Question 1:

1.

Result:

```
mean:  1.0489703904839585
variance:  5.421793461199845
skewness 0.8806086425277365
kurtosis: 26.122200789989737
```

2.

Result:

```
mean:  1.0489703904839585
variance:  5.421793461199845
skewness 0.8819320922598397
kurtosis: 23.244253469616208
```

3.

Null hypothesis: The results of skewness and kurtosis calculated by the package are unbiased

Alternative hypothesis: The results of skewness and kurtosis calculated by the package are biased

Significance level: 0.05

Result:

```
t test for skewness:  0.6533838631912604
t test for kurtosis:  1.7776760055644436
P value for skewness:  0.5136591876902288
P value for kurtosis:  0.07576123143828475
```

According to the p-values. We should reject the null hypothesis since both are higher than the significance level (especially P-value for skewness is significantly higher). That means that the results we get are unbiased. This makes sense to us since the dataset in problem1.csv contains large samples. The larger the sample is the less biased the results are.

Question 2:

1.

Result from OLS:

```
Beta: 0.7752740987226111
Intercept: -0.08738446427005074
Standard Error: 1.003756319417732
```

Result from MLE (assumed Normal Distribution):

```
Beta: 0.7752740987226111
Intercept: -0.08738446427005074
Standard Error: 1.003756319417732
```

Findings:

These two models generate the exact same Beta, Intercept and Standard Error. The reason for that is because linear regression assumes normally distributed errors. Under the setting provided by linear regression, though MLE maximizes likelihood while OLS minimizes the sum, they provide the same results.


```
Beta: 0.7752740987226111
Intercept: -0.08738446427005074
Standard Error: 1.003756319417732
```

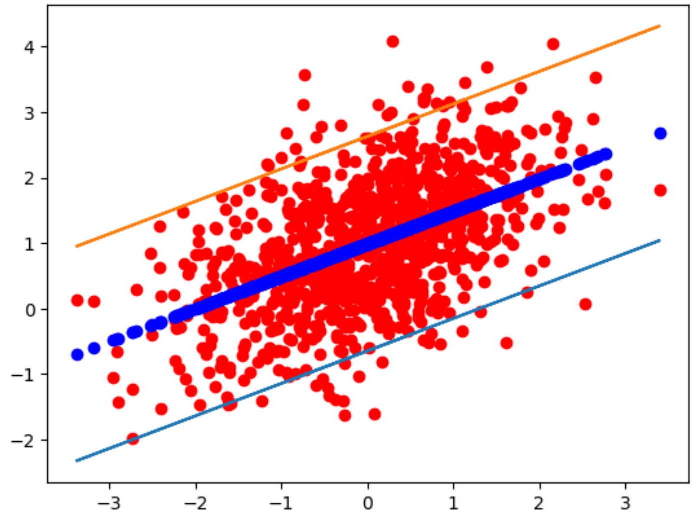

2.

Result from MLE (assumed t-distribution)

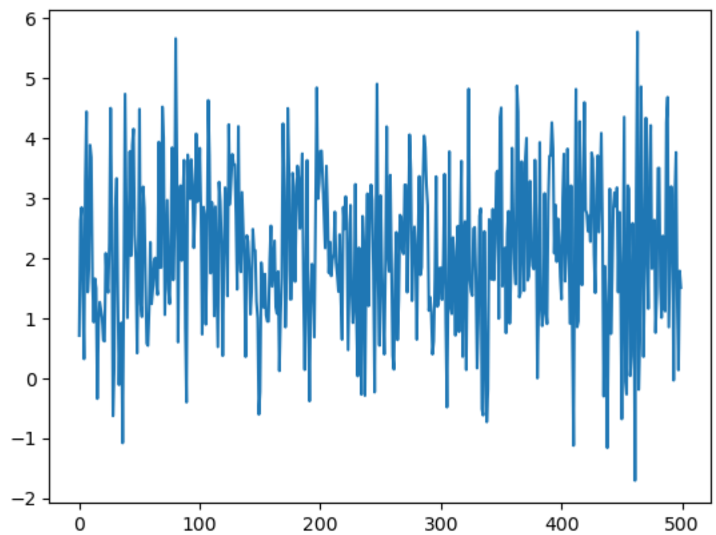
```
Beta:  0.6696415912961888
Intercept:  -0.0969499916785006
Standard error:  0.8522056148535601
Degree of freedom:  6.949450065580704
```

Findings: by comparing, one can find that t-distribution provides slightly lower Beta, higher Intercept and lower standard error. Besides, its degree of freedom is close to 7. Since t distribution is built to get rid of the effect from a fatter tail. As a result, the stand error of residual is lower in t distribution. Who is the best fit depending on the dataset we are using. If the distribution is close to normal distribution we can fit it with MLE assumed normalization. If the distribution has a fatter tail, we can fit it with MLE assumed t-distribution.

3.



Question 3:



I plotted the graph for the data in problem3.csv. Since the graph does not show any trend or recurring pattern. My initial guess on this problem is that MA model could probably be a better fit since it was good at processing and fitting random data.

```
MA(1) AIC: 1567.4036263707894
MA(1) AIC: 1537.94120638074
MA(1) AIC: 1536.8677087350327
```

```
AR(1) AIC: 1644.6555047688473
AR(2) AIC: 1581.0792659049778
AR(3) AIC: 1436.6598066945862
```

However, after running the code, I found that AR (3) model got the lowest AIC, which meant that it was the best fit for the data. This was totally contradictory to my hypothesis. Thus, we can conclude that it is hard to see from raw data if current value depends on the past value (AR Model).