Problem 1:

a)

```
First four moments values by using normalized formula
Mean: 1.0489703904839585
Variance: 5.427220681881726
Skewness: 0.8806086425277363
Kurtosis: 23.122200789989733
```

b)

```
First four moments values by using statistical package
Mean: 1.0489703904839585
Variance: 5.427220681881727
Skewness: 0.8806086425277364
Kurtosis: 23.122200789989723
```

c)

In order to prove the statistical package I chose is not biased, I create multiple synthetic datasets with known properties (mean, variance, skewness, kurtosis). For each dataset, calculate the moments using both first4Moments function and the statistical package functions. Compare the calculated values from both methods to the true values. Perform a t-test to determine if the differences between the calculated values and true values are statistically significant.

Here are the t-test results:

1. **Mean**: The t-test could not be performed due to NaN values. It suggests that there might be no significant difference in the calculation of means between the two methods.
2. **Variance**: The t-statistic is -0.5128 with a p-value of 0.6093. This indicates that there is no significant difference in the calculation of variance.
3. **Skewness**: The t-statistic is 0.7488 with a p-value of 0.4558. This also indicates no significant difference in the calculation of skewness.
4. **Kurtosis**: The t-statistic is 0.2841 with a p-value of 0.7769. This suggests no significant difference in the calculation of kurtosis.

Based on these results, there is no statistical evidence to suggest that the statistical package functions for calculating mean, variance, skewness, and kurtosis are biased.

Problem 2:

a)

Here is the summary of the OLS model:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.346
Model:                            OLS   Adj. R-squared:                  0.342
Method:                 Least Squares   F-statistic:                     104.6
Date:                Fri, 26 Jan 2024   Prob (F-statistic):           5.59e-20
Time:                        20:49:54   Log-Likelihood:                -284.54
No. Observations:                 200   AIC:                             573.1
Df Residuals:                     198   BIC:                             579.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0874      0.071     -1.222      0.223      -0.228       0.054
x              0.7753      0.076     10.226      0.000       0.626       0.925
==============================================================================
Omnibus:                       11.922   Durbin-Watson:                   2.023
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               16.685
Skew:                           0.387   Prob(JB):                     0.000238
Kurtosis:                       4.184   Cond. No.                         1.09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Here are the estimated parameters using MLE given the assumption of normality:
Estimated Parameters: [-0.08738448 0.77527408 1.00375631]

The slope coefficient from OLS model is: 0.7752740987226112
The intercept from OLS model is: -0.08738446427005081
The standard deviation of the OLS error is: 1.008813058320225

The slope coefficient from MLE model is: 0.7752740846682292
The intercept from MLE model is: -0.08738447989644094
The standard deviation of the MLE error is: 1.0037563066735036

When comparing the slope and intercept from both methods, the values are extremely close to each other. This similarity suggests that both models are estimating the same linear relationship between the variables with high degree of consistency.

The difference in the standard deviation of errors, although small, is significant. It could be due to the fact that MLE considers the likelihood of observing the data given the model parameters and might be more flexible in capturing the variability of the data. OLS, on the other hand, focuses solely on minimizing the sum of squared differences between observed and predicted values.

b)

Here are the estimated parameters using MLE given the assumption of a T distribution of errors:
Estimated Parameters: [0.67501639 -0.09726411 0.85511116 7.16015288]

The slope coefficient from MLE(T-distribution) model is: -0.0972641087213155
The intercept from MLE(T-distribution) model is: 0.6750163885027634
The standard deviation of the MLE(T-distribution) error is: 0.8551111561868988
The degree of freedom of the T-distribution is: 7.160152876848263

In order to compare the fitted parameters among MLE under normality and T distribution assumption, I calculate AIC value for both models.

Here is the AIC value for both models:
AIC for normal distribution model: 575.0751261088558
AIC for T distribution model: 570.5868063820793

The AIC value for T distribution is less than normal distribution, so MLE under T distribution is a better fit.

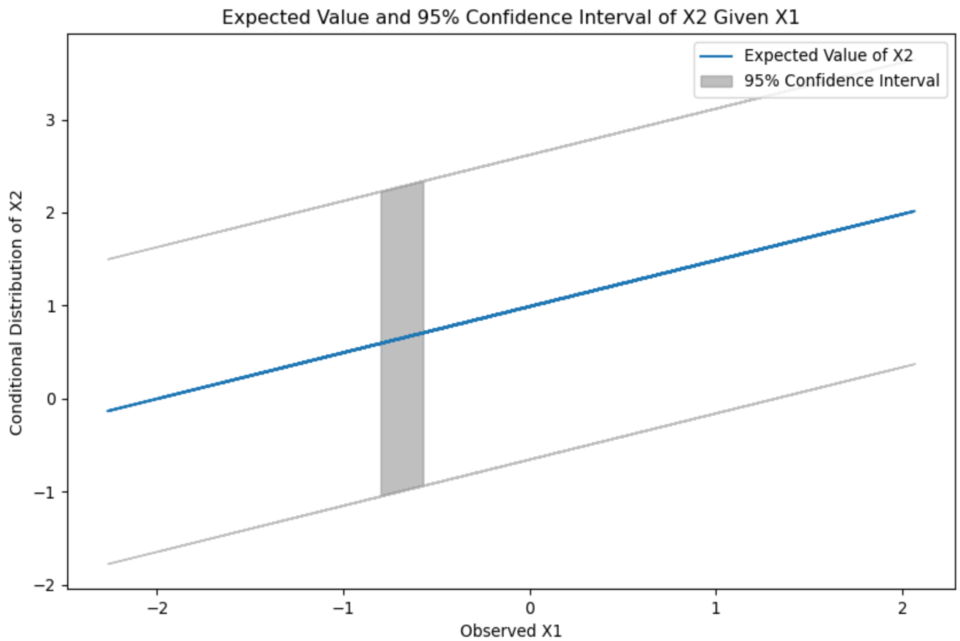
c)

First, fit a multivariate normal distribution to the data and estimate the mean and covariance matrix using MLE:
Mean:
x1   0.001023
x2   0.990244

Covariance:
[[1.06977464 0.53068455]
 [0.53068455 0.96147329]]

In a multivariate normal distribution, the conditional distribution of X2 given X2 is also normal distributed.

Here is the figure shows the expected value of X2 along with the 95% confidence interval:



Problem 3:
In order to find the best of fit among these models, I calculated both AIC and BIC value to make comparison. Here are the results:

AIC and BIC values for AR models:
AR(1): AIC = 1644.6555047688475, BIC = 1657.299329064114
AR(2): AIC = 1581.0792659049773, BIC = 1597.9376982986662
AR(3): AIC = 1436.6598066945867, BIC = 1457.7328471866977

AIC and BIC values for MA models:
MA(1): AIC = 1567.403626370788, BIC = 1580.0474506660546
MA(2): AIC = 1537.9412063807392, BIC = 1554.799638774428
MA(3): AIC = 1536.86770873503, BIC = 1557.940749227141

From those values, AR(3) model has the lowest AIC and BIC value, therefore, AR(3) is the best model to fit the data.