Assignment 1

Aditi Sahu
241030006
Civil Engineering
Email: aditisahu24@iitk.ac.in

July 1, 2025

Question 1: Probability Using Normal Distribution

A manufacturer claims the Z-Phone has a mean life of 42 months with a standard deviation of 8 months. We are to calculate the probability that a randomly selected Z-Phone lasts between 20 and 30 months.

$$z_1 = \frac{20 - 42}{8} = -2.75, \quad z_2 = \frac{30 - 42}{8} = -1.5$$

Using the standard normal distribution table:

$$P(20 < X < 30) = \Phi(-1.5) - \Phi(-2.75) = 0.0668 - 0.0030 = 0.0638$$

Answer: The probability is approximately 0.0638.

Question 2: Estimating Lifetime with Censored Data

Given failure times: 75, 63, 100+, 36, 51, 45, 80, 90 (with 100+ indicating censoring), we use Maximum Likelihood Estimation (MLE) assuming an exponential distribution.

Total observed lifetime: 440, Censored observation: 100, Number of failures: 7

$$\hat{\lambda} = \frac{7}{440 + 100} = 0.0130, \quad \text{Estimated Mean} = \frac{1}{\hat{\lambda}} = 76.90$$

Answer: Estimated mean lifetime is approximately 76.90 hours using MLE.

Question 3: Linear Regression on Age and Weight

Given data summary for age (x) and weight (y), we use simple linear regression.

$$\hat{\beta}_1 = 1.05, \quad \hat{\beta}_0 = 117.0$$

Fitted regression line:

$$\hat{y} = 117.0 + 1.05x$$

(b) Prediction for a 25-Year-Old:

$$\hat{y} = 117.0 + 1.05 \cdot 25 = 143.25$$

(c) Residual if Actual Weight is 170 lbs:

$$170 - 143.25 = 26.75$$

(d) Interpretation:

Since the residual is positive, the prediction was an **underestimate**.

Question 4: Descriptive Statistics and Box Plot

Gasoline Type 1 Data:

- Mean = 2.41 seconds
- Variance = 0.26
- Standard Deviation = 0.51 seconds

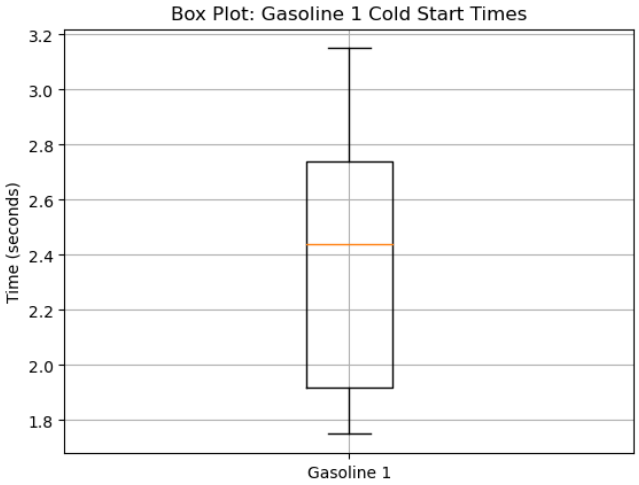

Figure 1: Box Plot: Gasoline Type 1

Gasoline Type 2 Data:

- Mean = 2.69 seconds
- Variance = 0.38
- Standard Deviation = 0.62 seconds

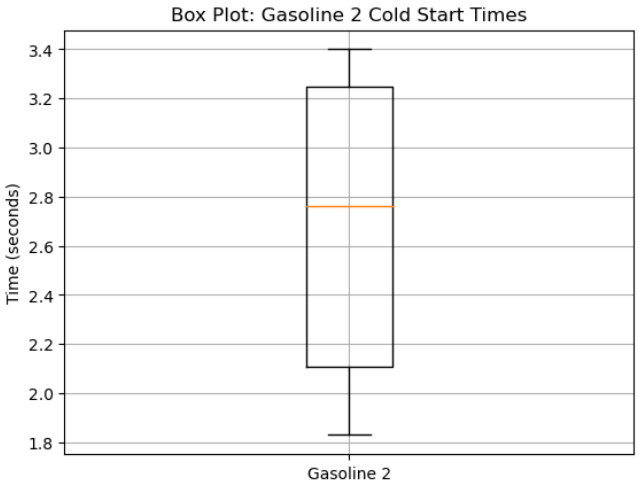

Figure 2: Box Plot: Gasoline Type 2

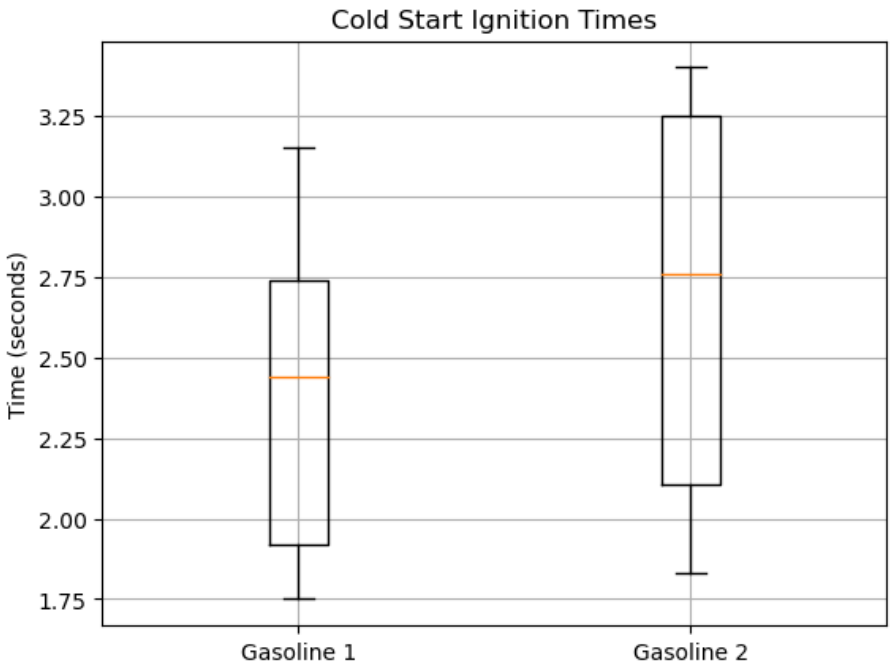

Figure 3: Comparative Box Plot: Gasoline Type 1 vs Type 2

Interpretation: Gasoline Type 2 shows slightly higher median and greater variability than Type 1, suggesting that Type 2 may ignite more slowly and less consistently.

Question 5: Linear and Polynomial Regression with Noise

(a) 100 data points were generated from a linear model with Gaussian noise. The data was split 80:20 into training and testing sets.

(b) Linear regression was applied to the training data using `sklearn.linear_model.LinearRegression`.

(c) Mean Squared Error (MSE) on the test set was computed.

(d) Polynomial regression models of degree 2 and 3 were also fit to the data, and their predictions compared visually and using MSE.

- Degree 1 (Linear): MSE = 0.8047
- Degree 2: MSE = 0.8499
- Degree 3: MSE = 0.8845

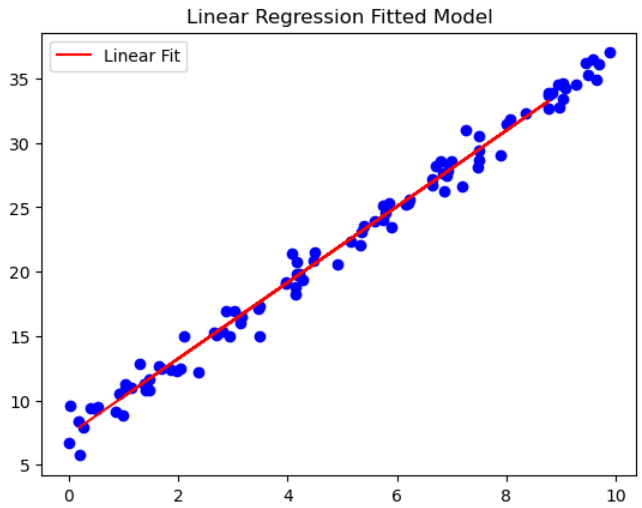

Figure 4: Best Fit Linear Model

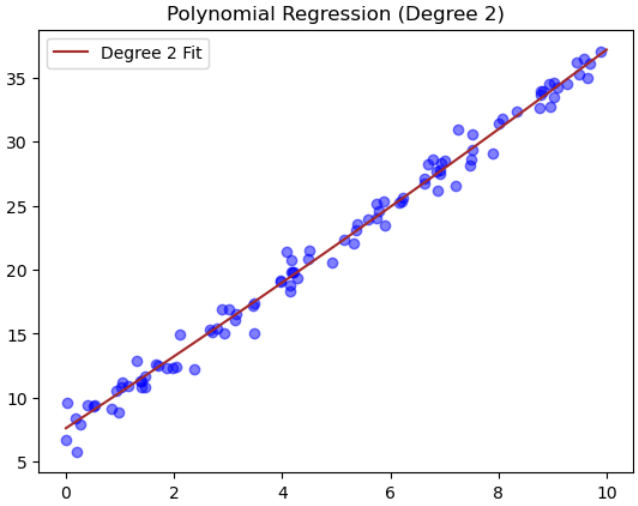

Figure 5: Best Fit Polynomial (degree = 2) Model

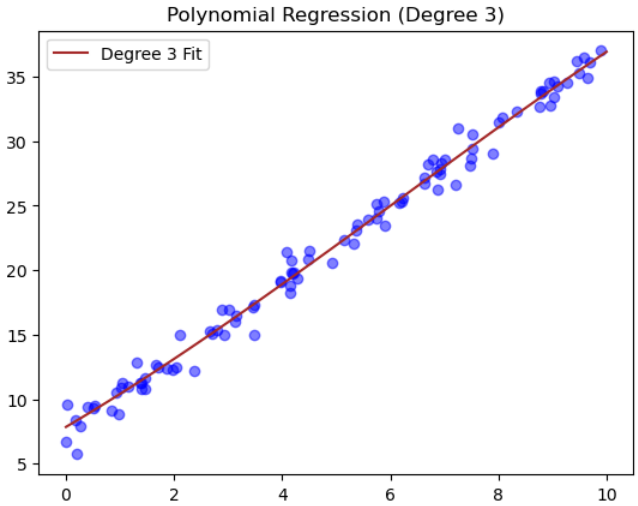

Figure 6: Best Fit Polynomial (degree = 3) Model

Conclusion: Higher-degree polynomials may fit the data better, but also risk overfitting if not regularized or validated properly.