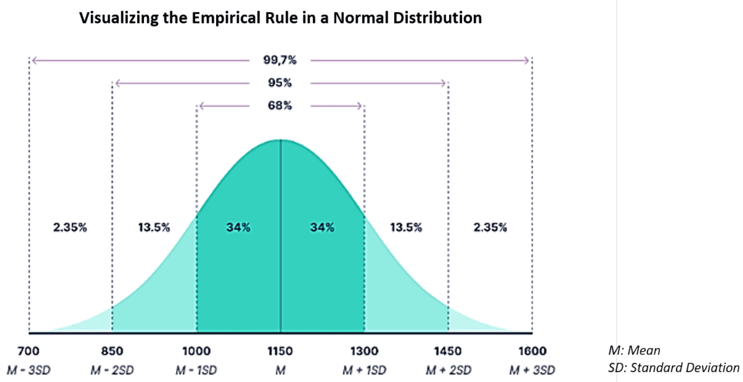
Distributions and probabilities go hand in hand. They help us understand how likely different outcomes are. For example, in a normal distribution, most data points are near the average, with fewer as you move away from it.

Characteristics of Common Distributions

Let’s look at some common distributions and their unique features.

Normal Distribution

The normal distribution, also known as the bell curve, is symmetric, and most of the data clusters around the mean (average). It follows the 68-95-99.7 rule (empirical rule), meaning 68% of data falls within one standard deviation of the mean, 95% within two, and nearly all (99.7%) within three standard deviations.
**Remember, a standard deviation tells us how spread out the numbers are in a set of data. It’s a way to measure how much the data varies from the average (mean).*



Binomial Distribution

This distribution comes into play when you're looking at the number of successes in a set number of tries (like flipping a coin). Its shape can be symmetric or skewed, depending on the likelihood of success in each try. Look at the symmetry below when the probability of the number of flipping 1 "heads" is around 0.5. Which way would it skew if it is closer to 0 or 1?

Table 1. Four Possible Outcomes.

Outcome	First Flip	Second Flip
1	Heads	Heads
2	Heads	Tails
3	Tails	Heads
4	Tails	Tails

Table 2. Probabilities of Getting 0, 1, or 2 Heads.

Number of Heads	Probability
0	1/4
1	1/2
2	1/4

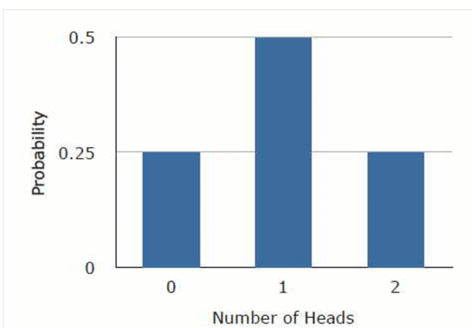
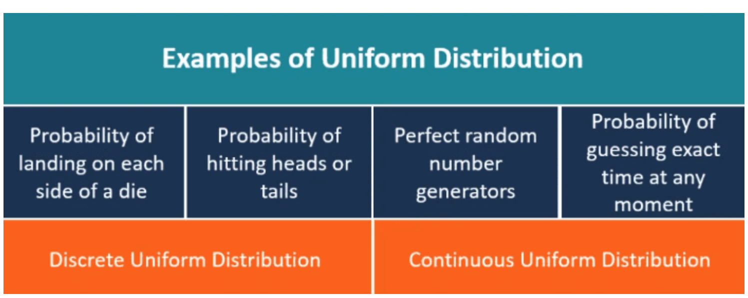
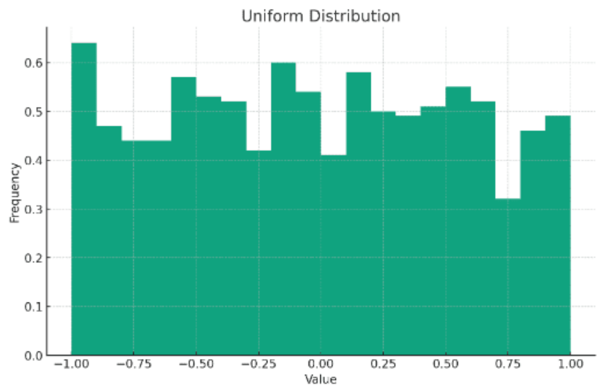


Figure 1. Probabilities of 0, 1, and 2 heads.

Uniform Distribution

Imagine every outcome having the same chance of happening – that's a uniform distribution for you. Its graph looks like a flat rectangle, showing equal probability for all outcomes.

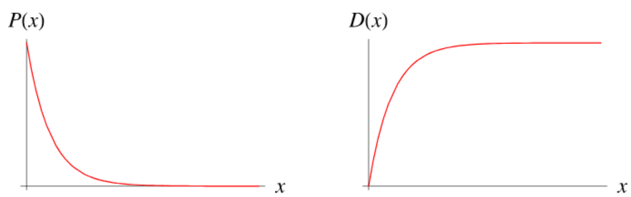




**Note: The above chart demonstrating uniform distribution may not appear to indicate the same probability for each value as it is not a flat rectangle. As the sample size increases, the distribution becomes flatter and flatter, eventually reaching the flat rectangle visualization. It is important to know that the sample size will impact the representation of the distribution, but it is still a uniform distribution.*

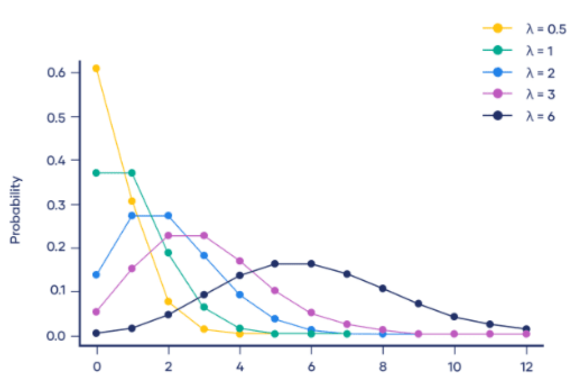
Exponential Distribution

Used to model the time between events, the exponential distribution is skewed to the right. It's useful in scenarios like measuring the time until the next earthquake or the wait time in a queue.



Poisson Distribution

Perfect for counting events in a fixed period, like the number of emails you receive in an hour. The Poisson distribution can be right skewed for low average numbers of events but becomes more symmetrical as the average increases.



Distributions are a fundamental concept in statistics, helping us make sense of data and predict future trends. Whether it's the bell-shaped curve of the normal distribution or the right-skew of the Poisson distribution, each type offers a unique lens through which we can view and understand our data.

Fun Fact Uniform Distribution of Penguins

Emperor penguins exhibit uniform distribution patterns: these territorial birds are equally spaced apart. Uniform distributions are found in populations where the distance between neighboring individuals is maximized. The need to maximize the space between individuals often arises from competition for resources, such as food or space, or as a result of direct social interactions between individuals within the population, such as territoriality. For example, Emperor penguins often exhibit uniform spacing by aggressively defending their territory among their neighbors.

Appendix D: Introduce Activity Script

Thank you all for sharing your thoughts on how heights in our school might be distributed. It's interesting to hear your perspectives on why heights vary and where most of them might cluster. Now, let's delve into a concept in statistics that helps us understand this kind of data distribution more scientifically. This concept is known as the normal distribution.

Imagine we actually measured everyone's height in the school and plotted them on a graph. You might expect some very tall and very short individuals, but most students would likely be around a certain height, right? This is where normal distribution comes into play.

Normal distribution, sometimes called the bell curve, is a way to show how data is spread out. It tells us that most of the data points - in our case, the heights - tend to cluster around a middle value, which we call the mean. This curve is symmetrical, and as we move away from the mean, the number of data points decreases.

Now, why is this important? Understanding normal distribution allows us to make predictions and conclusions about data. For instance, if we know the average height and the spread of heights in our school, we can predict how common or rare certain heights are.

But it's not just about heights. This concept applies to many areas - from test scores to measuring natural phenomena. Today, we'll explore how to calculate this mean, how to understand what these curves tell us, and how we can use this information in real-life situations. Let's dive in and see how normal distribution helps us make sense of the world around us through numbers and data."

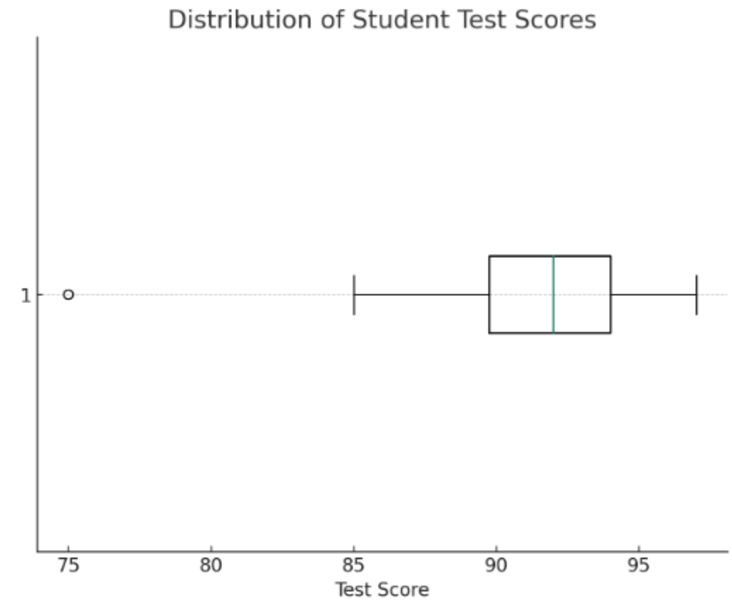
Appendix E: Educate Activity Script

Today, we're going to explore a fundamental concept in statistics – the normal distribution. Let's start by understanding why we look at distributions in the first place. When we have a set of data, like test scores or heights, we want to know how this data is spread out. Are most of the values clustered around a central point? Are there any extreme values, or what we call outliers?

Let's take an example from our compilation. Consider we have a dataset of test scores.

Student Name	Test Score	Student Name	Test Score
Aiden King	97	Isabella Perez	92
Alexander Young	96	James Rodriguez	91
Amelia Lopez	93	Liam Brown	89
Ava Jones	90	Lucas Gonzalez	94
Benjamin Lee	92	Mia Anderson	93
Charlotte Hernandez	93	Noah Smith	87
Elizabeth Walker	95	Olivia Williams	88
Emma Johnson	85	Sophia Martinez	91
Ethan Lewis	95	William Garcia	90
Harper Clark	94	Zoey Davis	75

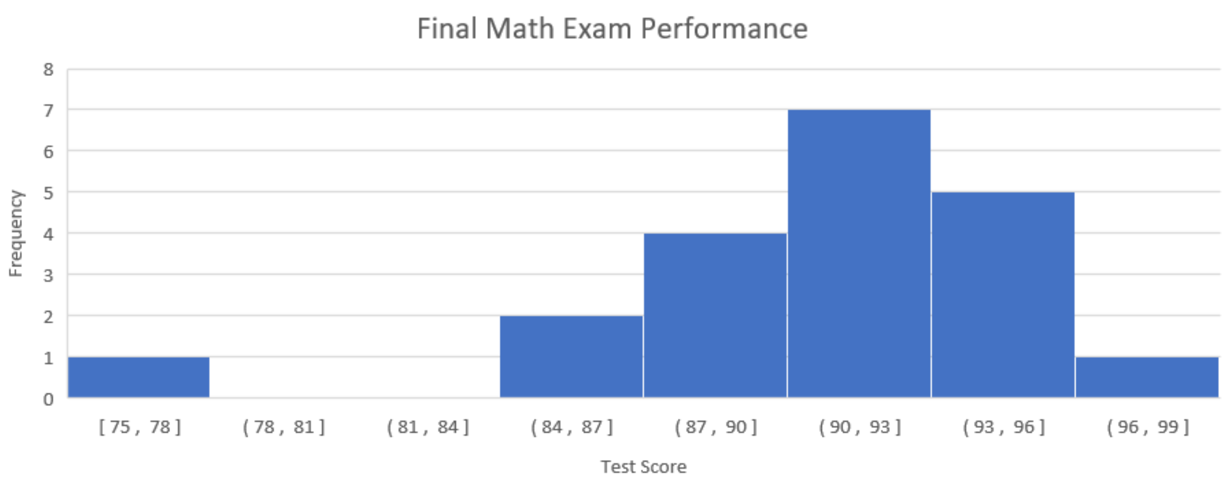
By exploring its distribution, we can see the range of scores, the most common scores, and any outliers. This helps us understand the performance of the class on that test and identify students who might need extra help.



Now, what if I told you that many datasets in nature and human behavior tend to follow a specific pattern of distribution? This is where the concept of normal distribution comes in.

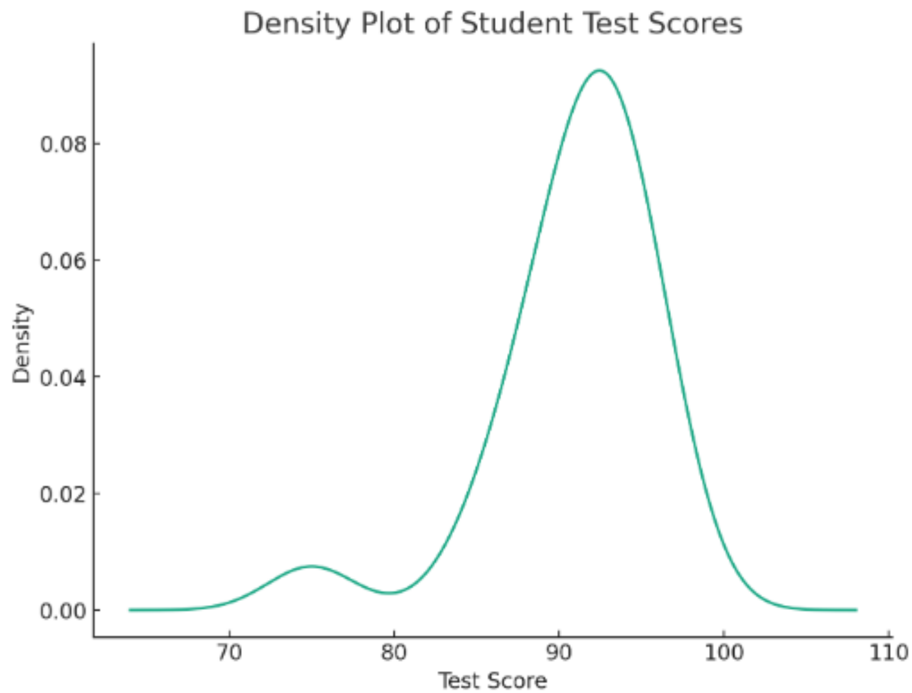
A normal distribution, also known as a bell curve, is a way to represent how data is spread out. It means most of the data points are around a central value (the mean), and as we move away from this center, the number of data points decreases.

Let's visualize this with a histogram from our resource compilation.



Notice the shape of the bars in this histogram. See how they create a bell-shaped curve? This is a classic representation of normal distribution. Most of the data falls around the center, and fewer data points exist as we move towards the extremes.

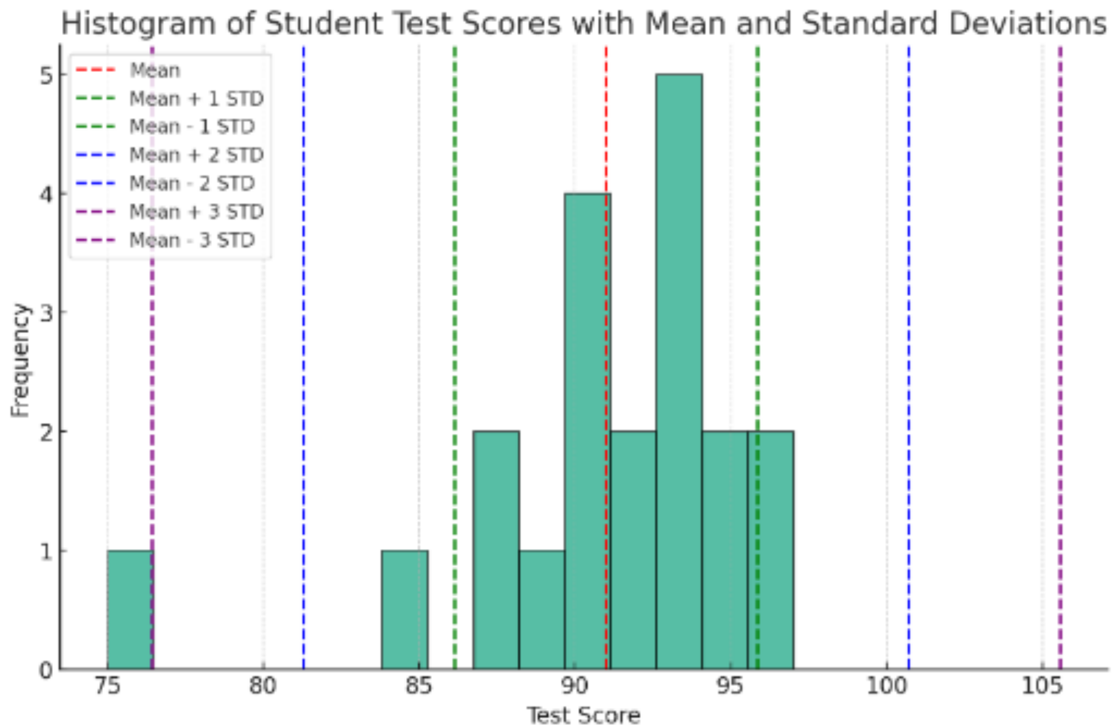
Another way to look at it is through a density plot.



Unlike histograms, which use bars, density plots show a smooth curve representing the spread of data. Here, you can see how the curve peaks in the middle and tapers off towards the ends, illustrating how most values are close to the mean.

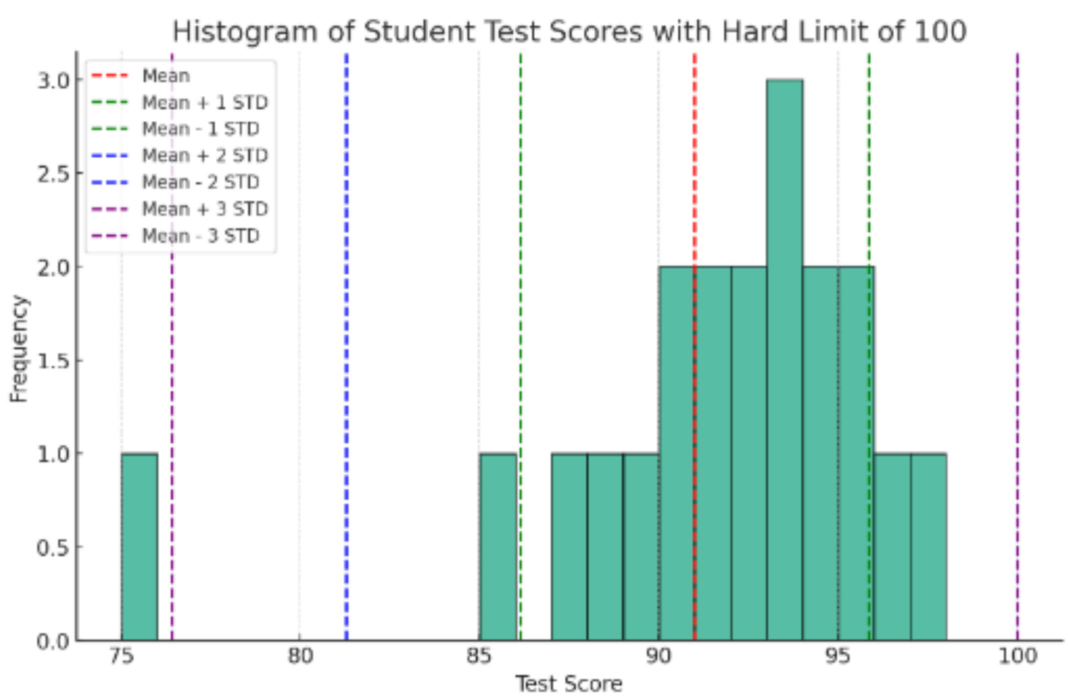
The beauty of normal distribution lies in its predictability. We have something called the empirical rule, which tells us that about 68% of the data falls within one standard deviation of the mean, about 95% within two standard deviations, and about 99.7% within three standard deviations.

To see this in action, let's apply these rules to our test scores data. By calculating the mean and standard deviations, we can predict where most of our scores are likely to fall.



Special cases of test scores:

When a dataset has a hard limit, like test scores that cannot exceed 100, the distribution may become skewed or truncated, affecting the applicability of the empirical rule.



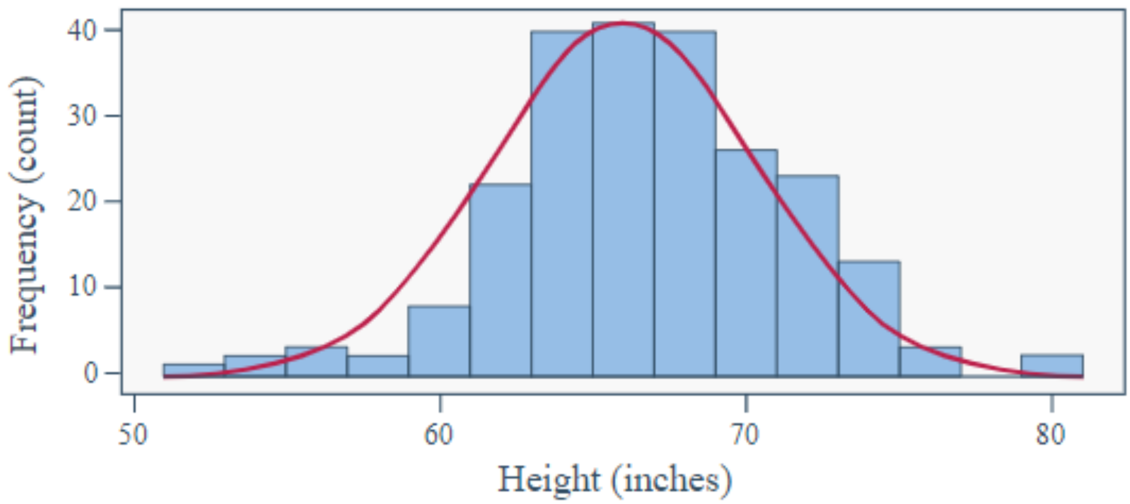
Skewness: If many students score near the maximum (e.g., in the high 90s), the distribution may become left-skewed. In a left-skewed distribution, the mean is pulled towards the higher scores, and more data points lie to the left of the mean. This skewness can distort the symmetry needed for the empirical rule to hold accurately.

Truncation: The hard cap at 100 truncates the distribution, meaning scores can't go beyond this point even if they would in a purely normal distribution. This truncation can lead to a pile-up of scores near the maximum, further distorting the distribution.

Impact on Standard Deviation and Mean: These factors can affect the calculation of the mean and standard deviation, which are central to the empirical rule. The mean might not accurately represent the "center" of the data, and the standard deviation might not reflect the true variability.

In practice, when a dataset is skewed or truncated, the empirical rule may not accurately describe the spread of the data. Alternative methods, such as non-parametric statistics or different models of distribution, might be more appropriate for understanding and interpreting such data.

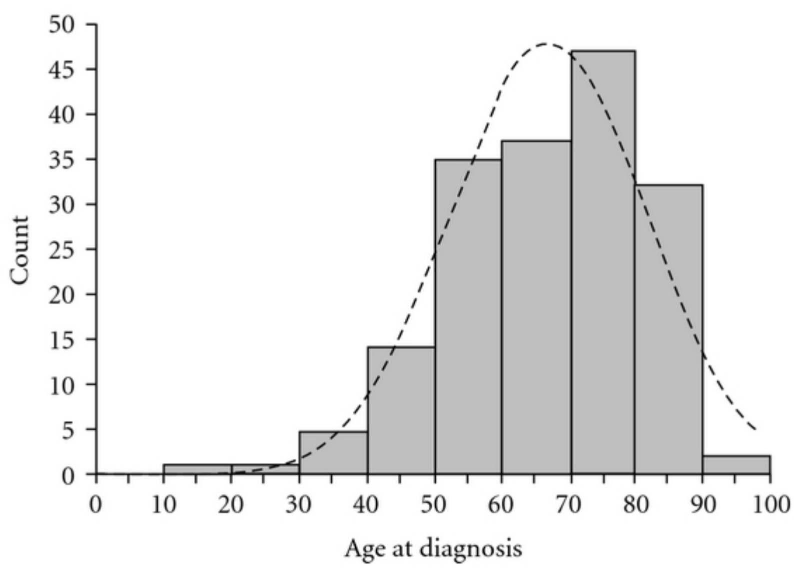
So, why does this matter? Understanding normal distribution helps us make sense of data and predict future trends. It's a powerful tool in fields ranging from psychology to economics. Today, we're going to delve deeper into this concept, understand how to calculate it, and see how it applies to various real-world scenarios."



Histogram of Height (Mean = 66.3 inches & Median = 66 inches)

*Red line is the density plot

Normal distribution no skew



Histogram of Age at Cancer Diagnosis

*Dashed line is density plot

Normal distribution - Skewed right

Appendix F: Inspire Independent Learner Dataset

Real-World Example

Avg Sleep Per Night for 13–18-year olds

Hours of Sleep per Night	**Number of Teenagers**
Less than 5	5
5 to 6	15
6 to 7	40
7 to 8	70
8 to 9	70
9 to 10	40
10 to 11	15
11 or more	5

In this table: The "Hours of Sleep per Night" column represents different ranges of sleep duration.

The "Number of Teenagers" column indicates how many teenagers typically sleep within each range.

Independent learners: Determine how many hours of sleep you get each night and increase the tally in the appropriate “Number or Teenagers” cell to reflect your own data.

Appendix G: Calculations Cheat Sheet

Dataset: 85, 90, 75, 88, 92

Calculating the Mean

First, we'll calculate the mean (average) of these test scores.

1. **Add up all the scores**: 85+90+75+88+92=430
2. **Divide by the number of scores**: There are 5 scores, so Mean=430/5=86

So, the average test score is 86.

Calculating the Sample Standard Deviation

Now, let's calculate the sample standard deviation. We'll use the sample formula since we're dealing with a small group, which could be part of a larger population.

1. **Subtract the mean from each score and square the result**:
   - $(85-86)^2 = (-1)^2 = 1$
   - $(90-86)^2 = 4^2 = 16$
   - $(75-86)^2 = (-11)^2 = 121$
   - $(88-86)^2 = 2^2 = 4$
   - $(92-86)^2 = 6^2 = 36$
2. **Add up these squared differences**:
   - 1+16+121+4+36=178
3. **Divide by the number of scores minus one** (this is Bessel's correction for a sample):
   - There are 5 scores, so Divisor=5−1=4
   - Sum of Squared Differences/Divisor=1784=44.5
4. **Take the square root of this result**:
   - Sqr root of 44.5 is approximately 6.67.

So, the sample standard deviation of the test scores is approximately 6.67.

Summary

- **Mean (Average) Score**: 86
- **Sample Standard Deviation**: Approximately 6.67

Appendix H: Calculation Worksheet

Central Tendency
Learning Outcome: Student will be able to calculate the mean, median, and mode.

Teacher Modeling Demonstration (I Do)

How to calculate the mean, median, and mode.

Example 1: Dataset: [5, 7, 3, 7, 10]

- **Mean Calculation**: Add all the numbers and divide by the number of items.
  5 + 7 + 3 + 7 + 10 = 12 + 10 + = 32
  32 / 5 = 6.4
  Mean = 6.4
- **Median Calculation**: Arrange numbers in order and find the middle value.
  Middle value in 3, 5, 7, 7, 10
  Median=7
- **Mode Calculation**: Identify the most frequent number.
  Mode=7 (appears most frequently)

Guided Practice (We Do)

Example 2: Dataset: [8, 10, 8, 10, 4]

1. Calculate the mean:

8 + 10 + ____ + _____ + ______ = ______. Mean = _____ / ______ = _______

2. Find the median:

4, 8, _____, ______, _______. Median =

3. Determine the mode:

Answers for Guided Discussion:
Mean: 8.4. Median: 8. Mode: 8

Independent Practice (You Do)

Exercise 3: Dataset: [6, 3, 6, 2, 9, 8, 4]

1. Calculate the mean:

2. Find the median:

3. Determine the mode:
