

Stat 443 Assignment 1: Exploratory Data Analysis

Caden Hewlett

January 21, 2024

Task 1: Analyzing Usual Hours Worked in Canada

a) Read in the data and create a time-series object. Plot the series and comment on any features of the data that you observe. In particular address the following points:

- Does the series have a trend?
- Is there seasonal variation and if so would an additive or multiplicative model be suitable? Explain your reasoning.
- Is the series stationary? Justify referring to the definition of a weakly stationary stochastic process.

```
# Code block for Task 1a
```

b) Create training and test datasets.

The training dataset should include all observations up to and including December 2021; this dataset will be used to fit ("train") the model. The test dataset should include all observations from January 2022 to December 2023; this dataset will be used to assess forecast accuracy. You can use the command `window()` on a ts object to split the data. Using a suitable decomposition model and the loess method (R function `stl()`) decompose the training series into trend, seasonal, and error components. Plot the resulting decomposition.

```
# Code block for Task 1b
```

c) Fit a linear model to the trend component (you can use R function `lm()`).

- Write down the fitted model for the trend component.
- Does the linear model provide evidence of a trend at the 95% confidence level?
- Without doing any further analysis, would you use this trend component to make predictions? Justify your answer using the linear model results and the trend component plot.

```
# Code block for Task 1c
```

d) Predict the monthly average values of the usual hours worked in Canada for the period from January 2022 to December 2023 using your seasonal decomposition model.

- Plot your predictions along with the actual observed values (on the same plot). Make sure to include a legend for your plot.
- Comment on the performance of your prediction method, explaining why or why not the method worked well for this data.

- How could the prediction method be improved?
- As a statistician, what other information would you like to add to your forecasts in addition to the point forecasts you produced above?

Code block for Task 1d

Task 2: Analyzing New York Temperature Data

a) Read the data into R and create an R object called `dat` for the data.

Code block for Task 2a

b) Create `zoo` objects for daily Max Temperature. Create monthly maxima time series. Plot the monthly maximum temperature series and comment on any features you observe.

Code block for Task 2b

c) Fit a suitable seasonal decomposition model to the monthly data using the moving average smoothing (R function `decompose`) and plot the estimates of the trend, seasonal, and error components.

Code block for Task 2c

d) Plot the correlogram for the deseasonalized series of monthly temperature maxima using the seasonal decomposition model you fit in part (c). Comment on the serial dependence of this series.

Code block for Task 2d

Task 3: Conducting a Simulation Study on the Autocorrelation Coefficient

i)

Simulate a time series of length $n = 2000$ from a white noise process with $Z_t \sim N(0, 1)$ (function `rnorm()`).

Code block for Task 3i

ii) Evaluate the sample autocorrelation coefficient.

At lag h for $h = 1$ and $h = 2$. Store these values.

Code block for Task 3ii

iii) Repeat steps (i) and (ii) $m = 8000$ times;

i.e. generate 8000 time series of length n and for each of them compute $r1$ and $r2$. You should now have two vectors of length m with estimates $r1$ and $r2$.

```
# Code block for Task 3iii
```

To summarize the results of the simulation study:

- Compute the mean and variance of $r1$ and $r2$ values from your simulation study.
- In two separate figures plot the two histograms for the sample of $r1$ and $r2$ values from the simulation study (function `hist()`) add the smoothed version of the histogram (function `density()`) and the theoretical asymptotic normal density (function `dnorm()`). Make sure your plots are well-presented including a suitable title, axes labels, curves of different type or colour, and a legend.
- Comment whether there is an agreement between the empirical estimates of the bias, variance, and sampling density of the estimator of the autocorrelation at lag h and their theoretical approximation.

```
# Code block for summarizing the simulation study results
```