# STAT 421/621 Project 1 - Modeling and Forecast ICE CREAM MANUFACTURING

Yuwen Liang

3/16/2023

Project 1 is worth 5% of final grade.

Honor Code: You may work with TWO other people on this project. If you do so, you should indicate all three authors on the paper, but submit individually.

**DUE DATE:** March 20, 2023 at 10pm.

**WHAT TO SUBMIT:** Solutions should be written using RMarkdown. You will submit

- The compiled html (or pdf if you prefer)
- The RMarkdown file

## Background

The time series used is IPN31152N.csv. This time series represents monthly ice cream production for the US since 1972 through 2022. The series is not seasonally adjusted but is indexed to 2017.

Data Citation: Board of Governors of the Federal Reserve System (US), Industrial Production: Manufacturing: Non-Durable Goods: Ice Cream and Frozen Dessert (NAICS = 31152) [IPN31152N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/IPN31152N, March 7, 2023.

Our goals for analyzing this data are

```
+ Describing and modeling to understand the dynamics of ice cream production
+ Forecasting monthly production through December of 2025
```

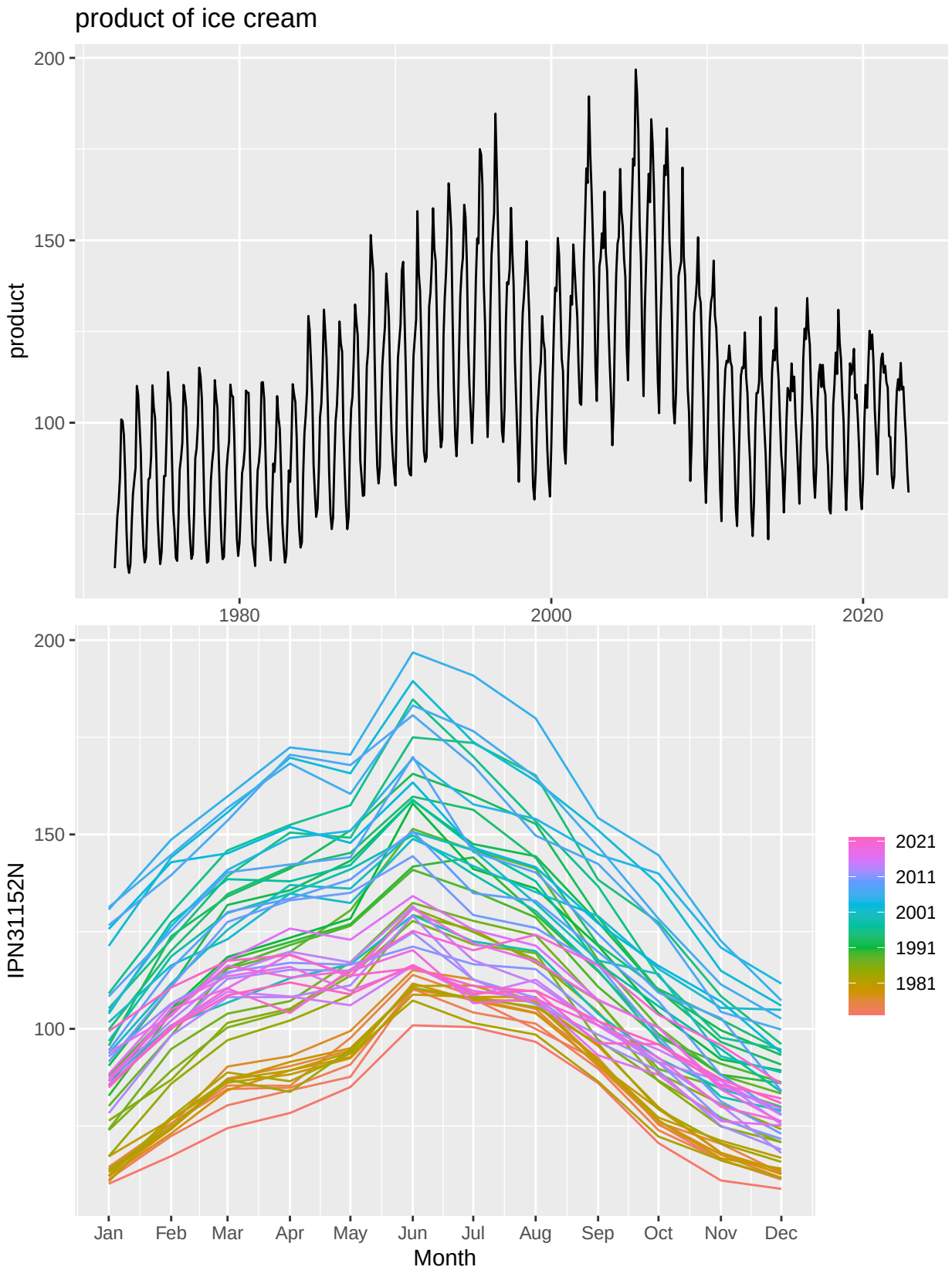
- **QUESTION 0** AFTER you answer questions 1-4, come back and ANSWER THIS QUESTION. Provide a succinct summary of your findings to address the above goals. Choose as the audience for your paragraph a manager at a company producing ice cream production, or an investment manager who focuses on ice cream, or something similar. For example, these managers WILL NOT CARE ABOUT order selection criteria but will expect that you did you analyst job well and the insight you are providing them is something that they can move forward with profitable decisions for the company. Do include uncertainty bounds in your narrative.

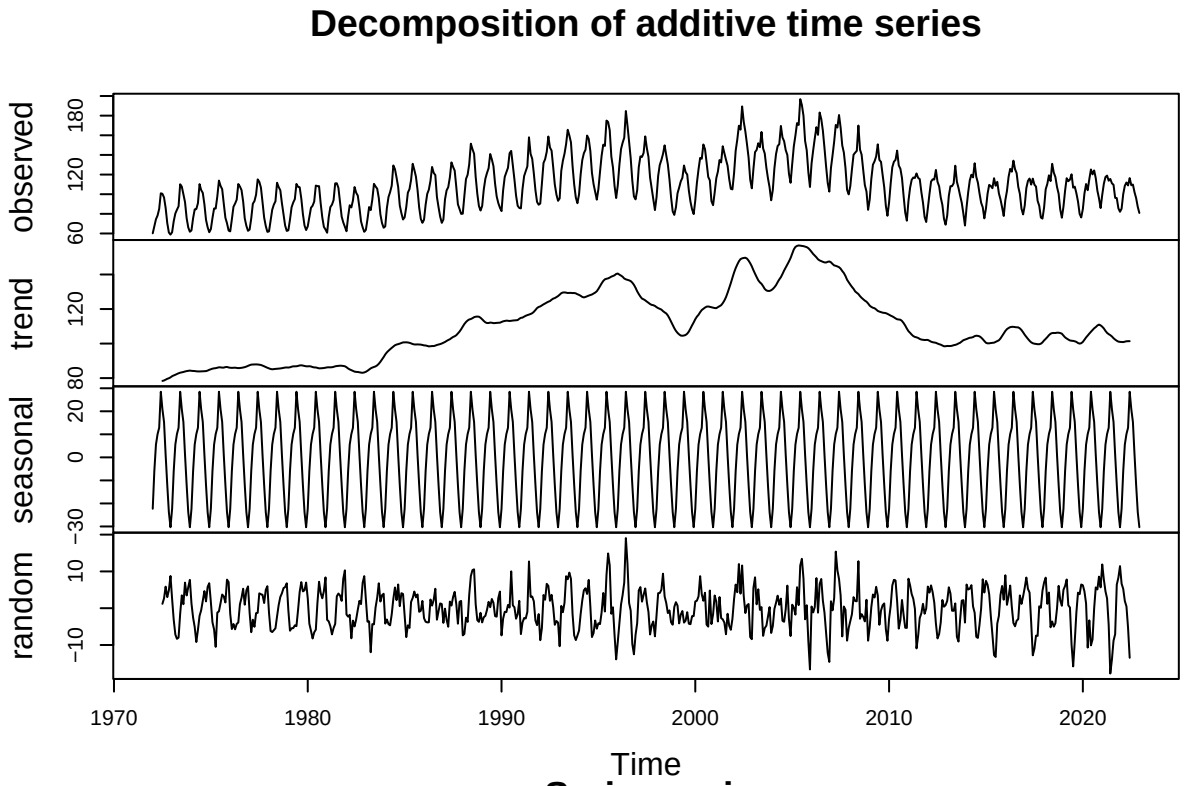
#After analyzing this time series, we can understand that the production of ice cream should conform to the seasonal component. For example, ice cream producers should increase ice cream production in June, July and August and decrease production in other months to prevent ice cream hoarding. In addition to seasonal factors, we also need to consider cycles and trends. For example, according to the time series chart, the two significant decreases in ice cream production could be due to people being more concerned about their health and reducing their sugar intake. So ice cream companies should also follow the trend and develop new varieties, such as making ice cream with less sugar, to increase sales without compromising taste.

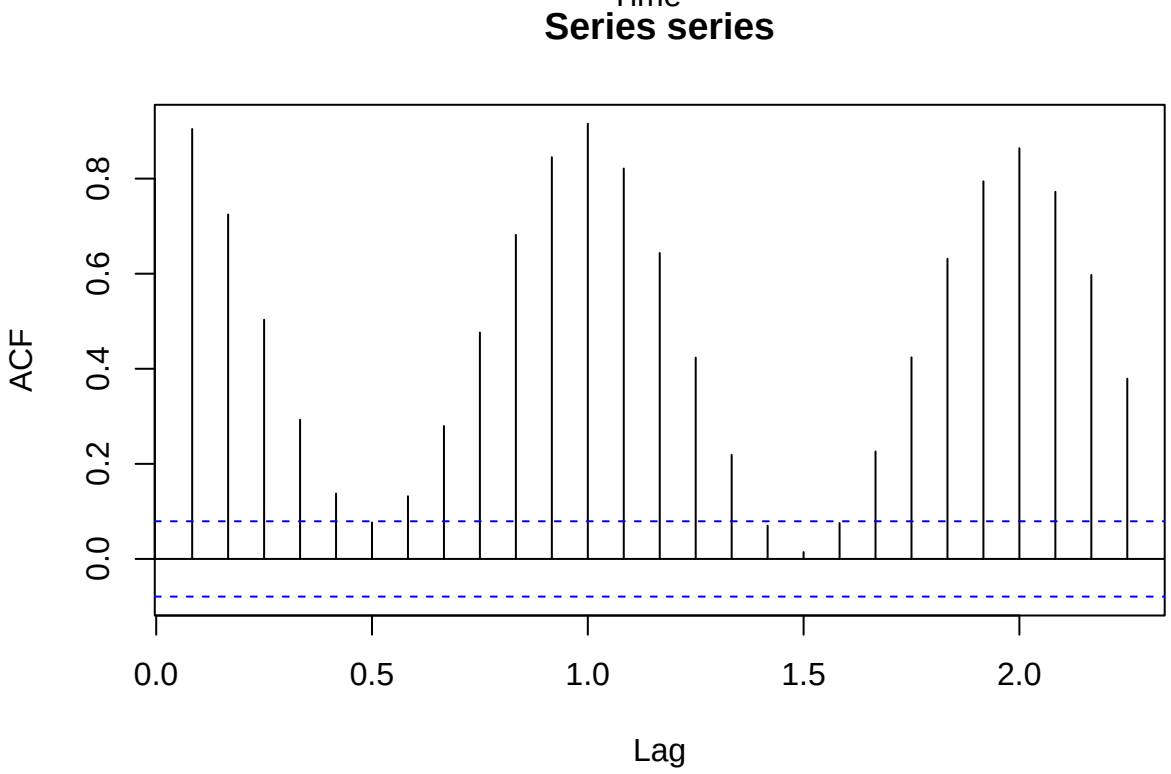
- **QUESTION 1** Produce descriptive plots and **{DISCUSS} what information you glean from each plot.**

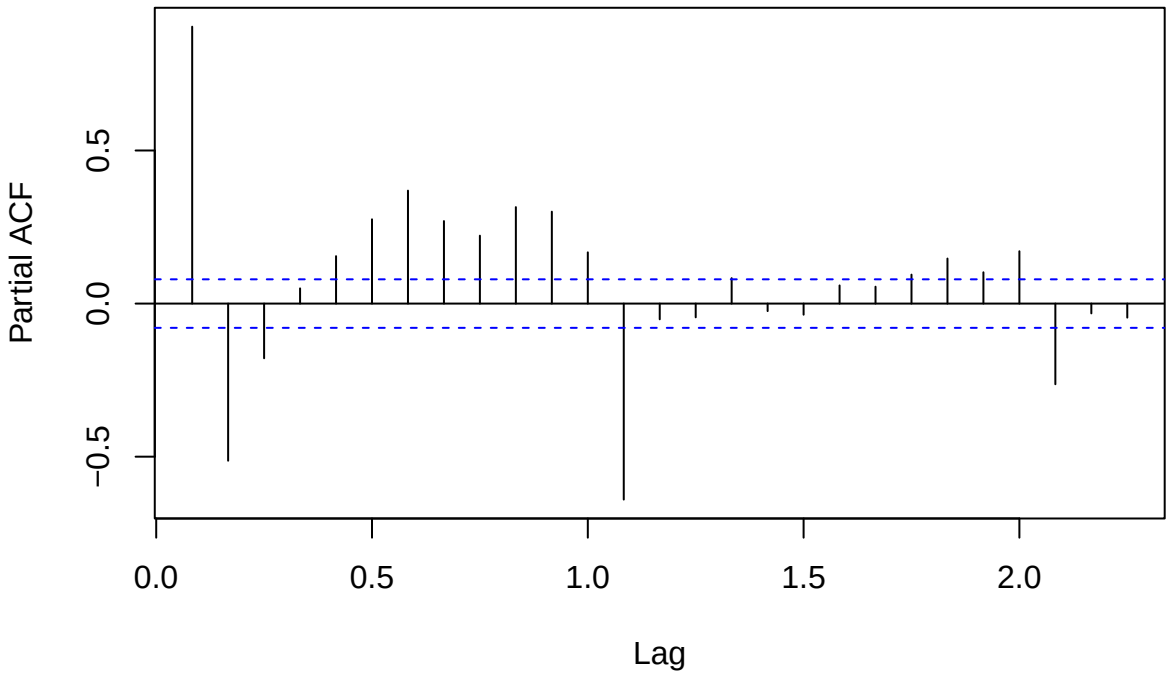
  - **Time series**
  - **Relevant seasonal plot**
  - **Decomposition plot**
  - **ACF, PACF and periodogram**

```
##
## Attaching package: 'tsibble'

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, union

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: ggplot2

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following object is masked from 'package:tsibble':
##
##     index

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar

## Loading required package: fabletools
```

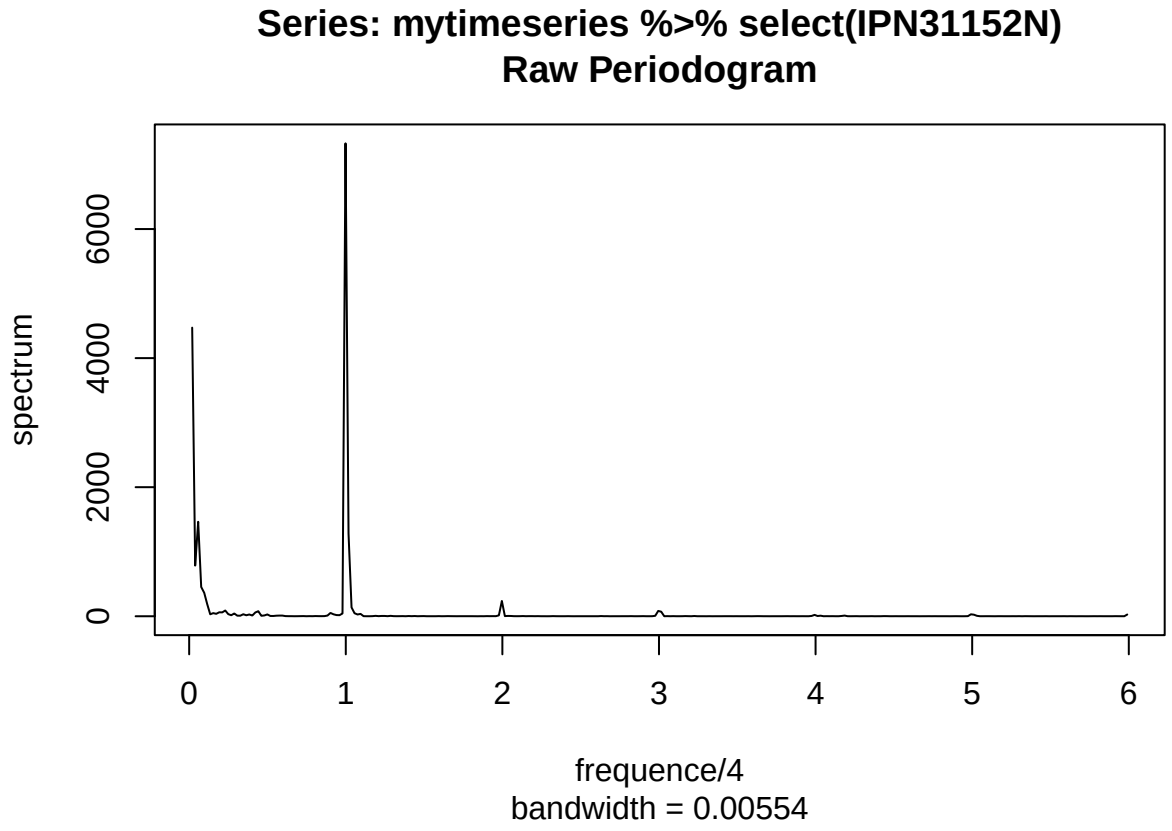

product of ice cream
product
200
150
100
1980
2000
2020
IPN31152N
200
150
100
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
2021
2011
2001
1991
1981


Decomposition of additive time series
observed
trend
seasonal
random
60
120
180
120
80
20
0
−30
10
−10
1970
1980
1990
2000
2010
2020
Time


Series series
ACF
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
Lag



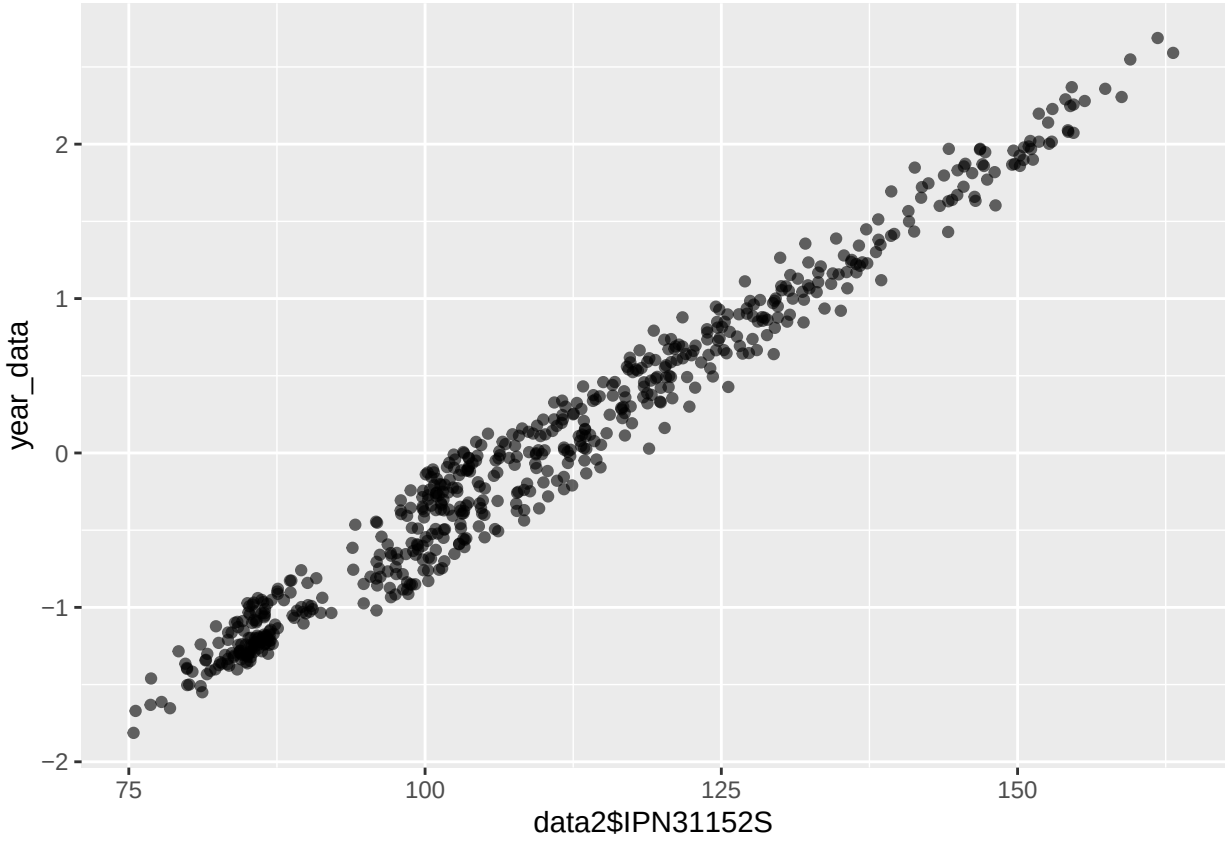

```
## Warning in plot.window(...): "dtrend" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "dtrend" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "dtrend" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "dtrend" is not a
## graphical parameter

## Warning in box(...): "dtrend" is not a graphical parameter

## Warning in title(...): "dtrend" is not a graphical parameter
```




# a.According to the time series graph it can be seen that ice cream production increased significantly from 1985 to 1988, and after experiencing a sudden decline around 2000, production rose again from 2003 to 2005. After that, ice cream production fell again. # b.According to the Relevant seasonal plot, it can be seen that there is a clear seasonal trend in the production of ice cream within a year. The production in June, July and August is significantly higher than the other months. # c.According to the Decomposition plot, it can be seen that this time series has trend and seasonal factors. # d.According to the ACF,PACF plots it can be seen that there is autocorrelation in this time series. The periodogram tells us that there is significant yearly circle in this time series.

- **QUESTION 2** Construct a **seasonally adjusted** time series by standardizing each month with the mean and standard deviation for that month across all years, and compare to the seasonally adjusted series from FRED, also in the folder as IPN31152S.csv. Use whatever graphs you deem helpful for this comparison (sometimes something simple like a scatterplot with colors for years or months works well - your choice). Again be sure and DISCUSS.

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:tsibble':
##
##     interval

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## `summarise()` has grouped output by 'Month'. You can override using the
## `.groups` argument.
```




#Based on the scatter plot, it can be seen that the SEASONALLY ADJUSTED series from FRED approximately follows a normal distribution.

- **QUESTION 3** Develop an appropriate SARIMA model, or trend stationary SARMA model for the original unadjusted series. Justify your model choice using
  - Preliminary tests (e.g. unit roots, trend stationary)
  - Model selection criteria (eg. AIC, AICc, BIC)
  - Diagnostics of standaridized residuals

Be sure and include a discussion not just the plots.

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## Registered S3 methods overwritten by 'forecast':
##   method                 from
##   autoplot.Arima         ggfortify
##   autoplot.acf           ggfortify
##   autoplot.ar            ggfortify
##   autoplot.bats          ggfortify
##   autoplot.decomposed.ts ggfortify
##   autoplot.ets           ggfortify
##   autoplot.forecast      ggfortify
##   autoplot.stl           ggfortify
##   autoplot.ts            ggfortify
##   fitted.Arima           TSA
##   fitted.ar              ggfortify
##   fortify.ts             ggfortify
```

```
##   plot.Arima             TSA
##   residuals.ar           ggfortify
```

```
##
## Attaching package: 'forecast'
```

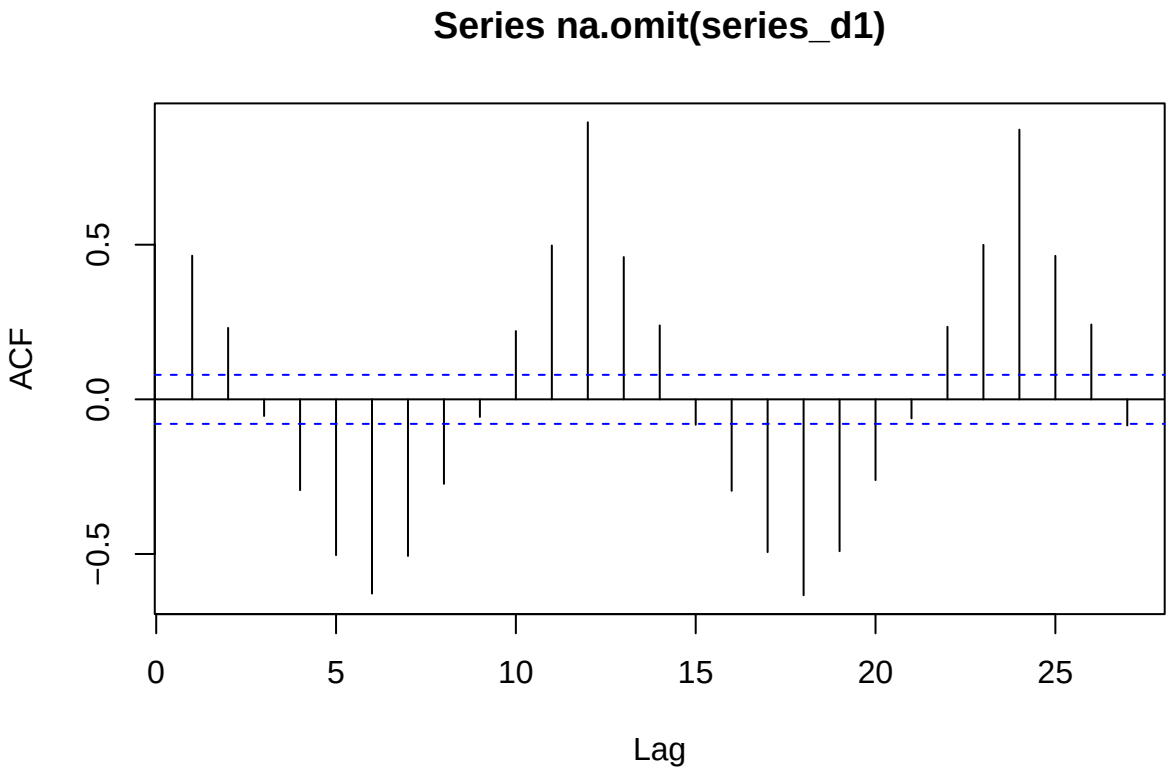
```
## The following object is masked from 'package:astsa':
##
##     gas
```

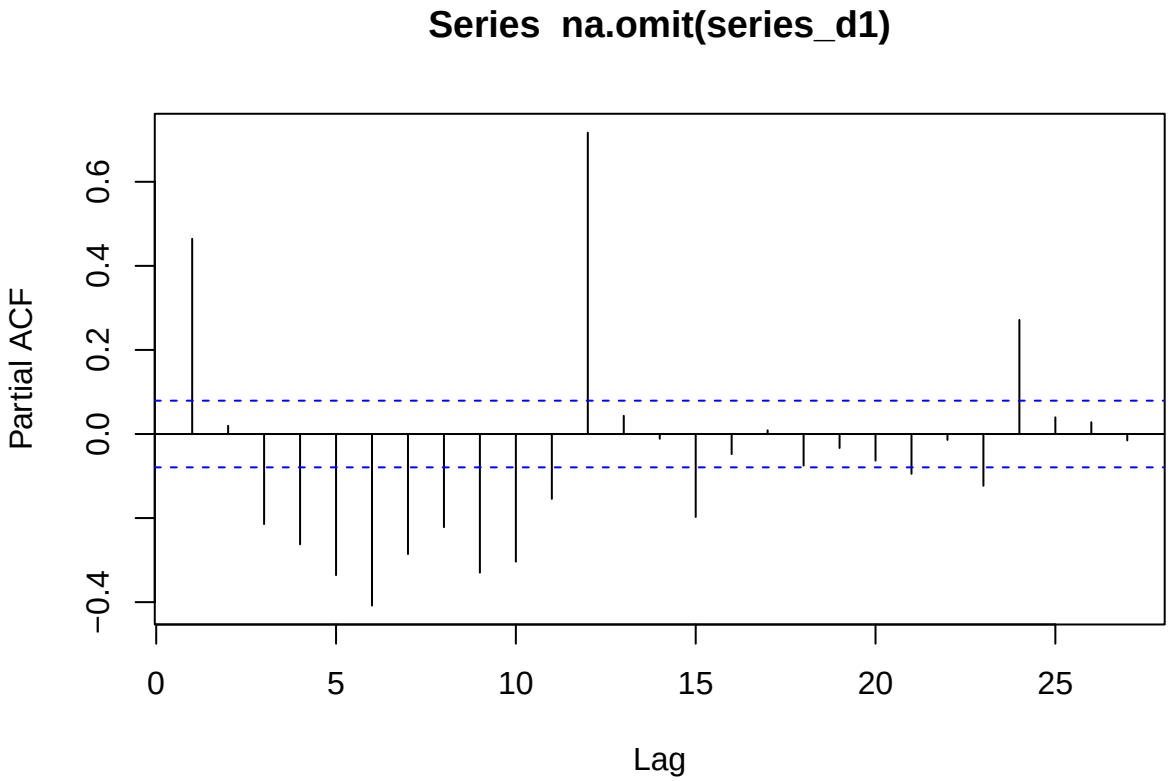
```
## The following objects are masked from 'package:fabletools':
##
##     accuracy, forecast
```

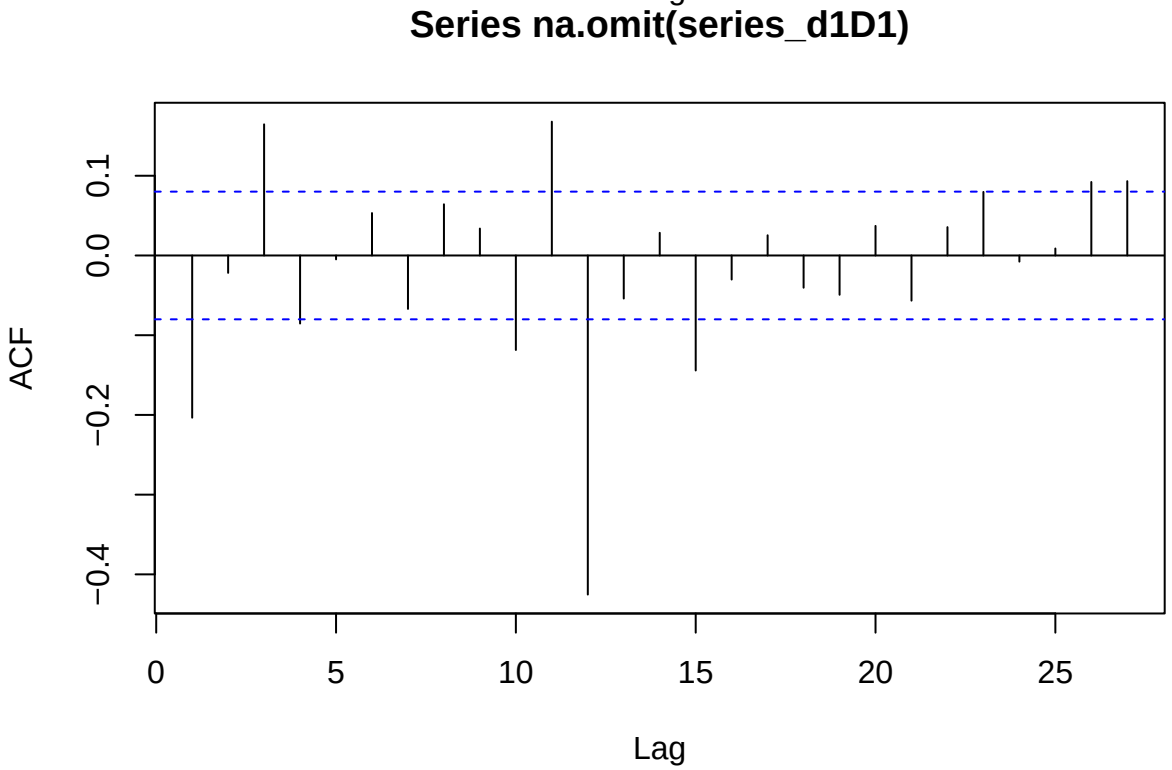
```
##
##  Augmented Dickey-Fuller Test
##
## data:  series
## Dickey-Fuller = -2.0396, Lag order = 8, p-value = 0.5615
## alternative hypothesis: stationary
```

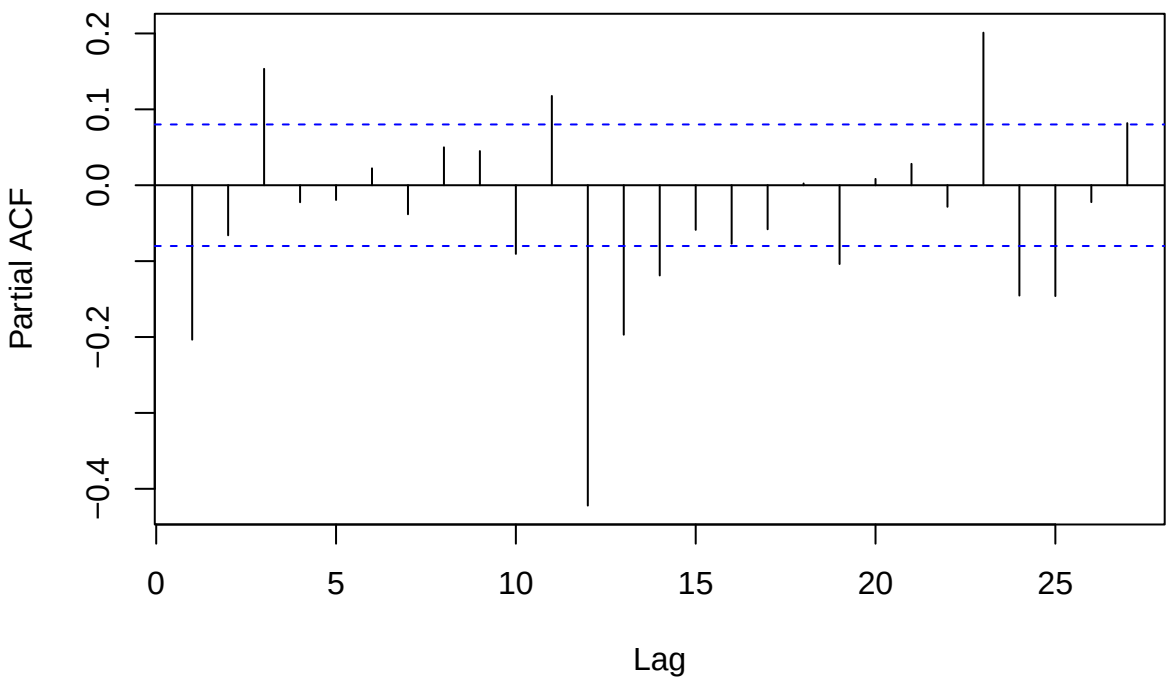
```
## Warning in adf.test(series_d1[2:length(series_d1)]): p-value smaller than
## printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  series_d1[2:length(series_d1)]
## Dickey-Fuller = -22.898, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```




Series na.omit(series_d1)
Partial ACF
0.6
0.4
0.2
0.0
−0.4
0
5
10
15
20
25
Lag


Series na.omit(series_d1D1)
ACF
0.1
0.0
−0.2
−0.4
0
5
10
15
20
25
Lag

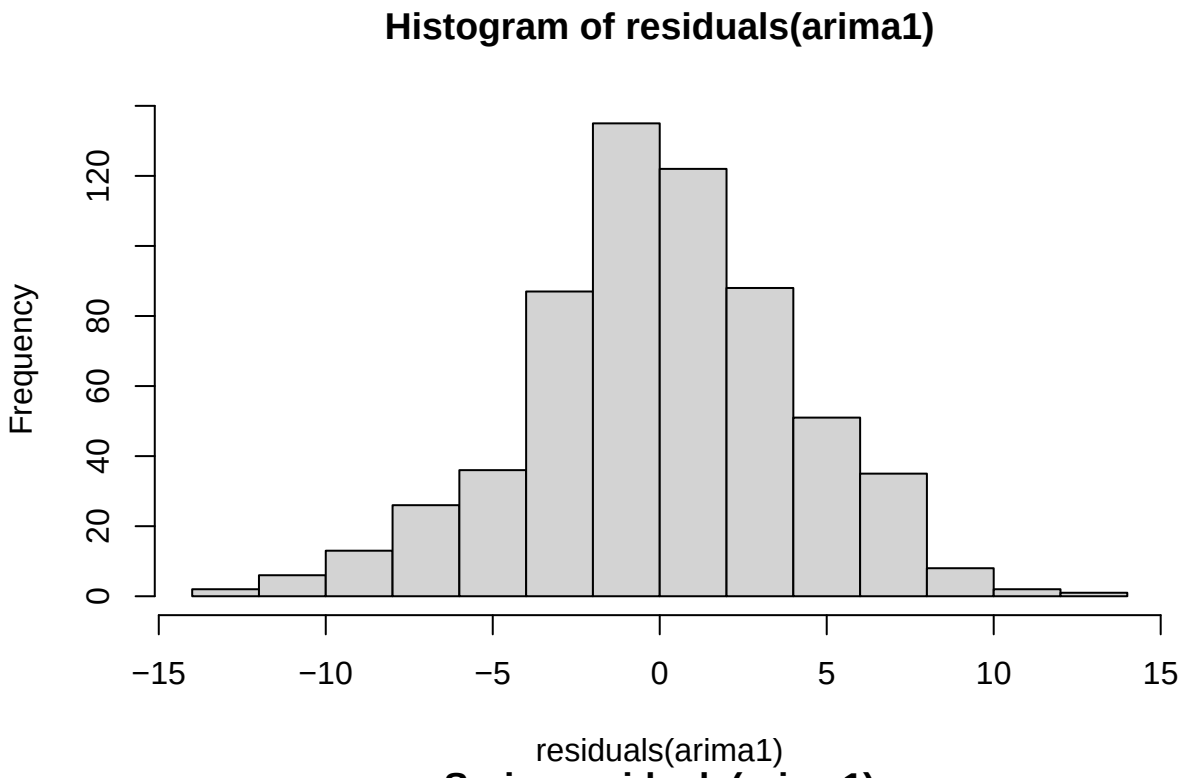


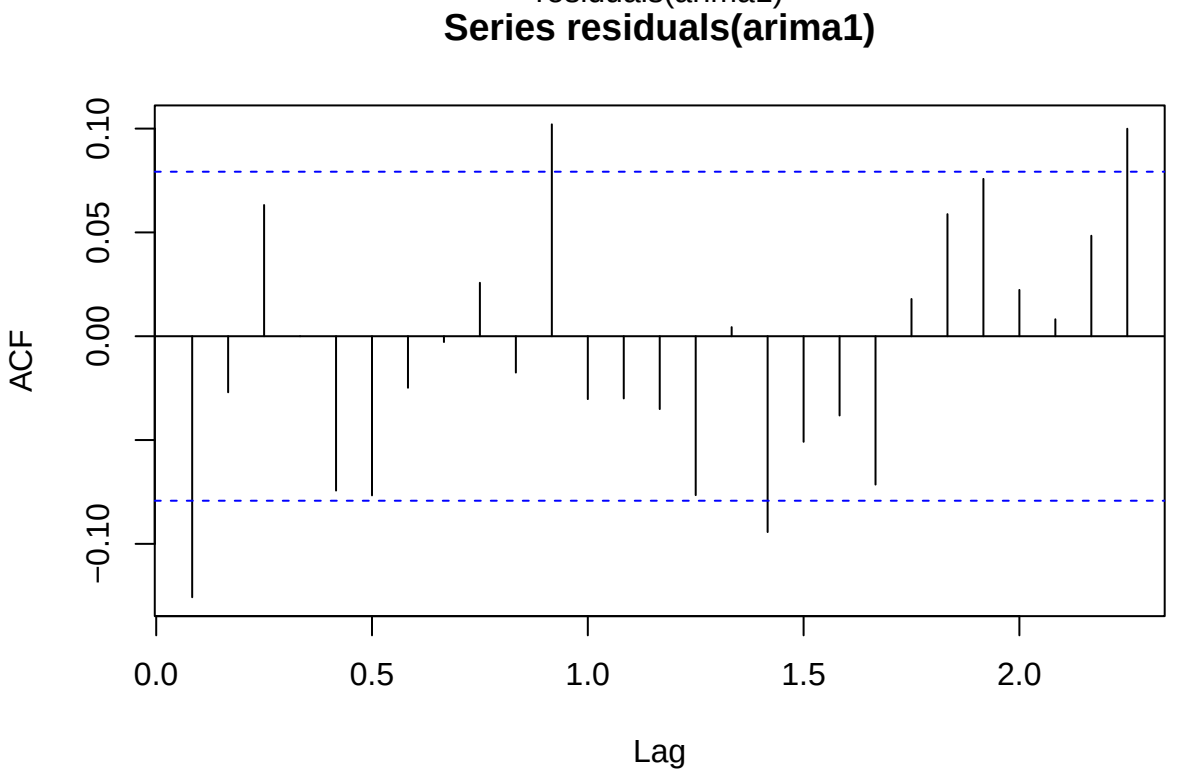

```
## Series: series
## ARIMA(2,1,2)(0,1,1)[12]
##
## Coefficients:
##           ar1      ar2     ma1     ma2     sma1
##       -1.1549  -0.9934  1.1261  0.9605  -0.6959
## s.e.   0.0062   0.0083  0.0227  0.0291   0.0298
##
## sigma^2 = 16.62:  log likelihood = -1694.1
## AIC=3400.2   AICc=3400.34   BIC=3426.57
```

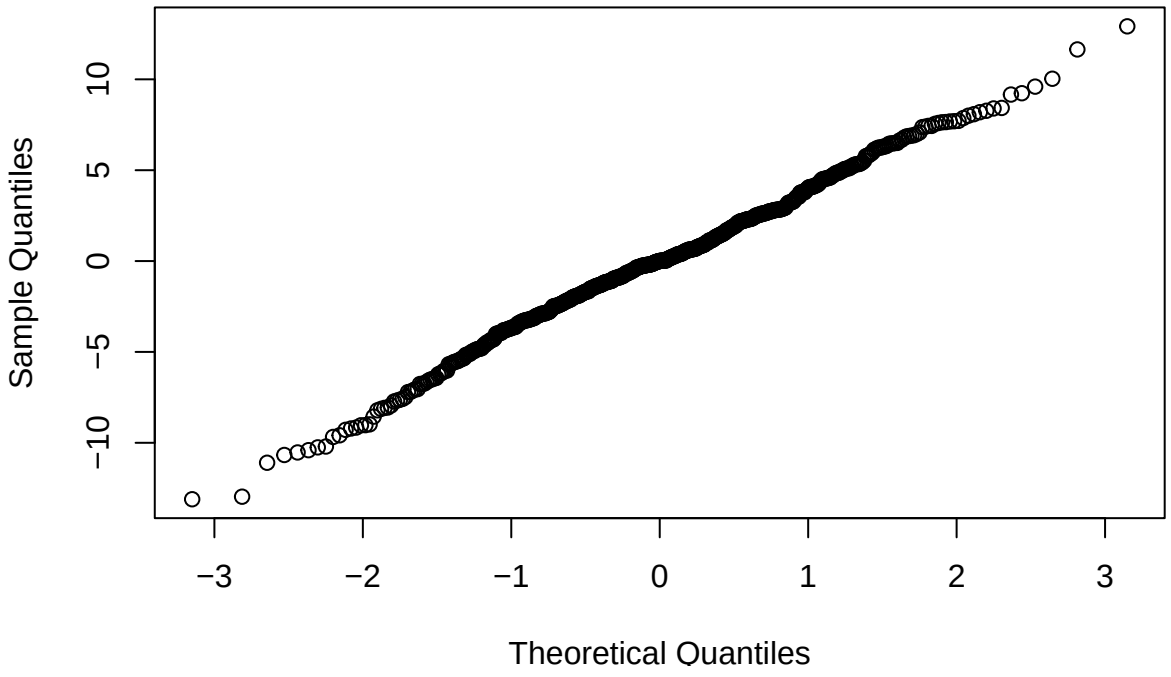
```
## Series: series
## ARIMA(2,0,4)(0,1,1)[12]
##
## Coefficients:
```

```
## Warning in sqrt(diag(x$var.coef)): NaNs produced
```

```
##         ar1    ar2     ma1      ma2    ma3    ma4     sma1
##       0.506  0.404  0.2783  -0.0449  0.197  0.057  -0.6859
## s.e.    NaN    NaN     NaN      NaN  0.040    NaN   0.0315
##
## sigma^2 = 16.92:  log likelihood = -1700.95
## AIC=3417.91   AICc=3418.15   BIC=3453.08
```


Histogram of residuals(arima1)
Frequency
0
20
40
60
80
120
−15
−10
−5
0
5
10
15
residuals(arima1)


Series residuals(arima1)
ACF
0.10
0.05
0.00
−0.10
0.0
0.5
1.0
1.5
2.0
Lag




# a.According to the ADF test, for the original time series, p value is larger than 0,05, we fail to reject the null hypothesis. The original series is not stationaryl;In order to make the series stationary, the first-order difference is done for the series, and the ADF test is performed again. p value is smaller than 0,05, we reject the null hypothesis. The original series is stationary. # b.According to the acf and pacf, there is an obvious spike on lag 12.Sothe 12th order seasonal difference will be done for the sequence. # c.Draw the ACF and PACF diagrams once again.I assume the model is ARIMA(2,1,2)(0,1,1)[12]. # d.According to the auto.arima(), the model is ARIMA(2,0,4)(0,1,1)[12], we can compare the 2 models. # e.According to AIC, ARIMA(2,1,2)(0,1,1)[12] has smaller AIC,BIC and AICc, so we choose model ARIMA(2,1,2)(0,1,1)[12]. # f.ased on the histogram, the autocorrelation plot and the QQ plot, we can see that the residuals of the model are not autocorrelated and approximately obey a normal distribution.The model is appropriate.

- **QUESTION 4** Forecast the unadjusted ice cream production series through December of 2025. Comment on the validity of your forecast.

```
##          Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
## Jan 2023       90.75535  85.53083  95.97986  82.76513  98.74556
## Feb 2023      103.99996  96.71680 111.28313  92.86132 115.13860
## Mar 2023      113.83417 104.95663 122.71172 100.25714 127.41120
## Apr 2023      115.46694 105.16678 125.76709  99.71421 131.21967
## May 2023      112.99209 101.51989 124.46429  95.44687 130.53731
## Jun 2023      121.87318 109.31697 134.42938 102.67012 141.07623
## Jul 2023      114.33519 100.73943 127.93095  93.54227 135.12812
## Aug 2023      113.02669  98.52863 127.52476  90.85382 135.19957
## Sep 2023      104.56561  89.18600 119.94522  81.04453 128.08669
## Oct 2023       97.78484  81.55063 114.01905  72.95676 122.61293
## Nov 2023       86.53496  69.54076 103.52916  60.54458 112.52535
## Dec 2023       81.88975  64.12966  99.64984  54.72803 109.05146
## Jan 2024       90.41116  71.40639 109.41593  61.34587 119.47645
## Feb 2024      104.21278  84.10552 124.32004  73.46138 134.96418
## Mar 2024      114.74287  93.54031 135.94542  82.31636 147.16938
## Apr 2024      115.01865  92.77102 137.26628  80.99384 149.04346
```

```
## May 2024      113.41965  90.22339 136.61592  77.94403 148.89528
## Jun 2024      122.63728  98.47791 146.79665  85.68872 159.58585
## Jul 2024      113.84056  88.76712 138.91400  75.49404 152.18708
## Aug 2024      113.65142  87.73047 139.57236  74.00876 153.29408
## Sep 2024      105.14806  78.35663 131.93950  64.17410 146.12203
## Oct 2024       97.30414  69.69332 124.91496  55.07704 139.53124
## Nov 2024       87.32406  58.93760 115.71051  43.91072 130.73740
## Dec 2024       82.26853  53.08208 111.45499  37.63171 126.90536
## Jan 2025       90.00237  59.67659 120.32815  43.62309 136.38165
## Feb 2025      105.12114  73.71759 136.52470  57.09355 153.14874
## Mar 2025      114.91248  82.42182 147.40313  65.22231 164.60264
## Apr 2025      114.73296  81.22198 148.24395  63.48233 165.98359
## May 2025      114.39366  79.89932 148.88799  61.63913 167.14819
## Jun 2025      122.60880  87.12108 158.09652  68.33501 176.88259
## Jul 2025      113.71843  77.29986 150.13700  58.02104 169.41583
## Aug 2025      114.63331  77.30237 151.96426  57.54056 171.72607
## Sep 2025      104.94798  66.69820 143.19777  46.44999 163.44598
## Oct 2025       97.37234  58.26158 136.48310  37.55760 157.18709
## Nov 2025       88.25661  48.28923 128.22399  27.13178 149.38144
## Dec 2025       81.93636  41.11185 122.76087  19.50066 144.37206
```

#According to the question 3, we already know the residuals are not autocorrelated and approximately obey a normal distribution, so the prediction is valid.