

ST5209/X Assignment 3

Due 4 Mar, 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. Pull changes from 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.

1. Holt-Winters, residuals, and forecast accuracy

Consider the antidiabetic drug sales time series which can be loaded using the following code snippet.

```
diabetes <- read_rds("../_data/cleaned/diabetes.rds") |>
  select(TotalC)
```

- a. Fit the following exponential smoothing models on the entire time series:
 - Holt-Winters with multiplicative noise and seasonality,

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

- Holt-Winters with a log transformation, with additive noise and seasonality,
 - Holt-Winters with multiplicative noise and seasonality, and damping.
- b. Make ACF plots for the innovation residuals of these three models. What can you say about stationarity of the residuals from the plot?
 - c. Calculate the p-value from a Ljung-Box test on the residuals with lag $h = 8$. What can you say about the stationarity of the residuals from the p-value? What does this mean about the model?
 - d. Perform time series cross-validation for the three methods, using `.init = 50` and `.step = 10`, and with the forecast horizon $h = 4$. Which method has the best RMSSE? How many data points is the error averaged over in total?

2. Moving averages and differences

Consider the linear trend model

$$X_t = \beta_0 + \beta_1 t + W_t.$$

Define a time series (Y_t) by taking a moving average of with a symmetric window of size 7. Define another times series (Z_t) by taking a difference.

- a. What is the mean function for (Y_t) ? What is the ACVF for (Y_t) ?
- b. What is the mean function for (Z_t) ? What is its ACVF?
- c. What is the CCF of (Y_t) and (Z_t) ?
- d. Are (Y_t) and (Z_t) jointly stationary?

3. Sample vs population ACF

Consider the signal plus noise model

$$X_t = \sin(2\pi t/5) + W_t.$$

- a. What is the ACF of (X_t) ?
- b. Simulate a time series X_1, X_2, \dots, X_{200} from this model and plot its sample ACF.
- c. Why does the sample ACF not look like the population ACF function?
- d. Why does the asymptotic normality theorem for the ACF not apply?