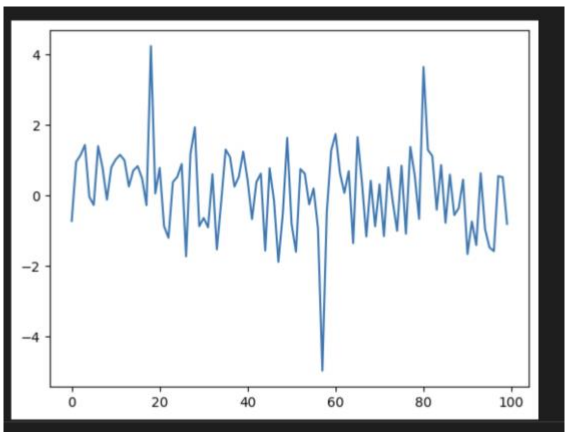
Problem 1:
My conclusion is that both my functions and packages will be biased due to small sample size. Firstly, I use package to test on the original data, and I normalized the data and use the same package to calculate again, I got the same answers, which means the package is accurate. However, when I randomly chose 30 out of 100 samples to calculate the skewness and kurtosis, there some difference between the data calculated previously. After that, I applied the same method on my own written function, and I got the same result, which was that there were difference in results between sample and original data.
Therefore, I believe that the bias come from the limited sample size, but if we have enough sample sizes, the bias should be gone.

Problem 2:
Firstly, I fit the data using OLS model, and I plot the distribution of error,



As we can see, except some outlines in 20, 60 and 80, most of the errors are within -2 and 2, which means they are identical and independent distributed. The the p value is smaller than 0.05, so we can reject the null hypothesis, and say the x is a significant variable

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                      y   R-squared (uncentered):                   0.193
Model:                            OLS   Adj. R-squared (uncentered):              0.185
Method:                 Least Squares   F-statistic:                              23.69
Date:                Fri, 27 Jan 2023   Prob (F-statistic):                    4.28e-06
Time:                        23:34:19   Log-Likelihood:                         -160.49
No. Observations:                 100   AIC:                                      323.0
Df Residuals:                      99   BIC:                                      325.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x              0.6052      0.124      4.867      0.000       0.358       0.852
==============================================================================
Omnibus:                       14.146   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               43.674
Skew:                          -0.267   Prob(JB):                     3.28e-10
Kurtosis:                       6.193   Cond. No.                         1.00
==============================================================================
```

Then I use MLE on the errors, which are also the residuals. The log-likehood for normal distribution is -159.99209669526914 and the log-likehood of T distribution is

159.29721858186196, which are very similar, because if the sample size is large, the T-distribution will be like the normal distribution. Usually, the larger LL is, the better the fit is. So in this case, T distribution perform a little better.

The parameter of normal distribution is mu and sigma. But the parameter of T distribution are mu, sigma and degree of freedom. The normal distribution has a fixed degree of freedom of infinity, whereas the T-distribution allows for a finite degree of freedom. The formula for the log-likelihood function will also be different, as it will include the probability density function (pdf) of the T-distribution, which is different from the pdf of the normal distribution.
Therefore, in this case, we can find that the T-distribution is more robust to outliers than the normal distribution, because it has heavier tails. This means that the probability of observing extreme values (outliers) is higher for the T-distribution compared to the normal distribution.

Problem 3:

The ACF is a measure of the correlation between a time series and a lagged version of itself. The PACF is a measure of the correlation between a time series and a lagged version of itself, controlling for the correlation of the time series with all of the intermediate lags.
In an MA process, the ACF will cut off after q lags, while the PACF will die out relatively quickly.
In an AR process, the ACF will die out relatively quickly, while the PACF will cut off after p lags.
This is the main difference between the ACF and PACF for MA and AR process.
If the PACF cuts off after lag p, then it is likely an AR process. If the ACF cuts off after lag q, then it is likely an MA process.