```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import poisson

# Load your CSV file
df = pd.read_csv("responceform.csv")



# Clean and map the main variable to numeric
usage_raw = df["How many times do you use AI tools (like ChatGPT, Bard) in a typical week? "]

# Replace "7 or more" with numeric 7
df['ai_usage_per_week'] = usage_raw.replace("7 or more", 7).astype(int)
df['ai_usage_per_week'].head()
```

```
0    3
1    4
2    7
3    2
4    7
Name: ai_usage_per_week, dtype: int32
```

```python
# Frequency Table
freq_table = df['ai_usage_per_week'].value_counts().sort_index()
print("Frequency Table:\n", freq_table)

# PMF
total = len(df)
pmf = freq_table / total
print("\nEmpirical PMF:\n", pmf)
```

```
Frequency Table:
 ai_usage_per_week
2    1
3    5
4    6
5    4
6    1
7    8
Name: count, dtype: int64

Empirical PMF:
 ai_usage_per_week
2    0.04
3    0.20
4    0.24
5    0.16
6    0.04
7    0.32
Name: count, dtype: float64
```

```python
plt.figure(figsize=(8, 5))
sns.barplot(x=pmf.index, y=pmf.values, color="skyblue")
plt.title("Empirical PMF: AI Tool Usage Per Week")
plt.xlabel("Number of Uses")
plt.ylabel("Probability")
plt.grid(True)
plt.show()
```

Empirical PMF: AI Tool Usage Per Week
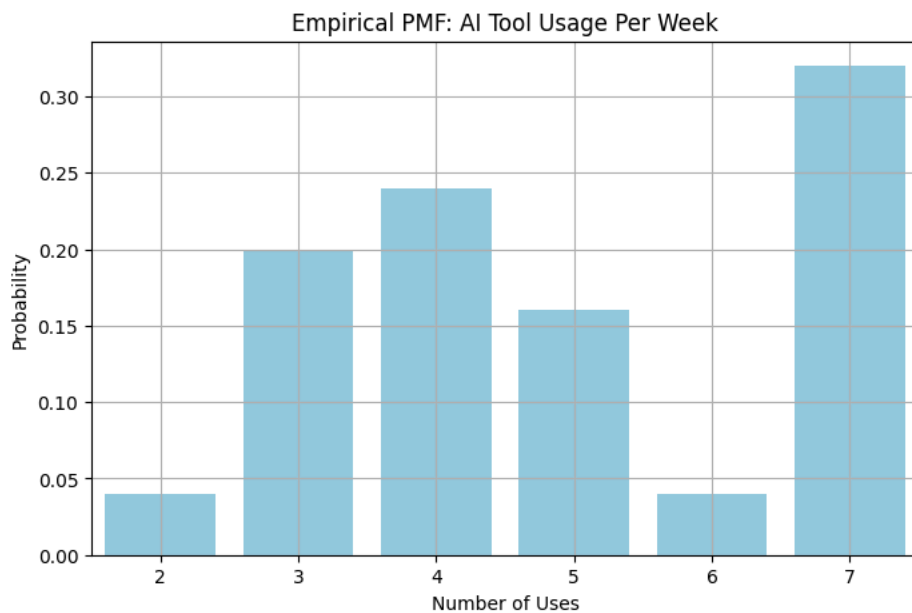

```python
mean = df['ai_usage_per_week'].mean()
var = df['ai_usage_per_week'].var()

print(f"Sample Mean: {mean:.2f}")
print(f"Sample Variance: {var:.2f}")
```

```
Sample Mean: 4.92
Sample Variance: 2.83
```

```python
# Generate Poisson PMF using sample mean as λ
x = np.arange(0, df['ai_usage_per_week'].max() + 1)
poisson_pmf = poisson.pmf(x, mu=mean)

# Plot both empirical and Poisson PMF
plt.figure(figsize=(8, 5))
plt.bar(pmf.index, pmf.values, label='Empirical PMF', alpha=0.6)
plt.plot(x, poisson_pmf, 'ro-', label='Poisson PMF (λ=mean)')
plt.title("Empirical vs. Poisson PMF")
plt.xlabel("Number of Uses Per Week")
plt.ylabel("Probability")
plt.legend()
plt.grid(True)
plt.show()
```

1. What your variable measured and why it matters

The variable measured how many times participants use AI tools like ChatGPT or Bard in a typical week. This is important as it reflects engagement with AI in everyday academic or personal tasks. Understanding usage frequency helps educators and developers assess the growing dependency on AI, inform curriculum updates, and guide responsible use policies.

2. Any challenges you faced in survey design or data collection

A major challenge was inconsistent formatting in responses—some participants wrote "7 or more" while others used exact numbers. This required data cleaning to extract usable numerical values. Additionally, some participants may have under- or over-estimated their usage due to recall bias or social desirability bias.

3. Whether your theoretical model fit the data well

The empirical distribution was compared to a Poisson model with λ = 4.92. Visually, the model captured the general shape but did not fit perfectly—especially at the extreme value (7 times/week), which had a higher-than-expected frequency. This suggests the Poisson model is a rough approximation but may not fully capture human usage behavior, which could be influenced by habit or assignment schedules.

˅ 4.What you would do differently if you repeated this experiment

If repeated, I would use predefined numeric choices (e.g., 0–10+) in the survey to reduce ambiguity. I would also collect more responses to improve statistical power and consider adding context-specific questions (e.g., "Was your usage related to assignments or curiosity?") for deeper insights. Additionally, I might explore alternative models like the Binomial or Negative Binomial for better fit.

Start coding or generate with AI.