Problem 1 (3pts)

Given the dataset in problem1.csv

1. Calculate the first 4 moment values using the normalized formulas in the Week 1 notes.

mean 1.04897
variance 5.427221
skew 0.883257
kurt 26.049938

2. Calculate the first 4 moment values using your chosen statistical package.

mean 1.04897
variance 5.427221
skew 0.881932
kurt 23.244253

3. Are your statistical package functions biased? Prove or disprove your hypotheses. Explain your conclusion.

- Mean is not biased, because expectation of sample mean is always equal to population mean
- Variance is not biased, because pandas use 1/n-1 to calculate variance, when we substitute the population mean into sample mean when calculating the variance, the expectation of variance will be smaller than when using population mean, so 1/n-1 instead of 1/n is a correctness of underestimate of variance.
- For skew and kurt, pandas in python use standard way to calculate them, so they are biased.

Problem 2 (5pts)

Assume the multiple linear regression model $Y = X\boldsymbol{\beta} + \epsilon$

1. Fit the data in problem2.csv using OLS. Then fit the data using MLE given the assumption of normality. Compare the beta values and the standard deviation of the OLS errors to the fitted MLE σ. What is your finding? Explain any differences.
2. Fit the data in problem2.csv using MLE given the assumption of a T distribution of errors. Show the fitted parameters. Compare the fitted parameters among the MLE under the normality assumption and T distribution assumption. Which is the best fit?
3. Fit a multivariate distribution to the data in problem2_x.csv. Given the values of X_1 what are the conditional distributions for X_2 for each observation. Plot the expected value along with the 95% confidence interval and the observed value.
4. (1 point Extra Credit). $Y = X\boldsymbol{\beta} + \epsilon \ and \ \epsilon \sim N(0, \sigma^2)$. Derive the maximum likelihood estimators for β and σ^2

1. Beta0,beta1,std

OLS:-0.08738446427005074 0.7752740987226111 1.0037563194177317

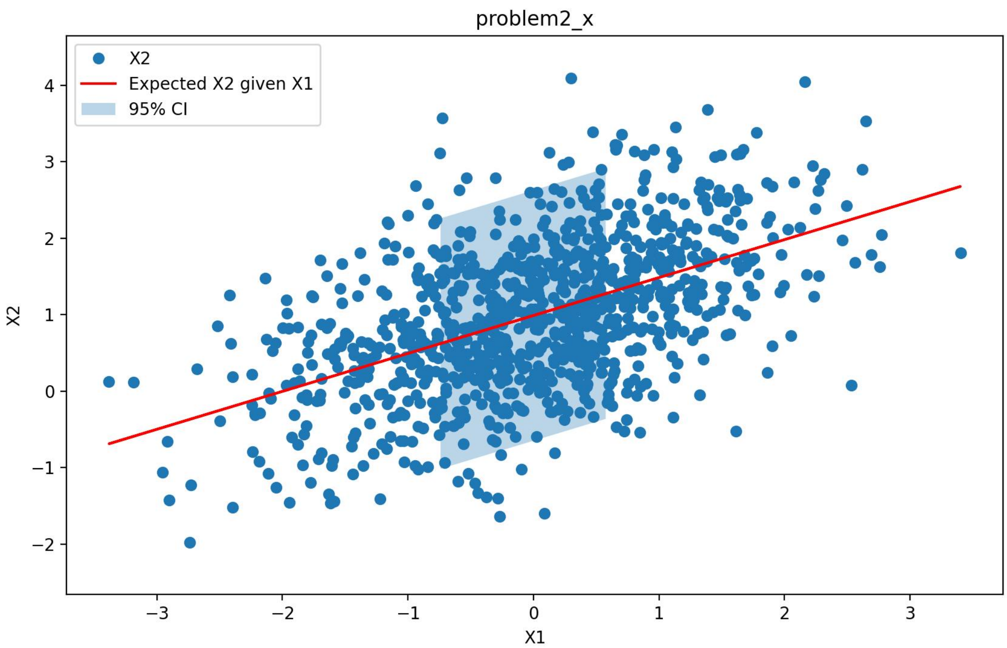
MLE: -0.08738450937788952 0.7752745345857192 1.0037562235144133

The beta and std are very similar, because in normal distribution assumption, the linear regression for MLE and OLS are essentially the same. If we put mathematical deviration here for MLE, we can find the L(beta,std|X,Y) after turning into log, we do dL/d beta = 0 to try to get beta formula, we can find it's the same as differentiate the RSS by using OLS.

2.
Beta0,beta1,std for this mle:
-0.09551509336364047 0.651442165303442 0.8132730960973427

R^2 for normal and t distribution mle:
0.34560688356470337 0.3366233202285491
We can see normal mle R^2 is better, so it fit better

3.



4. When error is normal distribution, the MLE beta has the same solution with OLS, so

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

beta = , the deriviation process is create L(beta,sigma|X,Y), and time log on both side, and solve dL/d beta =0 as an convex optimization.

Problem 3 (2pts)

Examine the data in problem3.csv; which AR(n) or MA(n) model do you expect to fit this data

best? Fit the data using AR(1) - AR(3) and MA(1) - MA(3) models. Which is the best fit and does this confirm your hypothesis?

We can see AR 3 has lowest AIC score, so AR(1) to AR(3) is best fit

```
{'AR(1)': 1644.6555047688473, 'AR(2)': 1581.0792659049778,
'AR(3)': 1436.6598066945862, 'MA(1)': 1567.4036263707894,
'MA(2)': 1537.94120638074, 'MA(3)': 1536.8677087350327}
```