

電影票房



國立台北大學統計學系

組長：統計三 陳庭安

組員：統計三 藍睿豪

統計三 詹芸嫻

統計三 周祖望

統計三 橋俊森

一、研究主題：電影票房

二、研究目的：

電影，相對於小說，是用一種新的語言詮釋一個故事的轉折，更透過場景的轉換、拍攝角度、聲光特效以及角色細膩的神情，引領觀眾走進故事的脈絡。電影中許多吸引人的元素，已經讓電影成為一個產業。電影產業數十年來的蓬勃發展，近幾年來所帶來的商機每年已超越 2000 億台幣。

產業的發展促使電影市場激烈的競爭，票房，是最直觀數據。欲在市場中獨占鰲頭取得票房佳績，必須針對影響票房的潛在因素去做適當的決策。故本研究主要在於探討影響票房高低的背後潛在原因為何。通路?發行商規模?抑或是電影曝光度。

三、參考資料：

1. 2015 電影列表

<https://zh.wikipedia.org/wiki/2015%E5%B9%B4%E9%9B%BB%E5%BD%B1%E5%88%97%E8%A1%A8%E7.BE.8E.E5.9C.8B.E9.9B.BB.E5.BD.B1.E5.88.97.E8.A1.A8>

2. 電影票房

<http://www.boxofficemojo.com/>

四、變數說明：

Obs 1-40 2015 的 40 部影片

(以分層抽樣方式，2015 年每一季個別隨機選取 10 部電影)

Movie 片名(中+英)

Box office 全球票房(萬美元)

First week box office 首周全球票房(萬美元)

Budget 預算(萬美元)

Theaters 在美國上映的電影院家數/通路

Publisher 發行商

Scale of publisher 發行商規模

五、議題與分析方法：

1. 2015 年 30 部電影票房與成本高低。(使用變數: movie, boxoffice, costs)
 - 利用群集分析，將電影分成低成本、高票房與高成本、低票房等各群。看 2015 年哪幾部電影為獲利高的電影，又哪幾部電影為低收益甚至是虧損的電影。
2. 探究製片成本、通路(電影院家數)、上映首周總票房、製片商規模是否會影響票房高低。
 - 繼 1.之後，(使用變數: movie, box office)利用群集分析分出高票房與低票房兩群，並將兩群結果命名為變數 box office group。
 - (使用變數: box office group、costs、theaters、opening gross、scale of studio)

邏輯斯迴歸

預期: 特效或卡司等製片成本增加，但可能吸引觀看人數=>增加票房。

上映電影院家數/通路增加，增加電影曝光度=>票房提高。

首周票房高，可能會增加曝光度，並吸引更多未觀看人數買票。

製片商的規模大，可能製作出商業大片，與大明星、名導合作=>高票房。

六、正式分析過程：

STEP1：

使用票房以中心點層次群集法+非層次法將資料觀察值分為兩群，群集 1 有 33 個觀測值，群集 2 有 7 個觀察值，因為是以票房做分群依據，再定義群集 1 為低票房群，群集 2 為高票房群。觀察報表中 R 平方=0.73845，解釋原本資料的部分頗多

STEP2：

分析方法：邏輯斯迴歸—逐步分析

分析目的：在群集分析將 40 部電影分為—高票房(7 部)、低票房(33 部) 兩群後，

欲探究影響電影票房高低的因素為何。藉由以下變數分析並預測票房的高低。

採用變數: cluster=票房的分群(1=高票房, 0=低票房); first_week_boxoffice=首周票房(百萬美元); budget=電影預算; theaters=上映電影院家數

採用以上變數的原因與預期結果: 上映首周票房即衝高, 表示許多民眾對該部電影極有興趣, 另外, 首周票房佳的消息可能將吸引未看過該部電影的民眾。再者, 不同規模的電影創造出的票房不同, 大規模、耗資龐大的電影效果可能更佳, 更吸引民眾。最後還考慮了上映的電影院數目, 電影播放的電影院越多、播放次數多、票房較佳。

邏輯斯逐步迴歸分析過程與結果:

boxoffice 分完後的分群結果為模型反映變數, 首先同時放入首周票房、電影預算、電影院數, 逐步分析選入與移除的門檻為 p 值=0.15。

第一步選入卡方值最大且 p 值小於 0.15 的變數: 首周票房, 選入模型後 p 值仍小於 0.15, 故保留在模型中。剩餘電影預算、電影院數中, 將卡方值較大且 p 值小於 0.15 的變數電影預算選入模型, 模型 p 值仍小於 0.15, 故不移除電影預算。最後尚未抉擇的變數剩下 p 值超過 0.15 的電影院數, 因此不將電影院數選入模型。逐步選擇程序停止, 模型包含首周票房與電影預算。

此時觀看預測模型, 發現在 Wald-test 檢定下, p 值均超過顯著水準, 故兩變數同時放入模型時, 皆無證據顯示電影預算和首周票房會顯著影響整體票房。但兩者單獨放入模型, 皆顯著影響整體票房。問題可能出於 1. 模型配適不足, 可能資料無法以此模型解釋。2. 模型是適合的, 但樣本數目太少, 誤差較大, 不容易正確判斷結果。3. 變數間有交互作用的影響尚未考慮。

STEP3:

由上述所說, 變數不顯著可能是因為模型配適不足或者首周票房與預算有交互作用, 使用加入交互作用的羅吉斯迴歸(logit 兩個變數)做分析, 由 SAS 報表得關聯性的各項指標顯示 $p\text{-hat}$ 與 y 關聯性高, 表示模型預測效度高, 排除是模型配

適不足造成的不顯著，接著我們看交互作用。

最大概似估計值的分析表中，變數交互作用檢定的 $p\text{-value}=0.4158>0.05$ ，拒絕虛無假設，表示沒有足夠的證據顯示首周票房與預算的交互作用顯著影響票房。

結論：變數的不顯著排除是因為模型配適不足也排除是交互作用造成的，接下來可能要考慮共線性的問題？

STEP4：

前面已經看過兩自變數的羅吉斯迴歸結果並不顯著，即羅吉斯迴歸模式不適用；在本節我們將採用單一變數的羅吉斯迴歸模式與雙變數線性複迴歸模式，以測試該兩模式下的配適度議題。

以首周票房作為羅吉斯迴歸的自變數

一樣，在此將總票房以中心點層次群集法 + 非層次法的方式將觀測資料分為兩群集，並定義為“高票房群(群集 1)”7 個觀測值、“低票房群(群集 2)”33 個觀測值。有了兩群之後，便可進行高低總票房機率(勝率)羅吉斯迴歸式的建構，而機率依據為高票房群。

從報表(p16)中可以得知最大概似比檢定、計分檢定、Wald 卡方檢定在顯著水準為 0.05 的設定下皆棄卻斜率=0 的虛無假設，即表示包含首周票房與截距的羅吉斯迴歸式是合適的；再者，報表(p17)，最大概似估計值的 Wald 卡方檢定也顯示出截距、首周票房的係數(斜率)為 0 的虛無假設有足夠的證據可拒絕，因此可建立{}的羅吉斯迴歸式。從報表(p17)，預測機率和觀測值的關聯性在 231 組配對中，和諧百分比高達 98.7%、不和諧僅 1.3%而繫結則為 0，這可說是個十分優異的結果。

雙變數線性複迴歸模式

在前面已知，以預算與首周票房兩變數建構出的羅吉斯迴歸並不適用，在此我們試著建立該兩變數對總票房的線性複迴歸模式，並簡單討論其結果。

報表(p13)，模型變異數檢定(H_0 : 斜率=0)結果顯著；報表(p14)， H_0 : 參數估計

值=0 檢定結果顯著；Radj square=0.9518，即模型配適度達 95.18%；從上述三結果得出總票房的雙變數線性複迴歸模式是合宜且不差的。

從圖(p15)，殘差對預測值的分布圖中可以看出，誤差變異數為常數的前提假設明顯有待商榷；殘差的常態機率百分位數(QQ)圖中，並無明顯違反誤差為常態分配前提假設的異狀；殘差直方圖中，殘差=0(平均數、中位數)之出現次數明顯較兩端多，因此誤差為平均值為 0 的常態分配是成立的。

七、額外的研究方向：

分析目的：

對電影發行者而言，除了票房，預算也是其關心的重大因素。本分析將預算與票房做分群，選出其最適當能分幾群來做額外的研究方向參考

分析過程：

1. 將資料分兩群時發現，兩個的資料的標準差差異極大。兩個資料中並沒有誰比較重要，因此將資料進行標準化。
2. 使用中心點層次群集法+非層次法

由圖表可得知，在集群數目 3 個到集群數目 2 個之間。RMSSTD、半偏 R 平方、R 平方皆有相當大幅度的差別。而且分三個群集時的各項統計量令人滿意

⇒ 分為三群是比較適當的選擇

3. 將三群定義為

- 1.低票房、低預算群
 - 2.中票房、中預算群
 - 3.高票房、高預算群
4. 可看出對兩個變數的R平方為0.816948(解釋資料的8成以上),群集分的不錯。

附錄一報表

STEP1

| 初始種子 | | | |
|------|-------------|--|--|
| 群集 | VAR5 | | |
| 1 | 23538.1515 | | |
| 2 | 136539.2857 | | |

| | | | |
|--------------|--|----------|--|
| 初始種子間的最小距離 = | | 113001.1 | |
|--------------|--|----------|--|

| 反覆運算歷史 | | | |
|--------|---------|-----------|---|
| 反覆運算 | 準則 | 集群種子的相對變更 | |
| | | 1 | 2 |
| 1 | 25552.9 | 0 | 0 |

| | | | |
|----------|--|--|--|
| 已滿足收斂準則。 | | | |
|----------|--|--|--|

以最終種子為基礎的準則 = 25552.9

| 集群摘要 | | | | | | |
|------|----|---------|----------------------|-----------|-------|--------------|
| 群集 | 次數 | RMS 標準差 | 從種子到 觀測值的 最大距離 | 半徑 已超過 | 最近的群集 | 群集重心 間的距離 |
| 1 | 33 | 21398.5 | 44694.8 | | 2 | 113001 |
| 2 | 7 | 43713.6 | 70282.7 | | 1 | 113001 |

| 變動的統計值 | | | | |
|----------|-------|-------|----------|-------------|
| 變動 | 總 STD | STD 內 | R 平方 | RSQ/(1-RSQ) |
| VAR5 | 50602 | 26217 | 0.738455 | 2.823439 |
| OVER-ALL | 50602 | 26217 | 0.738455 | 2.823439 |

虛擬 F 統計值 = 107.29

近似預期整體 R 平方 = 0.76363

立方群集準則 = -0.625

| 群集平均值 | |
|-------|-------------|
| 群集 | VAR5 |
| 1 | 23538.1515 |
| 2 | 136539.2857 |

| 群集標準差 | |
|-------|-------------|
| 群集 | VAR5 |
| 1 | 21398.52113 |
| 2 | 43713.57973 |

| 群集重心間的距離 | | |
|----------|-------------|-------------|
| 最接近的群集 | 1 | 2 |
| 1 | . | 113001.1342 |
| 2 | 113001.1342 | . |

| 群集=1 | | |
|------|-----|---------|
| Obs | obs | CLUSTER |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 3 | 1 |
| 4 | 4 | 1 |
| 5 | 5 | 1 |
| 6 | 6 | 1 |
| 7 | 7 | 1 |
| 8 | 8 | 1 |
| 9 | 9 | 1 |
| 10 | 10 | 1 |

| | | |
|----|----|---|
| 11 | 11 | 1 |
| 12 | 14 | 1 |
| 13 | 17 | 1 |
| 14 | 18 | 1 |
| 15 | 19 | 1 |
| 16 | 20 | 1 |
| 17 | 21 | 1 |
| 18 | 23 | 1 |
| 19 | 24 | 1 |
| 20 | 25 | 1 |
| 21 | 26 | 1 |

| | | |
|----|----|---|
| 22 | 27 | 1 |
| 23 | 28 | 1 |
| 24 | 29 | 1 |
| 25 | 30 | 1 |
| 26 | 31 | 1 |
| 27 | 32 | 1 |
| 28 | 33 | 1 |
| 29 | 34 | 1 |
| 30 | 35 | 1 |
| 31 | 36 | 1 |
| 32 | 38 | 1 |
| 33 | 40 | 1 |

| 群集=2 | | |
|------|-----|---------|
| Obs | obs | CLUSTER |
| 34 | 12 | 2 |
| 35 | 13 | 2 |
| 36 | 15 | 2 |
| 37 | 16 | 2 |
| 38 | 22 | 2 |
| 39 | 37 | 2 |
| 40 | 39 | 2 |

STEP2

LOGISTIC 程序

| 模型資訊 | | |
|--------|----------------|----|
| 資料集 | WORK.CLUSTER_H | |
| 回應變數 | CLUSTER | 群集 |
| 回應層級數目 | 2 | |
| 模型 | 二元對數優劣比 | |
| 最佳化技術 | Fisher 計分 | |

讀取的觀測值數目 40

使用的觀測值數目 40

回應概況

| 已排序的 值 | CLUSTER | 總次數 |
|-----------|---------|-----|
| 1 | 1 | 7 |
| 2 | 0 | 33 |

建立模型的機率是 CLUSTER=1。

步驟 0。輸入的截距:

模型收斂狀態

已滿足收斂準則 (GCONV=1E-8)。

-2 對數 L = 37.098

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 | -1.5506 | 0.4161 | 13.8851 | 0.0002 |

殘差卡方檢定

| 卡方 | 自由度 | Pr > ChiSq |
|---------|-----|------------|
| 28.1317 | 3 | <.0001 |

適合輸入的效果分析

| 效果 | 自由度 | 評分 卡方 | Pr > ChiSq |
|----------------------|-----|----------|------------|
| first_week_boxoffice | 1 | 26.1609 | <.0001 |
| budget1 | 1 | 21.6566 | <.0001 |
| theaters | 1 | 8.9666 | 0.0027 |

步驟 1。已輸入效果 first_week_boxoffice:

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 | 39.098 | 14.103 |
| SC | 40.787 | 17.481 |
| -2 對數 L | 37.098 | 10.103 |

| 檢定全域虛無假設: BETA=0 | | | |
|------------------|---------|-----|------------|
| 檢定 | 卡方 | 自由度 | Pr > ChiSq |
| 概度比 | 26.9952 | 1 | <.0001 |
| 計分 | 26.1609 | 1 | <.0001 |
| Wald | 5.1398 | 1 | 0.0234 |

STEP3

| 預測機率和觀測回應的關聯性 | | | |
|---------------|------|-----------|-------|
| 和諧百分比 | 98.7 | Somers' D | 0.974 |
| 不和諧百分比 | 1.3 | Gamma | 0.974 |
| 繫結百分比 | 0.0 | Tau-a | 0.288 |
| 配對 | 231 | c | 0.987 |

| 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 | -37.2824 | 43.4557 | 0.7361 | 0.3909 |
| first_week_boxoffice | 1 | 0.3305 | 0.3779 | 0.7649 | 0.3818 |
| budget1 | 1 | 0.1624 | 0.1957 | 0.6884 | 0.4067 |
| first_week_b*budget1 | 1 | -0.00132 | 0.00162 | 0.6620 | 0.4158 |

STEP4

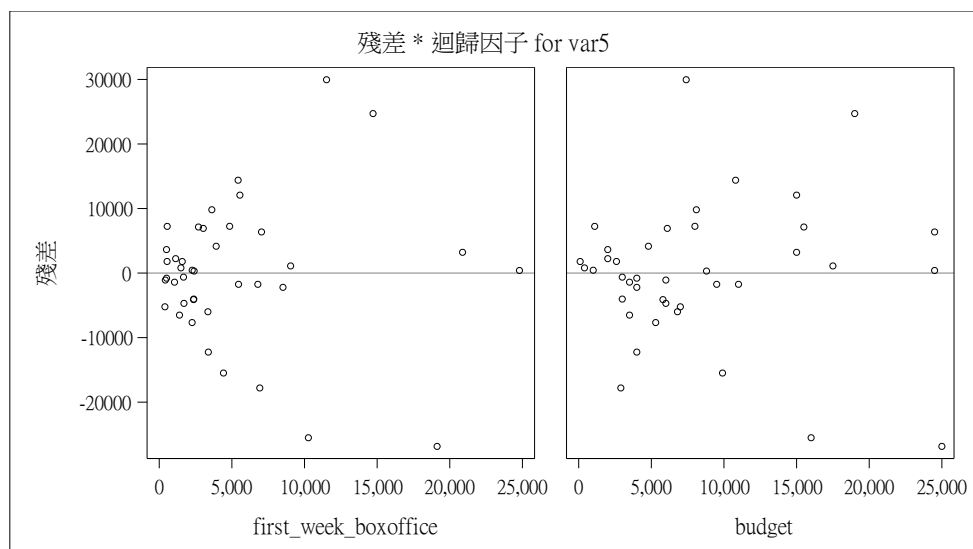
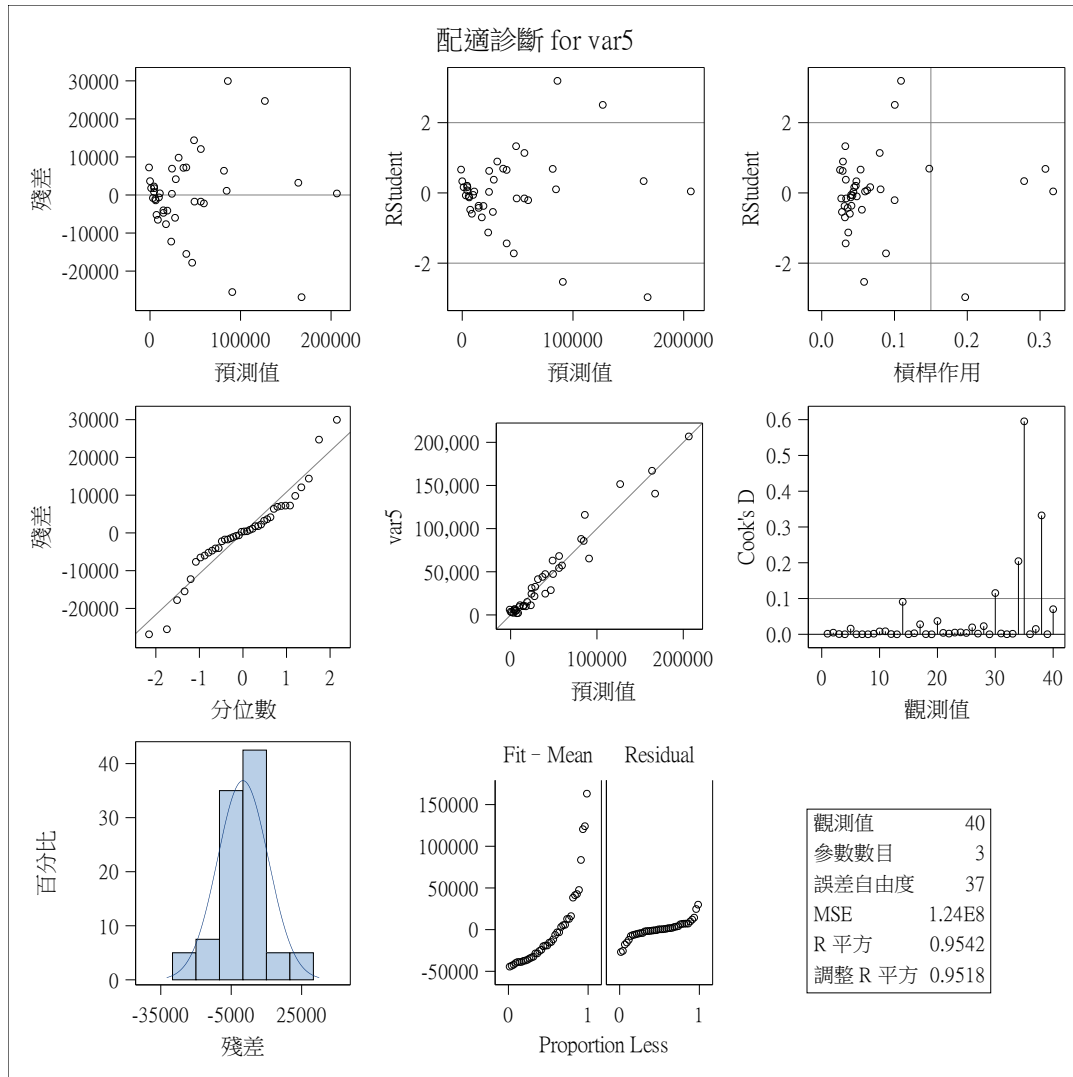
| | |
|----------|----|
| 讀取的觀測值數目 | 40 |
| 使用的觀測值數目 | 40 |

| 變異數分析 | | | | | |
|--------|-----|-------------|-------------|--------|--------|
| 來源 | 自由度 | 平方和 | 平均值 平方 | F 值 | Pr > F |
| 模型 | 2 | 95290836684 | 47645418342 | 385.79 | <.0001 |
| 誤差 | 37 | 4569575545 | 123502042 | | |
| 已校正的總計 | 39 | 99860412229 | | | |

| | | | |
|-------|----------|---------|--------|
| 根 MSE | 11113 | R 平方 | 0.9542 |
| 應變平均值 | 43313 | 調整 R 平方 | 0.9518 |
| 變異係數 | 25.65756 | | |

| 參數估計值 | | | | | | | | |
|-----------------------------|-----------------------------|-----|-------------|------------|-------|---------|---------|-----------|
| 變數 | 標籤 | 自由度 | 參數 估計值 | 標準 誤差 | t 值 | Pr > t | 允差 | 變異數 膨脹 |
| Intercept | Intercept | 1 | -6669.24108 | 2786.27157 | -2.39 | 0.0219 | . | 0 |
| <u>first_week_boxoffice</u> | <u>first_week_boxoffice</u> | 1 | 7.02434 | 0.47064 | 14.93 | <.0001 | 0.42506 | 2.35259 |
| <u>budget</u> | <u>budget</u> | 1 | 1.58839 | 0.40145 | 3.96 | 0.0003 | 0.42506 | 2.35259 |

| 共線性診斷 (截距調整) | | | | |
|--------------|---------|----------|-----------------------------|---------------|
| 數目 | 特徵值 | 條件 索引 | 變異的比例 | |
| | | | <u>first_week_boxoffice</u> | <u>budget</u> |
| 1 | 1.75825 | 1.00000 | 0.12088 | 0.12088 |
| 2 | 0.24175 | 2.69682 | 0.87912 | 0.87912 |



| 模型資訊 | | |
|--------|----------------|----|
| 資料集 | WORK.CLUSTER_H | |
| 回應變數 | CLUSTER | 群集 |
| 回應層級數目 | 2 | |
| 模型 | 二元對數優劣比 | |
| 最佳化技術 | Fisher 計分 | |

| | |
|----------|----|
| 讀取的觀測值數目 | 40 |
| 使用的觀測值數目 | 40 |

| 回應概況 | | |
|-----------|---------|-----|
| 已排序的 值 | CLUSTER | 總次數 |
| 1 | 1 | 7 |
| 2 | 0 | 33 |

建立模型的機率是 $CLUSTER=1$ 。

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

| 偏差及 Pearson 配適度統計值 | | | | |
|--------------------|---------|-----|--------|------------|
| 準則 | 值 | 自由度 | 值/自由度 | Pr > ChiSq |
| 偏差 | 10.1029 | 38 | 0.2659 | 1.0000 |
| Pearson | 10.6073 | 38 | 0.2791 | 1.0000 |

唯一設定檔數目: 40

| 模型配適統計值 | | |
|------------|--------|--------|
| 準則 | 僅限截距 | 截距和共變量 |
| AIC | 39.098 | 14.103 |
| SC | 40.787 | 17.481 |
| -2 對數 L | 37.098 | 10.103 |

| 檢定全域虛無假設: BETA=0 | | | |
|------------------|---------|-----|------------|
| 檢定 | 卡方 | 自由度 | Pr > ChiSq |
| 概度比 | 26.9952 | 1 | <.0001 |
| 計分 | 26.1609 | 1 | <.0001 |
| Wald | 5.1398 | 1 | 0.0234 |

| 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 | -7.3050 | 2.8960 | 6.3627 | 0.0117 |
| first_week_boxoffice | 1 | 0.0785 | 0.0346 | 5.1398 | 0.0234 |

| 勝算比估計值 | | | |
|----------------------|-------|------------------|-------|
| 效果 | 點估計值 | 95% Wald 信賴界限 | |
| first_week_boxoffice | 1.082 | 1.011 | 1.158 |

| 預測機率和觀測回應的關聯性 | | | |
|---------------|------|-----------|-------|
| 和諧百分比 | 98.7 | Somers' D | 0.974 |
| 不和諧百分比 | 1.3 | Gamma | 0.974 |
| 繫結百分比 | 0.0 | Tau-a | 0.288 |
| 配對 | 231 | c | 0.987 |

| 估計共變異數矩陣 | | |
|-----------------------|-----------|-----------------------|
| 參數 | Intercept | first_week_boxoffice1 |
| Intercept | 8.386814 | -0.09605 |
| first_week_boxoffice1 | -0.09605 | 0.001198 |

| 變數 | Mean | 標準 差 | Skewness | Kurtosis | 二元隨機變數 |
|--------|---------|---------|----------|----------|--------|
| VAR5 | 43308.4 | 50600.2 | 1.6918 | 2.4394 | 0.6794 |
| budget | 8260.0 | 6799.1 | 1.1438 | 0.5512 | 0.6080 |

CLUSTER 程序
重心階層式集群分析

| 變數 | Mean | 標準 差 | Skewness | Kurtosis | 二元隨機變數 |
|--------|---------|---------|----------|----------|--------|
| VAR5 | 43308.4 | 50600.2 | 1.6918 | 2.4394 | 0.6794 |
| budget | 8260.0 | 6799.1 | 1.1438 | 0.5512 | 0.6080 |

CLUSTER 程序
重心階層式集群分析

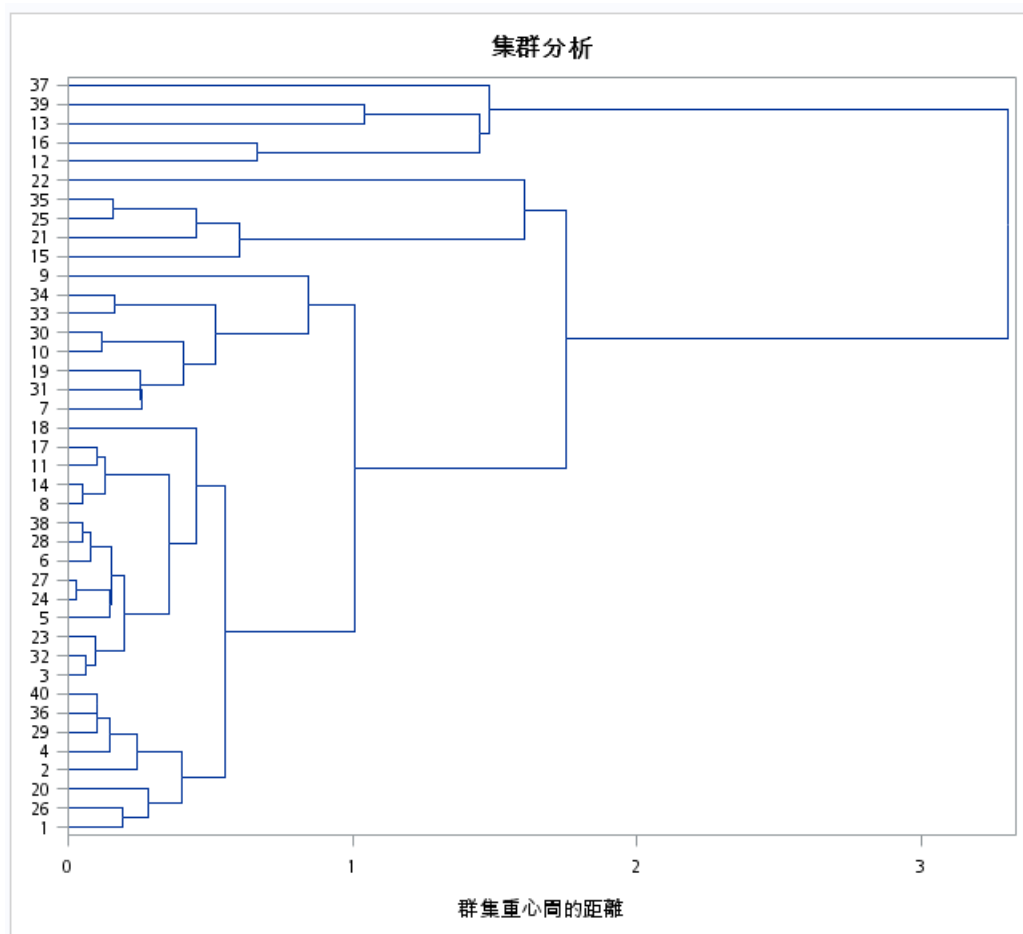
| 相關矩陣的特徵值 | | | | |
|----------|------------|------------|--------|--------|
| | 特徵值 | 差異 | 比例 | 累計 |
| 1 | 1.82394612 | 1.64789224 | 0.9120 | 0.9120 |
| 2 | 0.17605388 | | 0.0880 | 1.0000 |

已將資料標準化為平均值 0 和變異數 1

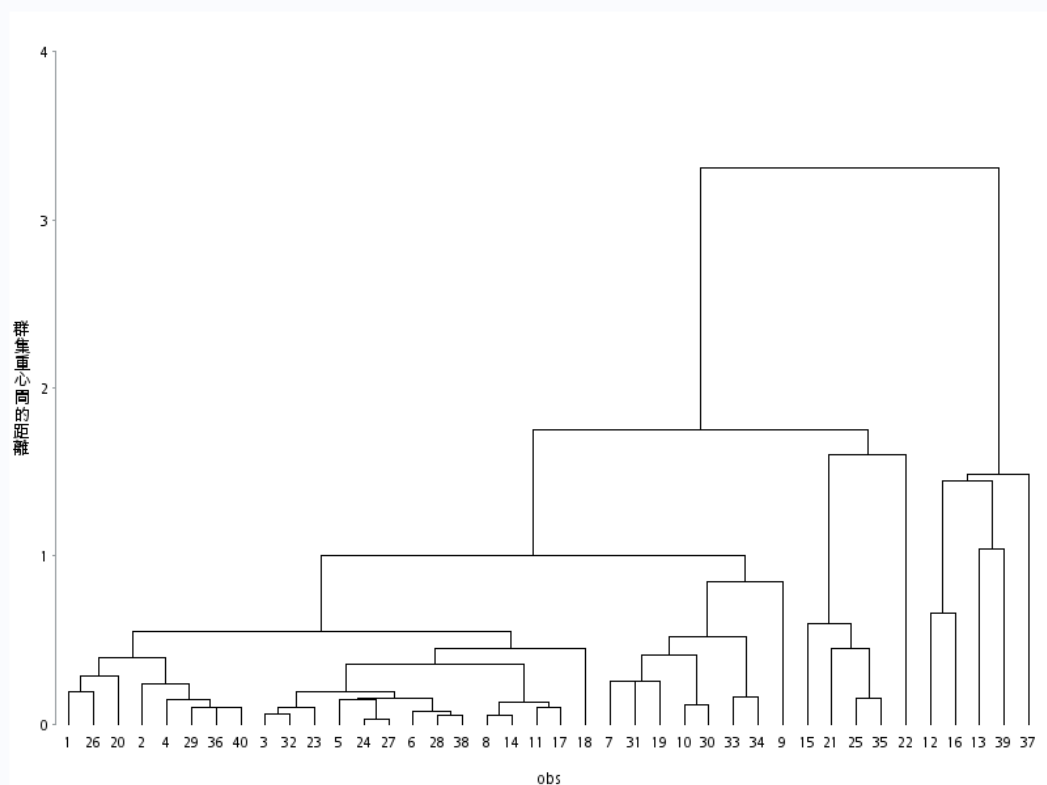
| | |
|-----------|---|
| 均方根總樣本標準差 | 1 |
|-----------|---|

| 集群歷史 | | | | | | | | |
|--------------|-------|------|----|----------------|------------|------|----------|----|
| 群集 數 目 | 聯結的集群 | | 次數 | 新集群 RMS 標準差 | 半偏 R 平方 | R 平方 | 重心 距離 | 繫結 |
| 39 | 24 | 27 | 2 | 0.0154 | 0.0000 | 1.00 | 0.0309 | |
| 38 | 8 | 14 | 2 | 0.0249 | 0.0000 | 1.00 | 0.0498 | |
| 37 | 28 | 38 | 2 | 0.0266 | 0.0000 | 1.00 | 0.0532 | |
| 36 | 3 | 32 | 2 | 0.0298 | 0.0000 | 1.00 | 0.0596 | |
| 35 | 6 | CL37 | 3 | 0.0365 | 0.0001 | 1.00 | 0.0767 | |
| 34 | CL36 | 23 | 3 | 0.0451 | 0.0001 | 1.00 | 0.0977 | |
| 33 | 36 | 40 | 2 | 0.0506 | 0.0001 | 1.00 | 0.1013 | |
| 32 | 29 | CL33 | 3 | 0.0549 | 0.0001 | 1.00 | 0.102 | |
| 31 | 11 | 17 | 2 | 0.0513 | 0.0001 | 1.00 | 0.1026 | |
| 30 | 10 | 30 | 2 | 0.0586 | 0.0001 | .999 | 0.1173 | |
| 29 | CL38 | CL31 | 4 | 0.0628 | 0.0002 | .999 | 0.1309 | |
| 28 | 4 | CL32 | 4 | 0.0680 | 0.0002 | .999 | 0.1447 | |
| 27 | 5 | CL39 | 3 | 0.0618 | 0.0002 | .999 | 0.149 | |
| 26 | CL27 | CL35 | 6 | 0.0741 | 0.0004 | .998 | 0.1511 | |
| 25 | 25 | 35 | 2 | 0.0789 | 0.0002 | .998 | 0.1578 | |
| 24 | 33 | 34 | 2 | 0.0809 | 0.0002 | .998 | 0.1618 | |
| 23 | 1 | 26 | 2 | 0.0966 | 0.0002 | .998 | 0.1932 | |
| 22 | CL34 | CL26 | 9 | 0.0936 | 0.0010 | .997 | 0.1965 | |
| 21 | 2 | CL28 | 5 | 0.0960 | 0.0006 | .996 | 0.2396 | |
| 20 | 7 | 31 | 2 | 0.1283 | 0.0004 | .996 | 0.2567 | |
| 19 | CL20 | 19 | 3 | 0.1383 | 0.0006 | .995 | 0.2556 | |
| 18 | CL23 | 20 | 3 | 0.1346 | 0.0007 | .995 | 0.2839 | |

| | | | | | | | | |
|----|------|------|----|--------|--------|------|--------|--|
| 17 | CL22 | CL29 | 13 | 0.1466 | 0.0045 | .990 | 0.3563 | |
| 16 | CL18 | CL21 | 8 | 0.1781 | 0.0038 | .986 | 0.3987 | |
| 15 | CL19 | CL30 | 5 | 0.1881 | 0.0026 | .984 | 0.4078 | |
| 14 | 21 | CL25 | 3 | 0.1917 | 0.0017 | .982 | 0.4492 | |
| 13 | CL17 | 18 | 14 | 0.1645 | 0.0024 | .980 | 0.45 | |
| 12 | CL15 | CL24 | 7 | 0.2377 | 0.0049 | .975 | 0.5171 | |
| 11 | CL16 | CL13 | 22 | 0.2539 | 0.0200 | .955 | 0.5537 | |
| 10 | 15 | CL14 | 4 | 0.2640 | 0.0035 | .951 | 0.6015 | |
| 9 | 12 | 16 | 2 | 0.3313 | 0.0028 | .948 | 0.6627 | |
| 8 | CL12 | 9 | 8 | 0.3053 | 0.0080 | .940 | 0.8461 | |
| 7 | CL11 | CL8 | 30 | 0.4144 | 0.0762 | .864 | 1.0067 | |
| 6 | 13 | 39 | 2 | 0.5198 | 0.0069 | .857 | 1.0396 | |
| 5 | CL9 | CL6 | 4 | 0.6893 | 0.0268 | .830 | 1.446 | |
| 4 | CL5 | 37 | 5 | 0.7591 | 0.0225 | .808 | 1.4827 | |
| 3 | CL10 | 22 | 5 | 0.5570 | 0.0265 | .781 | 1.6062 | |
| 2 | CL7 | CL3 | 35 | 0.6136 | 0.1687 | .613 | 1.7522 | |
| 1 | CL2 | CL4 | 40 | 1.0000 | 0.6127 | .000 | 3.3051 | |



TREE 程序
重心階層式集群分析



SAS 系統

MEANS 程序

CLUSTER=1

| 變數 | 標籤 | N | 平均值 | 標準差 | 最小值 | 最大值 |
|--------|-----------|----|------------|-----------|------------|-----------|
| VAR5 | boxoffice | 30 | -0.4613469 | 0.3686256 | -0.8170589 | 0.3894777 |
| budget | budget | 30 | -0.4745735 | 0.4555316 | -1.2001611 | 0.4029953 |

CLUSTER=2

| 變數 | 標籤 | N | 平均值 | 標準差 | 最小值 | 最大值 |
|--------|-----------|---|-----------|-----------|------------|-----------|
| VAR5 | boxoffice | 5 | 0.6434569 | 0.5307006 | 0.0148547 | 1.4354022 |
| budget | budget | 5 | 0.8854129 | 0.5821599 | -0.1264876 | 1.3590059 |

CLUSTER=3

| 變數 | 標籤 | N | 平均值 | 標準差 | 最小值 | 最大值 |
|--------|-----------|---|-----------|-----------|-----------|-----------|
| VAR5 | boxoffice | 5 | 2.1246247 | 0.8525260 | 0.8845547 | 3.2314818 |
| budget | budget | 5 | 1.9620280 | 0.6523892 | 0.9913095 | 2.4620951 |

| 集群清單 | | |
|------|----|--------|
| 觀測值 | 群集 | 從種子的距離 |
| 1 | 1 | 0.3318 |
| 2 | 1 | 0.4694 |
| 3 | 1 | 0.4032 |
| 4 | 1 | 0.3216 |
| 5 | 1 | 0.1159 |
| 6 | 1 | 0.2774 |
| 7 | 2 | 0.5602 |
| 8 | 1 | 0.6307 |
| 9 | 1 | 0.8132 |
| 10 | 1 | 0.7362 |
| 11 | 1 | 0.5280 |
| 12 | 3 | 0.3827 |
| 13 | 3 | 0.5397 |
| 14 | 1 | 0.6654 |
| 15 | 2 | 0.7702 |
| 16 | 3 | 1.0223 |
| 17 | 1 | 0.5137 |
| 18 | 1 | 0.3370 |
| 19 | 2 | 0.4889 |
| 20 | 1 | 0.3652 |
| 21 | 2 | 0.6144 |
| 22 | 2 | 1.2396 |
| 23 | 1 | 0.3728 |
| 24 | 1 | 0.2509 |
| 25 | 2 | 0.3181 |
| 26 | 1 | 0.3870 |
| 27 | 1 | 0.2356 |
| 28 | 1 | 0.3073 |
| 29 | 1 | 0.1755 |
| 30 | 1 | 0.8102 |
| 31 | 2 | 0.3155 |
| 32 | 1 | 0.4340 |
| 33 | 1 | 0.8216 |
| 34 | 1 | 0.6634 |
| 35 | 2 | 0.4681 |
| 36 | 1 | 0.2588 |
| 37 | 3 | 1.1862 |
| 38 | 1 | 0.2608 |
| 39 | 3 | 1.3114 |
| 40 | 1 | 0.2021 |

以最終種子為基礎的準則 = 0.4225

| 集群摘要 | | | | | | |
|------|----|---------|--------------|-------|-------|----------|
| 群集 | 次數 | RMS 標準差 | 從種子到觀測值的最大距離 | 半徑已超過 | 最近的群集 | 群集重心間的距離 |
| 1 | 27 | 0.3450 | 0.8216 | | 2 | 1.6066 |
| 2 | 8 | 0.4986 | 1.2396 | | 1 | 1.6066 |
| 3 | 5 | 0.7591 | 1.3114 | | 2 | 2.0830 |

| 變數的統計值 | | | | |
|----------|---------|---------|----------|-------------|
| 變數 | 總 STD | STD 內 | R 平方 | RSQ/(1-RSQ) |
| VAR5 | 1.00000 | 0.42569 | 0.828085 | 4.816826 |
| budget | 1.00000 | 0.45242 | 0.805811 | 4.149621 |
| OVER-ALL | 1.00000 | 0.43926 | 0.816948 | 4.462927 |

虛擬 F 統計值 = 82.56

近似預期整體 R 平方 = 0.69929

立方群集準則 = 4.814

WARNING: The two values above are invalid for correlated variables.

| 群集平均值 | | |
|-------|--------------|--------------|
| 群集 | VAR5 | budget |
| 1 | -0.538110411 | -0.562820630 |
| 2 | 0.488232193 | 0.673252117 |
| 3 | 2.124624712 | 1.962028016 |

| 群集標準差 | | |
|-------|--------------|--------------|
| 群集 | VAR5 | budget |
| 1 | 0.2975436202 | 0.3866236339 |
| 2 | 0.4622380179 | 0.5324474653 |
| 3 | 0.8525260283 | 0.6523891736 |

| 群集重心間的距離 | | | |
|----------|-------------|-------------|-------------|
| 最接近的群集 | 1 | 2 | 3 |
| 1 | . | 1.606628451 | 3.669471218 |
| 2 | 1.606628451 | . | 2.082960343 |
| 3 | 3.669471218 | 2.082960343 | . |

| SAS 系統 | | |
|--------|-----|---------|
| 群集=1 | | |
| Obs | obs | CLUSTER |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 3 | 1 |
| 4 | 4 | 1 |
| 5 | 5 | 1 |
| 6 | 6 | 1 |
| 7 | 8 | 1 |
| 8 | 9 | 1 |
| 9 | 10 | 1 |
| 10 | 11 | 1 |
| 11 | 14 | 1 |
| 12 | 17 | 1 |
| 13 | 18 | 1 |
| 14 | 20 | 1 |
| 15 | 23 | 1 |
| 16 | 24 | 1 |
| 17 | 26 | 1 |
| 18 | 27 | 1 |
| 19 | 28 | 1 |
| 20 | 29 | 1 |
| 21 | 30 | 1 |
| | | |
| 22 | 32 | 1 |
| 23 | 33 | 1 |
| 24 | 34 | 1 |
| 25 | 36 | 1 |
| 26 | 38 | 1 |
| 27 | 40 | 1 |
| 群集=2 | | |
| Obs | obs | CLUSTER |
| 28 | 7 | 2 |
| 29 | 15 | 2 |
| 30 | 19 | 2 |
| 31 | 21 | 2 |
| 32 | 22 | 2 |
| 33 | 25 | 2 |
| 34 | 31 | 2 |
| 35 | 35 | 2 |
| 群集=3 | | |
| Obs | obs | CLUSTER |
| 36 | 12 | 3 |
| 37 | 13 | 3 |
| 38 | 16 | 3 |
| 39 | 37 | 3 |
| 40 | 39 | 3 |