<div align="center">

STATS 3001 / STATS 4104 / STATS 7054
Statistical Modelling III
Assignment 2
2022

</div>

DEADLINE:

- Friday 8th April 2021 5pm (Week 6)

QUESTIONS:

1. For $y > 0$ prove that

$$\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \log y.$$

2. The file `companies.txt` contains the following variables for 79 US companies. The purpose of this exercise is to develop a model for predicting market value from the other variables.

Variable	Description
MarketValue	Market Value of the company (in millions)
Assets	Amount of assets (in millions)
Sales	Amount of sales (in millions)
Profits	Profits (in millions)
CashFlow	Cash Flow (in millions)
Employees	Number of employees (in thousands)

   (a) Read the data into R.
   (b) Perform an EDA of the data - `skim()` and histogram of response is fine.
   (c) Produce scatterplots of MarketValue against each of the other predictors.
   (d) Produce scatterplots of MarketValue against each of the other predictors with both the y-axis and x-axis on a log scale.
   (e) Fit the model,

   $$M1 : MarketValue \sim log(Assets) + log(Sales)$$
   $$+ Profits + CashFlow + log(Employees)$$

   (f) Why did we log some of the variables and not the others.
   (g) Check the assumptions of the linear model `M1`.

<div align="center">

1

</div>

(h) Use the Box-Cox method to find a suitable transformation of the data in the context of the model. State, with justification, your chosen value of λ.

(i) Refit the model, `M1` but with the transformed response variable: denote this `M2`.

(j) Check the assumptions for `M2`.

(k) What is your preferred model `M1` or `M2`? Justify your answer.

(l) Using your preferred model, obtain a 95% prediction interval for MarketValue for a company with

Assets	Sales	Profits	CashFlow	Employees
1065	642	30	59	3.5