DATS 6450 **Name (Print):** ____________________

Time series Anlysis & Modeling

Fall 2020

Exam #1

10/14/2020

Time Limit: 120 Minutes

This exam contains 10 pages (including this cover page) and 4 problem(s). Check to see if any pages are missing. Enter all requested information on the top of this page, and put your initials on the top of every page, in case the pages become separated.

- This is a closed book test and you must *not* use your books, notes on this test.
- You are allowed to use a*calculator*. You are *not* allowed to use cell phone or computer during the test (except for taking snap shots and uploading the solution through BB).
- This is an individual test and collaboration is *not* allowed.
- The exam is out of 80 points. The 20 *bonus points* are marked under footnotes. Maximum grade attained is 100.
- You are required to show all your work on each problem on this test.
- **Organize your work**, in a reasonably neat and coherent way, in the space provided.
- **Mysterious or unsupported answers will not receive full credit**.
- If you need more space, use the back of the pages; clearly indicate when you have done this.

Do not write in the table to the right.

Problem	Points	Score
1	6	
2	4	
3	44	
4	46	
Total:	100	

Formulas

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{T} (y_t - \overline{y})^2}$$

$$r = \frac{\sum (x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum (x_t - \overline{x})^2} \sqrt{\sum (y_t - \overline{y})^2}}$$

$$\hat{\sigma}_e^2 = \frac{1}{T - k - 1} (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$\hat{y} \pm 1.96 \hat{\sigma}_e \sqrt{1 + \mathbf{x}^* (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{x}^*)^T}$$

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha) \hat{y}_{t|t-1}$$

1. Consider the nine time series data plotted in Figure(1).

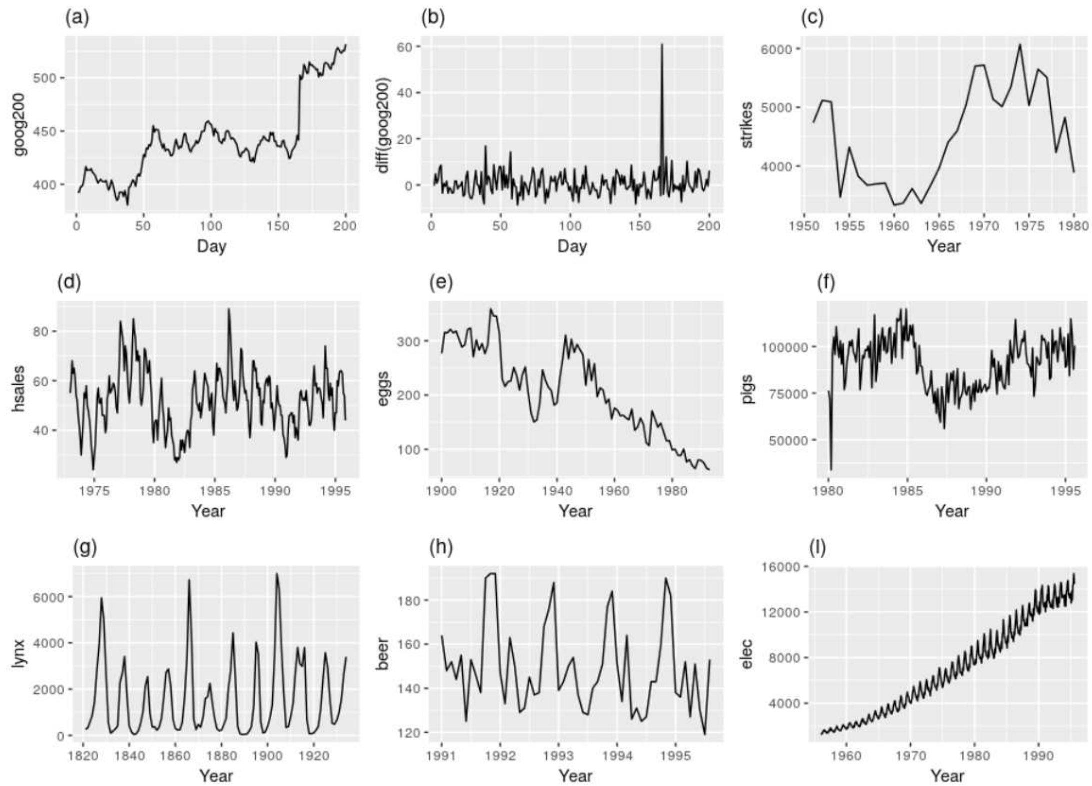


Figure 1: Time series data

(a) (2 points) Which <u>one</u> of above dataset shows a *seasonal pattern*? Justify your answer.

(b) (2 points) Which <u>one</u> of above dataset shows a *trend pattern*? Justify your answer.

(c) (2 points) Which <u>one</u> of above dataset is considered *stationary*? Justify your answer.

2. (4 points) The following time series and ACF plots corresponds to four different time series (Figure(2)). Your task is to match each time plot in the first row with one of the ACF plots in the second row. Justify your answer.

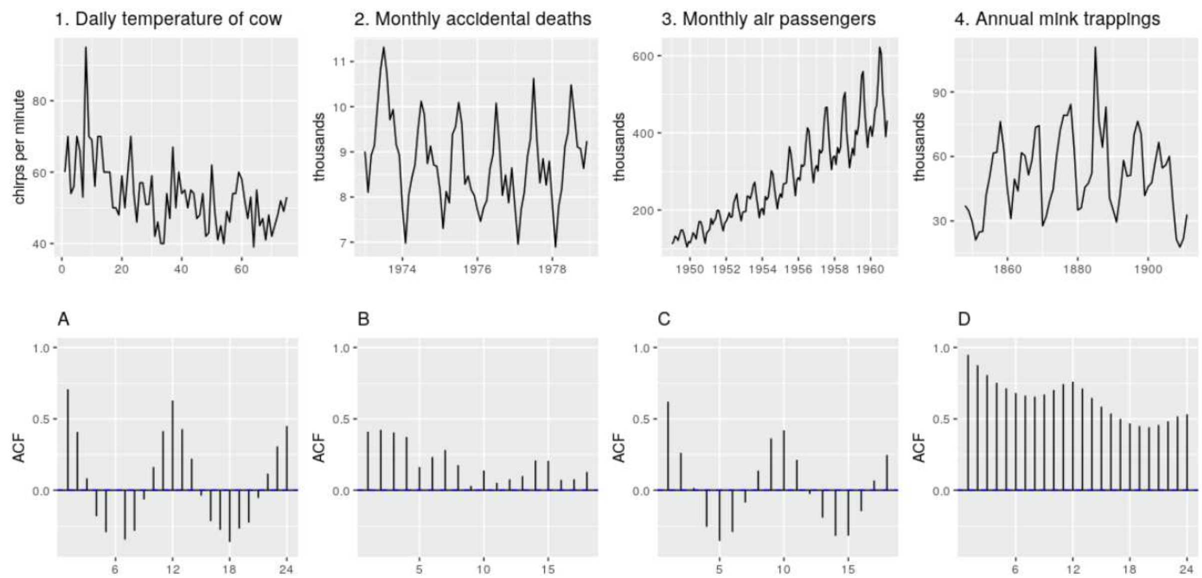
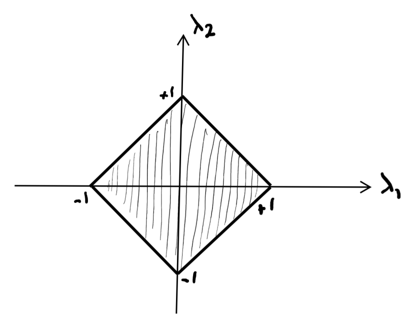


Figure 2: Time series data

3. Let X and Y be a continuous random variables with the following joint density function given as :



$$f_{X,Y}(\lambda_1, \lambda_2) = \begin{cases} c, & -1 < \lambda_1 < 1 \\ & -1 < \lambda_2 < 1 \\ 0, & \text{Else} \end{cases}$$

(a) (4 points) Determine the *constant* c.

(b) (4 points) Determine the *marginal density* $f_X(\lambda_1)$? Graph $f_X(\lambda_1)$ versus λ_1 and show this is a valid density function. *Hint: Find the side line equations first.*

(c) (4 points) [‡] Determine the *marginal density* $f_Y(\lambda_2)$? Graph $f_Y(\lambda_2)$ versus λ_2 and show this is a valid density function.

(d) (4 points) [§] Are random variable X and Y *independent*? Justify your answer.

(e) (4 points) Determine the *probability* of $Pr\{-0.5 < X < -.25\}$.

(f) (4 points) What is the *correlation coefficient* between random variable X and Y. *Hint: There is no need to for any mathematical calculations to answer this question. Pick an approximate value and justify your answer.*

[‡] Extra Credit

[§] Extra Credit

(g) (6 points) Determine the *conditional density function* $f_{X|Y}(\lambda_1|\lambda_2 = 0.5)$ and graph it.

(h) (6 points) Determine the *probability* of $Pr\{-0.5 < X < -.25|Y = 0.5\}$.

(i) (4 points) Calculate the $\mu_X = E[X]$.

(j) (4 points) * Calculate the *variance* of X using the following equation:

$$\sigma_X^2 = E[X^2] - \mu_X^2$$

*Extra Credit

4. A researcher wants to determine how employee salaries at a company are related to the length of employment. The researcher selects 3 employees from the company and obtains the data shown in the table: (Note: this is a dummy data set)

Sample #	Employee	Salary ($)	Employment (in years)
1	A	5	2
2	B	6	3
3	C	12	4

Table 1: Employee salary and length of employment

(a) (4 points) Find a *regression equation* that estimate the salary based on the length of employment.

(b) (4 points) Find the *prediction interval* of employee A with 95% confidence.

(c) (4 points) † Calculate the *coefficient of determination* for this estimate. What is the actual meaning of coefficient determination?

(d) (4 points) Calculate the *correlation coefficient* between true salaries and the estimated salaries and compare it with the root square of correlation of determination. What is the relationship?

(e) (3 points) †† Calculate the first three terms of the *residuals autocorrelation* and graph the result. Then calculate the Q value for this estimate.

†Extra Credit

††Extra Credit

(f) (3 points) [†] Calculate the *coefficient between residual errors and the employment*(in years). Interpret the actual meaning of this coefficient for this estimate.

(g) (4 points) Let suppose you are performing single-step-ahead prediction, utilizing the *Average* method, find $\hat{y}_{3|2}$?

(h) (4 points) Let suppose you are performing single-step-ahead prediction, utilizing the *Naïve* method , find $\hat{y}_{3|2}$?

(i) (4 points) Let suppose you are performing single-step-ahead prediction, utilizing the *Drift* method, find $\hat{y}_{3|2}$?

(j) (4 points) Let suppose you are performing single-step-ahead prediction, utilizing the *regression model*(derived in (a), find $\hat{y}_3$?

(k) (4 points) Let suppose you are performing single-step-ahead prediction, utilizing the *simple exponential smoothing* method, with zero initial condition and smoothing factor of 0.8. Find $\hat{y}_{3|2}$?

[†]Extra Credit

(l) (4 points) Pick the best one-step-ahead predictor for this dataset just by looking at the residuals for $\hat{y}_3$. Justify your answer.