Project ---- Week 02
Tianai Zhao

Problem 1

Multivariate Normal (2 variables): $\mathbb{E}(Y|X=x) = \mu_Y + \frac{Cov(X,Y)}{Var(X)}(x-\mu_X)$

OLS: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$$\mathbb{E}(\hat{Y}|X=x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

where $\begin{cases} \hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)} \\ \hat{\beta}_0 = \mu_Y - \hat{\beta}_1 \mu_X \end{cases}$

$$\therefore \mathbb{E}(\hat{Y}|X=x) = \mu_Y + \frac{Cov(X,Y)}{Var(X)}(x-\mu_X)$$

is same as the conditional mean in multivariant normal

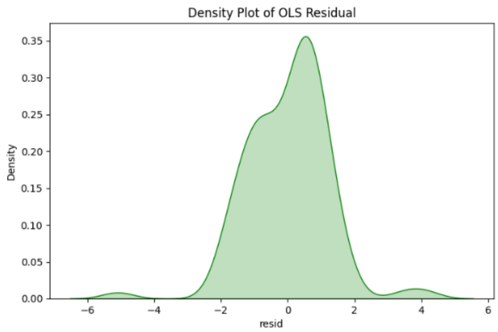
The values of the two approaches are the same.

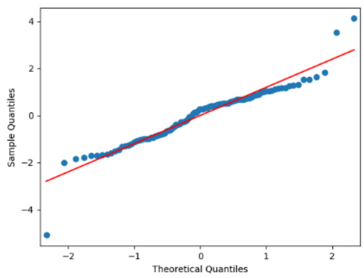
To see this empirically, first calculate the conditional mean of Y given X, implied by the multivariant normal distribution, which is $\mu_Y + \frac{Cov(X,Y)}{Var(X)}(x-\mu_X)$.
Then fit the data into OLS model and use the estimated parameters to calculate the predicted Y-hat. This would be the conditional value suggested by the OLS equation.
Ignoring rounding errors, the conditional mean and Y-hat are the same (the array holding differences in these two is empty).

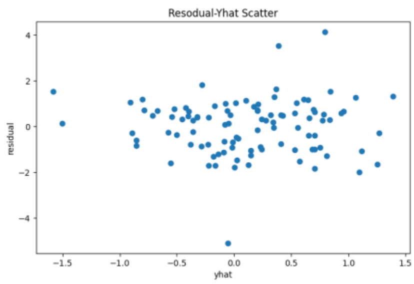
```
>>> print(df.loc[lambda x:x['yhat_multinorm']>x['yhat_OLS']])
Empty DataFrame
Columns: [x, y, yhat_multinorm, yhat_OLS]
Index: []
```

Problem 2

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.195
Model:                            OLS   Adj. R-squared:                  0.186
Method:                 Least Squares   F-statistic:                     23.68
Date:                Thu, 13 Jan 2022   Prob (F-statistic):           4.34e-06
Time:                        12:11:01   Log-Likelihood:                -159.99
No. Observations:                 100   AIC:                             324.0
Df Residuals:                      98   BIC:                             329.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1198      0.121      0.990      0.325      -0.120       0.360
x1             0.6052      0.124      4.867      0.000       0.358       0.852
==============================================================================
Omnibus:                       14.146   Durbin-Watson:                   1.885
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               43.673
Skew:                          -0.267   Prob(JB):                     3.28e-10
Kurtosis:                       6.193   Cond. No.                         1.03
==============================================================================
```







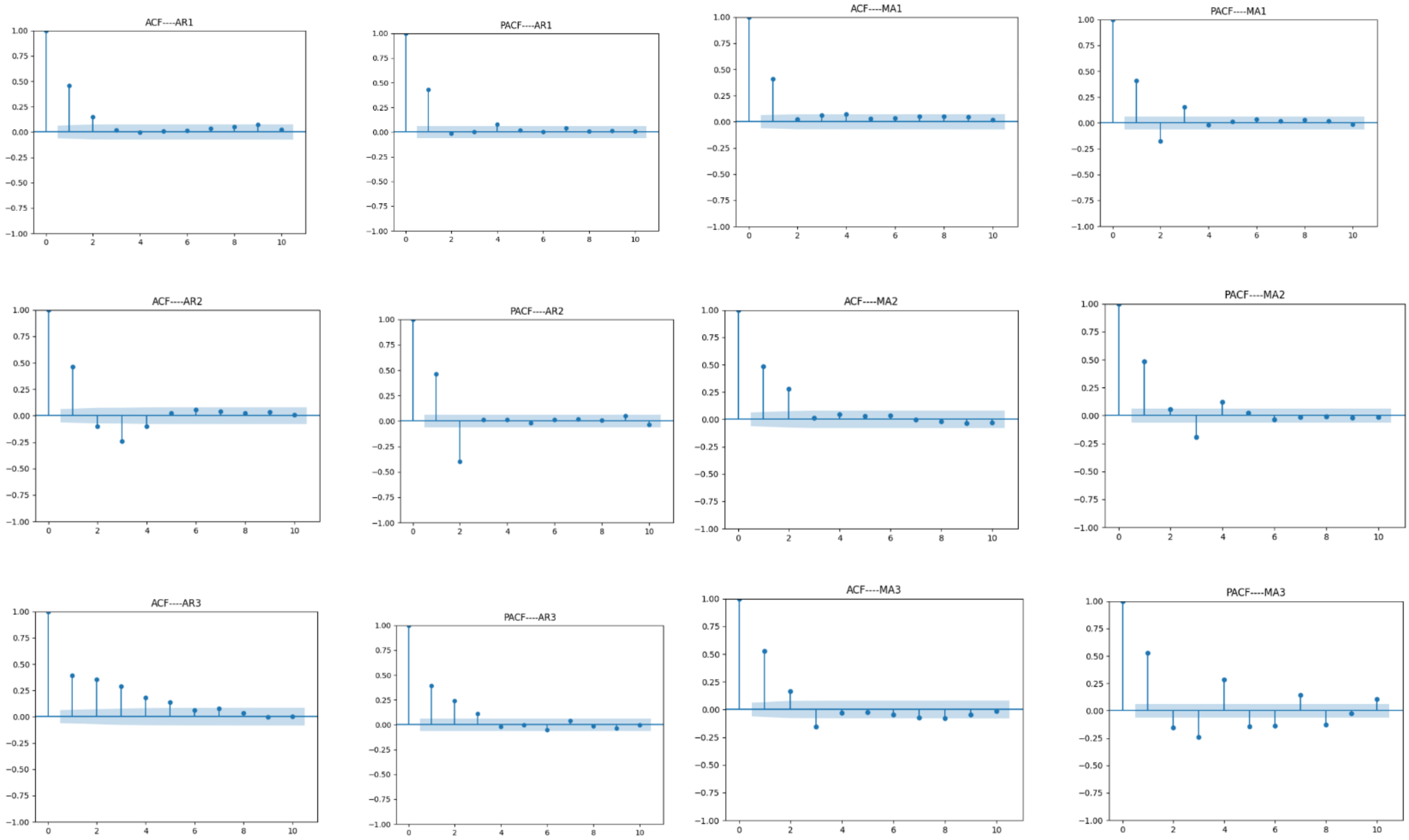
By looking at the distribution of the OLS residual, and its Q-Q plot, the error term is not normally distributed, especially in the tails.

```
OLS:  [0.1198362  0.60520482]
MLE-Assume normality: [0.11983621 0.6052048 ] Log Likelihood:  -163.70127759078036
MLE-Assume T distribution:  [0.12325311 0.59512445] Log Likelihood:  -161.75964926470715
```

Comparing the resulted value of the likelihood functions from the two MLE models with different assumptions, the one assuming T distribution of the residual is a better fit.

The OLS model gives the same parameters to those from the MLE where we assume normality, violation of the normality assumption in OLS will not affect the unbiasedness of estimated parameters, but it is not the maximum likelihood estimator.

Problem 3



A decaying ACF graph and a PACF graph with a clear cut off signature a AR process, while the opposite for a MA process.

AR(P) process will have a cut off after p lags in the PACF graph, and the MA(q) process will have a cut off after q lags in its ACF graph.

For identifying processes, first check if one of the PACF and ACF graphs has a cut off and the other is decaying. This tells us if the process is AR or MA. Then position of the cut off gives us the order. If the graph shows a mix of these two, the process is probably an ARIMA process.