Project

Nagarjuna sricharan Maram & Shardul Jani

6 December 2017

```
#install.packages("tseries")
library(tseries)
#install.packages("readxl")
library(readxl)
#install.packages("forecast")
library(forecast)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("tseries")
library(tseries)
#install.packages("Metrics")
library(Metrics)
#install.packages("TSA")
library(TSA)
```

Department 94

```
D94 <- read_excel("C:/Users/Charan/Desktop/Fall 17/Data Mining/Project/D94.xlsx")
```

```
D94$Month <- as.Date(D94$Month)
```

```
d94 <- D94[,-3]
```

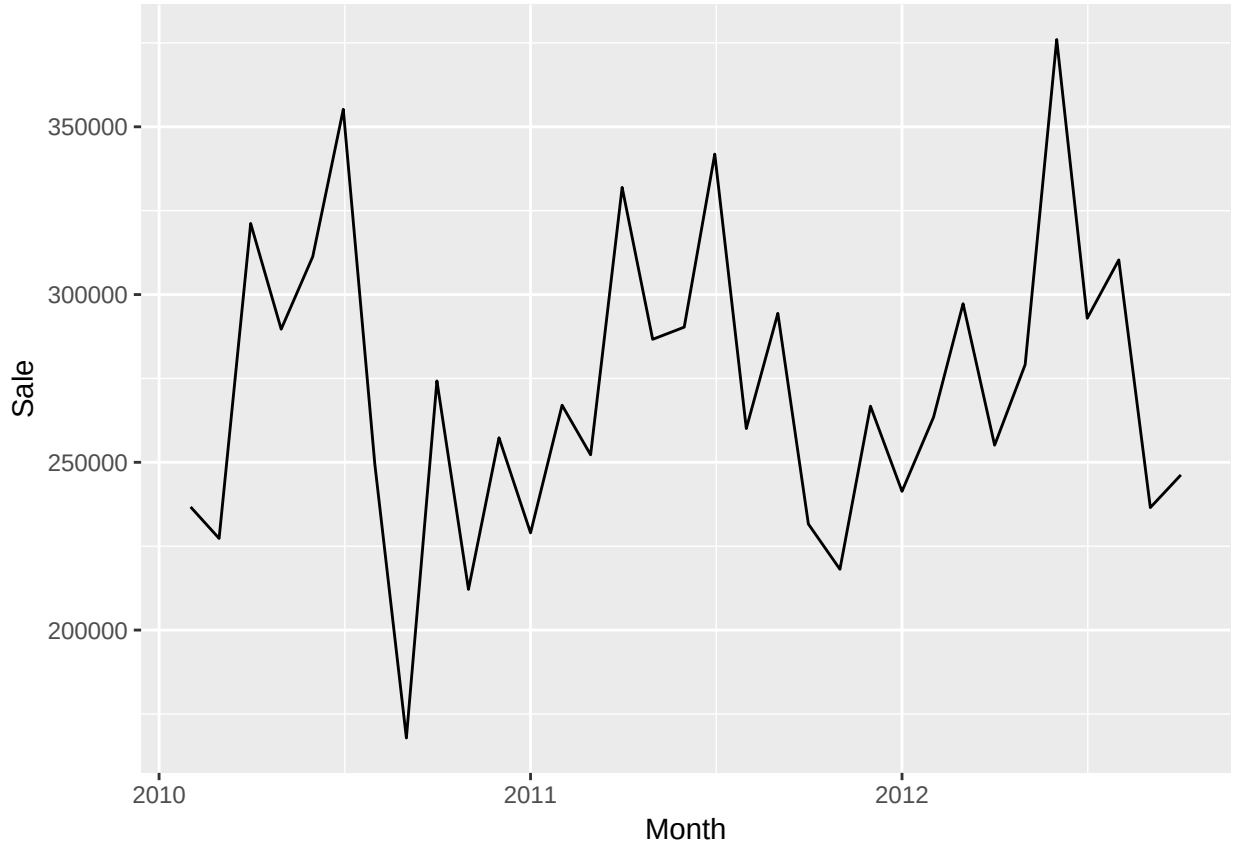
```
covariate94 <- D94[1:25,3]
```

```
d94.train <- d94[1:25,1:2]
d94.test <- d94[26:33,1:2]
```

Arima

Time series plot

```
library(ggplot2)
ggplot(D94 , aes(Month , Sale)) + geom_line()
```

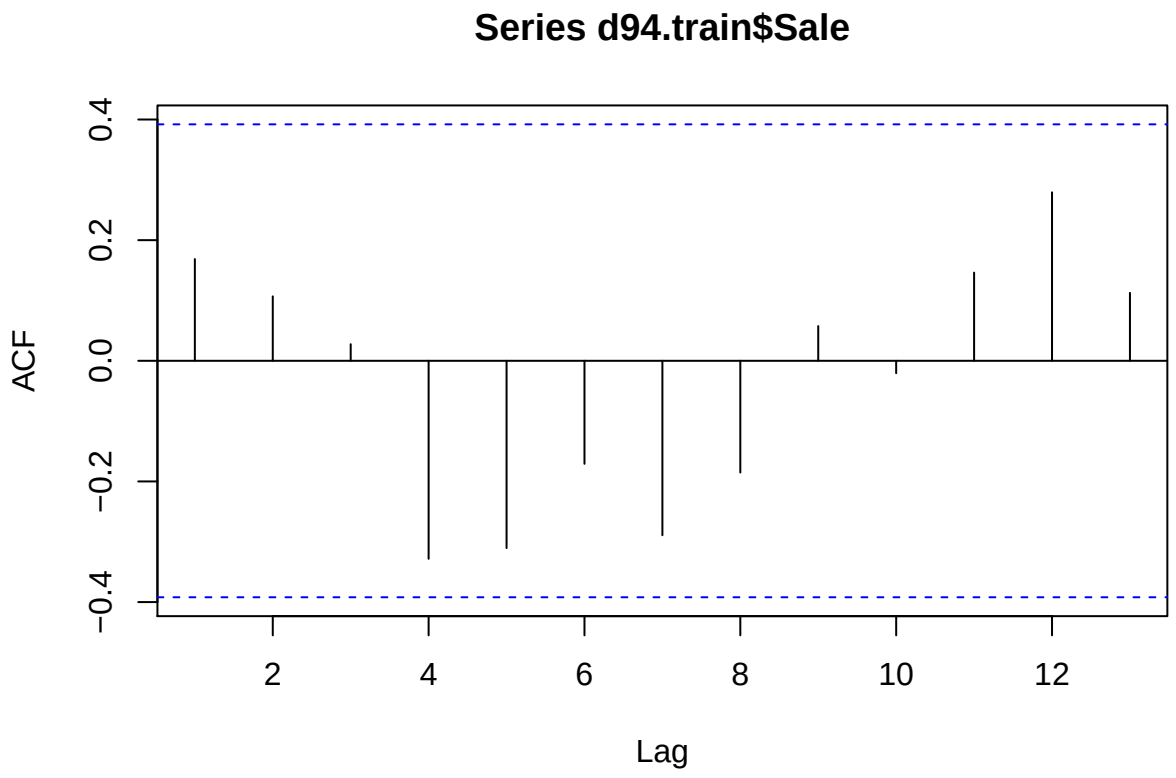


Test for stationary

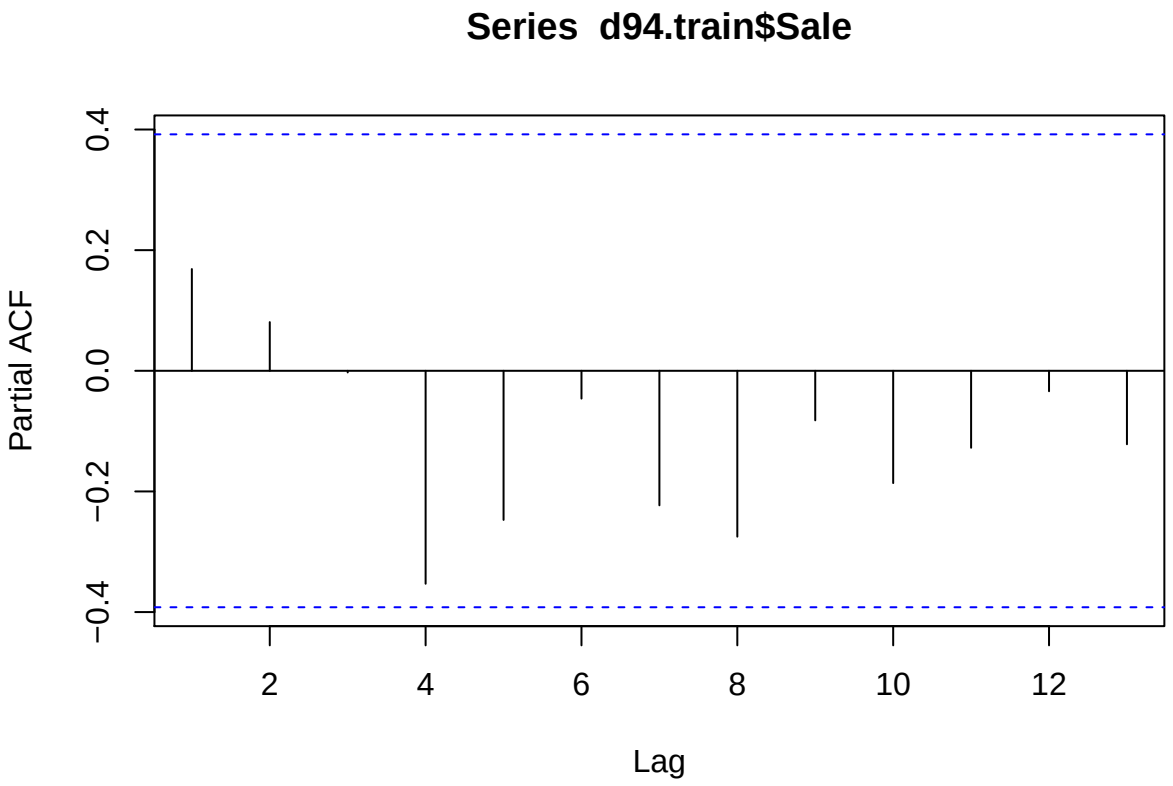
```
adf.test(diff(diff(d94.train$Sale)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(d94.train$Sale))
## Dickey-Fuller = -4.3505, Lag order = 2, p-value = 0.01103
## alternative hypothesis: stationary
```

```
acf(d94.train$Sale)
```



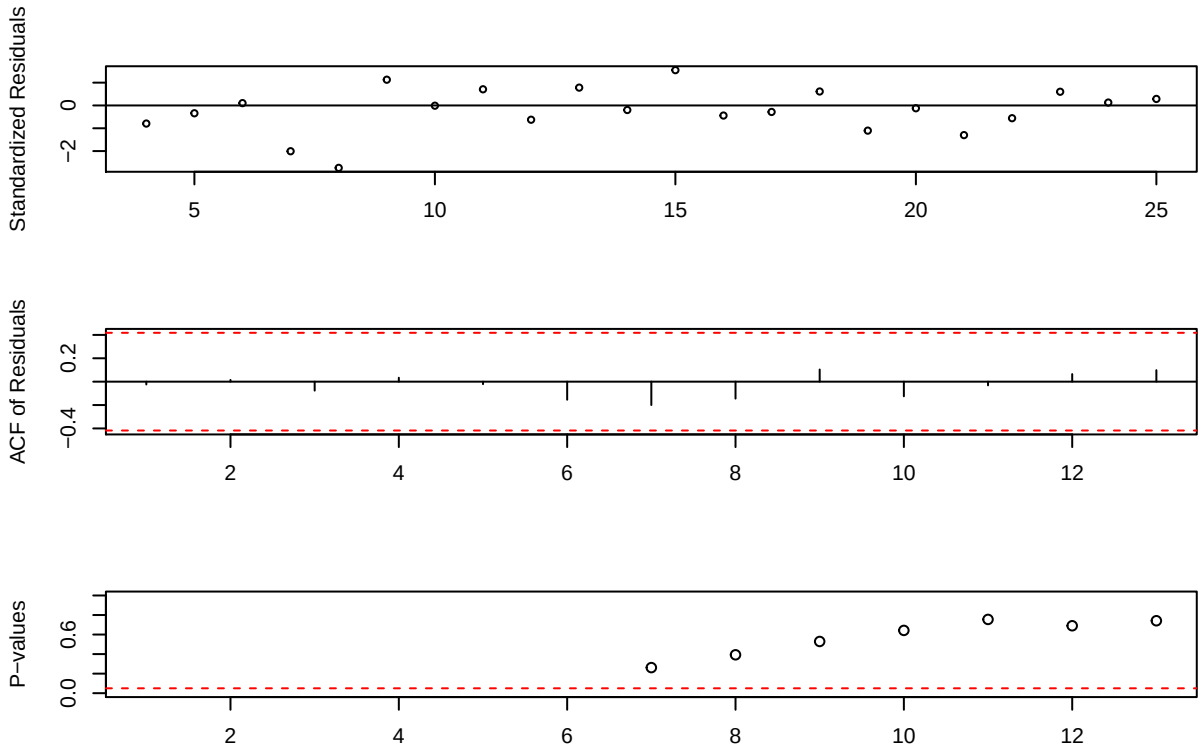

```
pacf(d94.train$Sale)
```



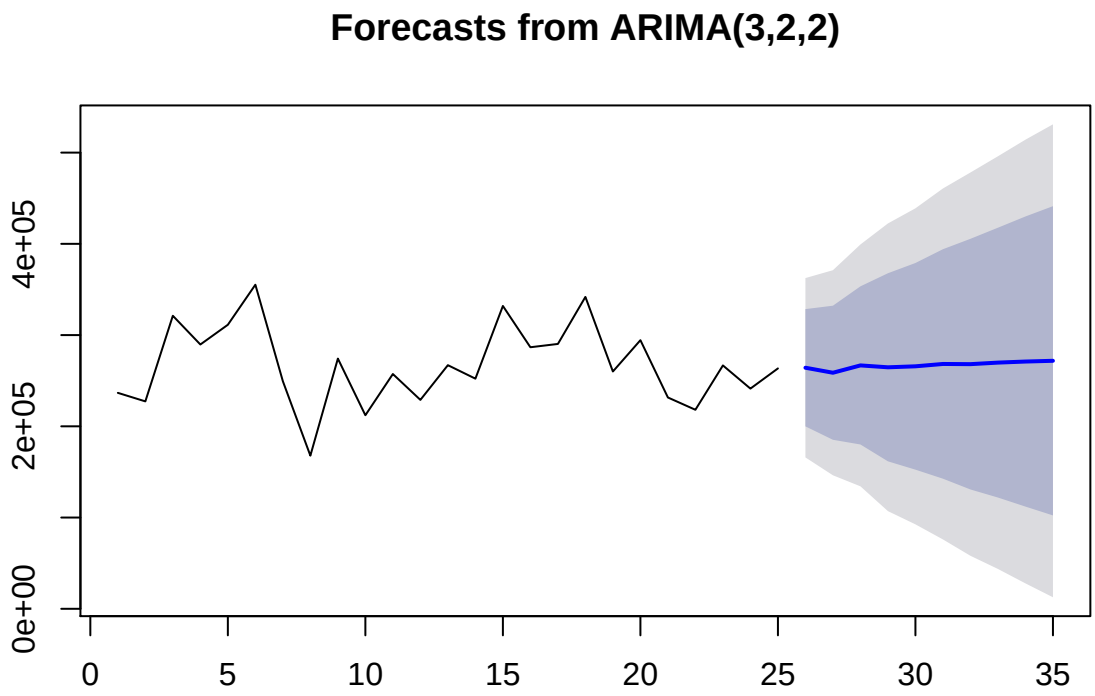

```
Model94.A <- stats::arima(d94.train$Sale , order = c(3,2,2))
Model94.A
```

```
##
## Call:
## stats::arima(x = d94.train$Sale, order = c(3, 2, 2))
##
## Coefficients:
##           ar1      ar2     ar3      ma1      ma2
##       -0.5924  -0.1467  0.1411  -0.8922  -0.1078
## s.e.   0.7003   0.4515  0.2636   0.6826   0.6691
##
## sigma^2 estimated as 2.412e+09:  log likelihood = -283.2,  aic = 578.41
```

```
tsdiag(Model94.A)
```




```
predict94.1 <- forecast(Model94.A , n.ahead = 8)
plot(predict94.1)
```




```
rmse(d94.test$Sale, predict94.1$mean)
```

```
## [1] 43213.91
```

ARIMAX

```
arimax94.A <- arimax(d94.train$Sale , order =  c(3,2,2) ,xtransf =  covariate)
predict94.2 <- predict(arimax94.A , n.ahead = 8)
 rmse(d94.test$Sale , predict94.2$pred)
```

```
## [1] 47020.89
```

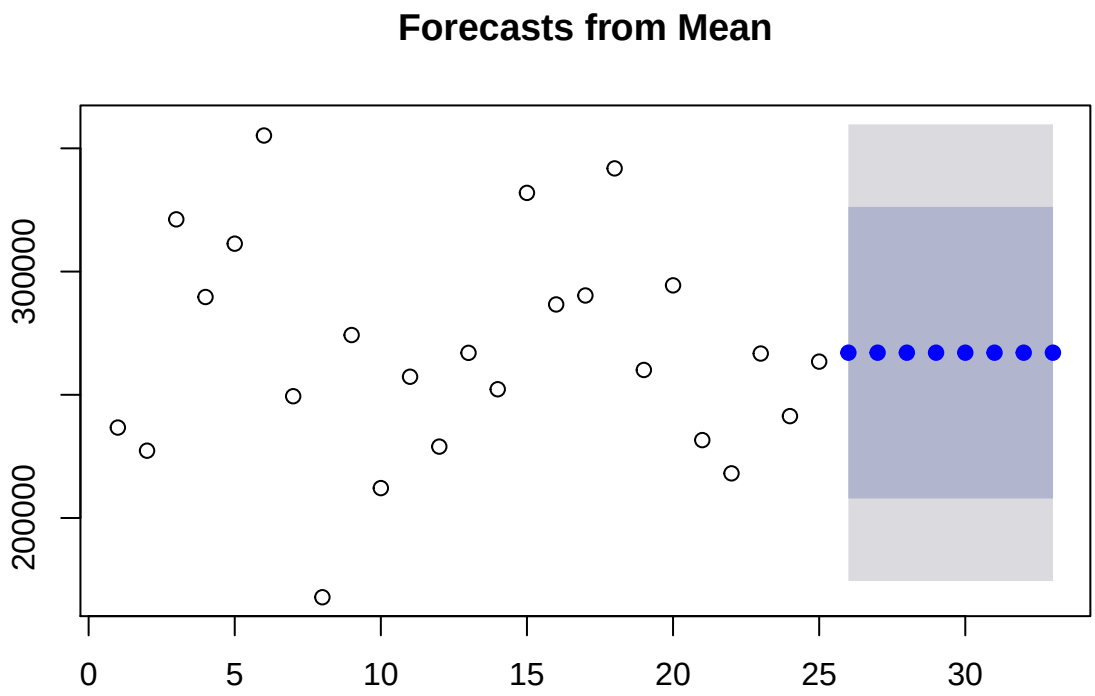
Average method

```
predict94.3 <- meanf(d94.train$Sale , h =8)

rmse(predict94.3$mean , d94.test$Sale)
```

```
## [1] 46058.15
```

```
plot(predict94.3)
```

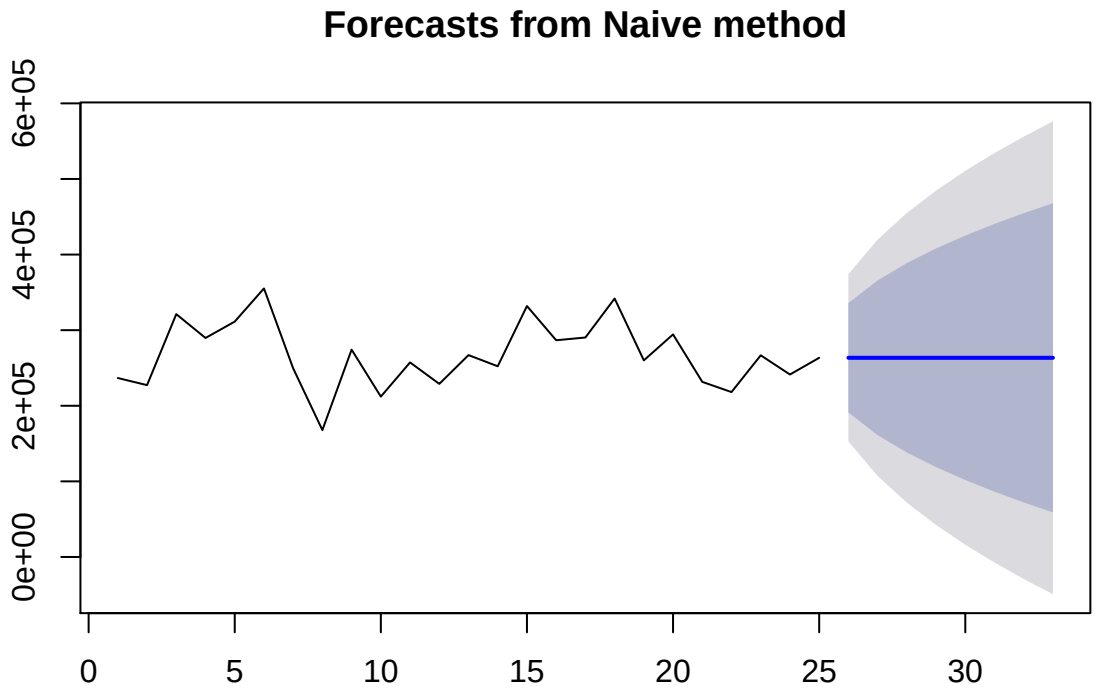


Naive method

```
predict94.4 <- naive(d94.train$Sale , h =8)
rmse(predict94.4$mean , d94.test$Sale)
```

```
## [1] 47714.02
```

```
plot(predict94.4)
```

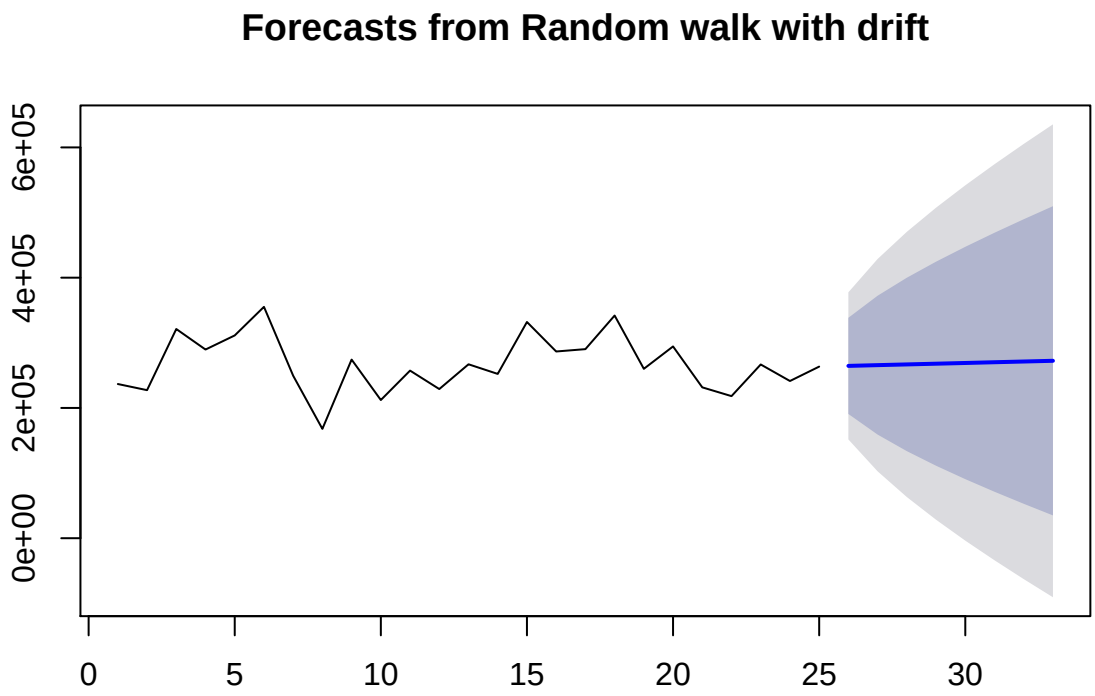


Drift method

```
predict94.5 <- rwf(d94.train$Sale , h = 8 , drift = TRUE)
rmse(predict94.5$mean , d94.test$Sale)
```

```
## [1] 46221.86
```

```
plot(predict94.5)
```



Department 92

Load Data

```
D92 <- read_excel("C:/Users/Charan/Desktop/Fall 17/Data Mining/Project/D92.xlsx")
```

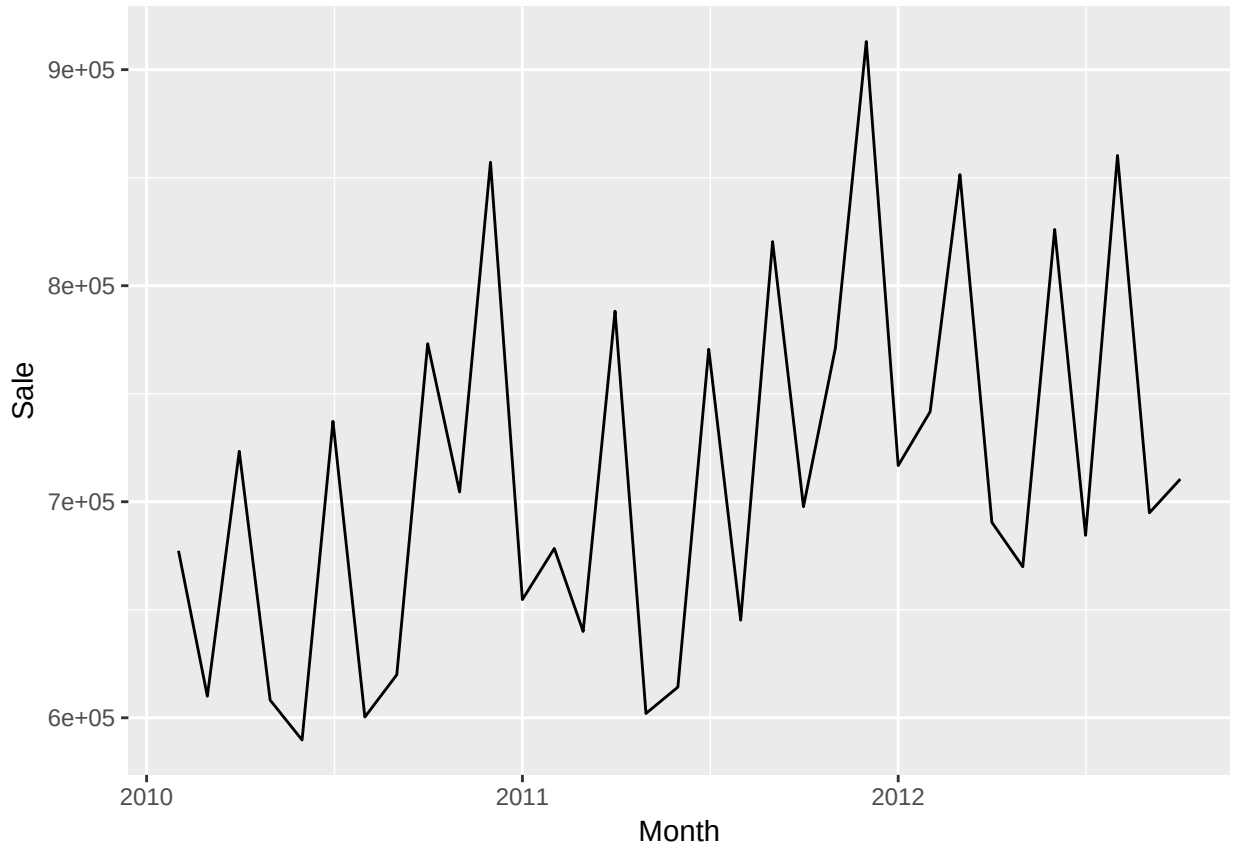
```
D92$Month <- as.Date(D92$Month)
```

```
d92 <- D92[,-3]
```

```
covariate92 <- D92[1:25,3]
```

```
d92.train <- d92[1:25,1:2]
d92.test <- d92[26:33,1:2]
```

```
library(ggplot2)
ggplot(D92 , aes(Month , Sale)) + geom_line()
```



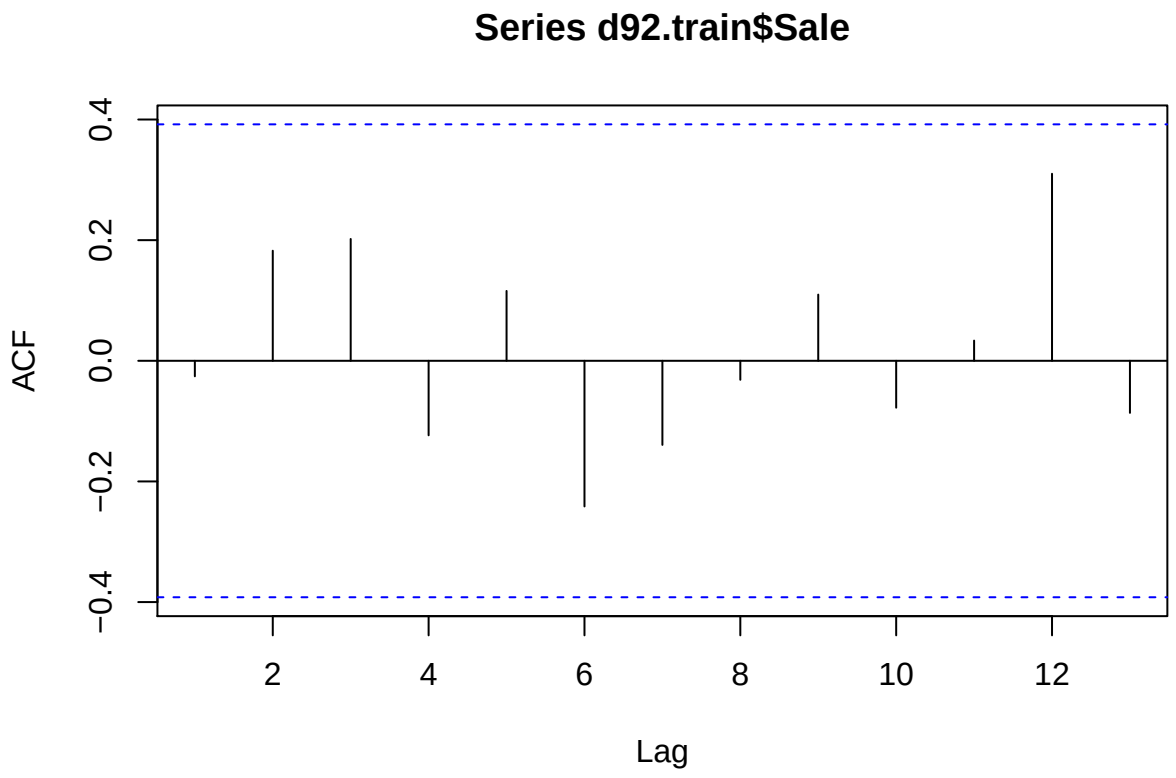
Arima

Test for stationary

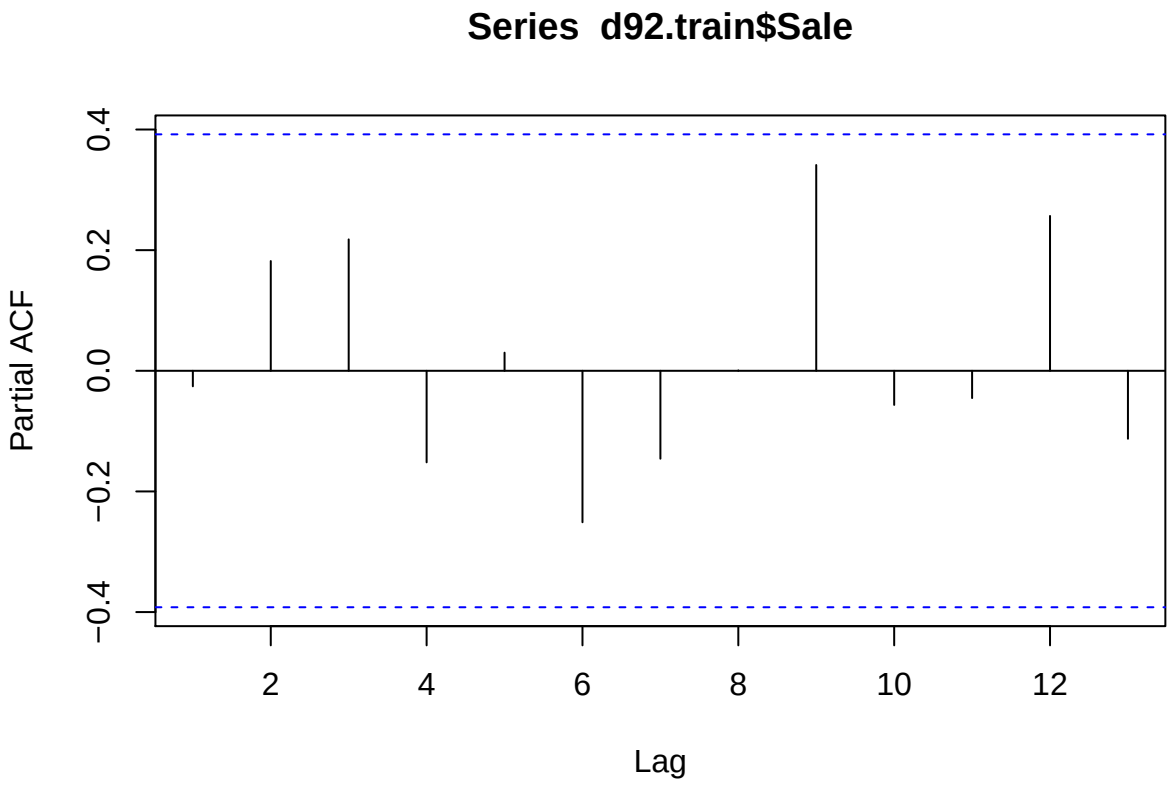
```
adf.test(diff(diff(d92.train$Sale)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(d92.train$Sale))
## Dickey-Fuller = -4.5706, Lag order = 2, p-value = 0.01
## alternative hypothesis: stationary
```

```
acf(d92.train$Sale)
```



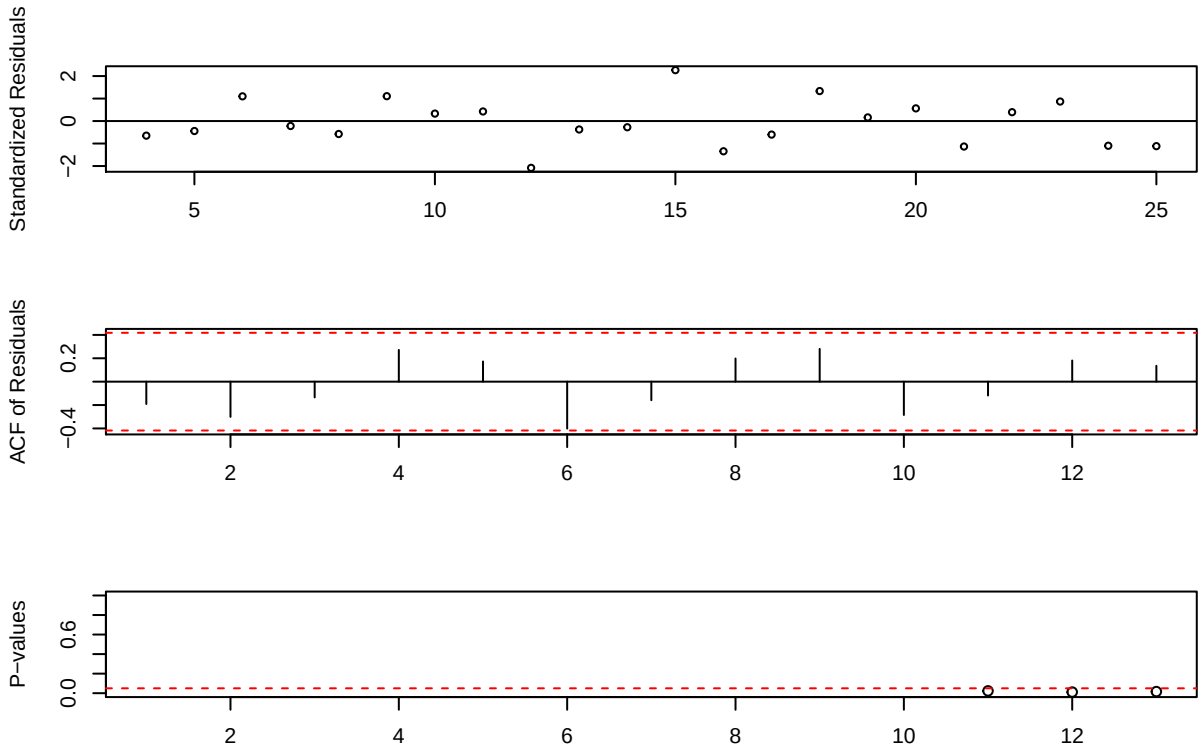

```
pacf(d92.train$Sale)
```



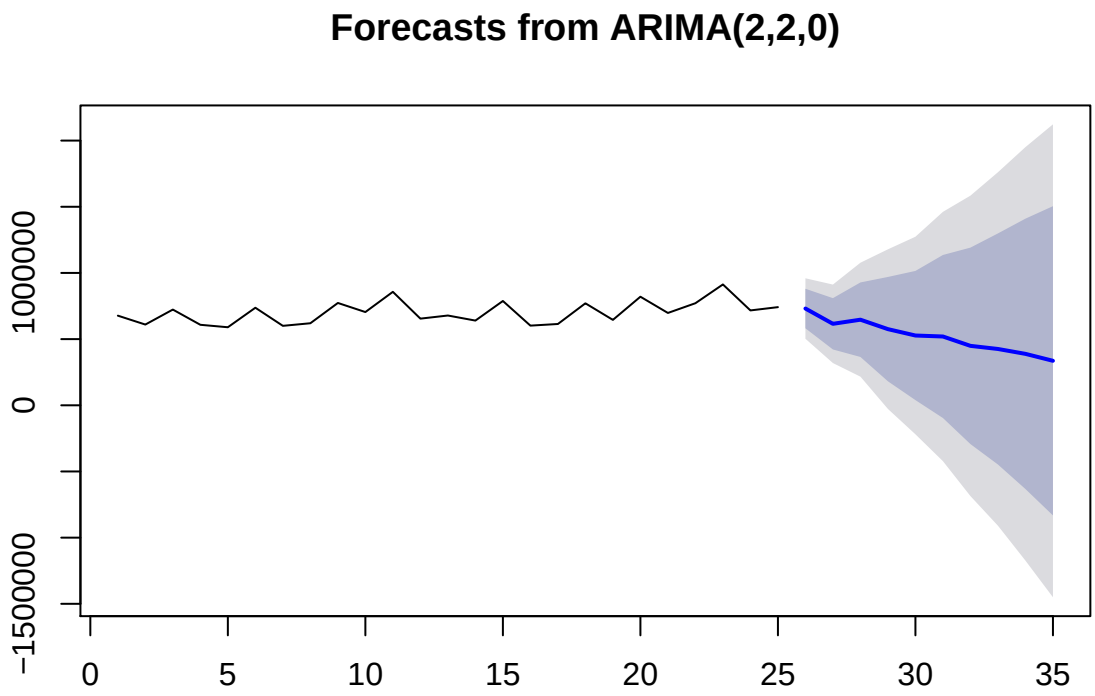

```
Model92.A <- stats::arima(d92.train$Sale , order = c(2,2,0))
Model92.A
```

```
##
## Call:
## stats::arima(x = d92.train$Sale, order = c(2, 2, 0))
##
## Coefficients:
##           ar1      ar2
##       -1.1729  -0.6625
## s.e.   0.1476   0.1509
##
## sigma^2 estimated as 1.358e+10:  log likelihood = -301.87,  aic = 609.74
```

```
tsdiag(Model92.A)
```




```
predict92.1 <- forecast(Model92.A , n.ahead = 8)
plot(predict92.1)
```




```
rmse(d92.test$Sale, predict92.1$mean)
```

```
## [1] 265584.2
```

Arimax

```
arimax92.A <- arimax(d92.train$Sale , order =  c(2,2,0) ,xtransf =  covariate92)
predict92.2 <-predict(arimax92.A , n.ahead = 8)
 rmse(d92.test$Sale , predict92.2$pred)
```

```
## [1] 213797.6
```

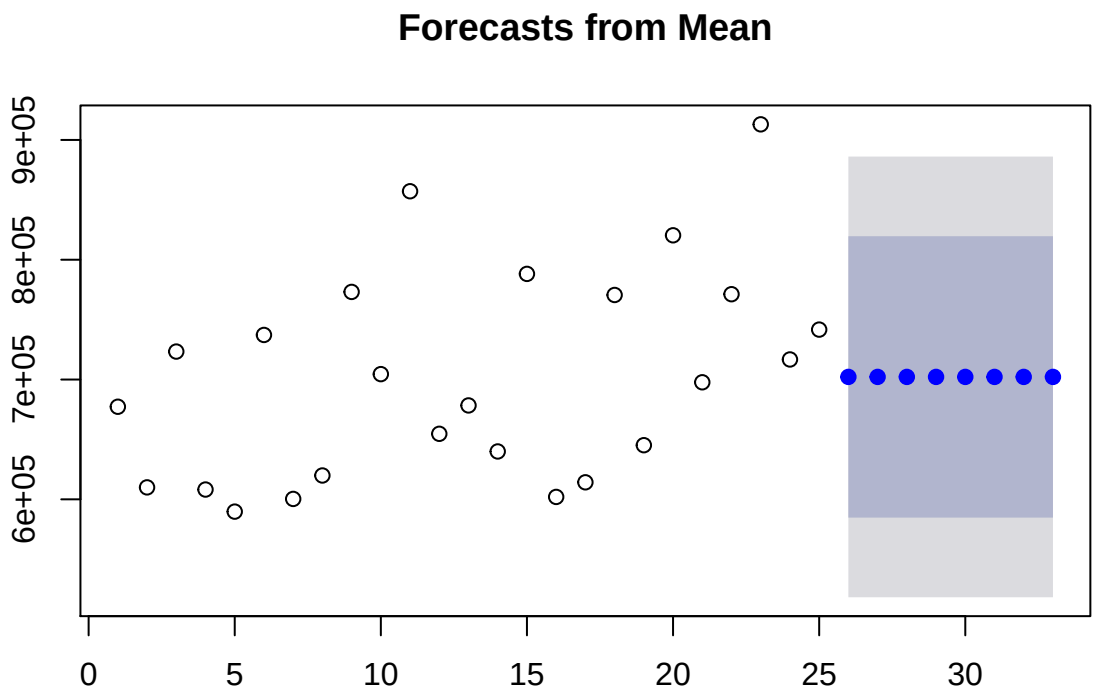
Average

```
predict92.3 <- meanf(d92.train$Sale , h =8)

rmse(predict92.3$mean , d92.test$Sale)
```

```
## [1] 89582.52
```

```
plot(predict92.3)
```

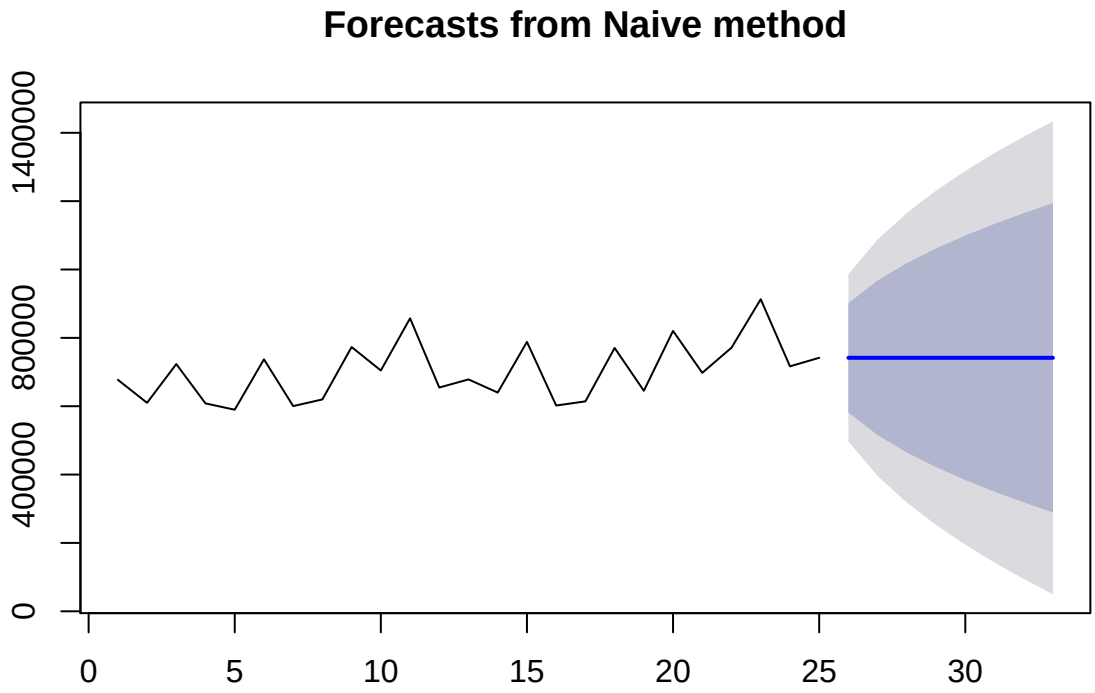


Naive

```
predict92.4 <- naive(d92.train$Sale , h =8)
rmse(predict92.4$mean , d92.test$Sale)
```

```
## [1] 76996.63
```

```
plot(predict92.4)
```

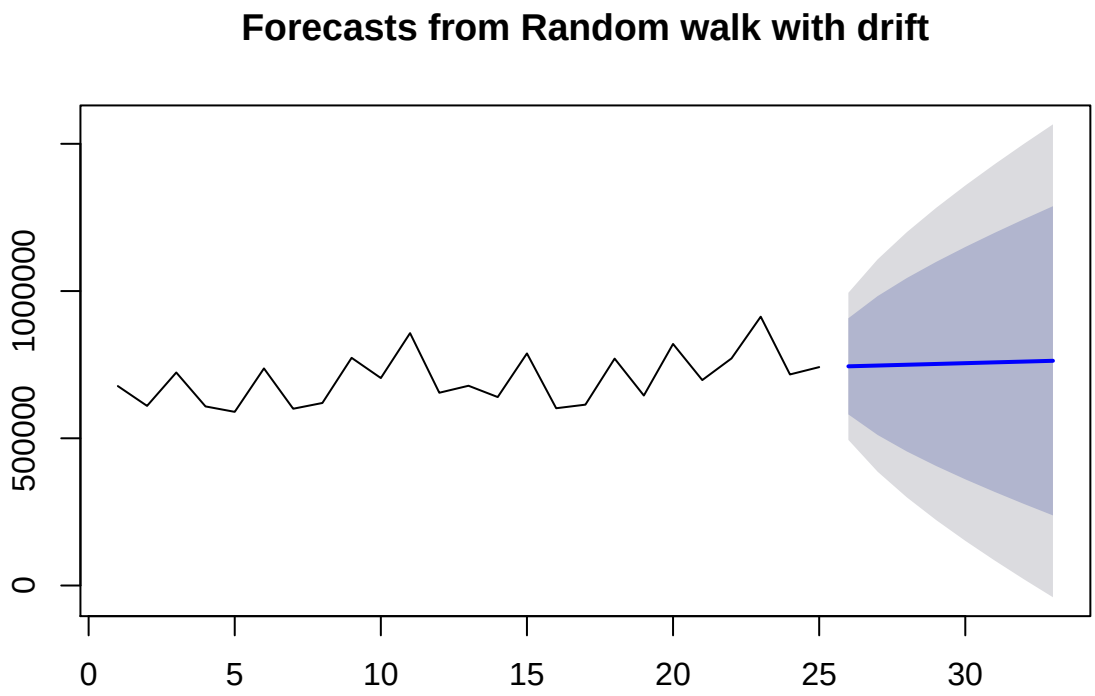


Drift Method

```
predict92.5 <- rwf(d92.train$Sale , h = 8 , drift = TRUE)
rmse(predict92.5$mean , d92.test$Sale)
```

```
## [1] 78276.65
```

```
plot(predict92.5)
```



Department 38

Loading dataset

```
D38 <- read_excel("C:/Users/Charan/Desktop/Fall 17/Data Mining/Project/D38.xlsx")
```

```
D38$Month <- as.Date(D38$Month)
```

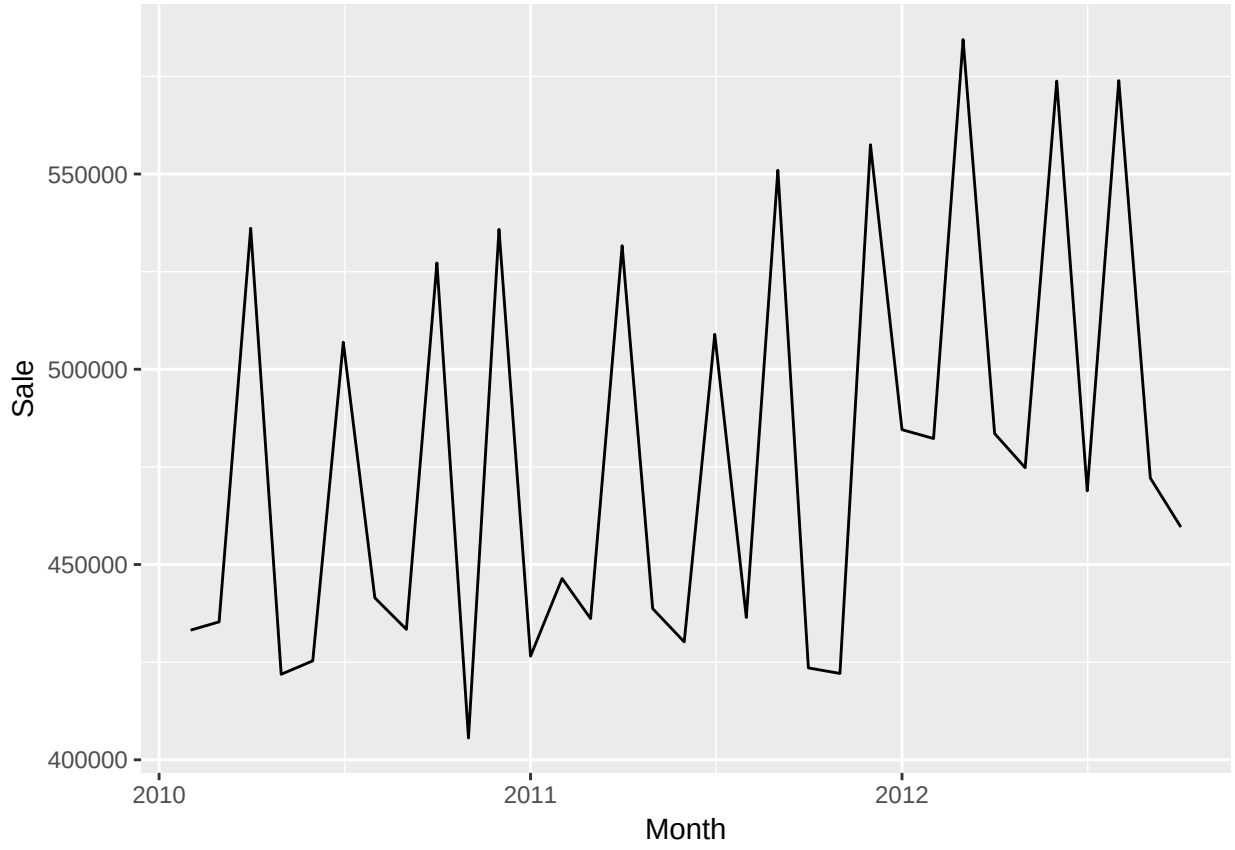
```
d38 <- D38[,-3]
```

```
covariate38 <- D38[1:25,3]
```

```
d38.train <- d38[1:25,1:2]
d38.test <- d38[26:33,1:2]
```

Time series plot

```
library(ggplot2)
ggplot(D38 , aes(Month , Sale)) + geom_line()
```



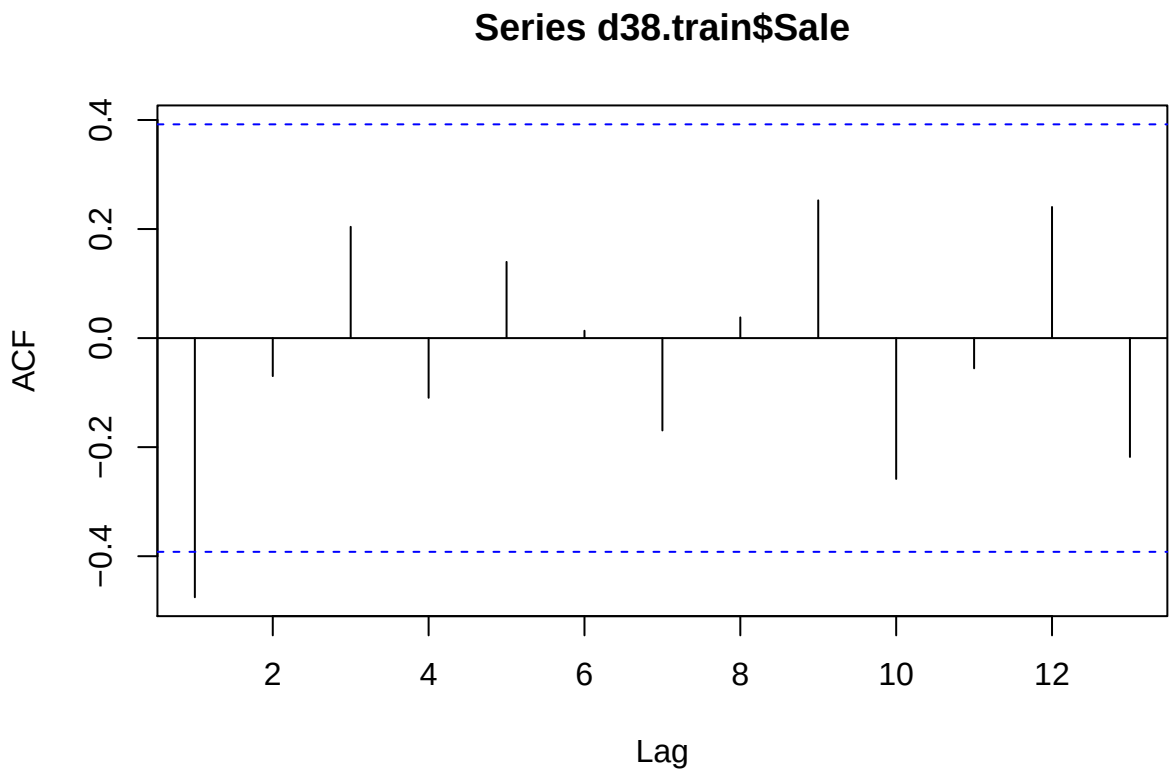
Arima

Test for stationary

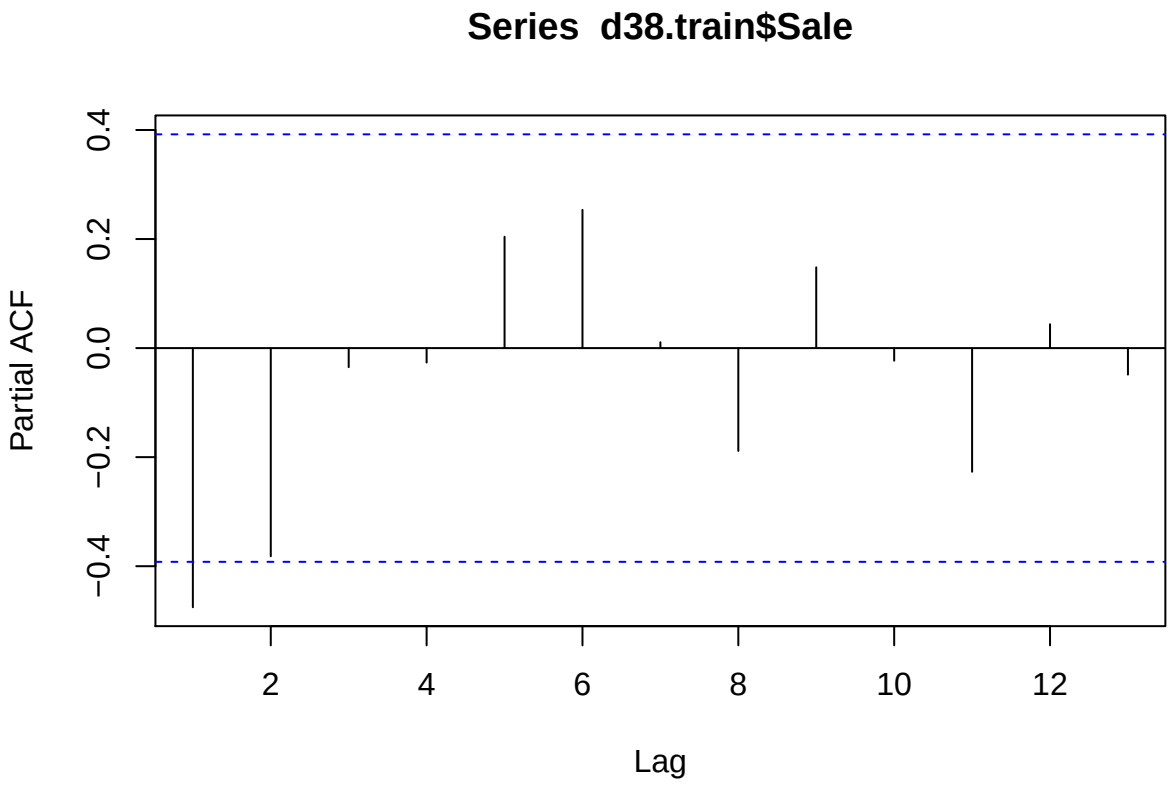
```
adf.test(d38.train$Sale)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  d38.train$Sale
## Dickey-Fuller = -4.3213, Lag order = 2, p-value = 0.01205
## alternative hypothesis: stationary
```

```
acf(d38.train$Sale)
```



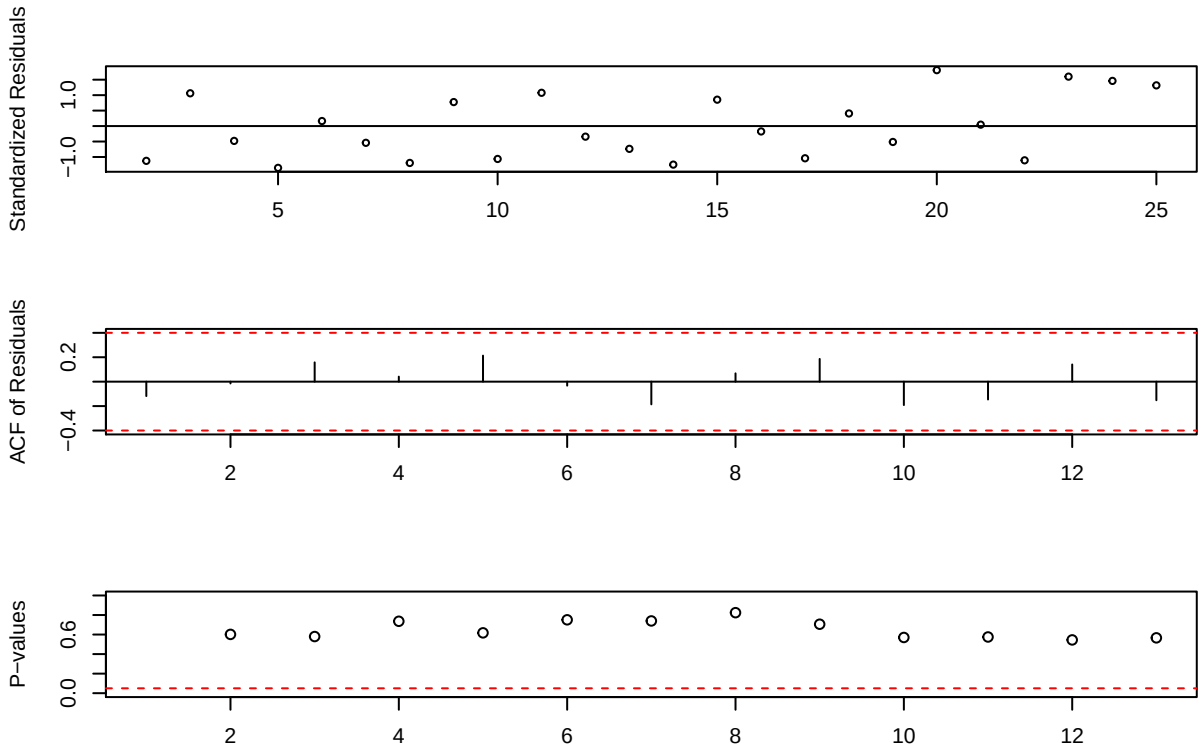

```
pacf(d38.train$Sale)
```



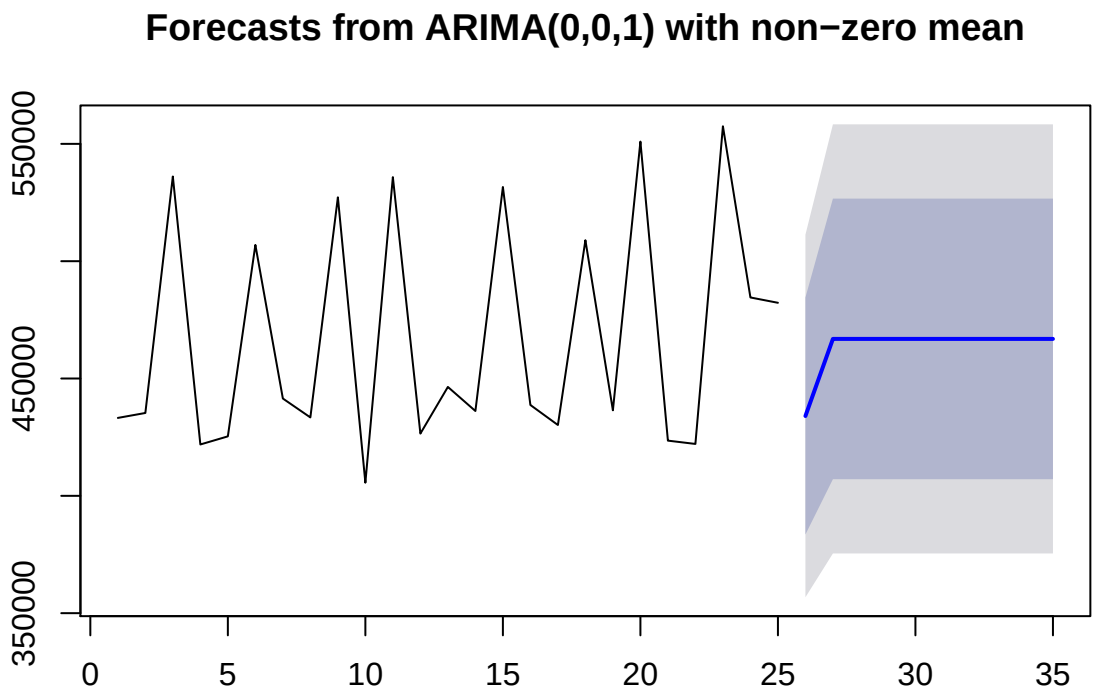

```
Model38.A <- stats::arima(d38.train$Sale , order = c(0,0,1))
Model38.A
```

```
##
## Call:
## stats::arima(x = d38.train$Sale, order = c(0, 0, 1))
##
## Coefficients:
##           ma1   intercept
##       -0.6337  466880.737
## s.e.   0.1555    3129.822
##
## sigma^2 estimated as 1.553e+09:  log likelihood = -300.27,  aic = 606.54
```

```
tsdiag(Model38.A)
```




```
predict38.1 <- forecast(Model38.A , n.ahead  = 8)
plot(predict38.1)
```



#

```
rmse(d38.test$Sale, predict38.1$mean)
```

```
## [1] 77480.53
```

Arimax

```
arimax38.A <- arimax(d38.train$Sale , order =  c(0,0,1) ,xtransf =  covariate38)
predict38.2 <-predict(arimax38.A , n.ahead = 8)
rmse(d38.test$Sale , predict38.2$pred)
```

```
## [1] 75775.32
```

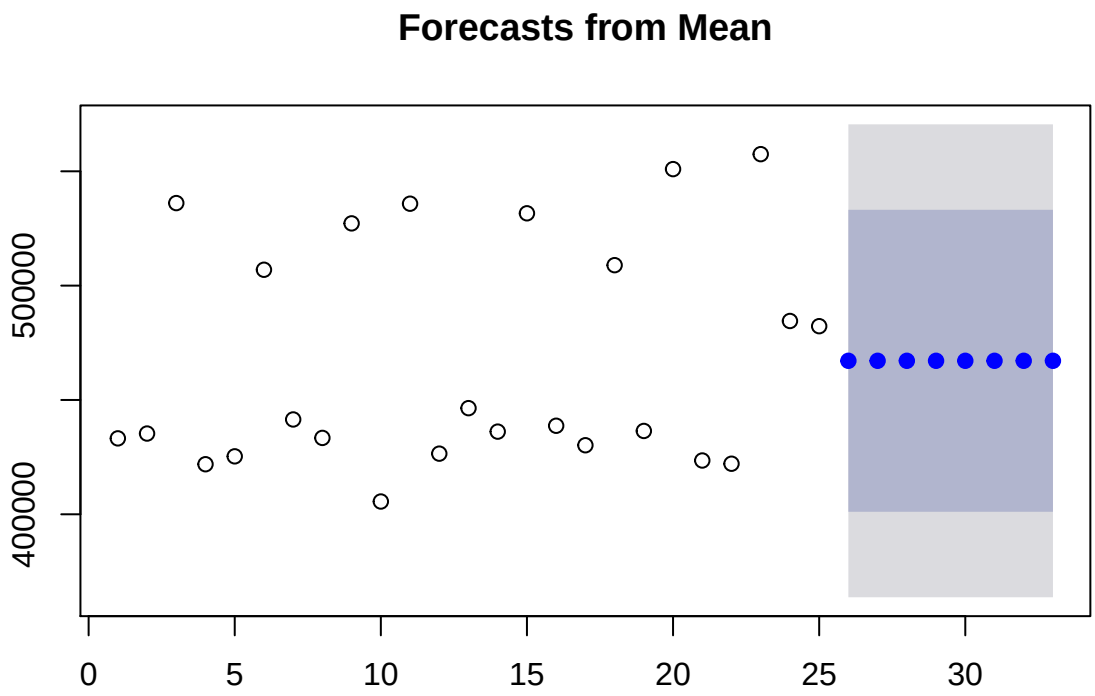
Average method

```
predict38.3 <- meanf(d38.train$Sale , h =8)

rmse(predict38.3$mean , d38.test$Sale)
```

```
## [1] 67963.83
```

```
plot(predict38.3)
```

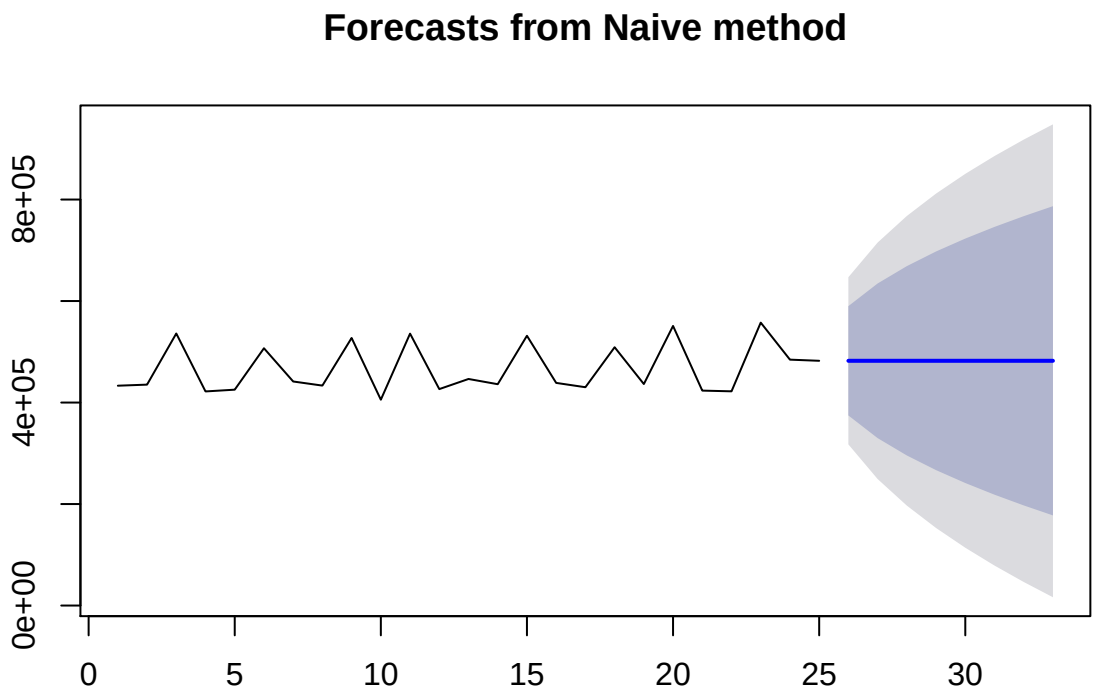


Naive

```
predict38.4 <- naive(d38.train$Sale , h =8)
 rmse(predict38.4$mean , d38.test$Sale)
```

```
## [1] 59231.87
```

```
plot(predict38.4)
```

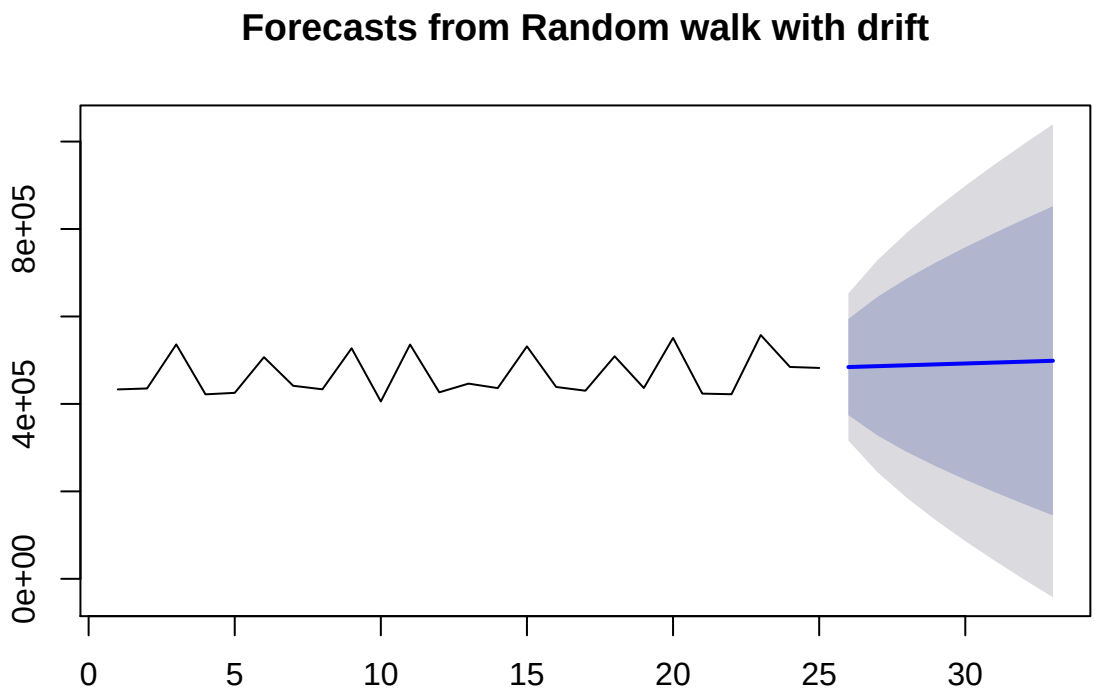


Drift method

```
predict38.5 <- rwf(d38.train$Sale , h = 8 , drift = TRUE)
rmse(predict38.5$mean , d38.test$Sale)
```

```
## [1] 57165.47
```

```
plot(predict38.5)
```



Department 95

```
D95 <- read_excel("C:/Users/Charan/Desktop/Fall 17/Data Mining/Project/D 95.xlsx")
```

```
D95$Month <- as.Date(D95$Month)
```

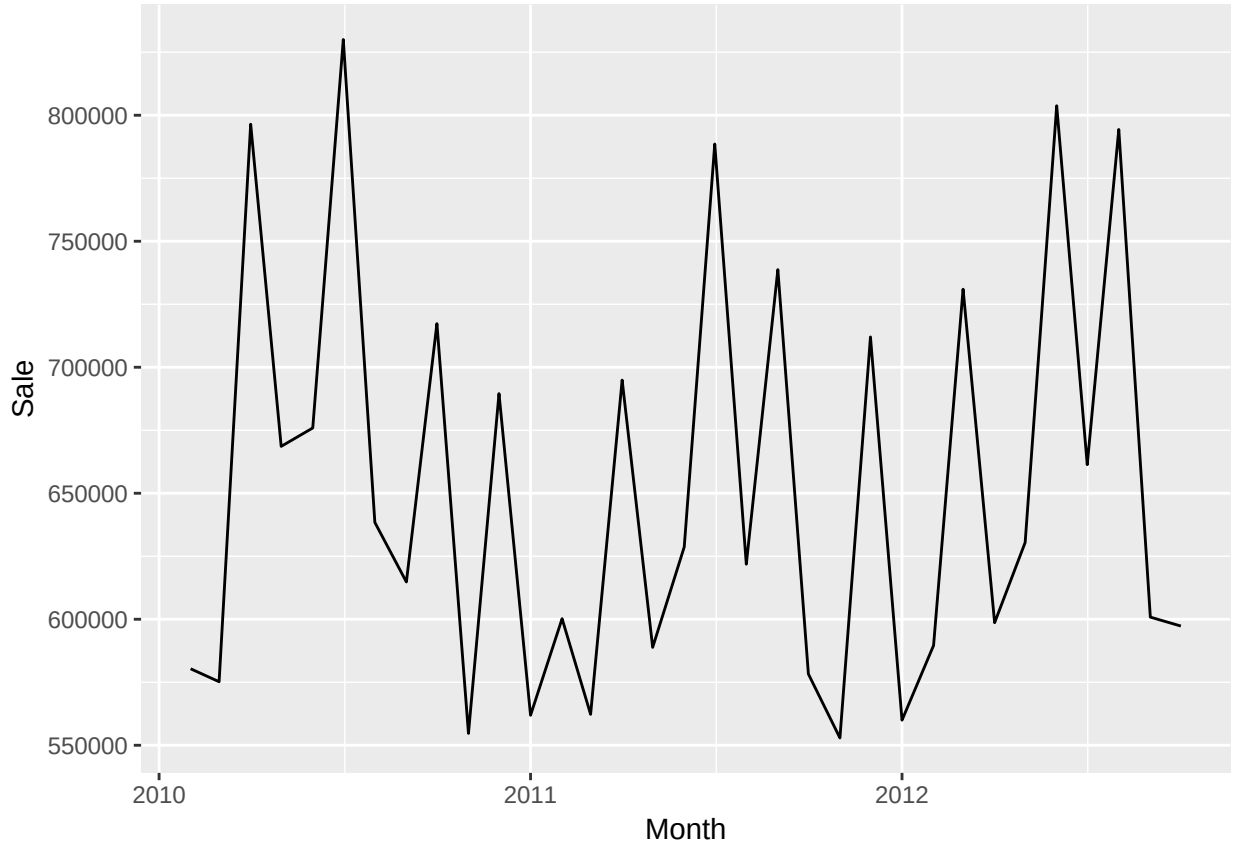
```
d95 <- D95[,-3]
```

```
covariate95 <- D95[1:25,3]
```

```
d95.train <- d95[1:25,1:2]
d95.test <- d95[26:33,1:2]
```

time series plot

```
library(ggplot2)
ggplot(D95 , aes(Month , Sale)) + geom_line()
```



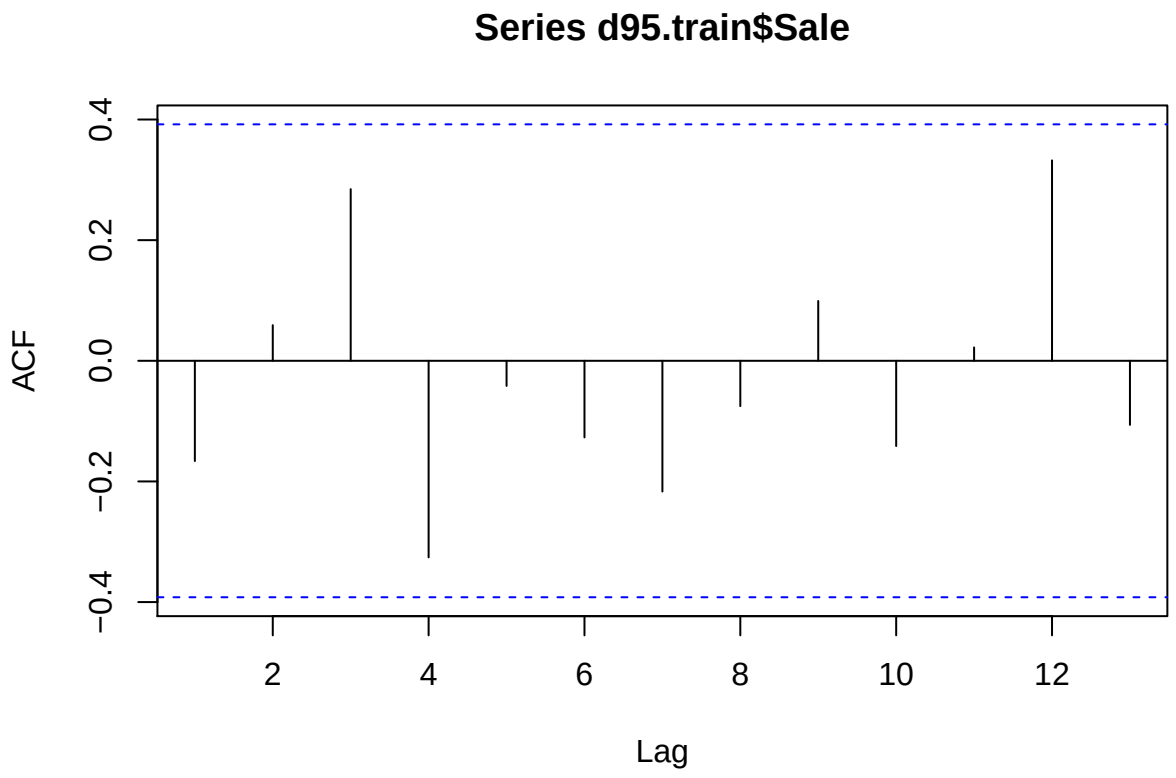
Arima

test for stationary

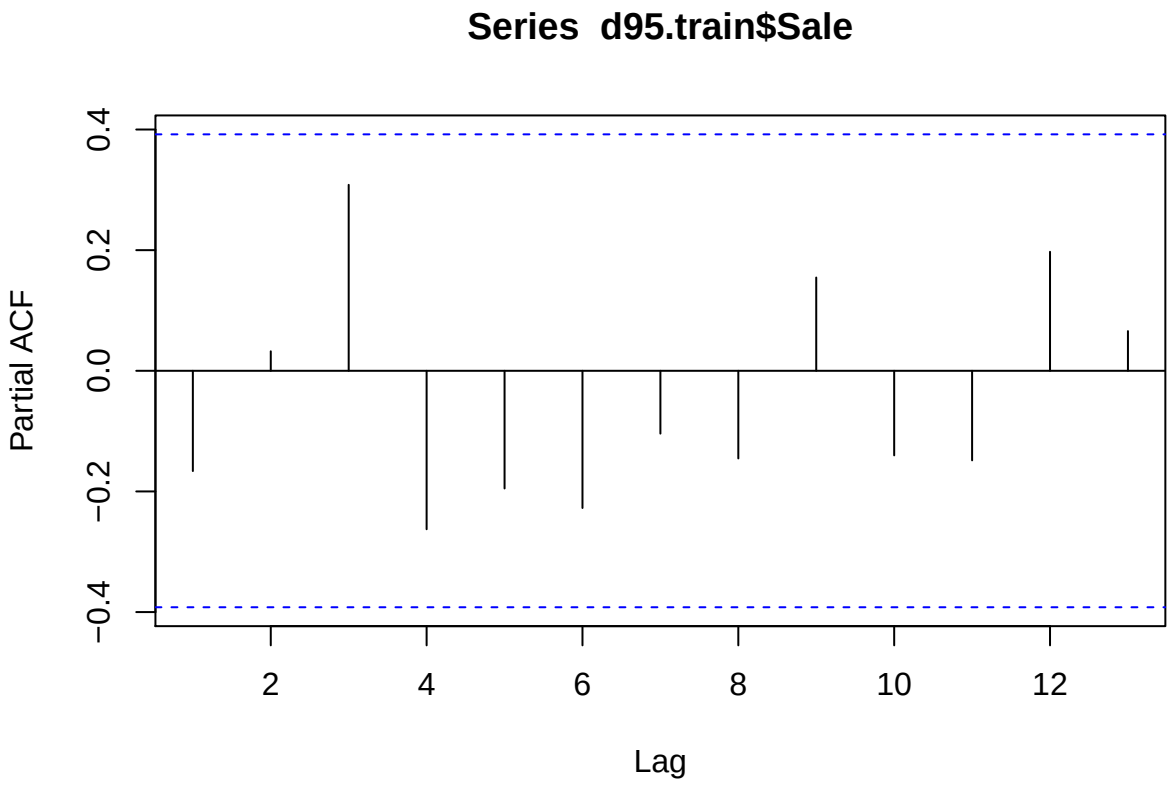
```
adf.test(diff(diff((d95.train$Sale))))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff((d95.train$Sale)))
## Dickey-Fuller = -5.3327, Lag order = 2, p-value = 0.01
## alternative hypothesis: stationary
```

```
acf(d95.train$Sale)
```



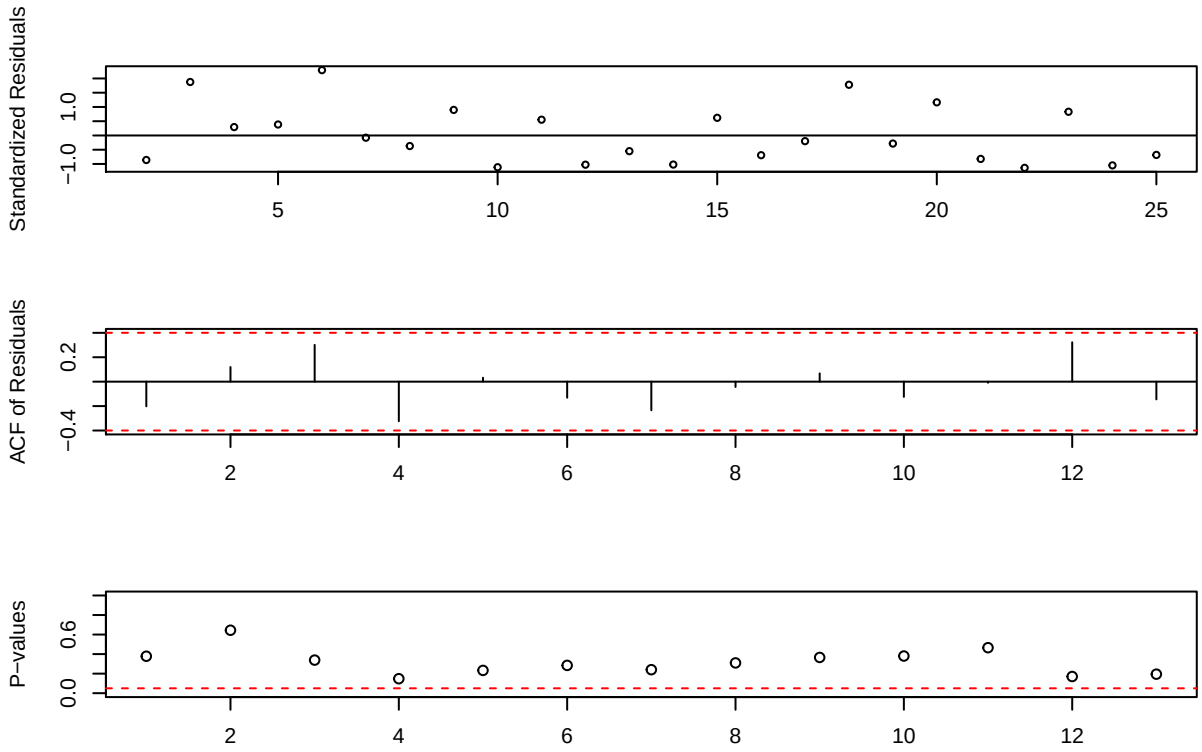

```
pacf(d95.train$Sale)
```



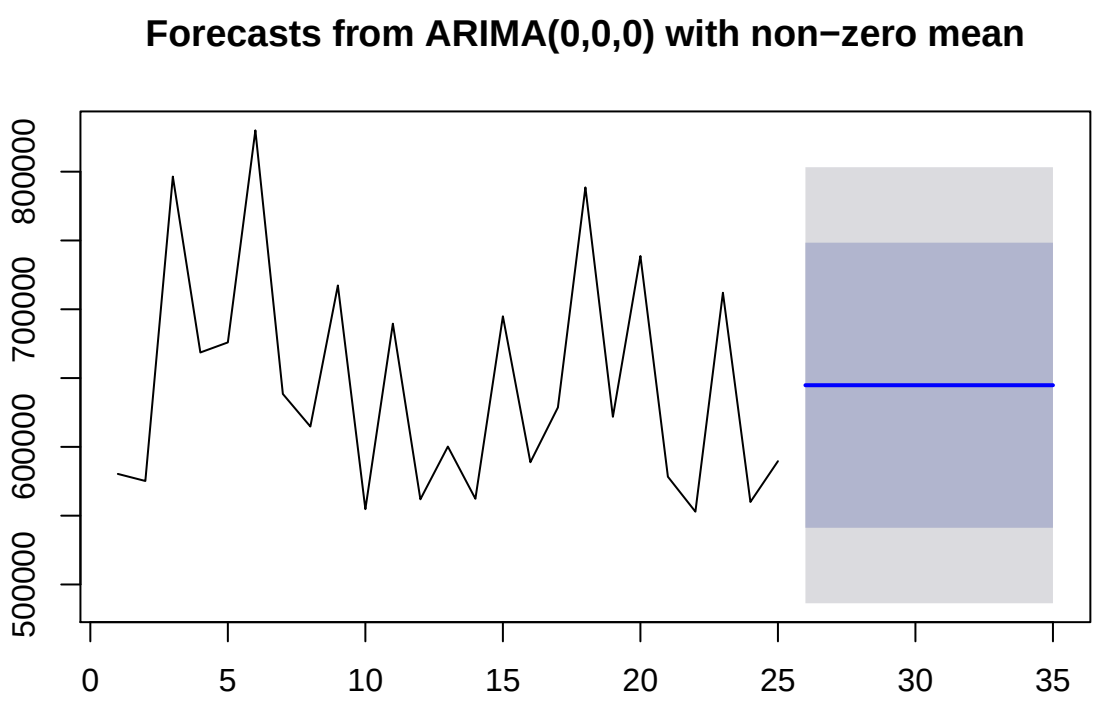

```
Model95.A <- stats::arima(d95.train$Sale, order = c(0,0,0))
Model95.A
```

```
##
## Call:
## stats::arima(x = d95.train$Sale, order = c(0, 0, 0))
##
## Coefficients:
##       intercept
##       644804.62
## s.e.   16168.21
##
## sigma^2 estimated as 6.535e+09:  log likelihood = -317.98,  aic = 639.96
```

```
tsdiag(Model95.A)
```




```
predict95.1 <- forecast(Model95.A , n.ahead = 8)
plot(predict95.1)
```




```
rmse(d95.test$Sale, predict95.1$mean)
```

```
## [1] 84502.82
```

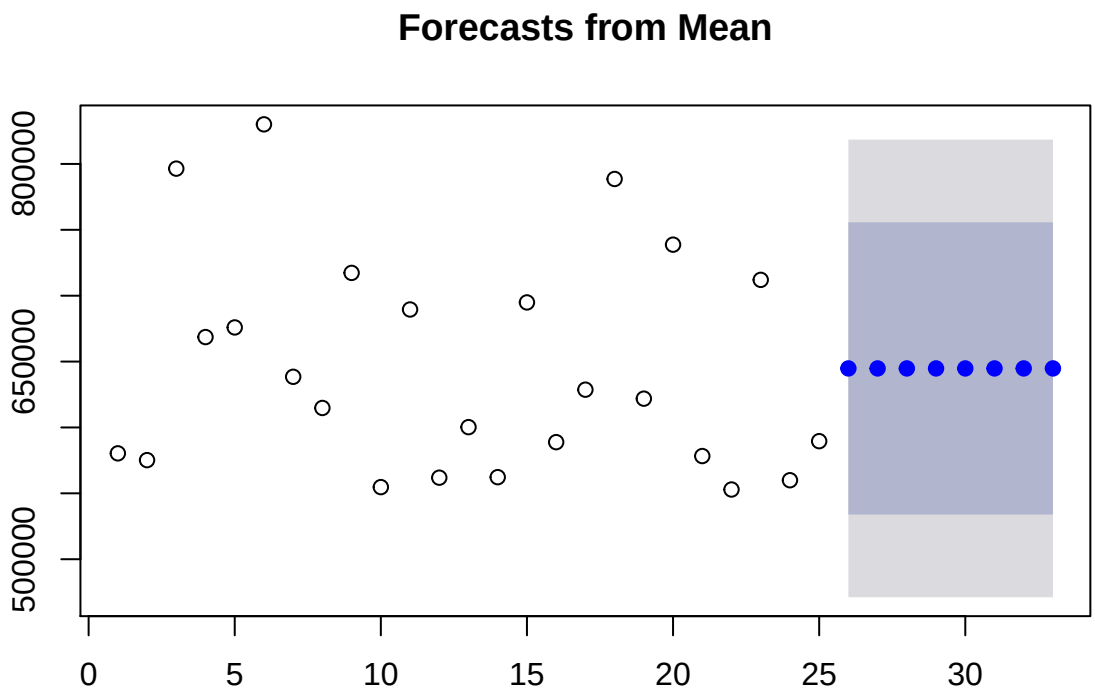
Average Method

```
predict95.3 <- meanf(d95.train$Sale , h =8)

rmse(predict95.3$mean , d95.test$Sale)
```

```
## [1] 87929.1
```

```
plot(predict95.3)
```

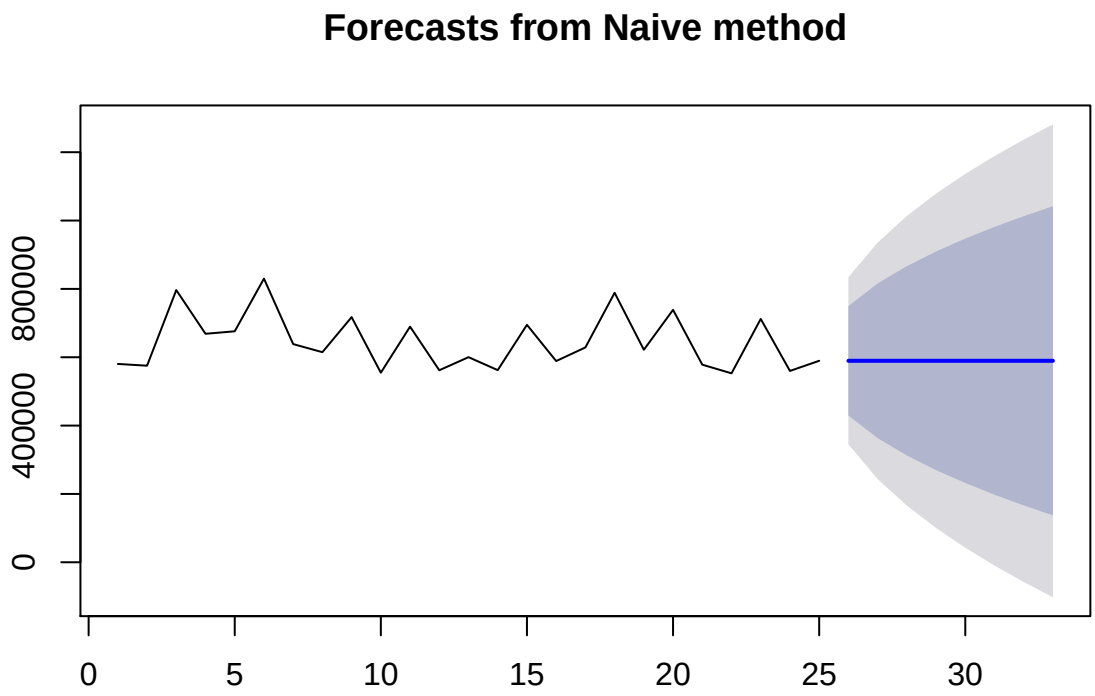


Naive Method

```
predict95.4 <- naive(d95.train$Sale , h =8)
rmse(predict95.4$mean , d95.test$Sale)
```

```
## [1] 119826
```

```
plot(predict95.4)
```

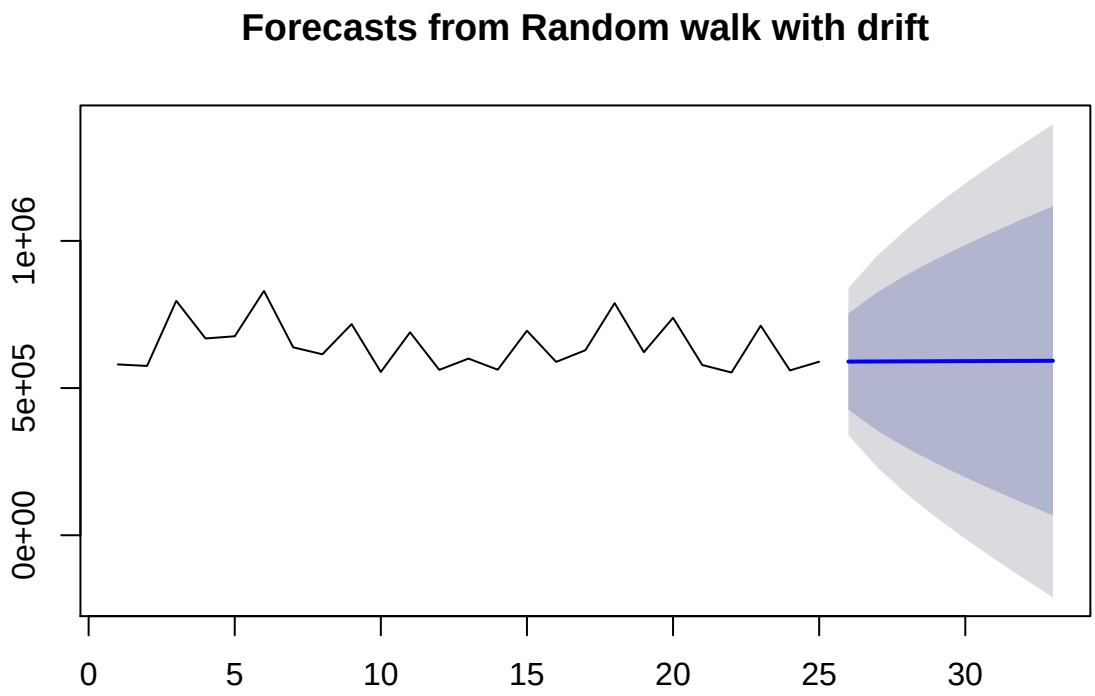


Drift Method

```
predict95.5 <- rwf(d95.train$Sale , h = 8 , drift = TRUE)
rmse(predict95.5$mean , d95.test$Sale)
```

```
## [1] 118685.1
```

```
plot(predict95.5)
```



Department/90

for dept 90

```
D90 <- read_excel("C:/Users/Charan/Desktop/Fall 17/Data Mining/Project/D 90.xlsx")
```

```
D90$Month <- as.Date(D90$Month)
```

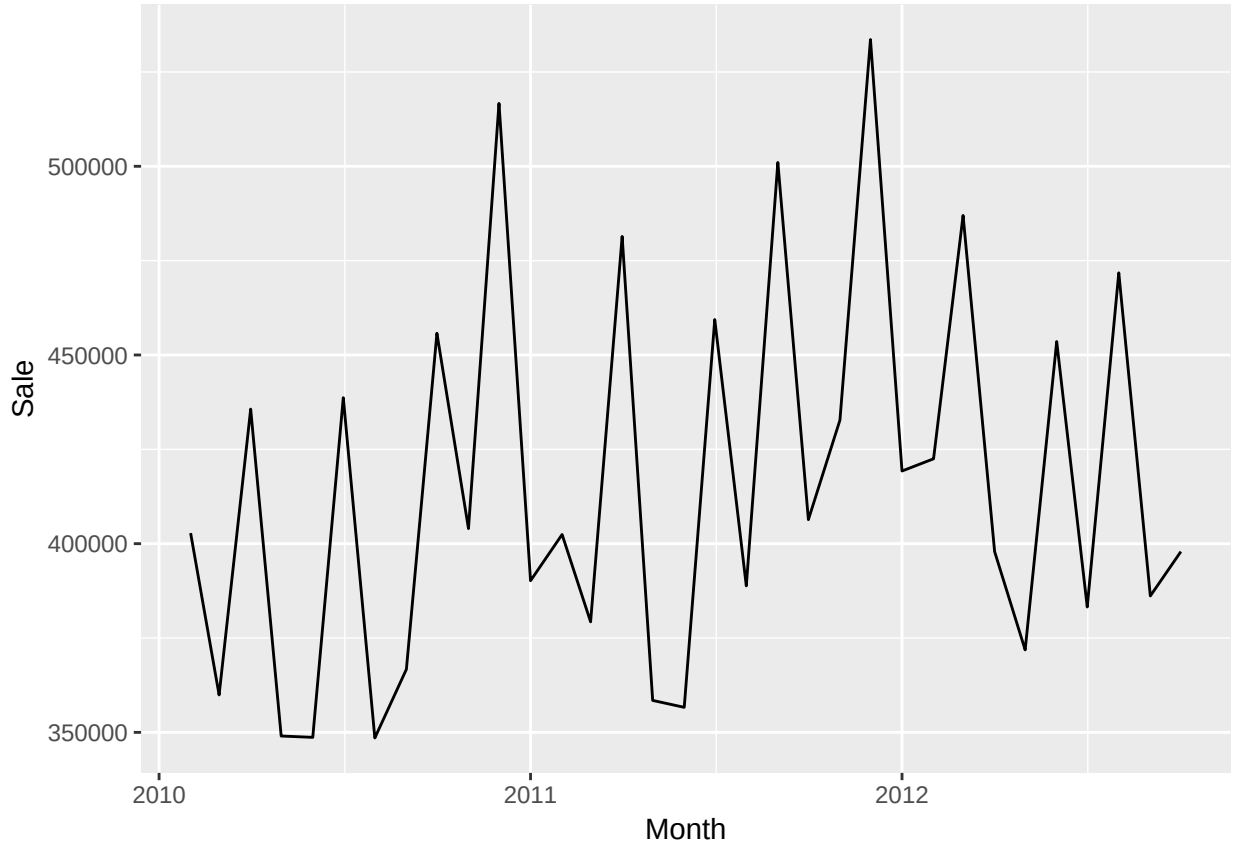
```
d90 <- D90[,-3]
```

```
covariate90 <- D90[1:25,3]
```

```
d90.train <- d90[1:25,1:2]
d90.test <- d90[26:33,1:2]
```

time series plot

```
library(ggplot2)
ggplot(D90 , aes(Month , Sale)) + geom_line()
```



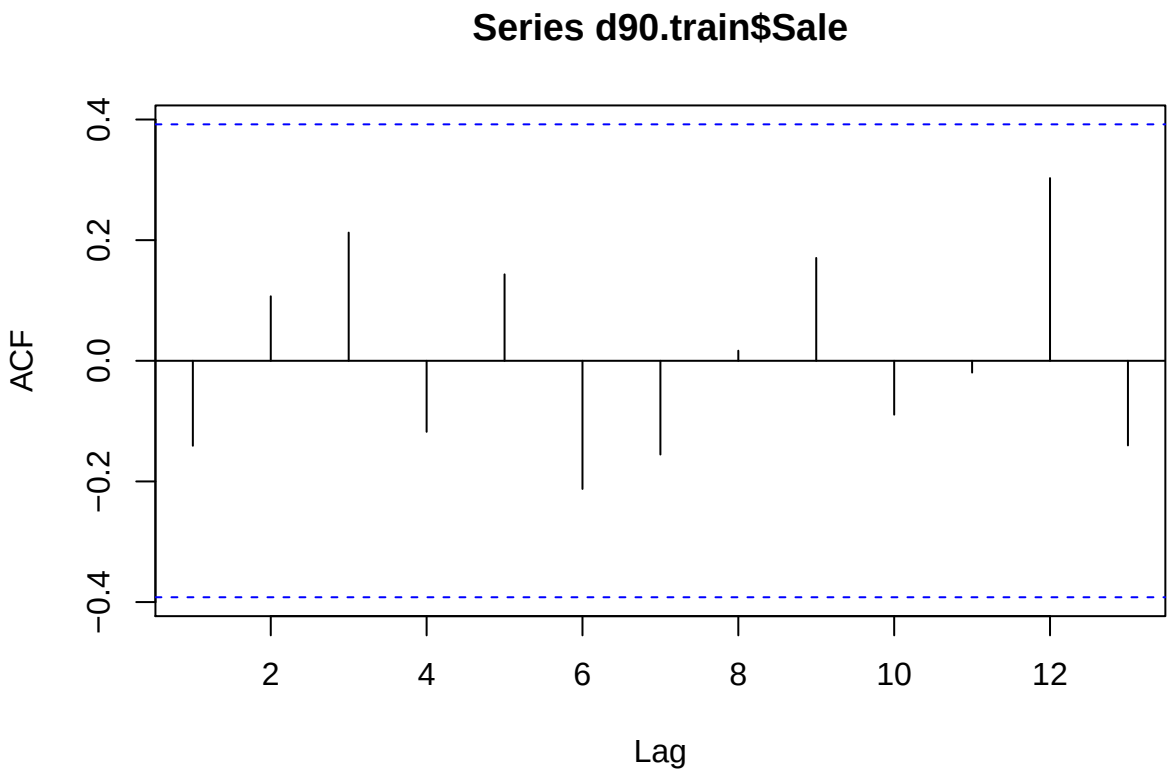
Arima

test for stationary

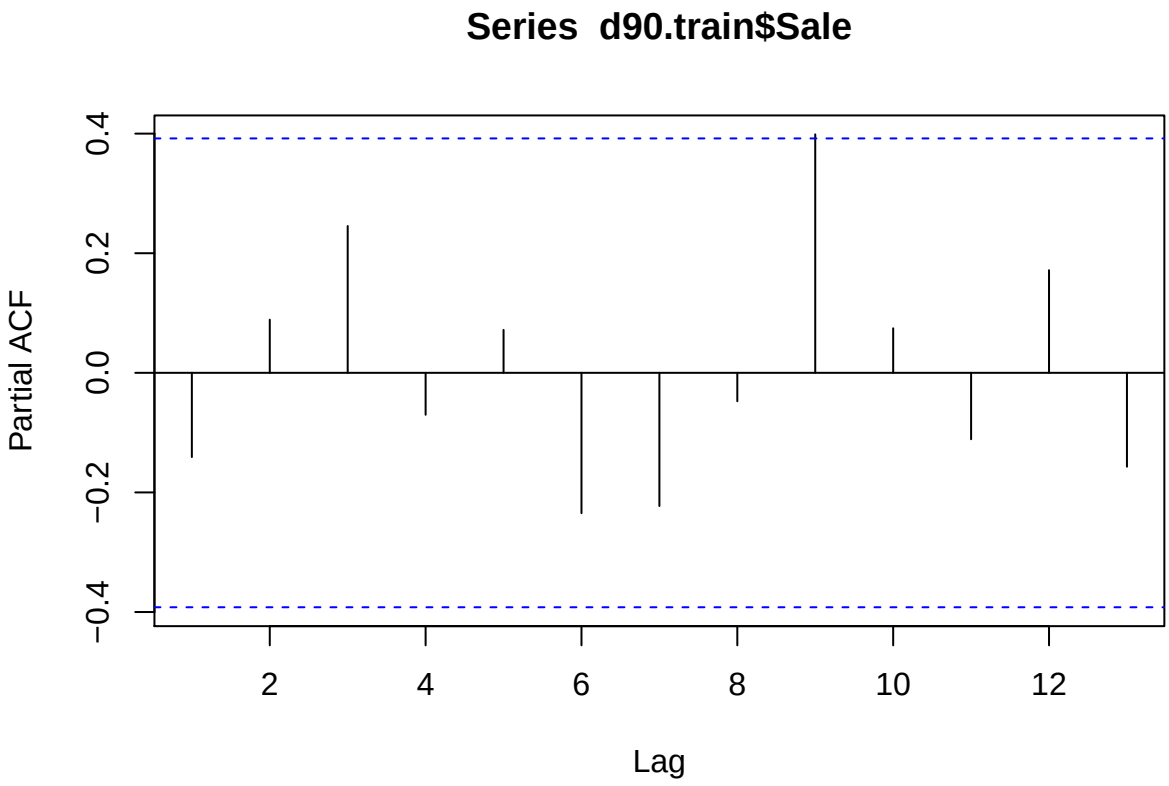
```
adf.test(diff((d90.train$Sale)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff((d90.train$Sale))
## Dickey-Fuller = -3.8819, Lag order = 2, p-value = 0.02987
## alternative hypothesis: stationary
```

```
acf(d90.train$Sale)
```



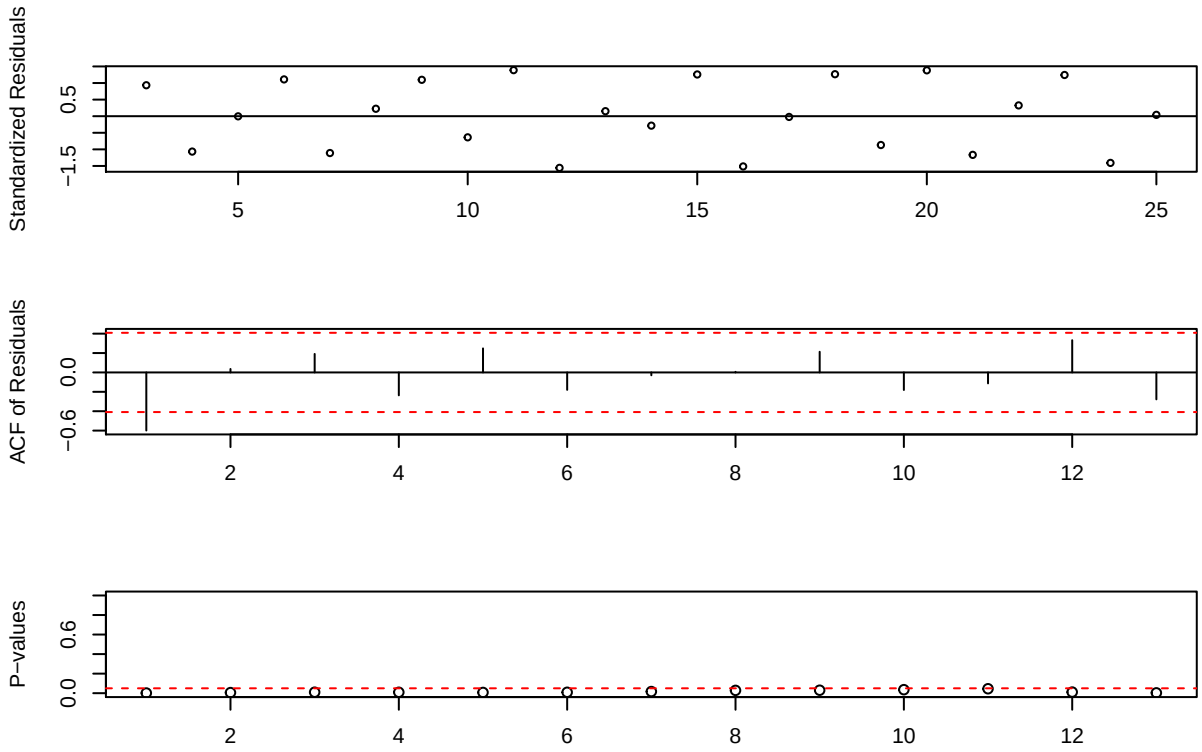

```
pacf(d90.train$Sale)
```



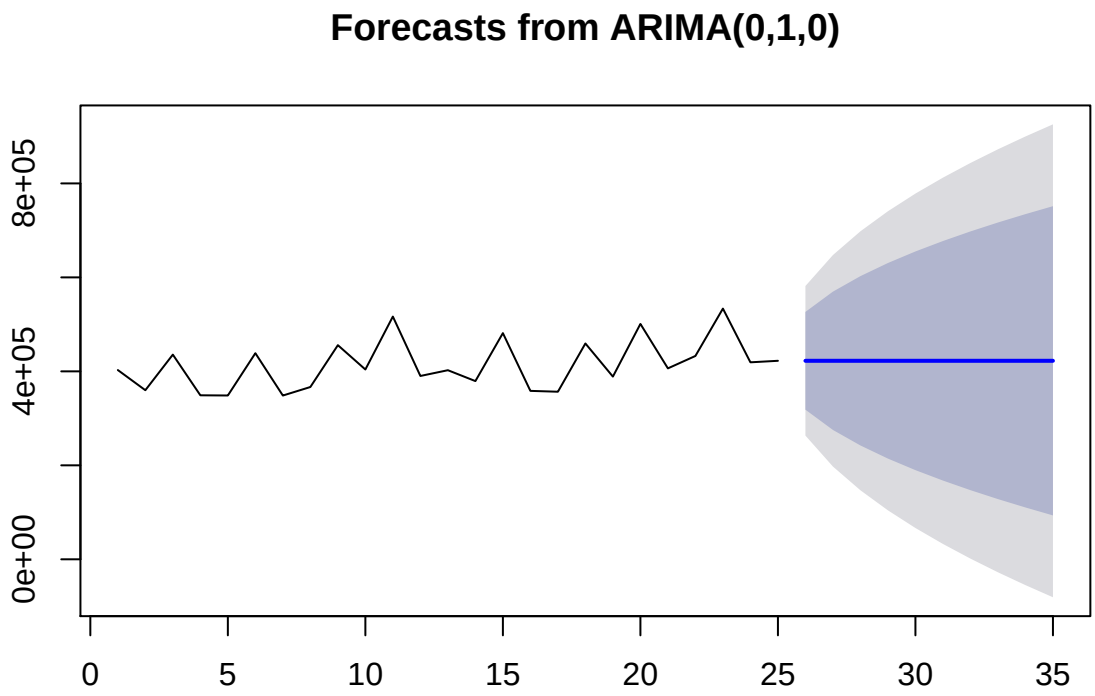

```
Model90.A <- stats::arima(d90.train$Sale , order = c(0,1,0))
Model90.A
```

```
##
## Call:
## stats::arima(x = d90.train$Sale, order = c(0, 1, 0))
##
##
## sigma^2 estimated as 6.592e+09:  log likelihood = -305.36,  aic = 612.73
```

```
tsdiag(Model90.A)
```




```
predict90.1 <- forecast(Model90.A , n.ahead = 8)
plot(predict90.1)
```




```
rmse(d90.test$Sale, predict90.1$mean)
```

```
## [1] 43531.8
```

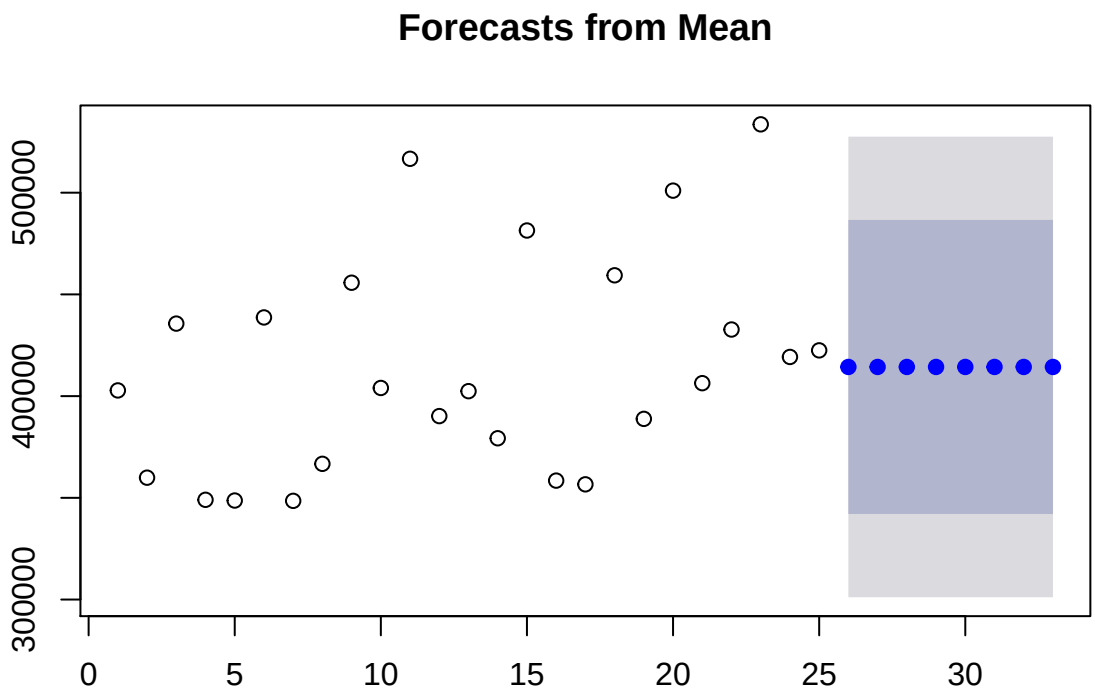
Average Method

```
predict90.3 <- meanf(d90.train$Sale , h =8)

rmse(predict90.3$mean , d90.test$Sale)
```

```
## [1] 42160.08
```

```
plot(predict90.3)
```

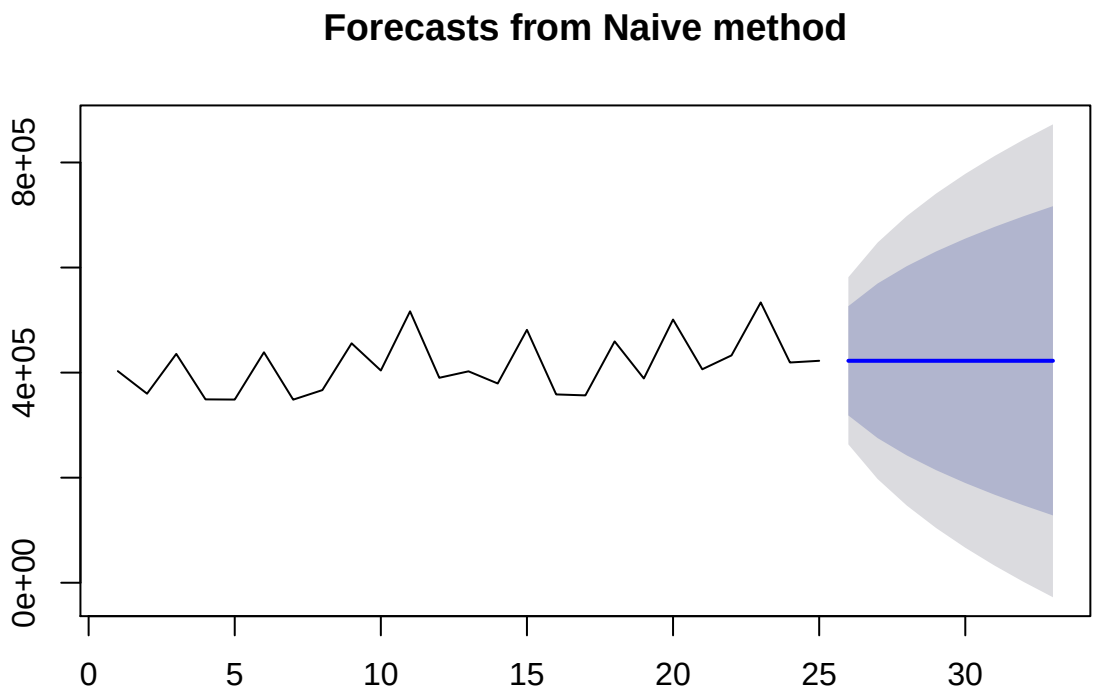


Naive Method

```
predict90.4 <- naive(d90.train$Sale , h =8)
rmse(predict90.4$mean , d90.test$Sale)
```

```
## [1] 42112.21
```

```
plot(predict90.4)
```

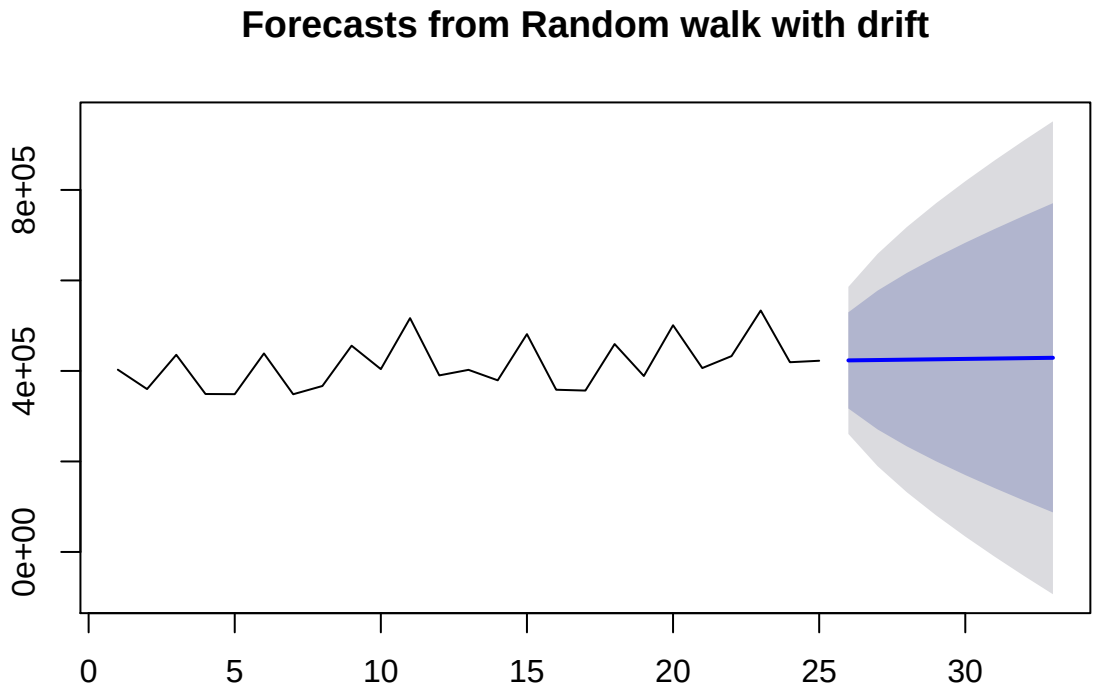


Drift Method

```
predict90.5 <- rwf(d90.train$Sale , h = 8 , drift = TRUE)
rmse(predict90.5$mean , d90.test$Sale)
```

```
## [1] 43189.28
```

```
plot(predict90.5)
```


Forecasts from Random walk with drift
8e+05
4e+05
0e+00
0
5
10
15
20
25
30

Based on the above results, we have chosen different forecasting methods for different departments based on the RMSE error.

Department 94 - Arima

Department 92 - naive

Department 38 - Drift

Department 95 - Arima

Department 90 - naive

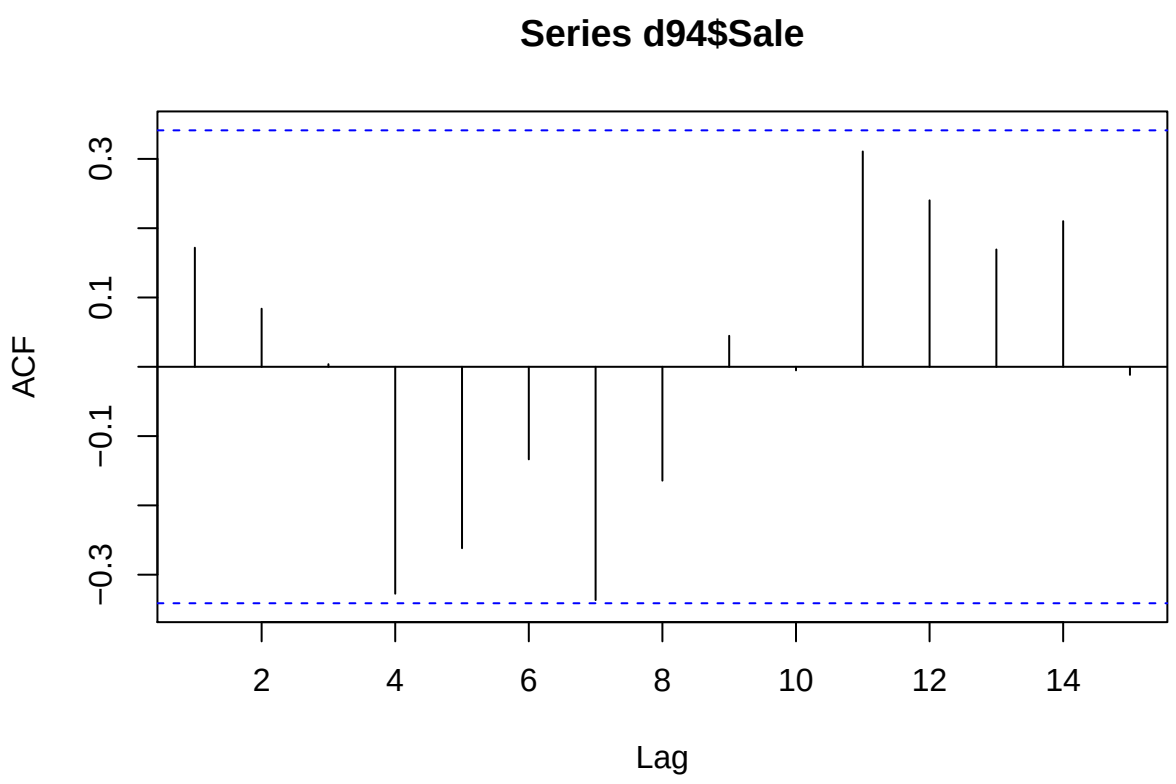
lets forecast future values of 4 months for the above departments.

Department 94

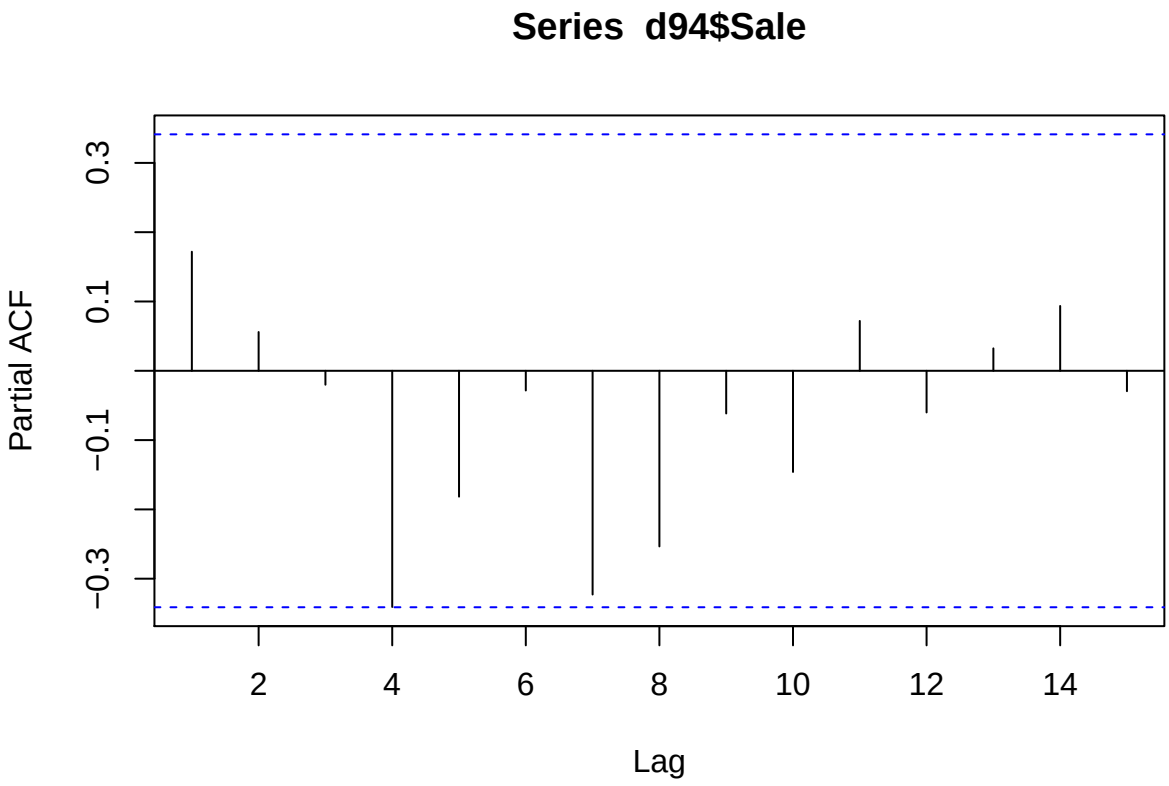
```
adf.test(diff(diff(d94$Sale)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(d94$Sale))
## Dickey-Fuller = -3.6017, Lag order = 3, p-value = 0.04841
## alternative hypothesis: stationary
```

```
acf(d94$Sale)
```



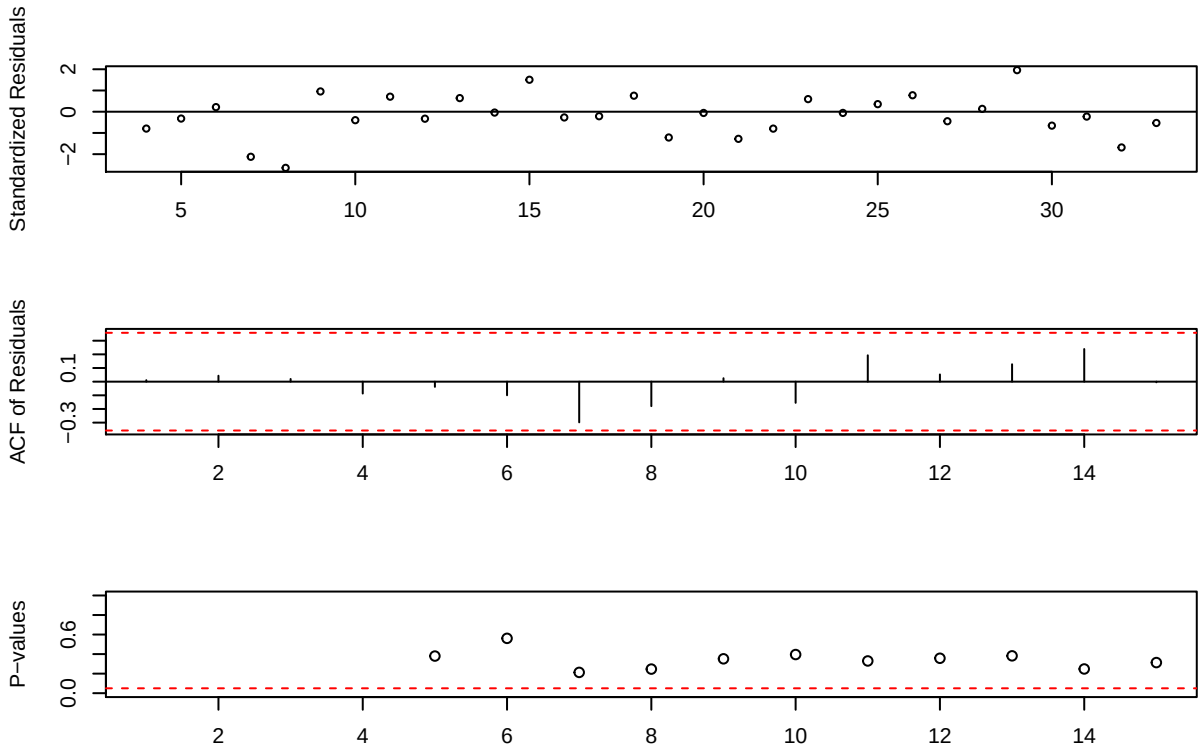

```
pacf(d94$Sale)
```




```
Model1 <- stats::arima(d94$Sale , order = c(2,2,1))
Model1
```

```
##
## Call:
## stats::arima(x = d94$Sale, order = c(2, 2, 1))
##
## Coefficients:
##           ar1      ar2      ma1
##       -0.5246  -0.1969  -1.0000
## s.e.   0.1772   0.1866   0.1023
##
## sigma^2 estimated as 2.47e+09:  log likelihood = -381.62,  aic = 771.24
```

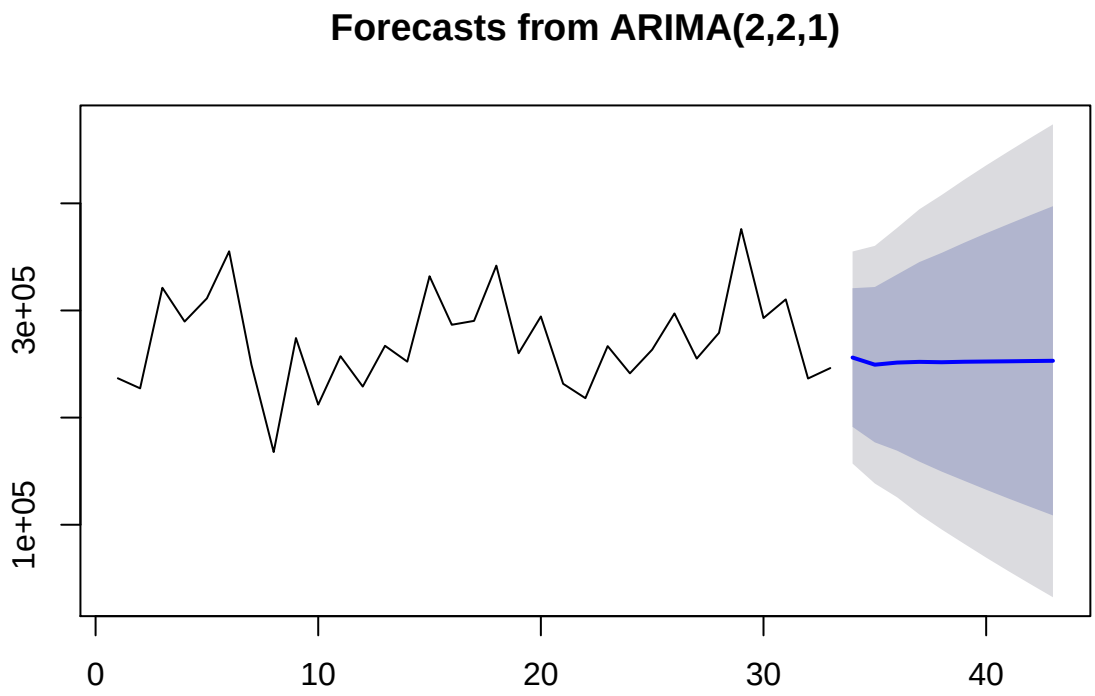
```
tsdiag(Model1)
```




```
predict1 <- forecast(Model1 , n.ahead = 4)
predict1$mean
```

```
## Time Series:
## Start = 34
## End = 43
## Frequency = 1
##  [1] 256072.6 249396.9 251359.6 252040.0 251692.1 252136.2 252367.3
##  [8] 252554.2 252806.2 253032.7
```

```
plot(predict1)
```

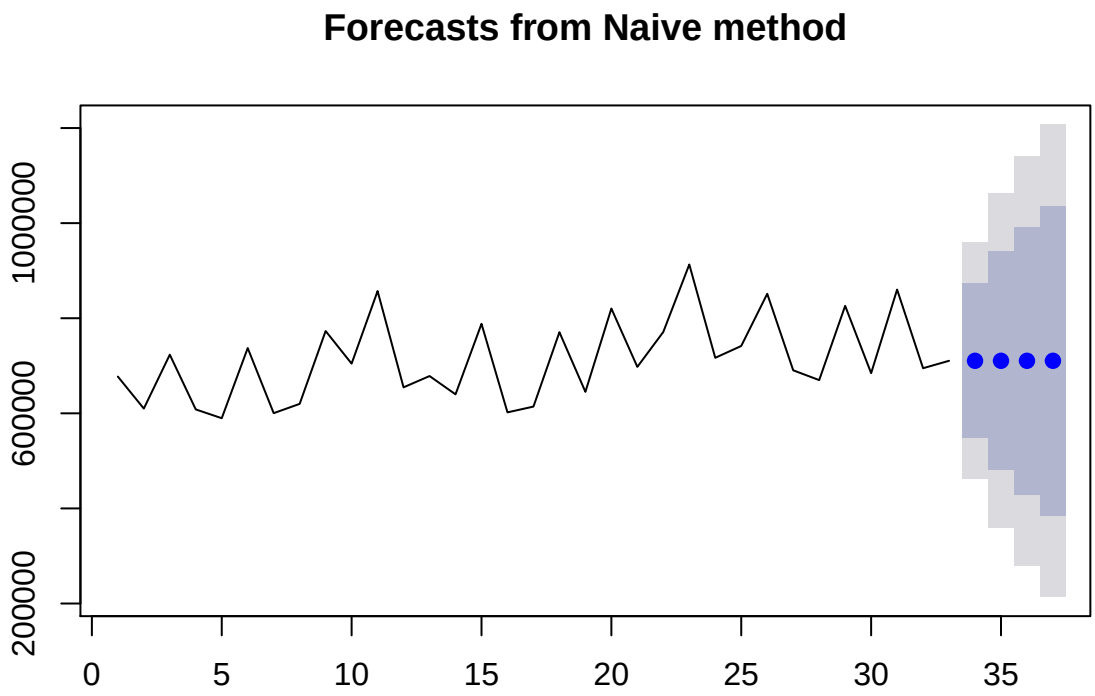


Department 92

```
predict2 <- naive(d92$Sale , h = 4)
predict2
```

```
##    Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
## 34       710487.3 547902.4  873072.3 461835.1  959139.6
## 35       710487.3 480557.5  940417.2 358840.0 1062134.7
## 36       710487.3 428882.0  992092.7 279809.1 1141165.6
## 37       710487.3 385317.5 1035657.2 213182.9 1207791.8
```

```
plot(predict2)
```

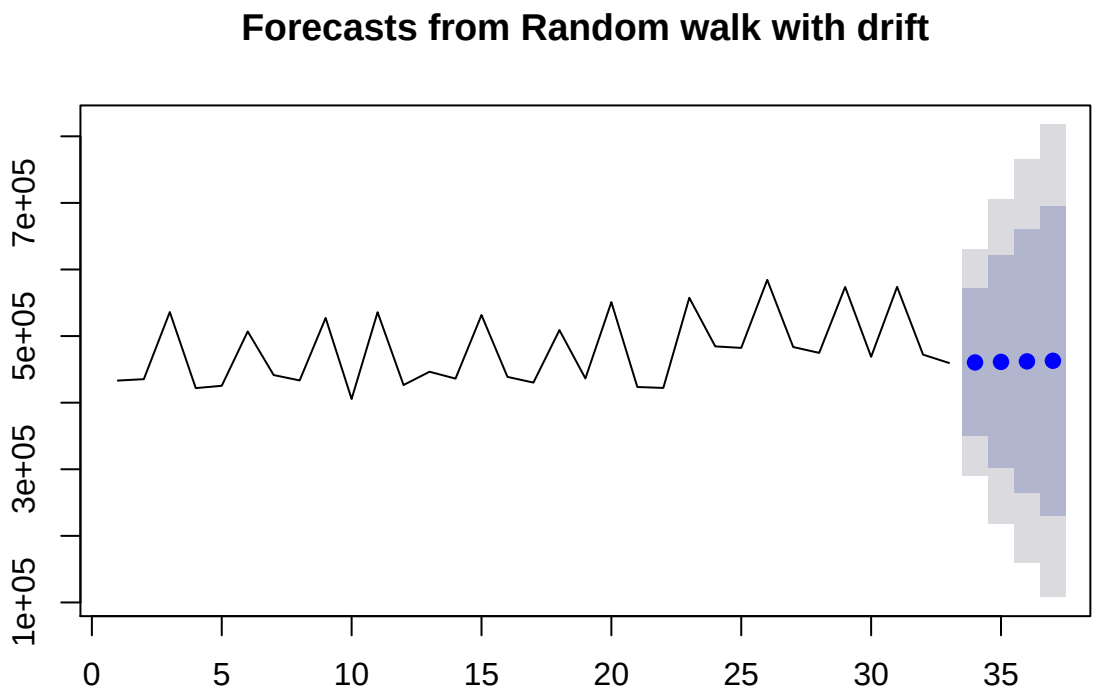


Department 38

```
predict3 <- rwf(d38$Sale , h = 4 , drift = TRUE)
predict3$mean
```

```
## Time Series:
## Start = 34
## End = 37
## Frequency = 1
## [1] 460421.1 461245.9 462070.7 462895.4
```

```
plot(predict3)
```

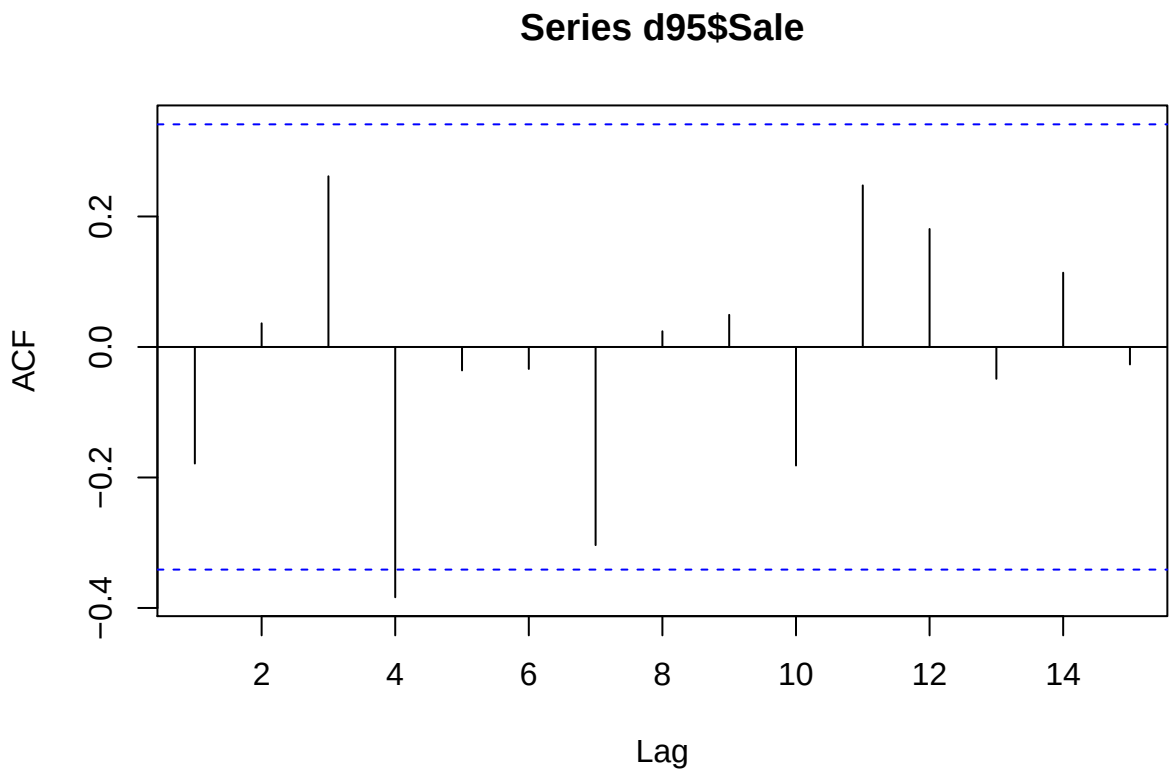


Department 95

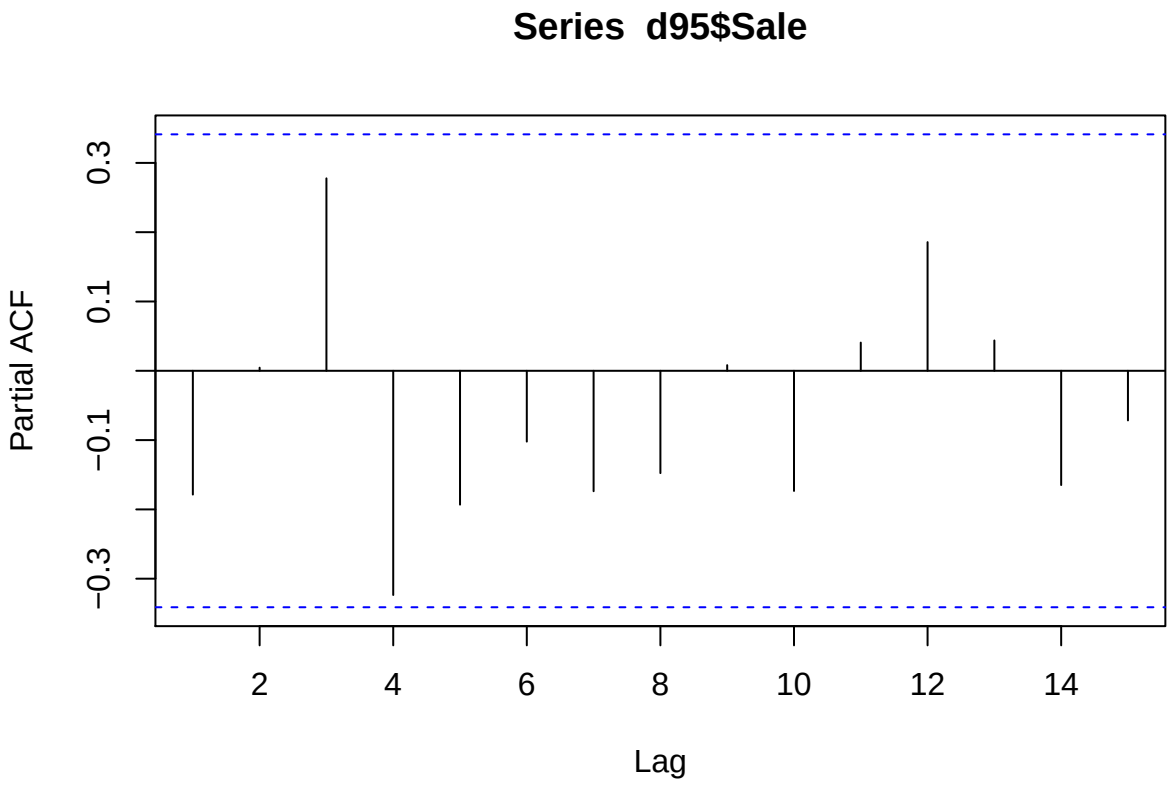
```
adf.test(diff(diff(d95$Sale)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(d95$Sale))
## Dickey-Fuller = -3.6596, Lag order = 3, p-value = 0.04415
## alternative hypothesis: stationary
```

```
acf(d95$Sale)
```



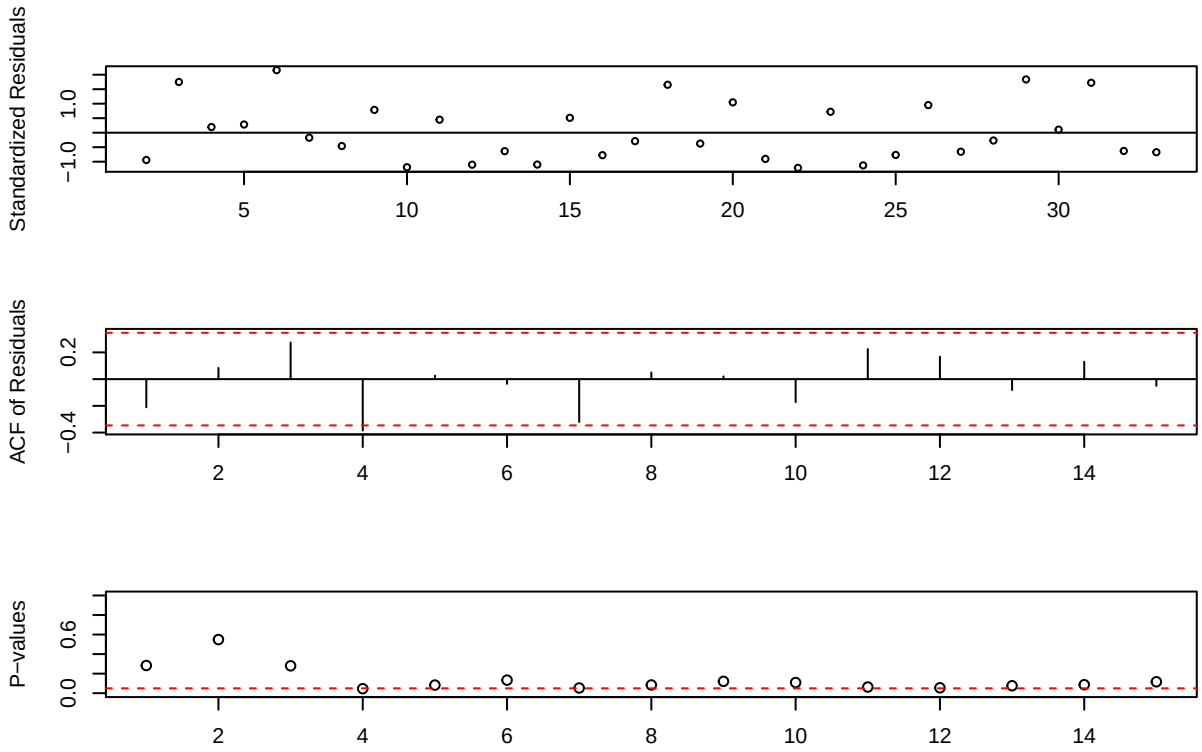

```
pacf(d95$Sale)
```




```
Model4 <- stats::arima(d95$Sale , order = c(0,0,0))
Model4
```

```
##
## Call:
## stats::arima(x = d95$Sale, order = c(0, 0, 0))
##
## Coefficients:
##       intercept
##       652660.32
## s.e.   14316.36
##
## sigma^2 estimated as 6.764e+09:  log likelihood = -420.3,  aic = 844.6
```

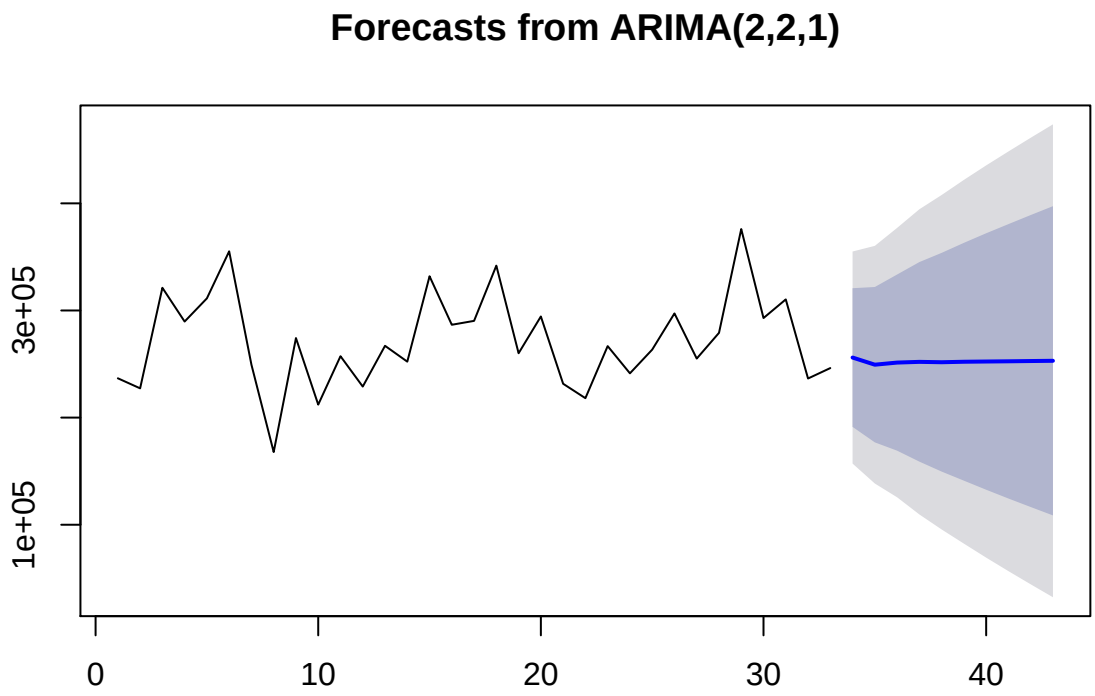
```
tsdiag(Model4)
```




```
predict4 <- forecast(Model4 , n.ahead = 4)
predict4$mean
```

```
## Time Series:
## Start = 34
## End = 43
## Frequency = 1
##  [1] 652660.3 652660.3 652660.3 652660.3 652660.3 652660.3 652660.3
##  [8] 652660.3 652660.3 652660.3
```

```
plot(predict1)
```

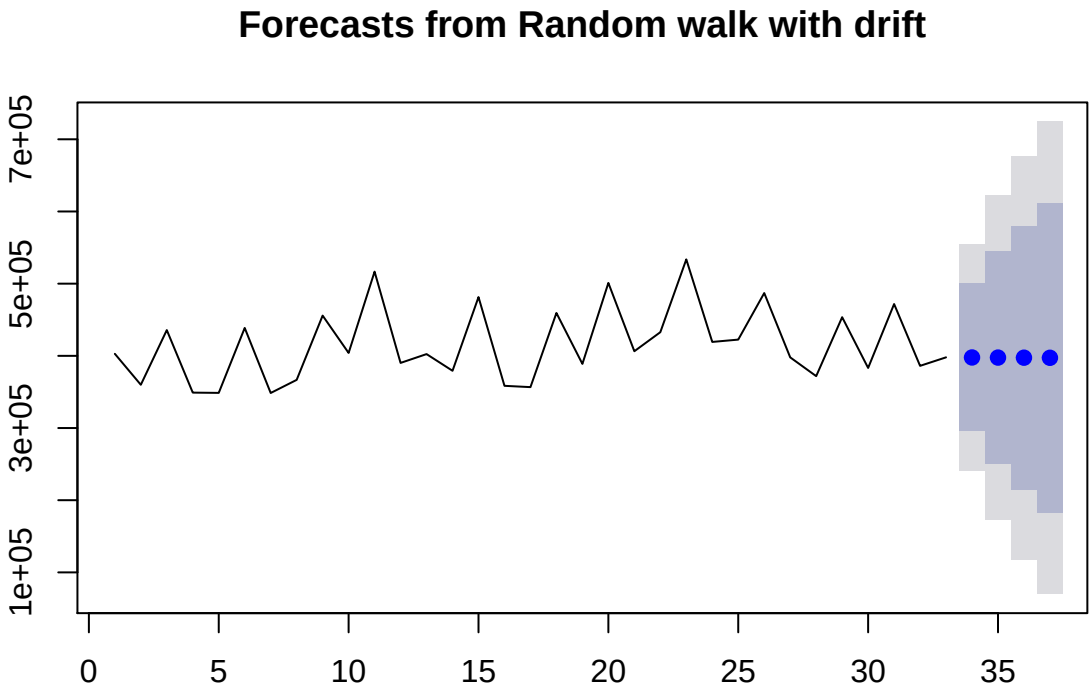


Department 90

```
predict5 <- rwf(d90$Sale , h = 4 , drift = TRUE)
predict5$mean
```

```
## Time Series:
## Start = 34
## End = 37
## Frequency = 1
## [1] 397748.5 397595.3 397442.1 397288.9
```

```
plot(predict5)
```


Forecasts from Random walk with drift
1e+05
3e+05
5e+05
7e+05
0
5
10
15
20
25
30
35