# QMB6304 - Assignment 5

Name: James Burnett (U00150685)  
Course: QMB6304  
Assignment: Module 5 - Time Series Analysis

## Preprocessing

The dataset '6304 Assignment 5 Data.xlsx' contains monthly beer production in Australia from January 1956 to December 1978. The following steps were executed for preprocessing:

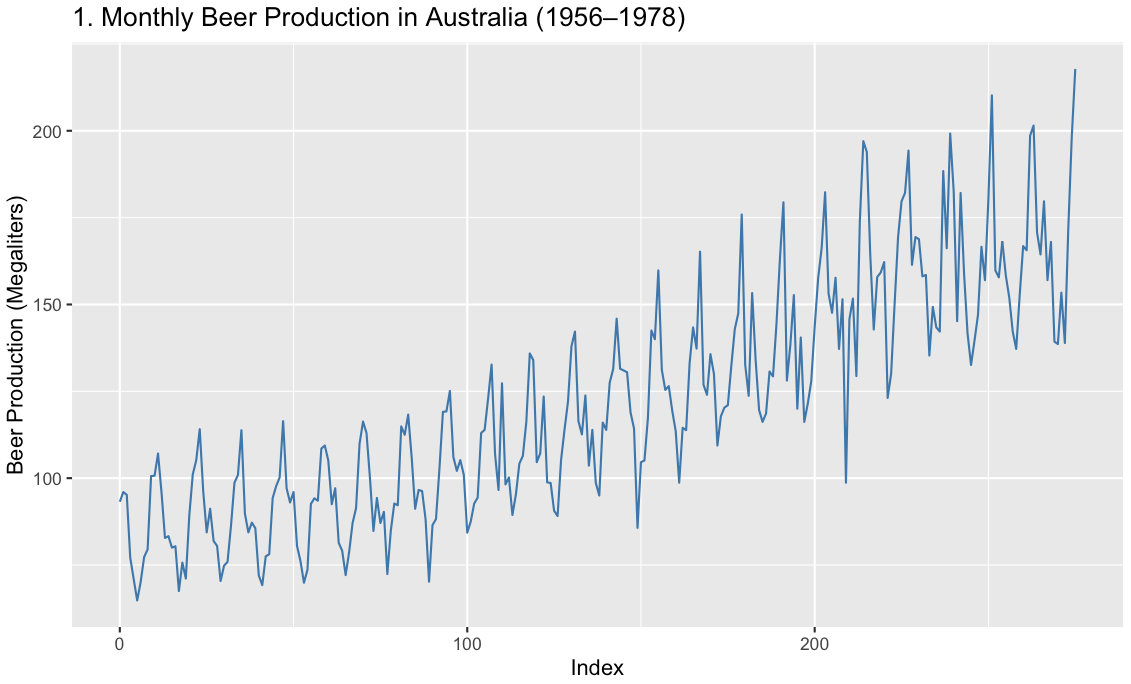
* Loaded the Excel file using `readxl::read\_excel()`.
* Renamed columns from X → index, Month → date, and Monthly.beer.production → production.
* Extracted the year and month from the date column using `lubridate`.

## Analysis

### Line Plot of Production Over Time

Instruction:  
Show a line plot of the data using the index as ‘x’ and production as 'y' variable. Show appropriate main titles and axis titles on the graph.

This plot shows the overall trend and seasonal patterns in beer production over the years.



### Simple Regression Model

Instruction:  
Using all the rows, parameterize a base time series simple regression model using 'index' as the independent variable and 'production' as the dependent variable. Show the summary of your regression output. From this state, the slope of your regression line and the correlation coefficient between actual and predicted values for production.

#### Output:

Slope of regression line: **0.3541**

Correlation coefficient: **0.8529**

### Time Series Plot with Regression Line

The red line shows the linear trend fitted over the time series.

A graph with a red line

AI-generated content may be incorrect.

### Durbin-Watson Test

#### Output

Durbin-Watson statistic: **0.8974**  
Autocorrelation: **0.5351**  
p-value: **0.000**  
  
This indicates significant positive autocorrelation in the residuals, violating the independence assumption of linear regression.

### Seasonal Indices and Deseasonalization

A cyclical pattern was evident in the data, so average production values were calculated for each month to represent seasonal indices. These indices were then used to deseasonalize the monthly production values by adjusting for recurring fluctuations.

### Comparison of Fitted Models

This plot compares the original beer production values with the reseasonalized predictions from both a simple linear model and a second-order polynomial model.

A graph showing a graph of a graph

AI-generated content may be incorrect.

## Appendix

# Repo: https://github.com/burnett013/stats\_mod5.git

# === Load Required Libraries ===

# install.packages(c("lubridate", "ggplot2", "dplyr", "car", "readxl", "httpgd"))

library(readxl)

library(ggplot2)

library(lubridate)

library(dplyr)

library(car) # for Durbin-Watson test

# library(httpgd)

# === Load Data ===

data <- read\_excel("6304 Assignment 5 Data.xlsx")

# === Rename Columns ===

colnames(data) <- c("index", "date", "production")

# === Create Year and Month Columns ===

data$year <- year(data$date)

data$month <- month(data$date)

**# === 1. Line Plot of Production Over Time ===**

ggplot(data, aes(x = index, y = production)) +

geom\_line(color = "steelblue") +

labs(

title = "1. Monthly Beer Production in Australia (1956–1978)",

x = "Index",

y = "Beer Production (Megaliters)"

)

**# === 2. Simple Regression Mdel ===**

model1 <- lm(production ~ index, data = data)

summary(model1)

# Slope and correlation

slope <- coef(model1)[["index"]]

predicted <- predict(model1)

correlation <- cor(data$production, predicted)

cat("Slope of regression line:", slope, "\n")

cat("Correlation coefficient:", correlation, "\n")

**# === 3. Plot with Regression Line ===**

ggplot(data, aes(x = index, y = production)) +

geom\_line(color = "gray") +

geom\_line(aes(y = predicted), color = "red") +

labs(

title = "3. Beer Production with Simple Linear Regression Line",

x = "Index",

y = "Beer Production (Megaliters)"

)

**# === 4. Durbin-Watson Test ===**

dw\_result <- durbinWatsonTest(model1)

print(dw\_result)

**# === 5. Seasonal Index and Deseasonalization ===**

data$month <- factor(data$month, levels = 1:12)

monthly\_avg <- data %>%

group\_by(month) %>%

summarise(avg\_monthly\_prod = mean(production))

# Join Seasonal Indicees

data <- left\_join(data, monthly\_avg, by = "month")

data$seasonal\_index <- data$avg\_monthly\_prod

data$deseasonalized <- data$production / data$seasonal\_index \* mean(data$seasonal\_index)

# === Two Regression Models on Deseasonalized Data ===

# Model A: Simple Linear

model\_a <- lm(deseasonalized ~ index, data = data)

# Model B: Polynomial

model\_b <- lm(deseasonalized ~ index + I(index^2), data = data)

# Reseasonalize Fited Values

data$fit\_a <- predict(model\_a)

data$fit\_b <- predict(model\_b)

data$reseason\_a <- data$fit\_a \* data$seasonal\_index / mean(data$seasonal\_index)

data$reseason\_b <- data$fit\_b \* data$seasonal\_index / mean(data$seasonal\_index)

**# === 6. Plot Original vs Reseasonalized Models ===**

ggplot(data, aes(x = index)) +

geom\_line(aes(y = production), color = "black", linewidth = 1, alpha = 0.6) +

geom\_line(aes(y = reseason\_a), color = "blue", linetype = "dashed") +

geom\_line(aes(y = reseason\_b), color = "green", linetype = "dotted") +

labs(

title = "6. Original vs Fitted Values from Two Models",

x = "Index",

y = "Beer Production (Megaliters)",

caption = "Black = Original, Blue = Linear Model, Green = Polynomial Model"

)