ST5209/X Assignment 1

Due 2 Feb, 11.59pm

Set up

1. Make sure you have the following installed on your system: LaTeX, R4.2.2+, RStudio 2023.12+, and Quarto 1.3.450+.
2. Clone the course repo.
3. Create a separate folder in the root directory of the repo, label it with your name, e.g. `yanshuo-assignments`
4. Copy the assignment1.qmd file over to this directory.
5. Modify the duplicated document with your solutions, writing all R code as code chunks.
6. When running code, make sure your working directory is set to be the folder with your assignment .qmd file, e.g. `yanshuo-assignments`. This is to ensure that all file paths are valid.[1]

Submission

1. Render the document to get a .pdf printout.
2. Submit both the .qmd and .pdf files to Canvas.

Question 1 (Quarto)

Read the guide on using Quarto with R and answer the following questions:

a) Write a code chunk that imports `tidyverse` and `fpp3`.
b) Modify the chunk so that only the following output is shown (i.e. the usual output about attaching packages and conflicts is not shown.)

[1] You may view and set the working directory using `getwd()` and `setwd()`.

```
library(tidyverse)
library(fpp3)
```

c) Modify the chunk so that it is executed but no code is shown at all when rendered to a pdf.
d) Modify the document so that your name is printed on it beneath the title.

Question 2 (Livestock)

Consider the `aus_livestock` dataset loaded in the `fpp3` package.

a) Use `filter()` to extract a time series comprising the monthly total number of pigs slaughtered in Victoria, Australia, from Jul 1972 to Dec 2018.
b) Make a time plot of the resulting time series.

Question 3 (Beer production)

Consider the `aus_production` dataset loaded in the `fpp3` package. We will study the column measuring the production of beer.

a) Make a time plot of the beer production time series.
b) Describe the observed trend.
c) Make a seasonal plot.
d) What is the period of the seasonality?
e) Describe the seasonal behavior.

Question 4 (Pelts)

Consider the `pelt` dataset loaded in the `fpp3` package, which measures the Hudson Bay Company trading records for Snowshoe Hare and Canadian Lynx furs from 1845 to 1935.

a) Plot both time series on the same axes. *Hint: Use `pivot_longer()` to create a key column.*
b) What happens when you try to use `gg_season()` to the lynx fur time series? What is producing the error?
c) Make a lag plot with the first 20 lags. Which lags display strong positive correlation? Which lags display strong negative correlation? Verify this with the time plot.
d) If you were to guess the seasonality period based on the lag plot, what would it be?

e) Use the provided function `gg_custom_season()` in `_code/plot_util.R`[2] to make a seasonal plot for lynx furs with the period that you guessed.[3] Does the resulting plot suggest seasonality? Why or why not?

Question 5 (Box-Cox, Q3.3 in FPP)

Why is the Box-Cox transform unhelpful for the `canadian_gas` data?

Question 6 (Decomposition with outliers, Q3.7 in FPP)

Consider the last five years of the Gas data from `aus_production` .

```
gas <- tail(aus_production, 5*4) |> select(Gas)
```

a. Plot the time series. Can you identify seasonal fluctuations and/or a trend-cycle?

b. Use `classical_decomposition` with `type=multiplicative` to calculate the trend-cycle and seasonal indices.

c. Do the results support the graphical interpretation from part a?

d. Compute and plot the seasonally adjusted data.

e. Change one observation to be an outlier by running the following snippet:

```
# Change to eval: TRUE in order to run
gas |>
  mutate(Gas = if_else(Quarter == yearquarter("2007Q4"), Gas + 300, Gas))
```

Recompute the decomposition. What is the effect of the outlier on the seasonally adjusted data?

f. Does it make any difference if the outlier is near the end rather than in the middle of the time series?

[2] You can load this function using `source("../_code/plot.util.R")`.
[3] Unfortunately, it seems '`gg_season()` does not allow this functionality.

Question 7 (STL decomposition, Q3.10 in FPP)

Consider the `canadian_gas` dataset.

a. Do an STL decomposition of the data.

b. How does the seasonal shape change over time? [Hint: Try plotting the seasonal component using `gg_season()`.]

c. Apply a calendar adjustment and compute the STL decomposition again. What is the effect on the seasonal shape?