

國立中央大學

統計研究所

碩士論文

兩條迴歸直線比值的區間估計

指導教授：陳玉英 博士

研究生：游明珠

中華民國九十五年六月



國立中央大學圖書館

碩博士論文電子檔授權書

(93年5月最新修正版)

本授權書所授權之論文全文電子檔，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

- ()同意 (立即開放)
- ()同意 (一年後開放)，原因是：_____
- ()同意 (二年後開放)，原因是：_____
- ()不同意，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動讀者間「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名： 游明珠

論文名稱： 兩條迴歸直線比值的區間估計

指導教授姓名： 陳玉英 博士

系所： 統計研究所 博士 碩士班

學號： 92225013

日期：民國 95 年 6 月 30 日

備註：

1. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
2. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
3. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

摘 要

為研究兩種處理效應的異同，可藉由處理平均數比值及其信賴區間加以衡量。本文則進一步研究，當兩個處理組的反應變數均受另一個共變數影響時，如何衡量此兩種處理效應之異同。本文考慮針對兩個處理組的資料，分別配適簡單線性迴歸模式，然後根據所估計之迴歸直線建構兩條迴歸直線的近似條件信賴區間及信賴束。文中除模擬研究外，也以一組資料說明其應用。

Abstract

To study the difference of two treatment effects, one can measure the ratio of the two treatment means together with its confidence interval. This article investigates the problem of comparing two treatment effects when the responses are affected by one covariate. The two groups of data are fitted separately by two simple linear regression models. Based on the fitted linear regression lines, the conditional confidence intervals and confidence band are constructed. This article presents the simulation results and demonstrates the use of the proposed procedures.

致 謝 辭

本論文得以順利完成由衷感謝指導教授 陳玉英 博士，陳老師在學生研究所期間的殷殷教誨與悉心指教，令我獲益匪淺。此外，還要感謝口試委員 嵇允嬋 教授、曾議寬 博士與 黃怡婷 博士在論文口試當天，不吝於對學生的提出指正，並提供眾多寶貴的意見與建議，致使本論文的內容更加充實完善，特此致上十二萬分的謝意。

在中央大學攻讀碩士這三年期間，首先，要感謝我親愛的家人，有你們不斷地給予鼓勵，才能使我在遇到挫折時再站起來，你們是我支持下去的原動力。當然，還要感謝陪我一起茁壯、一起成長的同窗好友們，你們的陪伴，為我的學生生涯增添許多歡笑與回憶；在彼此的互相扶持之下，我們一同經歷這人生的一個考驗。另外，我還要感謝春樹學長、玉媚學姊、啟軒學長，謝謝你們常常在我有疑惑時，不吝給予我意見，讓我在作抉擇之前，能有更詳盡的考量。

謹將本文獻給摯愛的父母、親友、以及所有愛護關心我的人。

游明珠 謹識於

中央大學統計研究所

中華民國九十五年六月

目 錄

第一章 緒論.....	1
1.1 研究動機及內容.....	1
1.2 文獻回顧.....	3
第二章 統計方法.....	7
2.1 兩條迴歸直線比值的條件信賴區間.....	8
2.2 兩條迴歸直線比值的信賴束.....	14
第三章 模擬研究.....	16
3.1 模擬方法.....	16
3.2 模擬結果.....	18
第四章 實例分析.....	19
第五章 結論及未來可行之研究.....	22
參考文獻.....	23
圖表	

圖 目 錄

圖 1：模擬研究中之迴歸直線。子圖(a)-(d)分別為迴歸模式 I~IV 中的 兩條迴歸直線.....	24
圖 2：男女性體內鉀離子含量與總水重的散佈圖及所估計之迴歸直線	25
圖 3：男性之體內鉀離子含量相對於總水重迴歸直線殘差圖及其常態 機率圖.....	26
圖 4：女性之體內鉀離子含量相對於總水重迴歸直線殘差圖及其常態 機率圖.....	27
圖 5：男女性鉀離子平均含量比值的 95%條件信賴區間及 95%信賴束	28

表 目 錄

表 1：兩條迴歸直線比值之 95% 條件信賴區間之覆蓋機率(期望長度)	29
表 2：兩條迴歸直線比值之 95% 條件信賴束之覆蓋機率(期望長度)	30
表 3：男女性體內水分與鉀離子含量	31

第一章 緒論

1.1 研究動機

為研究兩個母體是否有異，可藉由估計兩個母體平均數之差或比值或求其信賴區間加以衡量。一般而言，常態分布平均數的差異可以反應兩個母體的差異，但是當兩母體的平均數很接近的時候，兩個母體平均數的比值將會更適當的反映出兩者的差異。例如：母體一的平均數為 0.01，母體二的平均數為 0.0001，此兩者的差為 0.0099，但母體一平均數卻是母體二的平均數的一百倍。此外，比值在生活中的應用相當廣泛，例如外匯研究中美元對新台幣的比值。基因研究中，利用電腦軟體判讀，比較可能的病患與無病人員染色體螢光的強弱比值。

假設獨立的隨機變數 X 及 Y 均來自平均數為零，變異數為一的常態分布，則 X/Y 服從柯西 (Cauchy) 分布。但柯西分布的動差 (moment) 並不存在，所以我們無法應用擴展若干倍 X/Y 標準誤差的方法建立兩母體平均數比值的信賴區間。因此，Fieller (1954) 利用變數轉換方法求得兩個常態母體平均數比值的信賴區間。然而，實務上，可能有共變數 (covariates) 與 X 及 Y 皆有關。例如：鉀離子在人體中的總含量跟人體中的水份有關，但是男女之間，其鉀離子與水份的關係或有不同。為比較男性與女性體內鉀離子含量有何不同時，則需要先行配適

迴歸模式，然後比較兩條迴歸直線的異同。因此本文除推廣 Fieller (1954)及 Kempen 和 Vliet (2000)的研究，在有共變數的情形下建立兩母體平均數比值之條件信賴區間，也引用 Delta 方法建立近似條件信賴區間。本文甚進一步利用 Delta 方法建立兩母體平均數比值之信賴束。相關文獻的回顧列於下一節。本文第二章介紹具有共變數情形下，兩條迴歸直線比值的條件信賴區間及信賴束。第三章列表並且討論有關的模擬研究，第四章以實例說明所提方法的應用，在最後一章，第五章，則針對本文提出結論及討論未來研究。

1.2 文獻回顧

本章首先回顧 Fieller (1954) 的兩個常態母體平均數比值之信賴區間。假設 Y_1 和 Y_2 為服從二元常態分布之隨機變數，其平均數為 μ_1 及 μ_2 ，變異數為 σ_1^2 及 σ_2^2 ，共變數(covariance)則為 σ_{12} 。

假設 $\mu_2 \neq 0$ ，令 θ 為兩母體平均數的比值，即

$$\theta = \frac{\mu_1}{\mu_2}。$$

令 \bar{Y}_1 及 \bar{Y}_2 為樣本數為 n 的樣本平均數，則 θ 的估計式為

$$\hat{\theta} = \frac{\bar{Y}_1}{\bar{Y}_2}。$$

考慮下列變數：

$$Z_i = Y_{1i} - \theta Y_{2i} \quad , \quad i = 1, 2, \dots, n \quad ,$$

則

$$\bar{Z} = \bar{Y}_1 - \theta \bar{Y}_2。$$

\bar{Z} 之期望值及變異數分別為

$$E(\bar{Z}) = 0$$

和

$$V_{\theta}(\bar{Z}) = \frac{1}{n}(\sigma_2^2 - 2\theta\sigma_{12} + \theta^2\sigma_1^2)。$$

在 θ 已知情形下，此一變異數不偏估計式為

$$\hat{V}_{\theta}(\bar{Z}) = \frac{1}{n-1}(S_2^2 - 2\theta S_{12} + \theta^2 S_1^2)$$

其中 S_1^2 和 S_2^2 分別為 Y_1 樣本和 Y_2 樣本的變異數， S_{12} 則為 Y_1 和 Y_2 的樣本共變異數。Fieller (1954) 考慮統計式

$$T = \frac{\bar{Y}_1 - \theta \bar{Y}_2}{\sqrt{\frac{1}{n-1}(S_2^2 - 2\theta S_{12} + \theta^2 S_1^2)}}$$

並且求出其近似分布為自由度為 $(n-1)$ 之 t 分布，記作 t_{n-1} 。因此

$$P\{T^2 \leq t_{n-1, \alpha}^2\} = 1 - \alpha,$$

其中 $t_{n-1, \alpha}$ 為分布 t_{n-1} 的上 α 百分位。求解 $T^2 \leq t_{n-1, \alpha}^2$ ，即 θ 函數的二次式：

$$\left(\bar{Y}_2^2 - \frac{t_{n-1, \alpha}^2}{n-1} S_2^2\right) \theta^2 - 2\left(\bar{Y}_1 \bar{Y}_2 - \frac{t_{n-1, \alpha}^2}{n-1} S_{12}\right) \theta + \left(\bar{Y}_1^2 - \frac{t_{n-1, \alpha}^2}{n-1} S_1^2\right) \leq 0$$

可得知在信心水準 $100(1-\alpha)\%$ 之下 θ 的信賴上下限為

$$\frac{b \pm \sqrt{b^2 - ac}}{a}$$

其中 $a = \bar{Y}_2^2 - t_{n-1, \alpha}^2 S_2^2 / (n-1)$ ， $b = \bar{Y}_1 \bar{Y}_2 - t_{n-1, \alpha}^2 S_{12} / (n-1)$ ，且

$c = \bar{Y}_1^2 - t_{n-1, \alpha}^2 S_1^2 / (n-1)$ 。

Kempen 和 Vliet (2000) 則對 \bar{Y}_1 / \bar{Y}_2 在 μ_1 及 μ_2 處求出二階的泰勒展

開式：

$$\begin{aligned} \frac{\bar{Y}_1}{\bar{Y}_2} &\approx \frac{\mu_1}{\mu_2} + (\bar{Y}_1 - \mu_1) \frac{\partial}{\partial \bar{Y}_1} \left(\frac{\bar{Y}_1}{\bar{Y}_2} \right) \Bigg|_{\mu_1, \mu_2} + (\bar{Y}_2 - \mu_2) \frac{\partial}{\partial \bar{Y}_2} \left(\frac{\bar{Y}_1}{\bar{Y}_2} \right) \Bigg|_{\mu_1, \mu_2} \\ &+ \frac{1}{2} (\bar{Y}_1 - \mu_1)^2 \frac{\partial^2}{\partial \bar{Y}_1^2} \left(\frac{\bar{Y}_1}{\bar{Y}_2} \right) \Bigg|_{\mu_1, \mu_2} + \frac{1}{2} (\bar{Y}_2 - \mu_2)^2 \frac{\partial^2}{\partial \bar{Y}_2^2} \left(\frac{\bar{Y}_1}{\bar{Y}_2} \right) \Bigg|_{\mu_1, \mu_2} \end{aligned}$$

$$+(\bar{Y}_1 - \mu_1)(\bar{Y}_2 - \mu_2) \frac{\partial^2}{\partial \bar{Y}_1 \partial \bar{Y}_2} \left(\frac{\bar{Y}_1}{\bar{Y}_2} \right) \Bigg|_{\mu_1, \mu_2}$$

並求得 \bar{Y}_1/\bar{Y}_2 的近似期望值及變異數分別為

$$\begin{aligned} E\left(\frac{\bar{Y}_1}{\bar{Y}_2}\right) &\approx \frac{\mu_1}{\mu_2} + V(\bar{Y}_2) \frac{\mu_1}{\mu_2^3} - \frac{\text{Cov}(\bar{Y}_1, \bar{Y}_2)}{\mu_2^2} \\ &\approx \frac{\mu_1}{\mu_2} + \frac{1}{n} \left(\frac{\mu_1}{\mu_2^3} V(Y_2) - \frac{\text{Cov}(Y_1, Y_2)}{\mu_2^2} \right) \\ &= \frac{\mu_1}{\mu_2} \left\{ 1 + \frac{V(Y_2)}{n\mu_2^2} - \frac{\text{Cov}(Y_1, Y_2)}{\mu_1\mu_2} \right\} \end{aligned}$$

和

$$\begin{aligned} V\left(\frac{\bar{Y}_1}{\bar{Y}_2}\right) &= E\left\{ \left(\frac{\bar{Y}_1}{\bar{Y}_2} - E\left(\frac{\bar{Y}_1}{\bar{Y}_2}\right) \right)^2 \right\} \approx E\left\{ \left(\frac{\bar{Y}_1}{\bar{Y}_2} - \frac{\mu_1}{\mu_2} \right)^2 \right\} \\ &\approx \frac{V(\bar{Y}_1)}{\mu_2^2} + \frac{\mu_1^2 V(\bar{Y}_2)}{\mu_2^4} - \frac{2\mu_1 \text{Cov}(\bar{Y}_1, \bar{Y}_2)}{\mu_2^3} \\ &\approx \frac{1}{n} \left[\frac{V(Y_1)}{\mu_2^2} + \frac{\mu_1^2 V(Y_2)}{\mu_2^4} - \frac{2\mu_1 \text{Cov}(Y_1, Y_2)}{\mu_2^3} \right] \\ &= \frac{\mu_1^2}{n\mu_2^2} \left[\frac{V(Y_1)}{\mu_1^2} + \frac{V(Y_2)}{\mu_2^2} - \frac{2\text{Cov}(Y_1, Y_2)}{\mu_1\mu_2} \right] \end{aligned}$$

因為 μ_1/μ_2 不偏估計式為

$$\frac{\bar{Y}_1}{\bar{Y}_2} - \frac{1}{n} \left(\frac{\bar{Y}_1}{\bar{Y}_2^3} S_2^2 - \frac{S_{12}}{\bar{Y}_2^2} \right)$$

則根據上式，求得 \bar{Y}_1/\bar{Y}_2 的近似期望值及變異數分別為

$$E\left(\frac{\bar{Y}_1}{\bar{Y}_2}\right) \approx \frac{\mu_1}{\mu_2}$$

及

$$V\left(\frac{\bar{Y}_1}{\bar{Y}_2}\right) \approx \frac{\mu_1^2}{n\mu_2^2} \left[\frac{V(Y_1)}{\mu_1^2} + \frac{V(Y_2)}{\mu_2^2} - \frac{2Cov(Y_1, Y_2)}{\mu_1\mu_2} \right]$$

第二章 統計方法

假設兩個母體(處理組)的反應變數為 Y_1 及 Y_2 ，令 X 為考慮中的解釋變數。若分別配適的迴歸模式為

$$Y_1 = \beta_0 + \beta_1 X + \varepsilon = \mu_1(x) + \varepsilon$$

及

$$Y_2 = \gamma_0 + \gamma_1 X + \varepsilon = \mu_2(x) + \varepsilon$$

其中 $\mu_1(x) = \beta_0 + \beta_1 X$ ， $\mu_2(x) = \gamma_0 + \gamma_1 X$ ， β_0 、 β_1 、 γ_0 、 γ_1 為迴歸係數， ε 為服從 $N(0, \sigma^2)$ 的誤差項。假設 $\{(X_{11}, Y_{11}), \dots, (X_{1n_1}, Y_{1n_1})\}$ 及 $\{(X_{21}, Y_{21}), \dots, (X_{2n_2}, Y_{2n_2})\}$ 為兩組互相獨立的隨機樣本，在簡單線性迴歸模型之下，配適的最小平方迴歸直線分別為 $\hat{Y}_1(x) = \hat{\beta}_0 + \hat{\beta}_1 x$ 及 $\hat{Y}_2(x) = \hat{\gamma}_0 + \hat{\gamma}_1 x$ (Draper and Smith, 1998)。因為兩個迴歸模式中誤差項的變異數相同， σ^2 的估計式為

$$S^2 = \left[\sum_{i=1}^{n_1} (Y_{1i} - \hat{\beta}_0 - \hat{\beta}_1 x_{1i})^2 + \sum_{j=1}^{n_2} (Y_{2j} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{2j})^2 \right] / (n_1 + n_2 - 4)。$$

本章考慮分別推廣 Fieller (1954)及 Kempen 和 Vliet (2000)的研究及引用 Delta 方法，建立 $\theta(x) = \mu_1(x) / \mu_2(x)$ 的信賴區間。並且進一步利用 Delta 方法及 Kempen 和 Vliet (2000)的研究建立 $\theta(x)$ 的信賴束。

2.1 迴歸模式下兩母體平均數比值之條件信賴區間

令

$$Z(x) = \hat{Y}_1(x) - \theta(x)\hat{Y}_2(x)$$

則

$$\begin{aligned} E[Z(x)] &= E[\hat{Y}_1(x) - \theta(x)\hat{Y}_2(x)] \\ &= (\beta_0 + \beta_1 x) - \frac{\beta_0 + \beta_1 x}{\gamma_0 + \gamma_1 x}(\gamma_0 + \gamma_1 x) \\ &= 0, \end{aligned}$$

且

$$\begin{aligned} V_\theta[Z(x)] &= V(\hat{Y}_1(x) - \theta(x)\hat{Y}_2(x)) \\ &= V(\hat{Y}_1(x)) + [\theta(x)]^2 V(\hat{Y}_2(x)), \end{aligned}$$

$$\text{其中 } V[\hat{Y}_i(x)] = \sigma^2 \left[\frac{1}{n_i} + \frac{(x - \bar{x}_i)^2}{\sum_{j=1}^{n_i} (x_{ij} - x_i)^2} \right], \quad i=1,2$$

令 $SS_i = \sum_{j=1}^{n_i} (x_{ij} - x_i)^2$, $i=1,2$ 。以 S^2 估計 σ^2 , 則 $Z(x)$ 的變異數之估計

式為

$$\hat{V}_\theta[Z(x)] = S^2 \left\{ \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right) + \theta^2(x) \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right) \right\}$$

令樞紐量(pivotal quantity)

$$T(x) = \frac{\hat{\beta}_0 + \hat{\beta}_1 x - \theta(x)(\hat{\gamma}_0 + \hat{\gamma}_1 x)}{S \sqrt{\left\{ \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right) + \theta^2(x) \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right) \right\}}}$$

給定 $X = x$ 的情況下， $T(x)$ 服從自由度為 $(n_1 + n_2 - 4)$ 之 t 分布，因此

仿 Fieller(1954) 的求證，在

$$P\{T(x)^2 \leq t_{n_1+n_2-4, \alpha}^2\} = 1 - \alpha$$

之下我們將 $[T(x)]^2 \leq t_{n_1+n_2-4, \alpha}^2$ 改寫成 $\theta(x)$ 的函數如下：

$$\begin{aligned} & \left[(\hat{\beta}_0 + \hat{\beta}_1 x) - \theta(x)(\hat{\gamma}_0 + \hat{\gamma}_1 x) \right]^2 \\ & - t_{n_1+n_2-4, \alpha}^2 S^2 \left\{ \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right) + \theta^2(x) \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right) \right\} \leq 0 \end{aligned}$$

求解上式，得知當 $X = x$ 時， $\theta(x)$ 的 $100(1 - \alpha)\%$ 近似條件信賴區間為

$$(1) \quad \frac{B \pm \sqrt{B^2 - AC}}{A}$$

其中

$$A = (\hat{\gamma}_0 + \hat{\gamma}_1 x)^2 - t_{n_1+n_2-4, \alpha}^2 S^2 \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right),$$

$$B = (\hat{\beta}_0 + \hat{\beta}_1 x)(\hat{\gamma}_0 + \hat{\gamma}_1 x),$$

$$C = (\hat{\beta}_0 + \hat{\beta}_1 x)^2 - t^2_{n_1+n_2-4, \alpha} S^2 \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right)。$$

除推廣 Fieller 方法之外，當 $\mu_1(x)$ 及 $\mu_2(x)$ 皆為正數時，我們考慮先求得 $\eta(x) = \log[\mu_1(x)/\mu_2(x)]$ 的條件信賴區間。令 $\hat{\theta}(x) = \hat{Y}_1(x)/\hat{Y}_2(x)$ 則

$$\log \hat{\theta}(x) = \log \frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} = \log \hat{Y}_1(x) - \log \hat{Y}_2(x)。$$

引用 Delta 方法 (Casella and Berger, 2001) 求得 $\log \hat{\theta}(x)$ 的變異數為

$$\begin{aligned} V[\log \hat{\theta}(x)] &= V[\log \hat{Y}_1(x) - \log \hat{Y}_2(x)] \\ &= \left(\frac{1}{\hat{Y}_1(x)} \right)^2 V[\hat{Y}_1(x)] + \left(\frac{1}{\hat{Y}_2(x)} \right)^2 V[\hat{Y}_2(x)] \end{aligned}$$

則 $V[\log \hat{\theta}(x)]$ 之估計式為

$$\begin{aligned} \hat{V}[\log \hat{\theta}(x)] &= \left(\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x} \right)^2 S^2 \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right) + \left(\frac{1}{\hat{\gamma}_0 + \hat{\gamma}_1 x} \right)^2 S^2 \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right) \\ (2) \quad &= S^2 \left\{ \left(\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x} \right)^2 \left(\frac{1}{n_1} + \frac{(x - \bar{x}_1)^2}{SS_1} \right) + \left(\frac{1}{\hat{\gamma}_0 + \hat{\gamma}_1 x} \right)^2 \left(\frac{1}{n_2} + \frac{(x - \bar{x}_2)^2}{SS_2} \right) \right\} \end{aligned}$$

由 $\log \hat{\theta}(x)$ 之近似分布為常態分布得知

$$(3) \quad T_1(x) = \frac{\log \hat{\theta}(x) - \log \theta(x)}{\sqrt{\hat{V}[\log \hat{\theta}(x)]}} \xrightarrow{d} n(0,1)$$

其中 \xrightarrow{d} 代表分布收斂。因此 $\eta(x)$ 的近似 $100(1-\alpha)\%$ 條件信賴區間為

$$\log \hat{\theta}(x) \pm z_{\alpha/2} \sqrt{\hat{V}[\log \hat{\theta}(x)]}$$

經由轉換可得到 $\theta(x)$ 的 $100(1-\alpha)\%$ 近似條件信賴區間為

$$(4) \quad \exp\left\{\log \hat{\theta}(x) \pm z_{\alpha/2} \sqrt{\hat{V}[\log \hat{\theta}(x)]}\right\}$$

最後，仿 Kempen 和 Vliet (2000) 的研究，針對 $\hat{Y}_1(x)/\hat{Y}_2(x)$ 在 $\mu_1(x)$

及 $\mu_2(x)$ 處進行二階泰勒展開式：

$$\begin{aligned} \frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} &\approx \frac{\mu_1(x)}{\mu_2(x)} + \left[\hat{Y}_1(x) - \mu_1(x) \right] \frac{\partial}{\partial Y_1(x)} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \right) \Bigg|_{\mu_1(x), \mu_2(x)} \\ &\quad + \left[\hat{Y}_2(x) - \mu_2(x) \right] \frac{\partial}{\partial Y_2(x)} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \right) \Bigg|_{\mu_1(x), \mu_2(x)} \\ &\quad + \frac{1}{2} \left[\hat{Y}_1(x) - \mu_1(x) \right]^2 \frac{\partial^2}{\partial Y_1^2(x)} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \right) \Bigg|_{\mu_1(x), \mu_2(x)} \\ &\quad + \frac{1}{2} \left[\hat{Y}_2(x) - \mu_2(x) \right]^2 \frac{\partial^2}{\partial Y_2^2(x)} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \right) \Bigg|_{\mu_1(x), \mu_2(x)} \\ &\quad + \left[\hat{Y}_1(x) - \mu_1(x) \right] \left[\hat{Y}_2(x) - \mu_2(x) \right] \frac{\partial^2}{\partial Y_1(x) \partial Y_2(x)} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \right) \Bigg|_{\mu_1(x), \mu_2(x)} \end{aligned}$$

因此，

$$\begin{aligned} E\left(\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)}\right) &\approx \frac{\mu_1(x)}{\mu_2(x)} + \frac{\mu_1(x)}{\mu_2^3(x)} V[\hat{Y}_2(x)] \\ &= \frac{\mu_1(x)}{\mu_2(x)} \left\{ 1 + \frac{V[\hat{Y}_2(x)]}{\mu_2^2(x)} \right\} \end{aligned}$$

$$\begin{aligned}
(5) \quad V \begin{bmatrix} \hat{Y}_1(x) \\ \hat{Y}_2(x) \end{bmatrix} &= E \left\{ \begin{bmatrix} \hat{Y}_1(x) \\ \hat{Y}_2(x) \end{bmatrix} - E \begin{bmatrix} \hat{Y}_1(x) \\ \hat{Y}_2(x) \end{bmatrix} \right\}^2 \\
&\approx \frac{1}{\mu_2^2(x)} V[\hat{Y}_1(x)] + \frac{\mu_1^2(x)}{\mu_2^4(x)} V[\hat{Y}_2(x)] \\
&= \frac{\mu_1^2(x)}{\mu_2^2(x)} \left\{ \frac{V[\hat{Y}_1(x)]}{\mu_1^2(x)} + \frac{V[\hat{Y}_2(x)]}{\mu_2^2(x)} \right\}
\end{aligned}$$

$$\text{其中 } V[\hat{Y}_i(x)] = \sigma^2 \left[\frac{1}{n_i} + \frac{(x - \bar{x}_i)^2}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \right], \quad i=1,2$$

以 S^2 估計 σ^2 ，則 $\mu_2(x)/\mu_1(x)$ 不偏估計式為

$$\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \left\{ 1 - \frac{V[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)} \right\}$$

由 $\hat{\theta}(x)$ 之近似分布為常態分布得知

$$(6) \quad T_2(x) = \frac{\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \left\{ 1 - \frac{V[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)} \right\} - \frac{\mu_1(x)}{\mu_2(x)}}{\frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \sqrt{\frac{V[\hat{Y}_1(x)]}{\hat{Y}_1^2(x)} + \frac{V[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)}}} \xrightarrow{d} n(0,1)$$

其中 \xrightarrow{d} 代表分布收斂， $V[\hat{Y}_i(x)] = \sigma^2 \left[\frac{1}{n_i} + \frac{(x - \bar{x}_i)^2}{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \right]$ ， $i=1,2$ ，以

S^2 估計 σ^2 。因此，我們可求得 $\theta(x) = \mu_1(x)/\mu_2(x)$ 的近似 $100(1-\alpha)\%$ 條

件信賴區間為

$$\begin{aligned}
(7) \quad & \left\{ \frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} - \frac{1}{n_2} \left(\frac{\hat{Y}_1(x)}{\hat{Y}_2^3(x)} \hat{V}[\hat{Y}_2(x)] \right) \pm z_{\alpha/2} \frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \sqrt{\frac{\hat{V}[\hat{Y}_1(x)]}{\hat{Y}_1^2(x)} + \frac{\hat{V}[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)}} \right\} \\
& = \frac{\hat{Y}_1(x)}{\hat{Y}_2(x)} \left\{ 1 - \frac{1}{n_2} \left(\frac{\hat{V}[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)} \right) \pm z_{\alpha/2} \sqrt{\frac{\hat{V}[\hat{Y}_1(x)]}{\hat{Y}_1^2(x)} + \frac{\hat{V}[\hat{Y}_2(x)]}{\hat{Y}_2^2(x)}} \right\}
\end{aligned}$$

2.2 迴歸模式下之兩母體平均數比值之信賴束

在 2.1 節中曾引用 Delta 方法求出 $\log \hat{\theta}(x)$ 及推廣 Kempen 和 Vliet (2000) 的研究求出 $\hat{\theta}(x)$ 的近似常態分布，然後獲得在 x 已知時， $\theta(x)$ 的近似條件信賴區間。但是，為求得當 $x \in X$ 時 $\theta(x)$ 的信賴束，我們必須進一步求出

$$T_1 = \text{Sup}_{x \in X} T_1(x)$$

及

$$T_2 = \text{Sup}_{x \in X} T_2(x)$$

的近似抽樣分布，其中 $T_1(x)$ 如公式(3)所示， $T_2(x)$ 如公式(6)所示， X 為 x 的可能值所成集合。若能求得 $c_{i\alpha}$ ， $i=1,2$ 值使得

$$P\{T_i \leq c_{i\alpha}\} = 1 - \alpha。$$

當 $c_{1\alpha}$ 求得之後，則可以先求得 $\eta(x) = \log \theta(x)$ 之近似 $100(1-\alpha)\%$ 信賴束，然後轉換獲得 $\theta(x)$ 之近似 $100(1-\alpha)\%$ 信賴束，而 $c_{2\alpha}$ 求得之後，則可求得 $\theta(x)$ 之近似 $100(1-\alpha)\%$ 信賴束。本文應用自助法 (Bootstrapping) 求得 $c_{i\alpha}$ ($i=1,2$) 值，在此雖 T_i ($i=1,2$) 不同，但自助法是一致的，其步驟如下：

步驟一：根據配適迴歸直線的殘差進行重抽樣，其中重抽樣之樣本數與原來樣本數相同。以重抽樣本配適迴歸直線。

步驟二：就每一重抽樣樣本，在 x 值可能範圍 X 中，求出 $T(x)$ 的極大

值，然後據以計算其 T 值，記作 \tilde{T} 。

步驟三：重複步驟一與步驟二 1000 次，可獲得重抽樣樣本的 T 值，

記作 $\tilde{T}_1, \dots, \tilde{T}_{1000}$ 。此時將其自大至小排序為 $\tilde{T}_{(1)} > \dots > \tilde{T}_{(1000)}$ ，

則 c_α 值即為此一排序中第 1000α 位置的 T 值，即 $\tilde{T}_{(1000\alpha)}$ 。

因此， $\{\theta(x), x \in X\}$ 的近似 $100(1-\alpha)\%$ 信賴束如公式(4)及公式(7)，其

中公式(4)中的 $z_{\alpha/2}$ 為 $c_{1\alpha}$ 取代，公式(7)中的 $z_{\alpha/2}$ 為 $c_{2\alpha}$ 取代。

第三章 模擬研究

3.1 模擬方法

本章利用模擬研究本文所提估計 $\theta(x) = \mu_1(x)/\mu_2(x)$ 的三個條件信賴區間及二個信賴束之覆蓋機率及期望長度。其中 $F(1)$ 代表推廣 Filler 方法所得到的條件信賴區間， $D(4)$ 是引用 Delta 方法建立之條件信賴區間， $S(7)$ 則是根據二階泰勒展開式建立之條件信賴區間， BD 是引用 Delta 方法所得的信賴束， BS 則是根據二階泰勒展開式建立之信賴束。兩個迴歸模式中的誤差項為平均數是 0，標準差為 150 的常態隨機變數，解釋變數 X 則為均勻分布於 (10,40) 的隨機變數。

模式如下所述：

$$\begin{aligned} \text{模式 I : } & E(Y_1) = 280 + 90X \\ & E(Y_2) = 280 + 90X \end{aligned}$$

$$\begin{aligned} \text{模式 II : } & E(Y_1) = -250 + 90X \\ & E(Y_2) = 280 + 90X \end{aligned}$$

$$\begin{aligned} \text{模式 III : } & E(Y_1) = -250 + 90X \\ & E(Y_2) = 280 + 40X \end{aligned}$$

$$\begin{aligned} \text{模式 IV : } & E(Y_1) = -250 + 90X \\ & E(Y_2) = 280 + 60X \end{aligned}$$

上述模式 I~IV 對應的迴歸直線示於圖 1。模式 I 中，兩個迴歸模式為相同的，模式 II 中兩條迴歸直線斜率相同，截距不同，模式 III 及模式 IV 則包含兩條斜率及截距皆不同的迴歸模式。不過，模式 IV 中的

兩條迴歸直線為交叉的。模擬考慮的樣本數分別為 10, 30, 50 和 100。

相關的條件信賴區間之 x 值則為 $x=15, 25$ 及 35 。

此模擬中的均勻分布和常態分布之隨機變數是分別利用 R 的副程式庫中的 RUNIF 及 RNORM 生成。此一模擬重複 10000 次，每一次重複中計算本文所提的各種近似 95% 條件信賴區間及信賴束。若此一信賴區間或信賴束包含真實迴歸平均數比值，則視為成功覆蓋一次，以成功覆蓋相對次數，估計此一信賴區間的覆蓋機率。在 10000 次模擬下，此一覆蓋機率估計之標準誤差約為

$\sqrt{0.95 \times 0.05 / 10000} \approx 0.002$ 。就條件信賴區間而言，每一次重複計算所得之區間長度，然後以 10000 個區間的平均值，估計區間的期望長度。而就信賴束而言，在 1000 次模擬下，其成功覆蓋相對次數或覆蓋機率估計的標準誤差約為 $\sqrt{0.95 \times 0.05 / 1000} \approx 0.0069$ ，此時計算信賴束內的面積，估計其期望長度。所得結果列示於表 1 及表 2。

3.2 模擬結果

推廣 Fieller (1954)的方法所建立 95%條件信賴區間覆蓋機率大致都接近 0.95，而引用 Delta 方法及根據二階泰勒展開式所建立 95%條件信賴區間樣本數則必須 30 以上的時候覆蓋機率才會接近 0.95，但根據二階泰勒展開式所建立 95%條件信賴區間的平均長度是較窄的。大致上來說，推廣 Fieller (1954)的方法所建立 95%條件信賴區間的平均長度都比量兩種方法所建立 95%條件信賴區間的平均長度寬，而根據二階泰勒展開式所建立 95%條件信賴區間的平均長度是較窄的。

引用 Delta 方法及根據二階泰勒展開式所建立 95%條件信賴束在樣本數為 50 以上所得到的覆蓋機率都接近 0.95，但是就 95%條件信賴束的平均長度而言，在模式 I 及 II 之下，此兩種方法的平均長度是差不多的。但是在模式 I 及 II 之下，即兩條迴歸直線的斜率截距都不同的模式，根據二階泰勒展開式所建立 95%條件信賴束的平均長度都比引用 Delta 方法所建立 95%條件信賴束的平均長度窄。

第四章 實例分析

本文分析一筆研究男女性體內水分與鉀離子含量關係的資料 (Daniel, 1987)。資料中收集 6 歲到 18 歲 13 位男性及 14 位女性的水分總重(公升, L)與鉀離子總量(mEq)，如表三所示。

鉀離子是人體體液中與新陳代謝有關的電解質，要維持人體代謝功能正常的運作便需要適當的攝取鉀離子。鉀離子為人體細胞內液主要的陽離子。成年男性全身鉀離子總量約 3500mEq，血中濃度約為 3.5-4.8mEq/L，細胞內濃度為 150 mEq/L。事實上，鉀離子在人體的功能除幫助調節細胞內的滲透性，亦為細胞生長及代謝所必需（主要影響醣類代謝及蛋白質合成過程），更有助於神經衝動的傳導並且維持適當的骨骼肌、心肌及平滑肌的功能及人體酸鹼平衡的維持。

在青少年中，常患一種四肢和軀幹肌肉周期性發作鬆弛性癱瘓的疾病。這種病突然發生，經數小時或數天可恢復至健康人一樣，但過一段時間後又可復發，呈周期性癱瘓，因而得名。本病的病因尚未明瞭，可能與肌肉纖維的鉀離子代謝障礙有關，少數病例有家族遺傳史。周期性麻痺可分為低血鉀、高血鉀和正常血鉀三類，最常見的是低血鉀周期性麻痺，化驗時血液內鉀離子濃度降低，得病者以男青年居多。少數的病例為高血鉀和正常血鉀周期性麻痺，病者也以男性居多，其中高血鉀麻痺的發病年齡多在 10 歲以前，大多在 30 歲後趨向

好轉和逐漸中止。

針對此筆資料分別對男性與女性配適簡單線性迴歸模式得到估計的迴歸直線(圖 2)分別為

$$\text{男性： } \hat{y}_1 = -253.01 + 93.64x$$

及

$$\text{女性： } \hat{y}_2 = 280.42 + 60.71x$$

判定係數 R^2 分別為 0.9591 及 0.9031，估計的標準差則分別是 167 及 135。根據兩組殘差的 F 檢定(p 值=0.2286，統計量=1.2622)，結果顯示，資料無顯著證據指出兩條迴歸模式的誤差項變異數不同。此外，由殘差圖及常態機率圖(圖 3~圖 4)及 Shapiro-Wilk 檢定(男性： p 值=0.4053，統計量=0.9358，女性： p 值=0.1142，統計量=0.9004)結果得知，可以合理假設資料服從常態分布。

進一步求得體內具相同水分的男女性其鉀離子平均含量的 95% 信賴區間(圖 5)。正如一般預期的，95% 信賴束均比另外三個 95% 條件信賴區間寬，但推廣 Kempen 和 Vliet (2000) 的研究建立的信賴束比引用 Delta 方法所建立的信賴束窄。圖 5 更指出體內水分總量超過 20 公升的男性比具有相同水分的女性，體內含有顯著更多的鉀離子。當水分 20 公升時，男性與女性分別的平均鉀離子含量為 1619.813 mEq 及 1494.694 mEq。所以，此結果與醫學文獻所載，週期性高血

鉀麻痺患者以男性青年居多一致。

第五章 結論及未來可行之研究

若樣本數小的時候，本文建議採用推廣 Fieller (1954)的方法求得兩條迴歸直線比值的條件信賴區間。若是樣本數在 50 以上，則建議使用推廣 Kempen 和 Vliet (2000)的研究求得兩條迴歸直線比值的條件信賴區間。信賴束方面，則建議使用推廣 Kempen 和 Vliet (2000)的研究求得兩條迴歸直線比值的信賴束。

本文只就一個共變數之資料，建立兩母體平均數比值之 $100(1-\alpha)\%$ 條件信賴區間及信賴束。未來的可能研究，可考慮多個共變數的資料，建立兩母體平均數比值之 $100(1-\alpha)\%$ 條件信賴區間與信賴束。未來也可以就多個母體的資料，建立聯合的條件信賴域或聯合信賴束進行多個母體之間的比較。

參考文獻

1. Casella G. and Berger R.L. (2001). *Statistical Inference*. Duxbury.
2. Daniel, W. W. (1987). *Biostatistics: A foundation for Analysis in the Health Sciences*. John Wiley & Sons.
3. Draper, N. R. and Smith H. (1998). *Applied Regression Analysis*. John Wiley & Sons.
4. Fieller, E.C. (1932). The Distribution of the Index in a Normal Bivariate Population. *Biometrika* **24**, 428-440
5. Fieller, E.C. (1954). Some Problems in Interval Estimation. *Journal of the Royal Statistical Society. Series B* **16**,175-185.
6. van Kempen, G.M.P. and van Vliet, L.J. (2000). Mean and Variance of the Ratio Estimators Used in Fluorescence Ratio Imaging. *Cytometry* **39**, 300-305.

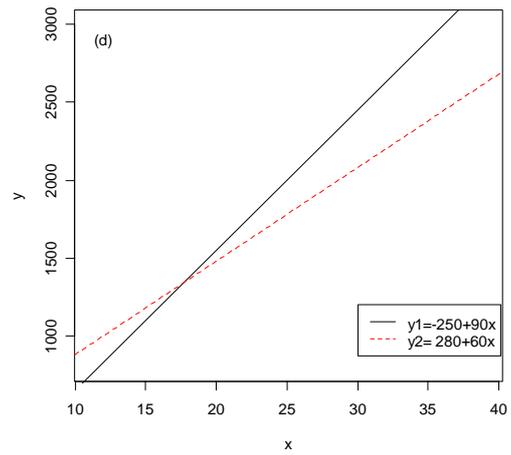
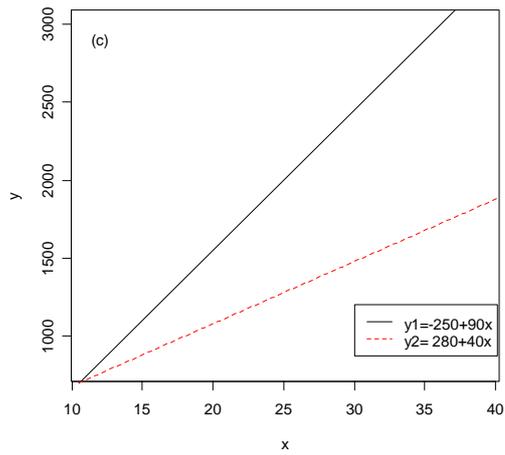
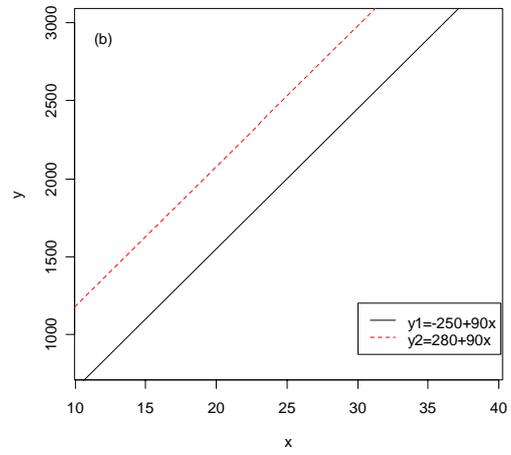
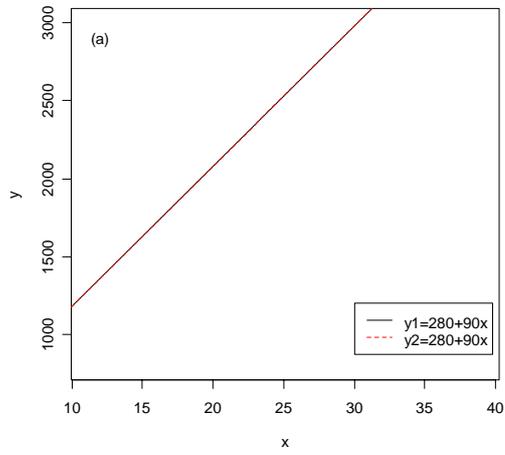


圖 1：模擬研究中之迴歸直線。子圖(a)-(d)分別為迴歸模式 I~IV 中的兩條迴歸直線

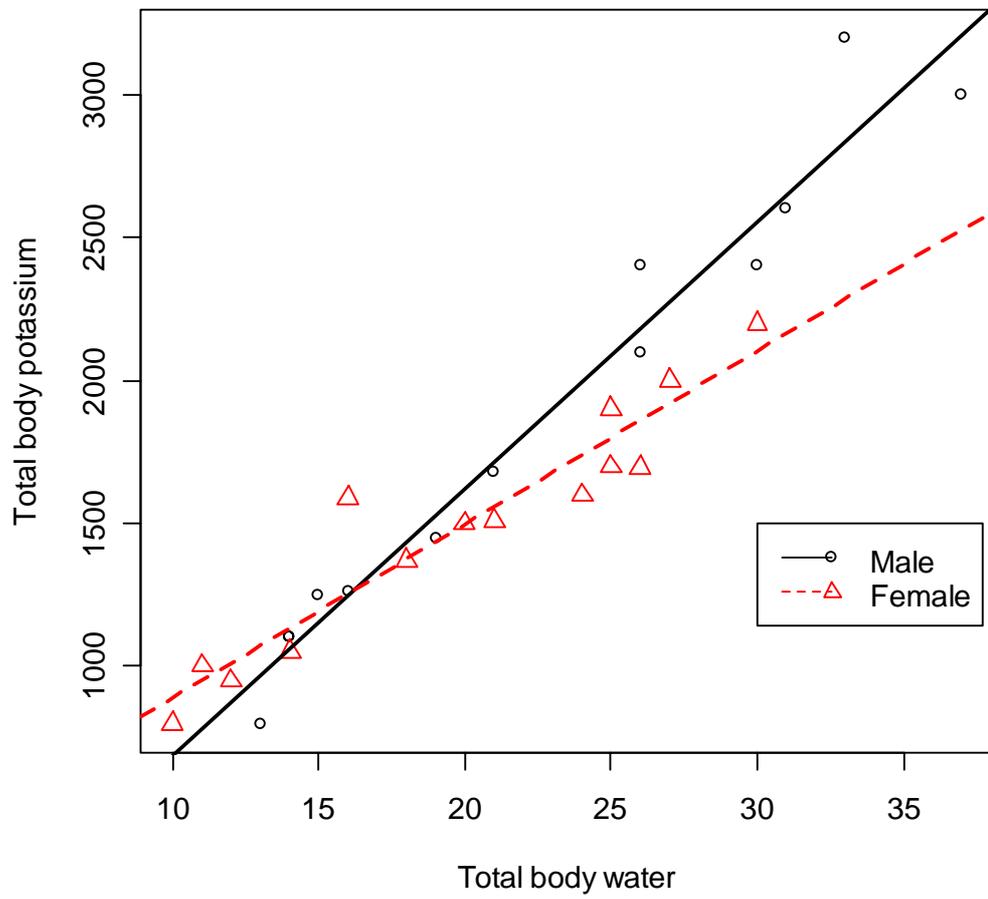


圖 2：男女性體內鉀離子含量與總水重的散佈圖及所估計之迴歸直線

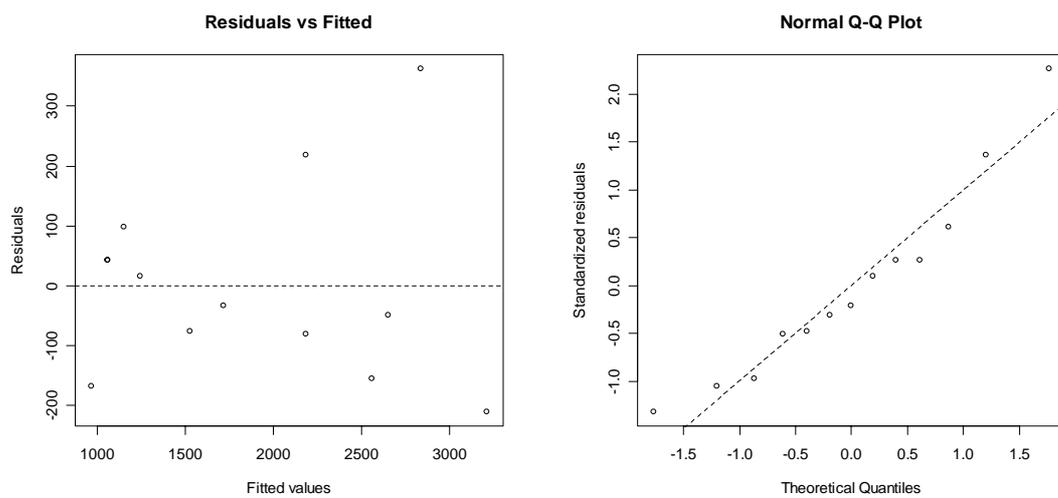


圖 3：男性之體內鉀離子含量相對於總水重迴歸直線殘差圖及其常態機率圖。左方所示為殘差圖，右方則為常態機率圖。

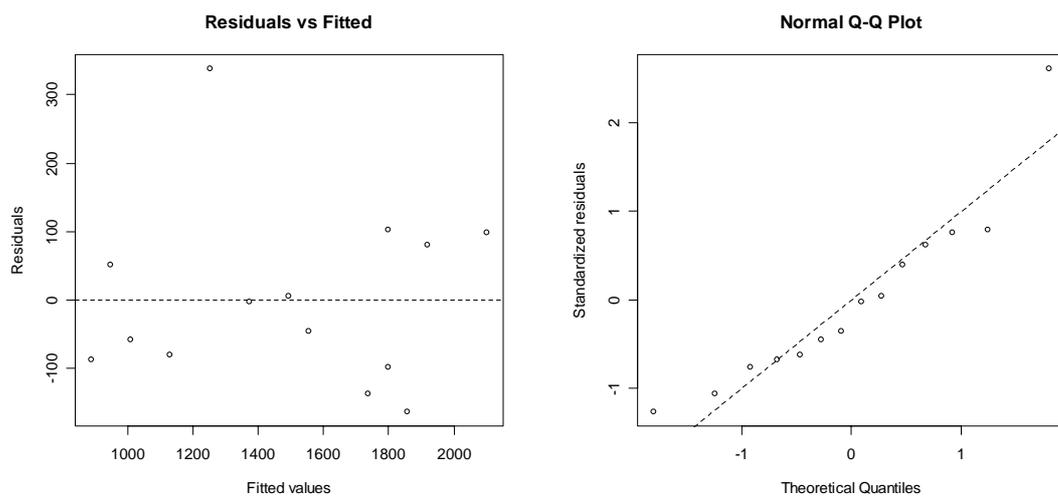


圖 4：女性之體內鉀離子含量相對於總水重迴歸直線殘差圖及其常態機率圖。左方所示為殘差圖，右方則為常態機率圖。

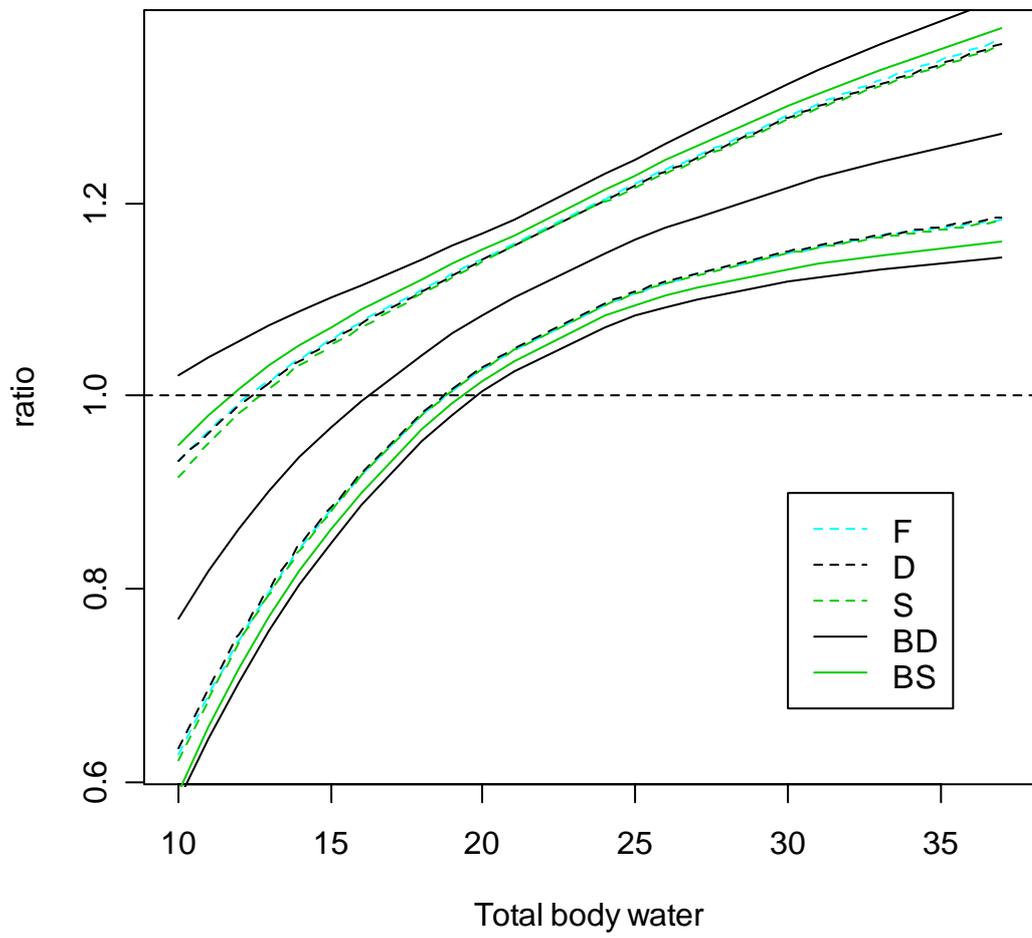


圖 5: 男女性鉀離子平均含量比值的 95% 條件信賴區間及 95% 信賴束。

表 1：兩條迴歸直線比値之 95% 條件信賴區間之覆蓋機率(期望長度)

模型	樣本數	$x = 15$			$x = 25$			$x = 35$		
		F	D	S	F	D	S	F	D	S
I	10	0.9534 (0.2843)	0.9382 (0.2577)	0.9374 (0.2585)	0.9512 (0.1092)	0.9343 (0.0992)	0.9343 (0.0993)	0.9508 (0.1093)	0.9328 (0.0992)	0.9318 (0.0991)
	30	0.9467 (0.1250)	0.9408 (0.1214)	0.9420 (0.1216)	0.9439 (0.0589)	0.9380 (0.0573)	0.9379 (0.0573)	0.9509 (0.0712)	0.9459 (0.0693)	0.9457 (0.0693)
	50	0.9505 (0.1035)	0.9476 (0.1018)	0.9474 (0.1018)	0.9488 (0.0449)	0.9465 (0.0442)	0.9465 (0.0442)	0.9501 (0.0515)	0.9475 (0.0507)	0.9484 (0.0507)
	100	0.9513 (0.0746)	0.9504 (0.0741)	0.9506 (0.0741)	0.9525 (0.0316)	0.9516 (0.0313)	0.9528 (0.0314)	0.9493 (0.0351)	0.9479 (0.0348)	0.9473 (0.0348)
II	10	0.9698 (0.2573)	0.9374 (0.3106)	0.9367 (0.2094)	0.9529 (0.0991)	0.9325 (0.1123)	0.9325 (0.0888)	0.9503 (0.1012)	0.9311 (0.1084)	0.9315 (0.0917)
	30	0.9421 (0.1051)	0.9414 (0.1552)	0.9413 (0.1046)	0.9455 (0.0533)	0.9391 (0.0652)	0.9383 (0.0515)	0.9538 (0.0665)	0.9463 (0.0755)	0.9455 (0.0637)
	50	0.9641 (0.0921)	0.9503 (0.1253)	0.9501 (0.0845)	0.9481 (0.0404)	0.9464 (0.0505)	0.9458 (0.0399)	0.9495 (0.0475)	0.9487 (0.0558)	0.9494 (0.0472)
	100	0.9576 (0.0650)	0.9496 (0.0922)	0.9493 (0.0622)	0.9527 (0.0285)	0.9515 (0.0357)	0.9509 (0.0282)	0.9458 (0.0322)	0.9473 (0.0383)	0.9465 (0.0324)
III	10	0.9412 (0.5916)	0.9402 (0.4509)	0.9340 (0.5713)	0.9484 (0.2817)	0.9360 (0.1669)	0.9360 (0.2616)	0.9502 (0.3170)	0.9354 (0.1664)	0.9329 (0.2874)
	30	0.9488 (0.2659)	0.9421 (0.2032)	0.9431 (0.2547)	0.9423 (0.1520)	0.9401 (0.0957)	0.9405 (0.1497)	0.9392 (0.1995)	0.9429 (0.1175)	0.9419 (0.2029)
	50	0.9412 (0.2122)	0.9454 (0.1737)	0.9456 (0.2177)	0.9501 (0.117)	0.9467 (0.0733)	0.9466 (0.1145)	0.9585 (0.1517)	0.9510 (0.0839)	0.9508 (0.1448)
	100	0.9483 (0.1550)	0.9517 (0.1256)	0.9510 (0.1571)	0.9525 (0.0819)	0.9508 (0.0520)	0.9513 (0.0813)	0.9574 (0.1037)	0.9485 (0.0571)	0.9478 (0.0985)
IV	10	0.9569 (0.3883)	0.9384 (0.3682)	0.9376 (0.3453)	0.9502 (0.1649)	0.9348 (0.1340)	0.9358 (0.1509)	0.9518 (0.1762)	0.9331 (0.1309)	0.9341 (0.1599)
	30	0.9449 (0.1667)	0.9410 (0.1745)	0.9421 (0.1629)	0.9420 (0.0888)	0.9390 (0.0773)	0.9375 (0.0867)	0.9461 (0.1131)	0.9443 (0.0919)	0.9445 (0.1120)
	50	0.9542 (0.1397)	0.9491 (0.1454)	0.9472 (0.1354)	0.9482 (0.0679)	0.9452 (0.0594)	0.9453 (0.0668)	0.9544 (0.0835)	0.9495 (0.0667)	0.9491 (0.0812)
	100	0.9525 (0.1001)	0.9506 (0.1059)	0.9500 (0.0987)	0.9531 (0.0478)	0.9519 (0.0422)	0.9516 (0.0474)	0.9529 (0.0570)	0.9485 (0.0456)	0.9480 (0.0555)

表 2：兩條迴歸直線比値之 95%信賴束之覆蓋機率(期望長度)

模型	樣本數	BD	BS
I	10	0.941(0.3539)	0.941 (0.2768)
	30	0.954(0.2143)	0.954 (0.2024)
	50	0.943(0.1790)	0.943(0.1728)
	100	0.950(0.1430)	0.950 (0.1401)
II	10	0.943(0.3446)	0.943 (0.2656)
	30	0.953(0.1973)	0.952(0.1982)
	50	0.948(0.1631)	0.949(0.1699)
	100	0.946(0.1279)	0.944(0.1376)
III	10	0.948(0.8152)	0.948(0.6482)
	30	0.948(0.5587)	0.951(0.4651)
	50	0.947(0.4662)	0.947(0.3875)
	100	0.965(0.3487)	0.965(0.3100)
IV	10	0.941(0.6903)	0.941(0.4104)
	30	0.952(0.3232)	0.952(0.2948)
	50	0.945(0.2605)	0.947(0.2500)
	100	0.954(0.2053)	0.956(0.2018)

表 3：男女性體內水分與鉀離子含量

男性		女性	
鉀離子總量	水分總重	鉀離子總量	水分總重
795	13	1590	16
1250	15	800	10
1680	21	1700	25
2100	26	1370	18
1260	16	1000	11
1100	14	1500	20
1450	19	950	12
1100	14	1600	24
2400	26	1695	26
2400	30	1510	21
3200	33	2000	27
2600	31	1050	14
3000	37	1900	25
		2200	30