Problem 1

1.

To address the first question, we manually calculated the first four moments of the data provided in problem1.csv using the normalized formulas from the Week 1 notes.

The mean is the expected value of the dataset. It represents the central point around which the data tends to cluster. The result is 1.049.

Variance measures the spread of data points around the mean, indicating the variability within the dataset. The result is 5.427, which indicates the spread of the data around the mean.

Skewness indicates the asymmetry of the data distribution around the mean. A skewness value greater than zero suggests a right-skewed distribution, while a value less than zero indicates a left-skew. The result is 0.880. The positive skewness shows that the distribution is skewed to the right, meaning the right tail is longer or fatter than the left.

Kurtosis measures the 'tailedness' of the distribution, providing insight into the data's propensity towards extreme values. The result is 23.096. This high kurtosis indicates that the distribution has heavy tails, suggesting a significant number of outliers compared to a normal distribution.

2.

For the second question, we utilized the scipy.stats package in Python, a robust tool for performing statistical calculations efficiently. We use the skew and kurtosis functions from the scipy.stats library to calculate these moments. Importantly, we set bias=False to obtain an unbiased estimate of these statistics, appropriate for sample data rather than population data. The results are 0.882 and 23.244 for skew and kurtosis respectively.

3.

To determine if the statistical package functions are biased, it's important to compare the results obtained from manual calculations and those derived from the statistical package. In this analysis, we will focus on skewness and kurtosis, as these are the metrics where bias is most commonly discussed and are particularly sensitive to sample size and distribution shape.

The hypothesis here is that the statistical functions provided by the package might exhibit bias. The typical expectation is that if the functions are unbiased, the results from manual calculations and the package calculations should be very close, if not identical, considering any minor discrepancies might be due to computational precision rather than systematic bias.

Based on the manual calculations and those derived from the scipy.stats package, here are the comparisons:

- Manual Calculation of Skewness: 0.88017
- Package Calculation of Skewness: 0.88193
- Manual Calculation of Kurtosis: 23.09608

- Package Calculation of Kurtosis: 23.24425

The differences between the manually calculated and package-derived values are Skewness difference: 0.00176 and Kurtosis difference: 0.14817. These differences are relatively small. In practical terms, such minor discrepancies are often considered negligible, especially in applied settings where such metrics are used for descriptive, rather than precise inferential, purposes. Thus, the hypothesis that the functions might be biased is not supported by this analysis.

Problem 2

1.

OLS Regression Results:
- Intercept (β_0): -0.0874
- Slope (β_1): 0.7753
- Standard Deviation of Residuals: 1.0038

MLE Results:
- Intercept (β_0): -0.0874
- Slope (β_1): 0.7753
- Estimated Standard Deviation (σ): 1.0038

Findings:

Intercept and Slope: The estimates for both the intercept and the slope are extremely close, with only minuscule numerical differences likely due to the computational precision and the algorithms used in the estimations. For all practical purposes, these differences are negligible and do not reflect a significant divergence in model interpretation.

Standard Deviation: Similarly, the standard deviation of the residuals from the OLS and the estimated standard deviation from the MLE are almost identical. The slight variance is, again, likely due to computational rounding and the inherent differences in how each method calculates these values.

Both methods aim to model the same underlying relationship in the data, and because they both assume normally distributed errors, they converge on similar estimates. The theoretical background supports that when the assumption of normally distributed errors holds true, OLS and MLE should yield consistent estimates, as seen here.

2.

MLE with Normal Distribution:

- Intercept (β_0): -0.08738447443979024
- Slope (β_1): 0.7752741738871116
- Scale (σ): 1.003756269318408

MLE with T-distribution:

- Intercept (β_0): -0.09726796971842995
- Slope (β_1): 0.6750102349179323
- Scale (σ): 0.8551027650098203
- Degrees of Freedom (ν): 7.159886081005919

The intercept and slope differ slightly between the two models. The T-distribution model has a lower intercept and a less steep slope. This suggests different interpretations of the relationship between x and y under different error distribution assumptions.

The standard deviation parameter for the normal distribution is higher than that for the T-distribution. A lower scale in the T-distribution model may indicate a tighter clustering of residuals around the regression line, possibly pointing to a better fit, especially if the data contains outliers or heavy tails.

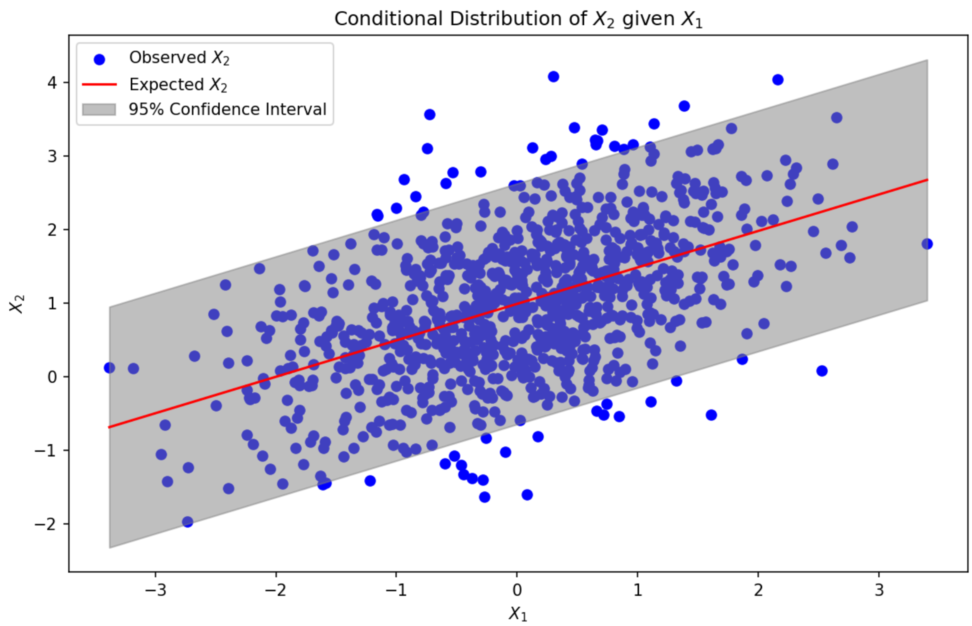
The log-likelihood values for the two MLE models are as follows:

- **MLE with Normal Distribution Errors**: Log-Likelihood = -284.54
- **MLE with T-Distribution Errors**: Log-Likelihood = -281.29

The MLE model assuming a T-distribution of errors has a higher log-likelihood (less negative) compared to the model assuming a normal distribution. This indicates that the T-distribution model fits the data better when only considering the likelihood of the observed data under each model, without taking model complexity into account.

This result suggests that the T-distribution model might be more appropriate for this dataset, potentially due to its ability to more effectively handle outliers or more accurately represent the underlying error structure in the data.

3.



The goal is to fit a multivariate normal distribution to data comprising two variables, X1 and X2, and to determine the conditional distributions of X2 given X1 for each observation. By assuming a multivariate normal distribution, we can utilize its properties to derive conditional means and variances. The conditional mean of X2 given X1 is calculated based on the overall means, variances, and covariance between X1 and X2. The conditional variance is determined from the variance and covariance terms, enabling predictions of X2 given X1 with quantified uncertainty.

For visualization and further analysis, we plot the expected values of X2 given X1 alongside the observed X2 values, incorporating 95% confidence intervals to visualize prediction accuracy and variability. This plot helps assess the predictive power and fit of the multivariate model, offering insights into the dependency and relationship between the two variables. Sorting the data by X1 ensures a coherent visual flow, aiding in better interpretation of how X2 varies with X1 within the given confidence bounds.

4.

The likelihood function for the model, given the normal distribution of ε, is:

$L(\beta, \sigma^2 \mid Y, X) = \prod (1 / \sqrt{(2\pi\sigma^2)}) \exp(- (y_i - x_i^T\beta)^2 / (2\sigma^2))$

where y_i is the ith observation and x_i^T is the ith row of X.

The log-likelihood function simplifies to:

$\log L(\beta, \sigma^2 \mid Y, X) = -(n/2) \log(2\pi\sigma^2) - (1/(2\sigma^2)) \Sigma (y_i - x_i^T\beta)^2$

where n is the number of observations.

Derivation of MLE for β and σ^2:

➢ Maximum Likelihood Estimator for β

The estimator for the coefficients β is derived by maximizing the log-likelihood function of the normal distribution of errors. The log-likelihood function is given by:

$\log L(\beta, \sigma^2 \mid Y, X) = -(n/2) \log(2\pi\sigma^2) - (1/(2\sigma^2)) \Sigma (y_i - x_i^T\beta)^2$

where n is the number of observations.

To find the MLE of β, the derivative of the log-likelihood function with respect to β is taken and set to zero:

$0 = -1/\sigma^2 X^T(Y - X\beta)$

$X^T Y = X^T X \beta$

Hence, the MLE for β, denoted as $\hat{\beta}$, is given by:

$\boldsymbol{\hat{\beta} = (X^T X)^{-1} X^T Y}$

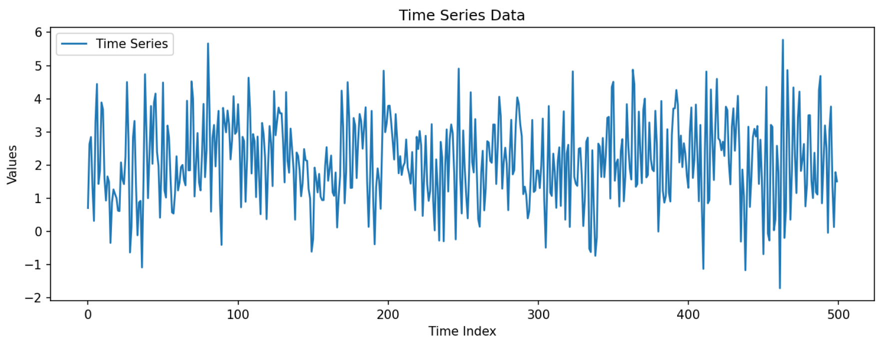
➢ Maximum Likelihood Estimator for σ^2

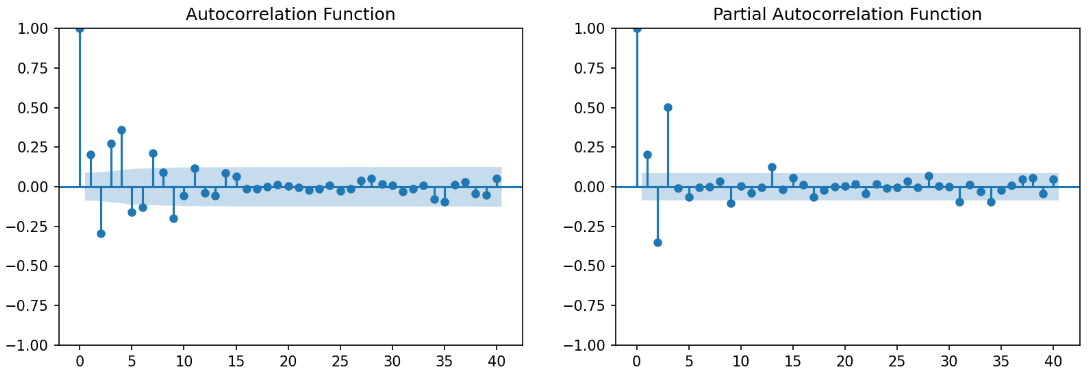
Using the estimator $\hat{\beta}$ derived above, we proceed to estimate σ^2. The estimator for σ^2 is derived by:

Taking the derivative of the log-likelihood with respect to σ^2 and setting it to zero, which simplifies to: $\boldsymbol{\hat{\sigma}^2 = (1/n) \Sigma (y_i - x_i^T\hat{\beta})^2}$

This expression gives the MLE for σ^2 based on the residuals of the regression.

Problem 3





ACF and PACF plots will help identify the possible order of AR or MA processes by observing how the autocorrelations and partial autocorrelations decay. The ACF plot shows a gradual decay. Also, the PACF plot shows a sharp cut-off after a lag 3, this typically indicates an AR process of order 3. Therefore, the AR(3) might fit the data best.

Here are the **AIC values** for each of the models fitted:

- AR(1): 1644.66
- AR(2): 1581.08
- AR(3): 1436.66
- MA(1): 1567.40
- MA(2): 1537.94
- MA(3): 1536.87

Based on the AIC, the AR(3) model has the lowest value (1436.66), suggesting it is the best fit among the models considered. This confirms hypothesis that the AR(3) model would fit the data best based on the initial examination of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Lower AIC values indicate a better fit model, taking into account both the model complexity and goodness of fit.