

# Assignment 2

Course: STAT2604 Introduction to R Programming and Elementary Data Analysis

Total marks: 100

Due date: 23:59, Nov. 12, 2021

**Please pack your Rmd source code file together with the output html file as one compressed file, and submit only that compressed file onto Moodle. Name the compressed file in the format (Name).(UID)\_A2.**

**Question 1** The `vitcap` data set contains data on vital capacity for 24 workers in the cadmium industry. It has the following columns:

<code>group</code>	group codes are, 1: exposed > 10 years, 3: not exposed
<code>age</code>	age in years
<code>vital.capacity</code>	vital capacity (a measure of lung volume) in liters

No missing values in this data set.

- Load the `vitcap` data set. Draw a scatter plot with `age` on the x-axis, and `vital.capacity` on the y-axis, and color the data points according to `group`. (20 marks)
- Build a simple linear regression model to predict `vital.capacity` from `age`. Plot the fitted linear curve and its confidence interval on the same figure in part a). (40 marks)
- Perform model checking by drawing the residual plot and Normal Q-Q plot. Draw your conclusion about the model checking. (40 marks)