Berkeley Engineers and Mentors

Take a Chance!

Destiny Ogu, Emily Yepez, Jared Puerta, Hannah Chou

Fall 2021

Field(s) of Interest: Statistics, Data Science

Brief Overview (1-3 sentences):
Students will learn the basics about statistics, the science of decision-making and uncertainty, and see first-hand how it can be applied to various aspects of our lives.

Agenda:

- Introduction (5 min)
- Module 1: Law of Large Numbers (15 min)
- Module 2: Probability (15 min)
- Module 3: Sampling (20 min)
- Conclusion (5 min)

Teaching Goals:

- **Central Limit Theorem:** If you take a lot of large samples, the distribution of the sample means (or sum of the samples) will be approximately normal
- **Law of Large Numbers:** Under certain conditions, if you repeat an experiment a large number of times and average the result, the average will converge to the true mean.
- **Probability:** How likely something is to happen
- **Sampling**: the process of selecting a predetermined number of observations from a population. There are different types of sampling: biased and un.
- **Population:** The group/collection of individuals/subjects we want to study
- **Sample:** a subset of the population

Background for Mentors

Module 1: Law of Large Numbers

- Mean
- Law of Large Numbers
- Sample Size
- Central Limit Theorem

Large numbers are used to model copious amounts of scientific and social phenomenons. The numbers are drawn from data collected through samples and populations. A population is the entire group that data is being collected from. A **sample size** is a specific group from the population. Sample size provides a smaller set of data that can be used to represent the whole population. However, samples can still be large and when working with a large data set, whether a sample or population, it can be useful to use one value to represent the data. The **mean** is the average in a given sample or population. The mean incorporates every value in a data set and sets the average as a representative of the entire data, which makes it useful for predicting future results when there are no extremes in the sample or population. The **Law of Large Numbers** states that as a sample size grows, the mean will converge to the average of the whole population, the true value.

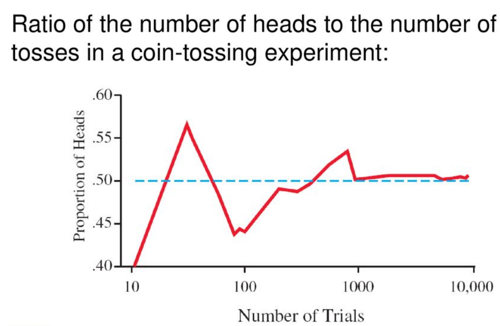

Figure 1: Law of Large Numbers Graph

The **Central Limit Theorem** states that the distribution of the sample means (or the sum of the samples) is approximately normally distributed as the sample size increases. This theorem is based on the law of large numbers as the mean of all sampled variables from the population will be almost equal to the mean of the whole population, as the sample size becomes larger.

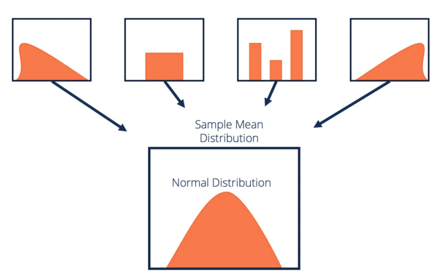

Figure 2: Central Limit Theorem

Module 2: Probability

- Chance
- Event
- Randomness

Probability expresses how likely an **event** is to occur. An event is a "desired" situation. A probability of 0 describes an impossibility because there is no **chance** the event will happen. On the other hand, a probability of 1 describes complete certainty. Not to mention, when the value falls between 0 and 1, there is some possibility for the event to occur.

On the surface of the Earth, if someone lets go of a ball in the air, the force due to gravity will cause the ball to fall downwards. This statement can be made with complete certainty. However, many situations like rolling dice or sharing birthdays with someone, all come down to **randomness** or unpredictableness.

Figure 1: Birthday Problem

In a room of 23 people, there is a 0.5 probability or 50% chance of two people sharing birthdays. The odds increase along with the number of people. Naturally, the probability reaches 100% with 367 people because there are 366 (February 29th included) different birthdays. This is based on the assumption that each birthday is equally probable (statistically, some birthdays are more likely than others and so the 50% threshold is 23 people or fewer).

Rolling one die only has 6 outcomes. When rolling two dice, there are 36 outcomes in total. The growth in outcomes is exponential; rolling three dice gives 260 outcomes. Focusing on the two dice for the activity, some outcomes are more likely than others. For instance, rolling doubles is less likely than rolling two even numbers. This is because there are only 6 ways to roll doubles (probability is ⅙). On the other hand, there are 9 ways to roll two even numbers (probability is ¼) .

Module 3: Sampling

- Simple Random Sample
- Biased Samples

In statistics, a **Simple random sample** is a group of individuals that are chosen at random from a larger group of individuals. With a simple random sample, it is important that each individual has the same probability to be selected. A simple random sample is meant to represent the entire data set, and the samples can be taken using methods like lotteries or random draws.

An example of a simple random sample would be using a random number generator to select a group of 10 individuals in a group of 25 boys and 25 girls. Because each individual has an equal chance of being selected, the sample should include close to 5 boys and 5 girls.

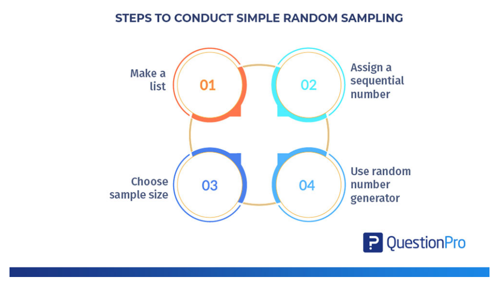

Figure 1: Simple Random Sampling Process

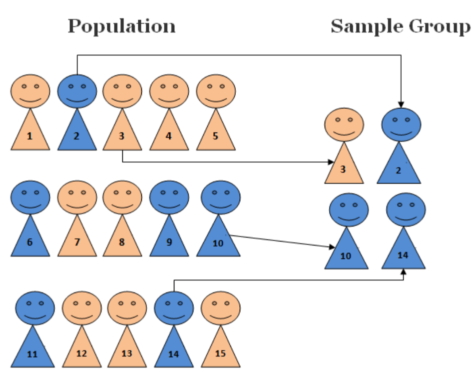

Figure 2: Sample vs. Population

Simple random samples differ heavily from **Biased samples**, which are samples that are collected in such a way that some members of the intended population have a higher or lower probability of being selected compared to others.

An example of a biased sample would be specifically choosing individuals on the left side of the room in a sample of 25 boys and girls. Because the sample is being collected with a bias, the result will represent said bias.

Introduction

Statistics can help us make informed decisions about various situations and even help us make predictions about upcoming events. Especially in the current age, having quantitative tools to analyze data and make informed decisions can help us understand the randomness in the world around us.

<table>
<tr>
<td>

Concepts to Introduce

- **Statistics** is a field of mathematics that focuses on the collection, description, and analysis of quantitative data.
- **Randomness** describes a phenomenon in which the outcome of a single repetition is uncertain
- **Probability** describes the likelihood of an event occurring.

</td>
<td>

Questions to Pique Interest

- When flipping a fair coin, what is the probability of getting heads?
  - 50%
- Say that you want to find out how many people in your school prefer cake over pie. How would you go about answering this question?
  - Take a simple random sample of the people at your school, ask them what they prefer, count up the number of people who prefer cake over pie.

</td>
</tr>
<tr>
<td>

Scientists, Current and Past Events

- Weather forecasting: Weather forecasting companies rely on gathering large amounts of data in order to forecast the weather for the day:
  - https://thisisstatistics.org/beyond-barometers-how-statisticians-help-to-predict-the-weather/
- Disease tracking: Statistics become very useful in tracking the spread of diseases and predicting the risk of transmission.
  - https://covid.cdc.gov/covid-data-tracker/#datatracker-home
- Election polls: Many polling organizations will conduct their own statistical analysis to predict outcomes of upcoming elections.

</td>
<td>

Careers and Applications

- Statistician
- Biostatistics/epidemiology
- Machine Learning/Artificial Intelligence
- Data Analyst

</td>
</tr>
</table>

Module 1: Law of Large Numbers!

In this module, students will learn how the law of large numbers works and why averages are important in statistical analysis.

Teaching Goals	Materials
1. **Mean**: the average in a given sample/population 2. **Law of Large Numbers**: Theorem that states that if you repeat an experiment independently a large number of times, the mean will converge to the true value **3. Sample Size**: The number of subjects/individuals in a sample **4. Central Limit Theorem:** The theorem that states that the distribution of the sample means (or sum of the samples) is approximately normally distributed as the sample size increases	• Coin • Paper • Pencil

Procedure

1. Students will first flip the coin 10 times, recording the number of times the coin lands on heads.
2. Students will then flip the coin 20 times, again recording the number of times the coin lands on heads
3. Repeat for 30 times, ask if students recognize a pattern

Figure 1: *Coin Flip*

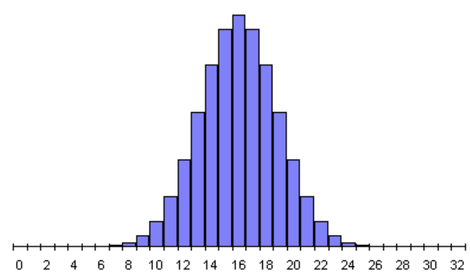

Figure 2: *Normal Distribution of Coin Flip (Central Limit Theorem)*

Module 2: Probability

Through the probability of shared birthdays and rolling dice, students will better understand how the world around them is unpredictable and learn about the concept of chance.

<table>
<tr>
<td>

Teaching Goals

1. **Chance:** The likelihood of an event occurring
2. **Event:** The situation we want to observe
3. **Randomness:** Unpredictableness; especially in this module, it is important to emphasize that we can never be sure that something will actually happen unless the probability is 1

</td>
<td>

Materials

- 2 dice
- Printout of Dice Combinations chart (11 sheets)

</td>
</tr>
</table>

Procedure

1. Have each student share their birthday to see if there is a pair.
2. Explain the following: In a group of 23 people, there is a 50% chance that two people share the same birthday. (***Figure 1***) More people increases the chances of the event.
3. If I have 6 yellow balls and 4 red balls, what are the chances of picking a red ball? Explain probability of the situation (# of desired of events/ # total events possible).
4. Next, introduce the dice activity by asking if they can predict the exact numbers the dice will land on. What is the probability of rolling those numbers? Roll the dice to test their hypotheses.
5. Ask the class the following questions:
    a. Which is more likely: rolling two ones or rolling two odd numbers?
        i. Answer: rolling two odd numbers is more likely
    b. Which is more likely: rolling doubles or rolling two even numbers?
        i. Answer: Rolling two even numbers is more likely.
    c. Which is more likely: rolling two numbers that sum to 12 or rolling two numbers that sum to 7?

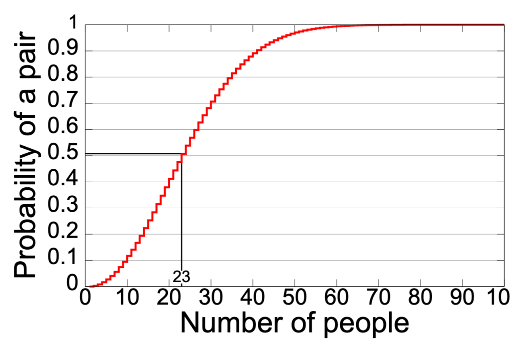

Figure 1: *Birthday Problem*

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Figure 2: *Combinations of Rolling 2 Dice*

i. Answer: rolling two numbers that sum to 7

Classroom Notes

Refer to the two dice combinations chart so that students can better visualize probability of each event

Module 3: Sampling

By predicting the number of red lollipops within a bag, students are able to visualize the difference between using different sampling techniques.

Teaching Goals	Materials
1. **Simple Random Sample:** Individuals are chosen completely at random from a population. This type of sampling is the most unbiased approach. 2. **Biased Samples:** A sample in which one or more sub-groups are over/under represented.	• Bag of candy (135 pieces)

Procedure

1. Pair students up. Each pair will either simulate a simple random sample or a biased sample.
2. The group simulating random sampling will reach into the bag and randomly select 30 pieces of candy.
3. The group simulating the biased sample will be provided with a (biased) sample. (**Figure 1)**
4. Have students count the number of the specific type of candy in their sample and report back to the mentors.
5. Come together for discussion!
6. Ideally, the simple random sample would be the most accurate (i.e. the number of each type of candy will be closest to the actual number of each type in the bag). **(Figure 2)**

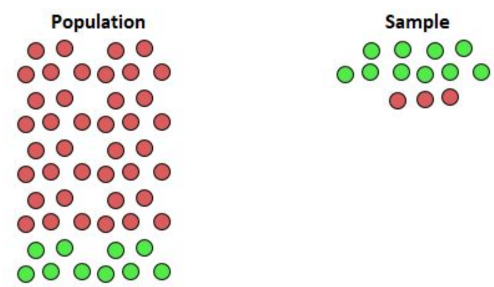

Figure 1: *An example of a biased sample.*

Figure 2: *An example of a simple random sample.*

Classroom Notes

Remember to speak loud

Conclusion

In this lesson, students learned about the Law of Large Numbers, probability, and sampling. Statistics as a whole is an extremely broad field that can be applied to multiple disciplines. Whether it be for disease tracking, sports analysis, or machine learning, having a solid understanding of statistics can help students interpret the meaning of large datasets and recognize the randomness that exists in our lives.

References

- Probabilities for Rolling Two Dice, Thought Co. https://www.thoughtco.com/probabilities-of-rolling-two-dice-3126559
- Probability and the Birthday Paradox, Scientific American https://www.scientificamerican.com/article/bring-science-home-probability-birthday-paradox/
- Law of Large Numbers https://www.probabilitycourse.com/chapter7/7_1_1_law_of_large_numbers.php
- Statistics, Investopedia https://www.investopedia.com/terms/s/statistics.asp
- COVID data tracker, CDC https://covid.cdc.gov/covid-data-tracker/#datatracker-home

Summary Materials Table

Material	**Amount per Site**	**Expected $$**	**Vendor (or online link)**
Dice	2	$0 (at Bechtel)	Target Amazon
Printout of Chart	1	$0	
Coins	25	$0.25	Withdraw from bank
Candy	1 bag		Walgreens (acquired)
