1. **令 Y, X1, X2 分別代表各區投票率、每人平均所得水準、藍綠政黨上屆得票率差距(取絕對值)， β0, β1, β2 為迴歸係數，請說明：**

**• 線性迴歸模型的數學設定 (請將所得變數 X∗ 1 採中心化操作，即 X∗ 1 = X1 − X1)**



Y = 𝛽0 + 𝛽1X1 + 𝛽2X2

**• 各迴歸係數的意義。**

𝛽0為截距

𝛽n為斜率，表自變數每增加一單位時，因變數Y的改變量

**• 其「線性」的意義。**

在這邊的線性模型代表投票率與每人平均所得水準和藍綠政黨上屆得票率差距的函數關係圖呈現線段關係。線性關係存在2種不同的含義。其一，若某[數學函數](https://zh.wikipedia.org/wiki/%E5%87%BD%E6%95%B0)或數量關係的[函數圖形](https://zh.wikipedia.org/wiki/%E5%87%BD%E6%95%B0%E5%9B%BE%E5%BD%A2)呈現為一條直線或線段，那麼這種關係就是一種線性的關係。其二，在代數學和[數學分析](https://zh.wikipedia.org/wiki/%E6%95%B0%E5%AD%A6%E5%88%86%E6%9E%90)學中，如果一種運算同時滿足特定的「加性」和「齊性」，則稱這種運算是線性的。

1. **我們知道採用最小平方法 (最大概似法亦同) 的迴歸係數解為 βˆ = (X′X)**

**−1X′Y ，請求出其解，並且解釋其結果。**

|  |  |  |
| --- | --- | --- |
| beta0 | beta1 | beta2 |
| 67.488 | -7.215 | -16.836 |

由所得的係數可推論:

每人平均所得水準與藍綠政黨上屆得票率差距的增加都會降低投票率，藍綠政黨上屆得票率差距影響較大。

1. **我們知道投票率的理論預測可由模型推出 (Yˆ = Xβˆ)，請算出各選區的預測**

**投票率值，並檢驗是否有不合理的情況。**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **預測投票率Yˆ** | **觀測投票值Y** | **偏差e** |
| **1** | **68.1137** | 43 | 25.1137 |
| **2** | **75.9545** | 80 | -4.0455 |
| **3** | **60.2008** | 66 | -5.7992 |
| **4** | **47.3092** | 48 | -0.6908 |
| **5** | **65.6605** | 76 | -10.3395 |
| **6** | **64.7946** | 57 | 7.7946 |
| **7** | **72.8279** | 81 | -8.1721 |
| **8** | **58.5412** | 79 | -20.4588 |
| **9** | **69.8454** | 58 | 11.8454 |
| **10** | **63.1110** | 64 | -0.8890 |
| **11** | **63.3996** | 69 | -5.6004 |
| **12** | **63.2555** | 58 | 5.2555 |
| **13** | **62.8707** | 61 | 1.8707 |
| **14** | **57.1702** | 45 | 12.1702 |
| **15** | **79.0091** | 84 | -4.9909 |
| **16** | **61.0185** | 62 | -0.9815 |
| **17** | **64.0972** | 54 | 10.0972 |
| **18** | **57.4590** | 59 | -1.5410 |
| **19** | **67.2959** | 71 | -3.7041 |
| **20** | **68.0656** | 75 | -6.9344 |

1. **倘若我們知道經驗上 (上表 20 個選區實際資料顯示)X∗ 1 , X2 各有其最大最小值，請利用經驗上兩自變量所有出現的可能數值，來討論第二小題所解出的最小平方解是否滿足依變項的值域限制。**

Max(X\*1) = 1.98

Min(X\*1) = -1.26

Max(X2) = 0.56

Min(X1) = 0.01

Y = 𝛽0 + 𝛽1Max(X\*1) + 𝛽2Max(X2) = 43.7737

Y = 𝛽0 + 𝛽1Min(X\*1) + 𝛽2Min(X2) = 79.0091

預測的最大與最小值為: 79.0091, 47.3092, 皆滿足依變項的值域限制。

1. **請用邏輯上兩自變量所有出現的可能數值，假定各區每人平均所得是在 [1**

**萬, 15萬] 之間，再重新評估第四題的問題。**

若平均所得為1萬，則預測投票率為: 94.8832%

若平均所得為15萬，則預測投票率為: -6.1338% (不滿足值域限制，且不合理)

1. **請運用老師附件中所準備的 Matlab 制限最佳化程式，評估滿足依變項界域**

**情況下，再考量自變量的經驗和邏輯值域，其各自算出的迴歸係數解為何？**

beta0:67.555, beta1:-7.286, beta2:-16.988

1. **再一次，請用邏輯上兩自變量所有出現的可能數值，即假定各區每人平均所**

**得是在 [1 萬,15 萬] 之間，再重新評估一次預測投票率值，並與第五題的預測結果相比較。**

beta0:67.204, beta1:-5.195, beta2:-14.322

若平均所得為1萬，則預測投票率為: 87.04809 %

若平均所得為15萬，則預測投票率為: 14.32163 %

結果皆合理。

第五題結果:

若平均所得為1萬，則預測投票率為: 94.8832%

若平均所得為15萬，則預測投票率為: -6.1338% (不滿足值域限制，且不合理)