

# MS4S08-Applied Statistics for Data Science

## 2020/2021 Assessment 2

Deadline: Tuesday 12th January 2021, 21:00 PM

This assessment is worth 50% of your overall mark for this module. This is to be submitted via Turnitin before the deadline in PDF format. You have to use SAS UE to perform your statistical analyses and a formal report summarising your output and results should be produced. Your findings should be interpreted and valid conclusions drawn.

You are required to consider a topic of interest to you and find 3 data-sets from different sources relevant to this topic. The data sets need not be merged but must each provide a different source of information and provide meaningful insights.

### Task 1- Modelling (30%)

- Using at least one of the data-sets you have selected, you should use at least one modelling method (linear or logistic regression).
- State clearly the method you use, the **assumptions** you make, the model you obtain, the results you produce and your conclusions.

### Task 2 Analysis (50%)

- Using at least two of the data-sets you have selected, you should do a Cluster Analysis and at least one different multivariate technique (PCA or Factor Analysis).
- State clearly the research question you want to answer, the methods you use, the results you produce and your conclusions.

## **Report (20%)**

The report must be well structured, well written and grammatically correct. Titles, heading and figures should be correct and labelled in a meaningful way and referenced accordingly. An important skill of statistical reporting is to be clear and concise; therefore, a page limit has been set (there is a word limit of 3,000 words. This doesn't include figures/tables/references etc.)

Use either New Roman or Arial font, minimum size 11pt font.

Note: Include a reference for the data-sets in your report explaining why they are relevant for your analysis.

Send a copy of these data-sets to **angelica.pachon@southwales.ac.uk** before the deadline in a convenient format (csv, Excel, etc.). The subject of the email must be: MS4S08 CW2 Data-sets.

## Marking Guidelines

	80-100	70-79	60-69	50-59	40-49	30-39	0-29
	Exceptional First	First	Upper 2nd	Lower 2nd	Third	Narrow Fail	Fail
<b>Analysis outline</b>	Professional outline of analysis presented	Detailed purpose of analysis provided	Adequate outline of analysis provided.	Outline of analysis provided but with some flaws	Simple outline of analysis provided, but lacking key detail.	Inadequate outline of analysis provided.	No outline of analysis provided.
<b>Methods used and assumptions made</b>	Sophisticated investigation of assumptions and description of method used	Comprehensive investigation of assumptions and description of method used	Adequate and correct investigation of assumptions and description of methods used.	Investigation of assumptions and description of methods used is provided but with some flaws	Limited investigation of assumptions and description of methods used.	Inadequate investigation of assumptions and description of methods used.	No investigation into assumption and no description of methods used.
<b>Key results and correctness of content</b>	Unanticipated results and implementations presented. Appropriate, substantial, correct and sophisticated nature.	Comprehensive results and implementations, presented and employed well. Appropriate, substantial and correct.	Expected results and implementations presented. All appropriate, largely correct, with few flaws	Not all expected results and implementations presented. All appropriate, largely correct, with few flaws.	Few or simple results and implementations presented. Much appropriate material, but flawed.	Seriously flawed results or no implementation. Appropriate but seriously flawed material.	No results or implementation. Incorrect or inappropriate content.
<b>Conclusions</b>	Deep and critical understanding provided	Thorough understanding shown	Good understanding shown.	Key concepts generally understood.	Some evidence of understanding.	Little of superficial understanding shown.	No evidence of understanding
<b>Report</b>	Like a publishable report, virtually error-free.	Like a publishable report with isolated minor errors.	Can be followed easily with very few errors.	Can be followed easily with some weaknesses	Can be followed with difficulty.	Poor structure or containing significant errors.	Unstructured and with many errors.