# STATS 2107
# Statistical Modelling and Inference II
# Project

*Jono Tuke*

*Semester 2 2017*

## Details

- Due date: 27th October 2017, 5pm.
- Group submission (groups of up to five people with a minimum of 3) self-select yourselves into groups. If needed I will reallocate people to groups to ensure at least 3 in each group.
- A single report is to be submitted for each group.
- The report is to be typed in latex, Rmarkdown, or Word.
- The submitted project must be a pdf.
- Each person also completes an online quiz to indicate the percentage completed by each team member.

## Goal

Find a predictive model for the price of dogs on gumtree by considering the following predictors:

- age of dog,
- whether the dog is a cross breed,
- whether the dog is vaccinated,
- whether the dog is microchipped,
- whether the dog is desexed,
- whether the dog is relinquished, and
- the state the dog is sold in.

## Marks

The project is worth 10% of the final mark of SMI. The breakdown is as follows

| Section | Marks | Allocation |
|---|---|---|
| Introduction | 5 | Group |
| Data description | 5 | Group |
| Cleaning of data | 10 | Group |
| Variable description | 15 | Group |
| Bivariate analysis | 15 | Group |
| Model fitting | 15 | Group |
| Final model | 5 | Group |
| Assumption checking | 15 | Group |
| Prediction | 5 | Group |
| Conclusion | 5 | Group |
| Formatting | 10 | Group |
| Contribution | 10 | Individual |

For all but one of the sections, each team member will receive the same mark that has been allocated to the group. For the section Contribution, after referring to the contribution quiz the marks will be allocated individually. If all team members are deemed to have contributed equally and have made a significant contribution to the team, then all team members will receive a mark of 10. If this is not the case, each individual will be given a mark out of 10 that reflects their level of contribution.

# Description of sections

In the following subsections, an indication is given of what each section in the final report should contain as a minimum.

## Introduction

The problem is introduced with an outline of the steps involved in the analysis.

## Data description

The data is described including identifying the subjects, the variables, which of the variables are predictors, and which are response variables. The number of subjects and variables must also be given. For each variable, it and its levels are explained in context.

## Cleaning of data

Each step of the cleaning performed is described with illustratory code and its output. The reasons and effect of the cleaning is described. For example, any excluded subjects should be noted and the numbers excluded given along with summary statistics both before and after cleaning.

## Variable description

For each variable considered, there should be a section giving the type of variable, summary statistics and a plot to illustrate its distribution. Summary statistics are best given in a table and referred to in the text. A discussion of the distribution of each variable must be given.

## Bivariate analysis

The relationship between each of the predictor variables and the response variable are explored with an appropriate plot. The relationship between each predictor variable and the response variable must also be described in the text.

## Model fitting

The model fitting process is described with the type of algorithm used, and the choice of heuristic discussed. The various models explored are compared.

## Final model

The final model is given and the coefficients interpreted in context.

## Assumption checking

The assumptions of the final model are checked with accompanying tables / figures to support the checking.

## Prediction

A prediction is made for a South Australian, cross-breed dog that is 1 year old. The dog is vaccinated, desexed and unfortunately has to be re-homed because the owner is moving interstate.

## Conclusion

Summarise your analysis and findings.

## Formatting

- All figures and tables should be appropriately captioned and cross-referenced in the text.
- Correct use of grammar and spelling.
- You need a title page with
  - title,
  - authors, and
  - date.

## Code

R code can be included in the text to illustrate how the analysis is done, but the majority of the code can be put into an appendix.