



QQI

HIGHER DIPLOMA IN SCIENCE IN DATA ANALYTICS

FINAL EXAMINATIONS

<i>Module Code:</i>	B8IT109
<i>Module Description:</i>	Advanced Data Analytics
<i>Examiner:</i>	Dr Shahram Azizi Sazi, Paul Laird
<i>Internal Moderator:</i> <i>External Examiner:</i>	Dr Amir Esmaeily

INSTRUCTIONS TO CANDIDATES

1. *Solve all questions with R. R code and necessary outputs need to be saved in R format. Please submit the R code and the required graphs/plots/curves in the zipped folder.*
2. *Select only one question from questions 3 and 4. Explicitly mention your optional question on the top of exam paper.*
3. *In the case, where the explanation of your output is in R, it is required to mention explicitly on the paper exam. Notice that your explanation should be consistent with your output. This exam assesses you on the development of the advanced techniques for data analytics mainly covering:*

- *Descriptive analysis*
- *probability models*
- *Decision making techniques*
- *Time series Analysis*
- *GLM Analysis*
- *Multivariate Analysis*
- *Data Analysis in Multi-agent systems*

Question 1

In a traffic management system (TMS), two traffic sensors are designed in a specific street. The first sensor records the waiting time and the second sensor records the buffer (the number of cars). On average, the rate of buffer is 7.

- (a) What is the appropriate probability model of data in each sensor?
(5 Marks)
- (b) Express the relation between your proposed models?
(5 Marks)
- (c) Generate 100 samples for each sensor. Frame all samples into one dataset.
(5 Marks)
- (d) Provide the descriptive analysis for your dataset (e.g. summary, boxplot, ...).
Interpret your quick insight about the simulated dataset.
(5 Marks)
- (e) In a multi-agent system (MAS), two agents are active. Each agent A_j , is Bernoulli with $p_j = \frac{j}{1+j}, j = 1, 2$. Simulate 40 samples for two agents and test whether the population proportion of the first agent (p_1), is significantly different from the proportion of the second agent (p_2) at the level $\alpha = 0.05$. To do so,
 - I. State the null and alternative hypotheses.
(5 Marks)
 - II. What is your decision rule and explain your decision?
(10 Marks)
 - III. Provide the value of Biasedness for each agent.
(5 Marks)

(TOTAL: 40 Marks)

Question 2

Use the AirPassengers built in dataset.

Examine the data visually and descriptively, recording your commands and your conclusions drawn from the output.

(5 marks)

Investigate whether the data are stationary, and if not, make the data stationary, describing the actions you took to achieve this.

(5 marks)

Investigate the (partial) autocorrelation functions, and fit the most appropriate AR(I)MA model to the data

(5 marks)

Use the first ten years' data to predict the last two using your model, and comment on the result.

(5 marks)

(Total: 20 Marks)

Question 3

Use the dataset available at:

<http://users.stat.ufl.edu/~winner/data/ingots.dat> , with description at:
<http://users.stat.ufl.edu/~winner/data/ingots.txt> .

Note that there are two dependent variables.

Test whether there is a significant relationship between each independent variable and the breaking strain at $\alpha=0.05$

(7 marks)

Construct a model including all variables, and refine iteratively, noting your justification for each refinement.

(8 marks)

Investigate whether there are any interactions between the predictors, and produce a final model.

(5 marks)

(Total: 20 Marks)

Question 4

Use dataset available on http://users.stat.ufl.edu/~winner/data/hotel_energy.csv , then:

- (a) Train the model using 80% of this dataset and suggest an appropriate GLM to model **effrooms** to **other quantitative variables** variables.

(5 Marks)

- (b) Specify the significant variables on **effrooms** at the level of $\alpha=0.05$, and estimate the parameters of your model.

(5 Marks)

- (c) Predict the test dataset using the trained model.

(5 Marks)

- (d) Propose an appropriate measure of prediction performance and evaluate this for predicted values.

(5 Marks)

(Total: 20 Marks)

Question 5

Five agents $A_j = j, j = 1, \dots, 5$, are connected in a MAS. Generate 100 samples from exponential model with $\lambda_j = j, j = 1, \dots, 5$,

- (a) Adopt a centralized scheme so that the central node is the fifth agent.

(10 Marks)

- (b) Compute the normalized weights and find the global arithmetic and geometric mean for the agent A5.

(10 Marks)

(Total: 20 Marks)

#####End of Examination #####