

---

# MATH 4044 – Statistics for Data Sciences

Take-home Exam SP5 2022

Due Sun 13 Nov 2022 by 11:55pm

---

## Instructions

- This take-home is worth 30% of your final mark. It is due no later than 11:55pm on Sunday 13 November 2022.
- You will need to submit your assignment via learnonline.
- The submitted assignment needs to be a **single file**, in either a Microsoft Word (doc or docx) or pdf file format, **25 pages at most excluding any appendices**.
- The assignment is out of 100 marks. To achieve maximum marks for each question, you should aim to:
  - Complete the requested statistical analysis in SAS using appropriate tasks or procedures (40%).
  - Include only the output most relevant to the question and interpret all key results (40%). Do not include every piece of output produced by SAS!
  - Discuss the results more broadly in the context of the given scenario (20%).
- Work submitted late, without an extension being granted, will not be accepted.

## Data Description

### HR Attrition

HR Analytics helps company to identify trends and pattern in employing, helping company to set strategies and policies for a profitable and healthy working environment. In this case study, we will explore factors that affect employee attrition. Understanding these features will help companies developing appropriate policies to maintain their workforce.

The data is modified from the data downloaded from <https://www.kaggle.com/datasets/itssuru/hr-employee-attrition>.

### Data file for this take-home exam

The data file for this exam is called `HRAAttritionExam.sas7bdat`. Variables in that file are as follows:

Variable	Description
age	Age of the Employee
Gender	employee gender (female/ male)
Department	Employee's department ( <i>Human Resource, Research &amp; Development and Sales</i> )
TotalWorkingYears	Total working years of employee
Education	Employee's highest qualification 1:Below college; 2: College; 3: Bachelor; 4: Master; 5: Doctor
EducationField	Fields of education: Human Resource, Life Sciences, Marketing, Medical, Tehcnical Degree, and Other
MonthlyIncome	Employee's monthly income
BusinessTravel	<i>Non-travel, Travel_Rarely, Travel_Frequently</i>
JobRole	<i>Healthcare Representative, Human Resources, Laboratory Technician, Manager, Manufacturing Director, Research Director, Research Scientist, Sales Executive, Sales Representative</i>
MaritalStatus	<i>Single, Married, Divorced</i>
NumCompaniesWorked	number of companies worked before current.
OverTime	<i>Yes, No</i>
YearsAtCompany	number of years work at current company
Attrition	<i>Yes, No</i>

## Take-home Exam Tasks

For questions 1–3 below:

- Identify appropriate statistical measures and procedure;
- Check the necessary conditions;
- Use SAS to generate appropriate output;
- Interpret results.

**Question 1 (20 marks)** Is there a statistically significant relationship between `attrition` and each of the following variables: `age` and `TotalWorkingYear`?

Perform appropriate hypothesis test.

**Question 2 (30 marks)** Is there a statistically significant relationship between `attrition` and each of the following variables: `EducationField`, `BusinessTravel` and `JobRole`?

Perform appropriate hypothesis tests and illustrate statistically significant relationships using mosaic plots.

**Question 3 (30 marks)**

Build a logistic model to predict attrition events. Investigate the factors that affect attrition.

You can use backward selection to help you find the simplest model with the highest predictive power. Report and interpret in details only your final model, but do indicate how it was obtained and why it was considered the best.

Consider as many numerical and categorical variables as possible. You may also wish to fit a model with interactions but this is not strictly required.

Interpret the odds ratios produced by your model.

*Warning: including too many interaction term may cause SAS to freeze.*

**Question 4 (20 marks)** Write a summary of your findings from Questions 1–3. In particular, indicate the important factors that affect attrition, and what policies company can implement to address it. Keep the technical details of the analysis that led you to these conclusions to the absolute minimum. Rather, focus on practical significance and present your findings in non-specialist terms. One to two paragraphs (up to a page) will be sufficient.