

ST662 Topics in Data Analytics
2018-19 Semester 2
Assignment Sheet 4

Due at 2pm Monday 8th April 2019

1. The dataset Dates.csv contains 3000 dates from years 2000 to 2015. Read it into SAS.
 - (a) Create a new variable which contains the date in format DD/MM/YYYY.
 - (b) Write code to screen the dataset.
 - (c) List any errors identified.

2. The dataset Bricks.csv contains information on Australian quarterly clay brick production from 1956 to 1994. Read the data into SAS.
 - (a) Create a single date variable from the year and quarter variables, and format it so that it reads as quarterly data. Hint: explore the YYQ function and format 'yyqs8.'.
 - (b) Create a time series plot of the data and comment (briefly - one to two sentences) on the effects (or not) of season, cycle and trend.
 - (c) Use an appropriate exponential smoothing method to forecast to the end of 1996. In your answer, state which type of exponential smoothing you used and why, provide a graph illustrating the forecasts, and give a table of the forecasts with confidence limits.

3. The dataset LakeHuron.csv contains annual depth measurements at a specific site on Lake Huron from 1875 to 1972. Read the data into SAS.
 - (a) Create four new variables that contain the time series depth measurements at lag 1 to 4.
 - (b) Generate scatterplots of depth versus each lag variable.
 - (c) Comment on autocorrelation in the data.

Details on what you have to submit for this assignment

Submission of this assignment is in two parts:

1. Submit on Moodle the SAS programme (code only, no output) you created to address Qu's 1-3 above. This must be done before the start of class, i.e. BEFORE 2PM. Do not leave this until the last minute as Moodle submission will close at this time.
2. Submit a printed hard copy of your programme (code only, no output), a hard copy of your answer to Qu 1 (c), Qu 2 (b) and (c), and Qu 3 (b) and (c). This will be submitted at the beginning of the lecture at 2pm.