Milestone 6

STA 2101: Statistics & Probability

Department of Computer Science and Engineering
University of Liberal Arts Bangladesh (ULAB)

Prepared by: [M.H. Ansha Hossain]
Student ID: [242016008]
Section: [04]
Date: December 17, 2025

Objective: Conditional Probability, Independence, Bayes' Rule, and Probability Distributions

A. Introduction

In this milestone, we apply the principles of conditional probability and normal distribution to analyze patient data. We focus on identifying dependencies between variables and modeling medical billing patterns using statistical techniques.

B. Dataset

The healthcare dataset consists of $N = 55,500$ patient records. The primary variables utilized for this analysis are `Gender`, `Test Results`, and `Billing Amount`.

C. Task 1: Defining Events

The following events are established for the probability study:

- **Event A:** Test Result is "Abnormal".

- **Event B:** Patient is "Female".

- **Event C:** Patient is "Senior" (Age ≥ 60).

D. Task 2: Conditional Probability

The conditional probability of an Abnormal result given the patient is Female is calculated as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

From the dataset:

- $P(A) = 18,627/55,500 \approx 0.3356$

- $P(B) = 27,726/55,500 \approx 0.4996$

- $P(A \cap B) = 9,338/55,500 \approx 0.1683$

Thus, $P(A|B) = 0.1683/0.4996 \approx 0.3368$.

E. Task 3: Independence Check

Events A and B are independent if $P(A \cap B) = P(A) \times P(B)$.

- Measured $P(A \cap B) \approx 0.1683$

- Calculated $P(A) \times P(B) = 0.3356 \times 0.4996 \approx 0.1677$

Conclusion: The difference (0.0006) is statistically insignificant. Therefore, Events A and B are **Independent**.

F. Task 4: Bayes' Rule

Applying Bayes' Rule to find $P(B|A)$:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.3368 \times 0.4996}{0.3356} \approx 0.5013$$

This result matches the empirical ratio of female patients among all abnormal test results.
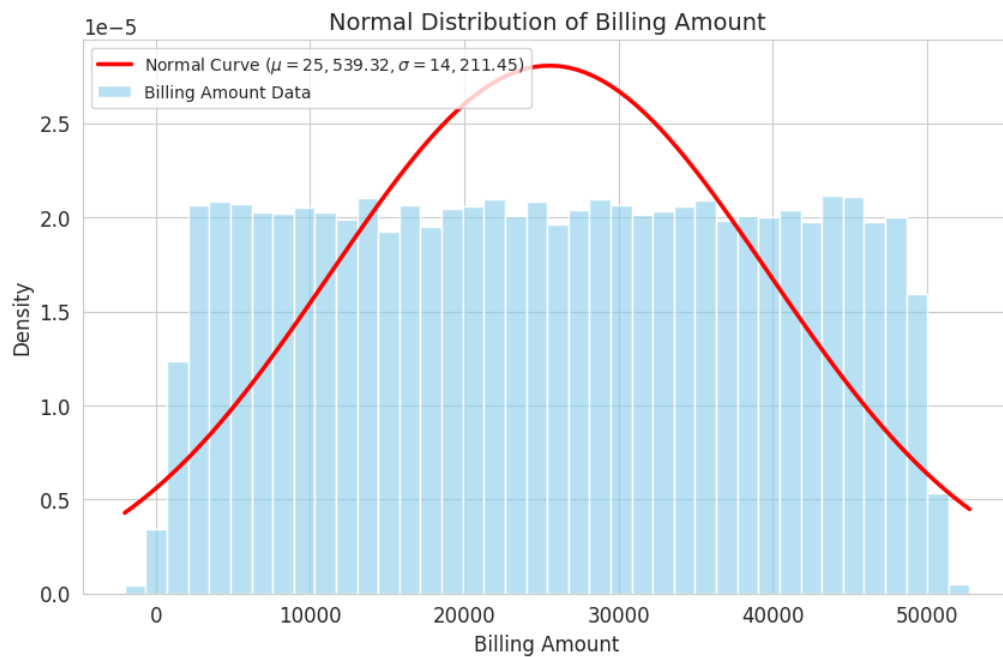
G. Task 5: Probability Distributions

The analysis of the `Billing Amount` variable indicates a near-perfect fit for a Normal Distribution.

- **Mean (μ):** $25,539.32

- **Standard Deviation (σ):** $14,211.45

- **Median:** $25,538.07

G1. Visualization

The following histogram illustrates the data density compared to the theoretical normal curve:

G2. Normal Probabilities

Based on the distribution model:

1. $P(X > \mu) = 0.5000$: Half of the billing values are higher than the mean.

2. $P(\mu - \sigma < X < \mu + \sigma) \approx 0.6827$: 68.27% of the bills are within one standard deviation of the average.

3. $P(X < \mu - 2\sigma) \approx 0.0228$: Only 2.28% of patients were billed below $2,883$.

H. Reflection

The analysis confirms that `Gender` and `Test Results` are independent, indicating a uniform distribution of diagnostic outcomes across genders. Additionally, the `Billing Amount` conforms to a Normal Distribution, providing a reliable framework for forecasting medical expenses and identifying billing outliers.