**Australian Monthly Gas Production**

**Objective:** The data in hand is from a package “forecast” which is in a time series format, data shows the gas production of Australia from year 1956/01-1995/08. The objective is to forecast the gas production for the next 12 months.

1. **Reading the data and visualizing the time series.**

***Code:***

library(ggplot2)

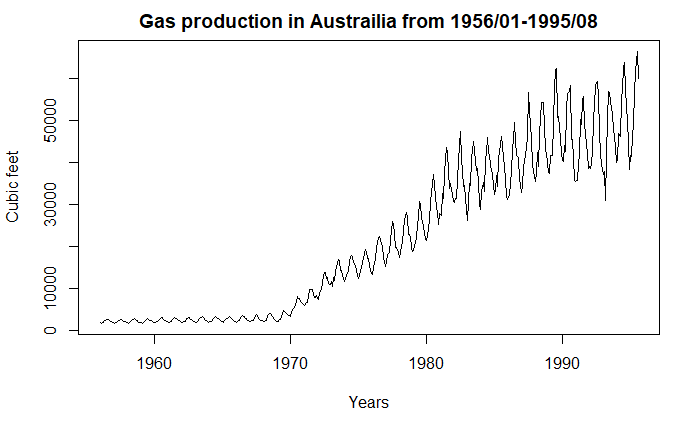
library(forecast)

library(tseries)

library(stats)

?gas

plot(gas,xlab="Years",ylab="Cubic feet",main="Gas production in Austrailia from 1956/01-1995/08")



**Observation:** The series shows trend and seasonality, the initial data from 1956 to 1970 seems to have no trend but only slight amount of seasonality as the peaks and drops appear in a uniform manner.

1. **Descriptive analysis.**

***Code:***

class(gas)

start(gas)

end(gas)

frequency(gas)

summary(gas)

cycle(gas)

**##Aggregate gas production by quarter**

gas.qr=aggregate(gas,nfrequency=4)

plot(gas.qr,main="Gas Production by quarter",xlab="quarter",ylab="Cubic feet")

**##Aggregate gas production by year**

gas.yr=aggregate(gas,nfrequency = 1)

plot(gas.yr,main="Gas Production by year",xlab="year",ylab="Cubic feet")

**###Boxplot Monthly**

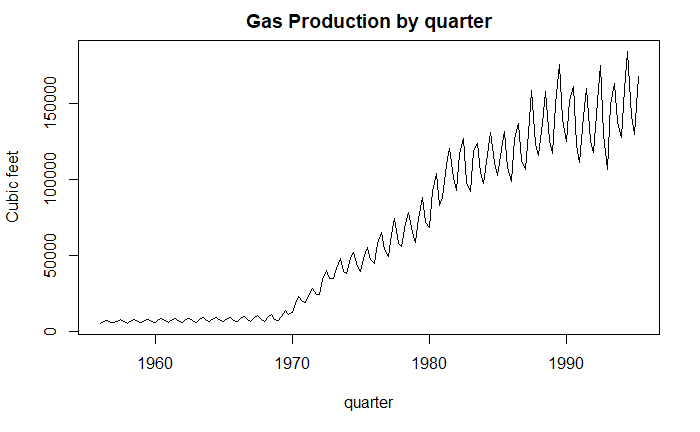
boxplot(gas~cycle(gas),ylab="Cubic feet",xlab="Month",col=1:40,pch=19,main="Monthly gas production in Australia")

**###seasonplot**

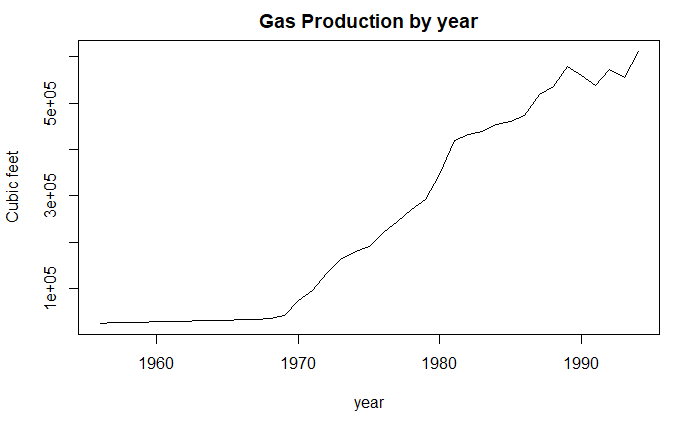
seasonplot(gas, year.labels = TRUE, year.labels.left=TRUE, col=1:40,pch=19, main = "Monthly gas Production in Australia - seasonplot", xlab = "Month", ylab = "Cubic feet")

**####Monthly plot**

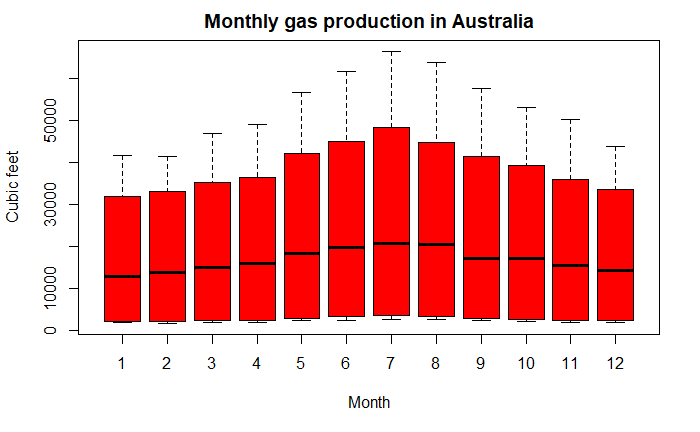
monthplot(gas, col=1:40,pch=19, main = "Monthly gas Production in Australia - seasonplot",xlab = "Month", ylab = "Cubic feet")



**Observation:** The quarterly plot aggregated over the years shows trend and seasonality with seasonality probably being multiplicative in nature.



**Observation:** The yearly plot aggregated over the years shows clear trend.



**Observation:** The Periodicity of the data is monthly and the gas production gradually increases up to July and falls gradually towards the end of the year, the rise and fall happens to be symmetric in nature and July being the highest gas production month across all years.

1. **Decomposing the data to check the individual components.**

***Code:***

decomp\_gas=decompose(gas,type="additive")

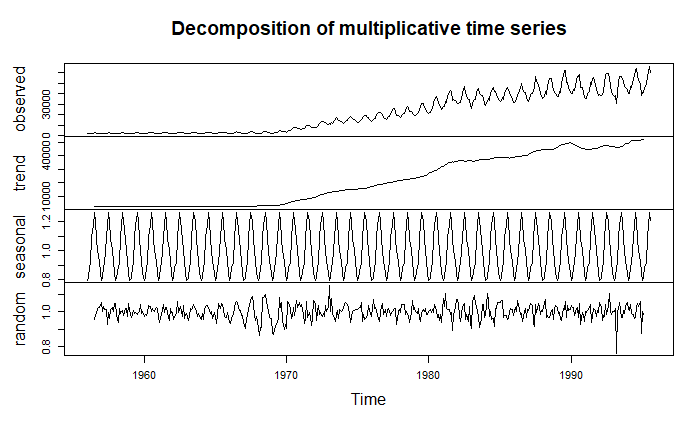
plot(decomp\_gas)

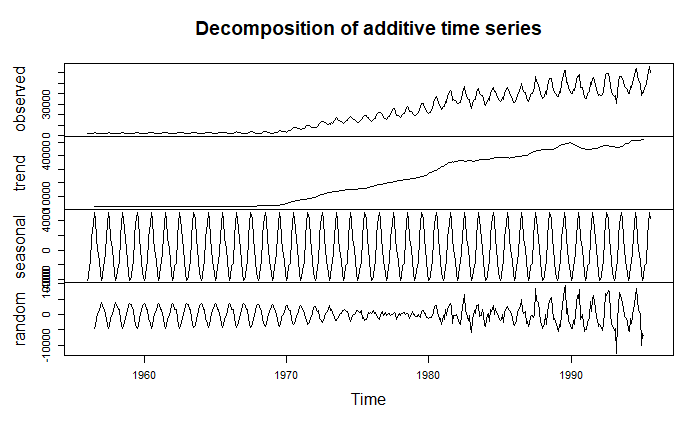
decomp\_gas=decompose(gas,type="multiplicative")

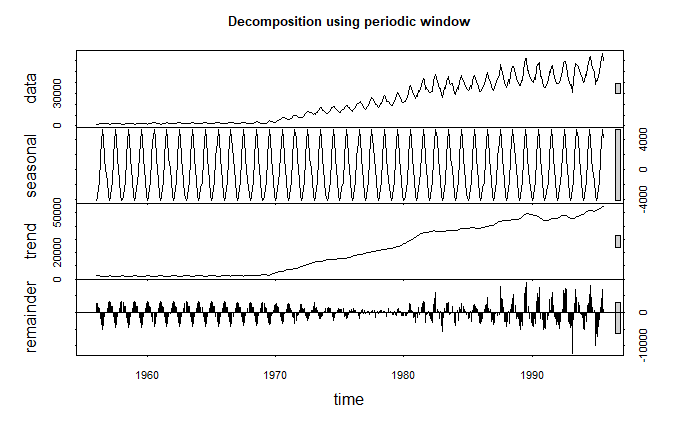
plot(decomp\_gas)

dd=stl(gas,s.window="p")

plot(dd)







**Observation:** Plots clearly show multiplicative seasonality as the random component shows a pattern and trend has the most prominent impact on the time series.

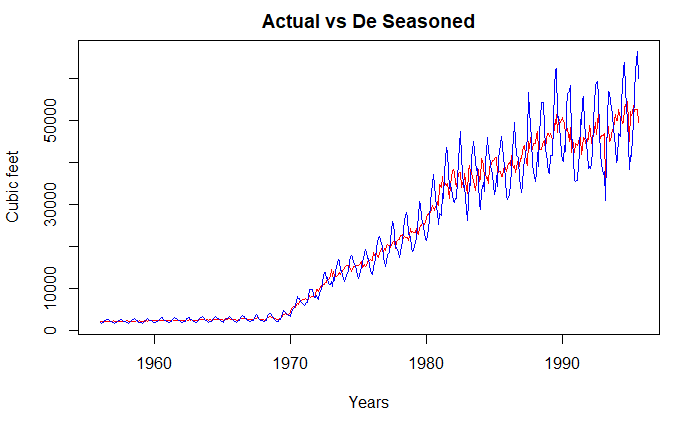
1. **Adjusting Seasonality component and plotting it against the actual data.**

***Code:***

deseasonal\_gas=seasadj(decomp\_gas)

plot(deseasonal\_gas)

ts.plot(deseasonal\_gas,gas,col=c("red","blue"),xlab="Years",ylab="Cubic feet",main="Actual vs De Seasoned")



1. **Checking for stationarity by using Augmented Dickey-Fuller**

**H0= Not Stationary**

**HA= Is Stationary**

***Code:***

plot(gas)

adf.test(gas, alternative = "stationary")

Augmented Dickey-Fuller Test

data: gas

Dickey-Fuller = -2.7131, Lag order = 7, p-value = 0.2764

alternative hypothesis: stationary

**Observation:** At a 95% confidence interval we conclude that the series is not stationary.

1. **Building an Arima model manually by checking for Auto-correlation and Partial auto-correlation, leaving out the first 14 years to build the model.**

***Code:***

acf(gas)

acf(gas,lag.max = 50,main="ACF for the original series")

pacf(gas,lag.max = 50,main="PACF for the original series")

**###Differencing to stationarize at lag 1**

count\_d1 = diff(deseasonal\_gas,differences = 1)

plot(count\_d1)

adf.test(count\_d1, alternative = "stationary")

acf(count\_d1,lag=50,main="ACF for the differenced series")

pacf(count\_d1,lag=50,main="ACF for the differenced series")

**##Splitting the de-seasonalized series into train and test , using data from 1970-1990 as train and remaining as test**

gasTStrain = window(deseasonal\_gas, start=c(1970,1), end=c(1990,12),frequency=12)

gasTStest= window(deseasonal\_gas, start=c(1991,1),end=c(1995,8),frequency=12)

plot(gasTStest)

plot(gasTStrain)

plot(gasTStest)

start(gasTStrain)

end(gasTStrain)

start(gasTStest)

end(gasTStest)

**#arima(4,1,1)**

gasARIMA = arima(gasTStrain, order=c(4,1,1))

gasARIMA

tsdisplay(residuals(gasARIMA), lag.max=10, main='Model Residuals')

**###Checking if residuals are independent and normally distributed**

**#H0: Residuals are independent**

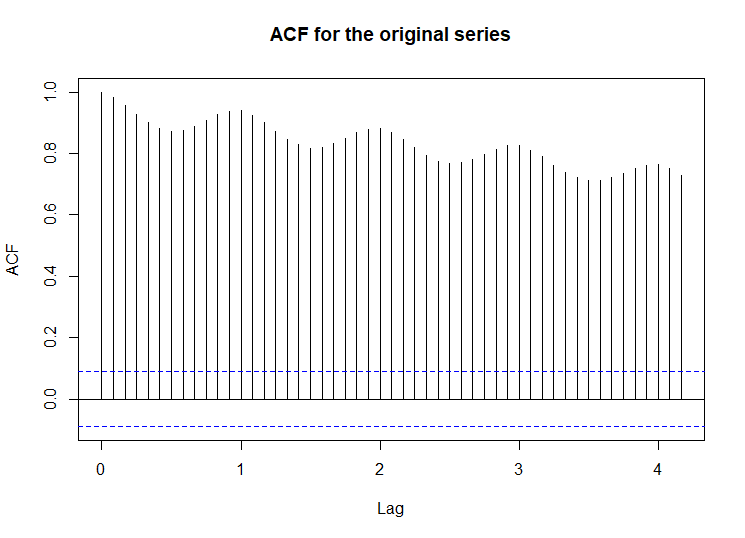
**#Ha: Residuals are not independent**

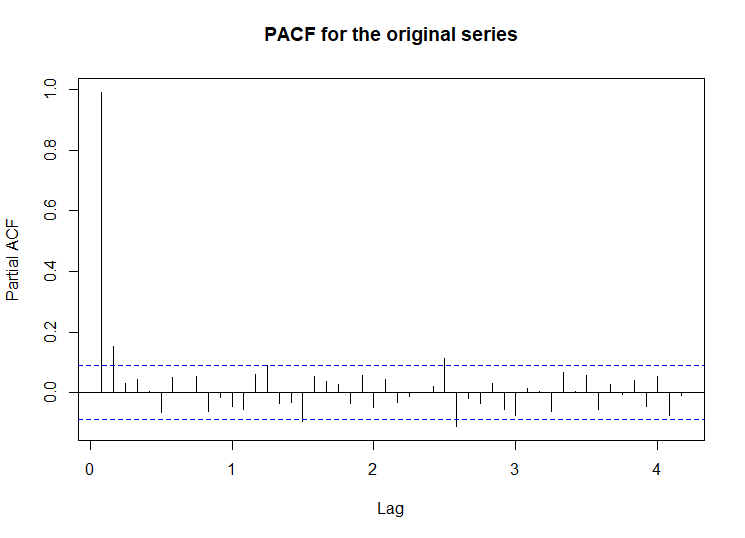
library(stats)

Box.test(gasARIMA$residuals,type="Ljung")

hist(gasARIMA$residuals,col="blue")

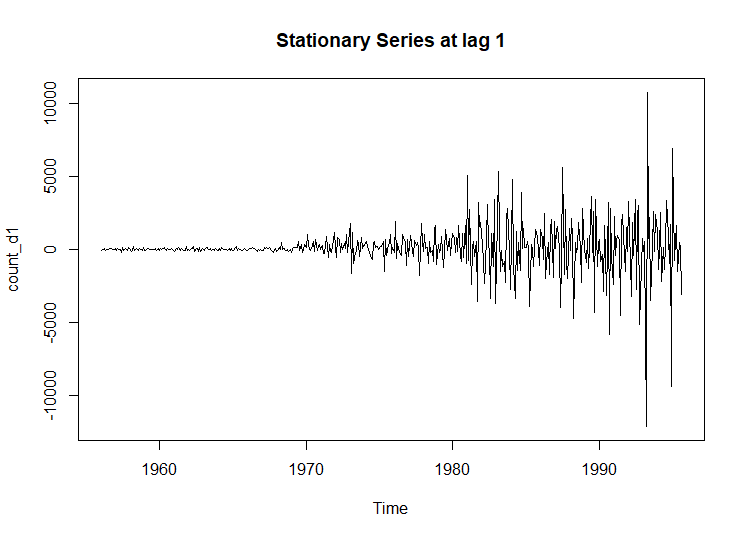
1. **Plotting Acf and Pacf for the original series .**





**Observation:** As shown previously the series is not stationary and the above plot is on the non-stationary series. There is significant amount of auto correlation in the series.

1. **Plotting Acf and Pacf by converting the series to be stationary (using the de-seasonalized) at lag 1 and validating the stationarity by Augmented Dickey-Fuller test.**



Augmented Dickey-Fuller Test

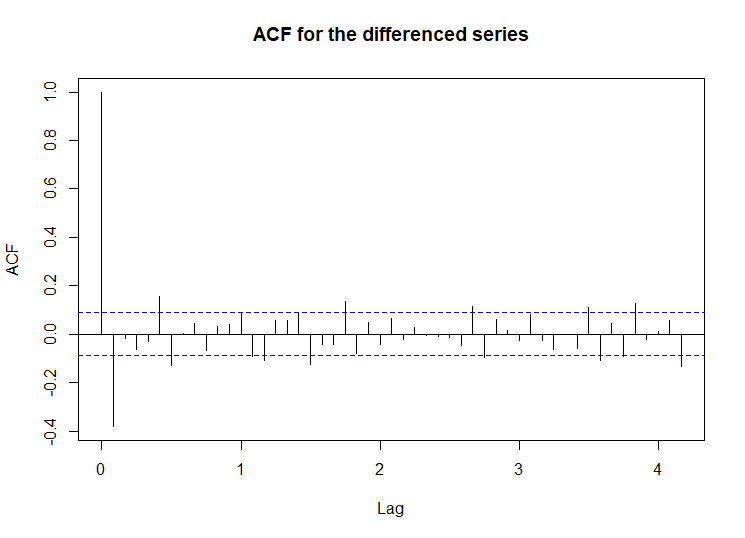
data: count\_d1

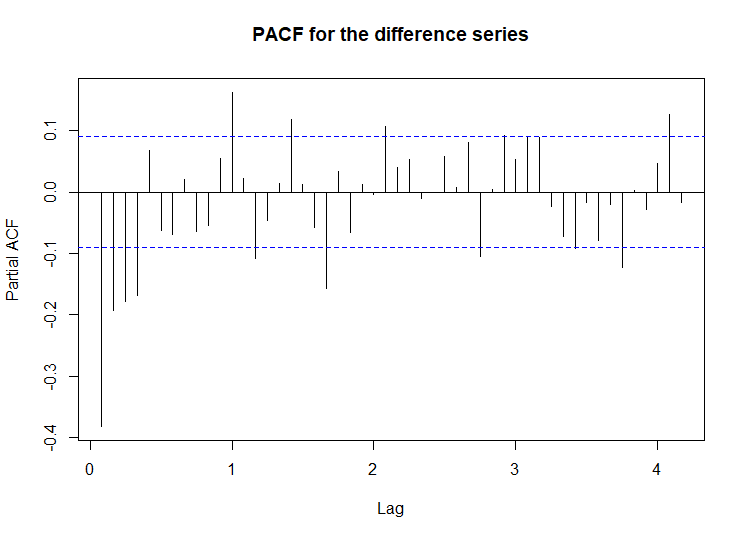
Dickey-Fuller = -9.3588, Lag order = 7, p-value = 0.01

alternative hypothesis: stationary

**Observation:** The above plot shows the stationary series at lag 1 and the Augmented Dickey-Fuller test validates that the series is stationary at lag 1 at a 95% confidence interval.

1. **Plotting Acf and Pacf for the stationary series to come up with the order (p d, q) for the Arima model.**

****

****

**Observation:** The above plots show that the Acf for the stationary series cuts-off after 1 and Pacf cuts-off after 4. Hence, we manually select the order as (4,1,1).

Call:

arima(x = gasTStrain, order = c(4, 1, 1))

Coefficients:

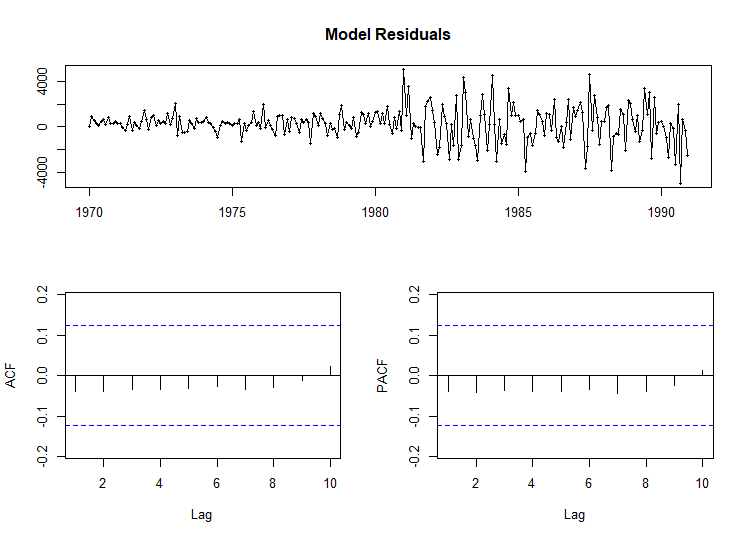
ar1 ar2 ar3 ar4 ma1

-0.9677 -0.4183 -0.2595 -0.1788 0.5960

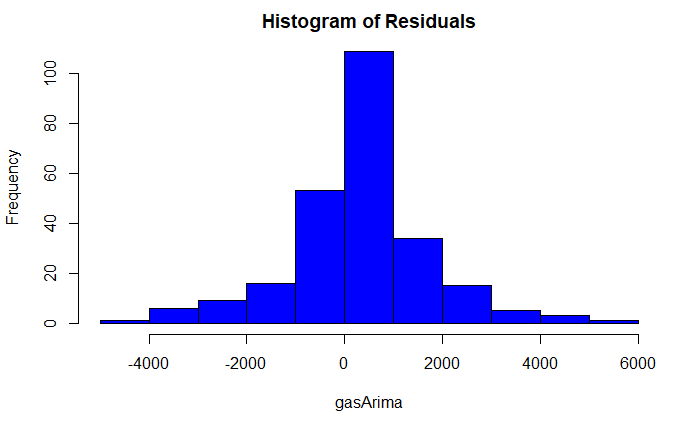
s.e. 0.2614 0.1241 0.0938 0.0641 0.2618

sigma^2 estimated as 2099993: log likelihood = -2183.25, aic = 4378.5

1. **Checking the residuals, Acf and Pacf of built Arima model.**

****

**Observation:** The Residuals of the model seem to be like white noise and no significant auto correlation.

****

Box-Ljung test

data: gasARIMA$residuals

X-squared = 0.39449, df = 1, p-value = 0.5299

**Observation:** The Residuals are normally distributed and are independent.

1. **Validating and checking the accuracy of the manually built Arima model against the hold out test data.**

***Code:***

fcast <- forecast(gasARIMA, h=56)

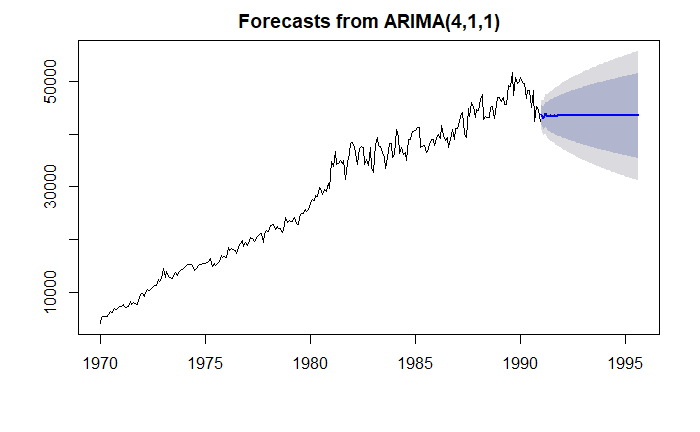
accuracy(fcast, gasTStest)

plot(fcast,gasTStest)

ME RMSE MAE MPE MAPE MASE

Training set 274.8294 1446.257 1042.145 1.439905 3.860483 0.4000523

Test set 4616.8622 5847.003 4983.657 9.060288 10.084658 1.9130960

****

**Observation:** The MAPE value on the testis 10.08 and as seen earlier the major contributor of the series is trend and the auto Arima model doesn’t seem to capture that, it gives a fairly straight line for the hold out data set.

1. **Building an Auto-Arima Model by using the same train and test data.**

***Code:***

fit<-auto.arima(gasTStrain, seasonal=F)

**###Checking if residuals are independent and normally distributed**

**#H0: Residuals are independent**

**#Ha: Residuals are not independent**

tsdisplay(residuals(fit), lag.max=10, main='Auto ARIMA Model Residuals')

Box.test(fit$residuals,type="Ljung")

hist(fit$residuals,col="blue")

Series: gasTStrain

ARIMA(1,1,1) with drift

Coefficients:

ar1 ma1 drift

0.2708 -0.6947 158.5120

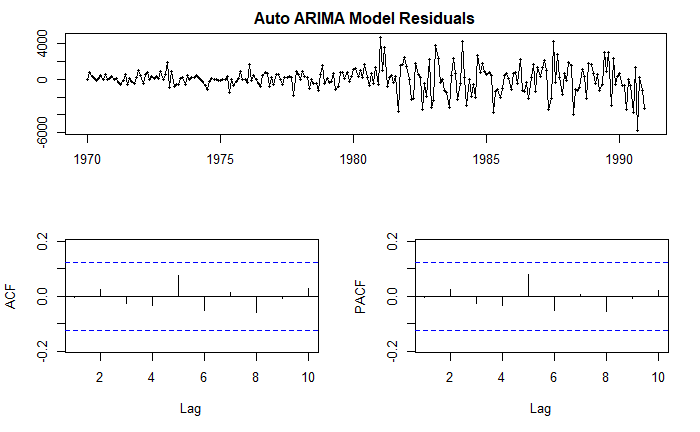
s.e. 0.1196 0.0921 38.0179

sigma^2 estimated as 2056583: log likelihood=-2179.14

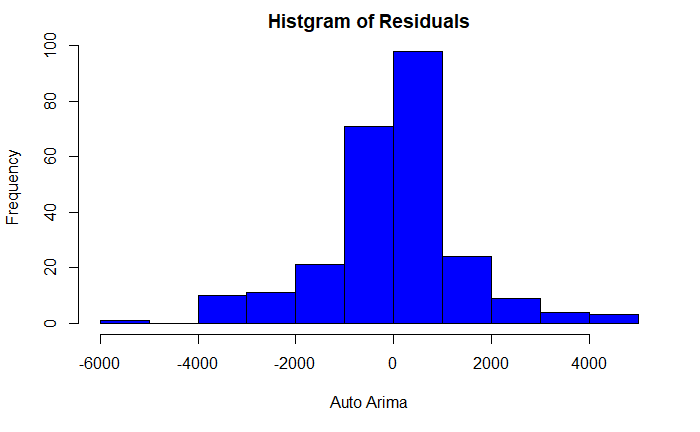
AIC=4366.28 AICc=4366.45 BIC=4380.38

**Observation:** The Auto Arima model has chosen the p, d, q values as 1,1,1 receptively, also has a slightly lower AIC as compared to Manual Arima.

1. **Checking the residuals, Acf and Pacf of built Auto-Arima model.**



**Observation:** The Residuals of the model seem to be like white noise and no significant auto correlation.



Box-Ljung test

data: fit$residuals

X-squared = 0.0086802, df = 1, p-value = 0.9258

**Observation:** The Residuals are normally distributed and are independent.

1. **Validating and checking the accuracy of the built Auto-Arima model against the hold out test data.**

***Code:***

fcast1 <- forecast(fit, h=56)

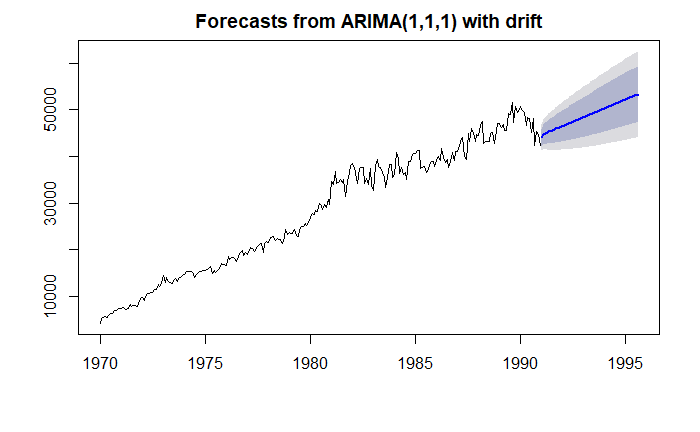
accuracy(fcast1,gasTStest)

plot(fcast1,gasTStest)

ME RMSE MAE MPE MAPE MASE

Training set 2.941562 1422.652 979.7707 0.1171342 3.457422 0.3761084

Test set -908.041570 2773.793 1805.7770 -2.2275847 3.988248 0.6931908



**Observation:** The MAPE value on the testis 3.98 which is much better as compared to Manually built Arima model. Auto-Arima captures trend which is major contributor of the series. Hence, Auto- Arima will be used to forecast the gas production for the next 12 periods.

1. **Forecasting into the next 12 periods on the complete data set by using auto Arima.**
2. **Forecasting on the de-seasonalized data considering no seasonality.**

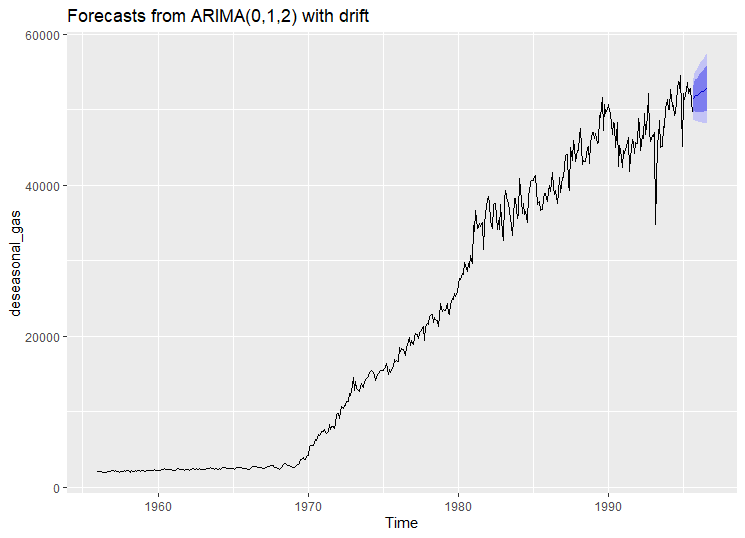
***Code:***

fit1<-auto.arima(deseasonal\_gas,seasonal=F)

fcast2=forecast(fit1, h=12)

fcast2

autoplot(fcast2)



**Observation:** The forecast done on the complete dataset successfully captures trend and the p, d, q values at taken as (0, 1, 2) receptively, the forecasted values are shown below.

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

Sep 1995 58640.96 55290.53 61991.39 53516.92 63764.99

Oct 1995 56168.61 50785.28 61551.94 47935.52 64401.70

Nov 1995 56168.61 48650.49 63686.73 44670.64 67666.58

Dec 1995 56168.61 46999.97 65337.25 42146.38 70190.83

Jan 1996 56168.61 45604.24 66732.97 40011.81 72325.41

Feb 1996 56168.61 44372.52 67964.70 38128.05 74209.17

Mar 1996 56168.61 43257.78 69079.44 36423.20 75914.02

Apr 1996 56168.61 42231.92 70105.30 34854.28 77482.94

May 1996 56168.61 41276.55 71060.66 33393.18 78944.04

Jun 1996 56168.61 40378.89 71958.33 32020.32 80316.90

Jul 1996 56168.61 39529.59 72807.63 30721.42 81615.80

Aug 1996 56168.61 38721.58 73615.64 29485.68 82851.54

1. **Forecasting on the original data considering seasonality.**

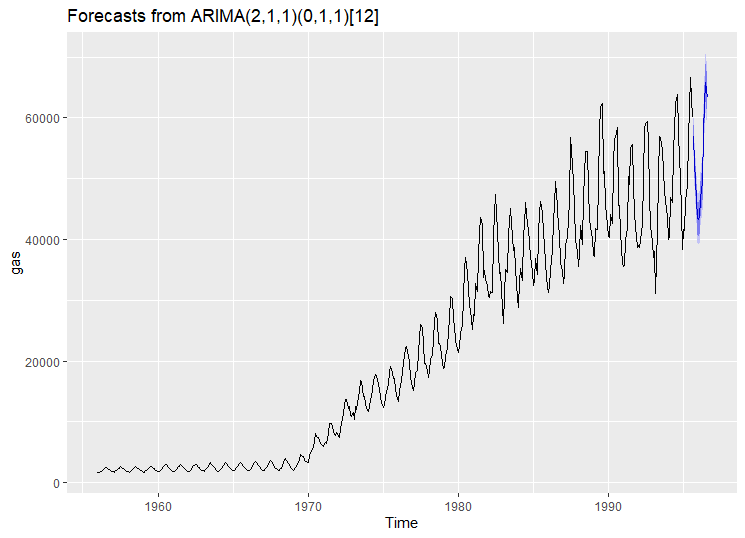
***Code:***

fit1<-auto.arima(gas,seasonal=T)

fcast2=forecast(fit1, h=12)

fcast2

autoplot(fcast2)



**Observation:** Auto-Arima with seasonality captures both the trend and seasonality components with the trend components of Arima as (2,1,1) and the seasonality component as (0,1,1) and the periods as 12, the forecasted values are shown below.

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

Sep 1995 56907.83 54846.53 58969.13 53755.35 60060.32

Oct 1995 52476.28 50158.98 54793.57 48932.28 56020.27

Nov 1995 49719.79 47202.83 52236.76 45870.42 53569.16

Dec 1995 43473.70 40830.29 46117.11 39430.96 47516.45

Jan 1996 43318.41 40575.77 46061.05 39123.91 47512.91

Feb 1996 43601.71 40776.80 46426.62 39281.38 47922.04

Mar 1996 46668.11 43770.45 49565.78 42236.52 51099.71

Apr 1996 49376.36 46411.96 52340.76 44842.70 53910.02

May 1996 57536.25 54509.03 60563.46 52906.53 62165.97

Jun 1996 62184.69 59097.40 65271.98 57463.09 66906.29

Jul 1996 65795.74 62650.40 68941.08 60985.35 70606.13

Aug 1996 63391.54 60189.71 66593.37 58494.76 68288.31