

國立交通大學

統計學研究所

碩士論文

對稱型分組平均及分組平均方法之比較分析

**Symmetric Categorized Mean**

and

**Comparison of Categorized Mean Methods**

1896

研究生：王兆慶

指導教授：陳鄰安 教授

中華民國一〇三年六月

# 對稱型分組平均及分組平均方法之比較分析

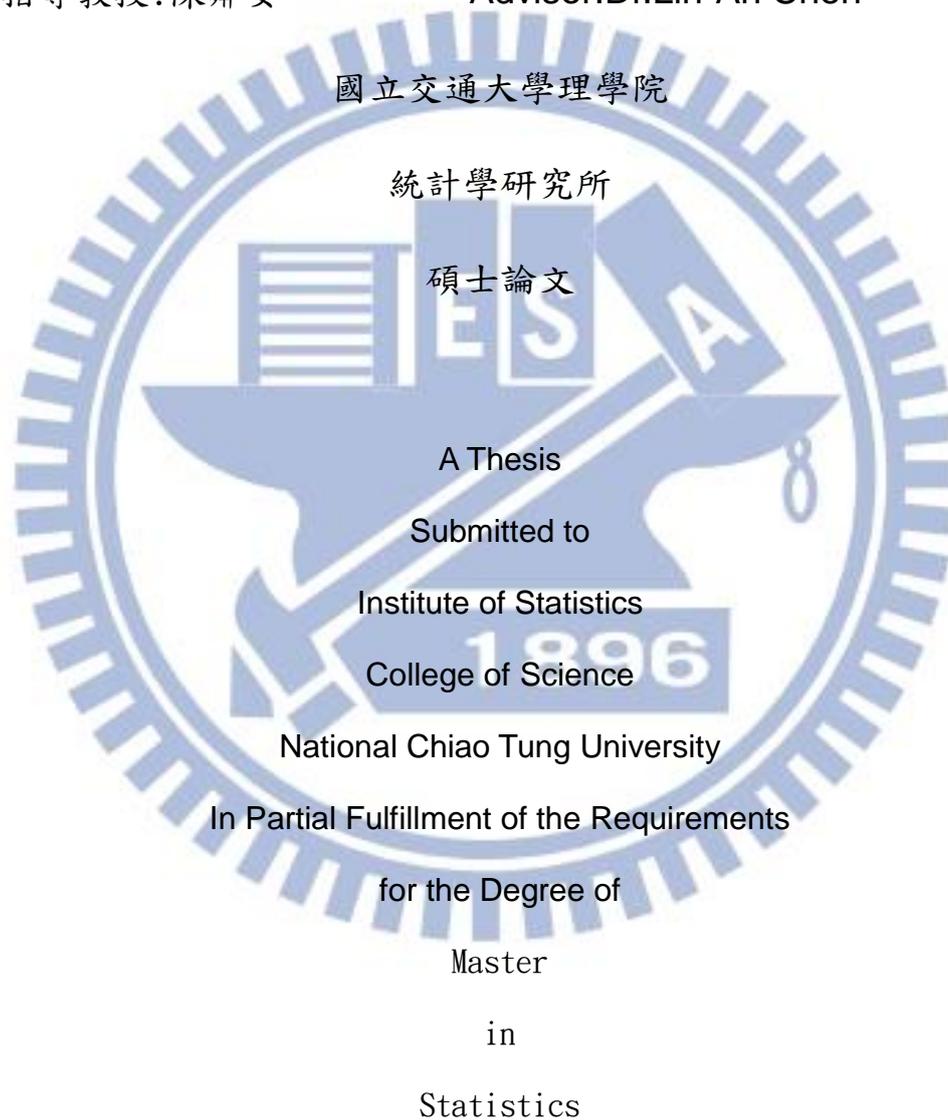
## Symmetric Categorized Mean and Comparison of Categorized Mean Methods

研究生:王兆慶

Student:Zhao-Ching Wang

指導教授:陳鄰安

Advisor:Dr.Lin-An Chen



June 2014

Hsinchu, Taiwan, Republic of China

中華民國一零三年六月

# 摘要

在流行病學中經常為了方便分析與解釋而把連續變數資料予以分組。此種以分組後的組資料所做統計分析近年來受到許多學者的批評及爭論。為了要修正那不被完全信任的統計方法，我們將會探討無母數對稱型分類變數在未知平均數下的統計效應。我們研究出無母數對稱型分類變數方法的大樣本理論以及其近似分配。同時，我們也將此新型的分類方法和有母數、無母數的情況做比對。此三種分類方法將會一一論述。



# 誌謝

在這篇論文產生之前，一路上受到很多人的幫助。首先是指導教授所給予的方向，讓我能夠思考在這個問題上能夠做出何種改善。同時也提出很多不一樣的想法，在分析上面能有更多的理解。還記得在某一次的模擬中，從看似沒有線索結果中找出值得深加探討的議題。另外也要感謝鄭宇傑學長提供一些在計算上程式碼的寫法要如何能更加有效率，並且在一些複雜的積分估計上所提供的建議。此外我也要感謝我的朋友們，在我遇到困難的時候能夠給予我鼓勵，並且分享一些他們自己的論文所遇到的難題，好讓我在這塊領域上能夠有所發揮。



# 目錄

緒論.....	p.1~p.2
無母數分組平均數.....	p.3~p.5
建立在分類變數之上的對稱型百分位數.....	p.6~p.8
模擬.....	p.9~p.12
資料分析.....	p.13~p.17
附錄資料.....	p.18~p.20
參考文獻.....	p.21~p.22





# 1. 緒論

研究學者們經常需要評斷連續型變數和解釋變數之間是否存在某種關係。在現代流行病學的做法，將連續型變數予以分類分組以便於資料呈現；例如：低、中、高危險的族群，再以分組資料做類似組平均是否相等之分析。Prock et al(2004) 的研究指出有百分之八十四出自前端科學期刊的流行病學研究都將連續變數轉變為分組變數再做統計分析。而將連續型變數轉變為分組變數的手法也遍布其他領域；例如：應用心理學(MacCallum et al (2002))、以及行銷應用 (Irwin and McClelland (2003))。研究者經常將分組資料再做迴歸分析以建立迴歸曲(直)線或者了解是否解釋變數值落在較高群組中的人是否伴隨較高危險(風險)之結果。

連續型變數的分類法一直飽受批評是源自於使用不受歡迎的統計方法：它被證明具有偏誤的估計以及因遺失訊息而導致較低的檢定力(詳見例子 Royston and Sauerbrei (2008) Taylor and Yu(2002) Walraven and Hart(2008) and Zhao and Kolonel(1992) )。

根據 Han(2008) 的研究發現，對於這些分類變數所做的多重比較，其背後的多個假設：常態性、獨立性以及常數變異數，均不正確(詳見 Bennette and Vickers(2012))。而那些不為樂見的統計性質如有偏誤的估計及過低的檢定力等，我們認為如果繼續用不被認可的統計推論方法來分析分組資料，則那些偏誤估計及低檢定力等缺點是無法避免的。

由於把連續變數資料分組再行分析相關性仍是普遍存在於學界。因此建立分組資料的正確分析方法是一刻不容緩的研究工作。黃子芸(2013)首先提出有母數及無母數分組平均數的統計分配理論並由此理論建立用於區間估計及假設檢定的正確方法。在 Chen and Chang(1996)所介紹的對稱型分位數(symmetrical)

quantile)在統計估計理論上經常會有不錯的估計結果。我們想利用此一分位數來建立分組平均數並研究其理論性質並建立統計推論方法。

由於黃子芸(2013)的研究中並未對無母數分組平均的大樣本分配之變異矩陣提出估計，我們將做此努力並更加廣泛地研究無母數、有母數及對稱型三種方法並進行模擬分析。

這篇論文第二章我們重述黃子芸(2013)的無母數分組平均的分配理論。我們也將建立其近似分配之參數(平均及變異矩陣)估計並與我們的新方法做比較分析。第三章將介紹 Chen and Chang(1996)基本對稱平均的觀念並由其擴展至分組對稱平均數之方法並推導其大樣本分配。模擬分析及資料分析的結果將呈現於本文中。



## 2. 無母數分組平均數

假設  $\mathbf{Y}$  與  $\mathbf{X}$  為從一個體(個人)所量出之反應變數(response variable)及解釋變數(explanatory variable)。在很多的研究中 (Shankar et al(2007), Luo et al(2007) 和 Letenneur et al (2007))，是由變數  $\mathbf{X}$  的百分位數建立  $k$  個區間再把變數  $\mathbf{Y}$  分割成  $k$  個區間，其  $\mathbf{X}$  的  $k$  區間為：

$$A_1 = (-\infty, F_X^{-1}(\alpha_1)], A_2 = (F_X^{-1}(\alpha_1), F_X^{-1}(\alpha_2)], \dots, A_k = (F_X^{-1}(\alpha_{k-1}), \infty)$$

而式子中  $F_X^{-1}(\alpha)$  為隨機變數  $\mathbf{X}$  的第  $\alpha$  母體百分位數。通常， $F_X^{-1}(\alpha_j)$  會是未知的。假設我們有隨機樣本  $\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix}, \dots, \begin{pmatrix} Y_n \\ X_n \end{pmatrix}$  且令  $\hat{F}_X^{-1}(\alpha)$  為  $\mathbf{X}$  之經驗百分位數，則上述  $\mathbf{X}$  之  $k$  區間可以估計，如下：

$$\hat{A}_1 = (-\infty, \hat{F}_X^{-1}(\alpha_1)], \hat{A}_2 = (\hat{F}_X^{-1}(\alpha_1), \hat{F}_X^{-1}(\alpha_2)], \dots, \hat{A}_k = (\hat{F}_X^{-1}(\alpha_{k-1}), \infty)。$$

因此我們會得到由無母數分類樣本所產生的平均：

$$\hat{\theta}_{qj} = \frac{\sum_{i=1}^n Y_i I(X_i \in \hat{A}_j)}{\sum_{i=1}^n I(X_i \in \hat{A}_j)}$$

其中  $q$  代表由百分位來當作截點。我們使用  $\hat{\boldsymbol{\theta}}_q = (\hat{\theta}_{q1}, \dots, \hat{\theta}_{qk})'$  以及  $\boldsymbol{\lambda}_j = \mathbf{E}(\mathbf{Y} - \mathbf{u}_y | \mathbf{X} = F_X^{-1}(\alpha_j))$

以下將介紹分類樣本平均向量  $\hat{\boldsymbol{\theta}}_j$  的近似分配以及表現方式。(黃子芸(2013))

**定理 2.1**

(a) 百分位截點根據分類樣本平均將會有以下的 Bahadur 分配：

$$\sqrt{n}(\hat{\theta}_q - \theta) = n^{-1/2} \sum_{i=1}^n \begin{pmatrix} \alpha_1^{-1}(\Psi_1(Y_i, X_i) - E(\Psi_1(Y, X))) \\ (\alpha_2 - \alpha_1)^{-1}(\Psi_2(Y_i, X_i) - E(\Psi_2(Y, X))) \\ \vdots \\ (1 - \alpha_{k-1})^{-1}(\Psi_k(Y_i, X_i) - E(\Psi_k(Y, X))) \end{pmatrix} + O_p(1)$$

其中

$$\Psi_1(Y, X) = \begin{cases} Y - \mu_y, & \text{if } X \leq F_X^{-1}(\alpha_1) \\ \lambda_1, & \text{if } X > F_X^{-1}(\alpha_1) \end{cases} \quad \Psi_k(Y, X) = \begin{cases} \lambda_{k-1}, & \text{if } X \leq F_X^{-1}(\alpha_{k-1}) \\ Y - \mu_y, & \text{if } X > F_X^{-1}(\alpha_{k-1}) \end{cases}$$

當  $j = 2 \dots k-1$  時：

$$\Psi_j(Y, X) = \begin{cases} \lambda_{j-1}, & \text{if } X \leq F_X^{-1}(\alpha_{j-1}) \\ Y - \mu_y, & \text{if } F_X^{-1}(\alpha_{j-1}) < X < F_X^{-1}(\alpha_j) \\ \lambda_j, & \text{if } X \geq F_X^{-1}(\alpha_j) \end{cases}$$

(b) 我們得到  $\sqrt{n}(\hat{\theta}_q - \theta)$  會是一個近似常態分配  $N_k(\mathbf{0}_k, \Sigma_q)$

其中  $\Sigma_q = (\sigma_{jm})$   $j, m = 1 \dots k$ ，並且：

$$\sigma_{11} = \frac{1}{\alpha_1^2} \{M_1 + (1 - \alpha_1)\lambda_1^2 - (m_1 + (1 - \alpha_1)\lambda_1^2)\}$$

$$\sigma_{1j} = \frac{1}{\alpha_1(\alpha_j - \alpha_{j-1})} \{ \lambda_1 \lambda_{j-1} (\alpha_{j-1} - \alpha_1) + \lambda_1 m_j + \lambda_1 \lambda_j (1 - \alpha_j) + \lambda_{j-1} m_1 - (m_1 + (1 - \alpha_1)\lambda_1)(\alpha_{j-1} \lambda_{j-1} + (1 - \alpha_j)\lambda_j + m_j) \}$$

$$\sigma_{1k} = \frac{1}{\alpha_1(1 - \alpha_{k-1})} \{ \lambda_1 \lambda_{k-1} (\alpha_{k-1} - \alpha_1) + \lambda_1 m_k + \lambda_{k-1} m_1 - (m_1 + (1 - \alpha_1)\lambda_1)(\alpha_{k-1} \lambda_{k-1} + m_k) \}$$

當  $j = 2 \dots k-1$

$$\sigma_{jj} = \frac{1}{(\alpha_j - \alpha_{j-1})^2} \{ \alpha_{j-1} \lambda_{j-1}^2 + (1 - \alpha_j)\lambda_j^2 + M_j - (\alpha_{j-1} \lambda_{j-1} + (1 - \alpha_j)\lambda_j + m_j)^2 \}$$

$$\sigma_{jj+1} = \frac{1}{(\alpha_{j+1} - \alpha_j)(\alpha_j - \alpha_{j-1})} \{ \lambda_j (\alpha_{j-1} \lambda_{j-1} + m_j + m_{j+1} + \lambda_{j+1}(1 - \alpha_{j+1})) - (\alpha_{j-1} \lambda_{j-1} + (1 - \alpha_j)\lambda_j + m_j)(\alpha_j \lambda_j + (1 - \alpha_{j+1})\lambda_{j+1} + m_{j+1}) \}$$

$$\sigma_j = \frac{1}{(\alpha_j - \alpha_{j-1})(\alpha_m - \alpha_{m-1})} \{ \lambda_{m-1} (\alpha_{j-1} \lambda_{j-1} + m_j + \lambda_j (\alpha_{m-1} - \alpha_j)) + \lambda_j (m_m + \lambda_m (1 - \alpha_m)) - (\alpha_{j-1} \lambda_{j-1} + (1 - \alpha_j) \lambda_j + m_j) (\alpha_{m-1} \lambda_{m-1} + (1 - \alpha_m) \lambda_m + m_m) \}$$

$$m = j + 2 \cdots k - 1$$

$$\sigma_{jk} = \frac{1}{(\alpha_j - \alpha_{j-1})(1 - \alpha_{k-1})} \{ \lambda_{k-1} (\alpha_{j-1} \lambda_{j-1} + m_j + \lambda_j (\alpha_{k-1} - \alpha_j)) + \lambda_j m_k - (\alpha_{j-1} \lambda_{j-1} + (1 - \alpha_j) \lambda_j + m_j) (\alpha_{k-1} \lambda_{k-1} + m_k) \}$$

$$\sigma_{kk} = \frac{1}{(1 - \alpha_{k-1})^2} \{ \alpha_{k-1} \lambda_{k-1}^2 + M_k - (\alpha_{k-1} \lambda_{k-1} + m_k)^2 \}$$

其中  $\mathbf{m}_j = \mathbf{E}((\mathbf{Y} - \mathbf{u}_y)\mathbf{I}(\mathbf{X} \in \mathbf{A}_j))$  ;  $\mathbf{M}_j = \mathbf{E}((\mathbf{Y} - \mathbf{u}_y)^2 \mathbf{I}(\mathbf{X} \in \mathbf{A}_j))$  ,  $j = 1 \dots k$  分別為第一、第二中央組的動差。

上面理論敘述於黃子芸(2013)。但是此一複雜近似變異矩陣如何估計以及所建造的檢定統計量效力如何均未進一步研究。我們將在這一方面做一點努力。

### 3. 建立在分類變數之上的對稱型 百分位數

令  $\mathbf{Y}$  和  $\mathbf{X}$  為兩隨機變數。定義對稱型百分位數我們需要一個中心點  $\mathbf{u}_x$ ，並且定義摺疊累積分配函數  $\tilde{F}_x(\mathbf{a}) = \mathbf{P}(|\mathbf{X} - \mathbf{u}_x| \leq \mathbf{a})$ ， $\mathbf{a} > 0$ 。我們可以得到對稱型百分位數  $\{\mathbf{u}_x - \tilde{F}_x^{-1}(\gamma), \mathbf{u}_x + \tilde{F}_x^{-1}(\gamma)\}$ ，其中  $\tilde{F}_x^{-1}(\gamma)$  為摺疊累積分配函數的第  $\gamma$  百分位數：

$$\tilde{F}_x^{-1}(\gamma) = \operatorname{arg\,inf}_{a>0} \{a: \tilde{F}_x(a) \geq \gamma\}$$

令  $0 < \gamma_j < 1$ ， $j = 0, 1, \dots, m-1$  滿足  $\mathbf{1} = \gamma_0 + 2 \sum_{j=1}^{m-1} \gamma_j$ 。我們將  $\mathbf{X}$  根據對稱型百分位數作分割，得到以下的區間：

$$\begin{aligned} A_1^- &= \left(-\infty, \mu_x - \tilde{F}_x^{-1}(1 - 2\gamma_1)\right), A_2^- = \left(\mu_x - \tilde{F}_x^{-1}(1 - 2\gamma_1), \mu_x - \tilde{F}_x^{-1}(1 - 2(\gamma_1 + \gamma_2))\right), \dots, A_m^- = \left(\mu_x - \tilde{F}_x^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j), \mu_x - \tilde{F}_x^{-1}(1 - 2 \sum_{j=1}^m \gamma_j)\right) \\ A_0 &= (\mu_x - \tilde{F}_x^{-1}(\gamma_0), \mu_x + \tilde{F}_x^{-1}(\gamma_0)), A_m^+ = \left(\mu_x + \tilde{F}_x^{-1}(1 - 2 \sum_{j=1}^m \gamma_j), \mu_x + \tilde{F}_x^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j)\right), \dots, A_2^+ = \left(\mu_x + \tilde{F}_x^{-1}(1 - 2(\gamma_1 + \gamma_2)), \mu_x + \tilde{F}_x^{-1}(1 - 2\gamma_1)\right) \\ ,A_1^+ &= (\mu_x + \tilde{F}_x^{-1}(1 - 2\gamma_1), \infty) \end{aligned}$$

並且覆蓋機率為  $\gamma_1, \dots, \gamma_{m-1}, \gamma_0, \gamma_{m-1}, \dots, \gamma_1$ ，則我們可以得到對稱型分類平均數：

$$\theta_j^- = E(Y|X \in A_j^-), j = 1 \dots m$$

$$\theta_0 = E(Y|X \in A_0)$$

$$\theta_j^+ = E(Y|X \in A_j^+), j = 1 \dots m$$

令  $\tilde{F}_n^{-1}(\gamma)$  為摺疊經驗分配  $\tilde{F}_n(\mathbf{a}) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(\hat{\mu}_x - \mathbf{a} < X < \hat{\mu}_x + \mathbf{a})$  的第  $\gamma$  百分位

數。則樣本對稱型分類平均數可得：

$$\begin{aligned} \hat{\theta}_1^- &= \frac{\sum_{i=1}^n Y_i I(X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2\gamma_1))}{\sum_{i=1}^n I(X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2\gamma_1))} \\ \hat{\theta}_2^- &= \frac{\sum_{i=1}^n Y_i I(\hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2\gamma_1) \leq X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2(\gamma_1 + \gamma_2)))}{\sum_{i=1}^n I(\hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2\gamma_1) \leq X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2(\gamma_1 + \gamma_2)))} \\ &\vdots \\ \hat{\theta}_m^- &= \frac{\sum_{i=1}^n Y_i I(\hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j) \leq X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^m \gamma_j))}{\sum_{i=1}^n I(\hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j) \leq X_i \leq \hat{\mu}_x - \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^m \gamma_j))} \\ \hat{\theta}_0 &= \frac{\sum_{i=1}^n Y_i I(\hat{\mu}_x - \tilde{F}_n^{-1}(\gamma_0) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(\gamma_0))}{\sum_{i=1}^n I(\hat{\mu}_x - \tilde{F}_n^{-1}(\gamma_0) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(\gamma_0))} \\ \hat{\theta}_m^+ &= \frac{\sum_{i=1}^n Y_i I(\hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^m \gamma_j) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j))}{\sum_{i=1}^n I(\hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^m \gamma_j) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2 \sum_{j=1}^{m-1} \gamma_j))} \\ &\vdots \\ \hat{\theta}_2^+ &= \frac{\sum_{i=1}^n Y_i I(\hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2(\gamma_1 + \gamma_2)) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2\gamma_1))}{\sum_{i=1}^n I(\hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2(\gamma_1 + \gamma_2)) \leq X_i \leq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2\gamma_1))} \\ \hat{\theta}_1^- &= \frac{\sum_{i=1}^n Y_i I(X_i \geq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2\gamma_1))}{\sum_{i=1}^n I(X_i \geq \hat{\mu}_x + \tilde{F}_n^{-1}(1 - 2\gamma_1))} \end{aligned}$$

由對稱型分組平均也可以定義對稱型的截尾平均如下

$$\hat{\mu}_{y,sym} = \frac{\sum_{i=1}^n y_i I(\hat{\mu}_x - \hat{F}_n^{-1}(\gamma) \leq X_i \leq \hat{\mu}_x + \hat{F}_n^{-1}(\gamma))}{\sum_{i=1}^n I(\hat{\mu}_x - \hat{F}_n^{-1}(\gamma) \leq X_i \leq \hat{\mu}_x + \hat{F}_n^{-1}(\gamma))}$$

來估計  $\mathbf{Y}$  之平均。

定義  $\tilde{F}_x^+(\gamma) = \mu_x + \tilde{F}^{-1}(\gamma)$ ,  $\tilde{F}_x^-(\gamma) = \mu_x - \tilde{F}^{-1}(\gamma)$ ,  $\tilde{\lambda}^+(\gamma) = E(Y - \mu_y | X = \tilde{F}_x^+(\gamma))$  以及  $\tilde{\lambda}^-(\gamma) = E(Y - \mu_y | X = \tilde{F}_x^-(\gamma))$ 。

### 定理 3.1

$$\begin{aligned} \text{(a)} \quad \sqrt{n}(\hat{\mu}_{y,sym} - \mu_y) &= \frac{1}{\gamma} [\tilde{\lambda}^+(\gamma) f_X(\tilde{F}_x^+(\gamma)) - \tilde{\lambda}^-(\gamma) f_X(\tilde{F}_x^-(\gamma)) + \\ &\quad (\tilde{\lambda}^+(\gamma) f_X(\tilde{F}_x^+(\gamma))) + \tilde{\lambda}^-(\gamma) f_X(\tilde{F}_x^-(\gamma)) (f_X(\tilde{F}_x^+(\gamma)) + \\ &\quad f_X(\tilde{F}_x^-(\gamma)))^{-1} (f_X(\tilde{F}_x^+(\gamma)) + f_X(\tilde{F}_x^-(\gamma))] \sqrt{n}(\hat{\mu}_x - \mu_x) + \\ &\quad (\tilde{\lambda}^+(\gamma) f_X(\tilde{F}_x^+(\gamma)) + \tilde{\lambda}^-(\gamma) f_X(\tilde{F}_x^-(\gamma))) (f_X(\tilde{F}_x^+(\gamma)) + \\ &\quad f_X(\tilde{F}_x^-(\gamma)))^{-1} n^{-1/2} \sum_{i=1}^n (\gamma - I(|X_i - \mu_x| \leq \tilde{F}^{-1}(\gamma)) + n^{-1/2} \sum_{i=1}^n (y_i - \\ &\quad \mu_y) I(|X_i - \mu_x| \leq \tilde{F}^{-1}(\gamma)) \} + O_p(1) \end{aligned}$$

(b) 若  $f_x$  對稱於  $\mu_x$  以及  $\tilde{\lambda}^-(\gamma) = \tilde{\lambda}^+(\gamma)$ ，則

$$\begin{aligned} \sqrt{n}(\hat{\mu}_{y,sym} - \mu_y) &= \frac{1}{\gamma} (2\tilde{\lambda}^+(\gamma) f_X(\tilde{F}_x^+(\gamma))) \sqrt{n}(\hat{\mu}_x - \mu_x) + n^{-1/2} \sum_{i=1}^n (y_i - \\ &\quad \mu_y) I(|X_i - \mu_x| \leq \tilde{F}^{-1}(\gamma)) \} + O_p(1) \end{aligned}$$

這個定理說明  $\sqrt{n}(\hat{\mu}_{y,sym} - \mu_y)$  將會有近似常態分配。

## 4. 模擬

下一步，我們考慮在有限樣本下，對有母數和無母數估計量 $\theta_p$ 的效率進行模擬，其中 $\mathbf{Y}$ 和 $\mathbf{X}$ 為聯合常態分布。定義 $\hat{\theta}_N^j$ 以及 $\hat{\theta}_p^j$ 為無母數和有母數在第 $j$ 次重複試驗時的值。我們計算以下的 MSE：

$$MSE_N = \frac{1}{m} \sum_{j=1}^m (\hat{\theta}_N^j - \theta_p)' ((\hat{\theta}_N^j - \theta_p))$$

$$MSE_P = \frac{1}{m} \sum_{j=1}^m (\hat{\theta}_p^j - \theta_p)' ((\hat{\theta}_p^j - \theta_p))$$

步驟如下：

1. 對 $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{yx} \\ \sigma_{xy} & 1 \end{pmatrix}\right)$ 作  $m=1000$  次抽樣。
2. 將每一次的抽樣做 MSE 的計算並記錄下來。
3. 最後再將  $m$  次結果取平均得出 **Table1** 之結果，其中分類段數為 4。

**Table1.**有母數、無母數與對稱型分組平均之均方誤差

Sample size	$\sigma_{yx} = 0.2$	0.3	0.5	0.7	0.8
$n = 30$					
$MSE_p$	0.265	0.252	0.233	0.212	0.218
$MSE_{np}$	0.904	1.095	1.770	2.981	3.641
$MSE_{snp}$	0.998	1.273	1.763	2.610	3.126
$n = 50$					
$MSE_p$	0.145	0.147	0.145	0.130	0.123
$MSE_{np}$	0.582	0.772	1.381	2.454	3.055
$MSE_{snp}$	0.711	0.870	1.352	2.239	2.716
$n = 100$					
$MSE_p$	0.077	0.072	0.073	0.068	0.066
$MSE_{np}$	0.309	0.512	1.096	1.923	2.447
$MSE_{snp}$	0.392	0.546	1.106	1.950	2.474

當分配為常態時有母數、無母數及對稱型三種估計量皆一致性收斂到同一個分組平均數 $\theta_p$ 。而我們的模擬分析結果顯示有母數方法比其他兩種無母數有效的

非常多。而無母數方法正是流行病學中常用的。兩種無母數方法似乎非對稱方法大致較好，但某些狀況卻是對稱型較佳。我們覺得應該是兩種無母數方法混合使用；也就是 $\alpha$ 較接近 0.5 的分位數要用無母數，而 $\alpha$ 較接近於 0 或 1 時要用對稱型將會建造一個效率較高的無母數分組平均估計量。

同時我們也計算了三種分類型態 covariance matrix 之 MSE

### (a) 有母數

1. 對 $(\begin{matrix} Y \\ X \end{matrix}) \sim N(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix})$ 作  $m=1000$  次抽樣。
2. 將每一次的抽樣，根據黃子芸(2013)研究結果計算其  $\Gamma$ 、 $V$  矩陣得出估計的 covariance matrix。
3. 將每一次的抽樣計算後的 covariance matrix 和該次抽樣所計算的 covariance matrix 理論值做方差，同時記錄每一次方差的結果。
4. 將  $m=1000$  次結果作平均得出 MSE matrix。

### (b) 無母數

1. 對 $(\begin{matrix} Y \\ X \end{matrix}) \sim N(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix})$ 作  $m=1000$  次抽樣。
2. 以 $\hat{\lambda}_j = \frac{\sum_{i=1}^n (Y_i - \hat{\mu}_y) I(\hat{F}_x^{-1}(\alpha_j) - a \leq X_i \leq \hat{F}_x^{-1}(\alpha_j) + a)}{\sum_{i=1}^n I(\hat{F}_x^{-1}(\alpha_j) - a \leq X_i \leq \hat{F}_x^{-1}(\alpha_j) + a)}$ 作為 $\lambda_j$ 的估計量。
3. 初始  $a$  值設定為 0.01，對於每一個迴圈，以漸進的方式向外擴張  $a$  值。每一次擴張增加 0.01 距離，直到 $\hat{\lambda}_j$ 和 $\lambda_j$ 的均方差在可接受的範圍內。
4. 計算 $m_j = (\alpha_j - \alpha_{j-1}) \int_{-\infty}^{\infty} \int_{x \in A_j} (y - \mu_y) f(x, y) dx dy$   
計算 $M_j = (\alpha_j - \alpha_{j-1}) \int_{-\infty}^{\infty} \int_{x \in A_j} (y - \mu_y)^2 f(x, y) dx dy$
5. 計算 $\hat{m}_j = (\alpha_j - \alpha_{j-1}) \frac{\sum_{i=1}^n (Y_i - \bar{Y}) I(X_i \in \hat{A}_j)}{\sum_{i=1}^n I(X_i \in \hat{A}_j)}$   
計算 $\hat{M}_j = (\alpha_j - \alpha_{j-1}) \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 I(X_i \in \hat{A}_j)}{\sum_{i=1}^n I(X_i \in \hat{A}_j)}$
6. 由  $\lambda_j$ 、 $m_j$ 、 $M_j$  以及  $\hat{\lambda}_j$ 、 $\hat{m}_j$ 、 $\hat{M}_j$ ，根據定理 2.1 可得單次抽樣的 covariance matrix 理論值  $\Sigma$  以及估計的 covariance matrix  $\hat{\Sigma}$ ，並記錄其方差結果。

7. 將  $m=1000$  次結果作平均得出 MSE matrix。

**(c) 對稱型**

1. 以 Bootstrap 法得出估計的 covariance matrix  $\hat{\Sigma}$
2. 計算  $\hat{\Sigma}$  和無母數理論值  $\Sigma$  之方差並記錄結果
3. 將  $m=1000$  次結果做平均 MSE matrix。

結果如下表(Table.2)：

SampleSize	MSEType	$(\sigma^2_y, \sigma^2_x)$	$\sigma_{xy}=0.2$	$\sigma_{xy}=0.3$	$\sigma_{xy}=0.5$	$\sigma_{xy}=0.7$	$\sigma_{xy}=0.8$
50	MSE <sub>p</sub>	(1, 1)	0.1470813	0.14218658	0.1448758	0.1307102	0.1336261
100			0.0728628	0.07412975	0.0689616	0.0643806	0.0619072
50	MSE <sub>np</sub>	(1, 1)	0.2778895	0.4599915	0.9831022	1.7409064	1.629092
100			0.2059092	0.3737305	0.9288406	1.2986642	1.371124
50	MSE <sub>snp</sub>	(1, 1)	0.3760505	0.5344480	1.0629342	1.8462645	2.375956
100			0.2648890	0.4470820	0.9600909	1.7757849	2.279510
50	MSE <sub>p</sub>	(1, 2)	0.1500156	0.1495779	0.1482542	0.1411011	0.1394231
100			0.0712700	0.0742531	0.0707686	0.0685668	0.0695003
50	MSE <sub>np</sub>	(1, 2)	0.6798328	0.6992730	0.9740117	1.2986640	1.6290920
100			0.4288380	0.4992917	0.8276544	1.1654620	1.3711241
50	MSE <sub>snp</sub>	(1, 2)	0.3305827	0.3902435	0.6651758	1.0484483	1.303644
100			0.2002260	0.2706314	0.5517891	0.9841396	1.223857
50	MSE <sub>p</sub>	(2, 1)	0.2852711	0.3056648	0.3068510	0.2868170	0.2758600
100			0.1433822	0.1406264	0.1456224	0.1473143	0.1350802
50	MSE <sub>np</sub>	(2, 1)	0.4148160	0.5902806	1.1331147	1.9523090	2.4314770
100			0.2699151	0.4421781	0.9967844	1.8130980	2.3209371
50	MSE <sub>snp</sub>	(2, 1)	0.6334337	0.8650501	1.3238753	2.0878886	2.582504
100			0.3966607	0.5512686	1.1237594	1.9319453	2.416740

三種分組平均的均方差均表示樣本變大時其近似變異矩陣的估計量會較準確。這也表示樣本數變大時利用三種分組平均所做的統計推論也會較準確。同時，在三種方法中有母數分組平均的變異矩陣最為準確。大致而言無母數分組平均在估計共變異矩陣較對稱型分組平均為好。但  $(\sigma^2_y, \sigma^2_x) = (1, 2)$  且  $\sigma_{yx}$  較大時(0.7、0.8)對稱型反而較佳。

接著我們分別計算三種分類型態的檢定力，並在 **Table3.(附錄資料)** 以  $\text{pow}_P$ 、 $\text{pow}_N$ 、 $\text{pow}_{\text{snp}}$  表示有母數、無母數以及對稱型之檢定力大小。

其過程如下：

欲檢定  $H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4$  v.s.  $H_1: \theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4$

$H_0$  可改寫為  $A\theta = 0$ ，其中  $A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

令  $\tau_j = \theta_j - \theta_{j-1}$ ,  $j=1,2,3$ ;  $\hat{\tau} = \widehat{A\theta} = A\hat{\theta}$

$\sqrt{n}(\widehat{A\theta} - A\theta) \sim N(0, A(\text{cov}(\theta)A'))$

1. 經模擬算出三種分類型態各自的  $\text{cov}(\hat{\theta})$

2. 對  $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}\right)$  作 5000 次重複抽樣。

3. 每一次抽樣均計算  $\hat{\tau}' \text{cov}(\hat{\theta})^{-1} \hat{\tau}$  並記錄結果

4. 將此 5000 筆紀錄中 0.95 百分位數令為  $a$

5. 重新進行對  $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}\right)$  5000 次重複抽樣

6. 在每一次抽樣中，檢驗  $T = n(A\hat{\theta})'(A\text{cov}(\hat{\theta})A')^{-1}(A\hat{\theta})$  是否大於等於  $a$

7. 計算 5000 次重複抽樣中  $T$  大於等於  $a$  之比率

根據 **Table3** 的結果我們有以下結論：

- 當一組資料來自於二維常態分配，在探討解釋變數與反應變數是否有相關時使用有母數分組平均來建立檢定統計量最為有效。
- 因為當資料不是來自於常態分配時採用有母數方法在理論上是不對的。從這個分析結果無母數及對稱型無母數兩種方法互有大小檢定力。
- 三種方法在共變異數  $\sigma_{yx}$  較大時均具有較大的檢定力。

## 5. 資料分析

由高血壓所產生的生理問題，包含心臟病、中風以及其他心血管疾病，已成為全球的健康議題。因此，研究造成高血壓的可能原因是現今需要探討的重要目標。

從美國人類健康中心，曾為了研究而進行全國性健康及營養調查 Trola

Triola(2006)，得來的男性和女性的資料表中，包含了十三個和健康相關的變數。

此筆資料讓我們能夠研究由 Kotsis et al.(2006)與 Kagan et al.(2007)所注意到的性別間的差異。在眾多變數之中，我們考慮一些和高血壓較為相關的變數，如：身體質量指數(BMI)以及膽固醇(CHOL)，來建立參考的分類。

BMI 和血壓之間的關係經常被發現，但還有更多等待著我們去發掘(Drowold et al.(2007), Agyemang et al.(2009), Nanaware, Gavekare, Surdi (2011))。

我們將會展示由有母數、無母數建立的百分位數對血壓樣本進行四段分類。其中不論男性或女性的資料； $\alpha_1 = 0.25, \alpha_2 = 0.50, \alpha_3 = 0.75$  並在 **Table4** 呈現。

**Table4.**有母數、無母數及對稱型分組血壓之樣本平均

Gender	Estimator	$A_1$	$A_2$	$A_3$	$A_4$
Male	P means	114.8	117.8	119.9	122.9
	NP means	113.4	118.2	119.1	124.4
	Sym means	114.0	118.6	121.2	122.8
Female	P means	97.7	107.4	114.1	123.8
	NP means	104.5	105.2	108.3	125.2
	Sym means	104.6	105.2	111.4	127.3

根據 **Table4** 我們有兩個結論：

- (a) 我們不斷地發現血壓和 BMI 的關聯似乎呈現正相關，不論是男性或女性，其有母數或無母數或對稱型無母數的分類平均數都是單調遞增。
- (b) 在男性和女性分割區間  $A_1$  以及  $A_2$  的平均數當中，我們發現血壓在正常以及偏低的男性和女性體重上有性別上的差異。對於到較瘦的人有著較高的血壓，我們考慮到年紀、飲酒以及缺乏運動。這樣的結果顯示了近來非肥胖的男性較女性在罹患高血壓的風險上有著更高的影響。而這項結果目前尚未被注意到。

根據以上結果無母數及對稱型無母數結果差異不大。以下我們將分析有母數及無母數方法的進一步研究。男性和女性在血壓上的比較可以利用假設檢定來進行正式的推論。我們以  $\mu_{y,m}$  和  $u_{y,f}$  表示男性和女性血壓的母體平均數。對於假設  $H_0: \mu_{y,m} = u_{y,f}$  v.s.  $H_1: \mu_{y,m} > u_{y,f}$ ，其 p-value 係根據傳統的兩樣本 T 檢定為 0.2839。

這說明了未分類的方法不夠顯著，並不能給予我們足夠的證據去評斷男性的高血壓大於與女性。我們接著將使用有母數分類法驗證更進一步的結果。

為了便於解釋，我們暫時令  $c=m$  為男性以及  $c=f$  為女性，並且重新標示

$\Sigma_{p,c}(\Lambda_c)$   $\theta_{jp,c}$  以及  $n_c$ ；其分別為母體共變異矩陣、第  $j$  個母體分類平均以及性別的母體樣本數。同時我們標示分配性參數的最大概似估計量  $\Lambda$  為  $\hat{\Lambda}_{mle,c}$ 。定理

**2.1** 指出以下的檢定統計量為有母數分類平均數所適用：

$$\frac{\hat{\theta}_{jp,m} - \hat{\theta}_{jp,f}}{\sqrt{\frac{1}{n_m} e_j' \Sigma_{p,m}(\hat{\Lambda}_{mle,m}) e_j + \frac{1}{n_f} e_j' \Sigma_{p,f}(\hat{\Lambda}_{mle,f}) e_j}}$$

用於檢定  $H_0: \theta_{jp,m} = \theta_{jp,f}$  v.s.  $H_1: \theta_{jp,m} > \theta_{jp,f}$

而 p-value 對於四段分類 ( $j = 1,2,3,4$ ) 的假設檢定為 0.043、0.023、0.06、

0.071。儘管兩性均在偏低以及正常體重的範圍內罹患高血壓的風險較小，但較低的 p-value 仍意味著男性在這兩個分類段落內罹患高血壓的風險較女性為更高。

然而，相較之下古典的 T 檢定卻無法發現此種跡象。

即使早期的研究關於膽固醇和高血壓有著許多不同的結論(Hjermann et al.(1978))，但近期流行病學的研究傾向於認為這兩個因子是息息相關的(Hjermann et al.(1978), Reaven(1988) Ferrara et al.(2002)) 於此，我們考慮迴歸模型的最小平方估計，其一為簡單模型： $BP = \beta_0 + \beta_1 CHOL + \epsilon$ ；其二為多重迴歸模型： $BP = \beta_0 + \beta_1 CHOL + \beta_2 BMI + \epsilon$ 。男性和女性的估計結果將呈現在 Table5。

Table5.迴歸參數之最小平方估計

Variables	Parameter estimates
CHOL	
Male	$\hat{\beta}_0 = 118.6, \hat{\beta}_1 = 0.0006$
Female	$\hat{\beta}_0 = 104.3, \hat{\beta}_1 = 0.026$
CHOL, BMI	
Male	$\hat{\beta}_0 = 94.94, \hat{\beta}_1 = -0.0005, \hat{\beta}_2 = 0.93$
Female	$\hat{\beta}_0 = 67.87, \hat{\beta}_1 = 0.00005, \hat{\beta}_2 = 1.667$

對於女性簡單迴歸迴歸參數 $\beta_1$ 的最小平方估計顯示了中等且正向的值(0.026)。其餘的 $\beta_1$ 均為接近 0 的值或者為負值，此結果可能會導致錯誤的血壓和膽固醇關係統計推論。

我們有興趣驗證是否血壓和膽固醇還存在著其他的關聯。我們使用由樣本分類平均所建立的四段百分位數，其中 $\alpha_1 = 0.25, \alpha_2 = 0.50, \alpha_3 = 0.75$ ，並在

Table6 呈現：

Table6.由膽固醇來分組的有母數差異與無母數、對稱型血壓分組平均估計值

Gender	Estimator	$A_1$	$A_2$	$A_3$	$A_4$
Male	P means	118.6	118.8	118.9	119.1
	NP means	119.8	120.8	115.3	119.7

	Sym means	119.6	118.5	117.0	119.7
Female	P means	104.4	109.1	112.4	117.1
	NP means	105.6	107.2	108.7	121.7
	Sym means	104.5	105.1	109.8	128.4

我們發現樣本的血壓平均數在有母數的方法下，不論男性女性均是持續地增加。但無母數的方法在男性資料上並無顯示此種趨勢。有母數分類樣本平均數顯示了迴歸方法所找不到的血壓和膽固醇的關係。這更進一步地顯示了在血壓上的性別差異。

總地來說，由本資料所得到的分析結果指出有效率的有母數分類法可以提供健康因子間更多的潛在的觀察。

作為重要的心血管疾病因子，高血壓和心臟病已成為美國人的趨勢[7,8]。然而肥胖是否為血壓和心血管疾病的重要角色一直備受爭論著。部份研究報告指出高血壓或者不斷上升的血壓在心血管疾病上，較瘦的人反而較肥胖者有著更高的罹患率[9-11]；而有些則指出兩者無任何關聯，甚至也出現兩者具有高相關性的報告。

一直以來在高血壓群中，較瘦者是否較肥胖者容易罹患心血管疾病仍然是一個未知數。

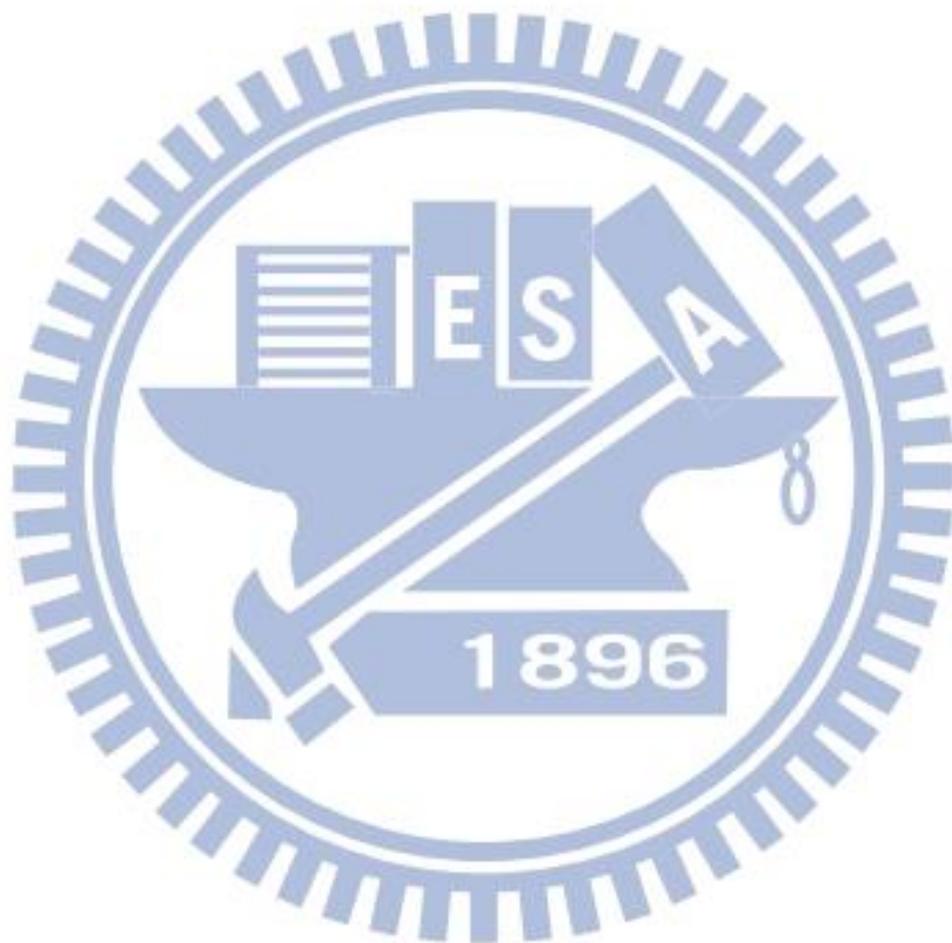
我們的研究認為：

(a)高血壓和心血管疾病的關聯性大小，一般來說，隨著 BMI 增加而增加。而高血壓不該在體重正常或偏低的族群中被認為和肥胖者相較之下為一個無關緊要的議題。

(b)在第一段以及第二段的分類區間內，50 歲以下的人口之中，男性較女性高血壓的普及率更高。

(c)在高齡層，女性則較男性較高的高血壓率

雖然重要的高血壓探究仍不明朗，我們卻很常看到高血壓影響男性人數多於女性；而在 55 歲之後，高血壓影響女性的人數多於男性。



## 6. 附錄資料

### 5.1 Table3

(i)有母數

$$(\sigma^2_Y, \sigma^2_X) = (1, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.1948	0.3542	0.7806	0.9822	0.9982
n=50	0.2994	0.5658	0.9490	0.9994	1.0000
n=100	0.5116	0.8544	0.9996	1.0000	1.0000

$$(\sigma^2_Y, \sigma^2_X) = (1, 2)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.1080	0.2052	0.4840	0.7542	0.8826
n=50	0.1632	0.3122	0.7038	0.9432	0.9864
n=100	0.2820	0.5604	0.9392	0.9992	1.0000

$$(\sigma^2_Y, \sigma^2_X) = (2, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.1130	0.2092	0.4718	0.76666	0.8760
n=50	0.1680	0.3116	0.6976	0.9440	0.9868
n=100	0.2876	0.5850	0.9436	0.9988	1.0000

(ii)無母數

$$(\sigma^2_Y, \sigma^2_X) = (1, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.123	0.214	0.509	0.871	0.964
n=50	0.188	0.320	0.809	0.995	1
n=100	0.332	0.665	0.990	1	1

$$(\sigma^2_Y, \sigma^2_X) = (1, 2)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.088	0.138	0.298	0.537	0.657
n=50	0.117	0.218	0.519	0.806	0.904
n=100	0.198	0.439	0.850	0.996	1

$$(\sigma^2_Y, \sigma^2_X) = (2, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.141	0.114	0.319	0.488	0.689
n=50	0.143	0.227	0.419	0.813	0.922
n=100	0.195	0.395	0.893	0.995	1

(iii) 對稱型

$$(\sigma^2_Y, \sigma^2_X) = (1, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.115	0.243	0.591	0.914	0.964
n=50	0.184	0.324	0.791	0.994	1
n=100	0.333	0.668	0.997	1	1

$$(\sigma^2_Y, \sigma^2_X) = (1, 2)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.089	0.161	0.287	0.550	0.686
n=50	0.096	0.236	0.490	0.794	0.904
n=100	0.150	0.360	0.840	0.992	1

$$(\sigma^2_Y, \sigma^2_X) = (2, 1)$$

Sample	$\sigma_{yx} = 0.2$	$\sigma_{yx} = 0.3$	$\sigma_{yx} = 0.5$	$\sigma_{yx} = 0.7$	$\sigma_{yx} = 0.8$
n=30	0.081	0.108	0.290	0.537	0.725
n=50	0.104	0.202	0.501	0.847	0.903
n=100	0.172	0.334	0.836	0.994	0.998



## 7. 參考文獻

1. 黃子芸(2013),Categorized Means,交通大學統計所碩士論文
2. Agyemang, C., Kearneyman, E., Zijmans, W., Wendte, J. and Stronks, K. (2009). Blood pressure and body mass index in an ethnically diverse sample of adolescents in Paramaribo, Suriname. *BMC Cardiovascular Disorders*, **9**, 1-9.
3. Drowold, W. B., Midthjell, K., Nilsen, T. I. and Holmen, J. (2005). Change in body mass index and its impact on blood pressure: a prospective population study. *International Journal of Obesity*, **29**, 650-655.
4. Ferrara, L. A., Guida, L., Iannuzzi, R., Celentano, A. and Lionello, F. (2002). Serum cholesterol affects blood pressure regulation. *Journal of Human Hypertension*, **16**, 337-343.
5. Hjermann, I., Helgeland, A., Holme, I., Lund-Larsen, P. G. and Leren, P. (1978). The association between blood pressure and serum cholesterol in healthy men: the Oslo study. *Journal of Epidemiology and Community Health*, **32**, 117-123.
6. Kagan, A., Faibel, H., Ben-Arie, G., Granevitze, Z. and Rapoport, J. (2007). Gender differences in ambulatory blood pressure monitoring

profile in obese, overweight and normal subjects. *Journal of Human Hypertension*, **21**, 128-134

7. Kearney, P. M., Whelton, M., Reynolds, K., Muntner, P., Whelton, P. K. and He, J. (2005). Global burden of hypertension: analysis of worldwide data. *Lancet*, **365**, 217-223.

8. Kotsis, V., Stabouli, S., Pitiriga, V., Toumanidis, S., Papamichael, C. and Zakopoulos, N. Ambulatory blood pressure monitoring and target organ damage: effects of age and sex. *Blood Pressure Monitoring*, **11**, 9-15.

9. Nanaware, N. L., Gavkare, A. M. and Surdi, A. D. (2011). Study of correlation of body mass index (BMI) with blood pressure in school going children and adolescents. *International Journal of Recent Trends in Science and Technology*, **1**, 20-26.

10. Reaven, G. M. (1988). Role of insulin resistance in human disease. *Diabetes*, **37**, 1595-1607.

11. Triola, M. M. and Triola, M. F. (2006). *Biostatistics for the Biological and Health Sciences*, Pearson Education, Inc.: Boston.