# ESMA 6205: Multiple Linerar Regression

Alejandro Ouslan

Spring 2025

## Contents

**1 Introduction** **1**

**2 Model Diagnostic** **1**

**3 Model Selection** **1**
3.1 Mallow's $C_p$ . . . . . 2
3.2 Introductionu to the likelihood funtion . . . . . 2
3.3 Akaike's Information Criterion (AIC) . . . . . 2
3.4 Bayesian Information Criterion (BIC) . . . . . 2

**4 Example: Rat population** **2**

# 1 Introduction

# 2 Model Diagnostic

1. It is possible that all of the predictors are associateded with the repsonse, but it is more often the case that the response is only related to a subset of the predictor.
2. The task of determining which predictors are associated with the reponce, in order to fit a single model involving only those predictors, is referred to as variable selection.
3. All subset regressions tests all posssibel subsets of the set of potebtial independent variables. If there are $K$ potential independent variables (beside the contant), then there are $2^K$ distinct subsets of them to be tested.

# 3 Model Selection

We can then select the best model out of all of the models that we have considered. How do we determine which model is best?

1. Mallow's $C_p$
2. Beyesian Information Criterion (BIC)
3. Akaike's Information Criterion (AIC)
4. Adjusted $R^2$

### 3.1 Mallow's $C_p$

Malllow's $C_p$ value is given by.

$$C_p = \frac{SSE_P}{\hat{\sigma}^2} - (n - 2p)$$

where $SSE_P$ is the sums of square of the error with $p$ predictors and $\hat{\sigma}^2$ is the estimated mean squared error

1. $SSE_P = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
2. $\hat{\sigma}^2 =$ MMSE of the model

### 3.2 Introductionu to the likelihood funtion

snuppose that $y = X\beta + \epsilon$ where $\epsilon \sim N(0, \sigma^2 I_n)$

1. $\epsilon \sim N(0, \sigma^2 I_n) \leftrightarrow y \sim N(X\beta, \sigma^2 I_n)$

$$\begin{aligned} f(y|X\beta, \sigma^2) &= (2\pi)^{-n/2}|\sigma^2 I_n|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right) \\ &= \max f(y|X\beta, \sigma^2) \\ &= \log f(y|X\beta, \sigma^2) \\ &= \max \log f(y|X\beta, \sigma^2) \\ &= \log(\hat{\theta}) = \max \log f(y|X\beta, \sigma^2) \end{aligned}$$

2. We are looking the ver

### 3.3 Akaike's Information Criterion (AIC)

The AIC is given by

$$AIC = -2\log(\hat{\theta}) + 2p \to \log f(y|\hat{\theta}) + 2p$$

### 3.4 Bayesian Information Criterion (BIC)

The BIC is given by

$$BIC = -2\log(\hat{\theta}) + p\log(n)$$

## 4 Example: Rat population