預測每坪租金高低



系級:統碩一

組員:蔣宛蓉、張庭禎、蔡承軒、

柯百翼、王楷文

目錄

* 主題
* 研究目的
* 資料與變數介紹
* 建立無交互作用項模型
* 建立有交互作用項模型
* 兩模型比較
* 預測Test data
* 結論

一、主題:預測每坪租金高低

二、研究目的:想知道哪些因素會影響每坪租金高低，判斷各個因素如何影響每

坪租金高低

三、資料介紹:資料來自於租屋網的網站，中山區的租屋資料，資料共有5064筆，分成train data有3544筆(70%)、test data有1520筆(30%)，變數有每坪租金、押金月數、提供車位、建築類型、物件類型、額外費用\_電費、額外費用\_水費、額外費用\_瓦斯、額外費用\_網路、額外費用\_第四台、附近有\_學校、附近有\_公園、附近有\_超商、附近的捷運站數、附近的公車站數、有身份限制、有性別限制、可炊、可寵、有產權登記、刊登者類型、提供家具、提供網路、提供第四台、提供天然瓦斯

|  |  |  |  |
| --- | --- | --- | --- |
| 變數名稱 | 變數類型 | 變數名稱 | 變數類型 |
| 每坪租金 | 連續型 | 附近有\_捷運站 | 類別型(有、沒有) |
| 押金月數 | 連續型 | 附近有\_公車站 | 類別型(有、沒有) |
| 提供車位 | 類別型(有、沒有) | 有身份限制 | 類別型(有、沒有) |
| 建築類型 | 類別型(公寓、電梯大樓) | 有性別限制 | 類別型(有、沒有) |
| 物件類型 | 類別型(整層住家、獨立套房、分租套房、雅房) | 可炊 | 類別型(可、不可) |
| 額外費用\_電費 | 類別型(有、沒有) | 可寵 | 類別型(可、不可) |
| 額外費用\_水費 | 類別型(有、沒有) | 有產權登記 | 類別型(有、沒有) |
| 額外費用\_瓦斯 | 類別型(有、沒有) | 刊登者類型 | 類別型(屋主、其他人) |
| 額外費用\_網路 | 類別型(有、沒有) | 提供家具 | 類別型(有、沒有) |
| 額外費用\_第四台 | 類別型(有、沒有) | 提供網路 | 類別型(有、沒有) |
| 附近有\_學校 | 類別型(有、沒有) | 提供第四台 | 類別型(有、沒有) |
| 附近有\_公園 | 類別型(有、沒有) | 提供天然瓦斯 | 類別型(有、沒有) |
| 附近有\_超商 | 類別型(有、沒有) |  |  |

應變數:將每坪租金以1700元為界分成每坪租交高和低(高設為1，低設為0)

自變數:24個變數

(1)1個連續型變數:押金月數

(2)23個類別型變數:提供車位(沒有設為0，有設為1)

建築類型(公寓設為0，電梯大樓設為1)

物件類型(整層住家設為0，獨立套房設為1，分租套房設

為2，雅房設為3)

額外費用\_電費(沒有設為0，有設為1)

額外費用\_水費(沒有設為0，有設為1)

額外費用\_瓦斯(沒有設為0，有設為1)

額外費用\_網路(沒有設為0，有設為1)

額外費用\_第四台(沒有設為0，有設為1)

附近有\_學校(沒有設為0，有設為1)

附近有\_公園(沒有設為0，有設為1)

附近有\_超商(沒有設為0，有設為1)

附近有\_捷運站(沒有設為0，有設為1)

附近有\_公車站(沒有設為0，有設為1)

有身份限制(沒有設為0，有設為1)

有性別限制(沒有設為0，有設為1)

可炊(不可設為0，可以設為1)

可寵(不可設為0，可以設為1)

有產權登記(沒有設為0，有設為1)

刊登者類型(屋主設為0，其他人設為1)

提供家具(沒有設為0，有設為1)

提供網路(沒有設為0，有設為1)

提供第四台(沒有設為0，有設為1)

提供天然瓦斯(沒有設為0，有設為1)

四、無交互作用模型

1.向後逐步回歸建立羅吉斯模型

| **型收斂狀態** |
| --- |
| 已滿足收斂準則 (GCONV=1E-8)。 |

| **模型配適統計值** | | |
| --- | --- | --- |
| **準則** | **僅限截距** | **截距和共變量** |
| **AIC** | 4912.848 | 4419.529 |
| **SC** | 4919.021 | 4518.292 |
| **-2 對數 L** | 4910.848 | 4387.529 |

| **檢定全域虛無假設: BETA=0** | | | |
| --- | --- | --- | --- |
| **檢定** | **卡方** | **自由度** | **Pr > ChiSq** |
| **概度比** | 523.3193 | 15 | <.0001 |
| **計分** | 485.0834 | 15 | <.0001 |
| **Wald** | 426.8328 | 15 | <.0001 |

| **殘差卡方檢定** | | |
| --- | --- | --- |
| **卡方** | **自由度** | **Pr > ChiSq** |
| 18.9378 | 11 | 0.0622 |

|  |  |
| --- | --- |
| Note: | 沒有 (其他) 效果符合從模型予以移除的 0.05 顯著層級。 |

| **向後消去的摘要** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **步驟** | **移除的 效果** | **自由度** | **數目 於** | **Wald 卡方** | **Pr > ChiSq** |
| **1** | **network1** | 1 | 23 | 0.1194 | 0.7297 |
| **2** | **TV1** | 1 | 22 | 0.0563 | 0.8125 |
| **3** | **month** | 1 | 21 | 0.3485 | 0.5550 |
| **4** | **gas** | 1 | 20 | 1.3113 | 0.2522 |
| **5** | **car** | 1 | 19 | 1.3218 | 0.2503 |
| **6** | **TV** | 1 | 18 | 1.3936 | 0.2378 |
| **7** | **school** | 1 | 17 | 2.2749 | 0.1315 |
| **8** | **register** | 1 | 16 | 2.6848 | 0.1013 |
| **9** | **people** | 1 | 15 | 2.8810 | 0.0896 |
| **10** | **park** | 1 | 14 | 3.1812 | 0.0745 |
| **11** | **supermarket** | 1 | 13 | 3.1500 | 0.0759 |

| **效果的第三型分析** | | | |
| --- | --- | --- | --- |
| **效果** | **自由度** | **Wald 卡方** | **Pr > ChiSq** |
| **building** | 1 | 80.1139 | <.0001 |
| **object** | 3 | 39.9651 | <.0001 |
| **electric** | 1 | 7.3505 | 0.0067 |
| **water\_** | 1 | 7.0616 | 0.0079 |
| **network** | 1 | 12.8938 | 0.0003 |
| **station** | 1 | 4.1100 | 0.0426 |
| **bus** | 1 | 12.2551 | 0.0005 |
| **sex** | 1 | 5.4986 | 0.0190 |
| **cook** | 1 | 19.3690 | <.0001 |
| **pet** | 1 | 30.0546 | <.0001 |
| **put** | 1 | 57.4521 | <.0001 |
| **Furniture** | 1 | 75.6214 | <.0001 |
| **gas1** | 1 | 12.0189 | 0.0005 |

| **最大概度估計值的分析** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **參數** |  | **自由度** | **估計值** | **標準 誤差** | **Wald 卡方** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -2.6454 | 0.5422 | 23.8006 | <.0001 |
| **building** |  | 1 | 0.9808 | 0.1096 | 80.1139 | <.0001 |
| **object** | **1** | 1 | 0.5470 | 0.0913 | 35.9270 | <.0001 |
| **object** | **2** | 1 | 0.6753 | 0.1568 | 18.5501 | <.0001 |
| **object** | **3** | 1 | 0.3037 | 0.2428 | 1.5641 | 0.2111 |
| **electric** |  | 1 | -0.8269 | 0.3050 | 7.3505 | 0.0067 |
| **water\_** |  | 1 | 0.3710 | 0.1396 | 7.0616 | 0.0079 |
| **network** |  | 1 | -0.4708 | 0.1311 | 12.8938 | 0.0003 |
| **station** |  | 1 | 0.8252 | 0.4070 | 4.1100 | 0.0426 |
| **bus** |  | 1 | -0.2669 | 0.0762 | 12.2551 | 0.0005 |
| **sex** |  | 1 | 0.4064 | 0.1733 | 5.4986 | 0.0190 |
| **cook** |  | 1 | 0.5271 | 0.1198 | 19.3690 | <.0001 |
| **pet** |  | 1 | -0.5039 | 0.0919 | 30.0546 | <.0001 |
| **put** |  | 1 | 0.7370 | 0.0972 | 57.4521 | <.0001 |
| **Furniture** |  | 1 | 1.1199 | 0.1288 | 75.6214 | <.0001 |
| **gas1** |  | 1 | -0.2861 | 0.0825 | 12.0189 | 0.0005 |

| **勝算比估計值** | | | |
| --- | --- | --- | --- |
| **效果** | **點估計值** | **95% Wald 信賴界限** | |
| **building** | 2.667 | 2.151 | 3.306 |
| **object 1 與 0 之間的關係** | 1.728 | 1.445 | 2.066 |
| **object 2 與 0 之間的關係** | 1.965 | 1.445 | 2.671 |
| **object 3 與 0 之間的關係** | 1.355 | 0.842 | 2.181 |
| **electric** | 0.437 | 0.241 | 0.795 |
| **water\_** | 1.449 | 1.102 | 1.905 |
| **network** | 0.624 | 0.483 | 0.807 |
| **station** | 2.282 | 1.028 | 5.068 |
| **bus** | 0.766 | 0.660 | 0.889 |
| **sex** | 1.501 | 1.069 | 2.109 |
| **cook** | 1.694 | 1.340 | 2.142 |
| **pet** | 0.604 | 0.505 | 0.723 |
| **put** | 2.090 | 1.727 | 2.528 |
| **Furniture** | 3.065 | 2.381 | 3.945 |
| **gas1** | 0.751 | 0.639 | 0.883 |

| **預測機率和觀測回應的關聯性** | | | |
| --- | --- | --- | --- |
| **和諧百分比** | 70.3 | **Somers' D** | 0.425 |
| **不和諧百分比** | 27.8 | **Gamma** | 0.434 |
| **繫結百分比** | 1.9 | **Tau-a** | 0.213 |
| **配對** | 3137510 | **c** | 0.713 |



結論: (1)建築類型為電梯大樓每坪租金大於1700元的機率較建築類型為公寓

大

(2)物件類型為獨立套房每坪租金大於1700元的機率較物件類型為整層住

家大

(3)物件類型為分租套房每坪租金大於1700元的機率較物件類型為整層住家大

(4)物件類型為雅房每坪租金大於1700元的機率較物件類型為整層住家大

(5)電費為額外費用每坪租金大於1700元的機率較電費不為額外費用小

(6)水費為額外費用每坪租金大於1700元的機率較水費不為額外費用大

(7)網路為額外費用每坪租金大於1700元的機率較水費不為額外費用小

(8)附近有超商每坪租金大於1700元的機率較附近沒有超商小

(9)附近有捷運站每坪租金大於1700元的機率較附近沒有捷運站大

(10)附近有公車站每坪租金大於1700元的機率較附近沒有公車站小

(11)有身分限制每坪租金大於1700元的機率較沒有身分限制大

(12)有性別限制每坪租金大於1700元的機率較沒有性別限制大

(13)可烹每坪租金大於1700元的機率較不可烹大

(14)可寵每坪租金大於1700元的機率較不可寵小

(15)刊登者不為屋主每坪租金大於1700元的機率較刊登者為屋主大

(16)有提供家具每坪租金大於1700元的機率較沒有提供家具小

(17)有提供瓦斯每坪租金大於1700元的機率較沒有提供瓦斯小

五、有交互作用項

| **模型收斂狀態** |
| --- |
| 已滿足收斂準則 (GCONV=1E-8)。 |

| **模型配適統計值** | | |
| --- | --- | --- |
| **準則** | **僅限截距** | **截距和共變量** |
| **AIC** | 4912.848 | 4332.540 |
| **SC** | 4919.021 | 4493.031 |
| **-2 對數 L** | 4910.848 | 4280.540 |

| **檢定全域虛無假設: BETA=0** | | | |
| --- | --- | --- | --- |
| **檢定** | **卡方** | **自由度** | **Pr > ChiSq** |
| **概度比** | 630.3084 | 25 | <.0001 |
| **計分** | 545.3893 | 25 | <.0001 |
| **Wald** | 438.1222 | 25 | <.0001 |

| **殘差卡方檢定** | | |
| --- | --- | --- |
| **卡方** | **自由度** | **Pr > ChiSq** |
| 46.3979 | 24 | 0.0040 |

|  |  |
| --- | --- |
| Note: | 沒有 (其他) 效果符合從模型予以移除的 0.05 顯著層級。 |

| **向後消去的摘要** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **步驟** | **移除的 效果** | **自由度** | **數目 於** | **Wald 卡方** | **Pr > ChiSq** |
| **1** | **sex\*object** | 2 | 24 | 1.1180 | 0.5718 |
| **2** | **bus\*object** | 3 | 23 | 2.1884 | 0.5342 |
| **3** | **electric\*object** | 3 | 22 | 3.6108 | 0.3067 |
| **4** | **station\*object** | 2 | 21 | 2.6136 | 0.2707 |
| **5** | **network\*object** | 3 | 20 | 6.9380 | 0.0739 |
| **6** | **sex** | 1 | 19 | 2.9274 | 0.0871 |
| **7** | **water\_\*object** | 3 | 18 | 6.8187 | 0.0779 |
| **8** | **pet\*object** | 3 | 17 | 6.6260 | 0.0848 |
| **9** | **station** | 1 | 16 | 3.5161 | 0.0608 |
| **10** | **cook\*object** | 3 | 15 | 7.5348 | 0.0567 |

| **聯合檢定** | | | |
| --- | --- | --- | --- |
| **效果** | **自由度** | **Wald 卡方** | **Pr > ChiSq** |
| **building** | 1 | 52.2293 | <.0001 |
| **object** | 3 | 76.0066 | <.0001 |
| **building\*object** | 3 | 26.9108 | <.0001 |
| **electric** | 1 | 5.7898 | 0.0161 |
| **water\_** | 1 | 6.0120 | 0.0142 |
| **network** | 1 | 13.7277 | 0.0002 |
| **bus** | 1 | 10.1289 | 0.0015 |
| **cook** | 1 | 21.6841 | <.0001 |
| **pet** | 1 | 22.5697 | <.0001 |
| **put** | 1 | 29.9987 | <.0001 |
| **put\*object** | 3 | 12.8623 | 0.0049 |
| **Furniture** | 1 | 95.9158 | <.0001 |
| **Furniture\*object** | 3 | 40.1411 | <.0001 |
| **gas1** | 1 | 21.5695 | <.0001 |
| **gas1\*object** | 3 | 14.3727 | 0.0024 |

|  |  |
| --- | --- |
| Note: | 在完整排名參數化下， 類型 3 效果檢定會取代為聯結檢定。 進行效果的聯結檢定時， 會將所有與該效果相關聯的參數都設為零。 這樣的聯結檢定可能不會等同於 GLM 參數化下的類型 3 效果檢定。 |

| **最大概度估計值的分析** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **參數** |  | **自由度** | **估計值** | **標準 誤差** | **Wald 卡方** | **Pr > ChiSq** |
| **Intercept** |  | 1 | -3.9913 | 0.4960 | 64.7545 | <.0001 |
| **building** |  | 1 | 2.2359 | 0.3094 | 52.2293 | <.0001 |
| **object** | **1** | 1 | 4.0732 | 0.4816 | 71.5305 | <.0001 |
| **object** | **2** | 1 | 5.3976 | 1.3125 | 16.9115 | <.0001 |
| **object** | **3** | 1 | 16.3737 | 286.5 | 0.0033 | 0.9544 |
| **building\*object** | **1** | 1 | -1.5022 | 0.3394 | 19.5875 | <.0001 |
| **building\*object** | **2** | 1 | -1.9488 | 0.3923 | 24.6772 | <.0001 |
| **building\*object** | **3** | 1 | -1.7353 | 0.5158 | 11.3193 | 0.0008 |
| **electric** |  | 1 | -0.7209 | 0.2996 | 5.7898 | 0.0161 |
| **water\_** |  | 1 | 0.3455 | 0.1409 | 6.0120 | 0.0142 |
| **network** |  | 1 | -0.4994 | 0.1348 | 13.7277 | 0.0002 |
| **bus** |  | 1 | -0.2471 | 0.0776 | 10.1289 | 0.0015 |
| **cook** |  | 1 | 0.5545 | 0.1191 | 21.6841 | <.0001 |
| **pet** |  | 1 | -0.4488 | 0.0945 | 22.5697 | <.0001 |
| **put** |  | 1 | 1.3569 | 0.2477 | 29.9987 | <.0001 |
| **put\*object** | **1** | 1 | -0.7249 | 0.2745 | 6.9735 | 0.0083 |
| **put\*object** | **2** | 1 | -1.0068 | 0.3313 | 9.2373 | 0.0024 |
| **put\*object** | **3** | 1 | -1.6001 | 0.5910 | 7.3300 | 0.0068 |
| **Furniture** |  | 1 | 1.6952 | 0.1731 | 95.9158 | <.0001 |
| **Furniture\*object** | **1** | 1 | -1.8951 | 0.3128 | 36.7059 | <.0001 |
| **Furniture\*object** | **2** | 1 | -2.9006 | 1.2539 | 5.3513 | 0.0207 |
| **Furniture\*object** | **3** | 1 | -14.6026 | 286.5 | 0.0026 | 0.9593 |
| **gas1** |  | 1 | -0.5725 | 0.1233 | 21.5695 | <.0001 |
| **gas1\*object** | **1** | 1 | 0.5761 | 0.1845 | 9.7500 | 0.0018 |
| **gas1\*object** | **2** | 1 | 0.5094 | 0.2802 | 3.3049 | 0.0691 |
| **gas1\*object** | **3** | 1 | 1.0587 | 0.4062 | 6.7931 | 0.0092 |

| **勝算比估計值** | | | |
| --- | --- | --- | --- |
| **效果** | **點估計值** | **95% Wald 信賴界限** | |
| **electric** | 0.486 | 0.270 | 0.875 |
| **water\_** | 1.413 | 1.072 | 1.862 |
| **network** | 0.607 | 0.466 | 0.790 |
| **bus** | 0.781 | 0.671 | 0.909 |
| **cook** | 1.741 | 1.379 | 2.199 |
| **pet** | 0.638 | 0.530 | 0.768 |

| **預測機率和觀測回應的關聯性** | | | |
| --- | --- | --- | --- |
| **和諧百分比** | 72.0 | **Somers' D** | 0.460 |
| **不和諧百分比** | 26.0 | **Gamma** | 0.469 |
| **繫結百分比** | 2.0 | **Tau-a** | 0.230 |
| **配對** | 3137510 | **c** | 0.730 |



結論:(1)物件類型為獨立套房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(2)物件類型為分租套房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(3)物件類型為雅房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(4)物件類型為獨立套房，則刊登者不為屋主每坪租金高的機率上升幅度

較物件類型為整層住家小

(5)物件類型為分租套房，則刊登者不為屋主每坪租金高的機率上升幅度

較物件類型為整層住家小

(6)物件類型為雅房，則刊登者不為屋主每坪租金高的機率較刊登者為屋

主小

(7)物件類型為獨立套房，則提供家具每坪租金高的機率較沒有提供家具

小

(8)物件類型為分租套房，則提供家具每坪租金高的機率較沒有提供家具

小

(9)物件類型為雅房，則提供家具每坪租金高的機率較沒有提供家具小

(10)物件類型為獨立套房，則提供瓦斯每坪租金高的機率較沒有提供瓦斯

大

(11)物件類型為分租套房，則提供瓦斯每坪租金高的機率下降幅度較物件

類型為整層住家小

(12)物件類型為雅房，則提供瓦斯每坪租金高的機率較沒有提供瓦斯大

六、兩模型比較





結論:有交互作用項的模型較無交互作用項的模型好，因此最後選擇的模型為有加交互作用項的模型

七、Test data

1.

|  |  |  |
| --- | --- | --- |
|  | **實際為低租金(y=0)** | **實際為高租金(y=1)** |
| 預測為低租金(y=0) | 420 | 169 |
| 預測為高租金(y=1) | 355 | 575 |

* 敏感性=575/(169+575)=0.7728495
* 特異性=420/(420+355)=0.5419355
* 準確度=(420+575)/(420+355+169+575)=0.6550362

預測

實際 0 1

0 420 169

1 355 575

> mean(predict==test$rent)

[1] 0.6550362

> #計算核準卡正確率

> cm[4] / sum(cm[, 2])

[1] 0.7728495

> cm[1] / sum(cm[, 1])

[1] 0.5419355

>

> #整體準確率(對角線元素總和/所有觀察值總和)

> accuracy <- sum(diag(cm)) / sum(cm)

> accuracy

[1] 0.6550362

2.ROC curve



八、結論

1.無交互作用項模型解釋

(1)建築類型為電梯大樓每坪租金大於1700元的機率較建築類型為公寓

大

(2)物件類型為獨立套房每坪租金大於1700元的機率較物件類型為整層住

家大

(3)物件類型為分租套房每坪租金大於1700元的機率較物件類型為整層住家大

(4)物件類型為雅房每坪租金大於1700元的機率較物件類型為整層住家大

(5)電費為額外費用每坪租金大於1700元的機率較電費不為額外費用小

(6)水費為額外費用每坪租金大於1700元的機率較水費不為額外費用大

(7)網路為額外費用每坪租金大於1700元的機率較水費不為額外費用小

(8)附近有超商每坪租金大於1700元的機率較附近沒有超商小

(9)附近有捷運站每坪租金大於1700元的機率較附近沒有捷運站大

(10)附近有公車站每坪租金大於1700元的機率較附近沒有公車站小

(11)有身分限制每坪租金大於1700元的機率較沒有身分限制大

(12)有性別限制每坪租金大於1700元的機率較沒有性別限制大

(13)可烹每坪租金大於1700元的機率較不可烹大

(14)可寵每坪租金大於1700元的機率較不可寵小

(15)刊登者不為屋主每坪租金大於1700元的機率較刊登者為屋主大

(16)有提供家具每坪租金大於1700元的機率較沒有提供家具小

(17)有提供瓦斯每坪租金大於1700元的機率較沒有提供瓦斯小

2.有交互作用項模型解釋

(1)物件類型為獨立套房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(2)物件類型為分租套房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(3)物件類型為雅房，則建築類型為電梯大樓每坪租金高的機率上升

幅度較物件類型為整層住家小

(4)物件類型為獨立套房，則刊登者不為屋主每坪租金高的機率上升幅度

較物件類型為整層住家小

(5)物件類型為分租套房，則刊登者不為屋主每坪租金高的機率上升幅度

較物件類型為整層住家小

(6)物件類型為雅房，則刊登者不為屋主每坪租金高的機率較刊登者為屋

主小

(7)物件類型為獨立套房，則提供家具每坪租金高的機率較沒有提供家具

小

(8)物件類型為分租套房，則提供家具每坪租金高的機率較沒有提供家具

小

(9)物件類型為雅房，則提供家具每坪租金高的機率較沒有提供家具小

(10)物件類型為獨立套房，則提供瓦斯每坪租金高的機率較沒有提供瓦斯

大

(11)物件類型為分租套房，則提供瓦斯每坪租金高的機率下降幅度較物件

類型為整層住家小

(12)物件類型為雅房，則提供瓦斯每坪租金高的機率較沒有提供瓦斯大