

1. 簡介：

- 說明資料內容，包含哪些變數。
- 問題陳述：想要預測的response是什麼？

2. 初步資料調查：

- 檢查變數間的相關性
 - Scatter plot matrix
 - Correlation matrix
- 利用所有潛在的解釋變數(data set 裡所有的解釋變數)的一階迴歸模型當成起點
 - 檢查此模型的residual plots (residual v.s. predicted values, residual v.s. predictor variables)
 - 線性關係是否ok，是否可能存在曲線關係
 - Constant variance 假設是否合理
 - 檢查 normal plot for predicted values
 - 常態假設是否合理
- 經由這些調查決定是否要加入interaction term，或是對應變變數 Y 作轉換(若做了轉換重新fitting做檢查是否有改善)

3. 將資料分為 training and testing 兩部分：

- training data — 訓練模型
- Testing data — 模型性能評估
- 解釋如何分的，兩個size 各為多少

4. 模型訓練：描述模型的訓練過程，包括任何參數調整或交叉驗證的步驟。

- 模型選擇的方法(model selection/ variable selection)
 - Best Subset algorithm (9.4)
 - 根據results選出3-4個你覺得好的models
 - 必須有圖與表格的結果(參考P. 45, 46 in chap 9 slides)

5. 驗證 (Validation) (9.6)

- 利用testing data 做驗證。
- 驗證 Regression model, Regression coefficients, Predictive ability

6. 模型結果：

- 將最終選擇模型重新 fit 所有資料
- 模型評估：報告模型的表現指標，例如均方誤差 (Mean Squared Error) 或決定係數 (Coefficient of Determination) 。
- 對於coefficients 係數的解釋，包括顯著性測試和效應大小。
- 通過圖表或圖形展示模型的預測結果，例如實際值與預測值之間的比較。
- Residual plots

7. Diagnostic(chap10)(加分項目)：

- added-variable plot, outliers, influential points

8. 結論：

- 總結報告的主要結果和發現。