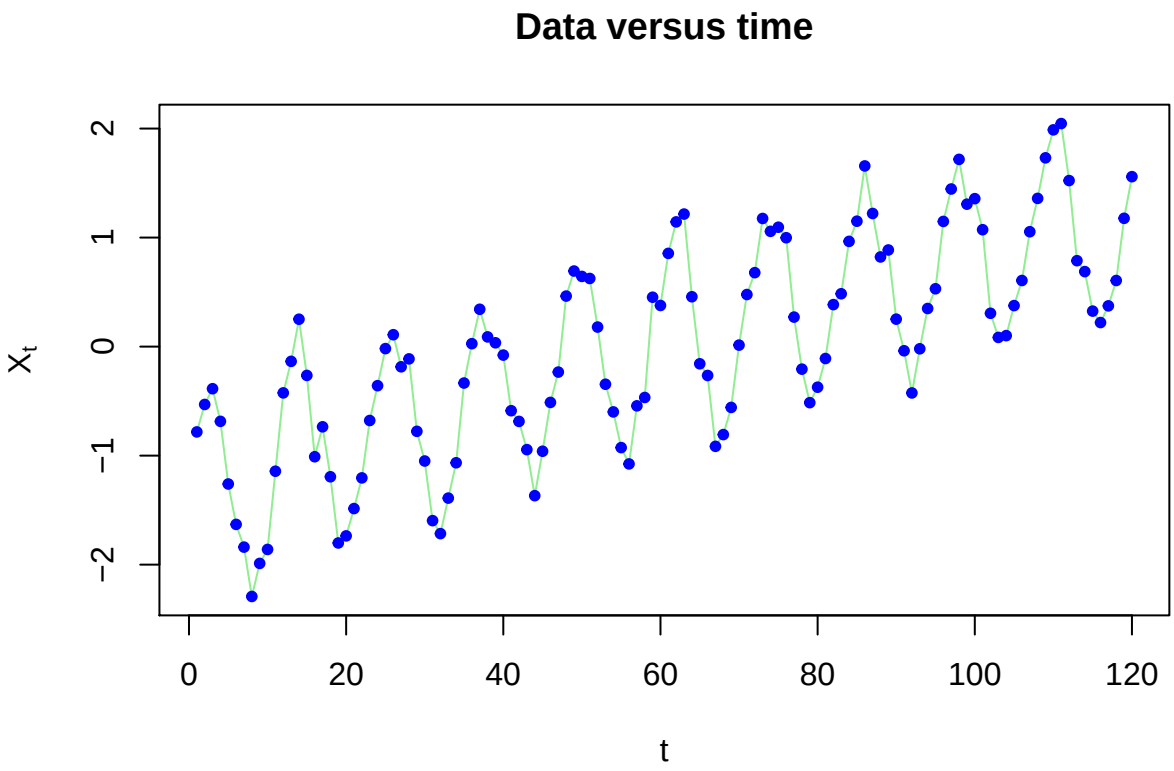
1.

```
data = read.table('project5_data.txt')
plot(data$V1, type = 'l', col = 'lightgreen',
     xlab = 't', ylab = expression(X[t]),
     main = 'Data versus time')
points(data$V1, pch = 20, col = 'blue')
```



We observe an upward trend, and seasonality on a 12 month interval.

2. Least square estimation

(a)

```
t = 1:120
trigest = lm(data$V1 ~ t + cos(pi*t/6) + sin(pi*t/6))
coef(trigest)
```

```
##   (Intercept)             t cos(pi * t/6) sin(pi * t/6)
##   -1.29642763    0.02142856    0.48893652    0.74596584
```

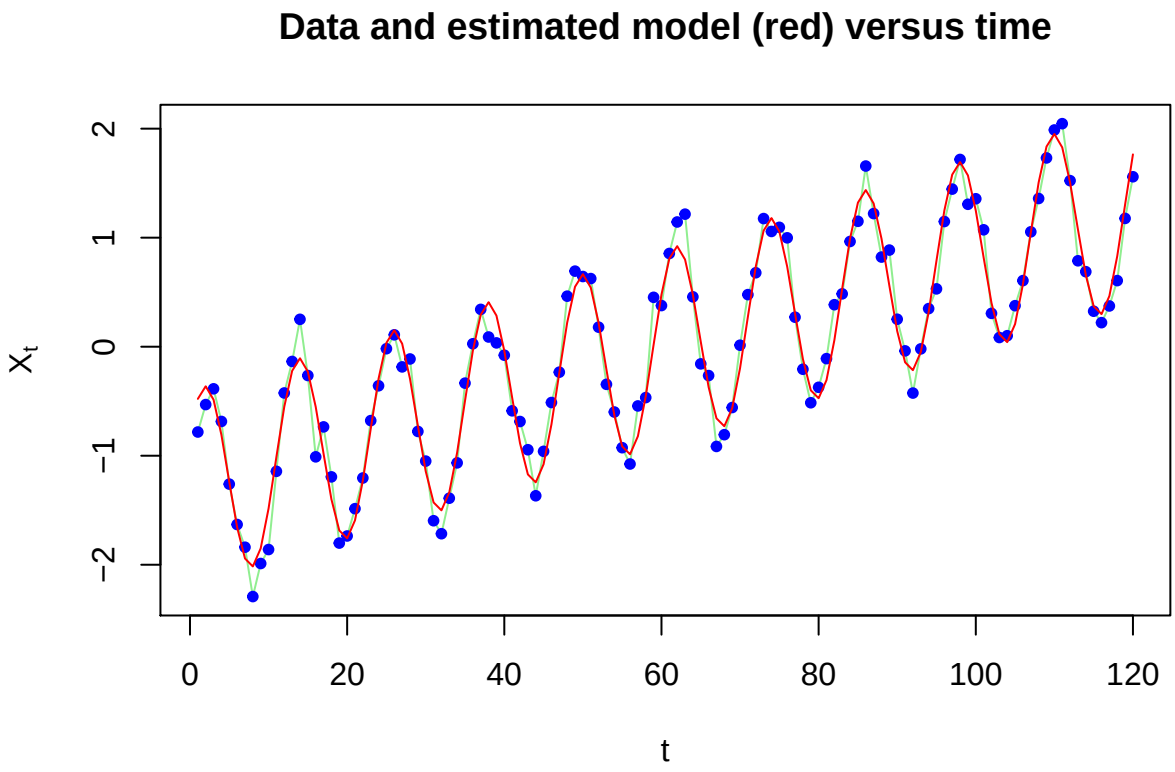
a least squares estimate for a linear trend model and a single trigonometric polynomial

$$a_0 + a_1 t + b_1 cos(\pi t/6) + b_2 sin(\pi t/6)$$

a_0 = -1.29642763, a_1 = 0.02142856, b_1 = 0.48893652, and b_2 = 0.74596584.

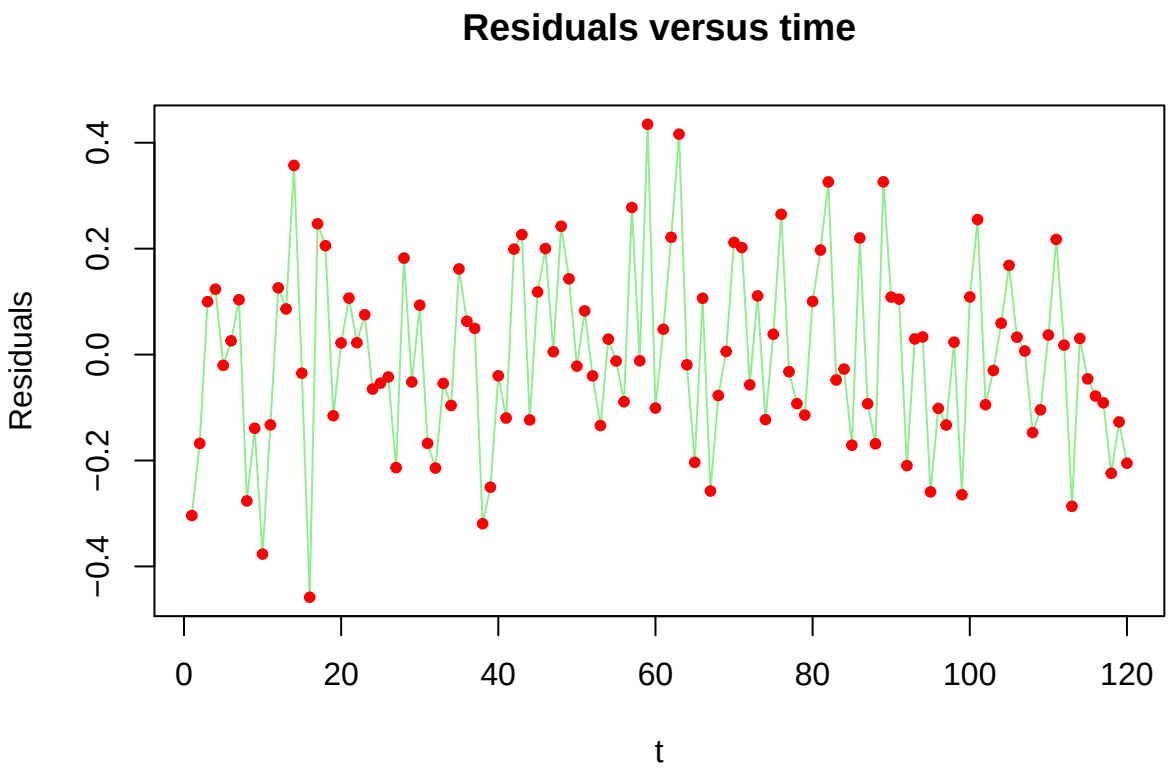
(b)

```
plot(data$V1, type = 'l', col = 'lightgreen',
     xlab = 't', ylab = expression(X[t]),
     main = 'Data and estimated model (red) versus time')
points(data$V1, pch = 20, col = 'blue')
lines(fitted(trigest), col = 'red')
```



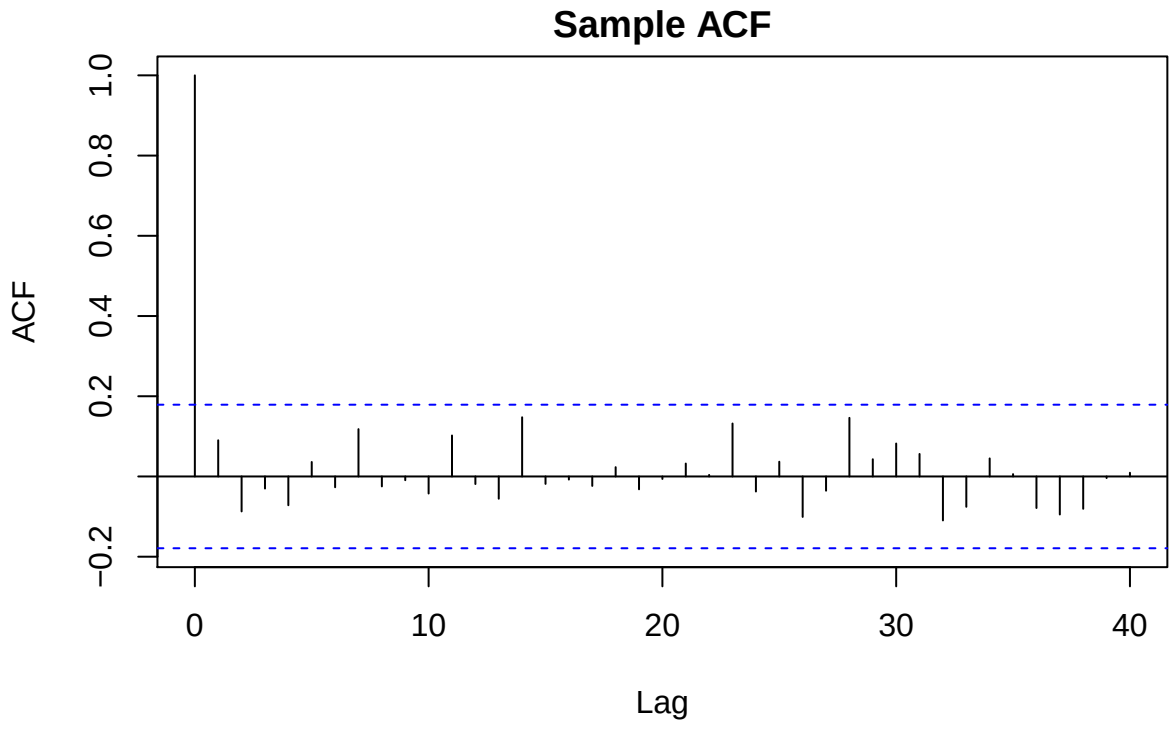
(c)

```
plot(resid(trigest), type = 'l', col = 'lightgreen',
     xlab = 't', ylab = 'Residuals',
     main = 'Residuals versus time')
points(resid(trigest), pch = 20, col = 'red')
```



(d)

```
acf1 = acf(resid(trigest), lag.max = 40, plot = F)
plot(acf1, main = "")
title("Sample ACF",line=0.5)
```

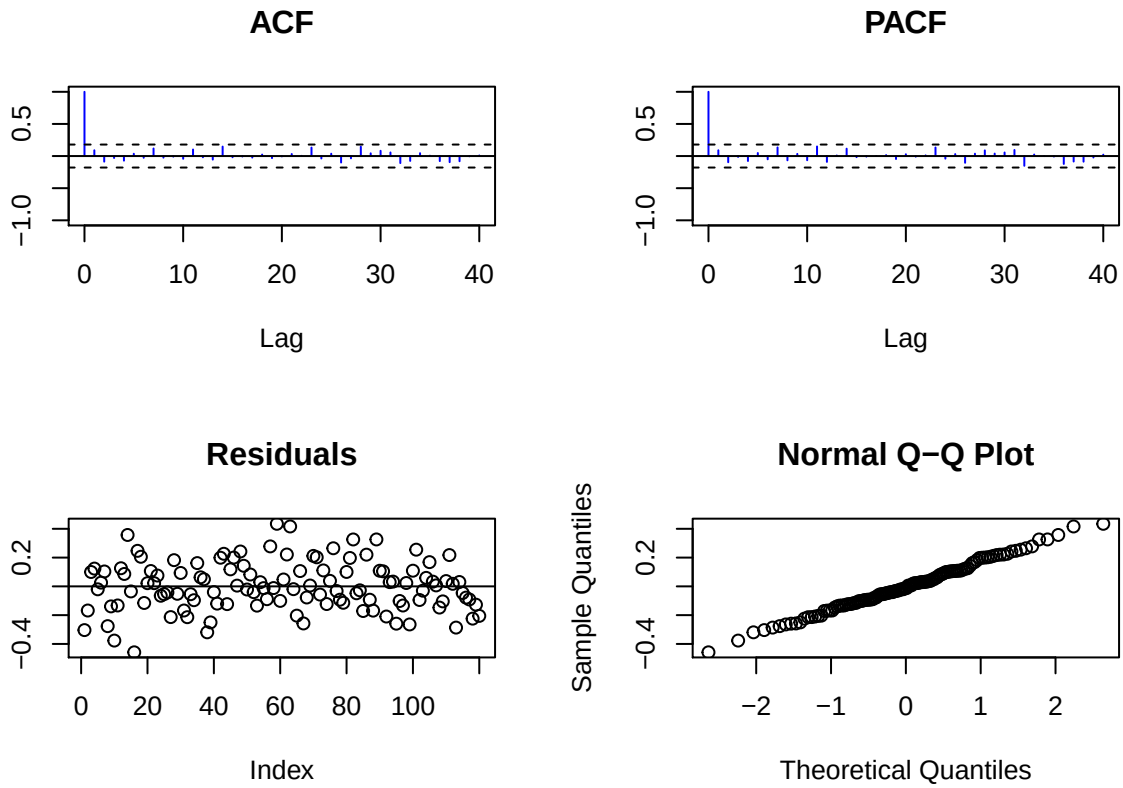


According to the sample acf plot, we fail to reject the hypothesis that the residuals are samples from an iid time series.

(e)

```
library(itsmr)
test(resid(trigest))
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     10.35    0.9614
## McLeod-Li Q                Q ~ chisq(20)     21.41    0.3731
## Turning points T   (T-78.7)/4.6 ~ N(0,1)        75    0.4238
## Diff signs S       (S-59.5)/3.2 ~ N(0,1)        60    0.8749
## Rank P           (P-3570)/220.4 ~ N(0,1)      3482    0.6897
```



$Q_{LB} = 10.35$ with p-value 0.9614 and $Q_{ML} = 21.41$ with p-value 0.3731. These do not support rejection of the hypothesis that the residuals are samples from an iid time series.

(f)

According to the sample acf plot and statistical tests, they all do not support the rejection of the hypothesis that the residuals are samples from an iid time series, so i would say the residuals are samples from an iid series.

3. Differencing

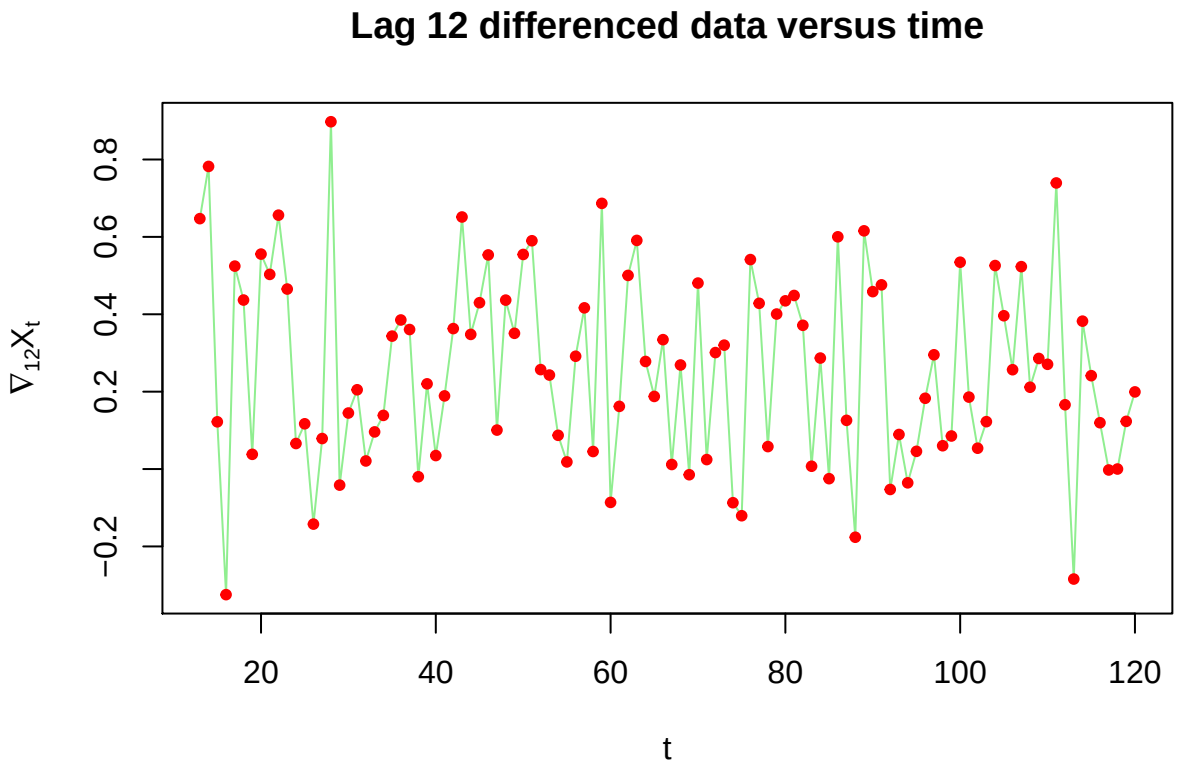
(a)

```
diff_data = diff(data$V1, lag = 12)
```

(b)

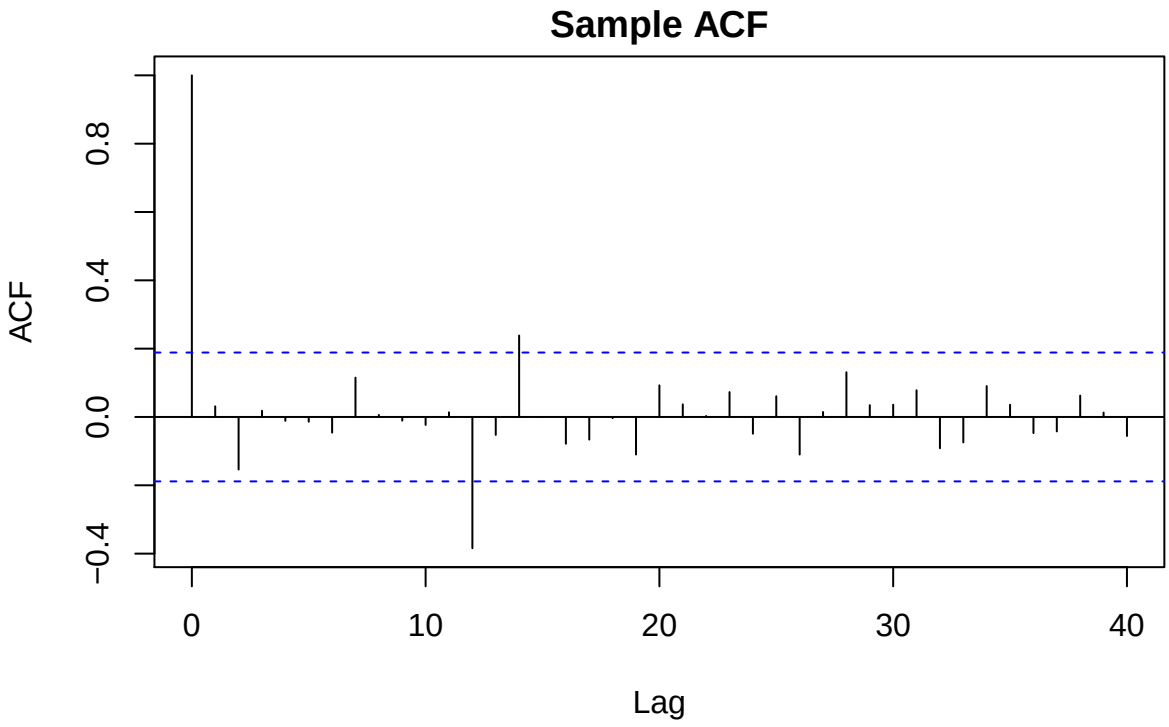
```
plot(x = 13:120, y = diff_data,
     type = 'l', col = 'lightgreen',
     xlab = 't', ylab = expression(paste(nabla[12],X[t])),
```

```
     main = 'Lag 12 differenced data versus time')
points(x = 13:120, y = diff_data, pch = 20, col = 'red')
```



(c)

```
acf2 = acf(diff_data, lag.max = 40, plot = F)
plot(acf2, main = "")
title("Sample ACF",line=0.5)
```

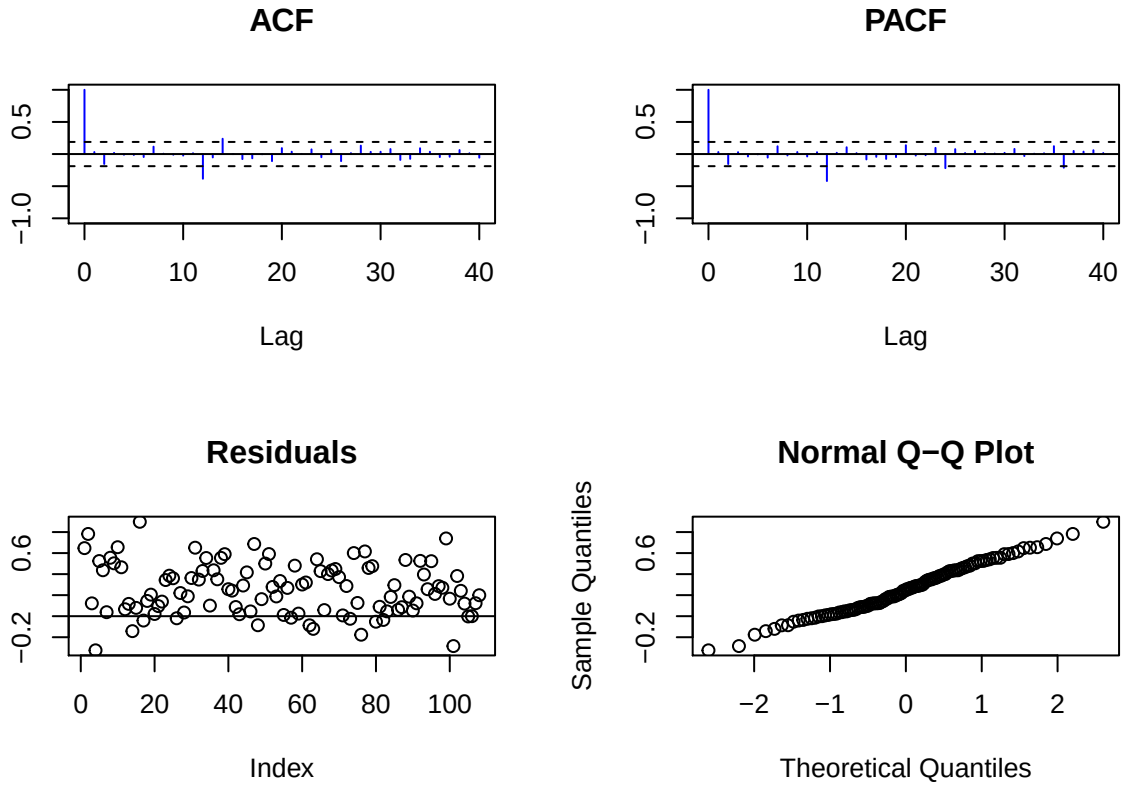


According to the sample acf plot, we do not have evidence against the hypothesis that the differenced data are samples from an iid time series.

(d)

```
test(diff_data)
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     34.68    0.0219 *
## McLeod-Li Q                Q ~ chisq(20)     22.86    0.2958
## Turning points T   (T-70.7)/4.3 ~ N(0,1)        68    0.5394
## Diff signs S         (S-53.5)/3 ~ N(0,1)        56    0.4068
## Rank P           (P-2889)/188.3 ~ N(0,1)      2653    0.2102
```



$Q_{LB} = 34.68$ with p-value 0.0219. This support rejection of the hypothesis that the differenced data are samples from an iid time series at level 0.05.

$Q_{ML} = 22.86$ with p-value 0.2958. This do not support rejection of the hypothesis that the differenced data are samples from an iid time series at level 0.05.

(e)

According to the sample acf plot and statistical tests, most of the methods do not support rejection of the hypothesis that the differenced data are samples from an iid time series, so i would concluded that the differenced data are samples from an iid time series.