

User Instructions: Bernoulli Distribution

...

By Naseeba Faiza
CSE300 Fall 2023
Stony Brook University

To Text or Not to Text?

...

A **Bernoulli Distribution** Example



Target audience



21+ non-computer science majors
with curiosity for statistics and
decision-making

Why?

- Switch careers to Data Science
- Or just be more decisive!

Table of Contents

1. Introduction
2. Bernoulli Distribution Overview
3. When to Use Bernoulli Tests
4. Step-by-Step Instructions:
 1. Define Your Hypotheses
 2. Collect Data
 3. Calculate Probability
 4. Conduct Bernoulli Test using Statistical Software
 5. Interpret Results
5. Conclusion

Introduction

Have you ever wondered if you should send a text message to your ex or not while you are intoxicated? In this tutorial, we will use the idea of deciding to text or not to text as an example to understand **Bernoulli Distributions** in data science. By the end, you will understand how Bernoulli distribution works and how it can help you make decisions based on probability.



Bernoulli Distribution Overview

What Is It?

Whenever you are running an experiment which might lead to either a “**success**” or a “**failure**,” you can associate your success with p and failure with $(1-p)$.

“Success”

Texting your ex = p

“Failure”

Not texting your ex = $(1-p)$

How do we define the values of success and failure? Let's apply the Bernoulli Distribution overview to our decision of texting an ex.



Step 1: Define Your Hypotheses

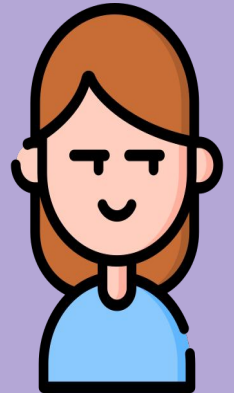
What is Hypotheses in Data Science?

Hypothesis originates from the Greek words ‘hupo’ (translation: under) and ‘thesis’ (translation: placing), which means an idea you have that is generated from limited evidence. This serves as the starting point for further investigation that can validate the accuracy of your idea.

“Therefore, a hypothesis can be defined as a statement that is an informed guess but cannot be believed until proven true. When proven true, it becomes a fact.” [\[Future Learn - Hypothesising in Data Science\]](#)

Null Hypothesis (H_0)

Sending a text message to your ex (success) will have a positive outcome to the situation.



Alternative Hypothesis (H_0)

Sending a text message to your ex (success) will not have a positive outcome to the situation.



Step 2: Collect Data

Ask your friends who are of legal drinking age about similar situations.

Keep a record of whether texting their ex had a positive outcome (success) or not (failure). Create a simple table to keep track of your data.

Texted his/her/their text	Number of times event occurred
Success	6
Failure	3

Step 3: Calculate p

$$p = \text{Number of Successes} \div \text{Total Number of Attempts}$$

Based on the data you collected, calculate p, which is the likelihood of a positive outcome when you send a text message to your ex.

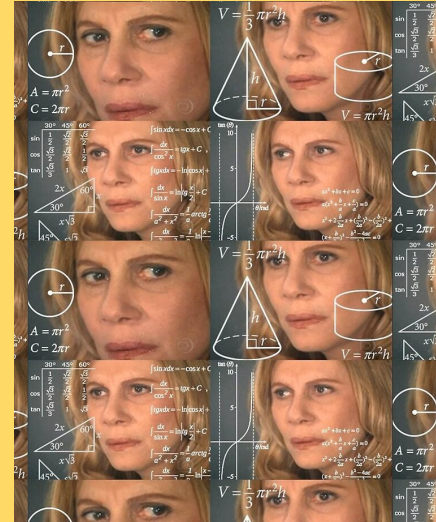
$$6 \div 9 = 0.67$$

Step 4: Conduct Bernoulli Test

- Choose a commonly used significance level (0.05) to determine what is considered statistically significant.
- Statistical significance is a determination made by an analyst that the results in the data are not explainable by chance alone. Therefore, a **p**-value of less than 0.05 is considered statistically significant. [\[Bernoulli Test\]](#)
- Perform Bernoulli test based on **p** using statistical software [\[SciPy Python\]](#)

Step 5: Interpret Results

If the value of **p** is significantly different from **(1-p)**, it suggests that a text message has a statistically significant impact on the situation. If the value of **p** is high, you might want to text your ex!



Happy Decision Making!

Congratulations! You've just learned about Bernoulli Distribution—a powerful tool for decision-making based on probabilities. By understanding the probabilities of success and failure and conducting a Bernoulli test, you can make more decisive about whether to text your ex-partner or not.

