1. Let S_t be the sales at time t for the gadget IG-11, with the data in **Item2.csv**. Using $X_t = \ln S_t$, do the following in **R**.

   (a) Is the data stationary? Use appropriate test(s) to justify your answer. If the data is not stationary, perform the necessary transformation to make the data stationary. Verify that the transformed data is indeed stationary.

   (b) Let Y_t be the transformed data in (a). Is there serial correlation? Use appropriate test(s) to justify your answer.

   (c) Generate correlograms of the ACF and PACF of the transformed data Y_t, and use these to identify one possible AR(p), one MA(q), and one ARMA(p, q) model, where p and q are both less than or equal to 5. Justify your choices.

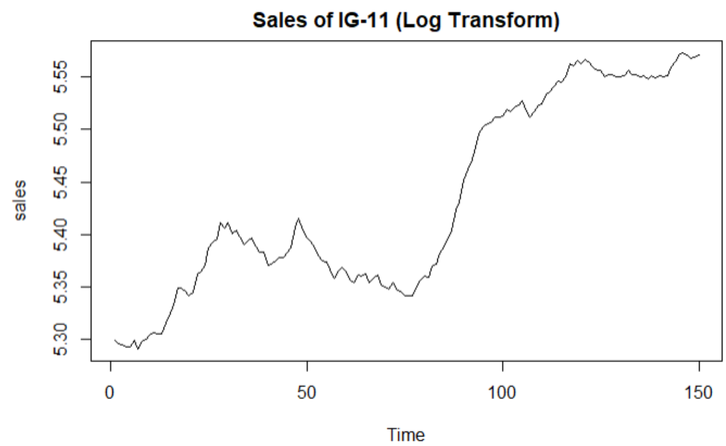
   (d) For each of these models, write down the resulting models in functional form.

   (e) Given only the three models in (c), which should you choose? Explain.

<u>**Solution**</u>.

All the R codes can be found in the attached R notebook.

The data from the csv file was first loaded into the R notebook. The original data was log-transformed according to the given instructions.



(a) We do the Augmented Dickey-Fuller Test to test for stationarity.

```
	Augmented Dickey-Fuller Test

data:  X_t
Dickey-Fuller = -2.1717, Lag order = 5, p-value = 0.5048
alternative hypothesis: stationary
```

Figure 1: ADF test on the log-transformed data

For the log-transformed data, we obtained a p-value of 0.5048, which is greater than 0.05. This means that we don't reject the null hypothesis of the ADF test that the time series is non-stationary. Hence, we perform differencing.

```
        Augmented Dickey-Fuller Test

data:  Y_t
Dickey-Fuller = -3.3453, Lag order = 5, p-value = 0.06638
alternative hypothesis: stationary
```

Figure 2: ADF test on the differenced data

After differencing, we obtained a p-value of 0.06638, which is still greater than 0.05. Thus, we need to perform differencing again.

```
Warning: p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  Y_t2
Dickey-Fuller = -6.5965, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Figure 3: ADF test on the twice differenced data

Finally, after differencing twice, we obtained a p-value of less than 0.01 on the ADF test, meaning that we reject the null hypothesis that the time series contains a unit root. We conclude that the time series for the log-transformed data twice differenced is stationary.

(b) We use the LJung Box Test to test for serial correlation or autocorrelation. The null hypothesis for the LJung Box Test is that all the autocorrelations of the time series are equal to zero.

```
        Box-Ljung test

data:  Y_t2
X-squared = 33.806, df = 1, p-value = 6.09e-09
```

Figure 4: LJung Box Test

Since the p-value is less than 0.05, we reject the null hypothesis. We conclude that the time series exhibits serial correlation.

(c) The following are the ACF and PACF correlograms generated using Rstudio.

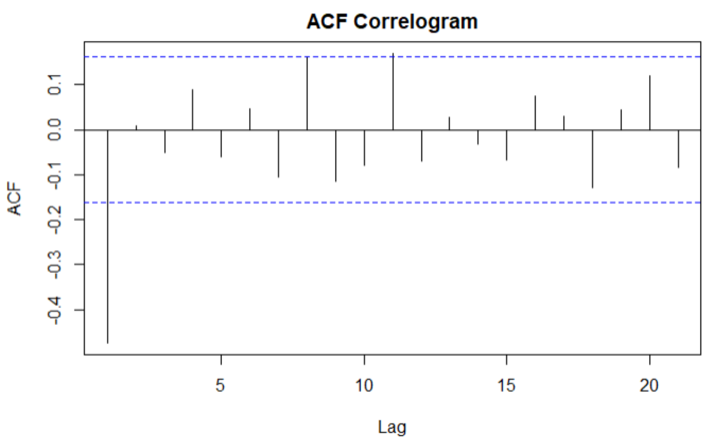


Figure 5: ACF Correlogram

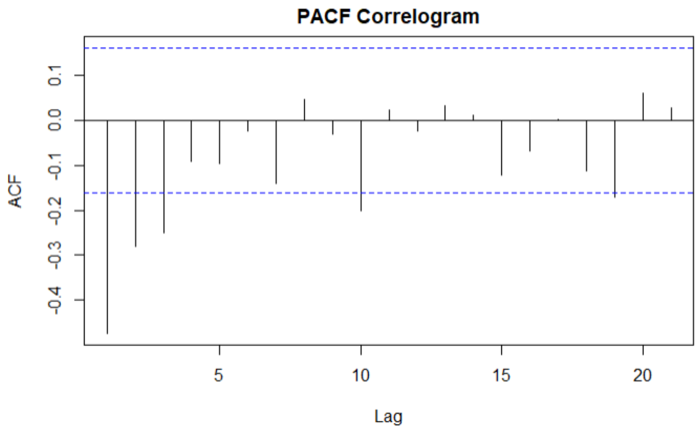


Figure 6: PACF Correlogram

Looking at the ACF Correlogram, the ACF cuts off at lag 1. Thus, we select MA(1) as one of the candidate models. On the other hand, looking at the PACF Correlogram, the PACF cuts off at lag 3. Thus, we select AR(3). We will also test ARMA(3,1).

(d) The following are the models in functional form:

i. AR(3): $Y_t = -0.6728Y_{t-1} - 0.4286Y_{t-2} - 0.2485Y_{t-3} + Z_t$

ii. MA(1): $Y_t = -0.7512Z_{t-1} + Z_t$

iii. ARMA(3,1): $Y_t = 0.2340Y_{t-1} + 0.1734Y_{t-2} + 0.1176Y_{t-3} - Z_{t-1} + Zt$

Note that Z_t is white noise, $Z_t \sim WN(0, \sigma^2)$.

(e) We evaluate the models using the Akaike Information Criterion (AIC). The model with the lowest AIC is the best-fitting model.

Model	AIC
AR(3)	-1077.74
MA(1)	-1085.74
ARMA(3,1)	-1083.21

Table 1: AIC Values for Candidate Models

We select MA(1) as the best-fitting model since it has the lowest AIC.