

# 基於貝氏階層（非參數）模型之分群

四寶雅數據為例與統計學派比較

統計三 111304019 林承佑

# Contents

1 摘要	1
2 研究目的	2
3 研究資料介紹	3
3.1 實驗背景	3
3.1.1 寶雅據點介紹	3
3.1.2 皮帶品項定位	3
3.1.3 皮帶款式類別	4
3.2 資料整理與研究設計	5
4 模型設計	7
4.1 模型一	7
4.1.1 模型一的數學推導	9
4.2 模型二	10
4.2.1 模型二的數學推導	11
4.3 模型三	12
4.4 模型四	13
4.4.1 模型四的數學推導	14
4.5 模型五	15
4.5.1 相關論文閱讀	15
4.5.2 階層與參數說明	16
5 分析結果	20
5.1 參數設計貢獻	20
5.2 款式與品項定位之其組合表現	21
5.3 季度與銷售情況分析	22
5.3.1 原始模型參照不同季度之結果	22
5.3.2 修正分佈與參數估計目標與其結果	23
5.4 模型之地區效應問題	24
5.4.1 原始模型地區估計結果	25
5.4.2 修正先驗參數與分組目標之結果	25
5.5 產品組合分析	26
5.6 進行整體模型的分組嘗試	26

5.7 改進模型，引入階層（無母數）分群	28
5.8 變異數分析與其交互作用	30
5.8.1 交互作用	32
5.9 事後檢定 Tukey-HSD	33
5.9.1 直轄市面向	34
5.9.2 一般縣市面向	34
<b>6 結論</b>	<b>35</b>
6.1 寶雅面向	35
6.2 廠商面向	35
6.3 貝氏分析之總結	36
<b>7 References</b>	<b>37</b>

# 1 摘要

本研究旨在利用貝氏階層模型探討零售業銷售數據中的分群結構，並比較統計學中貝氏學派與頻率學派的分析結果。以寶雅銷售數據為例，我們分析不同地區與季度下的產品組合與銷售模式，特別觀察款式與品項定位的交互影響。研究方法包括基於中國餐廳過程的貝氏非參數分群方法，結合階層貝氏線性模型，分析因子間的交互作用對銷售表現的影響。本研究進一步探討模型先驗分布的設置與假設修正，並延伸至統計學派之變異數分析實驗結果比較，評估兩種學派在分群結果上的差異與管理應用，與其中給予相對應的商業策略為目標。研究結果亦期望為台灣零售業提供創新的管理策略或意涵，並以貝氏為主軸作為核心驅動的優化管理經營建議。

## 2 研究目的

找出不同地區及季度的產品最佳組合。觀察在款式這項因子下，各個產品的銷售狀況與分佈是否相似。期待看出品項定位與款式的組合下，每組的銷售狀況的分組的情形；並藉此延伸在不同季度當中去比較與觀察，每季的銷售狀況（款式與品項定位之組合中）是否類似。

將先使用貝氏資料分析的方式，觀察款式在不同地區與季度中的銷售表現。並藉此作為基底模型延伸與修正，更動先驗分佈的參數與假設，並透過貝氏分群方法（基於中國餐廳過程和多項分布），對款式與品項定位的銷售表現進行分群。最後，應用階層貝氏線性模型分析因子（參數）間的交互作用與關係，對銷售總金額的影響。

最後，希望利用本項研究的資料內容與背景，在統計學中的兩大學派當中，即貝氏學派與頻率學派之變異數實驗設計分析，進行結果的比較與探討，期望得到對於台灣零售業管理模式下，與廠商倉儲及產品生命力的創新與生存，其為廠商及寶雅提出銷售策略，並提升各地區的總銷售額之結論。

### 3 研究資料介紹

#### 3.1 實驗背景

我們在 2024 年一整年中進行本場實驗，以同樣廠商的「皮帶產品」，在全台每間「實體的寶雅分店」中進行販售。

先行定義各個可被販售的皮帶產品中，每項的款式與品項定位；並在該年度中盡可能不主動或被動下架任一產品，意即仍由廠商進行產品販售的規劃安排，如在同樣品項定位下，可進行新增或修正不同的樣式，使其產品生命力延長，並成功持續於店鋪中販售。

由於寶雅公司的物流與管理模式緣故，所有產品非直接性寄送到各個分店進行上架，而是由寶雅的倉儲進行管理與分發；因此，我們記錄每筆寶雅的倉儲物流（及其倉儲系統下所提供的資料），與廠商進銷貨的狀況作為資料來源，比對各個分店的產品銷售狀況，藉此探討縣市間的銷售總額。

鑑於寶雅在 2024 年度中，仍於台灣各地進行據點的調整，不論是新增、合併或是減少，如新竹市於該 12 月時，開幕新分店，或寶雅將其子公司之特定店鋪的「寶家」轉型回寶雅等。然而，為了確保實驗的可靠性及其完整性，我們撇除所有寶雅營運上的問題與調整，只保留在 2024 該年度完整銷售一整年的店鋪的資料內容，其餘資料皆不採用。

在本研究分析當中，為求後續學派比較的環節具合理性，以及結果之可靠性；因此，會先對所有研究內容皆使用整筆實驗資料進行，並且以整理相同的結構與結果進行分析，再做後續的比較或結論。

##### 3.1.1 寶雅據點介紹

寶雅目前在全台灣中除了「嘉義市」與「連江縣」外，皆在各縣市中至少有一間分店數。而從我們取得的實驗資料中，經由整理與撇除不適合的數據後，總共留下 244 間分店的資料作為接下來分析使用。（此處的不適合指，該分店未於 2024 年一整年中進行完整銷售）

寶雅目前在全台灣中除了「嘉義市」與「連江縣」外，皆在各縣市中至少有一間分店數。而從我們取得的實驗資料中，經由整理與撇除不適合的數據後，總共留下 244 間分店的資料作為接下來分析使用。（此處的不適合指，該分店未於 2024 年一整年中進行完整銷售）

根據寶雅官網的分區規劃，將東北部列為北區，離島列為中區等，從 Figure 1 中，可以明顯發現北區的分店數明顯較多，而由於南部被劃分為兩個區塊，所以分店數量較少，這裡推估是鑑於寶雅其子公司寶家的分店狀況緣故，寶雅才於官網中以這樣的分區進行呈現。若將南區與高屏區合併觀看，北中南三區的分區狀況會落在平均 80 間左右。

而從 Figure 2 中，可以很明顯發現直轄市的明顯比一般縣市多出至少兩倍多的分店數；藉此我們可以推估，寶雅在西半部的分店狀況，北中南三區分店數平均，且以各區的直轄數分店數最多，而一般縣市分布零星的分店，以服務與推廣在各縣市的潛在與各類型的客群。

##### 3.1.2 皮帶品項定位

該廠商的皮帶分為三種品項定位，如 Figure 3 所示；分為男性、女性與中性。隨著時代的變化與開放的社會氛圍，其中的男性與女性的意涵亦逐漸廣義化，不再單單局限於性別上的差異，進而轉為風格或適用場合等的解釋。

此處仍以廠商對於品項定位的定義作為介紹與分類標準。左圖的男性品項較為嚴肅與穩重，

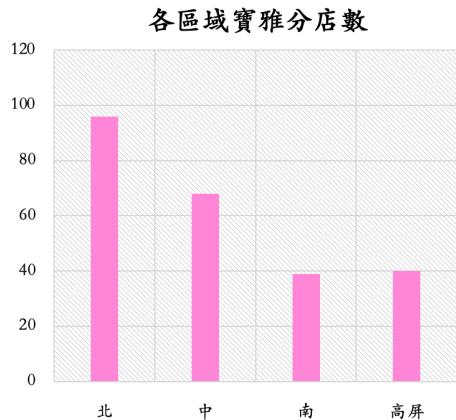


Figure 1: 各區域寶雅分店數

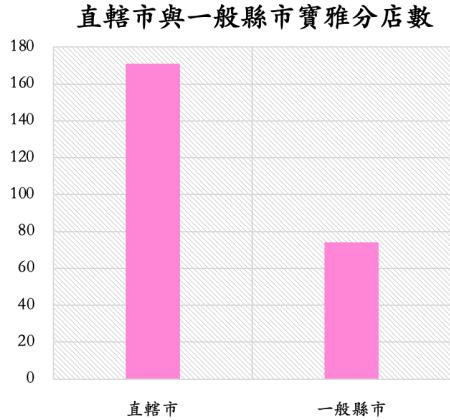


Figure 2: 直轄市與一般縣市寶雅分店數

花紋與顏色表現上較為單一與正式，不會有過多的變化，適合商務與金融行業等類型的對象。

中間圖片則為女性品項，相較於其他兩種而言，女性品項的長度會較短一些，而不論花紋、顏色、帶頭的款式乃至皮帶的寬度與種類，都會有多樣的變化（圖中僅呈現寬度較窄，花紋不同的造型），適合喜歡穿搭與風格多變的類型。

右圖則為中性的品項，介於男性與女性款之間，風格上較於女性品項保守，而在男性品項上開放，在皮帶本身會有不同的顏色呈現，且造型上較為耐看非嚴肅的類型，適合大部分的休閒風格與褲裝等，適合的場合數與使用的廣度也大幅提升，可謂實用且耐用型。



Figure 3: 品項定位圖（左：男性；中：女性；右：中性）

### 3.1.3 皮帶款式類別

皮帶的款式可以有很多種方式進行分類，如皮帶的種類或類型、帶頭的款式與使用方式等。本實驗以「帶頭的款式」作為款式的主要分類標準，如 Figure 4 所示；由上而下分為自動扣、一般款、針棒。

首先，圖中最上橫列為「自動扣」，其使用方式與另外兩款不同，主要依靠皮帶尾巴有一塊車入帶身的塑膠片（如該列最右側的圖片），當帶身尾巴插入帶頭後會自動卡住無法向後退出，即可藉由尾巴向內插入的長度來調節適合的緊度；而按下帶頭背面第二橫列處的一個機關按鈕（如該列中間的圖片），即可順利自動將帶身向外退出解開腰帶，因此稱其為「自動扣」。

其次，圖中下二橫列則依序分別為「一般款」與「針棒」，使用方式與大眾認知相同，藉由帶身上的孔洞進行腰帶的緊度調整；而一般款的設計，是依靠其帶頭背面的一個凸出的圓頭（點）進行調節，針棒款則是依靠中間那根針棒進行調節。藉此由於外表上的明顯差異，因此我們將其分為兩種款式。

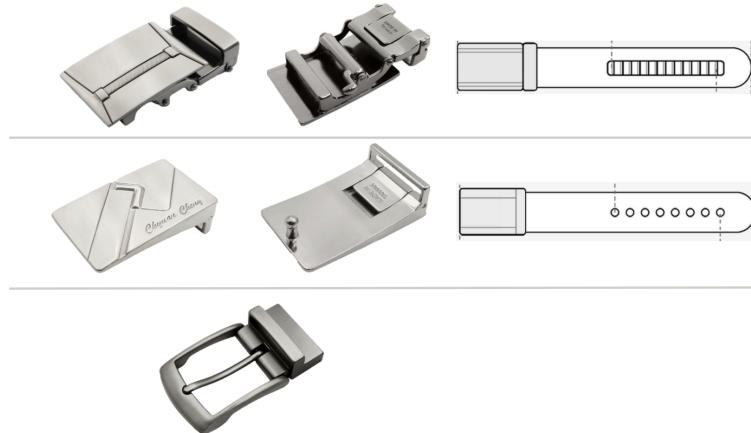


Figure 4: 款式圖 (上: 自動扣; 中: 一般款; 右: 針棒)

最後，我們其實可以發現款式的類型與使用方式具有密不可分的關係，而此處我們僅以帶頭的款式作為分類標準，而其於樣式或皮帶本身的花紋造型，則由前一節的品項定位進行分類；剩餘關於皮帶本身仍可被分類的標準或內容皆不在本實驗中可控的範圍內，歸於實驗背景中描述之「廠商進行產品販售的規劃安排，如在同樣品項定位下，可進行新增或修正不同的樣式，使其產品生命力延長」所條件與實驗設定。

### 3.2 資料整理與研究設計

根據實驗目標之訂定與描述，我們的目標變數  $y$  為「銷售總金額」，即我們於十實驗背景與資料收集與研究設計之兩節中，所共同描述之探討縣市間或提升各地區的銷售總額。(即實驗中所收集，且經由整理後所得到的數據值。) 各項因子及其符號表示如 Table 1 所呈現與下方列點描述與解釋。(Level 該欄位中，括號內的數字為該 Level 在其因子下，本研究中之代碼編號。)

Table 1: 各因子說明表。

EnglishName	Parameter	Level
Quarter	季度	第一季 (1)、第二季 (2)、第三季 (3)、第四季 (4)
ProductType	品項定位	中性 (1)、特定性別 (男性與女性合併) (2)
Style	款式	一般款 (1)、自動扣 (2)、針棒 (3)
Region	地區	一般縣市 (1)、直轄市 (2)

- 季度：

以三個月為一單位，1 月至 3 月為第一季，4 月至 6 月為第二季，以此類推至 12 月。

- 品項定位：

鑑於前節的介紹與說明，關於男性與女性的定位意涵已於時代的變化逐漸廣義化，而非指特定生理（或心理）性別上的意義；因此，我們將其品項定位分為兩種，將具有特地風格（男性或女性）合併為特定性別的定位，中性則保持不變。

- 款式：

保持如前節的介紹，三種款式。

- 地區：

依照臺灣各行政區類型劃分，分為直轄市與一般縣市。

Table 2: 品項定位與款式的關係。

	自動扣	一般款	針棒
男性	V	V	V
女性			V
中性	V	V	V

另外，補充說明品項定位與款式的關係部分，如 Table 2 所示，打勾 (V) 處為該品項定位存在該款式。我們可以從表格中明顯看出，品項定位為女性的部分僅有針棒該款式。因此，為了讓實驗與分析上更加順利，我們以前述列點之品項定位所描述，因其兩種定位皆具有「特定風格」的意涵存在，故以特定性別將兩種品項定位合併。

經由數據的整理，我們將每一個縣市下，各個分店的營業總額進行計算與合併，再依照我們實驗因子規劃進行整合與計算，實驗數據之每個 Treatment 的 mean 結果如 Table 4 所示；其中括號內所顯示之反覆數，恰為直轄市與一般縣市的數量。(與前節介紹相同，由於有兩縣市寶雅無分店開設，故一般縣市中僅有 14 個縣市。)

Table 3: 實驗數據表。

		第一季				第二季			
		中性		特定性別		中性		特定性別	
一般縣市	一般款	1374.15	(14)	1352.85	(14)	1438.6875	(14)	1298.4375	(14)
	自動扣	967.2375	(14)	915.864	(14)	1107.225	(14)	1057.35	(14)
	針棒	1895.475	(14)	1510.5	(14)	1373.1	(14)	1560.1875	(14)
直轄市	一般款	8270.675	(6)	7603.575	(6)	9153.9	(6)	6302.8875	(6)
	自動扣	5396.286	(6)	5294.919	(6)	5690.7375	(6)	5167.05	(6)
	針棒	8183.875	(6)	8950.375	(6)	9298.275	(6)	8950.375	(6)

Table 4: 實驗數據表。

		第三季				第四季			
		中性		特定性別		中性		特定性別	
一般縣市	一般款	1632.975	(14)	1428.0375	(14)	1900.9875	(14)	1467.5625	(14)
	自動扣	887.775	(14)	1157.1	(14)	1201.9875	(14)	798	(14)
	針棒	1462.65	(14)	1440.9375	(14)	1865.625	(14)	1401.1875	(14)
直轄市	一般款	11079.42	(6)	7324.7125	(6)	10704.575	(6)	7173.1625	(6)
	自動扣	6330.8	(6)	5632.55	(6)	6377.35	(6)	5586	(6)
	針棒	9298.275	(6)	8509.8125	(6)	10377.85	(6)	8996.75	(6)

註：因表格寬度過長，故在第二季與第三季中分段，以兩個表格進行呈現。

註：每一格為該 Treatment 的 mean；括號內則為其 Replication。

## 4 模型設計

### 4.1 模型一

第一個模型的建立相當直覺和簡單，引入回歸分析的概念；而其中由於本研究期待與變異數分析的概念與結果做比較，因此，此處所指回歸分析概念大致為「Kuehl, R.O. (2000) Design of Experiments: Statistical Principles of Research Design and Analysis」該本書第二章節所論證之方式，從複回歸的角度推導至變異數分析的概念，並從中延伸與擴展至 Main effect 或 Mixed effect 等；而本研究對於貝氏分析的第一模型，便是由此而生，將所有在意的因子（如前節說明）放入模型中進行估計。

其中，為了達成我們的研究目的，找到「最佳組合」的目標，在此之於前節對於皮帶的描述之論，我們先將「款式」與「品項定位」作為產品組合的關鍵因子（與其 Level），並形成一個交互作用項，作為另一估計的參數。

觀測模型假設：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

其中：

$$\mu[i] = \beta_0 + \beta_{\text{quarter}}[\text{Quarter}[i]] + \beta_{\text{region}}[\text{RegionType}[i]] + \beta_{\text{product}}[\text{ProductType}[i]]$$

$$+ \beta_{\text{style}}[\text{Style}[i]] + \beta_{\text{interaction}}[\text{ProductType}[i], \text{Style}[i]]$$

先驗參數之分佈如下，我們將所有參數都遵照於常態分佈之下，在我們關注的總銷售額而言，季度就好似一整個月中每日銷售的狀況的縮影，而某縣市之所有賣場分店的銷售狀況，當然不會日日相同或相異，因此，推斷為一種常態分佈之現象，而右側極端狀況，則可歸類於如活動或促銷等發生機率；而其他關於產品的部分，則是認為該廠商經營數年的經驗，各項產品類別應當因應市場狀態，有所調整或平衡，因此亦為常態分配之概念。

$$\beta_0 \sim \mathcal{N}(0, 1000)$$

$$\beta_{\text{quarter}}[j] \sim \mathcal{N}(0, 1000), \quad j = 1, 2, 3, 4$$

$$\beta_{\text{region}}[k] \sim \mathcal{N}(0, 1000), \quad k = 1, 2$$

$$\beta_{\text{product}}[m] \sim \mathcal{N}(0, 1000), \quad m = 1, 2$$

$$\beta_{\text{style}}[n] \sim \mathcal{N}(0, 1000), \quad n = 1, 2, 3$$

$$\beta_{\text{interaction}}[p, q] \sim \mathcal{N}(0, 1000), \quad p = 1, 2, \quad q = 1, 2, 3$$

$\sigma^2$  與自由度的先驗分布：

$$\tau = \frac{1}{\sigma^2}, \quad \sigma \sim \text{Uniform}(0, 100)$$

$$df \sim \text{Exponential}(1)$$

根據貝氏定理，後驗分布為：

$$p(\boldsymbol{\theta} \mid \text{data}) \propto p(\text{data} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$$

其中， $\boldsymbol{\theta}$  包含所有模型參數，此處會利用後小節所將提及之模型更新的想法：

$$\boldsymbol{\theta} = \{\beta_0, \beta_{\text{quarter}}, \beta_{\text{region}}, \beta_{\text{product}}, \beta_{\text{style}}, \beta_{\text{interaction}}, \tau, df\}$$

#### 4.1.1 模型一的數學推導

模型一中的 Likelihood 函數，為給定模型中的觀測數據  $Sales[i]$  之下的情況，其中可以表示為：

$$p(Sales|\theta) = \prod_{i=1}^N f(Sales[i]|\mu[i], \tau, df)$$

其中， $f(x|\mu, \tau, df)$  是  $t$  分布的概率密度函數，定義為：

$$f(x|\mu, \tau, df) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\Gamma\left(\frac{df}{2}\right)\sigma} \left[1 + \frac{1}{df}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{df+1}{2}}$$

where  $\sigma = \frac{1}{\sqrt{\tau}}$ 。

對 Likelihood 函數取對數後為：

$$\log p(Sales|\theta) = \sum_{i=1}^N \left[ \log \Gamma\left(\frac{df+1}{2}\right) - \log \sqrt{df\pi} - \log \Gamma\left(\frac{df}{2}\right) - \frac{1}{2} \log \tau - \frac{df+1}{2} \log \left(1 + \frac{(Sales[i] - \mu[i])^2\tau}{df}\right) \right]$$

根據貝氏定理，後驗分布為：

$$p(\theta|Sales) \propto p(Sales|\theta)p(\theta)$$

其中：

- $p(Sales|\theta)$  為似然函數，見上述推導。
- $p(\theta)$  為先驗分布的乘積：

$$p(\theta) = p(\beta_0) \prod_{j=1}^4 p(\beta_{\text{quarter}[j]}) \prod_{k=1}^2 p(\beta_{\text{region}[k]}) \prod_{m=1}^2 p(\beta_{\text{product}[m]}) \\ \prod_{n=1}^3 p(\beta_{\text{style}[n]}) \prod_{p=1}^2 \prod_{q=1}^3 p(\beta_{\text{interaction}[p,q]}) p(\tau) p(df)$$

其中參數的分佈將會有以下關係存在：1. 固定效果參數  $\beta$  的後驗分布：

$$p(\beta|Sales) \propto \prod_{i=1}^N f(Sales[i]|\mu[i], \tau, df) \cdot \prod_{\text{coeff}} \mathcal{N}(\beta; 0, 1000)$$

2. 精確度  $\tau$  的後驗分布：

$$p(\tau|Sales) \propto \prod_{i=1}^N f(Sales[i]|\mu[i], \tau, df) \cdot \frac{1}{\sigma^2}$$

3. 自由度  $df$  的後驗分布：

$$p(df|Sales) \propto \prod_{i=1}^N f(Sales[i]|\mu[i], \tau, df) \cdot \text{Exp}(df; 1)$$

## 4.2 模型二

此模型為模型一的修正版本，關鍵更動之處為「季度」的參數，我們不再對每一季的狀態都做估計，而是將其視為一個整體因子，只對整體做一次的估計，並將其結果以相同效應影響其他因子的估計，藉此來達成我們的研究過程的想法與目標。

觀測模型：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

其中：

$$\begin{aligned} \mu[i] = & \beta_0 + \beta_{\text{quarter}} + \beta_{\text{region}[RegionType[i]]} + \beta_{\text{product}[ProductType[i]]} \\ & + \beta_{\text{style}[Style[i]]} + \beta_{\text{interaction}[ProductType[i], Style[i]]} \end{aligned}$$

先驗分布的部分如同模型一，包含修正後的每季參數的估計，一樣是遵照常態分佈。

$$\begin{aligned} \beta_0 & \sim \mathcal{N}(0, 0.001) \\ \beta_{\text{quarter}} & \sim \mathcal{N}(0, 0.001) \\ \beta_{\text{region}[k]} & \sim \mathcal{N}(0, 0.001), \quad k = 1, 2 \\ \beta_{\text{product}[m]} & \sim \mathcal{N}(0, 0.001), \quad m = 1, 2 \\ \beta_{\text{style}[n]} & \sim \mathcal{N}(0, 0.001), \quad n = 1, 2, 3 \\ \beta_{\text{interaction}[p, q]} & \sim \mathcal{N}(0, 0.001), \quad p = 1, 2; q = 1, 2, 3 \end{aligned}$$

$\sigma^2$  與自由度的先驗分布（同模型一）：

$$\begin{aligned} \tau & = \frac{1}{\sigma^2} \\ \sigma & \sim \text{Uniform}(0, 100) \\ df & \sim \text{Exponential}(1) \end{aligned}$$

最後，在此解釋使用  $t$ -分布進行建模的緣故，在於其對於異常值具有比較大的包容性，因為我們知道寶雅的市場定位與目標客群的限制，皮帶產品不一定是其主力產品，另外，鑑於設點與各縣市的分佈都將間接影響銷售總金額；因此，我們需要容錯率相對較高的分佈來進行模型的建立，確保我們不忽略表現不好的分店狀況，以致模型與結果的偏頗解釋。

#### 4.2.1 模型二的數學推導

相關的數學算是如同模型一該小節的介紹，而這裡將本模型有不同之參數，及其影響之參數估計的後驗分佈列出。

$\beta_0$  和  $\beta_{\text{quarter}}$  的後驗分布：

$$p(\beta_0, \beta_{\text{quarter}} | Sales) \propto \prod_{i=1}^N f(Sales[i] | \mu[i], \tau, df) \cdot \mathcal{N}(\beta_0; 0, 1000) \cdot \mathcal{N}(\beta_{\text{quarter}}; 0, 1000).$$

$\beta_{\text{region}}, \beta_{\text{product}}, \beta_{\text{style}}, \beta_{\text{interaction}}$  的後驗分布：

$$p(\beta_{\text{region}}, \beta_{\text{product}}, \beta_{\text{style}}, \beta_{\text{interaction}} | Sales) \propto \prod_{i=1}^N f(Sales[i] | \mu[i], \tau, df) \prod_{\text{coeff}} \mathcal{N}(\beta; 0, 1000).$$

很明顯我們可以觀察出，其簡化假設的影響；與模型一相比，模型二通過對  $\beta_{\text{quarter}}$  設置單一參數簡化了季度效應的建模，減少了自由參數的數量，從而期待提升模型穩定性，可從模型中預期能夠適用的環境為，數據量有限或季度間效應差異不大的情況之結果。

### 4.3 模型三

這裡仍然是從模型一進行延伸的不同模型，是為解釋在研究過程中發現與遇見的情況；然而，我想亦為「預見」之結果來形容該模型的調整。本模型將「地區」的兩個 Level 假設為不同的分佈情況，其餘則皆不變（包含季度）。

觀測模型：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

其中：

$$\begin{aligned} \mu[i] = & \beta_0 + \beta_{\text{quarter}}[\text{Quarter}[i]] + \beta_{\text{region}}[\text{RegionType}[i]] + \beta_{\text{product}}[\text{ProductType}[i]] \\ & + \beta_{\text{style}}[\text{Style}[i]] + \beta_{\text{interaction}}[\text{ProductType}[i], \text{Style}[i]] \end{aligned}$$

參數之先驗分布設置如下所示，皆服從常態分配之模型一的說明。

$$\begin{aligned} \beta_0 & \sim \mathcal{N}(0, 0.001) \\ \beta_{\text{quarter}}[j] & \sim \mathcal{N}(0, 0.001), \quad j = 1, 2, 3, 4 \\ \beta_{\text{product}}[m] & \sim \mathcal{N}(0, 0.001), \quad m = 1, 2 \\ \beta_{\text{style}}[n] & \sim \mathcal{N}(0, 0.001), \quad n = 1, 2, 3 \\ \beta_{\text{interaction}}[p, q] & \sim \mathcal{N}(0, 0.001), \quad p = 1, 2, q = 1, 2, 3 \\ \tau & = \frac{1}{\sigma^2}, \quad \sigma \sim \text{Uniform}(0, 100) \\ df & \sim \text{Exponential}(1) \end{aligned}$$

而有所改變之分佈的參數為「地區」因子，我們認為由於寶雅在一般縣市與直轄市的數量分佈上有著明顯的差距，而根據數理統計學的推論與說明，我們亦可得知必然能讓其顯著；甚至於貝氏統計的概念中，他們將無法分出明缺的差異之估計結果；因此，我們將對於分店銷售狀況會有明顯差異，變異相對可能較大的一般地區寶有較大的變異數，而直轄市則縮小接受變異的可能值，藉此期待分出明顯的地區因子估計結果。

$$\beta_{\text{region}}[k] \sim \begin{cases} \mathcal{N}(0, 0.002) & \text{if } k = 1 \text{ (一般地區)} \\ \mathcal{N}(0, 0.0005) & \text{if } k = 2 \text{ (都市地區)} \end{cases}$$

而利用使用貝氏定理進行推導，與其中的後驗分佈皆無更動。

$$p(\boldsymbol{\theta} | \text{data}) \propto p(\text{data} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$$

其中：

$$p(\text{data} | \boldsymbol{\theta}) = \prod_{i=1}^N t(Sales[i] | \mu[i], \tau, df),$$

$$p(\boldsymbol{\theta}) = p(\beta_0) \cdot p(\beta_{\text{quarter}}) \cdot p(\beta_{\text{region}}) \cdot \dots \cdot p(\tau) \cdot p(df).$$

## 4.4 模型四

該模型的結構，我們從模型一進行延伸，加入新的一個參數內容放入產品組合的因子當中，希望藉由貝氏統計的分組概念，嘗試將產品銷售組合的狀況再總銷售額下，做分組的估計嘗試。

觀測模型：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

其中：

$$\begin{aligned} \mu[i] = & \beta_0 + \beta_{\text{quarter}}[\text{Quarter}[i]] + \beta_{\text{region}}[\text{RegionType}[i]] + \beta_{\text{product}}[\text{ProductType}[i]] \\ & + \beta_{\text{style}}[\text{Style}[i]] + z[i] \cdot \beta_{\text{interaction}}[\text{ProductType}[i], \text{Style}[i]] \end{aligned}$$

潛在模型，定義為白努力事件，藉此來達成分組的效果（此處不以產品組合作為分組的想法，而是由整題模型與實驗目的之售出狀況（銷售總金額）的概念思考）：

$$z[i] \sim \text{Bernoulli}(\pi[i])$$

$\pi[i]$  的建模方式：

$$\text{logit}(\pi[i]) = \alpha + \beta_{\text{latent}}[\text{RegionType}[i]]$$

其中， $\text{logit}(\pi[i])$  定義為：

$$\text{logit}(\pi[i]) = \log \left( \frac{\pi[i]}{1 - \pi[i]} \right)$$

先驗分布：

$$\begin{aligned} \beta_0 & \sim \mathcal{N}(0, 0.001) \\ \beta_{\text{quarter}[j]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } j = 1, 2, 3, 4 \\ \beta_{\text{region}[k]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } k = 1, 2 \\ \beta_{\text{product}[m]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } m = 1, 2 \\ \beta_{\text{style}[n]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } n = 1, 2, 3 \\ \beta_{\text{interaction}[p,q]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } p = 1, 2, q = 1, 2, 3 \\ \alpha & \sim \mathcal{N}(0, 0.001) \\ \beta_{\text{latent}[k]} & \sim \mathcal{N}(0, 0.001) \quad \text{for } k = 1, 2 \end{aligned}$$

$$\tau = \frac{1}{\sigma^2}, \quad \sigma \sim \text{Uniform}(0, 100), \quad df \sim \text{Exponential}(1)$$

#### 4.4.1 模型四的數學推導

引入潛在變量  $z[i]$ ，對於  $z[i]$  的條件分布，概率為：

$$p(z|\pi) = \prod_{i=1}^N \pi[i]^{z[i]} (1 - \pi[i])^{1-z[i]},$$

其中：

$$\pi[i] = \frac{\exp(\alpha + \beta_{\text{latent}}[\text{RegionType}[i]])}{1 + \exp(\alpha + \beta_{\text{latent}}[\text{RegionType}[i]])}.$$

Likelihood 函數為：

$$p(\text{Sales}|z, \theta) = \prod_{i=1}^N f(\text{Sales}[i]|\mu[i] + z[i] \cdot \beta_{\text{interaction}}[\text{ProductType}[i], \text{Style}[i]], \tau, df).$$

根據貝氏定理，後驗分布為：

$$p(z, \theta|\text{data}) \propto p(\text{Sales}|z, \theta) \cdot p(z|\pi) \cdot p(\theta),$$

其中：

- $p(\text{Sales}|z, \theta)$ ：條件 Likelihood 函數，描述在參數  $\theta$  和潛在變量  $z$  下，觀測到數據的可能性。
- $p(z|\pi)$ ：潛在變量的先驗分布。
- $p(\theta)$ ：所有參數的先驗分布，包括  $\beta_0, \beta_{\text{quarter}}, \beta_{\text{region}}, \beta_{\text{product}}, \beta_{\text{style}}, \beta_{\text{interaction}}, \alpha, \beta_{\text{latent}}, \tau, df$ 。

## 4.5 模型五

### 4.5.1 相關論文閱讀

第一篇我閱讀了邱逸彥的「利用貝氏理論於旅行時間推估之研究」，其中主要提及貝氏更新方法的概念，將歷史資料視為先驗資訊，假設其服從常態分配。而即時速度資料作為觀察值，通過貝氏更新修正歷史速度資料。以下是他的核心公式推導：

參數定義：

- 歷史速度的均值： $\theta_0$ ，方差： $\phi_0$
- 即時數據的均值： $x$ ，方差： $\phi$
- 更新後的均值： $\theta_1$ ，方差： $\phi_1$

公式：

$$\begin{aligned}\phi_1 &= \left( \frac{1}{\phi_0} + \frac{1}{\phi} \right)^{-1} \\ \theta_1 &= \phi_1 \left( \frac{\theta_0}{\phi_0} + \frac{x}{\phi} \right)\end{aligned}$$

該方法他實現了將歷史與即時數據的整合，而權重的決定則交由方差大小決定。而為計算長路徑旅行時間，加入了延遲加總的方法。此方法逐段考慮每段路徑的旅行時間，將其累積以得到總旅行時間。

而他的實驗本體中則是，使用高速公路車輛偵測器（VD）數據，結合歷史與即時數據進行模型測試。或得到的結論與關鍵的貝氏更新意涵整理如下：

- 不更新法在長路段和中等路段中效果最佳，但依賴即時數據的完整性和穩定性。
- 貝氏更新法更適合應對數據不穩定情況，且權重（方差或標準差）對結果影響不大。
- 路段長度越短，旅行時間的推估準確性越差。

利用貝氏更新的概念整合歷史數據與即時數據的推估方式，其概念與我期待的模型中，所謂的階層結構，可能具有一些重疊之處。如該論文的貝氏更新方法可以當作我於前四個模型中地區、季度等固定效果的先驗分布的調整基礎；而論文強調異質數據的整合（如不同時間區間、路段），與我模型中不同因子（品項定位、款式）的交互效應的概念有所類似。而我僅成功使用最為簡單的「變異數」與「 $\tau$ 」，作為權重調整的方式。

第二篇我則閱讀林哲遠的「基於中國餐廳過程之貝氏探索分群法」。而所提及之核心思想，即為中國餐廳的模型，其為將資料點類比為顧客進入餐廳選擇座位的行為，而每位客人有以下選擇：

- 選擇已有的桌子，其概率與該桌人數成正比；
- 選擇新桌子，其概率與參數  $\alpha$  成正比。

轉為數學公式上則為：

- 選擇已有桌子的概率：

$$P(z_i = j | z_{-i}) = \frac{n_j}{\alpha + N - 1}$$

- 選擇新桌子的概率：

$$P(z_i = \text{new}) = \frac{\alpha}{\alpha + N - 1}$$

而作者則在貝氏探索分群法上，基於中國餐廳的模型概念，並嘗試修正為：

- 基於切比雪夫不等式，使用資料分佈的特徵（均值與標準差）設計更精確的先驗分布（Prior Parameters）。
- 提出了新群判斷機制，利用熵（Entropy）判斷資料點是否適合現有群或需要新建群。

此處提及的熵，定義為：

$$H = - \sum_j P(z = j) \log P(z = j)$$

熵值與門檻值  $\epsilon$  的關係決定是否創建新群：

$$H > \epsilon \Rightarrow \text{新建群}$$

該作者利用常見的資料集（包含文字型態資料與數值行資料，並使用不同的混合模型），做了多次實驗與設計，其提出的方法通過改進先驗參數設計和新群判斷機制，有效提升了分群效果。利用熵作為新群判斷標準，減少了傳統中國餐廳模型中，參數設置不當對分群結果具有顯著影響的問題。

而我則利用模型中的交互效應分群（基於中國餐廳過程）讓模型加入無參數的概念進行實作語言勁，其中熵的計算與應用在模型則是擴展了這樣的概念生成。模型中先驗分佈的設定，本期待與論文提及的基於切比雪夫不等式優化先驗設定類似，事實上於最後一個模型中沒有成功實現，仍先基於常態分佈而向下探討。

#### 4.5.2 階層與參數說明

本模型是一個貝氏階層模型，目標仍是分析銷售總金額的產品狀況。而核心包括以下幾部分：

1. 數據的觀測模型：以  $t$ -分布為基礎，考慮觀測數據的異值性。
2. 固定效應：包括季節、地區、產品類型、款式等的主效應，以及其交互效應。
3. 熵的計算：透過對交互效應進行正規化以推導概率分布，並計算熵值。
4. 先驗分布：為所有參數指定先驗分布。

模型觀測數據  $Sales[i]$  被假設服從  $t$ -分布：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

均值的表達式為：

$$\begin{aligned}\mu[i] = & \beta_0 + \beta_{\text{quarter}}[\text{Quarter}[i]] + \beta_{\text{region}}[\text{RegionType}[i]] + \beta_{\text{product}}[\text{ProductType}[i]] \\ & + \beta_{\text{style}}[\text{Style}[i]] + \beta_{\text{interaction}}[\text{ProductType}[i], \text{Style}[i], \text{RegionType}[i]]\end{aligned}$$

熵的計算正規化交互效應，對每組 (ProductType, Style, RegionType) 的交互效應進行正規化：

$$\begin{aligned}\text{exp\_beta\_interaction}[p, s, r] &= \exp(\beta_{\text{interaction}}[p, s, r]) \\ \text{prob}[p, s, r] &= \frac{\text{exp\_beta\_interaction}[p, s, r]}{\sum_{r'} \text{exp\_beta\_interaction}[p, s, r']}\end{aligned}$$

熵的計算如下，而從算式可得其作用是衡量交互效應的分布均勻性，並可用於模型擴展。

$$\text{entropy}[p, s, r] = -\text{prob}[p, s, r] \cdot \log(\text{prob}[p, s, r] + 10^{-10})$$

先驗分布：

$$\beta_0 \sim \mathcal{N}(0, 0.001)$$

$$\beta_{\text{quarter}}[j] \sim \mathcal{N}(\mu_{\text{quarter}}[j], \text{prec}_{\text{quarter}}[j])$$

$$\beta_{\text{region}}[k] \sim \mathcal{N}(\mu_{\text{region}}[k], \text{prec}_{\text{region}}[k])$$

$$\beta_{\text{product}}[m] \sim \mathcal{N}(\mu_{\text{product}}[m], \text{prec}_{\text{product}}[m])$$

$$\beta_{\text{style}}[n] \sim \mathcal{N}(\mu_{\text{style}}[n], \text{prec}_{\text{style}}[n])$$

$$\beta_{\text{interaction}}[p, q, r] \sim \mathcal{N}(0, 0.001)$$

$$\tau = \frac{1}{\sigma^2}, \quad \sigma \sim \text{Uniform}(0, 100)$$

$$\text{df} \sim \text{Exponential}(1)$$

模型中階層結構的想法為，期待層與層之間參數有條件相依的關係；化為基礎與原始的形式及想法為：

- 第一層：觀測數據層

$$y_i \sim \text{Likelihood}(y_i | \theta_i)$$

其中  $\theta_i$  是個體層級參數。

- 第二層：群組層

$$\theta_i \sim P(\theta_i | \phi)$$

其中  $\phi$  是群組層級參數。

- 第三層：全域層

$$\phi \sim P(\phi)$$

上述這樣的結構核心是引入共享的先驗分布  $P(\phi)$  實現群組內的數據共享資訊。

模型中的階層結構：

1. 觀測數據層：

$$Sales[i] \sim t(\mu[i], \tau, df)$$

2. 固定效應與交互效應的先驗分布層：固定效應有條件獨立的正態先驗分布：

$$\beta_{\text{quarter}}[j] \sim \mathcal{N}(\mu_{\text{quarter}}[j], \text{prec}_{\text{quarter}}[j])$$

$$\beta_{\text{region}}[k] \sim \mathcal{N}(\mu_{\text{region}}[k], \text{prec}_{\text{region}}[k])$$

$$\beta_{\text{interaction}}[p, q, r] \sim \mathcal{N}(0, 0.001)$$

3. 全域參數層：超參數  $\phi$  描述固定效應與交互效應的全域屬性：

$$\mu_{\text{quarter}}[j] \sim \mathcal{N}(0, \sigma_{\text{quarter}}^2)$$

$$\text{prec}_{\text{quarter}}[j] = \frac{1}{\sigma_{\text{quarter}}^2}, \quad \sigma_{\text{quarter}} \sim \text{Uniform}(0, 10)$$

(參考論文中所敘述之中國餐廳模型過程：分群的貝氏非參數模型，用於自適應決定群組數量  $K$ 。數學形式為：

$$P(z_i = k \mid z_{-i}, \alpha) = \begin{cases} \frac{n_k}{N+\alpha-1} & \text{已有群組 } k \\ \frac{\alpha}{N+\alpha-1} & \text{新群組} \end{cases}$$

其中：

- $z_i$  是樣本  $i$  的群組標籤。
- $n_k$  是已有群組  $k$  的樣本數。
- $\alpha$  是超參數，控制新群組的生成概率。

在模型中，交互效應  $\beta_{\text{interaction}}[p, q, r]$  可能根據  $p, q, r$  的分群結果進行更新。

每組的分布狀態以多項分布表達：

$$\text{prob}[p, s, r] \sim \text{Multinomial}(\exp\_beta\_interaction[p, s, :])$$

)

後驗分布推導，根據貝氏定理，模型的後驗分布為：

$$P(\theta \mid \text{Data}) \propto P(\text{Data} \mid \theta)P(\theta)$$

其中：

- $\theta = \{\beta_0, \beta_{\text{quarter}}, \beta_{\text{region}}, \beta_{\text{product}}, \beta_{\text{style}}, \beta_{\text{interaction}}, \tau, df\}$ 。
- $P(\text{Data} \mid \theta)$  為似然函數，由  $t$ -分布給出：

$$P(\text{Data} \mid \theta) = \prod_{i=1}^N t(Sales[i] \mid \mu[i], \tau, df)$$

- $P(\theta)$  為先驗分布的聯合分布。

## 5 分析結果

根據實驗目標的建立與說明，我們將環繞且抽絲剝繭地為這段描述進行一步步的分析：

- 找出不同地區及季度的產品最佳組合，並為廠商及寶雅提出銷售策略，並提升各地區的總銷售額。

這段目標當中所提及的「最佳組合」為何呢？由我們前各小節中的介紹與說明，我們認為在我們的實驗設計與因子的規劃下，最為重要亦影響著產品本身的即「款式」因子；不同的使用方式與其款式，很有可能會直接與間接性地影響潛在客戶。

對於本研究而言，我們最為關心的便為「款式」，然而，款式又分為三款；因此，我們期待亦認為：

- 預期每種款式的售出狀況應該相似。

因為該廠商在過去幾十年的經營下，我們認為廠商應該已經對於市場的定位與潛在客戶具有一定的調整與修正，當今能在各大通路販售的產品與其款式，即為具有一定的潛在客戶與銷量，否則該款式並不會保留於此，乃至於寶雅販售。

更進一步地可以從寶雅進貨與陳列的演算法說明，由於寶雅為美妝生活百貨，能給予各個產品類別的空間具有限制，而皮帶並非主要且熱門產品（此處指於寶雅實體店的市場定位），因此皮帶於寶雅中具有品項和數量的雙重限制。甚至基於營運管理與演算法的計算，無法即時且充分補足於各個分店的各種款式。

藉此我們可以得知，能夠被廠商列在寶雅販售的商品在市場中，或寶雅的潛在客群中皆具有一定的市場定位。因此，撇除各個產品的單價差異（亦可透過該款式產品的銷量進行平衡），我們便認為且預期每種款式的售出狀況應該相似。

### 5.1 參數設計貢獻

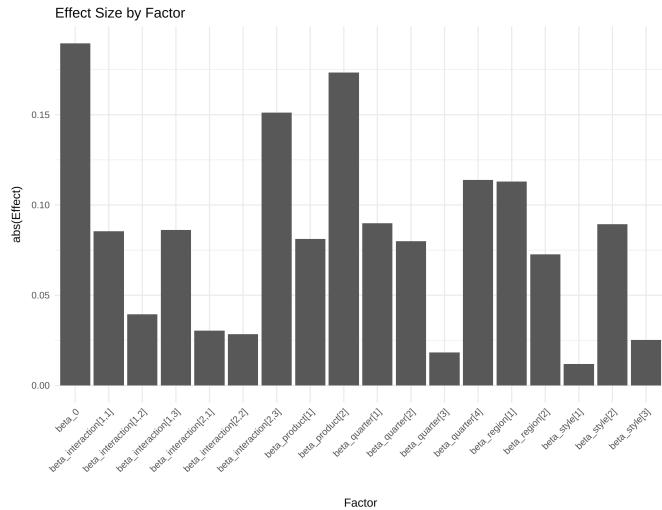


Figure 5: 各參數的貢獻 Effect。

從 Figure 5 中，長條圖中使用了效應值的絕對值，期待能夠比較不同參數對模型結果的影響大小，而不考慮方向（正向或負向）。而高 Effect 值，直覺而言，其該參數對銷售數據具有顯著影響，是需要關注的主要因素。

而高 Effect 值的參數對 Likelihood 的貢獻亦更大，這些參數調整時會對 Likelihood 造成較大改變。如圖中款式和品項定位的值較高，則表明其選擇對銷售數據有一定顯著影響，而不亦不排除為迭代與抽樣後的結果；然而，我們仍可藉此為出發點，開始針對參數的調整與事後的分佈，達成我們的研究目的。

意即，在此從 Figure 5 中，我們獲得了「款式」與「品項定位」的重要程度，而季度在其中的表現極高又極低，則可部分推估其並不是影響或被影響售出狀況的關鍵（Likelihood 相對整體拉低或偏高）。

## 5.2 款式與品項定位之其組合表現

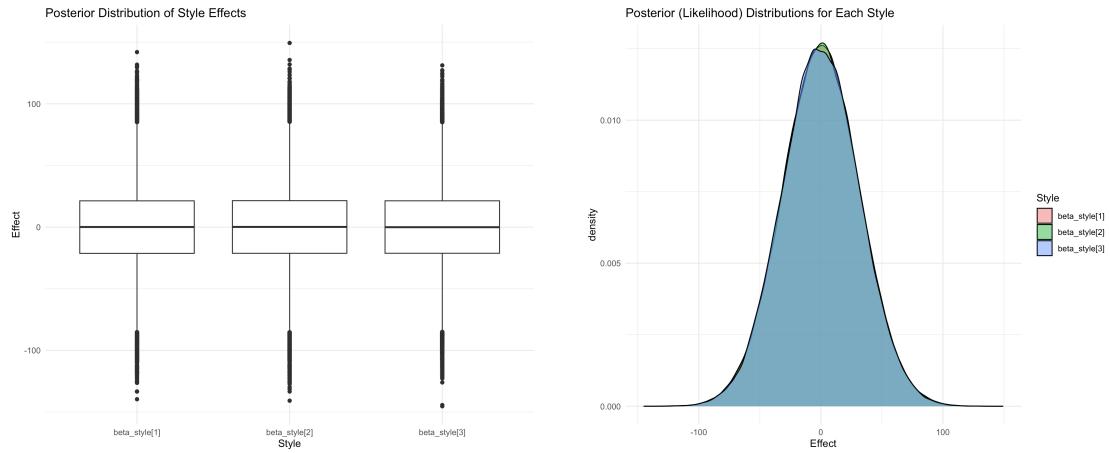


Figure 6: 模型一中款式的後驗與 Effect 狀態。

我們可以從模型結果之發現，亦從 Figure 6 各項分佈的狀況可以得知，在該模型下，事實上我們所在意的款式，並沒有直接性的影響了銷售總金額的情況，他近乎重疊的結果，代表顧客願意花錢與在意的因子並沒有直接性地落在產品的使用方式（如研究資料介紹該節中的說明）乃至外觀的款式。

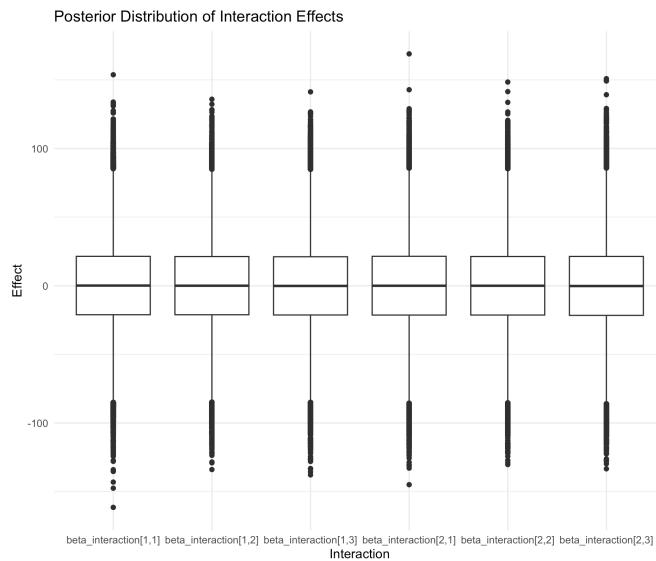


Figure 7: 模型一中款式與品項定位的後驗與 Effect 狀態。

而從 Figure 7 中，我們亦可得到類似的結果，發現在產品組合下的 Effect 却是相似且

Normal 的結果，連異常值或偏離箱型圖的值域皆不多；藉此同時，鑑於貝氏統計的概念，我將懷疑這項研究中，可能存在先驗分佈的猜測錯誤，或有其他相關參數間接影響估計結果，又或在機率概念下，被其他分母（因子）間接解釋掉不應當，或期待看見卻被解釋的部分。

### 5.3 季度與銷售情況分析

因此，由於我們從 Figure 5 中，發現季度貢獻上的偏頗情況的緣故；在此我們來觀察季度在本模型中的結果為何。

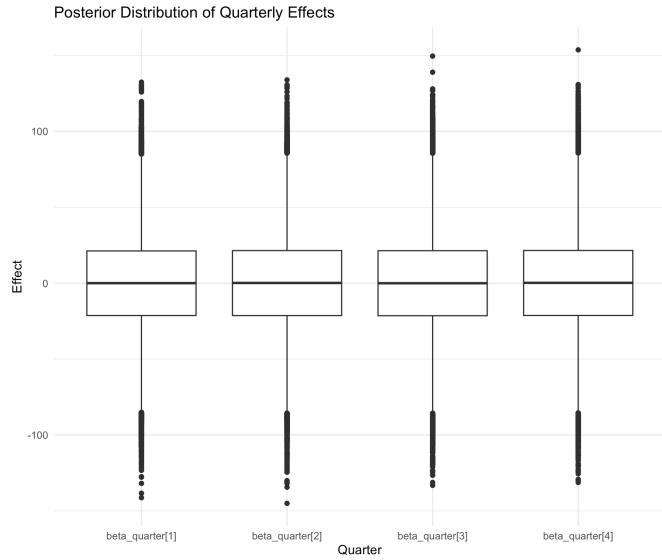


Figure 8: 模型一中季度的後驗與 Effect 狀態。

結果如 Figure 8 所呈現，事實上季度的表現一樣為類似的結果，發現四個季度下的 Effect 是相似且 Normal 的結果；然而，這樣的情況從季度的角度來思考，會發現事實上我們並不需要每個月的資料進行參數的估計，因為每季的資料結果對於我們實驗目的的銷售總金額而言，是提供相似的 Likelihood 狀態，並不直接性地影響我們所期待的售出狀況，而藉此我們可能並不需要這麼多無辜的參數存在在模型當中。而此推估尚不完全與可靠，因此我們將繼續向下探討與討論。

#### 5.3.1 原始模型參照不同季度之結果

由於上節的推論與想法，在此我們將其延伸討論為，在不同季度當中去比較與觀察，每季的銷售狀況（組合下）是否類似？

從 Figure 9 可以明顯看出，每季的各個銷售組合下的狀況是非常均勻平均的，甚至無法看出超過一種組合的銷售結果，或其提供的後驗分佈與 Likelihood（Effect）。

因此，我們應當減少不必要的模型估計，讓模型更加穩定與完善，並達成我們的研究目的，找到影響銷售狀況的產品組合，或為與廠商倉儲及產品生命力的創新與生存給出具有建設性的提議。

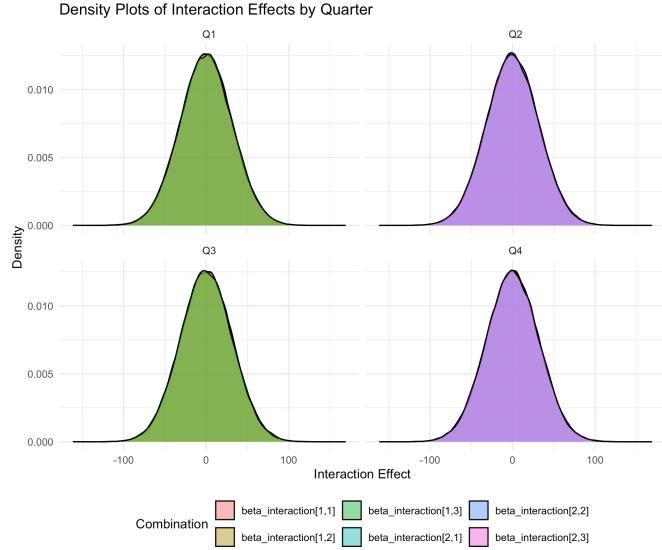


Figure 9: 模型一中參照不同季度與產品組合之結果。

### 5.3.2 修正分佈與參數估計目標與其結果

首先，我們將模型中的季度，改為單一參數（不分成四個季度進行參數估計），以形成模型二的結果，並再次觀察其後驗分佈的狀態是否正常。如 Figure 10 所示，表現如同分開估計一般，類似且完整亦不重複提供相同的 Likelihood 值。

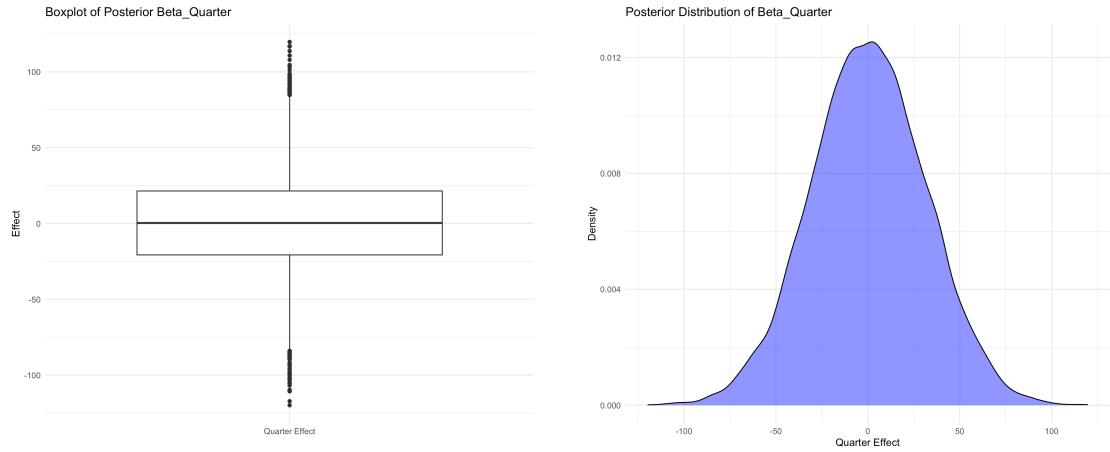


Figure 10: 模型二中季度參數估計分佈。

而減少重複且貢獻不穩的季度參數後，我們再回到原本對於模型一的操作，並環繞著我們的研究目的進行分析和探討。

Table 5: 原本的款式參數估計

Style	MLE	CI_Lower	CI_Upper
beta_style[1]	-0.217	-62.1	62.6
beta_style[2]	0.160	-62.6	62.5
beta_style[3]	0.322	-62.0	61.9

從 Figure 11，我們可以開始看出每個款式下的後驗分佈有所差異，表現上仍呈現 Normal 的狀態，但實際確實有不同的貢獻和對銷售總額上的影響，我們亦可從 Table 5 9，得到實際的

Table 6: 模型後的款式參數估計

Style	MLE	CI_Lower	CI_Upper
beta_style[1]	106.00	48.8	163.0
beta_style[2]	-1.20	-55.2	54.7
beta_style[3]	75.10	15.4	135.0

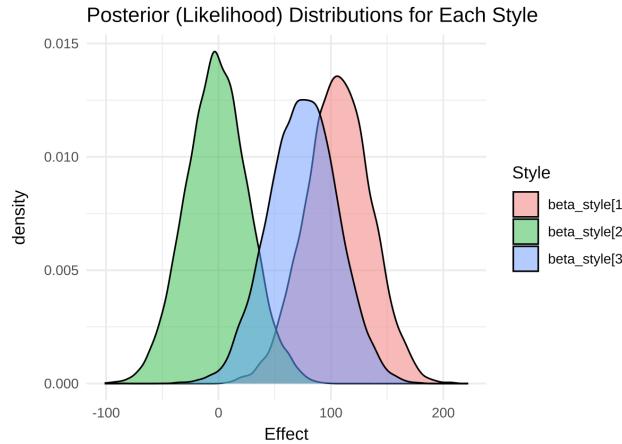


Figure 11: 模型二中款式的後驗與 Effect 狀態。

數值，發現季度因子的估計減少，確實擴大了款式下彼此的差異。

Table 7: 原本的品項定位參數估計

Product	MLE	CI_Lower	CI_Upper
beta_product[1]	0.147	-61.7	61.2
beta_product[2]	0.0489	-62.0	62.6

Table 8: 模型後的品項定位參數估計

Product	MLE	CI_Lower	CI_Upper
beta_product[1]	112.00	53.5	172.0
beta_product[2]	68.70	14.1	121.0

另外，我們一樣關心在產品組合下的另一個因子，品項定位，而從 Figure 23，我們可以開始看出不同品項定位下的後驗分佈有部分的差異，重疊的狀態仍然不少，而這似乎亦可推斷與我們將特定性別分為一組的緣故導致，但實際確實有不同的貢獻和對銷售總額上的影響，我們亦可從 Table 10 11，得到實際的數值，發現季度因子的估計減少，確實整體放大了品項定位的估計。

這項模型的結果，也間界對我們先前的疑問與推估解答，其季度因子事實上，在被貝氏估計上，若不直接性地影響我們所期待的售出狀況，而我們將多餘且無辜的參數存在在模型當中，定會直接性地影響模型的成效與估計。

#### 5.4 模型之地區效應問題

回到我們原始的模型中，對資料直觀而言我們可以發現地區的劃分會有明顯的差異（一般縣市和直轄市）那在先驗分佈的參數設定上，是不是可以給予不同的分佈再做分析呢？

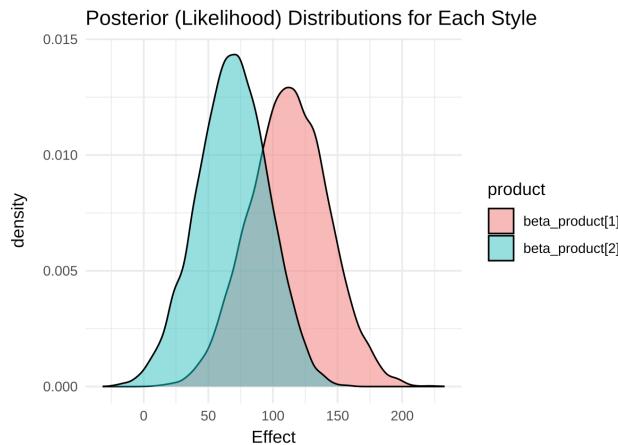


Figure 12: 模型二中品項定位的後驗與 Effect 狀態。

#### 5.4.1 原始模型地區估計結果

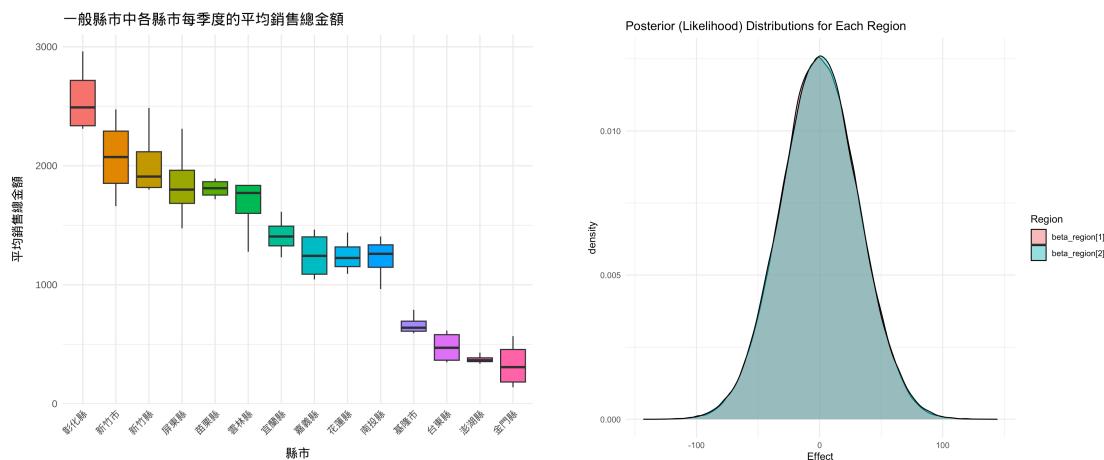


Figure 13: 左圖為一般縣市原始的銷售狀況，右圖為原始模型地區因子的後驗 Effect 狀態。

在 Figure 24我們可以從原始數據的圖中先證明了，一般縣市內定然存在明顯的差異；而在原始模型（即模型一）中，亦將結果以相同的推測呈現。

#### 5.4.2 修正先驗參數與分組目標之結果

接著，為了達成與在貝氏模型中破解前述的問題與概念，我們在先驗分佈中就將地區的參數估計做調整，使其已經先有差異化，再來觀察結果會不會有明顯的差異。

而我們執行完新的模型後，如 Figure 25所示；我們可得我們得假設與對原始資料的觀察結果為真，由於我們對先驗分佈做了些微的參數修正，使其在一般縣市接受比較大的變異性，藉此體現一般縣市內與直轄市內本身的差異，而結果如我們猜想般，確實呈現了分組且不同的現象；故然，我們亦可透過這項模型做後續的參數估計，然而，我們認為事實上他的顯著，同時亦造就了他對我們目標的銷售總金額（同時指產品組合）並不是主要緣故，而是各地分店、工作類型與客群等間接影響該因子的狀態。因此，我們將帶著這些結果再次回到原始的模型當中。

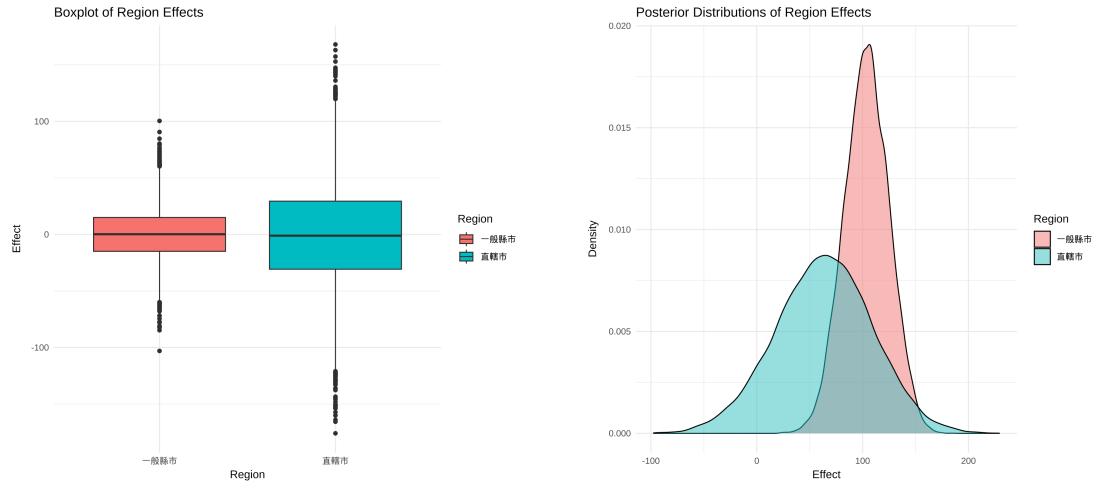


Figure 14: 模型三的地區後驗與 Effect 結果。

## 5.5 產品組合分析

回到一開始的模型當中，可以發現，事實上季度在組合之間的效應並沒有直接性的影響，而是整體偏向穩定，若有特定的組合劇烈波動，甚至我們可以歸類於寶雅門市對於產品與市場的活動規劃（即模型中提及的常態下的極端結果）。

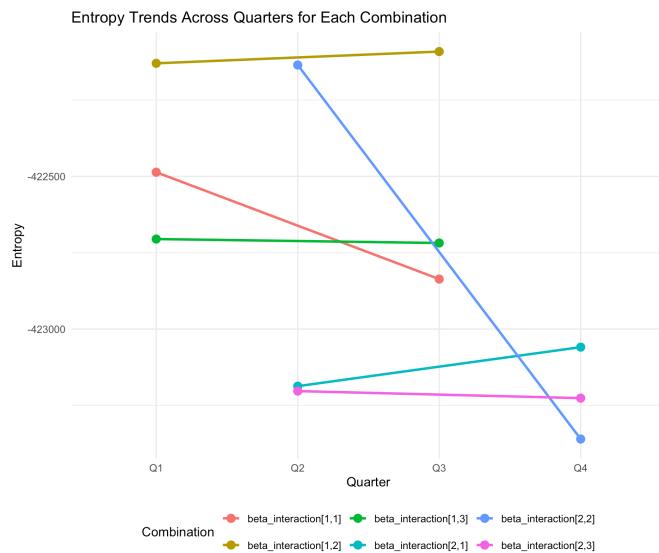


Figure 15: Dummy Caption A.

至此，我們仍還沒獲取我們所關注的研究目標；然而，事實上轉向站於消費者的角度思考，今日類似相同的產品陳列於寶雅當中，不論款式是否相同，亦不論品項定位的定奪，進而事實上在乎的僅是實用性與性價比，故然，弱我們將轉向對於目標與參數們的關係，加入新的參數做分組的估計，或許可以得到不同的結果。

## 5.6 進行整體模型的分組嘗試

我們根據先前的想法修正了我們的模型，從 Figure 16 可以發現確實有分組的現象，而我們在針對結果進一步的處理與分類。從 Figure 17 可以看出各個產品組合下確實具有不同分組的結果。

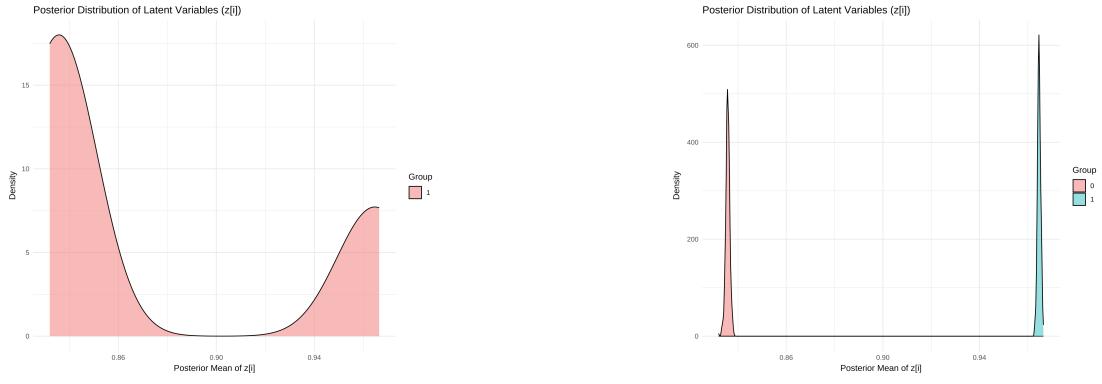


Figure 16: 模型四的後驗參數分組狀態。

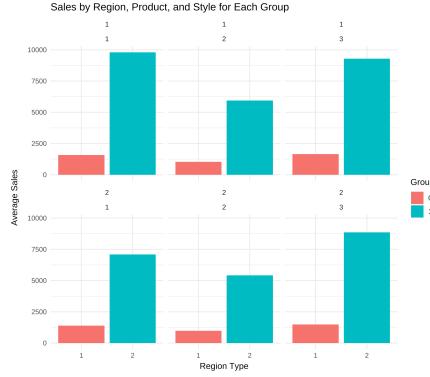


Figure 17: 各個產品組合的分組結果

而其結果為何，我們可以從 Figure 18 中得知，事實上價格得高低決定了組別，但即使價格較高的結果，不論何種產品組合在各地區，都具有一定的購買估計結果；而我們猜想這與該廠商在市場中的定位則有關係，亦或是賣雅的客群限制。而我們更信於前者的猜想，如前述曾說名相妨，該廠商成立於市場中已有近二十餘年，於品牌或品質上具有一定的口碑或回流客戶。

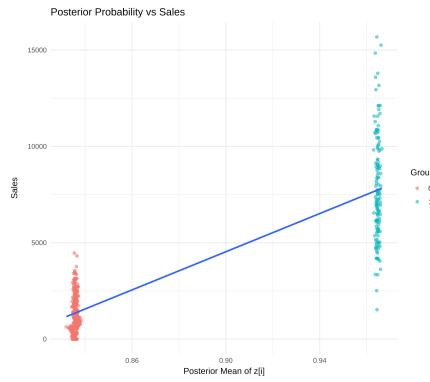


Figure 18: 分組結果 (Y 軸為銷售總金額)

最後，我們仍然要讓數字說話來論證我們的觀點，因此我們引用先前所描述的論文提及的模型方式，並融入後形成模型五進行分析。

## 5.7 改進模型，引入階層（無母數）分群

鑑於階層模型的概念更為複雜，因此我們進行了同樣模型兩次的結果，而其中我們採用課堂中所提及的抽樣概念。



Figure 19: 迭代次數少，無抽樣的結果。

我們可以從 Figure 19 和 Figure 21 中看出，事實上迭代次數差距十倍的狀態是稍微較擬合的程度；而我們每十次進行抽取一次的方式（即程式碼中的 `thin = 10`）使得結果差距可以更加明顯與提升，並同時減少原本隨機抽取全部資料時，可能會有兩兩結果高相關的問題，藉此盡可能實現階層分析的結果。

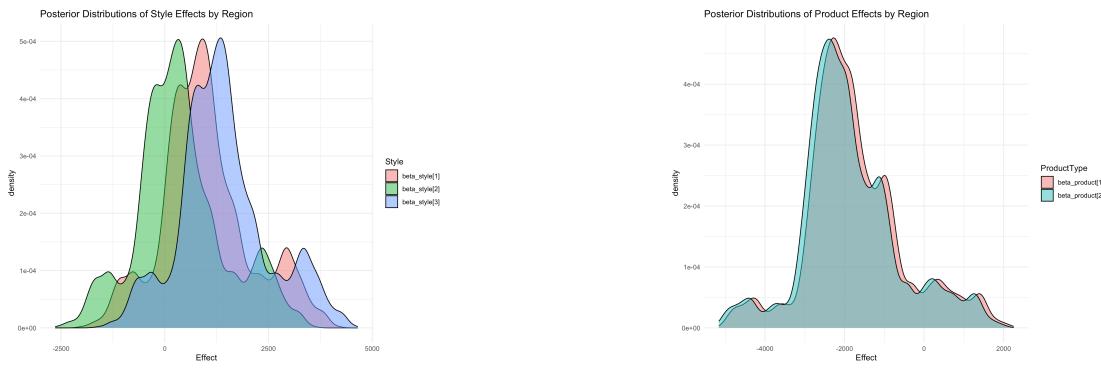


Figure 20: 階層後的分群狀態



Figure 21: 迭代次數多，有抽樣的結果。

從 Figure 20 和 Figure 22 中，亦可得到相對應的結果，我們於抽樣的結果中撇除了高相依性的問題；藉此發現，款式在兩圖中表現相似，確實有一定的差異性存在，這也與我們先前的多次分析下的結果相同；而品項定位，在迭代多次中也獲得了與我們先前的討論結果相同的結論，鑑於實驗上的分組將特兩種特定性別歸屬在一起，造成重疊性高的結果，不論在階層（無母數）或一般分群上，應當難以分出大多重疊現象之參數的分組結果。

無論如何，這項嘗試讓我們發現，在階層中的分析省略了我們前述大多的推測與繁瑣的反覆分析結果，除了對於銷售總額的分群想法為反覆分析下，所間接啟發外；事實上，我們只需要簡單地帶入一季乃至一個月的資料，就可以對市場與寶雅的各項定位做產品上的初步分群探討。

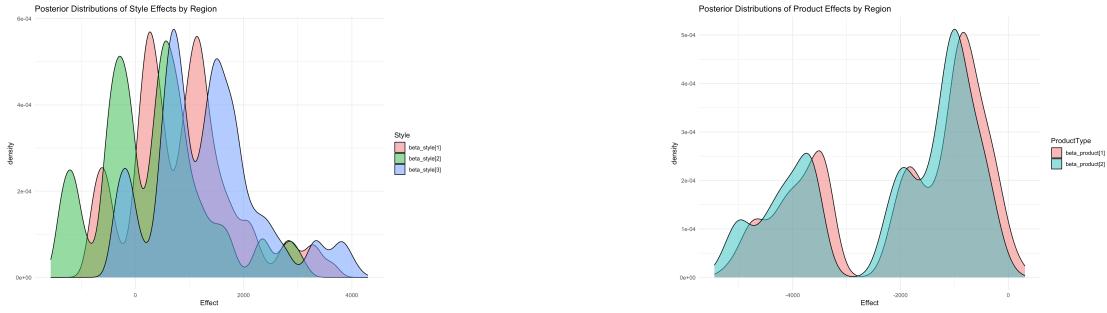


Figure 22: 階層後的分群狀態

## 5.8 變異數分析與其交互作用

已針對款式因子做過 Mix Model 的建立，發現其中款式因子顯著，代表款式因子所遵照的分佈之各個 Level 的變異數不相同。因此，將其改為 Fixed effect，並建立為 Fixed Model 如下，繼續研究與說明。

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} + \\ (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\alpha\gamma\delta)_{ikl} + (\beta\gamma\delta)_{jkl} + (\alpha\beta\gamma\delta)_{ijkl} + \epsilon_{ijklm}$$

- $y_{ijklm}$ ：第  $i j k l$  組的第  $m$  次測量的銷售總金額。
- $\mu$ ：整體平均效應。
- $\alpha_i$ ：季度 Fixed effect。 $(i = 1, 2, 3, 4$ ，對應第一季、第二季、第三季、第四季)
- $\beta_j$ ：品項定位 Fixed effect。 $(j = 1, 2$ ，對應特定性別、中性)
- $\gamma_k$ ：款式 Fixed effect。 $(k = 1, 2, 3$ ，對應自動扣、一般款、針棒)
- $\delta_l$ ：地區 Fixed effect。 $(l = 1, 2$ ，對應直轄市、一般縣市)
- $(\alpha\beta)_{ij}$ ：季度與品項定位的雙因子交互作用。
- $(\alpha\gamma)_{ik}$ ：季度與款式的雙因子交互作用。
- $(\alpha\delta)_{il}$ ：季度與地區的雙因子交互作用。
- $(\beta\gamma)_{jk}$ ：品項定位與款式的雙因子交互作用。
- $(\beta\delta)_{jl}$ ：品項定位與地區的雙因子交互作用。
- $(\gamma\delta)_{kl}$ ：款式與地區的雙因子交互作用。
- $(\alpha\beta\gamma)_{ijk}$ ：季度、品項定位與款式的三因子交互作用。
- $(\alpha\beta\delta)_{ijl}$ ：季度、品項地位與地區的三因子交互作用。
- $(\alpha\gamma\delta)_{ikl}$ ：季度、款式與地區的三因子交互作用。
- $(\beta\gamma\delta)_{jkl}$ ：品項定位、款式與地區的三因子交互作用。
- $(\alpha\beta\gamma\delta)_{ijkl}$ ：四因子交互作用。

- $\epsilon_{ijklm}$  : Experimental error  $\begin{cases} l = 1 (\text{直轄市}), m = 1, 2, \dots, 6 \\ l = 2 (\text{一般縣市}), m = 1, 2, \dots, 14 \end{cases}$  。 Assume  $\epsilon \sim N(0, \sigma^2)$

從 Table 9 中，我們可以得知 Fixed Model 的分析結果。

Table 9: ANOVA Table with 季度 \* 品項定位 \* 款式 \* 地區

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
季度	3	10,009,792	3,336,597	1.7620	0.1537213
地區	1	4,105,056,538	4,105,056,538	2167.8003	$< 2.2 \times 10^{-16}$ ***
品項定位	1	26,178,704	26,178,704	13.8245	0.0002271 ***
款式	2	180,055,356	90,027,678	47.5419	$< 2.2 \times 10^{-16}$ ***
季度: 地區	3	12,598,370	4,199,457	2.2177	0.0854288 .
季度: 品項定位	3	9,193,615	3,064,538	1.6183	0.1844055
地區: 品項定位	1	29,347,908	29,347,908	15.4981	0.00009624 ***
季度: 款式	6	6,394,010	1,065,668	0.5628	0.7599973
地區: 款式	2	152,087,784	76,043,892	40.1573	$< 2.2 \times 10^{-16}$ ***
品項定位: 款式	2	14,063,044	7,031,522	3.7132	0.0251813 *
季度: 地區: 品項定位	3	13,591,644	4,530,548	2.3925	0.0679663 .
季度: 地區: 款式	6	5,853,474	975,579	0.5152	0.7969179
季度: 品項定位: 款式	6	4,064,062	677,344	0.3577	0.9053269
地區: 品項定位: 款式	2	25,648,217	12,824,109	6.7722	0.0012707 **
季度: 地區: 品項定位: 款式	6	3,478,457	579,743	0.3062	0.9336849
Residuals	432	818,057,083	1,893,651		

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

從上述說明可知，四維交互作用不顯著，因此我們下一步將針對四維 ANOVA 中，近似顯著的「季度: 地區: 品項定位」、「地區: 品項定位: 款式」這兩種三因子交互作用的組合結果進行 ANOVA 分析，不過前者顯著的結果相對薄弱，我們仍將依序觀察與分析探討其因子間的交互作用。

然而，如果是三維的 ANOVA 分析中，三維的交互作用顯著，則要把整個資料集按照某一個因素不同取值拆開，然後在得到的多個「子資料集」中做兩因子 ANOVA，對另兩個因子的二維交互效應進行考察；如果三維交互效應不顯著，則進一步考察整個樣本中三個因素兩兩組合得到的三個二維交互效應的顯著性。同樣，如果二維交互效應顯著，則需要繼續在一個因素不同取值下拆分資料集，分別考慮另一因素的主效應；如果二維交互效應不顯著，則直接考慮整個樣本中單個因素的主效應。

因此，由於我們是四維的 ANOVA 分析，因此我們會先拆成兩組三因子 ANOVA 進行分析，並針對兩者（四因子與三因子）的結果做更近一步的推估與解釋及說明。

Table 10: ANOVA Table with 品項定位 \* 季度 \* 地區

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
品項定位	1	26,178,704	26,178,704	10.0413	0.0016318 **
地區	1	4,105,056,538	4,105,056,538	1574.5589	$< 2.2 \times 10^{-16}$ ***
季度	3	10,009,792	3,336,597	1.2798	0.2806772
品項定位: 地區	1	29,347,908	29,347,908	11.2569	0.0008584 ***
品項定位: 季度	3	9,193,615	3,064,538	1.1755	0.3185894
地區: 季度	3	12,598,370	4,199,457	1.6108	0.1860543
品項定位: 地區: 季度	3	13,591,644	4,530,548	1.7378	0.1584154
Residuals	464	1,209,701,487	2,607,115		

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

鑑於我們需要針對四因子中的三因子做進一步的分析，模型的部分如同 Fixed Model 中的假設，並扣除款式因子及其 Fixed effect，包含其交互作用。因此，我們可以來觀看該模型的三因子 ANOVA 結果，如 Table 10。我們可以發現三因子的部分與四因子的結果有所差距，在本表中三因子的交互作用並不顯著，而其中我們可以再觀察兩因子的交互作用情況。

季度與品項定位相對非常不顯著；季度與地區的部分結果中規中矩，近似於四因子當中，可推測並沒有因為因子的增加或減少有被多或被少的解釋掉部分的內容；最後，品項定位與地區則非常顯著，這個結果如同四因子當中的結果，且包含在另一三因子交互作用之中，並且於四因子的結果亦可得到相同的結果，而接下來我們也將為該三因子交互作用進行分析與說明。

Table 11: ANOVA Table with 品項定位 \* 款式 \* 地區

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
品項定位	1	26,178,704	26,178,704	13.8712	0.0002196 ***
地區	1	4,105,056,538	4,105,056,538	2175.1340	< 2.2e <sup>-16</sup> ***
款式	2	180,055,356	90,027,678	47.7027	< 2.2e <sup>-16</sup> ***
品項定位: 地區	1	29,347,908	29,347,908	15.5505	9.262e <sup>-5</sup> ***
品項定位: 款式	2	14,063,044	7,031,522	3.7258	0.0248122 *
地區: 款式	2	152,087,784	76,043,892	40.2931	< 2.2e <sup>-16</sup> ***
品項定位: 地區: 款式	2	25,648,217	12,824,109	6.7951	0.0012330 **
Residuals	468	883,240,508	1,887,266		

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

接著，我們需要針對四因子中的另一個三因子做進一步的分析。模型的部分如同 Fixed Model 中的假設，並扣除季度因子及其 Fixed effect，包含其交互作用。我們可以來觀看該模型的三因子 ANOVA 結果，如 Table 11。我們可以發現三因子的部分與四因子的結果相似，都具有相對顯著的結果。在本表中不論是三因子的交互作用，還是其之下的兩兩因子的交互作用都相當顯著的情形。

根據以上的結果與分析，我們可以發現「款式」內的銷售情況有所不同外，也可以推估季度在本實驗中能夠被解釋的程度相對較低或薄弱；而回到我們的研究目標之下，我們期待找到產品的最佳組合，且我們仍所關注與在意的是款式的因子，因此我們將依據且針對交互作用與款式的關係再進一步地分析與探討，以達成我們期待的目標與預期。

### 5.8.1 交互作用

接著，仍然環扣我們的實驗目標，期待找到產品的最佳組合，而我們延續本節中的 Fixed Model 結果，觀察我們實驗中的各項因子對於銷售總金額的關係與結果。從中我們已可獲得除了季度外的因子具有交互作用。因此，本小節中我們將針對三因子的交互作用進行繪圖呈現，使其結果更加明確，並為後續的分析進行鋪陳。

從 Figure 23 結果中可觀察出以下的論點。在直轄市下，一般款的中性銷售總金額高於特定性別非常多；而在一般縣市下，一般款的中性銷售總金額與特定性別相差不遠。然而，鑑於該圖表呈現上的限制，我們將其中一因子拆為各個 Level，進行與其他因子間的交互作用呈現，故然會有些許不夠明確或難以解釋之處，並且同於前面小節的分析，我們應當針對其中兩兩因子的交互作用進行觀察，因此，結果如 Figure 24 呈現。

針對兩兩因子的交互作用部分 (Figure 24)，依序我們具有以下的解釋與說明：

- 品項定位與款式：中性的一般款與針棒銷售總金額差距不大，但特定性別的針棒高於一般款不少。

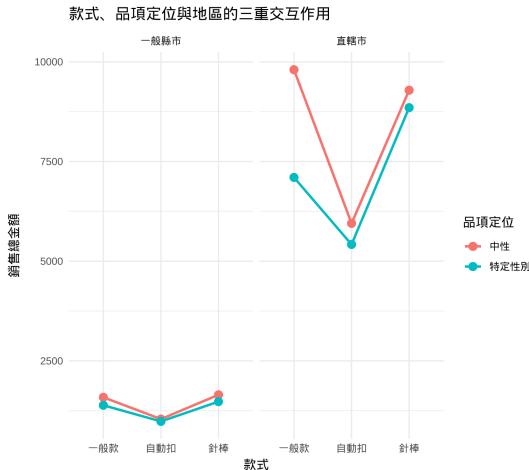


Figure 23: 款式 \* 品項定位 \* 地區之交互作用圖。

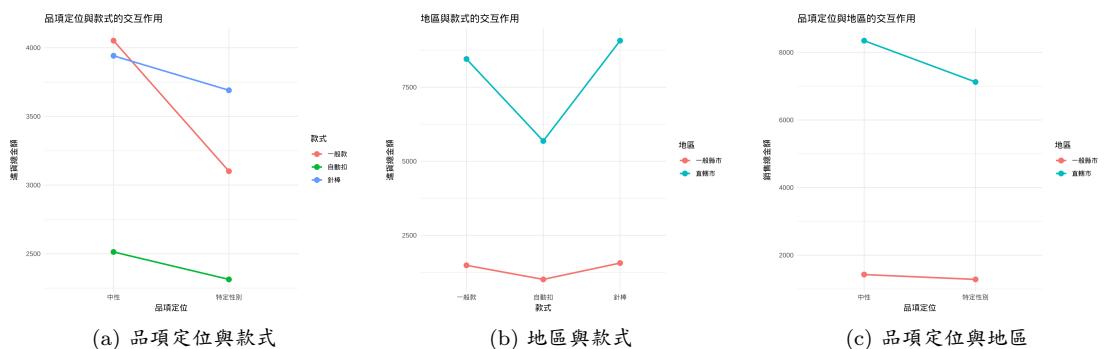


Figure 24: 款式 \* 品項定位 \* 地區，三因子下兩因子交互作用圖。

- 地區與款式：不論在直轄市或一般縣市，自動扣的銷售總金額都是所有款式裡最少，但一般縣市中三者之間差異不大，直轄市中自動扣與另外兩個有明顯落差。
- 品項定位與地區：一般縣市下，中性銷售總金額與特定性別相差不遠；直轄市的中性明顯高於特定性別。

由這些圖可得知，這三個因子之間確實有明顯的交互作用，不過具體到底有沒有顯著差異，還需透過後面的事後檢定驗證。

## 5.9 事後檢定 Tukey-HSD

最後，同樣地回到實驗目標中的說明，其中提及希望「並為廠商及寶雅提出銷售策略」；因此，我們將利用交互作用的結果（同前節所描述），來查看顯著之相關的因子與其交互作用的結論，來檢定其中不同的 Level 比較結果。本研究我們選用檢定嚴謹度適中的 Tukey - HSD 進行事後檢定。

而我們將同時站在廠商或寶雅總公司立場，分別探討直轄市或一般縣市之下，何種「品項定位」或「款式」產品的總銷售額較佳的情況。

Table 12: 直轄市之事後檢定結果。

Contrast	Estimate	SE	df	t.ratio	p.value
中性 - 特定性別	1222	229	432	5.330	<0.0001 ***

Table 13: 直轄市之事後檢定結果。

Contrast	Estimate	SE	df	t.ratio	p.value
一般款 - 自動扣	2767	281	432	9.851	<.0001 ***
一般款 - 針棒	-619	281	432	-2.204	0.2380
自動扣 - 針棒	-3386	281	432	-12.055	<0.001 ***

### 5.9.1 直轄市面向

在直轄市下，我們對品項定位與款式進行事後檢定，結果如 Table 13。在品項定位中具有顯著結果；其中從 Figure 25 (左上) 中可得之，中性的商品銷售總額狀況較佳。而在款式因子當中，僅「一般款」與「針棒」之結果較為不顯著；從 Figure 25 (右上) 中仍可獲得相對的結果，其中針棒跟一般款的銷售總額仍比自動扣佳。

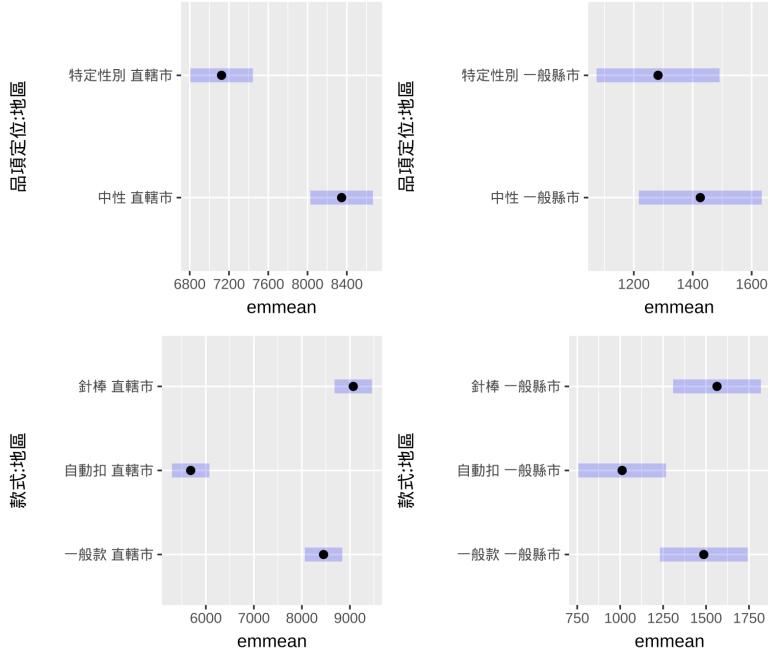


Figure 25: 事後檢定結果圖。

### 5.9.2 一般縣市面向

接著，我們將結果專注於一般縣市之下，檢定結果如 Table 13。在品項定位該因子中，結果不具有顯著結果，無法得知品項定位的差異，亦無法從圖中做出其他的解讀或推測；而在款式當中：僅「針棒」與「自動扣」具顯著結果；同樣地，從 Figure 25 (右下) 中，僅能部分解釋針棒銷售總額明顯大於自動扣。

Table 14: 一般縣市之事後檢定結果。

Contrast	Estimate	SE	df	t.ratio	p.value
中性 - 特定性別	143	150	432	0.955	0.7753

Table 15: 一般縣市之事後檢定結果。

Contrast	Estimate	SE	df	t.ratio	p.value
一般款 - 自動扣	475	184	432	2.584	0.1034
一般款 - 針棒	-77	184	432	-0.419	0.9984
自動扣 - 針棒	-552	184	432	-3.003	0.0335 *

## 6 結論

我們前節的分析中，我們藉由 Contrast 款式的部分 (Table 16)，獲得了以下的結論：

- 「針棒」的銷售總金額明顯與「外觀相似的其他組」存在顯著差異。

我們藉由變異數的相關分析，如交互作用與其事後分析等，我們可以獲得更近一步的結果，並且達成我們對本研究的期許與目標。其中我們將事後檢定的結果簡易地整合在 Table 16。

並且我們為了對廠商及寶雅提出銷售策略，我們將分別為其提供不同面向的剖析與建議，並給予相對應的結論說明，期待這場實驗可以帶給不同對象的正向效應，同時能夠改變管理或營運上的瓶頸或困難，以提升各地區的總銷售額達到雙贏的未來與結果。

Table 16: 事後檢定之表格結果。

變數	款式			品項定位	
	一般款	自動扣	針棒	特殊性別	中性
直轄市			V		V
一般縣市			V		

### 6.1 寶雅面向

對於寶雅而言，物流成本與銷售額的狀況是主要關注的內容。首先，鑑於直轄市的針棒和中性銷售狀況較佳且顯著，建議可以針對直轄市地區的倉庫或店面增加這類型的產品；其次，由於直轄市與一般縣市的分店數有落差，針對一般縣市仍可以相對顯著的針棒款式進行進貨與販售，而品項則不限制；另外，由於本實驗中地區分法的限制，而從 Nested Model 的結果亦可得知，事實上，縣市的產業分布與寶雅本身商店的定位，亦會對於皮帶該產品的銷量具有一定層面的直接或間接性影響。

另外，如同第一節中對於寶雅進貨方式的描述，可知其管理模式與資訊系統的演算法緣故，無法對於特定地區補齊特定產品，各分店僅能針對已被售出過的產品進行進貨。因此，我們僅從此角度給予不同的回饋與建議。

### 6.2 廠商面向

對於廠商而言，壓低產品的成本與提高售出減少退貨機率（即提升迴轉率）為關鍵。同時，在台灣中由於量販店中的管理模式，與社會風俗息息相關之緣故，注重顧客的感受。因此，管理模

式與寶雅之資訊系統的演算法緣故，無法對於特定地區增加特定產品，將會是廠商的倉儲或中盤商上管理的瓶頸點。

因此，首先，為了讓廠商能夠持續延長或創新產品的生命力，根據我們的結果，亦建議廠商可以針棒款式的皮帶為未來販售或製作的主要產品，增加數量或調整製造比例，以降低成本。而由於直轄市仍為寶雅大宗的分店位置，廠商可以中性產品為主進行設計或主要販售，以同時滿足直轄市與一般縣市的販售需求（中性商品仍可低度滿足購買特定性別產品的顧客）；另外，考量到縣市的產業分布，本處建議廠商以更寬廣的角度思考，可將台灣未來產業發展作為切入點，部分改變或影響縣市產業分布與整體數量，進而影響皮帶產品的銷量結果。（本處將皮帶視為非流行性產品，或將皮帶之款式或品項定位的流行定義為週期性流行，不受此時代潮流影響）

### 6.3 貝氏分析之總結

本研究中的模型與貝氏分析之結果，雖然沒辦法像變異數分析的結論如此清晰可見，甚至給予產品上相當實質的建議；然而，我們事實上可以從中發現一些神奇的環節，在我們研究資料介紹該節中，便間接描述寶雅的管理模式與其演算法的概念，既使廠商有心欲改變現況推陳出新，為各地區可謂客製化的概念進行產品的調整並提升銷售總額，實際上卻會被零售業端的演算法計算模式，推估為不適合或是不需補貨的結論。故然，我們可以發現他便是以頻率學派的觀點進行分析，利用過去的資料直接計算相對應的數值（如回歸分析概念），僅能在特定的值域範圍中進行推測和說明，倘若期待加入不同的花樣或改變現況時，便會被計算為超出 Residual 可負荷的範圍內，被排除。

而這樣的現象正為台灣產業市場中的弊端，我們不在給予廠商乃至工廠端不同的機會與調整，而是以零售業甚至消費者至上的觀點，管理公司物流與以及垂直與水平整合的倉儲及供應鏈；然而，在本研究貝氏學派中的階層貝氏分析，卻間接地打破了這樣的盲點，他可以獲得相似於頻率學派的結果，畢竟我們仍有在事後分析的環節中存在不顯著的部分無法解釋（即使我們做了一整年的實驗，仍無法使其顯著），卻可以利用相當短的時間成本獲得分組的結果且並且對應了款式與品項定位的狀態，甚至亦判斷出季度並不顯著的相對結果。

事實上，我們可以優雅地用這簡單的貝氏估計，得到與變異數分析中相似的答案，並且給予對應的商業策略，如我們亦可從品項定位的重疊結果，推薦寶雅或廠商將品項定位往中性發展；我們亦可從款式的因子在分群的結果中，推薦廠商可以重新審視自動扣所熱賣的縣市或地區，因為分群結果下販售狀況不算均勻等。

最後，我們其實可以藉由貝氏估計的先驗分佈與 Likelihood 的估計，節省大量的時間成本來滿足現今已顧客至上，且以無母數的快速分組，使其相對即時的方式來為每一季乃至每一個月進行分組推估；而在實現這一切的前提為市場與社會氛圍願意為這些，雖有一定的品牌與品質的台灣在地製造的「夕陽產業」，接納與打造起「再昇產業」。

## 7 References

林哲遠 (2015)。基於中國餐廳過程之貝氏探索分群法。交通大學資訊科學與工程研究所學位論文。國立交通大學。

邱逸彥、王晉元等人 (2013)。利用貝氏理論於旅行時間推估之研究。博士論文。

Kuehl, R. O. (2000). *\*Design of experiments: Statistical principles of research design and analysis\**.