STATS 3001 / STATS 4104 / STATS 7054
Statistical Modelling III
Practical 1 - Linear model recap

Week 1

The following exercises are intended as revision of basic regression calculation and matrix manipulation functions in R.

For the purpose of this exercise, we will use built-in data set `Rubber` that is provided with the `MASS` library. The data set comprises three variables recorded on thirty samples of tyre rubber that were being tested for durability:

- `loss`: The abrasion loss in grams per hour;
- `hard`: The hardness in Shore units; and
- `tens`: The tensile strength in kg per square metre.

STEPS

- Load packages using the command

```
pacman::p_load(tidyverse, gglm)
```

- Load the `rubber` dataset from the `MASS` package

```
data(Rubber, package = "MASS")
```

- Obtain scatter-plots of loss against each of the other predictors.
- Use the `lm()` function to fit the following model.

$$E(\texttt{loss}_i) = \beta_0 + \beta_1 \times \texttt{hard}_i + \beta_2 \times \texttt{tens}_i.$$

- Interpret the output
- What are the assumptions of the linear regression model. Produce appropriate plots to check them.
- Predict the loss for the following points:

hard	tens
50	200
65	190

- Calculate 95% CI and PI for the above points.
- Get the design matrix of the model using the command `model.matrix()`.
- Assign the response variable `loss` to a variable `Y`.

- Using the R commands `%*%`, `solve()`, and `t()`, calculate

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T Y$$

  Is it the same as the answer from `lm()`?
- Calculate the fitted values

$$\hat{\boldsymbol{\eta}} = X\hat{\boldsymbol{\beta}}$$

- Calculate the residual variance, s_e^2 directly from the observed and fitted values. Compare the result to the `residual standard error` produced by `lm()`
- Calculate the estimated variance matrix for $\hat{\boldsymbol{\beta}}$ using

$$(X^T X)^{-1} \times s_e^2$$

  Compare this to the result of the built-in calculation `vcov()`.