clicks x	frequency fx
0	90
1	180
2	120
3	110

$N = 90 + 180 + 120 + 110 = 500$
Total users $= 500 \times 3 = 1500$
Total clicks $= 0 \times 90 + 1 \times 180 + 2 \times 120 + 3 \times 110 = 750$

Empirical PMF.

$\hat{P}_X(x) = f_x / N$

x	f_x	$\hat{P}_X(x)$
0	90	$90/500 = 0.18$
1	180	$180/500 = 0.36$
2	120	$120/500 = 0.24$
3	110	$110/500 = 0.22$

3) Estimated probability exactly 2 users click in a batch.
$\hat{P}(X=2) = 0.24$

4) $E[X] = \sum_x x \hat{p}_X(x) = 0 \cdot 0.18 + 1 \cdot 0.36 + 2 \cdot 0.24 + 3 \cdot 0.22$
$= 1.5$.

Week-3&4

2) Short Answers

1) Random ~~vale~~ variable is a function that maps each outcome of a random experiment to a real number. It formalizes uncertain quantities so we can reason about their probabilities, expectations & variability.

2) Empirical PMF — observed relative frequencies of outcomes computed from data.

Theoretical PMF — PMF implied by a probabilistic model with specified parameters.

$\hat{P}$ = sample proportion
$P(X=x)$ probability mass function value.

3) Case-Based Questions.

3.1) Case - Email Marketing Campaign.

Given.
$N = 500$
$n = 3$ users.

1) Empirical PMF

$\hat{P}_K(k) = f_k / N$

k	f_k	empirical PMF $\hat{P}_K(k)$
0	240	240/600 = 0.40
1	190	190/600 = 0.3167
2	110	110/600 = 0.1833
3	40	40/600 = 0.0667
4	15	15/600 = 0.0250
5	5	5/600 = 0.0083

3) $\hat{P}(K=2) = 110/600 = 0.1833$ (about 18.33%)

$E[K] = \sum_k k \hat{p}_K(k) = 1.025$.

4) $n = 5$.

$\hat{p} = \frac{\text{total errors}}{\text{total items}} = \frac{615}{3000} = 0.205$.

k	theoretical
0	~0.318
1	~0.409
2	~0.210
3	~0.055
4	~0.007
5	~0.0005.

Binomial assumption and estimate $\hat{p}$

$X \sim \text{Binomial}(n=3, p)$.

$\hat{p} = \frac{\text{total clicks}}{\text{total users}} = \frac{750}{1500} = 0.5$.

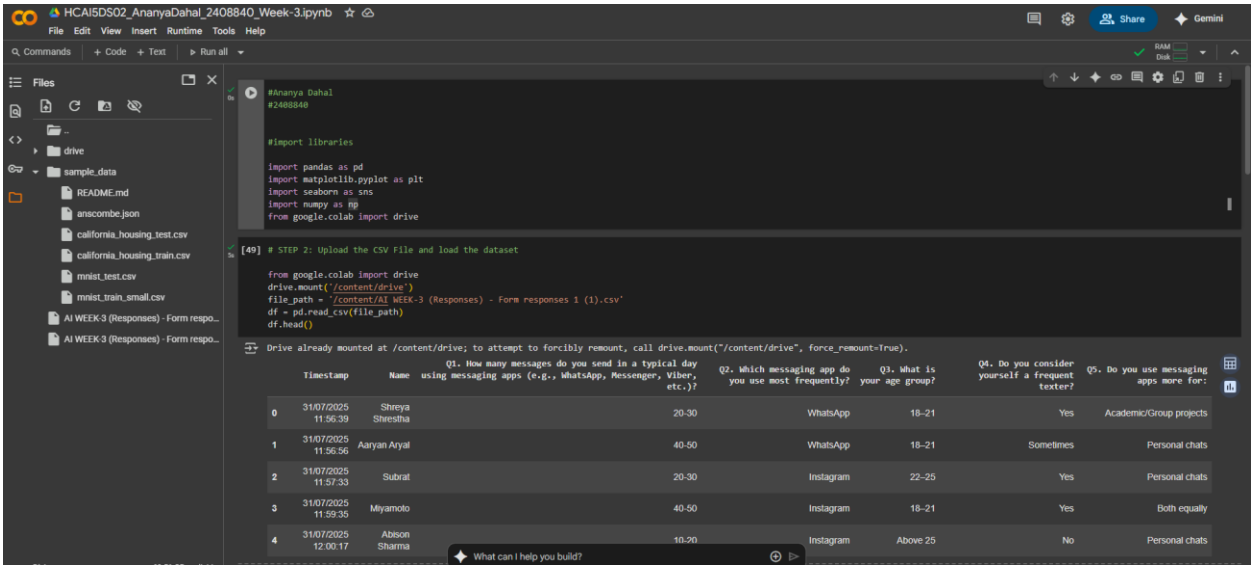
Empirical : [0.18, 0.36, 0.24, 0.22]
Theoretical (Binomial $n=3$, $p=0.5$): [0.125, 0.375, 0.375, 0.125]

2. Case - Online Order Fulfillment Center

$N = 600$ $n = 5$ items

Errors	frequency
0	240
1	190
2	110
3	40
4	15
5	5

Total items inspected $= 600 \times 5 = 3000$
Total errors $= 0 \cdot 240 + 1 \cdot 190 + 2 \cdot 110 + 3 \cdot 40 + 4 \cdot 15 + 5 \cdot 5 = 615$


HCAI5DS02_AnanyaDahal_2408840_Week-3.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text Run all
Files
drive
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
AI WEEK-3 (Responses) - Form respo...
AI WEEK-3 (Responses) - Form respo...
#Ananya Dahal
#2408840
#import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from google.colab import drive
[49] # STEP 2: Upload the CSV File and load the dataset
from google.colab import drive
drive.mount('/content/drive')
file_path = "/content/AI WEEK-3 (Responses) - Form responses 1 (1).csv"
df = pd.read_csv(file_path)
df.head()
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
Timestamp
Name
Q1. How many messages do you send in a typical day using messaging apps (e.g., WhatsApp, Messenger, Viber, etc.)?
Q2. Which messaging app do you use most frequently?
Q3. What is your age group?
Q4. Do you consider yourself a frequent texter?
Q5. Do you use messaging apps more for:
0 31/07/2025 11:56:39 Shreya Shrestha 20-30 WhatsApp 18–21 Yes Academic/Group projects
1 31/07/2025 11:56:56 Aaryan Aryal 40-50 WhatsApp 18–21 Sometimes Personal chats
2 31/07/2025 11:57:33 Subrat 20-30 Instagram 22–25 Yes Personal chats
3 31/07/2025 11:59:35 Miyamoto 40-50 Instagram 18–21 Yes Both equally
4 31/07/2025 12:00:17 Abison Sharma 10-20 Instagram Above 25 No Personal chats
What can I help you build?


HCAI5DS02_AnanyaDahal_2408840_Week-3.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text Run all
Files
drive
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
AI WEEK-3 (Responses) - Form respo...
AI WEEK-3 (Responses) - Form respo...
[50] # Rename the column name for easier use
df = df.rename(columns={
df.columns[2]: "Messages_Per_Day"
})
[51] #STEP 4: Create Frequency Table
frequency_table = df["Messages_Per_Day"].value_counts().sort_index()
total_responses = frequency_table.sum()
[52] #Compute Empirical PMF
empirical_pmf = frequency_table / total_responses
[53] # Combine into a DataFrame
pmf_table = pd.DataFrame({
"Frequency": frequency_table,
"Empirical PMF": empirical_pmf
})

HCAI5DS02_AnanyaDahal_2408840_Week-3.ipynb

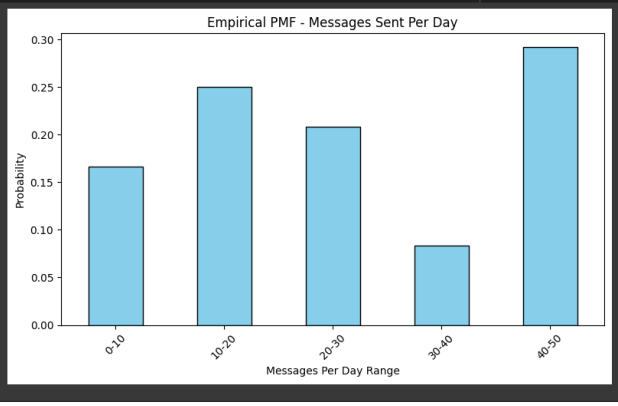
File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

..
drive
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
AI WEEK-3 (Responses) - Form respo...
AI WEEK-3 (Responses) - Form respo...

```
# STEP 6: Plot Empirical PMF
plt.figure(figsize=(8, 5))
pmf_table["Empirical PMF"].plot(kind='bar', color='skyblue', edgecolor='black')
plt.title("Empirical PMF - Messages Sent Per Day")
plt.xlabel("Messages Per Day Range")
plt.ylabel("Probability")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



HCAI5DS02_AnanyaDahal_2408840_Week-3.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Run all

RAM Disk

Files

..
drive
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
AI WEEK-3 (Responses) - Form respo...
AI WEEK-3 (Responses) - Form respo...

```
[55] # STEP 7: Estimate Mean and Variance using Midpoints
     midpoints = {
         "0-10": 5,
         "10-20": 15,
         "20-30": 25,
         "30-40": 35,
         "40-50": 45
     }
```

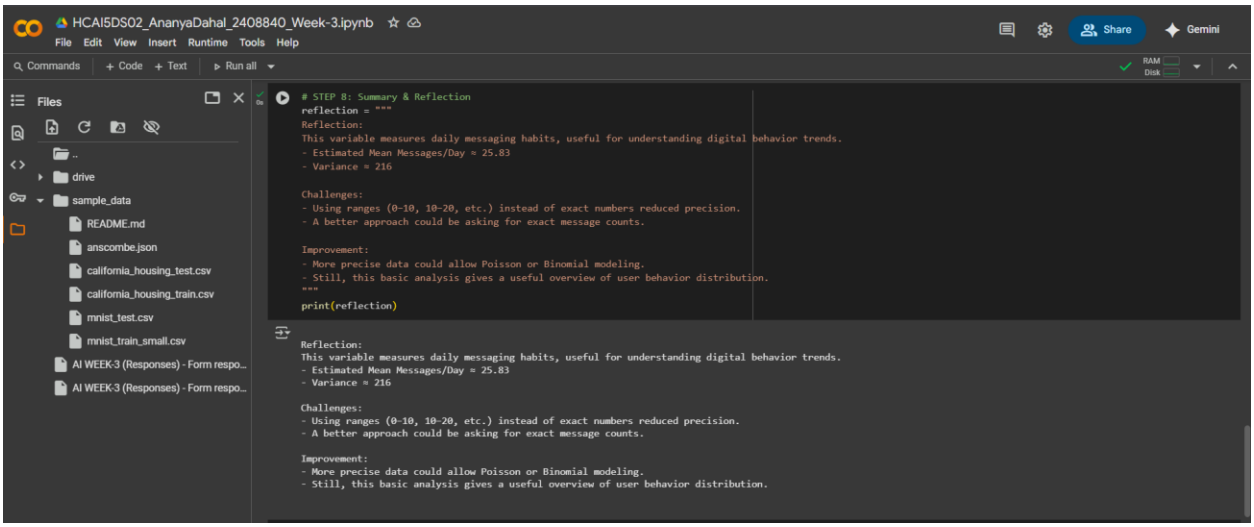
```
[56] # Add midpoints to the PMF table
     pmf_table["Midpoint"] = pmf_table.index.map(midpoints)
```

```
[57] # Estimate sample mean
     mean_estimate = (pmf_table["Midpoint"] * pmf_table["Empirical PMF"]).sum()
```

```
# Estimate sample variance
E_X2 = ((pmf_table["Midpoint"]**2) * pmf_table["Empirical PMF"]).sum()
variance_estimate = E_X2 - (mean_estimate**2)

print(f"Estimated Mean: {mean_estimate:.2f}")
print(f" Estimated Variance: {variance_estimate:.2f}")
```

```
Estimated Mean: 25.83
 Estimated Variance: 215.97
```



Link:
https://docs.google.com/spreadsheets/d/1yMn_a5SyY9SO6obBlDt6sGbVcD7kUDooIf9thAWNfik/edit?usp=sharing