

Human vs. Machine: Unveiling Randomness with Data Visualization and Stats

Venessa N. Singhroy and Rommel Robertson, Queensborough Community College

Overview of Lesson

This lesson plan demonstrates the foundational importance of ‘randomness’ in the random selection process by contrasting human decision-making with computer-generated choices, illustrating that humans struggle to make truly random selections even when presented with information on the random selection process (Singhroy, Robertson & Stroumbakis, 2025). The lesson explores key statistical principles, including the law of large numbers, sampling variation, and the Chi-Square Goodness of Fit test, through a name selection activity where students collect data, create graphical visualizations, and perform statistical analyses. Recognizing that while 82% of U.S. school districts have sufficient internet access (CoSN, 2021), this lesson utilizes Google Forms and Google Sheets, aiming to broaden inclusivity by leveraging accessible and cost-free technology to ensure that students from a wide range of classroom settings can engage fully with the material. The Chi-Square Goodness of Fit test, introduced in the second stage of the lesson focusing on data analysis, can be used as an enrichment activity, leaving it to the discretion of the instructor to determine whether it aligns with their curriculum and students’ readiness. Educators who choose to omit the Chi-Square test can focus on the data visualization component introduced in the first stage of the lesson, ensuring accessibility for a broader audience while maintaining the activity’s educational value. Additionally, the lesson need not be taught over four successive days but can instead be integrated throughout a statistics curriculum, allowing instructors to align it with their pacing and instructional goals. To enhance comprehension, the activity also emphasizes critical reflection on the inherent biases in human decision-making. This is bolstered by the findings of Singhroy et al. (2025), who highlighted that even when students understood the importance of random selection and were directed to make selections randomly, they frequently exhibited a selection bias, often favoring names familiar to them (Singhroy, Robertson & Stroumbakis, 2025). Extensions of this activity encourage students to examine how these biases affect their decision-making processes.

Type of Data

- One categorical variable
- Data generated or collected as a class

Learning Objectives

- Explore the concept of “randomness”.
- Understand the concept of random selection and its comparison between human and computer-generated selections (i.e., without replacement).
- Demonstrate the law of large numbers through graphical and statistical analysis.
- Perform and interpret the Chi-Square Goodness of Fit test.
- Identify biases in human random selection.

Audience

- Grade 10-12 students at Level C from GAISE II (GAISE II Reference https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf)
- *Prerequisites:*

First Stage of the Lesson - Data Visualization: A Graphical Approach

- Understanding variables (categorical and quantitative).
- Familiarity with measures of central tendency (mean, median, mode).
- Knowledge of measures of dispersion (range, variance, standard deviation).
- Ability to create and interpret bar charts, histograms, pie charts, scatter plots, and box plots.
- Basic understanding of probability and randomness.

Second Stage of the Lesson - Data Analysis: A Statistical Approach (Days 3 & 4)

- All prerequisites from the first stage.
- Familiarity with sampling methods and the potential for sampling bias.
- Prior experience with statistical concepts such as the Chi-Square Goodness of Fit test.

Time Required

Implementing this lesson plan requires approximately four 55-minute class periods. On Day 1, students will collect data using an instructor created Google Forms name selection surveys activity, generate bar charts from their selections, and discuss human selections. Day 2 will focus on comparing human vs. computer-generated selections (i.e., random selections without replacement), followed by initial reflections and discussions on the differences observed. On Day 3, students will engage in statistical analysis by performing a Chi-Square Goodness of Fit test using Google Sheets. Finally, Day 4 will be dedicated to interpreting the test results, reflecting on the implications of their findings, and discussing real-world applications of these statistical

concepts. It is at the instructor's discretion to conduct the lessons on four successive days or space them out to cover prerequisite skills students need to implement the activities on each day.

Technology and Other Materials

- **Technology:** Google Forms, Google Sheets, Internet access, Projector or screen
- **Materials:** Instructor handout to construct the Google Form surveys that will be administered to students, student handouts with instructions and directions using Google Sheets to implement the Chi-Square Goodness of Fit test, create bar graphs, and randomly select names. While access to the instructor created Google form survey does not require students having access to Google account, working in the Google Sheets application and accessing all its features will require a Google account. Students who already have a personal Google account will already have access to the Google Sheets app while students who do not have a Google account can be encouraged to sign up for a free account prior to the lesson or they can be paired to work with a student who already owns a Google account.

Lesson Plan

In statistics education, it is crucial to emphasize the foundational significance of randomness to random selection, which ensures unbiased representation of populations in samples and underpins statistical methodologies. However, students often struggle to grasp the complexities of this concept, leading to challenges in effectively applying random selection methods. This lesson aims to clarify the concept of randomness in the selection process through a practical and engaging instructional activity. By providing hands-on experience in comparing human-generated and computer-generated selections, the lesson seeks to illuminate the inherent biases in human decision-making and enhance students' understanding of objective data analysis.

The 4-day structure ensures students gain a comprehensive understanding of data collection, visualization, and statistical analysis within the context of random selection. By organizing the lesson activities in this manner, the lesson plan adheres to the GAISE four-step statistical problem-solving process, providing a structured and thorough approach to teaching statistical concepts. However, as noted in the lesson overview, instructors who decide to omit the Chi-Square test can shift their focus to the data visualization component introduced in the first stage, ensuring the activity remains accessible to a wider audience while retaining its educational value. The alignment with GAISE's four-step process is as follows:

- **Formulate a Question:** Day 1 focuses on understanding random selection and its challenges.
- **Collect Data:** Day 1 and Day 2 involve gathering data on human and computer-generated selections.

- **Analyze Data:** Day 2 and Day 3 are dedicated to visualizing and analyzing the collected data.
- **Interpret Results:** Day 3 and Day 4 involve interpreting the results, discussing implications, and understanding biases.

Background for GAISE Level C and Statistical Problem-Solving Process

Students with a Level C understanding, according to GAISE II, should possess the foundational knowledge necessary to meet the prerequisites outlined above. Level C corresponds to high school grades (9-12), where students typically have been exposed to basic statistical concepts and data visualization techniques as part of their curriculum. While they may not have mastered these concepts at an advanced level, they should have sufficient understanding to engage with the lesson plan effectively. However, it's essential to consider individual variations in students' prior learning experiences and provide additional support or review as needed to ensure all students can participate fully in the lesson.

The initial instructional task, centered on visualizing data through Google Forms to create bar charts, acts as scaffolding for subsequent activities. It acquaints students with the process of data collection and the visual depiction of data distributions. Moving to the Chi-Square Goodness of Fit test in Google Sheets builds on this support structure, further exploring statistical analysis and contrasting observed data with anticipated theoretical distributions, with a particular focus on the uniform distribution within this framework. While uniformity can serve as an indication of randomness in contexts such as rolling dice or shuffling cards, it is neither definitive proof of randomness nor a strict requirement. To enhance students' understanding, the lesson plan includes discussion questions on additional criteria necessary to establish randomness, such as independence and unpredictability, while using the uniform distribution as a practical tool for illustrating randomness in the subsequent activities. The use of Google Forms and Google Sheets offers accessibility advantages, making these tools suitable for all classrooms, regardless of resource levels, thus ensuring that both low-resource and high-resource educational settings can implement this lesson effectively.

First Stage of the Lesson

Data Visualization: A Graphical Approach (Days 1 & 2)

Objective: Demonstrate the concept of random selection and compare human selections to computer-generated selections using data visualization techniques.

Materials Needed:

1. Google Sheets app accessed via Free Google account
2. Computer with internet access for students
3. Projector or screen for displaying the bar chart

Day 1:

On Day 1 of the lesson plan, the goal is to introduce students to the concept of randomness in random selection and the potential biases in human decision-making. The teacher will begin with a discussion on what it means to make a truly random selection, emphasizing its importance in various fields such as research and lotteries. To engage students and gauge their initial understanding, the teacher will pose questions such as:

- "What do you think it means for something to be 'random'?"
- "Why do you think researchers use random selection?"

Following this, the teacher will show a YouTube video [Is Anything Truly Random](#), which suggest that neither humans nor computers can generate truly random selections. After the video, the class will be asked to formulate questions. It is anticipated they will arrive at:

- "Which is better at making random choices: humans or computers?"
- "Are there certain contexts where humans may be better at making random choices than computers and vice versa?"

In response, the teacher will query:

- "How can we tell if the outcomes are random?"

The teacher should attempt to elicit from the students that a uniform distribution of outcomes could be indicative of random selection. For example, they may recall that when tossing a coin or rolling a die, each outcome appears equally likely in the long run. However, the teacher should also guide students to understand that uniformity alone is not sufficient to establish randomness. Additional criteria such as independence and unpredictability must also be considered. To foster deeper understanding, the following discussion questions can be introduced:

- "If we observe a uniform distribution, what additional evidence might we need to confirm randomness?"
- "Can a system that is not random produce a uniform distribution? Why or why not?"
- "How can we test for independence in a sequence of outcomes?"

Students will then participate in name selection activities using a pre-prepared Google Form (see Instructor Handout). These will consist of two separate surveys—one for male-oriented names and one for female-oriented names—each comprising 20 options. The surveys will ask students to select 5 female-oriented and 5 male-oriented names from a list of 20 female and 20 male-oriented names, respectively. The teacher will underscore the instructions directing students to attempt making random choices. The teacher will guide students through submitting their responses. To help students reflect on their choices and the potential biases involved, the teacher will ask students to answer the following questions for homework:

- "What factors influenced your name selections?"

- "What do you notice about the distribution of name selections?"
 - Elicit:
 - Variability in the frequency of name selections.
 - Possible patterns, such as clustering around certain names or avoidance of others.
 - Factors that may influence these patterns, such as familiarity, cultural associations, or perceived desirability of names.

Students should revisit their earlier responses to the questions: "Which is better at making random choices: humans or computers?" and "Are there certain contexts where humans might excel at making random choices compared to computers, and vice versa?" This reflection allows them to consider whether their perspectives have evolved based on the activity and subsequent discussion. The day will conclude with a preview of the next lesson, which will involve generating computer selections to the surveys given to students.

Day 2

On Day 2 of the lesson plan, students will delve deeper into understanding random selection by comparing human choices with computer-generated selections. The teacher will begin by briefly reviewing the data collected from Day 1, then either demonstrate how to visualize the data using bar charts in Google Sheets or have students use the provided handout to create their own visualizations. Focusing on the bar charts that visually represent students' name selections, students will be encouraged to interpret the charts and discuss any patterns or biases they observe in their selections. Questions to facilitate this discussion may include:

- "What do you notice about the distribution of name selections?"
- "Are there any patterns or trends in how names were chosen?"
- "Do you think your selections were truly random? Why or why not?"

Next, the teacher will introduce the concept of computer-generated random selections. Students will learn how to simulate random selections using Google Sheets. The teacher can provide students with an instructional handout on randomly selecting names in Google Sheets and offer guidance on setting up and executing these simulations, helping students generate datasets similar to their selections from Day 1. Specifically, students should be encouraged to use Google Sheets to generate datasets with the same number of observations as those collected on Day 1, ensuring that subsequent comparisons are made using equivalent sample sizes.

Students will then compare their own selections with the computer-generated datasets by juxtaposing. They will analyze and discuss differences in distributions. Discussion questions may include:

- "How do the distributions of human selections compare with those generated by a computer?"

- "What can we infer about human decision-making compared to computer-generated randomness?"

At this stage, students should use Google Sheets to generate progressively larger sample sizes (e.g., 200, 500, 1000) and create corresponding bar graphs. They should observe that the distribution becomes increasingly uniform as the sample size grows. If the concept has been previously introduced, the instructor can connect this observation to the law of large numbers, which explains how random selections tend toward uniformity with larger sample sizes.

The day will conclude with a whole-class discussion summarizing key insights, comparing the advantages of the human vs. computer selections and preparing for the statistical analysis they will conduct in the following lesson.

Second Stage of the Lesson

Data Analysis: A Statistical Approach (Days 3 & 4)

Objective: Understand random selection through statistical analysis and the Chi-Square Goodness of Fit test.

Caveat: We acknowledge that the statistical power of the Chi-Square Goodness of Fit test may be limited when working with small sample sizes, as is often the case in smaller classes. To mitigate this, we recommend collaborating with other instructors teaching the same topic or involving students from additional classes. This collaboration can help increase the sample size, leading to more robust and reliable results. While definitive conclusions cannot be drawn without an adequate sample size, the application of the Chi-Square test still holds value as an accessible and practical tool, fostering students' exploration of statistical principles, hypothesis generation, and critical thinking about the adaptability of these methods in real-world contexts.

Materials Needed:

1. Google Sheets app accessed via Free Google account
2. Computer with internet access for students
3. Projector or screen for displaying data and analysis

Day 3:

On Day 3 of the lesson plan, students will transition from data visualization to statistical analysis, focusing on the Chi-Square Goodness of Fit test to further explore the concept of random selection. The teacher will begin by reviewing the datasets from Days 1 and 2, emphasizing the distributions of both human selections and computer-generated selections. Students will be prompted to reflect on their previous discussions and observations about randomness and biases in decision-making.

The teacher reminds students that the Chi-Square Goodness of Fit test can be used as a statistical tool used to compare observed frequencies with expected frequencies under a uniform

distribution assumption. Questions to help students recall the theoretical background behind the test and its relevance to assessing the fit of observed data to expected patterns may include:

- "Why do we use the Chi-Square Goodness of Fit test in this context?"
- "What do 'observed frequencies' and 'expected frequencies' represent in our dataset?"
- "How does the test help us evaluate whether our data follows a uniform distribution?"

Next, students will apply their knowledge by conducting the Chi-Square Goodness of Fit test on the student observations and the computer-generated observations of the same sample size that they collected on Day 2, using Google Sheets. They will be given the attached handouts which will guide them through setting up the test, entering data, calculating the test statistic, and interpreting the results. Students will work in pairs or small groups to complete the analysis, with the teacher providing support as needed.

After completing the statistical analysis, students will engage in a discussion to interpret their findings. They will analyze the results of the Chi-Square test, discuss any deviations from expected distributions, and reflect on the implications for random selection and decision-making. The day will conclude with a reflection on the entire lesson sequence, emphasizing connections between data visualization, statistical analysis, and real-world applications of random selection concepts.

Discussion and Summary

1. Summarize what students learned about random selection, the law of large numbers, and statistical analysis.
2. Discuss real-world applications and the importance of recognizing biases in data interpretation.

Discussion Questions:

1. How does the variability in human selections compare to computer-generated selections?
 - a. Example of incomplete understanding: "Human selections are completely random."
 - b. Example of correct response: "Human selections show more variability and bias compared to computer-generated selections, which tend towards a uniform distribution."
2. How can biases in decision-making affect data interpretation and analysis?
 - a. Example of Incomplete Understanding: "Biases don't really affect data interpretation because the data speaks for itself."
 - b. Example of Correct Response: "Biases in decision-making can significantly affect data interpretation and analysis by skewing the results and leading to incorrect conclusions. For instance, if a researcher has a confirmation bias, they might give more weight to data that supports their hypothesis while ignoring data that

contradicts it. This can result in an inaccurate representation of the true nature of the data and flawed analyses."

Day 4:

On Day 4, students will reinforce their understanding by collaboratively analyzing the larger datasets that were generated by the computer on Day 2 and discuss their findings. This culminating activity synthesizes their understanding of random selection, the law of large numbers, statistical analysis, and the implications of biases in decision-making.

The teacher will begin by reviewing the results of the Chi-Square Goodness of Fit tests conducted on Day 3, emphasizing key findings and interpretations from the statistical analysis. Next, students will be instructed to work in pairs or small groups to perform the Chi-Square Goodness of Fit test on the various sample sizes generated by the computer on Day 2, analyzing whether the larger samples indeed tend toward a uniform distribution. The lesson will conclude with a summary discussion guided by the following questions:

1. What does the law of large numbers predict about random selections as the sample size increases?
 - a. Example of Incomplete Understanding: "The law of large numbers means that the results will always be perfectly random regardless of the sample size."
 - b. Example of Correct Response: "The law of large numbers predicts that as the sample size increases, the average of the results obtained from random selections will get closer to the expected value. This means that the outcomes will more closely resemble a uniform distribution over time, even if individual selections can vary greatly in smaller samples."

Students will be encouraged to draw connections between their findings from the activities throughout the lesson and real-world applications. The teacher will facilitate the discussion, guiding students to articulate their understanding and insights gained from the lesson activities.

By the end of Day 4, students will have comprehensively explored random selection, statistical principles, and biases in data interpretation, equipping them with critical thinking skills applicable across various disciplines.

Attached Materials

- For instructor:
 - Instructions for Creating Google Form to collect student responses
- For students:
 - Instructions for Randomly Selecting Names and Counting Frequencies in Google Sheets
 - Instructions for Creating Bar Charts in Google Sheets
 - How to Perform a Chi-Square Goodness of Fit Test in Google Sheets

Reflections and Additional Recommendations

- Notes about possible extensions: Explore other forms of random selection and their applications in different fields. Encourage students to engage in a qualitative analysis of their responses and reasoning behind their selections. Examine the necessary adherence to the assumptions of the Chi-Square Goodness of Fit test, allowing it to be used as an inferential tool rather than solely as an exploratory one.
- Ideas for differentiation: Provide additional support for students struggling with statistical concepts through one-on-one guidance or additional resources.
- Teacher reflections: Discuss student reactions and consider potential changes for future iterations of the lesson.

References

Singhroy, V., Robertson, R., & Stroumbakis, K. (2025). Does a Name Make a Difference? Teaching Random Selection in the Classroom. *Numeracy*, 18(1), 2.

Consortium for School Networking. (2021). 2021 Annual Infrastructure Survey Report. CoSN.