

## DA5020 – Assignment 9

This assignment provides you with an opportunity to create several forecasting models and evaluate the predictions.

The Organization for Economic Co-operation and Development (OECD) publishes data on the annual meat consumption for several countries.

### Question 1 — (0 points)

Load the data into your R environment directly from the URL. Ensure that you inspect the data, so that you know how to identify the necessary columns.

### Question 2 — (5 points)

Extract the **poultry** consumption data, from 1994 to 2014, for **Mexico**, that is measured in thousand tonnes of carcass weight. Pay close attention to the SUBJECT and MEASURE fields to filter the appropriate type of meat and the correct measurement. Visualize the extracted data, using a line chart, and comment on the trend.

**Use the extracted poultry data to answer the questions below.**

### Question 3 — (5 points)

Forecast the poultry consumption for 2014, using a simple moving average of the following four time periods: 2010, 2011, 2012 and 2013. After which, calculate the error (i.e. the difference between the actual and the predicted values). Evaluate the results; how does it compare to the actual data for 2014?

### Question 4 — (5 points)

Forecast the poultry consumption for 2014, using a three year weighted moving average. Apply the following weights: 5, 7, and 15 for the respective years 2011, 2012, and 2013. After which, calculate the error and evaluate the result from your prediction.

### Question 5 — (10 points)

Forecast the poultry consumption for 2014 using exponential smoothing (alpha is 0.9). Comment on the prediction for 2014 with the actual value. Note: use data from 2004 to 2013 to build your model.

### Question 6 — (15 points)

Build a simple linear regression model using the TIME and VALUE for all data from 2004 to 2013. After which, forecast the poultry consumption for 2014 to 2016. Comment on the results. **Note: Your predictions should be calculated using the coefficients. Do not use any libraries to make your predictions.**

### Question 7 — (10 points)

Calculate the mean squared error (MSE) for the models in (5 and 6) above based on the data from 2004 to 2014. Perform this step by step, using each model to make a forecast for each given time period, then calculate the squared error for each observation. After which average the squared errors. Which model has the smallest (MSE)? **Note: do not use any libraries in your calculations.**

### Question 8 — (+5 bonus points)

Visualize the actual data and the forecast for each model that you created in questions 3, 4, 5 and 6, using a line chart. First, create the line chart using the original data from 1994 to 2014. Then, add a `geom_point()` to each chart to indicate the forecast from each model (color the point red). You should

produce four charts, one for each model. Reflect on all the models that you created in this assignment, and explain which model gave the most accurate forecast for 2014, when compared to the actual value.

### Submission Details

- Your submission must contain two files: the .Rmd notebook and a knitted PDF or HTML (from the notebook). Name your .Rmd R Notebook, DA5020.A9.FirstName.LastName.Rmd and your PDF/HTML DA5020.A9.FirstName.LastName.{pdf,html}, where *FirstName.LastName* is your first and last name.
- The .Rmd file must be fully commented and properly "chunked" R code and detailed explanations. Make sure that it is easy to recognize which question you answer and that your code runs from beginning to end (because that is how we will test it). Code that doesn't execute, stops, throws errors will receive no points. If the TAs have to "debug" your code or spend any effort getting it to run, substantial points will be deducted.
- Not submitting a knitted PDF or HTML will result in reduction of 30 points.
- Not submitting the .Rmd file (or both) will result in a score of 0.