



SCSI/SECI2143: PROBABILITY & STATISTICAL DATA ANALYSIS
2020/2021 – SEMESTER 2

PROJECT 2

In this Project 2, you and team members are required to conduct an inference statistical analysis that covers the following analysis based on selected dataset (refer Appendix 1):

- ✓ **Compulsory:**
 - Hypothesis testing 1-sample or 2-sample
 - Correlation
 - Regression
- ✓ **Optional (at least 1):**
 - Goodness of fit test
 - Chi Square test of independence
 - ANOVA

The findings are represented in report (refer Project Report.pdf for the format), video and e-portfolio. Details as follows:

- ✓ One submission folder will be created in e-learning for you to upload a **ZIP file** containing
 - Dataset
 - R Script file
 - Pdf version of project report
- ✓ Presentation video of your project, explaining your findings. Duration of the video is 3 minutes.
- ✓ E-portfolio reflection on what you have learnt while doing this project. This is an individual work.

The due date for this project is **20 June 2021; 11.59 P.M.**

APPENDIX 1: Sources of data (suggestion only):

<https://www.kaggle.com/datasets>

<https://guides.emich.edu/data/free-data>

<https://rsp.uni.edu/secondary-data-sets>

<https://www.dosm.gov.my/v1/>

<https://perso.telecom-paristech.fr/eagan/class/igr204/datasets>

<http://statweb.stanford.edu/~sabatti/data.html>

<https://www.springboard.com/blog/free-public-data-sets-data-science-project/>