Section 1: Sampling and Data

1.1 Population vs. Sample

Define population and sample:

- **Population**: All 10,000 active BrightRetail customers.
- **Sample**: The 200 customers BrightRetail surveys to represent the opinions of the entire customer base.

Why sampling is necessary:

- Surveying all 10,000 customers would be expensive, time-consuming, and impractical. Sampling allows BrightRetail to gather insights efficiently while still maintaining reliability, as long as the sample is representative.

Parameter vs. Statistic:

- The result from surveying 200 customers is a **statistic**.
- A **statistic** summarizes a sample, while a **parameter** summarizes a whole population (which we don't have here).

1.2 Ethical Survey Design

Sampling Method:

- **Stratified Random Sampling**: Divide customers into relevant strata (e.g., based on purchase frequency, age, location) and randomly sample within each group.
- **Justification**: Ensures that the sample reflects important subgroups of customers and improves representativeness.

Ethical Consideration:

- **Privacy Protection**: Ensure anonymity — avoid collecting names, addresses, or other personally identifiable information (PII).

Avoiding Bias:

- Use **neutral wording** in questions. Example: Instead of asking, “How much do you dislike our checkout process?”, ask, “How satisfied are you with the checkout process?”

- Randomize question order if possible to minimize order effects.

Section 2: Descriptive Statistics

2.1 Data Summary

[3, 5, 2, 7, 4, 6, 4, 5, 8, 4, 5, 3, 6, 2, 5]

Number of Complaints	**Frequency**
2	2
3	2
4	3
5	4
6	2
7	1
8	1

Mean (Average):

Mean = (3+5+2+7+4+6+4+5+8+4+5+3+6+2+5)/15
= 69/15 ≈ 4.6

Median (Middle Value when ordered):

Ordered list: [2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 8]

Because the total number of occurrences is odd 15 we take the 8th value, the middle one, which classifies the data into two parts.

Middle value = 8th value = **5**

Mode (Most frequent value):

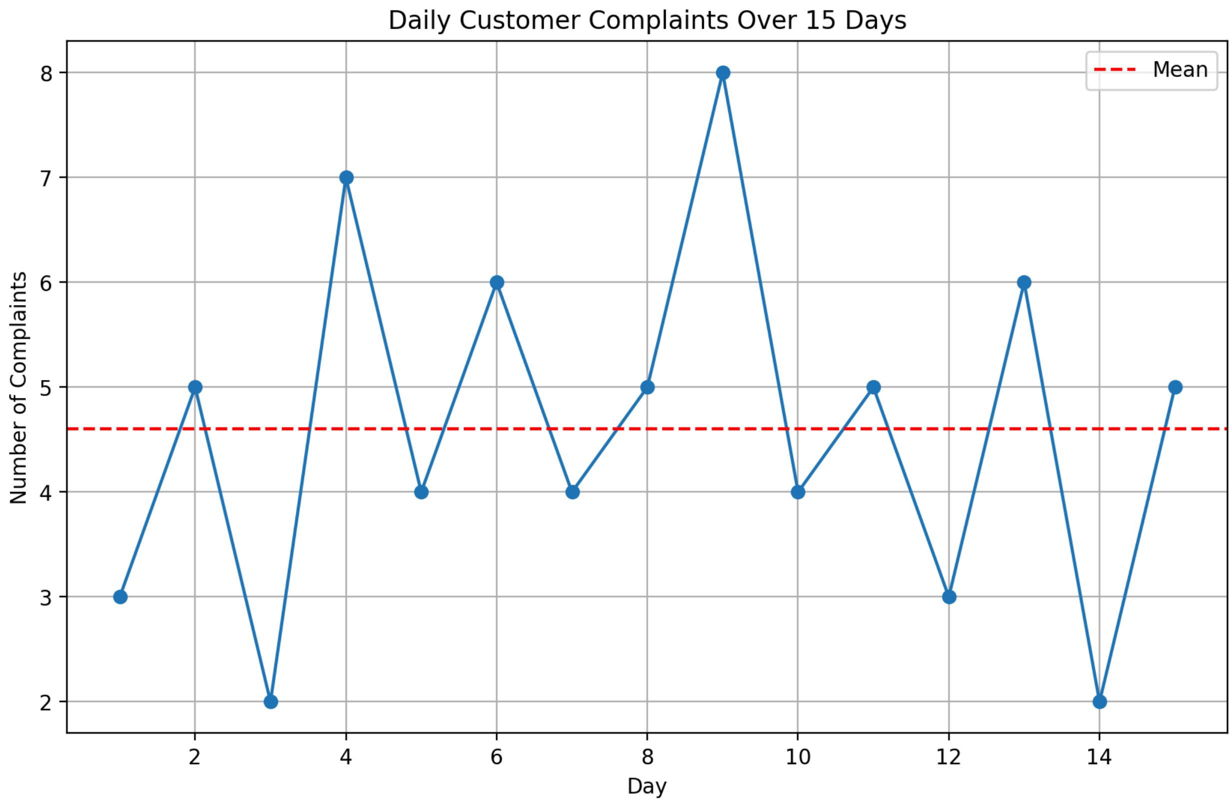
5 appears 4 times → **Mode = 5**

Final Summary:

- Mean ≈ 4.6
- Median = 5
- Mode = 5

2.2 Data Visualization

I chose a line plot with the mean trend line to show the daily fluctuations and overall pattern:



2.3 Pattern Identification

The data shows considerable day-to-day variation, with complaints ranging from 2 to 8 per day. Most days (4 days) had 5 complaints, suggesting this is a typical volume. There are two notable peaks (days with 7 and 8 complaints) that might warrant investigation.

Section 3: Probability

3.1 Probability Basics

Definitions using customer complaint example:

- **Outcome**: A possible result of an experiment.
  Example: A customer files a complaint or does not file a complaint.

- **Event**: A set of outcomes we are interested in.
  Example: The event that a customer files a complaint.

- **Sample Space**: The set of all possible outcomes.
  Example: {Customer files a complaint, Customer does not file a complaint}

3.2 Complaint Analysis

Given:

- Probability a customer files a complaint = 10% = 0.1

- 15 customers selected randomly

This follows a **binomial distribution**:

$$P(x = k) = \langle \frac{n}{k} \rangle P^{K} (1 - P)^{n-k}$$

where:

- n=15

- k=2

- p=0.1
  (a) Probability exactly 2 file complaints

$$P(x = 2) = \binom{15}{2} (0.10)^{2} (0.90)^{13}$$

$$P(x=2) = 105 \times 0.01 \times 0.2542 \approx 0.267$$

Answer: ≈ 0.267

(b) Probability at least 1 files a complaint

Use the complement rule: P(at least 1)=1−P(0)

$$P(x = 2) = \binom{15}{0} (0.10)^{0} (0.90)^{15}$$

$$= 1 \times 1 \times (0.90)^{15} \approx 0.2059$$

: P(at least 1)=1−P(0) = 1-0.2059 = 0.7941≈0.794

Answer: ≈ 0.794

3.3 Risk Assessment (4 Marks)

Given:

- Chatbot error rate p = 5% = 0.05
- 20 users interact with it. n = 20
- Using binomial distribution

we are calculating:

$$P(x = 3) = \left(\frac{20}{3}\right)(0.05)^3(0.95)^{17}$$

=1140×0.000125×0.416≈0.0594

Answer: ≈ 0.0594 (or 5.94%)

Section 4: Discrete Distributions

4.1. Distribution Identification

Counting website crashes per hour: This scenario fits the Poisson distribution. The Poisson distribution is ideal for modeling the number of times an event occurs within a fixed interval of time or space, assuming the events happen independently and at a constant average rate. Since we are counting crashes per hour, it's a classic use case for Poisson.

Finding the probability the first complaint occurs on the 5th customer: This fits the Geometric distribution. The geometric distribution models the number of independent trials until the first success (in this case, the first complaint). We're specifically interested in when the first complaint happens, and the 5th customer is that trial.

Selecting 3 dissatisfied customers from a group of 30: This matches the Hypergeometric distribution. The hypergeometric distribution applies when selecting items from a finite population without replacement. Since we're choosing 3 dissatisfied customers from a total of 30, without replacement, this is not binomial—it's hypergeometric.

4.2. Staffing Strategy

Given:

- Average calls per hour: $\lambda=5$
- Distribution: Poisson $P(x = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

a) Probability of exactly 8 calls in an hour

$$P(x = 8) = \frac{e^{-5}5^8}{8!}$$

$$= \frac{e^{-5}(390625)}{40320} \approx 0.0653$$

Answer: 0.0653 (6.53%)

b) Probability of at least 2 calls in 30 minutes

30 minutes = 0.5 hours → New λ = 5 × 0.5 = 2.5

We want $P(X \geq 2) = 1 - P(0) - P(1)$

$$P(x = 0) = \frac{e^{-2.5}2.5^0}{0!} = e^{-2.5} \approx 0.0821$$

$$P(x = 1) = \frac{e^{-2.5}2.5^1}{1!} = e^{-2.5} \times 2.5 \approx 0.2053$$

P (X≥2) = 1−P(0)−P(1) = 1 - 0.0821 - 0.2053 = 0.7126

Answer: 0.7126 (71.26%)

4.3. Business Application

Using the binomial distribution to predict repeat purchases:

Bright Retail can model the probability of repeat purchases using the binomial distribution, where:

Each customer has a fixed probability p of making a repeat purchase.

There are n customers (trials), and outcomes are independent.

Example:

If Bright Retail finds that 30% of customers typically make a repeat purchase, and they want to estimate how many of the next 20 customers will return:

Let:

- n = 20
- p = 0.30

The probability of exactly k repeat customers is:

$$P(x = k) = \binom{20}{k}(0.30)^k(0.70)^{20-k}$$

This allows the company to forecast inventory, plan marketing campaigns, and allocate staff to loyalty programs.