# Dublin Business School

# Assessment Brief

# Assessment Details

|  |  |
| --- | --- |
| Module Title: | Statistics for Data Analytics |
| Module Code: | B9DA101 |
| Module Leader: | Dr. Shahram Azizi Sazi, Dr. Muhammad Alli |
| Stage (if relevant): |  |
| Assessment Title: | CA two |
| Assessment Number (if relevant): |  |
| Assessment Type: |  |
| Restrictions on Time/Length : | Submission before deadline |
| Individual/Group: | Individual |
| Assessment Weighting: |  |
| Issue Date: |  |
| Hand In Date: | See Moodle page |
| Planned Feedback Date: |  |
| Mode of Submission: | Online |

**Guideline:**

* This CA assesses students on core concept in GLM, time series analysis, and Bayesian analytics.
* All questions are mandatory.
* Use R/Rstudio to solve questions and perform analytics.
* Any submission after deadline will not be considered and scored.

**Question 1**

Consider a relational dataset and specify your input and output variables , then:

* 1. Train the model using 80% of this dataset and suggest an appropriate GLM to model **ouput** to **input** variables.

**(10 Marks)**

* 1. Specify the significant variables on the **output** variable at the level of 𝛼=0.05 and explore the related hypotheses test. Estimate the parameters of your model.

**(10 Marks)**

* 1. Predict the output of the test dataset using the trained model. Provide the functional form of the optimal predictive model.

**(10 Marks)**

* 1. Provide the confusion matrix and obtain the probability of correctness of predictions.

**(5 Marks)**

**(Total: 35 Marks)**

**Question 2**

Let are identically independently distributed (iid) with Poisson().

1. Compute the likelihood function (LF**). (10 Marks)**

1. Adopt the appropriate conjugate prior to the parameter (Hint: Choose hyperparameters optionally within the support of distribution). (**10 Marks)**
2. Using (a) and (b), find the posterior distribution of . **(10 Marks)**
3. Compute the minimum Bayesian risk estimator of . **(5 Marks)**

**(Total: 35 Marks)**

**Question 3**

Use the a particular stock market dataset and apply the following steps to accomplish the time series analysis:

1. Check whether the time series is stationary in mean and variance.

**(5 Marks)**

1. Use acf() and pacf() functions to identify the order of AR and MA.

**(10 Marks)**

**c)** Use auto.arima() to learn the best ARIMA model.   **(5 Marks)**

**d)** Forecast h=10 step ahead prediction of the time series variable and plot it with the original time series.     **(10 Marks)**

**(Total: 30 Marks)**