Model_building_2

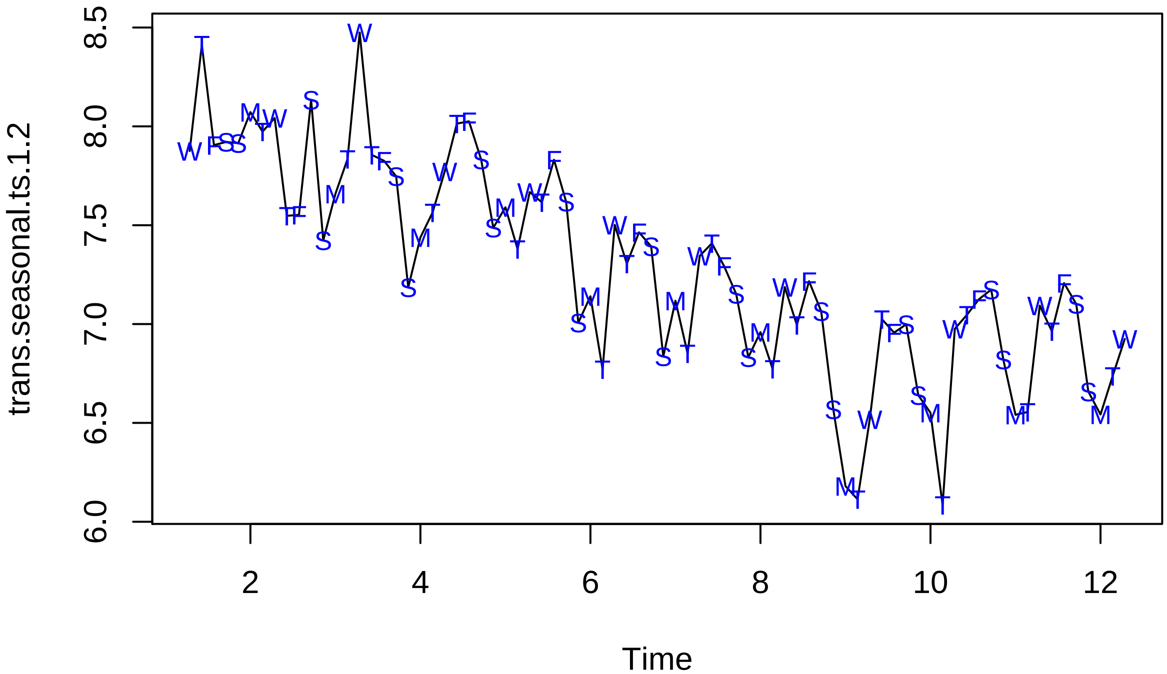
Annie Cohen

4/25/2021

```r
# pulling the data from the Los Angeles County GitHub
casedata <- read.csv(text = getURL("https://raw.githubusercontent.com/datadesk/california-coronavirus-da
  filter(county == "Los Angeles") %>%
  mutate(date = date(date), month = month(date)) %>%
  map_df(rev) %>%
  filter(!is.na(new_confirmed_cases) & between(date, date("2020-04-01"),date("2021-03-31")))

# creating the time series
case.ts <- ts(casedata$new_confirmed_cases, start = 1,  frequency = 1)

# averaging dec 25th and 26th
case.ts[269] <- 14711
case.ts[270] <- 14712

# jul 15 - sep 30
case.ts.1.2 <- ts(case.ts[106:183], start = 1, frequency = 1)
```

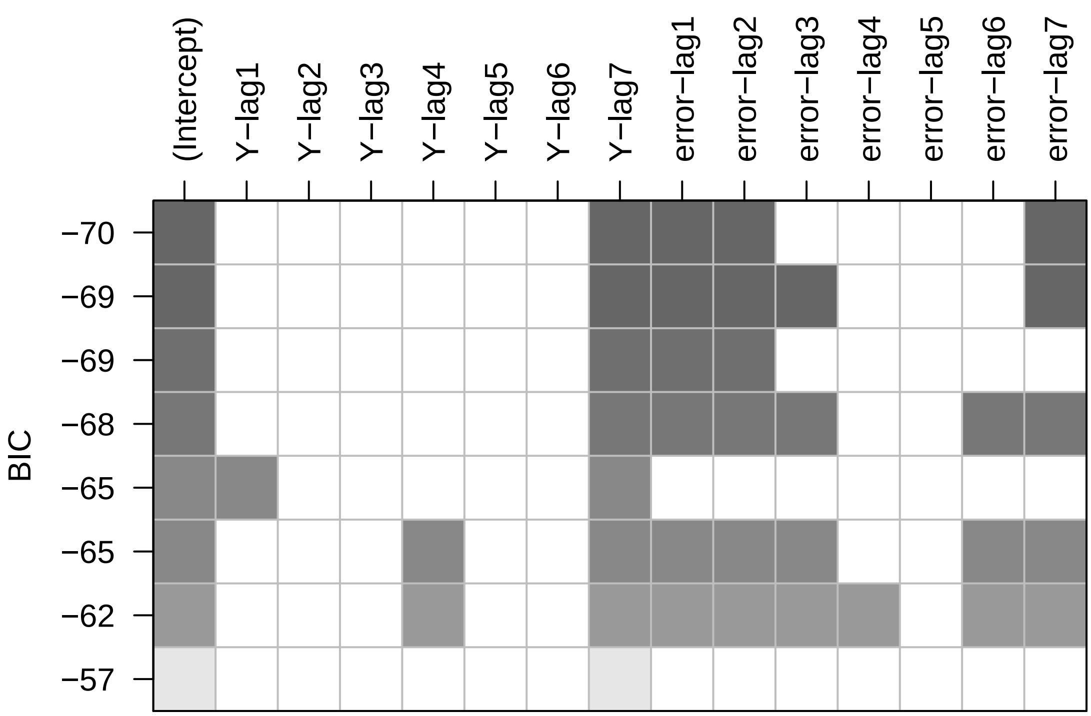
July 15 to September 30

```r
trans.seasonal.ts.1.2 <- ts(log(case.ts.1.2), frequency = 7, start = c(1,3))
trans.ts.1.2 <- ts(log(case.ts.1.2))

plot(trans.seasonal.ts.1.2, type = "l")
points(trans.seasonal.ts.1.2, pch = as.vector(season(trans.seasonal.ts.1.2)), cex = 0.8, col = "blue")
```



This plot displays a weekly trend, given that the higher values are usually on Wednesday, Thursday, and Friday, while lower values seem to occur on Sunday, Monday, and Tuesday. This suggests that we may want to look into a model that incorporates this seasonality.

```
plot(armasubsets(trans.ts.1.2, nar = 7, nma = 7))
```




```
eacf(diff(trans.ts.1.2, lag = 7))
```

```
## AR/MA
```

```
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o  o  o  o
## 1 o o o o o o o o o o o  o  o  o
## 2 x o o o o o o o o o o  o  o  o
## 3 o o o o o o o o o o o  o  o  o
## 4 o o o o o o o o o o o  o  o  o
## 5 o x o o o o x o o o o  o  o  o
## 6 o x o o o o o o o o o  o  o  o
## 7 x x x o o o o o o o o  o  o  o
```

```
auto.arima(trans.seasonal.ts.1.2, max.P = 5, max.Q = 5)
```

```
## Series: trans.seasonal.ts.1.2
## ARIMA(1,0,0)(0,1,1)[7] with drift
##
## Coefficients:
##          ar1     sma1    drift
##       0.5185  -0.8115  -0.0176
## s.e.  0.1035   0.1717   0.0027
##
## sigma^2 estimated as 0.05459:  log likelihood=0.15
## AIC=7.7   AICc=8.31   BIC=16.75
```

```
auto.arima(trans.ts.1.2, max.P = 5, max.Q = 5)
```

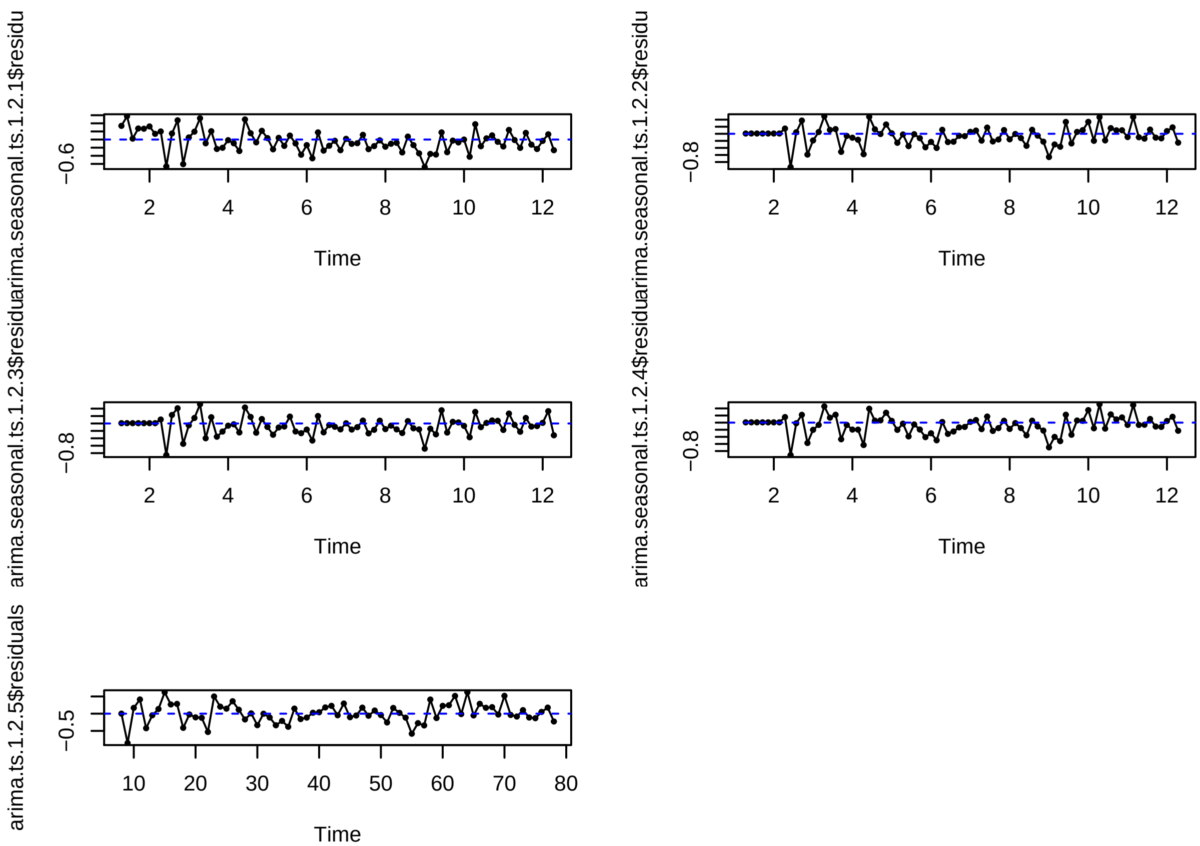
```
## Series: trans.ts.1.2
## ARIMA(0,1,2) with drift
##
## Coefficients:
##           ma1      ma2    drift
##       -0.5855  -0.2959  -0.0172
## s.e.   0.1017   0.0997   0.0048
##
## sigma^2 estimated as 0.08927:  log likelihood=-15.39
## AIC=38.77   AICc=39.33   BIC=48.15
```

These functions help narrow down model choices. Our candidate models are:

- ARIMA(0,0,2)x(1,0,1)[7]
- ARIMA(1,0,0)x(0,1,0)[7]
- ARIMA(1,0,0)x(0,1,1)[7]
- ARIMA(0,0,1)x(0,1,0)[7]
- ARIMA(0,1,2)

```
arima.seasonal.ts.1.2.1 <-arima(trans.seasonal.ts.1.2, order = c(0,0,2), seasonal = list(order = c(1,0,
arima.seasonal.ts.1.2.2 <- arima(trans.seasonal.ts.1.2, order = c(1,0,0), seasonal = list(order = c(0,1
arima.seasonal.ts.1.2.3 <- arima(trans.seasonal.ts.1.2, order = c(1,0,0), seasonal = list(order = c(0,1
arima.seasonal.ts.1.2.4 <- arima(trans.seasonal.ts.1.2, order = c(0,0,1), seasonal = list(order = c(0,1
arima.ts.1.2.5 <- arima(diff(trans.ts.1.2, lag = 7), order = c(0,1,2))
```

```r
par(mfrow = c(3,2))

plot(arima.seasonal.ts.1.2.1$residuals, type = "o", pch = 20)
abline(a=0,b=0,lty=2,col="blue")
plot(arima.seasonal.ts.1.2.2$residuals, type = "o", pch = 20)
abline(a=0,b=0,lty=2,col="blue")
plot(arima.seasonal.ts.1.2.3$residuals, type = "o", pch = 20)
abline(a=0,b=0,lty=2,col="blue")
plot(arima.seasonal.ts.1.2.4$residuals, type = "o", pch = 20)
abline(a=0,b=0,lty=2,col="blue")
plot(arima.ts.1.2.5$residuals, type = "o", pch = 20)
abline(a=0,b=0,lty=2,col="blue")
```




```r
arima.seasonal.ts.1.2.1$aic
```

```
## [1] 26.68219
```

```r
arima.seasonal.ts.1.2.2$aic
```

```
## [1] 21.78533
```

```r
arima.seasonal.ts.1.2.3$aic
```

```
## [1] 15.22738
```

```r
arima.seasonal.ts.1.2.4$aic
```

```
## [1] 26.04906
```

```r
arima.ts.1.2.5$aic
```

```
## [1] 25.75294
```

Based on the AIC values, the third model, the ARIMA(1,0,0)x(0,1,1)[7] has the lowest AIC value of 15.23.

```r
Box.test(arima.seasonal.ts.1.2.1$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.seasonal.ts.1.2.1$residuals
## X-squared = 0.024176, df = 1, p-value = 0.8764
```

```r
Box.test(arima.seasonal.ts.1.2.2$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.seasonal.ts.1.2.2$residuals
## X-squared = 0.78523, df = 1, p-value = 0.3755
```

```r
Box.test(arima.seasonal.ts.1.2.3$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.seasonal.ts.1.2.3$residuals
## X-squared = 5.3272, df = 1, p-value = 0.02099
```

```r
Box.test(arima.seasonal.ts.1.2.4$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.seasonal.ts.1.2.4$residuals
## X-squared = 0.0037858, df = 1, p-value = 0.9509
```

```r
Box.test(arima.ts.1.2.5$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arima.ts.1.2.5$residuals
## X-squared = 0.18201, df = 1, p-value = 0.6696
```

Here we can see that all of the models pass the Ljung-Box test except for the third model, which is interesting because it had the lowest AIC.

```
shapiro.test(arima.seasonal.ts.1.2.1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  arima.seasonal.ts.1.2.1$residuals
## W = 0.99003, p-value = 0.8096
```

```
shapiro.test(arima.seasonal.ts.1.2.2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  arima.seasonal.ts.1.2.2$residuals
## W = 0.96865, p-value = 0.05141
```

```
shapiro.test(arima.seasonal.ts.1.2.3$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  arima.seasonal.ts.1.2.3$residuals
## W = 0.97783, p-value = 0.1914
```

```
shapiro.test(arima.seasonal.ts.1.2.4$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  arima.seasonal.ts.1.2.4$residuals
## W = 0.97752, p-value = 0.183
```

```
shapiro.test(arima.ts.1.2.5$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  arima.ts.1.2.5$residuals
## W = 0.98361, p-value = 0.4824
```

We see here that all of them pass the Shapiro-Wilk test, although the second model is right on the border, and I would be hesitant to move forward with it, considering the residuals may not be normally distributed.

```
runs(arima.seasonal.ts.1.2.1$residuals)
```

```
## $pvalue
## [1] 0.655
##
## $observed.runs
## [1] 36
##
## $expected.runs
## [1] 38.35897
##
## $n1
## [1] 47
##
## $n2
## [1] 31
##
## $k
## [1] 0
```

```
runs(arima.seasonal.ts.1.2.2$residuals)
```

```
## $pvalue
## [1] 0.227
##
## $observed.runs
## [1] 34
##
## $expected.runs
## [1] 39.76923
##
## $n1
## [1] 42
##
## $n2
## [1] 36
##
## $k
## [1] 0
```

```
runs(arima.seasonal.ts.1.2.3$residuals)
```

```
## $pvalue
## [1] 0.891
##
## $observed.runs
## [1] 38
##
## $expected.runs
## [1] 39.07692
##
## $n1
## [1] 45
##
## $n2
```

```
## [1] 33
##
## $k
## [1] 0
```

```
runs(arima.seasonal.ts.1.2.4$residuals)
```

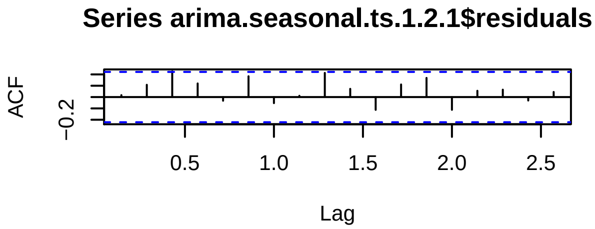
```
## $pvalue
## [1] 0.102
##
## $observed.runs
## [1] 32
##
## $expected.runs
## [1] 39.58974
##
## $n1
## [1] 43
##
## $n2
## [1] 35
##
## $k
## [1] 0
```

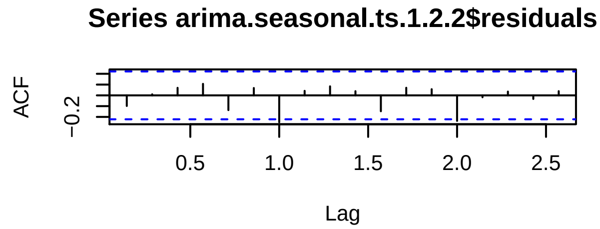
```
runs(arima.ts.1.2.5$residuals)
```

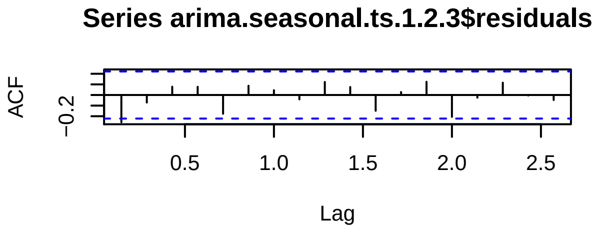
```
## $pvalue
## [1] 0.997
##
## $observed.runs
## [1] 36
##
## $expected.runs
## [1] 36.49296
##
## $n1
## [1] 36
##
## $n2
## [1] 35
##
## $k
## [1] 0
```

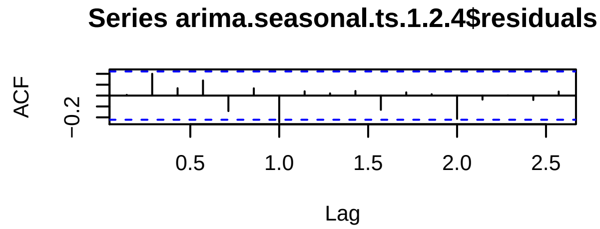
All of these pass the runs test, suggesting that the residuals are all fairly independent.

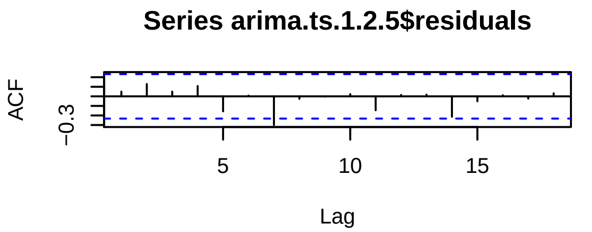
```
layout(matrix(c(1,2,3,4,5,0), nrow = 3, ncol = 2, byrow = TRUE))

acf(arima.seasonal.ts.1.2.1$residuals)
acf(arima.seasonal.ts.1.2.2$residuals)
acf(arima.seasonal.ts.1.2.3$residuals)
acf(arima.seasonal.ts.1.2.4$residuals)
acf(arima.ts.1.2.5$residuals)
```



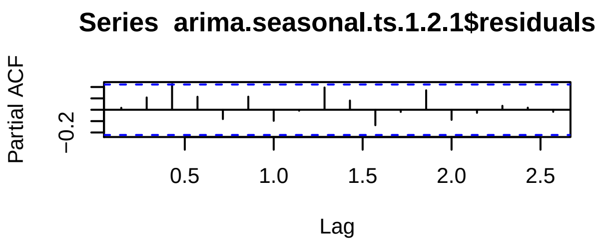


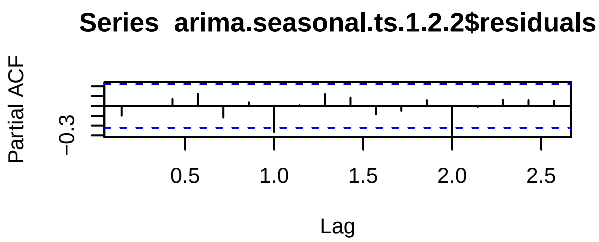


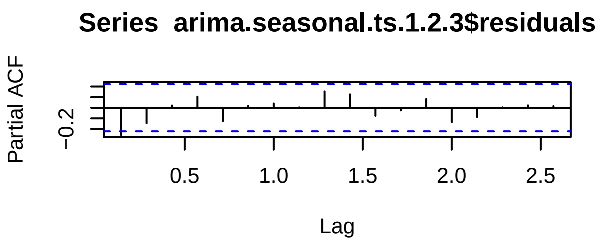


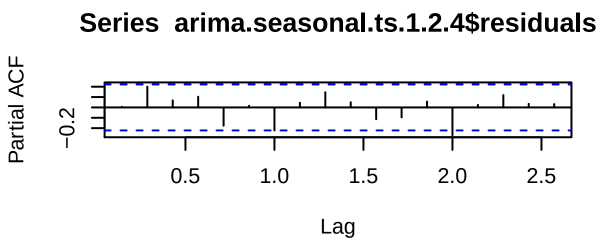


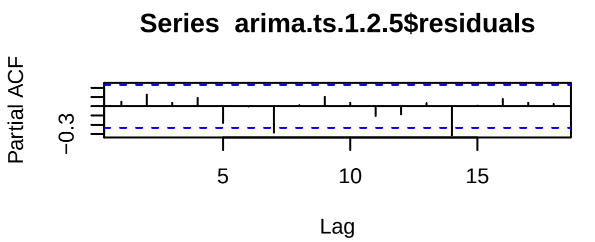

```
pacf(arima.seasonal.ts.1.2.1$residuals)
pacf(arima.seasonal.ts.1.2.2$residuals)
pacf(arima.seasonal.ts.1.2.3$residuals)
pacf(arima.seasonal.ts.1.2.4$residuals)
pacf(arima.ts.1.2.5$residuals)
```



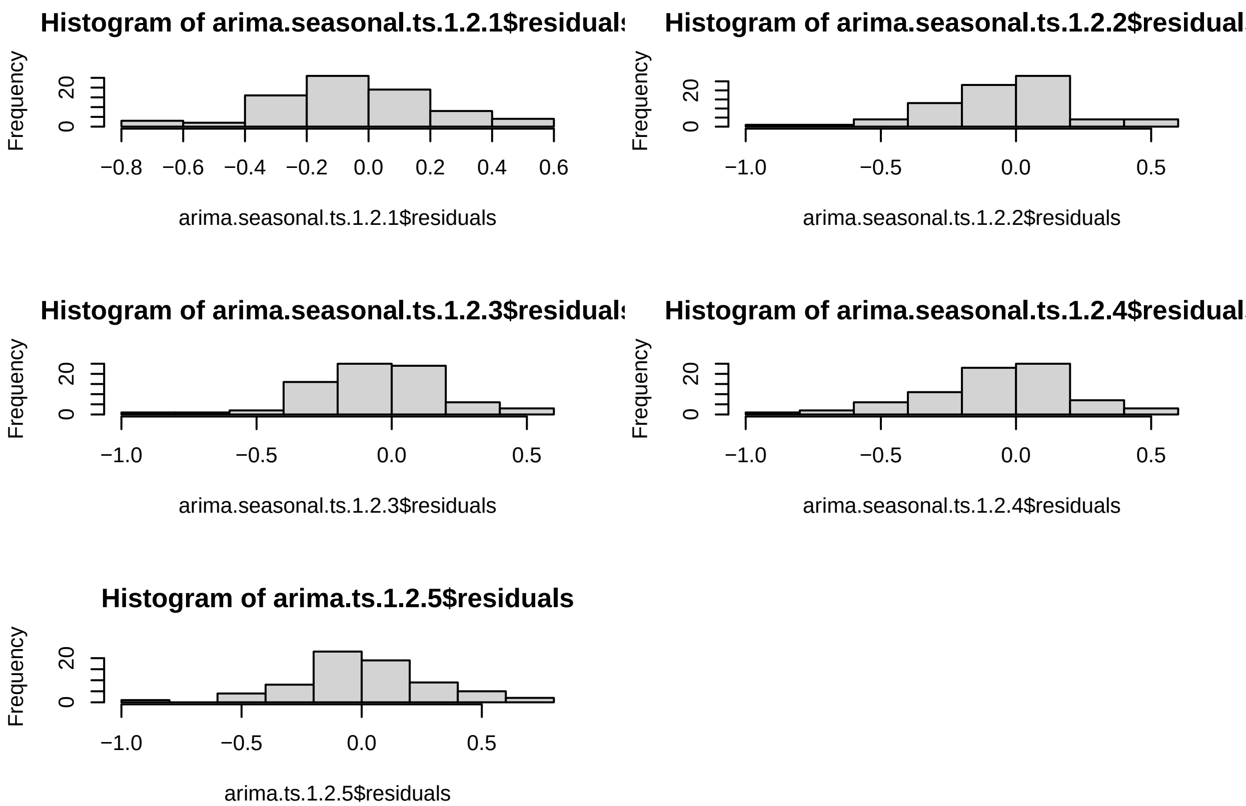




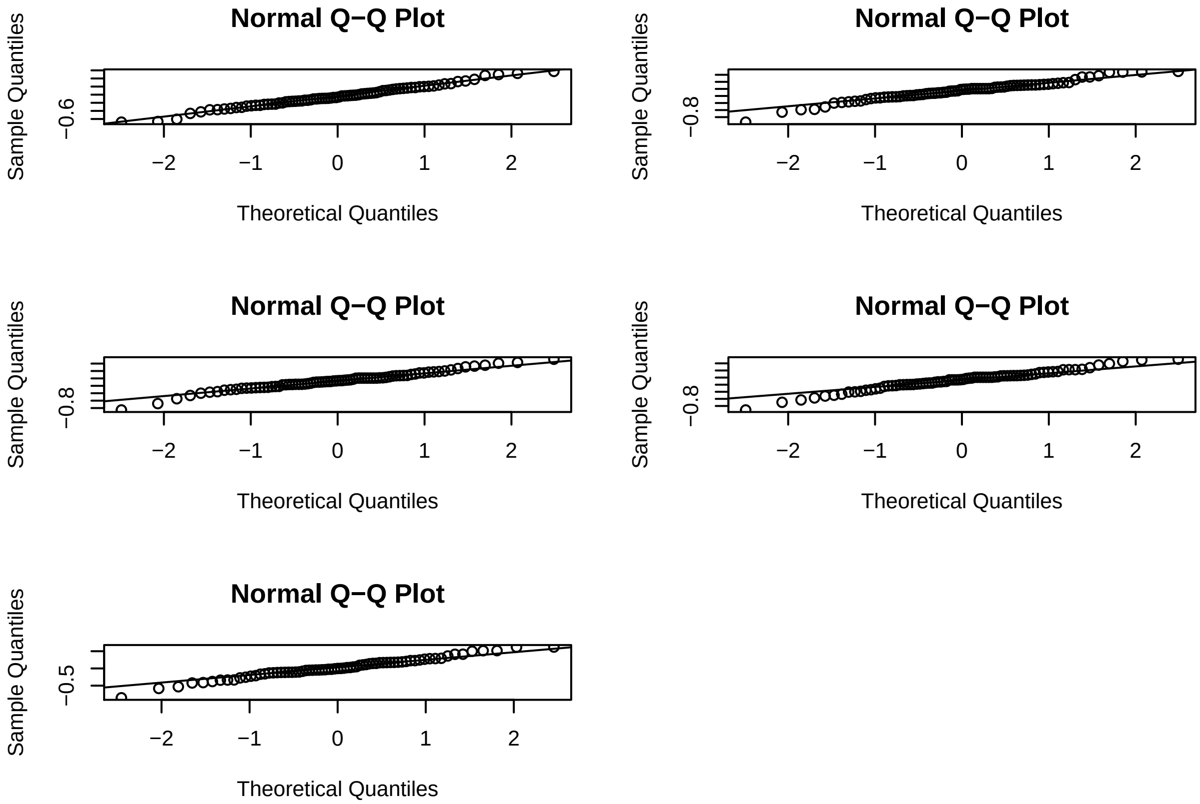





```
hist(arima.seasonal.ts.1.2.1$residuals)
hist(arima.seasonal.ts.1.2.2$residuals)
hist(arima.seasonal.ts.1.2.3$residuals)
hist(arima.seasonal.ts.1.2.4$residuals)
hist(arima.ts.1.2.5$residuals)
```




```r
qqnorm(arima.seasonal.ts.1.2.1$residuals)
qqline(arima.seasonal.ts.1.2.1$residuals)

qqnorm(arima.seasonal.ts.1.2.2$residuals)
qqline(arima.seasonal.ts.1.2.2$residuals)

qqnorm(arima.seasonal.ts.1.2.3$residuals)
qqline(arima.seasonal.ts.1.2.3$residuals)

qqnorm(arima.seasonal.ts.1.2.4$residuals)
qqline(arima.seasonal.ts.1.2.4$residuals)

qqnorm(arima.ts.1.2.5$residuals)
qqline(arima.ts.1.2.5$residuals)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−0.6
−0.8
−0.5
−2
−1
0
1
2