# Assessment Brief

# Assessment Details

|  |  |
| --- | --- |
| Module Title: | Statistics for Data Analytics |
| Module Code: | B8IT152 |
| Module Leader: | Dr Shahram Azizi |
| Stage (if relevant): |  |
| Assessment Title: | CA One |
| Assessment Number (if relevant): |  |
| Assessment Type: |  |
| Restrictions on Time/Length : | Submission before deadline |
| Individual/Group: | Group |
| Assessment Weighting: |  |
| Issue Date: |  |
| Hand In Date: |  |
| Planned Feedback Date: |  |
| Mode of Submission: | Online |

**Guideline:**

* This CA assesses students on core concept in descriptive analytics, discrete and continuous probability models and hypotheses tests.
* All questions are mandatory.
* Use R/Rstudio to solve questions and perform analytics.
* Any submission after deadline will not be considered and scored.

Consider a real-world, relational dataset. This dataset must have at least 2 categorical and 2 continuous variables.

**Question 1 (35 Marks)**

* 1. Describe the dataset using appropriate plots/curves/charts,… (7)
  2. Consider one of continuous attributes, and compute central and variational measures. (8)
  3. For a particular variable of the dataset, use Chebyshev's rule, and propose one-sigma interval. Based on your proposed interval, specify the outliers if any. (10)
  4. Explain how the box-plot technique can be used to detect outliers. Apply this technique for one attribute of the dataset (10)

**Question 2 (35 Marks)**

1. Select four variables of the dataset, and propose an appropriate probability model to quantify uncertainty of each variable.  **(10)**

1. For each model in part (a), estimate the parameters of model. (10)

1. Express the way in which each model can be used for the predictive analytics, then find the prediction for each attribute. (15)

**Question 3 (30 Marks)**

(a) Consider two categorical variables of the dataset, develop a binary decision making strategy to check whether two variables are independent at the significant level alpha=0.01. To do so, (10)

1. State the hypotheses.
2. Find the statistic and critical values.
3. Explain your decision and Interpret results.

(b) Consider one categorical variable, apply goodness of fit test to evaluate whether a candidate set of probabilities can be appropriate to quantify the uncertainty of class frequency at the significant level alpha=0.05. (10)

(c) Consider one continuous variable in the dataset, and apply test of mean for a proposed candidate of at the significant level alpha=0.05. (10)