UNIVERSITY of ARKANSAS AT PINE BLUFF 1873

DIVISION OF GRADUATE STUDIES
AND CONTINUING EDUCATION

Notification to Registrar of Project/Thesis/Dissertation Completion

This is to notify the Office of the Registrar that Ashish Chaudary
(print student's name)

did on 06/24/2025 (mm/dd/yyyy) successfully complete the Project (Project/Thesis/Dissertation) Requirement in the

Graduate Program in

Aquaculture and Fisheries	X Computer Science and Technology
Agricultural Regulations	Master of Business Administration
Addiction Studies	

DocuSigned by: Anna Harris 2DAC61080C0C4C3... — 6/24/2025
Committee chair — Date

Signed by: 9B882CC00CD04D4... — 6/24/2025
Graduate Coordinator — Date

Signed by: 9B882CC00CD04D4... — 6/25/2025
Department Chair — Date

DocuSigned by: E07BD25D7F00442... — 6/25/2025
Dean of School — Date

DocuSigned by: David Fernandez AE07057AEC1E459... — 6/25/2025
Dean of Graduate Studies and Continuing Education — Date

DocuSigned by: Sonya Lockett 8A6035784A7C473... — 6/25/2025
Thesis copies (printed & PDF) received in Watson Memorial Library — Date

Duplication Release

I hereby authorize the University of Arkansas at Pine Bluff Library to duplicate this project report when needed for research and/or scholarship.

Agree: ………………………………………………………
(Signature of Student)

Refused: ……………………………………………..
(Signature of Student)

ENHANCING STUDENTS' LEARNING EXPERIENCE IN INTRODUCTORY PROBABILITY AND STATISTICS CLASS THROUGH THE INTEGRATION OF TECHNOLOGY, PYTHON PROGRAMMING, AND GOOGLE COLAB

A Project

Presented to the faculty of the Department of Mathematics and Computer Science
University of Arkansas at Pine Bluff

Submitted in partial satisfaction of
the requirements for the degree of

Master of Science

in

Computer Science and Technology

by

Ashish Chaudhary

May 2025

University of Arkansas at Pine Bluff

THIS PROJECT IS APPROVED FOR
RECOMMENDATION TO THE GRADUATE COUNCIL

Project Advisor:

Anna Harris	6/20/2025
Dr. Anna Harris	Date

Project Committee:

Dr. Vinay Raj	Date
	6/23/2025
Dr. Qinglong Jiang	Date
	06/20/2025
Dr. Eddie Branch Graduate Coordinator, Department of Math and Computer Science	Date
	06/20/2025
Dr. Eddie Branch Chair, Department of Mathematics and Computer Science	Date

ABSTRACT

The traditional teaching methods are based on lectures, static visuals on textbooks and boards, and manual or simpler tools for calculations, which may hinder understanding the abstract concepts of mathematics, statistics, and science, limiting the students' engagement and the ability to apply statistical concepts in real life. This project aims to modernize the teaching approach by utilizing the technologies available to students by integrating Python programming and Google Colab as interactive learning tools. Python offers strong and plenty of statistical and visualization libraries that allow students to explore probability and statistical concepts, and analyze real-world datasets, thus improving their computational thinking and data literacy.

Google Colab, a cloud-based web application, eliminates hardware capability and software installation barriers. This feature removes the drawbacks of students' technical backgrounds when installing the required complex software and hardware specifications. This project has benefits like interactive coding in exercises, getting real-time feedback, and visualizations, which are very helpful for students in understanding statistical methods better. Students can manipulate data, alter parameters, and observe the effects in real time, which is only possible in an active learning environment. Google Colab and Python can help create one. This links the gap between theory and practical applications of statistical concepts. This project contributes to the modernization of statistics education by presenting an expandable and adaptable model for incorporating technology-driven learning in university-level coursework. The idea behind this project is not just teaching statistics but also preparing students to compete in the data-driven world by providing the skills they need.

ACKNOWLEDGMENTS

My heartfelt appreciation goes to Dr. Anna Harris for her continuous guidance, unwavering support, and intuitive feedback throughout the building of this project. Her expertise has been extremely beneficial in completing my project. Along with my advisor, I am grateful to my committee members, Dr. Vinay Raj and Dr. Qinglong Jiang, for their invaluable time and positive criticism. These encouragements have significantly improved this project. I also appreciate all the faculty and staff in the Department of Mathematics and Computer Science for providing the necessary learning resources and environment.

Special thanks to my fellow graduate students for their participation and feedback, which helped refine the integration of technology in teaching statistics and probability. I am grateful to my family and friends for their steady encouragement and patience, which kept me motivated throughout this journey.

Finally, I acknowledge the contributions of the open-source community and developers behind Python and Google Colab, whose innovations have enabled interactive and accessible learning experiences.

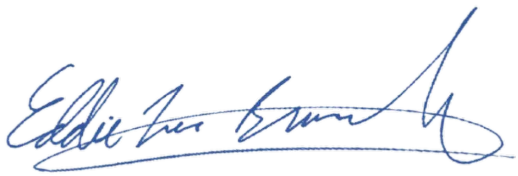
Thank you all for your support and inspiration.

Sincerely,
Ashish Chaudhary

CERTIFICATION PAGE

Student: Ashish Chaudhary

I certify that this student has met the requirements for the format contained in the University format manual and that this project is suitable for shelving in the library and credit is to be awarded for the project.



06/20/2025

______________________________ ____________________

Department Chair: Dr. Eddie Branch Date

Department of Mathematics and Computer Science

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

In recent years, educational systems and institutions have integrated technology, prioritizing it, particularly in teaching statistics and probability. The world is becoming more data-driven, so students must upgrade their firm statistical mastery and computational skills to keep up in the competitive environment. Using technologies like Python programming and Google Colab boosts student learning capabilities in introductory probability and statistics courses. Recent technological advances have already transformed our lives, along with education. Technology can make conceptual concepts more realistic and easily understandable to students in fields with more abstract concepts, like statistics and probability. With its extensive data analysis and visualization libraries, the Python programming language has emerged as a dominant tool for teaching statistics (VanderPlas, 2016).

Likewise, cloud service-based platforms like Google Colab provide handy and collaborative environments for students to participate in statistical concepts and data analysis (Carneiro et al., 2018). Integrating technology in education aligns with the goals outlined in the United Nations Sustainable Development Goals (SDGs), particularly SDG 4, which aims to ensure inclusive and equitable quality education for all (UNESCO, 2023). As noted by UNESCO (2023), technology appears in six out of the ten targets in SDG 4, recognizing its potential to affect education through various channels, including as an input, means of delivery, skill, tool for planning, and providing a social and cultural context. However, the effective integration of technology in education is not without challenges. As highlighted by (Delgado et al., 2015), the mere presence of technology in classrooms does not guarantee improved learning outcomes. The success of technology integration depends on various factors, including the type of technology used, the context in which it is implemented, and the pedagogical approaches employed.

1.2 Problem Statement

Despite the numerous benefits of integrating technology into education for better and more effective teaching and learning, there are some challenges that hinder the wide adoption in educational institutions. Some of the challenges are discussed below.

1. Inadequate Implementation Strategies: Innovating technologies don't bring success on their own. The bottleneck is implementing them in effective and sustainable ways. There is limited research in building strategies for implementing technology to teach courses like statistics (Hardin et al., 2015).
2. Gap in technological skills between students and instructors: The teaching methods so far are based on similar skills of writing, reading, and navigating through books among students and instructors. But, integrating technology is a new method, and students and instructors should be able to use the technology with comparable proficiency to teach and learn at the same level. But in most scenarios, this case is not ideal, which can cause problems in the classroom (Horton et al., 2015).
3. Using Technology in balance with statistical concepts: Overuse of technology over the cost of core statistical concepts is not good. So, there should always be a balance between using technology and understanding concepts. Technology may not be used in understanding all the concepts, which remains a challenge (Cobb, 2015).

4. Limited access to technology causing digital divide: Still in many parts of the world access to technology is significantly limited. If access to technology is not provided to all, using technology in education can create a substantial division (UNESCO, 2023).

5. Inadequate research on long-term impacts: There is much research on the positive result of integrating technology in education, but this research is only based on one or two decades of data. However, there is still no proper data on the impact of integrating Python and Colab in teaching statistics in the long run in learning outcomes and career readiness (Baumer, 2015).

6. Resistance to change: Teachers may be reluctant to use technology in teaching statistics, rather they may be comfortable using their traditional methods in teaching. Similarly, students might also not be ready to adopt new learning methods (Chance et al., 2007).

1.3 Justification of Study

The integration of Python programming and Google Colab in introductory probability and statistics courses supports an in-depth study for several reasons:

1. Addressing the skills gap: As we know, data is very important for every business, and they need newcomers with strong statistical skills as well as computational. If students already possess skills in using Colab and Python, businesses would be happy to hire them, as this can also help in bridging the skill gap (Donoghue et al., 2021).

2. Enhancing student engagement and understanding: As in Colab, students can write their code and see the instant results, also they can modify the code to get interactive visuals, which makes students active and engaged in learning. This can also help students understand the concepts better, as they can see how the results change with different parameters and codes. With visuals, one can play with code and create different graphs or diagrams with their personal colors and designs, helping the student to spend more time learning (Tintle et al., 2015).

3. Preparing students for the digital age: With the advancement in technology and its easy availability, everything is getting digital in the upcoming decades. If students understand the programming language and cloud-based technology, they can be prepared for future careers and skills (Hardin et al., 2015).

4. Promoting data literacy: Data literacy is very important for every individual, as everyone is talking and utilizing big data and data-driven decisions. Python and Colab give hands-on experience in data analysis, visualization, and manipulation. This increases data literacy among students (Pruim et al., 2017).

5. Addressing educational inequalities: By building effective implementation strategies, we can decrease the gap between developing and developed areas in understanding and learning statistics. This can provide equitable education and decrease existing educational inequalities (UNESCO, 2023).

6. Informing pedagogical approaches: Utilizing this technology, researchers can develop new approaches to teaching subjects with abstract concepts like statistics because, in the digital age, the teaching and learning ways can be different than

traditional ones. This can be helpful in developing new guidelines and approaches to teaching in the rapidly changing world (Wild et al., 2018).

7. Contributing to the body of knowledge: As mentioned earlier, there is much research needed to study the long-term impact of utilizing digital technologies in learning statistics. This project can be one of the research which can be utilized by future educators, researchers, and policymakers (Nolan & Temple Lang, 2010).

8. Aligning with global educational goals: If we can implement it successfully by developing proper strategies, it can be helpful in achieving UNESCO's Sustainable Development Goal (SDG) 4, which is about education, in improving the quality and significance of education (UNESCO, 2023).

9. Addressing the challenges of remote learning: It can contribute to remote learning, as some education is getting remote through online degrees. Utilizing technology in statistical learning can be more important for online education (Moon et al., 2023).

10. Promoting interdisciplinary connections: The integration of technologies, Python, and Colab in teaching statistics can also connect with other fields. Those fields can be computer science and data science, which are crucial in interdisciplinary connections (Kim, 2023).

1.4 Objectives

The primary objective of this project is to integrate Python and Google Colab as teaching tools to enhance the students' learning experience in Introductory Probability and Statistics classes.

The specific objectives of the project are:

a) Enhance conceptual understanding of statistical and probabilistic theories.

b) Improve student engagement and active learning.

c) Bridge the gap between abstract theoretical knowledge and practical applications.

d) Promote collaborative and self-paced learning among students.

e) Eliminate software and hardware barriers to using technology in learning.

f) Develop computational thinking and data literacy among students.

g) Contributes to the modernization of statistics education.

2. LITERATURE REVIEW

2.1 Technology Integration in Education

The lives of people are changing rapidly, from their interactions to communication and conducting business. This is all due to advancements in digital technology. As stated earlier, the environment in educational institutions is no exception. A huge investment has been made to integrate technology into K-12 classrooms to prepare students with the skills they need before joining college and careers. The schools with more technology integrated into the class have shown significantly higher scores on math achievements and overall grades (Delgado et al., 2015). Technology has increased engagement and learning achievement for students. Incorporating technologies like apps and educational games has improved students' understanding of the subject. Despite all these benefits of using technology in teaching, there are challenges like technology accessibility and students' and teachers' skills in using it properly (Larisang, 2024). Providing teachers with sufficient training and support will allow them to understand the benefits of digital technologies. The teachers' efficiency can significantly increase in preparing and teaching materials. It helps in reducing inequalities and gaps by promoting inclusion. Technologies support educational attainment, particularly math and science (ICF International (Firm), 2015). Understanding the concepts in science, especially physics, is not feasible by only reading the books and teachers' explanations; it requires an interactive approach. Significant learning achievement exists between conventional methods and technology-based learning, where a problem-based approach is needed. This helps attract students' interest and helps them understand the subject matter in a more effective way. (Kurniawan et al., 2015). The high-performing institutions are more likely to invest more in technology for more effective and efficient learning, which involves collaboration between students and teachers through the digital platform (Larisang, 2024). Multimedia-based instructions promote self-learning, offer more freedom, reduce students' effort in processing traditional learning methods, and, more importantly, make students self-dependent, improving their sense of responsibility (Sousa et al., 2017).

Research has shown the positive impact of integrating multimedia technology in learning dentistry, nursing, science, pharmacy, political science, psychology, biology, business, computer science, statistics, and mathematics (Rackaway, 2010; Terrana et al., 2016; Vagg et al., 2020; Wahidah et al., 2020). Knowing the importance of technology in learning, in 2005, the American Statistical Association recommended the use of technology in learning and teaching introductory statistics at the college level (Franklin et al., 2007). Mathematics and statistics learning can be enhanced by motivating students. Technology-enhanced learning (TEL) environment is effective in creating motivation in students. However, this is only one of a few other factors, such as teachers' skills and their understanding of the current stage of motivation in students, that can sustain motivation in the long run (Wong & Wong, 2017). Research among over 500 undergraduate students from two Portuguese universities showed significant effectiveness in learning mathematics and statistics by using technology (Silva et al., 2021). One of many advantages of using technology in classrooms is increasing task-based instruction, where students are actively involved in analyzing real-world data sets and interpreting results to help in decision-making instead of memorizing formulas and procedures (Sousa et al., 2017).

2.2 Python in Statistics Education

Students prefer Python programming skills before stepping into the workplace for better career opportunities and growth. Learning Python programming is easier for students with no programming skills than for those who already have some programming skills (Xu & Frydenberg, 2021). Python programming is one of the most powerful languages in computer science. Due to its simplicity and multiple libraries, Python is used extensively in data analysis. Students who use Python for educational purposes have developed improved problem-solving skills and data literacy. Its versatile nature in statistical problems makes it one of the best tools for probability and statistics courses (VanderPlas, 2016). Deep learning teaching is effective in Python, specifically focusing on data analysis (Kim, 2023). Studies have proved that Python enhances the computational thinking of students. The sixth-grade students improved computational cognition, elaboration, and fluency after using Python for data visualization. Robot-based Python learning model significantly improved in thinking skills of elementary school students (Bai et al., 2021). Python programming is updating in data analytics and allied fields, as a result, its demand is increasing in curricula. In a survey of 64 students at the undergraduate level, business school students wanted to learn Python programming for a better future career in data analytics (Xu & Frydenberg, 2021).

2.3 Google Colab as an Educational Platform

Google Colaboratory, also known as Google Colab in short, is an as-a-service edition of Jupyter Notebook where we can write and run Python programming code in a web browser (Burke, 2023). Jupyter Notebook is a web-based application for writing, developing, running code, and sharing results. It has two components: a web application

– a program that runs on a browser such as Chrome, Safari, Firefox, Opera, and Edge, for a fast-interactive environment for writing and explaining code, discovering and visualizing data, and revealing ideas with others, and computational notebook documents – a document which has computer code, plain language descriptions, visualizations such as interactive controls, charts, 3D models, mathematics, and graphs and figures, and data. There are three types of cells in the notebook: code cells, where we write code; markdown cells, allowing for descriptive text using rich text; and the last one is raw cells, allowing us to write output directly, which will not be assessed by the notebook. The notebook supports a wide range of programming languages (Python, Julia, R, Ruby, Haskell, Scala, Node.js, and Go) to write and run the code (*The Jupyter Notebook — Jupyter Notebook 7.3.2 Documentation*, n.d.). The key features of Google Colab are:

1. It provides a full-fledged of Jupyter Notebook in the cloud, thus requiring no setup or installation. It can be accessed from any type of device connected to the internet.
2. It provides free access to Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), allowing students to run computationally demanding work without setting up expensive hardware systems.
3. With the integration of Google Drive, students can save and share their notebooks with their friends and instructors.
4. Google Colab comes with pre-installed advanced libraries, which include TensorFlow, Keras, PyTorch, Pandas, NumPy, Matplotlib, and Scikit-learn, which are helpful for data science and machine learning (Bisong, 2019; "Google Colab for Data Science," 2024).

3. METHODOLOGY

The project aims to help students learn abstract concepts of statistics and probability with the help of available technology in three different steps.

3.1 Collection of Information

To provide the foundation of this project, "Elementary Statistics, a Step-by-Step Approach, Ninth Edition", a book written by Professor Dr. Allan G. Bluman, was utilized. The book is very easy to understand, which is my key aim to replicate this for the students who find it hard to grasp concepts in statistics. The definition and examples are easy to understand and relate to the students, especially in the United States. The information to be included in the project was gathered using this book.

3.2 Coding Machine and Accessories

After gathering all the information to be included in the project, a personal computer, Samsung Galaxy Book 360, with a 13th Gen Intel(R) Core i7-1360P processor, 16 GB installed RAM operating in Windows 11 Home system, was used. Since Colab is a cloud-based web browser application, a stable internet connection was a must throughout the building of this project. An external monitor with an external mouse and keyboard for better visualization and easy navigation during this project was used.

3.3 Writing Code in Colab

Building chapters on statistics and probability in Colab and saving them in Google Drive using my personal Google account was started. In Colab, first, some information about the topics and chapters for students to help them refresh their knowledge of the topics was provided. If the students needed more information on the topic and Python programming, links to external sources were provided. All this information is provided in the markdown cells. The tables and mathematical formulas are easier in markdown cells, and Google Colab supports LaTeX notation. The Python code with examples to solve statistical problems is written in the code cell after the markdown. The codes are made as simple as possible for the students to learn easily. To help them further, additional comments are provided for lines of code. The codes are editable, so if the students need to understand more about codes, they can easily change them according to their enthusiasm.

These three steps to build this project are explained in Figure 1 as a flow chart. The flow chart is a pictorial representation of the methodology. After this section, students will work on exercises to get experience in writing code on their own. The questions and code cells are provided for the students with additional information and guides for each line as comments in the code.

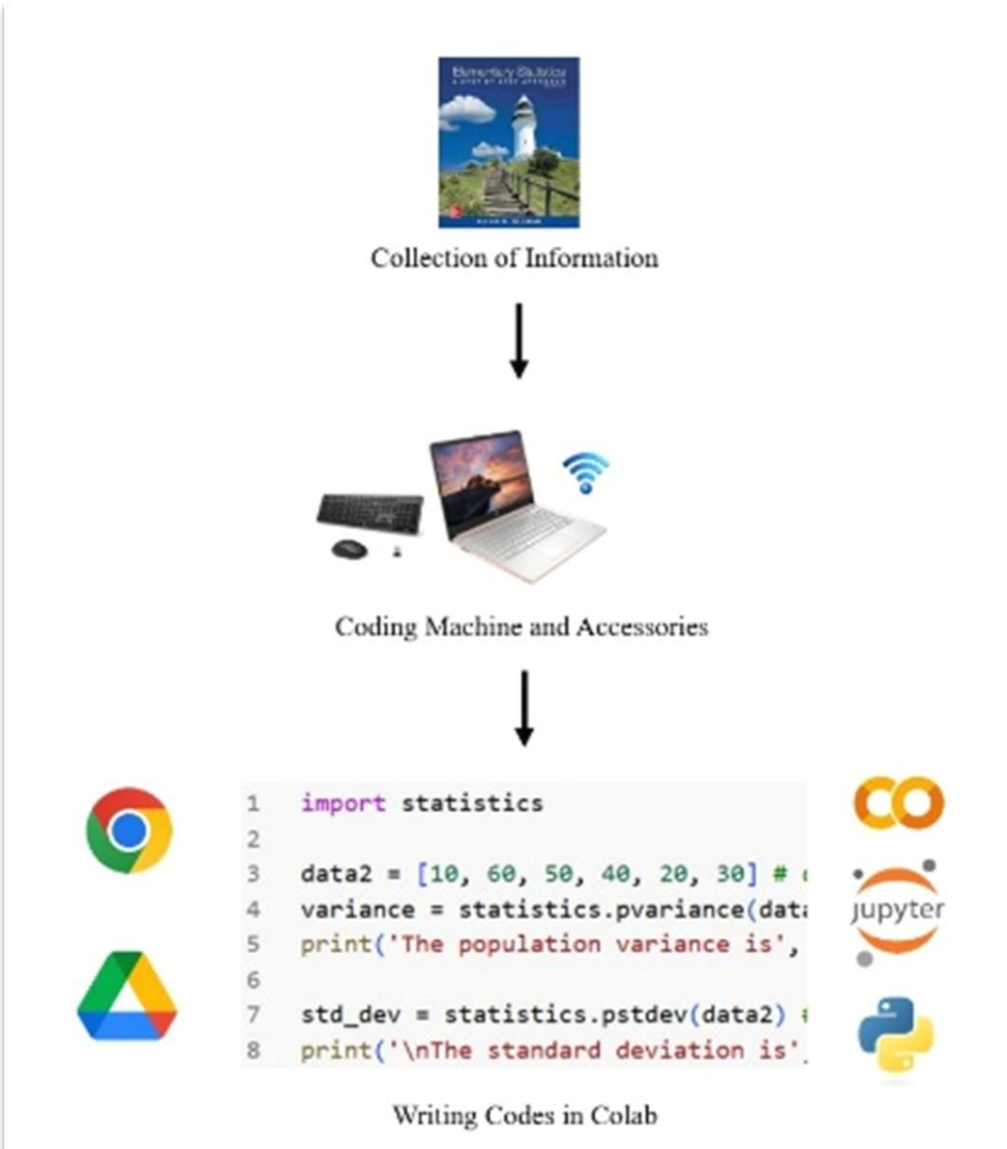


Figure 1: Methodology and Workflow Diagram.

4. SYSTEM ARCHITECTURE

Building an innovative platform to facilitate teaching and learning cannot be achieved using a single technology. This system utilizes the functions of Google Colab and aligns itself with the concept of 'Elementary Statistics', a book written by Bluman. This probability and statistics course was developed using three main components:

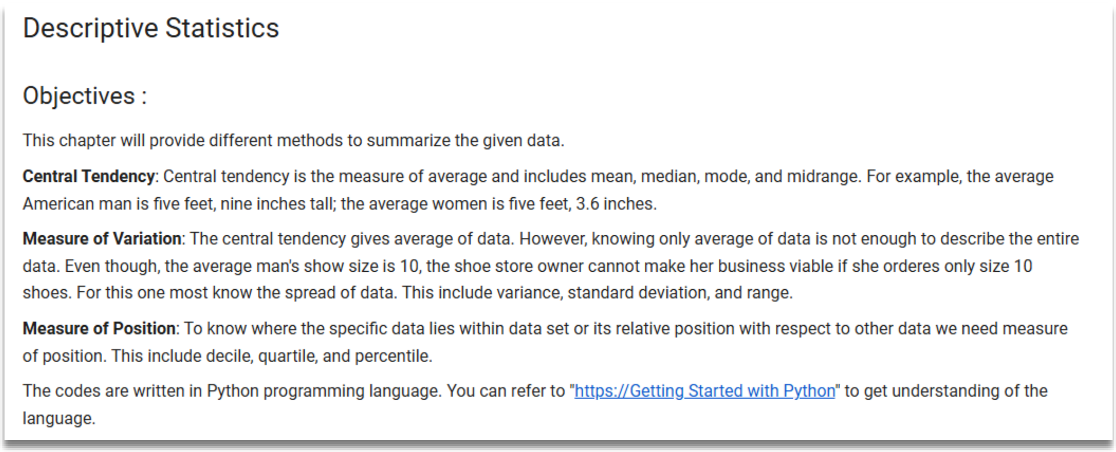
4.1 Students' Information

Colab contains two types of cells: text and code. In the text cell, which is shown in Figure 2, lightweight markup language, Markdown, is used to write text in a plain text editor.

```
#Descriptive Statistics
##Objectives :
This chapter will provide different methods to summarize the given data.

**Central Tendency**: Central tendency is the measure of average and includes
mean, median, mode, and midrange. For example, the average American man is five
feet, nine inches tall; the average women is five feet, 3.6 inches.

**Measure of Variation**: The central tendency gives average of data. However,
knowing only average of data is not enough to describe the entire data. Even
though, the average man's show size is 10, the shoe store owner cannot make her
business viable if she orderes only size 10 shoes. For this one most know the
spread of data. This include variance, standard deviation, and range.

**Measure of Position**: To know where the specific data lies within data set
or its relative position with respect to other data we need measure of
position. This include decile, quartile, and percentile.

The codes are written in Python programming language. You can refer to "
[https://Getting Started with Python](https://ethanweed.github.io/pythonbook/02.
01-getting_started_with_python.html)" to get understanding of the language.
```

Figure 2: Markdown Language in Text Cell.

To display them in the browser through HTML, first, the markdown parser converts the markdown files to HTML. The HTML version of the markdown cell is shown in Figure 3.



Descriptive Statistics

Objectives :

This chapter will provide different methods to summarize the given data.

Central Tendency: Central tendency is the measure of average and includes mean, median, mode, and midrange. For example, the average American man is five feet, nine inches tall; the average women is five feet, 3.6 inches.

Measure of Variation: The central tendency gives average of data. However, knowing only average of data is not enough to describe the entire data. Even though, the average man's show size is 10, the shoe store owner cannot make her business viable if she orderes only size 10 shoes. For this one most know the spread of data. This include variance, standard deviation, and range.

Measure of Position: To know where the specific data lies within data set or its relative position with respect to other data we need measure of position. This include decile, quartile, and percentile.

The codes are written in Python programming language. You can refer to "https://Getting Started with Python" to get understanding of the language.

Figure 3: Markdown Rendered as HTML.

Unlike markdowns used by Jupyter and GitHub, Colab supports LaTeX mathematics for equations. Figure 4 represents how the LaTeX notation is supported in Colab, making it easier to write mathematical formulas. The figure shows the mathematical formula for Population Variance and Standard Deviation. The code in LaTeX, after execution, gives the mathematical symbols and formulas as shown in Figure 5. The mathematical formula is easy to write using LaTeX notation.

```
## Population Variance and Standard Deviation

Data variation is the difference of the data from its mean.
The population variance is the average of the squares of the distance each
value is from the mean. The symbol for the population variance is
$\sigma^2$ . The formula for the population variance is:

$\sigma^2 = \frac {\sum(X - \mu)^2}{N} $


where,

X = individual value

$µ$ = population mean

N = population size

The population standard deviation is the square root of the variance. The
symbol for the population standard deviation is $\sigma$.
Mathematically, population standard deviation is:


$\sigma = \sqrt{\frac{Σ(X-µ)^2}{N}} $
```

Figure 4: LaTeX Notation in Text Cell.

Population Variance and Standard Deviation

Data variation is the difference of the data from its mean. The population variance is the average of the squares of the distance each value is from the mean. The symbol for the population variance is σ^2 . The formula for the population variance is:

$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$

where,

X = individual value

μ = population mean

N = population size

The population standard deviation is the square root of the variance. The symbol for the population standard deviation is σ. Mathematically, population standard deviation is:

$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$

Figure 5: Rendered LaTeX Notation.

To refresh the learners' knowledge and concepts of the topic, the text cell is used to introduce the topic and guide them before directly jumping to the Python code. The plain text with the introduction of the topic in Markdown is shown in Figure 6. The figure shows the introduction of the topic "Means of Groups", including a table structure in Markdown and a mathematical formula in LaTeX notation. The output of the topic in Figure 6 is shown in Figure 7. It shows the HTML version rendered from the markdown language in Figure 6. The table is created perfectly with the mathematical formula of the mean of groups.

```
##Mean of Groups
The students in a class are carrying money which is divided into following
classes.

Calculate the mean money in the class per students.

| Class Interval | Frequency |
|----------------|-----------|
| 0 - 10         | 3         |
| 10 - 20        | 5         |
| 20 - 30        | 3         |
| 30 - 40        | 9         |
| 40 - 50        | 6         |
| 50 - 60        | 3         |
| 60 - 70        | 5         |
| 70 - 80        | 3         |
| 80 - 90        | 4         |
| 90 - 100       | 2         |

To calculate mean, first we need to calculate midpoint of all the money group.
The formula to calculate midpoint is:

$mid \space point = \frac{lower \space point \space + \space higher \space
point}{2}$

The fomula to calculte total money is:

$total \space sum = Σ (\space midpoint \space * \space number \space of \space
students) $

$total \space students = Σ \space frequency$

Finally,

$\bar{x} = \frac{total \space sum}{total \space students}$
```

Figure 6: Topic Overview in Markdown.

Mean of Groups

The students in a class are carrying money which is divided into following classes.

Calculate the mean money in the class per students.

Class Interval	Frequency
0 - 10	3
10 - 20	5
20 - 30	3
30 - 40	9
40 - 50	6
50 - 60	3
60 - 70	5
70 - 80	3
80 - 90	4
90 - 100	2

To calculate mean, first we need to calculate midpoint of all the money group. The formula to calculate midpoint is:

$$mid\ point = \frac{lower\ point + higher\ point}{2}$$

The fomula to calculte total money is:

$$total\ sum = \Sigma(\ midpoint\ *\ number\ of\ students)$$

$$total\ students = \Sigma\ frequency$$

Finally,

$$\bar{x} = \frac{total\ sum}{total\ students}$$

Figure 7: Rendered HTML Version.

4.2 Python Programming Examples

When the students get acquainted with the topics, introductions, and formulas in markdown cells, Python code is written in code cells. A proper guide is provided in code cells, including comments for extra help using Python programming. Figure 8 shows one example using Python to solve a statistical problem. Also, it provides a proper guide to executing code.

To run the code below click on play button or press "Shift + Enter" on keyboard.

```python
# Python Code for to calculate the above question
data1 = [3, 5, 7, 2, 9, 12, 15, 1] # Given data
total_sum = sum(data1) # Calculating sum of the given number
number = len(data1) # Counting total number of the data
mean = total_sum/number # Formula of the data
print ("The mean of record broken per day is ", mean) # Displaying the mean
```

```
The mean of record broken per day is  4.5
```

Figure 8: Statistical Problem Solved in Python.

Figure 9 shows code to import Python modules and libraries, and each line of code is explained using comments. Python has a built-in module to calculate mathematical statistics of numeric data.

Python ignores the characters written after "#". It is used to add comment in the code.

Using Python built in funtion to calculate mean.

```python
import statistics # Importing module
data2 = [3, 5, 7, 2, 9, 12, 25, 1]
arithmetic_mean = statistics.mean(data2) # Using in-built function of the module
print ("The mean of record broken per day is", arithmetic_mean) # Displaying the result
```

```
The mean of record broken per day is 8
```

Figure 9: Using Statistical Module in Python.

This will help even the students who are new to programming. The keyword in Python programming is similar to English words, helping students understand better. In Figure 10, students can also edit the code to see how the codes execute, like changing data for different results, changing color in the graph, boundary color, axes titles, labels, and their angles, editing elements of the figures, and so on.

```
# Plotting the frequency distribution
frequency_table.plot(kind='bar', figsize=(10, 4), color='green',
edgecolor='black')
plt.title('Frequency Distribution of Ages of Billionaires')
plt.xlabel('Age Class')
plt.ylabel('Frequency')
plt.xticks(rotation=90)
plt.show()
```

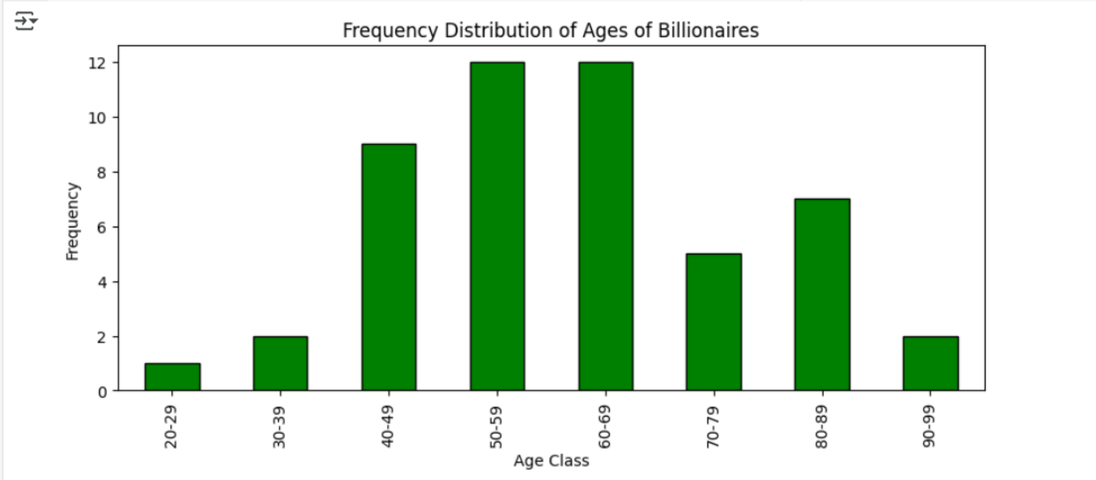


Figure 10: Graphs for Data Visualization.

The code in Colab can be executed in any of the following ways (*Google Colab*, 2025):

a) Click the Play icon in the left gutter of the cell.

b) Type Cmd / Ctrl + Enter to run the cell in place.

c) Type Shift + Enter to run the cell and move focus to the next cell; or

d) Type Alt + Enter to run the cell and insert a new code cell immediately below it.

After they execute the code, just like the Jupyter Notebook environment, the result will be displayed below the code.

4.3 Python Programming Exercises

When the students are familiar with the topic and Python code to solve statistical questions, Figure 11 shows practice questions that have been provided to provide them with hands-on experience. Additional information is provided using comments in the exercise section's code cells.

Figure 11: Exercise for Students.

These line-to-line comments guide them to what is required to solve the questions provided in the exercises. It is more helpful for students new to programming. The students can get the results below the code after executing it in one of the four different ways. If there are mistakes in the code, as a feature of Jupyter Notebook shown in Figure 12, it will give an error with a reason, which is helpful for debugging the code. The error is also pointing to the location of the error with the "^" symbol and the type of error.

```python
# Class boundaries and frequencies
class_boundaries = [99.5, 104.5, 109.5, 114.5, 119.5, 124.5, 129.5, 134.5]
frequency = [2, 8, 18, 13, 7, 1, 1]
midpoints = [] # empty list for midpoints

# calculating midpoints
for i in range(len(class_boundaries)-):
    midpoint = (class_boundaries[i]+class_boundaries[i+1])/2
    midpoints.append(midpoint) # append added the values to the 'midpoints' list

# ploting frequency polygon
plt.plot(midpoints, frequency, marker = 'o', linestyle='-')
plt.title('Frequency Polygon of Record High Temperatures')
plt.xlabel('Temperature Class (°F)')
plt.ylabel('Frequency')
plt.xticks(midpoints) # x-axis data labels
plt.grid()
plt.show()

```

```
  File "<ipython-input-4-9bd2a2cd0ea3>", line 7
    for i in range(len(class_boundaries)-):
                                         ^
SyntaxError: invalid syntax
```

Figure 12: Showing Error in Python Code.

4.4 Future Work

This project has successfully demonstrated the integration of Python and Google Colab in enhancing the teaching of Introductory Probability and Statistics. However, there are several areas for future improvement and expansion to further enhance its impact and scalability. The plan is to continue working on the project after submitting the report. Some of the future works include:

a) Expanding Content Coverage: Additional chapters like hypothesis testing, ANOVA, regression analysis, Bayesian statistics, and others can be included to make statistics learning more comprehensive.

b) Integrating Auto grading: Tools like Otter grader or nbgrader can streamline assessments, reducing the manual workload for instructors. These automated grading systems can provide students with their achieved scores immediately or ask them to rewrite the code and execute it unless the required answer is returned.

c) Scalability and Accessibility: The project can be further optimized to support large-scale deployment through GitHub and similar platforms. This ensures that students globally can benefit from this interactive learning approach.

d) Enhancing Visualization: Adding more active visualizations, interactive widgets, and dashboards can improve students' engagement and motivation. This enhances their learning achievement.

These future advancements will help statistics education ideal with the technologies and tools available for the students, making instructors efficient in teaching. This will make statistics learning more efficient, accessible, and according to the students' requirements.

5. REFERENCES

Bai, H., Wang, X., & Zhao, L. (2021). Effects of the Problem-Oriented Learning Model on Middle School Students' Computational Thinking Skills in a Python Course. *Frontiers in Psychology*, *12*, 771221. https://doi.org/10.3389/fpsyg.2021.771221

Baumer, B. (2015). A Data Science Course for Undergraduates: Thinking With Data. *The American Statistician*, *69*(4), 334–342. https://doi.org/10.1080/00031305.2015.1081105

Bisong, E. (2019). Google Colaboratory. In E. Bisong (Ed.), *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (pp. 59–64). Apress. https://doi.org/10.1007/978-1-4842-4470-8_7

Burke, J. (2023, September 25). *Why and how to use Google Colab | TechTarget*. TechTarget Search Enterprise AI. https://www.techtarget.com/searchenterpriseai/tutorial/Why-and-how-to-use-Google-Colab

Carneiro, T., Medeiros Da NóBrega, R. V., Nepomuceno, T., Bian, G.-B., De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, *6*, 61677–61685. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2874767

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics. *Technology Innovations in Statistics Education*, *1*(1). https://doi.org/10.5070/T511000026

Cobb, G. (2015). Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum from the Ground Up. *The American Statistician*, *69*(4), 266–282. https://doi.org/10.1080/00031305.2015.1093029

Delgado, A., Wardlow, L., O'Malley, K., & McKnight, K. (2015). Educational Technology: A Review of the Integration, Resources, and Effectiveness of Technology in K-12 Classrooms. *Journal of Information Technology Education. Research*, *14*, 397–416. https://doi.org/10.28945/2298

Donoghue, T., Voytek, B., & Ellis, S. E. (2021). Teaching Creative and Practical Data Science at Scale. *Journal of Statistics and Data Science Education*, *29*(sup1), S27–S39. https://doi.org/10.1080/10691898.2020.1860725

Franklin, C. A., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Sheaffer, R. (Eds.). (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. American Statistical Association.

Google Colab. (2025). Overview of Colaboratory Features. https://colab.research.google.com/notebooks/basic_features_overview.ipynb

Google Colab for Data Science. (2024, August 29). *Locus*. https://locusit.se/techpost/technology/google-colab-for-data-science/

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D., & Ward, M. D. (2015). Data Science in Statistics Curricula: Preparing Students to "Think with Data." *The American Statistician*, *69*(4), 343–353. https://doi.org/10.1080/00031305.2015.1077729

Horton, N. J., Baumer, B. S., & Wickham, H. (2015). Setting the stage for data science: Integration of data management skills in introductory and second courses in statistics. *CHANCE*, *28*(2), 40–50. https://doi.org/10.1080/09332480.2015.1042739

ICF International (Firm), I. C. F. (2015). *Literature review on the impact of digital technology on learning and teaching*. Scottish Government.

Kim, D. H. (2023). A Study on an Effective Teaching of AI using Google Colab-Based DCGAN Deep Learning Model Building for Music Data Analysis and Genre Classification. *International Journal of Recent Technology and Engineering (IJRTE)*, *11*(6), 13–25. https://doi.org/10.35940/ijrte.E7351.0311623

Kurniawan, T., Rokhmat, J., & Ardhuha, J. (2015). Perbedaan hasil belajar melalui penerapan model pembelajaran berbasis masalah berbantuan komik fisika dengan pembelajaran konvensional pada siswa kelas viii smpn 1 labuapi tahun ajaran 2013/2014. *Jurnal Pendidikan Fisika Dan Teknologi*, *1*(2), 123–128.

Larisang, L. (2024). Integrating Technology in Learning: A Literature Review. *Technical and Vocational Education International Journal (TAVEIJ)*, *4*(1), Article 1. https://doi.org/10.55642/taveij.v4i1.603

Moon, P. F., Israel-Fishelson, R., Tabak, R., & Weintrop, D. (2023). The Tools Being Used to Introduce Youth to Data Science. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 150–159. https://doi.org/10.1145/3585088.3589363

Nolan, D., & Temple Lang, D. (2010). Computing in the Statistics Curricula. *The American Statistician*, *64*(2), 97–107. https://doi.org/10.1198/tast.2010.09132

Pruim, R., Kaplan, D., & Horton, N. (2017). The mosaic package: Helping students to "think with data" using R. *R Journal*, *9*(1), 77–102. https://doi.org/10.32614/rj-2017-024

Rackaway, C. (2010). Video Killed the Textbook Star? Use of Multimedia Supplements to Enhance Student Learning. *SSRN Electronic Journal*, *8*. https://doi.org/10.2139/ssrn.1547142

Silva, O., Sousa, Á., & Nunes, J. (2021). *Factors that Affect Students' Attitudes Towards Statistics and Technology and Their Interrelationships* (pp. 157–166). https://doi.org/10.1007/978-3-030-68418-1_17

Sousa, L. de, Richter, B., & Nel, C. (2017). The effect of multimedia use on the teaching and learning of Social Sciences at the tertiary level: A case study. *Yesterday and Today*, *17*, 1–22. https://doi.org/10.17159/2223-0386/2017/n17a1

Terrana, A., Leung, A., & Jerzak, S. (2016). Use of online resources by physics students in Canada: A cross-sectional study. *Canadian Journal of Physics*, *95*, 201–208. https://doi.org/10.1139/cjp-2016-0779

The Jupyter Notebook—Jupyter Notebook 7.3.2 documentation. (n.d.). Jupyter. Retrieved February 9, 2025, from https://jupyter-notebook.readthedocs.io/en/stable/notebook.html

Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., & VanderStoep, J. (2015). Combating Anti-Statistical Thinking Using Simulation-Based Methods Throughout the Undergraduate Curriculum. *The American Statistician*, *69*(4), 362–370. https://doi.org/10.1080/00031305.2015.1081619

UNESCO. (2023, July 25). *Technology in education: GEM Report 2023 | Global Education Monitoring Report*. UNESCO. https://www.unesco.org/gem-report/en/technology

Vagg, T., Balta, J. Y., Bolger, A., & Lone, M. (2020). Multimedia in Education: What do the Students Think? *Health Professions Education*, *6*(3), 325–333. https://doi.org/10.1016/j.hpe.2020.04.011

VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data* (First). O'Reilly Media, Inc.

Wahidah, N. I., Sari, W. R., Festiana, I., & Nasir, N. (2020). Game-Based Learning: Design a Multimedia with DDD-E Model for Mathematics Education. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(21), Article 21. https://doi.org/10.3991/ijet.v15i21.16353

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What Is Statistics? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 5–36). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_1

Wong, S. L., & Wong, S. L. (2017). Motivation towards Mathematics Learning in the Technology-enhanced Environment. *International Conference on Computers in Education*. https://library.apsce.net/index.php/ICCE/article/view/2169

Xu, J., & Frydenberg, M. (2021). Python Programming in an IS Curriculum: Perceived Relevance and Outcomes. *Information Systems Education Journal*, *19*(4), 37–54.

6. APPENDICES

The project is supported by a publicly available GitHub repository containing all the related materials, including:

- Python scripts and Jupyter/Colab notebooks
- Visualizations
- Presentation file
- Instructions for reproducing and using the notebooks

GitHub Repository Link: https://github.com/ashish-chaudhary024/StatsLearningWithPython

Users can clone or download the repository to explore and run the notebooks independently. The materials are organized by chapter and topics, and include comments and explanations to support learning and reuse.

THIS PROJECT IS APPROVED FOR
RECOMMENDATION TO THE GRADUATE COUNCIL

Project Advisor:

Anna Harris 6/20/2025

Dr. Anna Harris Date

Project Committee:

6/23/2025

Dr. Vinay Raj Date

6/23/2025

Dr. Qinglong Jiang Date

06/20/2025

Dr. Eddie Branch
Graduate Coordinator, Department of Math and Computer Science

Date

06/20/2025

Dr. Eddie Branch
Chair, Department of Mathematics and Computer Science

Date