JPPS

ISSN 1607-7083

統計薪傳

一份有趣、有用、有創意、有人性之全方位統計期刊

JOURNAL OF PROPAGATIONS IN PROBABILITY AND STATISTICS

Volume 17　Number 2　December 2017

第十七卷　第二期　中華民國一百零六年十二月

統領世紀　薪傳天下

統 計 薪 傳

JPPS ISSN 1607-7083

JOURNAL OF PROPAGATIONS IN PROBABILITY AND STATISTICS

Aims and Scope: ***JPPS***, published semiannually in June and December (in February and August before 2003), is a multipurpose, comprehensive statistics journal that publishes articles of general interest to a broad audience of professionals, practicians, teachers, students, and any other users of statistics. All articles related to statistics or probability are welcome, including original research and review papers, notes on specific problems, technical and statistical practice reports, survey reports on public opinion or market research, essays on teaching methodology or experiences and the history of statistics, biographies of prominent statisticians, etc. It is our goal to disseminate the science of probability and/or statistics and its applications through the publication of ***JPPS***.

Publication Date of First Issue: August 1, 2000

Submission and Review Policies:

1) **PDF** document of the manuscript should be mailed to the **Editor-in-Chief** or **Managing Editor** through an email attachment. Similar to the *Canadian Journal of Statistics*, this journal welcomes articles written **either in English or in Chinese**.
2) A manuscript is accepted only with the understanding that the text has not appeared in any other publication, and that it is not being simultaneously reviewed by another journal.
3) Submitted manuscripts are reviewed by a mutually blind process, meaning that the reviewers will not know the names of the authors and vice versa.
4) If an article is approved for publication, the author(s) will be asked to provide an electronic copy of the paper, in **Micro-Soft WORD**, through an email attachment. The authors will also be required to transfer their copyright on certain conditions to Susan Rivers' Cultural Institute, Hsinchu City, Taiwan, ROC.

Last updated: **December 1, 2017**

JPPS ISSN 1607-7083

統 計 薪 傳

JOURNAL OF PROPAGATIONS IN PROBABILITY AND STATISTICS

宗 旨 本刊物為一綜合性多元取向之統計期刊，內容涵蓋與機率或統計有關之學術研究、技術報告、教學經驗與心得、問題探討、實務應用、人物介紹與專訪、書評書介、市調民調、就業經驗以及大專學生或研究生之研究報告與學習心得等等不一而足。讀者與邀稿對象，上至學者專家、政府官員或企業主管，下至大專學生與社會大眾。機率與統計是應用廣泛的科學，隨著社會的日新月異與進步，它們的重要性愈形增加，每個人在日常生活中都可能遭遇和機率或統計有關的事物。藉由本期刊之發行, 我們傳播機率與統計的知識與常識，使它們能更普遍化、大眾化，促進社會的更進步，而前人之經驗與成就，亦能薪火傳承，並發揚光大。

創刊年月 公元 2000 年 8 月

JPPS **ISSN 1607-7083**

統　計　薪　傳

***J**OURNAL OF **P**ROPAGATIONS IN **P**ROBABILITY AND **S**TATISTICS*

投稿須知

本期刊登載與統計或機率有關之各類文章，來稿請作者儘量自行事先歸類，如學術論文、應用文摘、教學心得、書評書介、散文雜記等等。若有特定之讀者對象(如高中生、大專生、研究生等)，亦請註明。稿件將送請學者專家雙向隱名審閱，審閱通過後，請作者依本期刊最近一期之刊登格式以 **Microsoft Word** 自行打字排版，再以電子郵件附加檔寄送本期刊總編輯，以利編輯作業。其他注意事項如下:

1. 來稿文字應流暢精確，以電子郵件附加 **PDF 檔**投稿。
2. 較學術或專技性文稿請儘量附摘要(中文及英文)、關鍵字詞與參考文獻。
3. 翻譯或轉載稿件請附原文及原著作所有權人同意授權書。
4. 來稿請註明作者姓名、地址、服務機關或就讀學校、系所與年級，歡迎提供作者之重要學經歷。
5. 本刊對來稿內容中之次要文句有修飾權，未能刊登稿件恕不退還。
6. 審核通過並刊登於本期刊之稿件，其出版權歸魏蘇珊文教事業機構所有。
7. 刊登之文章格式大致如下：
   (a)中文文字部份，第一頁之題目與作者姓名請用標楷體，大小分別為 18 與 15；摘要、關鍵字詞、及作者簡介請用新細明體，大小依次分別為 11.5、11.5、及 11；正文之字體請用新細明體 12。英文請一律使用 Times New Roman 體。
   (b)打字請採橫式單欄，每列間隔以固定行高 18 pt 為原則，用紙以 A4 規格為準。
   (c)參考文獻中文部份請依姓氏筆劃列於前，英文部份請依作者姓氏字母先後列於後。期刊名稱請儘量用全名及斜體，例如 JASA 之全名為 *Journal of the American Statistical Association*。
8. 來稿請寄本期刊之總編輯(或客席總編輯)。

總編輯　**盧宏益**　輔仁大學統計資訊系副教授　電子信箱: **069201@mail.fju.edu.tw**
客席總編輯　**陳宇宏**　展欣科技企業有限公司負責人　電子信箱: **techcom5054@hotmail.com**

發行暨編輯總監　**張光昭**　輔仁大學統計資訊系教授　電子信箱: **stat1016@mail.fju.edu.tw**
創辦人: 張光昭　前輔仁大學夜間部暨進修部統計系系主任(1991-1996, 1997-2001)
創刊年月: 公元 2000 年 8 月　**創刊發行單位**: 輔仁大學進修部統計系
發行次數: 每年出刊兩次(6 月與 12 月)(2003 年之前: 2 月與 8 月)
發行單位: 魏蘇珊文教事業機構/總公司: 新竹市建美路 2 巷 26 號/電話: (03)5716594
發行人: 陳啟興　魏蘇珊文教事業機構負責人　**創刊發行人**: 林吉基　前輔仁大學進修部部主任
創刊發行顧問: 呂漁亭　滕允中　前輔仁大學夜間部(進修部)部主任
電腦排版顧問: 鄭志強　國立中山大學電機系教授
封面畫作原創人: 何若蘭　中華心靈美全民推展協會理事長
零售價: 新台幣 300 元整(長期或大量訂購另有優待價)
創刊印刷者: 宏章彩色製版有限公司(台北縣中和市中山路三段 110 號 3 樓/電話: 02-82214567)

JPPS **ISSN 1607-7083**

統　計　薪　傳

Journal of Propagations in Probability and Statistics

Volume 17　Number 2　December 2017

第十七卷　第二期　　中華民國 106 年 12 月

Table of Contents / 目 次

© Published by: Susan Rivers' Cultural Institute, Hsinchu City, Taiwan, ROC（魏蘇珊文教事業機構 發行）

Journal of Propagations in Probability and Statistics
17(2), 97-104 Dec. 2017

統 計 薪 傳
第十七卷 第二期
民國106年12月97-104頁

關於抽樣調查之比值與迴歸估計的一堂簡短入門課程

陳宇宏
展欣科技企業公司

張光昭
輔仁大學

摘 要 在這一篇教學短文裡，作者們淺介抽樣理論中的比值與迴歸估計這兩種方法，並搭配一些有趣且易懂的計算例題，以期對於剛起步學習統計抽樣與調查方法的莘莘學子們有所助益。

關鍵字詞 最小平方估計量、比值估計、迴歸估計、樣本相關系數、簡單隨機抽樣。

□民國一百零六年六月收稿，一百零六年八月修訂、九月定稿。
□本文第一作者為展欣科技企業有限公司負責人，電子郵址: techcom5054@hotmail.com；第二作者為輔仁大學統計資訊學系專任教授；電子郵址: stat1016@mail.fju.edu.tw。

英文摘要/ English Abstract

A Short Introductory Lesson on Ratio/Regression Estimation in Survey Sampling

Ardor Chen
Techcom Information Corp.

Kuang-Chao Chang
Fu Jen Catholic University

ABSTRACT In this short article, we introduce the basic concepts and methods of ratio and regression estimation in the theory of survey sampling. In particular, we present some interesting numerical examples to illustrate the usages of the ratio and regression estimation methods. We hope the contents of this article can be useful to students who are beginners in learning survey sampling.

Keywords Least squares estimator; Ratio estimation; Regression estimation; Sample correlation coefficient; Simple random sampling.

□Received June 2017, revised August 2017, in final form September 2017.
□Ardor Chen is the founder and CEO of Techcom Information Corp., Taipei, Taiwan, ROC; email: techcom5054@hotmail.com. Kuang-Chao Chang is a Professor in the Department of Statistics and Information Science at Fu Jen Catholic University, Hsinchuang, New Taipei City, Taiwan, ROC; email: stat1016@mail.fju.edu.tw.

 JPPS: ISSN 1607-7083

1. 前言

在抽樣調查的理論中，比值估計(ratio estimation)是一種重要的推估方法，幾乎任何有關抽樣理論與調查方法的書籍，都會談到此一方法(例如: Scheaffer *et al*. 2012 一書第六章，或是 Cochran 1977 一書第六章)。此外，另有一種與比值估計頗爲類似且具有異曲同工之妙的推估方法，稱之爲迴歸估計(regression estimation)，也可算是抽樣調查的理論中的一種重要推估方法。此二種推估方法適用於雙變數的有限母體(finite population)。

在這一篇教學短文裡，作者們藉由 Scheaffer *et al*. 2012 一書第六章的一個計算例題，來介紹比值估計的基本概念。隨後，再搭配另一個人工有限母體的計算例題，來進一步說明比值與迴歸估計的用法。在第二節，作者們介紹比值估計的基本概念與理論。在第三節，作者們介紹迴歸估計的基本概念與理論。在第四節，作者們於幾個計算例題之中實際演算比值與迴歸估計量，並於第三個計算例題之中與簡單隨機抽樣之下的樣本平均數相互比較。第五節爲本文的總結。

2. 比值估計

抽樣調查的比值估計，是適用於雙變數有限母體的一種推估參數的方法。假設某一有限母體具有雙變數: 研究變數(study variable)以及輔助變數(auxiliary variable)，我們以 Y 來表示研究變數、X 表示輔助變數。假設這個母體的母體大小(population size)爲 N，也就是說，母體之中總共含有 N 個元素(element)；那麼母體之中的第 i 個元素的兩個變數值就可以用數對 (y_i, x_i) 來表示，$i = 1, 2, \cdots, N$。所以，兩個變數之母體平均數(population mean)的數學式可分別表示如下:

$$\text{研究變數 } Y \text{ 的母體平均數} = \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i \text{ ，}$$

$$\text{輔助變數 } X \text{ 的母體平均數} = \mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ 。}$$

其次，兩個變數之母體總和數(population total)的數學式可分別表示如下:

$$\text{研究變數 } Y \text{ 的母體總和數} = \tau_y = N\mu_y = \sum_{i=1}^{N} y_i \text{ ，}$$

$$\text{輔助變數 } X \text{ 的母體總和數} = \tau_x = N\mu_x = \sum_{i=1}^{N} x_i \text{ 。}$$

在比值估計的方法之中，有一個重要的母體參數，可稱爲比值參數(以 R 表示此參數)，它是兩個變數之母體平均數(或是母體總和數)的比值，如下所示:

$$R = \frac{\mu_y}{\mu_x} = \frac{\tau_y}{\tau_x} \text{ 。} \tag{1}$$

比值參數 R 通常是一個未知其值的母體參數，也是抽樣調查研究者想要估計的參數。假設我們使用簡單隨機抽樣(simple random sampling)從以上的雙變數母體之中抽出 n 個樣本，其樣本觀測值爲 (y_i, x_i) ，$i = 1, 2, \cdots, n$，那麼想要估計比值參數 R，就十分容易，只要將(1)式之中的

兩個變數之母體平均數換成樣本平均數即可，如下所示:

$$\text{比值參數 } R \text{ 的估計量} = \hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{(1/n)\sum_{i=1}^{n} y_i}{(1/n)\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \text{ 。} \tag{2}$$

以上比值參數 R 的估計量 $\hat{R}$ 是否能夠準確地估計 R，要取決於 Y 與 X 這兩個變數之間是否具有高度的正相關性，若正相關性很高，譬如說相關係數大於 0.9，那麼 $\hat{R}$ 就會是一個相當準確的估計量；但若相關性不夠高，譬如說相關係數小於 0.5，那麼就還是不要使用 $\hat{R}$ 爲妙；如果 Y 與 X 這兩個變數之間不具有正相關性，卻具有負相關性，那可就千千萬萬不能使用 $\hat{R}$ 這個估計量啦！此外，估計量 $\hat{R}$ 還有一個理論上的小缺點: 它是一個具有偏差的估計量(biased estimator)；不過，只要 Y 與 X 這兩個變數之間具有高度的正相關性，那麼這個小缺點的負面影響也就很小。

(2)式之中估計量 $\hat{R}$ 的一個主要用途，就是可以被間接地用來估計研究變數 Y 的母體平均數 μ_y 或是 Y 的母體總和數 τ_y 。我們首先將(1)式改寫成爲以下二式:

$$\mu_y = R\mu_x \quad \text{與} \quad \tau_Y = R\tau_x \text{ 。} \tag{3 與 4}$$

接著，如果我們假設 τ_x 是一個已知其値的母體參數，那麼只要將(3)式之中等號右邊的比値參數 R 換成它的估計量 $\hat{R}$ ，即可得到等號左邊之參數 μ_y 的估計量，如下所示:

$$\hat{\mu}_y = \hat{R}\mu_x \text{ 。} \tag{5}$$

以上(5)式之中的估計量 $\hat{\mu}_y$ 就是赫赫有名的**比值估計量**(ratio estimator)。同理，如果我們假設 τ_x 是一個已知其值的母體參數，那麼只要將(4)式之中等號右邊的比值參數 R 換成它的估計量 $\hat{R}$，即可得到等號左邊之參數 τ_y 的比值估計量如下:

$$\hat{\tau}_y = \hat{R}\tau_x \text{ 。} \tag{6}$$

以上(6)式的估計量 $\hat{\tau}_y$ 將被用於第四節的例題 4.1之中。

3. 迴歸估計

如同前一節的比值估計，迴歸估計也是一種適用於雙變數有限母體的推估方法。從線性迴歸的觀點來看待比值估計，研究變數和輔助變數之間的線性關係就像是一條通過原點之直線。那麼，如果將“通過原點之直線”的概念延伸至“未必通過原點之直線”，原先的比值估計就可能不太適用嘍！因此，取而代之的方法，就是迴歸估計啦！在與先前比值估計的相同假設條件之下，參數 μ_y 的**迴歸估計量**(regression estimator)爲

$$\hat{\mu}_{y\mathrm{L}} = \bar{y} + \hat{\beta}(\mu_x - \bar{x}) \text{ ，} \tag{7}$$

而參數 τ_y 的迴歸估計量爲

$$\hat{\tau}_{y\mathrm{L}} = N\hat{\mu}_y = N[\bar{y} + \hat{\beta}(\mu_x - \bar{x})] \text{ 。}$$

以上兩個估計式之中的 $\hat{\beta}$ 就是迴歸分析課程中赫赫有名的最小平方估計量(least squares estimator)，其數學式爲

$$\hat{\beta}=\frac{s_{xy}}{s_x^2}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} ,$$

其中

$$s_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) \quad 以及 \quad s_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 。$$

以上(7)式的估計量 $\hat{\mu}_{yL}$ 將被用於第四節的例題 4.3 之中，而該估計量之符號的下標之中的 L 代表 Linear Regression 之意。

4. 計算範例

在 Scheaffer *et al*. 2012 一書第六章的 176 頁，有一個關於比值估計的計算例題，十分有趣! 我們將這個例題翻譯爲中文之後稍加修飾與改寫，並加強解答部份的計算與說明。

例題 4.1 某水果商訂購了一卡車的柳丁，並打算採用比值估計之方法，來推估整輛卡車裝載之所有柳丁(如果全數被切開之後再榨取)的柳丁汁總重量。於是，該水果商採用不置回的簡單隨機抽樣之方法從整輛卡車的柳丁之中抽出了 10 顆柳丁。試說明該水果商應該如何利用被抽出的 10 顆柳丁來取得樣本數據以及如何進行比值估計，並寫出點估計量的數學式。

解答 從被抽出的10顆柳丁，水果商可用磅秤取得數對型態的樣本數據如下:

$$\{(y_i,\ x_i)\mid i=1,\ldots,10\} ,$$

其中 x_i 爲第 i 顆柳丁的總重量， y_i 爲第 i 顆柳丁被切開之後榨取的柳丁汁重量。接著，令

τ_y = 整輛卡車裝載所有的柳丁如果全數被切開之後榨取的柳丁汁總重量；

τ_x = 整輛卡車裝載所有尚未被切開之柳丁的總重量
= (裝滿柳丁的整輛卡車停在地磅之上秤得之重量) − (卡車的空車重量)。

那麼，τ_x 可被視爲一個已知其值的母體參數(亦即，輔助變數的母體總和數)。於是，依據第二節之中的(6)式，τ_y 的比值估計值爲

$$\bar{\tau}_y=(\frac{\bar{y}}{\bar{x}})\tau_x=\left(\frac{\sum_{i=1}^{10}y_i}{\sum_{i=1}^{10}x_i}\right)\tau_x ,$$

其中 $\bar{y}=\frac{1}{10}\sum_{i=1}^{10}y_i$ 以及 $\bar{x}=\frac{1}{10}\sum_{i=1}^{10}x_i$ 。 □

例題 4.2 假設前一例題解答之中數對型態的樣本數據如下表所示: (重量單位: 磅)

i	1	2	3	4	5	6	7	8	9	10
y_i	0.021	0.030	0.025	0.022	0.033	0.027	0.019	0.021	0.023	0.025
x_i	0.40	0.48	0.43	0.42	0.50	0.46	0.39	0.41	0.42	0.44

同時

$$\tau_x = \text{整輛卡車裝載所有尚未被切開之柳丁的總重量} = 1800 \text{ 磅。}$$

試求 τ_y 的比值估計值，並計算 y 與 x 的樣本相關係數(sample correlation coefficient)，r_{xy}，以便判斷採用比值估計之方法是否恰當。

解答　τ_y 的比值估計值為

$$\bar{\tau}_y = (\frac{\bar{y}}{\bar{x}})\tau_x = \frac{0.0246}{0.4350}\cdot(1800) \approx 101.79 \text{ 磅，}$$

其中

$$\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = \frac{1}{10}(0.021 + 0.030 + \cdots + 0.025) = 0.0246$$

以及

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(0.40 + 0.48 + \cdots + 0.44) = 0.4350 \text{ 。}$$

接著，樣本相關係數的計算如下:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)}} = \frac{\left(\sum_{i=1}^{n}(x_i y_i)\right) - n(\bar{x}\cdot\bar{y})}{\sqrt{\left[\left(\sum_{i=1}^{n}x_i^2\right) - n(\bar{x})^2\right]\left[\left(\sum_{i=1}^{n}y_i^2\right) - n(\bar{y})^2\right]}}$$

$$= \frac{0.10839 - (10)(0.4350)(0.0246)}{\sqrt{\left[1.9035 - (10)(0.4350)^2\right]\left[0.006224 - (10)(0.0246)^2\right]}}$$

$$= \frac{0.00138}{\sqrt{(0.01125)(0.0001724)}} \approx 0.991 \text{ ，}$$

其中

$$\sum_{i=1}^{n} y_i^2 = (0.021)^2 + (0.030)^2 + \cdots + (0.025)^2 = 0.006224 \text{ ，}$$

$$\sum_{i=1}^{n} x_i^2 = (0.40)^2 + (0.48)^2 + \cdots + (0.44)^2 = 1.9035 \text{ ，}$$

以及

$$\sum_{i=1}^{n}(x_i y_i) = (0.40)(0.021) + (0.48)(0.030) + \cdots + (0.44)(0.025) = 0.10839 \text{ 。}$$

由於 $r_{xy} \approx 0.991$ 十分接近於 1，表示 y 與 x 具有高度的正相關性，所以採用比值估計之方法是非常恰當的。 □

例題 4.3　某一個雙變數的人工有限母體含有 100 個元素，這 100 個元素的數對 (y_i, x_i) 資料陳列於 表 1 之中(見次頁)。因此，這個有限母體之研究變數的母體平均數為

$$\mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{100}(2 + 1 + \cdots + 2) = \frac{266}{100} = 2.66 \text{ ，}$$

其中

$$N = \text{母體大小(population size)} = \text{母體之中的元素總個數} = 100。$$

我們暫且假裝不知道母體平均數 μ_y 的答案是多少，然後使用簡單隨機抽樣之方法從母體抽出 $n = 8$ 個元素，來估計母體平均數 μ_y，那麼也就是要從 100 個元素之中隨機地抽出 8 個元素，而這就需要介於 1 與 100 之間的 8 個亂數，來選出 8 個元素。因此，我們從某一本統計學書籍中節錄一小部份的亂數表，如以下表 2 所示，以便稍後模擬抽樣之用。

表 1　某人工母體 100 個元素的數對 (y_i, x_i) 資料
（y_i : 第 i 個元素的研究變數值； x_i : 第 i 個元素的輔助變數值)

i	(y_i, x_i)	i	(y_i, x_i)	i	(y_i, x_i)	i	(y_i, x_i)	i	(y_i, x_i)
1	(2, 3)	21	(2, 4)	41	(3, 4)	61	(3, 5)	81	(3, 4)
2	(1, 2)	22	(3, 4)	42	(3, 5)	62	(2, 4)	82	(3, 5)
3	(2, 4)	23	(4, 5)	43	(4, 6)	63	(3, 4)	83	(4, 6)
4	(3, 4)	24	(3, 5)	44	(5, 6)	64	(3, 5)	84	(5, 6)
5	(1, 2)	25	(2, 3)	45	(1, 2)	65	(4, 6)	85	(1, 2)
6	(0, 1)	26	(1, 2)	46	(2, 3)	66	(4, 7)	86	(2, 3)
7	(2, 4)	27	(2, 3)	47	(2, 4)	67	(3, 5)	87	(2, 4)
8	(3, 4)	28	(3, 4)	48	(3, 5)	68	(2, 3)	88	(3, 5)
9	(3, 5)	29	(2, 4)	49	(3, 4)	69	(1, 2)	89	(3, 4)
10	(2, 4)	30	(4, 6)	50	(4, 5)	70	(2, 3)	90	(4, 5)
11	(3, 4)	31	(5, 6)	51	(3, 4)	71	(2, 3)	91	(0, 1)
12	(3, 5)	32	(1, 2)	52	(3, 5)	72	(3, 4)	92	(2, 3)
13	(4, 6)	33	(2, 3)	53	(2, 4)	73	(2, 4)	93	(3, 4)
14	(5, 6)	34	(2, 4)	54	(3, 4)	74	(4, 6)	94	(2, 4)
15	(1, 2)	35	(3, 5)	55	(3, 5)	75	(5, 6)	95	(4, 6)
16	(2, 3)	36	(0, 1)	56	(6, 8)	76	(1, 2)	96	(3, 5)
17	(2, 4)	37	(2, 3)	57	(1, 2)	77	(2, 3)	97	(3, 6)
18	(3, 5)	38	(3, 4)	58	(2, 3)	78	(3, 5)	98	(2, 4)
19	(3, 4)	39	(5, 7)	59	(2, 4)	79	(2, 4)	99	(1, 2)
20	(4, 5)	40	(4, 6)	60	(3, 5)	80	(3, 6)	100	(2, 3)

表 2　從統計學書籍中節錄的一小部份亂數表

	column							
line	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40
1	62956	95735	70988	86027	27648	65155	46301	27217
2	17143	50118	41681	87224	75674	43371	09846	83403
3	99285	01369	94610	71099	69207	01999	23931	34711

現在，我們就從表 2 的第一個橫列由左至右依序選取 8 個二位數字如後: 62、95、69、57、35、70、98、88，所以依據簡單隨機抽樣之方法而抽得 8 個元素的數對型態樣本數據爲

(2, 4)、(4, 6)、(1, 2)、(1, 2)、(3, 5)、(2, 3)、(2, 4)、(3, 5)。

試採用以下三種不同的估計量，來推估母體平均數 μ_y，並相互比較三種估計量之優劣。同時，試計算 y 與 x 的樣本相關係數，以便判斷採用比值估計之方法是否恰當。

(a) y 的樣本平均數 $\bar{y}$　　(b) 比值估計量 (假設 $\mu_x = 4.15$ 爲已知)　　(c) 迴歸估計量

<u>解答</u> (a) y 的樣本平均數 $\overline{y}$ 之值為

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{8}(2+4+\cdots+3)=\frac{18}{8}=2.25\ 。$$

(b) 依據(5)式，μ_y 的比值估計量之值為

$$\hat{\mu}_y=\hat{R}\mu_x=\frac{\overline{y}}{\overline{x}}\cdot\mu_x=(\frac{2.25}{3.875})(4.15)\approx 2.41\ ，$$

其中

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{8}(4+6+\cdots+5)=\frac{31}{8}=3.875\ 。$$

(c) 依據(7)式，μ_y 的迴歸估計量之值為

$$\hat{\mu}_{y\mathrm{L}}=\overline{y}+\hat{\beta}(\mu_x-\overline{x})=2.25+\frac{82}{119}(4.15-3.875)\approx 2.44\ ，$$

其中

$$\hat{\beta}=\frac{s_{xy}}{s_x^2}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$=\frac{(\sum_{i=1}^{n}x_iy_i)-n(\overline{x}\cdot\overline{y})}{(\sum_{i=1}^{n}x_i^2)-n(\overline{x})^2}=\frac{(4\cdot 2+\cdots+5\cdot 3)-(1/8)(4+\cdots+5)(2+\cdots+3)}{(4^2+\cdots+5^2)-(1/8)(4+\cdots+5)^2}$$

$$=\frac{80-(1/8)(31)(18)}{135-(1/8)(31)^2}=\frac{10.25}{14.875}=\frac{82}{119}\approx 0.689\ 。$$

最後，我們來計算 y 與 x 的樣本相關係數，如下:

$$r_{xy}=\frac{s_{xy}}{\sqrt{s_x^2s_y^2}}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\overline{y})^2\right)}}=\frac{\left(\sum_{i=1}^{n}(x_iy_i)\right)-n(\overline{x}\cdot\overline{y})}{\sqrt{\left[\left(\sum_{i=1}^{n}x_i^2\right)-n(\overline{x})^2\right]\left[\left(\sum_{i=1}^{n}y_i^2\right)-n(\overline{y})^2\right]}}$$

$$=\frac{(4\cdot 2+\cdots+5\cdot 3)-(1/8)(4+\cdots+5)(2+\cdots+3)}{\sqrt{(4^2+\cdots+5^2)-(1/8)(4+\cdots+5)^2}\sqrt{(2^2+\cdots+3^2)-(1/8)(2+\cdots+3)^2}}$$

$$=\frac{80-(1/8)(31)(18)}{\sqrt{135-(1/8)(31)^2}\sqrt{48-(1/8)(18)^2}}=\frac{10.25}{\sqrt{14.875}\sqrt{7.5}}\approx 0.97\ 。$$

由於 $r_{xy}\approx 0.97$ 十分接近於 1，表示 y 與 x 具有高度的正相關性，所以採用比值估計以及迴歸估計之方法是非常恰當的。由於母體平均數 μ_y 的標準答案是 $\mu_y=2.66$，所以三個點估計值來相互比較，很顯然的，比值估計以及迴歸估計的估計誤差較小，樣本平均數 $\overline{y}$ 的估計誤差較大，所以比值估計以及迴歸估計皆十分適用於這個例題，這也驗證了先前 y 與 x

具有高度的正相關性的樣本相關係數值。

讀者或許會質疑，說不定以上的比較結果只是一種碰巧的情況；如果將先前抽得的 8 個亂數換成另外 8 個亂數，也許兩個不同估計值的比較結果會恰好相反也很難說喔！那麼，有這種質疑的讀者，不妨自行去做一個計算性質的實驗，將先前抽得的 8 個亂數換成另外 8 個亂數，再重複以上整個抽樣與估計的過程，看看結果會怎樣！如果每一次換成另外 8 個亂數再重複整個抽樣與估計的過程被視為"一回合"的實驗，那麼本文作者們斗膽在此拍胸脯保證，十個回合的實驗之中，至少有八個回合的實驗結果，皆為採用比值估計以及迴歸估計較優於樣本平均數 $\bar{y}$！ □

5. 結語

在這一篇教學短文裡，作者們利用幾個有趣的計算例題，來說明比值估計以及迴歸估計的基本概念與用法。作者們希望本文的內容對於剛起步學習抽樣調查方法的莘莘學子們有所助益。

參考文獻

Cochran, W. G. (1977). Sampling Techniques, 3rd ed., John Wiley & Sons, INC.

Scheaffer, R. L., Mendenhall, W., Ott, R.L., and Gerow, K. G. (2012). Survey Sampling, 7th ed., Brooks/Cole, Cengage Learning.

JPPS: ISSN 1607-7083