Problem 1

1. Mean: 1.04897
   Variance: 5.42722
   Skewness: 0.87929
   Kurtosis: 23.06998

2. Mean: 1.04897
   Variance: 5.42722
   Skewness: 0.88061
   Kurtosis: 23.12220

3. Unbiased

・The mean and variance are identical in both the manual and scipy calculations, showing that these functions are unbiased.

・ The skewness and kurtosis calculations show slight differences: skewness differs by about 0.0013, and kurtosis by about 0.05. These small discrepancies can be attributed to floating-point precision rather than any inherent bias in the statistical functions. Therefore, we can conclude that these functions are also unbiased, and the differences are negligible.

Therefore, the functions in scipy used to calculate the mean, variance, skewness, and kurtosis are statistically unbiased.

Problem 2

1.

OLS Results:
Intercept (beta0): -0.08738446
Slope (beta1): 0.7752741
Standard Deviation of OLS Residuals: 1.008813

MLE Results (under normality assumption):
Intercept (beta0): -0.08738446427005084
Slope (beta1): 0.7752740987226112
Standard Deviation of MLE: 1.003756

Comparison and Analysis:

・Regression coefficients (beta0 and beta1):

The intercept and slope obtained from OLS and MLE are nearly identical. The intercept values are -0.08738446 (OLS) and -0.08738446427005084 (MLE), and the slope values are 0.7752741 (OLS) and 0.7752740987226112 (MLE). This shows that, under the assumption of normally distributed errors, OLS and MLE provide almost identical results for the regression coefficients. The differences between the two methods are negligible.

・Standard deviation (sigma vs OLS residual standard deviation):

The standard deviation of the residuals from the OLS model is 1.008813, while the MLE estimated standard deviation is 1.003756. The two values are very close, with a difference of approximately 0.005. This indicates that both OLS and MLE provide almost the same error estimates under the normality assumption.

Conclusion:
OLS and MLE yield very similar estimates for the regression coefficients and the standard deviation. This similarity is expected because OLS and MLE are theoretically equivalent under the assumption of normally distributed errors. The small difference in the standard deviation values may be due to computational precision or implementation details, but overall, the two methods perform equivalently.

2.

	MLE (Normality)	MLE (t-distribution)	Difference
Intercept (beta0)	-0.08738446	-0.09726885	-0.00988439
Slope (beta1)	0.77527410	0.67500872	-0.10026538
Standard Deviation (sigma)	1.00375632	0.85510089	-0.14865543
Degrees of Freedom (nu)	N/A	7.15964031	N/A

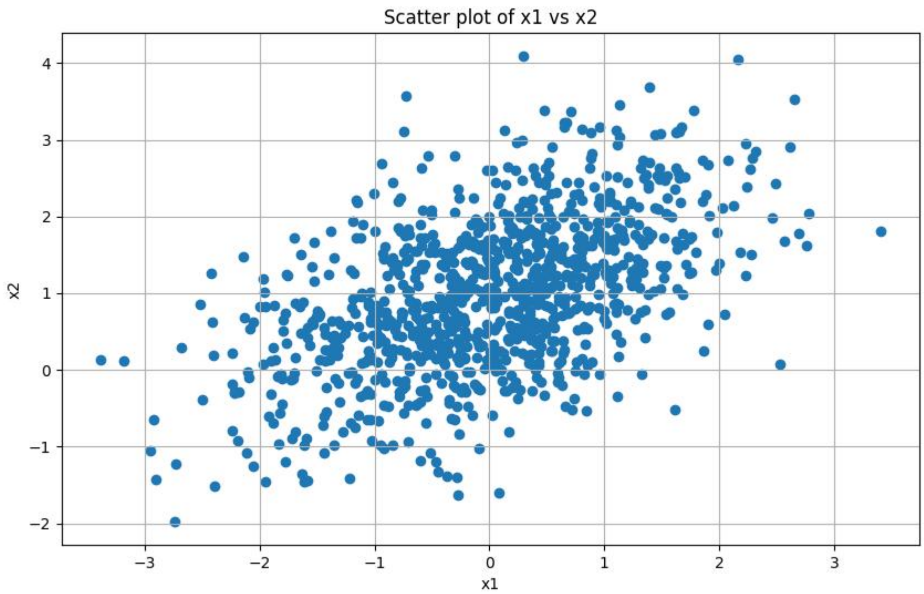
Comparison and Analysis:
・Intercept (beta0):
The intercept under the t-distribution is slightly smaller than under the normality assumption, with a difference of about -0.00988. This suggests that when we account for potential outliers (through the t-distribution), the intercept shifts slightly, but not drastically.
・Slope (beta1):
The slope under the t-distribution is 0.6750, compared to 0.7753 under the normality assumption. This indicates a larger deviation in the slope, showing that the t-distribution model, which is more robust to outliers, fits the data with a smaller weight on the independent variable (x). This suggests that under the t-distribution, the effect of x on y is slightly reduced, possibly due to the presence of outliers that are better accounted for by the t-distribution.
・Standard Deviation (sigma):
The standard deviation under the t-distribution is 0.8551, which is lower than the standard deviation under the normality assumption (1.0038). This reduction suggests that the t-distribution model, being more robust to outliers, can fit the data more tightly, resulting in smaller residuals and a lower overall error variance.
・Degrees of Freedom (nu):
The degrees of freedom, ν =7.16, indicate that the t-distribution has heavier tails than the normal distribution, but still not extremely heavy. A lower ν typically implies a distribution with thicker tails, meaning that the model can better account for extreme values (outliers) in the data.

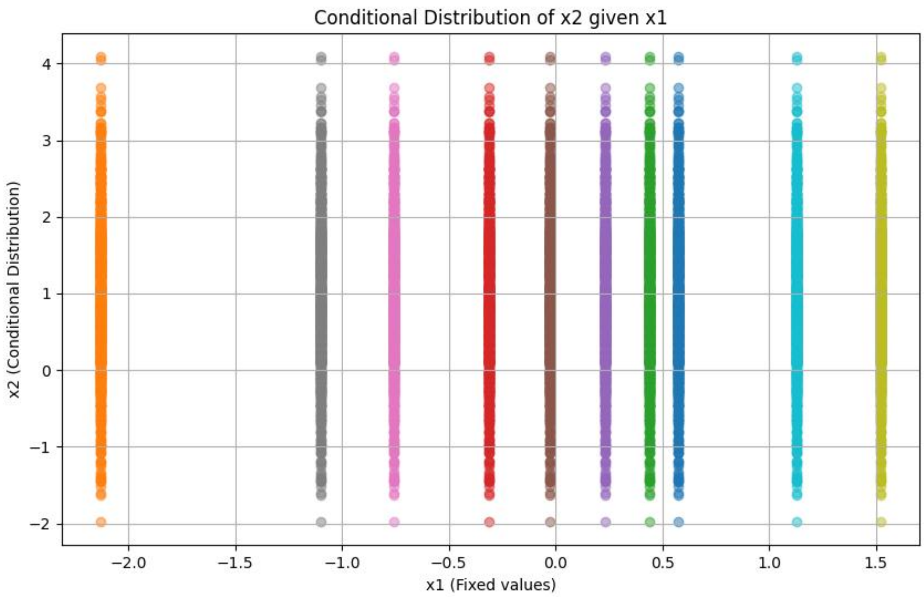
Conclusion:

The t-distribution results in a smaller intercept and a reduced slope compared to the normality assumption. The lower standard deviation (sigma) and the degrees of freedom (ν) indicate that the t-distribution is better at handling outliers, leading to a tighter fit with less variance in the residuals.

Differences between the two models suggest that there may be outliers or heavy-tailed characteristics in the data, which the t-distribution accounts for more effectively. Therefore, if the data exhibits outliers or is not perfectly normally distributed, the t-distribution provides a better model fit.

3.





4.

assume $\epsilon \sim N(0, \delta^2)$

$$f(Y \mid X, \beta, \delta^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\delta^2}} e^{\left(-\frac{(Y_i - X_i\beta)^2}{2\delta^2}\right)}$$

$$L(\beta, \delta^2 \mid Y) = \frac{1}{(2\pi\delta^2)^{\frac{n}{2}}} e^{\left(-\frac{1}{2\delta^2}(Y-X\beta)'(Y-X\beta)\right)}$$

$$\log L(\beta, \delta^2 \mid Y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\delta^2) - \frac{1}{2\delta^2}(Y-X\beta)'(Y-X\beta)$$

$$\frac{\partial}{\partial\beta}\log L(\beta, \delta^2 \mid Y) = -\frac{1}{\delta^2}X'(Y-X\beta)$$

$$X'(Y-X\beta) = 0$$

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'Y$$

$$\frac{\partial}{\partial\delta^2}\log L(\beta, \delta^2 \mid Y) = -\frac{n}{2\delta^2} + \frac{1}{2(\delta^2)^2}(Y-X\beta)'(Y-X\beta)$$

$$\frac{n}{2\delta^2} = \frac{(Y-X\beta)'(Y-X\beta)}{2(\delta^2)^2}$$

$$\delta^2_{MLE} = \frac{1}{n}(Y-X\hat{\beta})'(Y-X\hat{\beta})$$

$$\therefore \hat{\beta}_{MLE} = (X'X)^{-1}X'Y$$

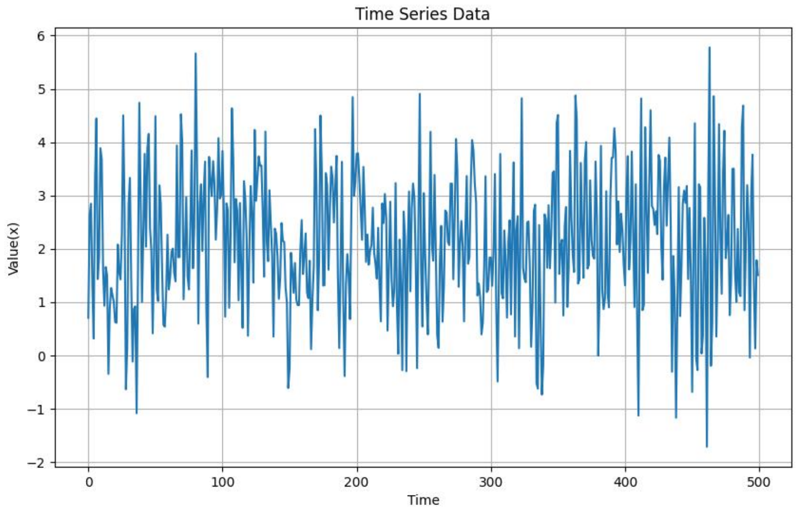
$$\hat{\delta}^2_{MLE} = \frac{1}{n}(Y-X\hat{\beta})'(Y-X\hat{\beta})$$

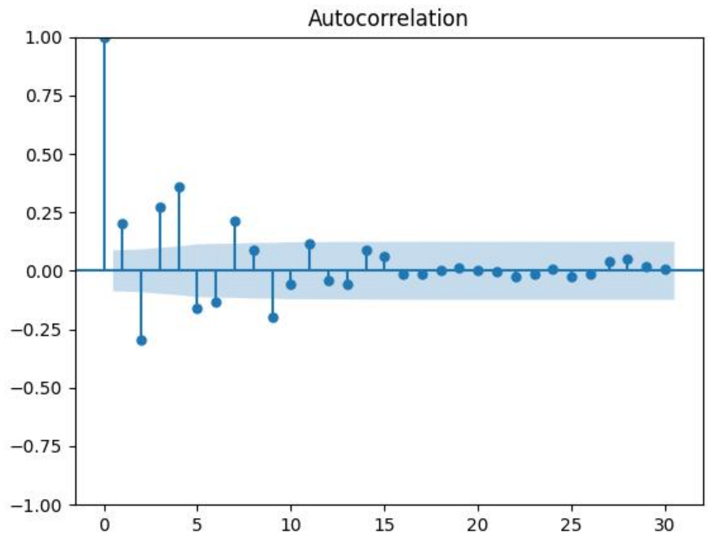
Problem3

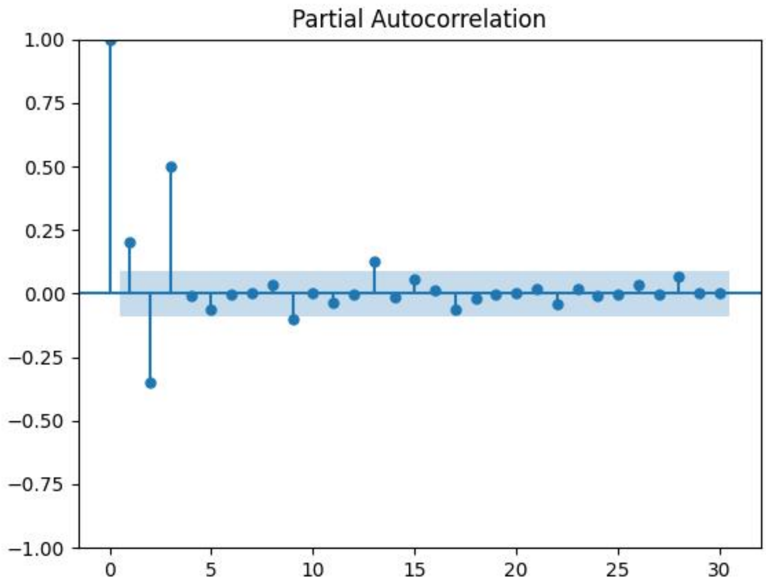
・ The first plot shows the time series data. The data fluctuates around a mean, suggesting potential auto regressive (AR) or moving average (MA) components.

・The second plot is the Auto correlation Function (ACF), showing how each value in the time series is correlated with previous values. There are significant auto correlations up to lag 3.

・ The third plot is the Partial Autocorrelation Function (PACF), which shows the correlation between values at specific lags, controlling for earlier lags. The PACF cuts off sharply after lag 3, indicating a possible AR(3) model might fit the data well.


Time Series Data
Value(x)
Time
6
5
4
3
2
1
0
−1
−2
0
100
200
300
400
500


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
25
30



Based on the ACF and PACF plots, both AR and MA models up to lag 3 were fitted. The AIC values for each model are as follows:

AR(1) AIC: 1641.09
AR(2) AIC: 1574.83
AR(3) AIC: 1428.26 (Lowest AIC among AR models)
MA(1) AIC: 1567.40
MA(2) AIC: 1537.94
MA(3) AIC: 1536.87 (Lowest AIC among MA models)

The AR(3) model provides the best fit to the data, based on both the AIC values and the insights from the ACF and PACF plots. This confirms the hypothesis that an auto regressive model, particularly AR(3), would be the most appropriate model for this dataset.