# You will try the different models we ran in class today on your dataset.

Done

Understand and Explain your model output

Attributes(Columns): Year, Month, Sales

Time series data formation

• Data converted to timeseries using ts() function to map sales values with time frame in order to compute different forecasting models. Ranging from (1964-Jan to 1972-Sept)

Syntax

time\_index <- as.Date(paste(df$Year, df$Month, "01", sep = "-"))

ts\_data <- ts(df$Sales, start = c(min(df$Year), min(df$Month)), frequency = 12)

time\_series <- ts(ts\_data, start = c(min(df$Year), min(df$Month)), frequency = 12)

names(time\_series) <- "Sales"

print(time\_series)

A screenshot of a computer

Description automatically generated

Auto-Correlation Function (ACF): displays the correlation coefficients between a time series and its lagged values.

The ACF is a useful tool for understanding the temporal dependencies and patterns in time series data that is underlying structure of the data, and identify potential seasonality.

A screenshot of a computer

Description automatically generated

Forecasting models:

# The Naive forecast/method/approach

Simplest time series forecasting techniques

Idea: future values of a time series will be the same as the most recent observed value.

Assumptions: there are no patterns, trends, or seasonality in the data.

the best estimate for the future is the most recent historical value.

Key Points about the Naive Forecast:

Simplicity

Limited Assumptions

Performance

Limitations: The naive forecast assumes that the future will be exactly like the most recent past.

A screenshot of a computer

Description automatically generated

# RWF(Random Walk with Drift)

# also known as a Random Walk Plus Drift

It is an extension of the simple random walk model and incorporates a linear trend or "drift" component in addition to the random fluctuation that is assumes that each future value in the time series is equal to the previous value plus a random error term.

Random Walk with Drift models can be valuable for forecasting when there is evidence of a long-term trend or systematic change in a time series.

A screenshot of a computer

Description automatically generated

# Moving averages

used to smooth out variations or noise in data, highlight trends, and make underlying patterns more apparent. Moving averages are calculated by taking the average of a set of data points within a sliding or "moving" window.

**Simple Moving Average (SMA):** To calculate an SMA, you sum a fixed number of data points (e.g., the last N data points) and then divide by the number of data points in the window (N).

**Exponential Moving Average (EMA):**

The Exponential Moving Average gives more weight to recent data points and less weight to older data points.

**Weighted Moving Average (WMA):**

The Weighted Moving Average assigns different weights to data points within the moving window.

Weights can be assigned based on various criteria, such as time or significance of data points.

A screenshot of a computer

Description automatically generated

# Plot the time series and different model forecasts in one chart

A screenshot of a computer

Description automatically generated

# Pick an accuracy measure, compare your models, and state the best model based on the accuracy comparison

Accuracy Measures:

● Mean Error (ME)

● Root Mean Squared Error (RMSE)

● Mean Absolute Error (MAE)

● Mean Percentage Error (MPE)

● Mean Absolute Percentage Error (MAPE)

● Mean Absolute Scaled Error (MASE)

● Autocorrelation of errors at lag 1 (ACF1)

**Selected Mean Absolute Scaled Error(MASE)**

It measures the relative accuracy of a forecast by comparing it to the mean absolute error (MAE) of a naive or benchmark model.

The interpretation of MASE is as follows:

If **MASE = 1**, it indicates that your forecasting model is as accurate as the naive model. In other words, your model's performance is comparable to a simple baseline model.

If **MASE < 1**, it suggests that your forecasting model is better than the naive model. A lower MASE indicates improved forecasting accuracy compared to the baseline.

If **MASE > 1**, it suggests that your forecasting model is less accurate than the naive model. A higher MASE indicates poorer forecasting performance compared to the baseline.

MASE for all forecast:

**naive\_forecast:** **2.78062**

**rwf\_forecast:** **2.78062**

**rwf\_forecast\_1:** **2.772454**

**snaive\_forecast:** **1**

**ets\_forecast:** **0.7569917**

**mean\_forecast:** **2.912396**

**holt winter:** **0.8611309**

**forecast\_SSE:2.972526**

**The Best Model, based on the MASE comparison is ETS FORECAST:** **0.7569917, since it has the lowest MASE value among the models i.e the most accurate forecast.**