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 18 Number 1 June 2018

第十八卷 第一期 中華民國一百零七年六月

統領世紀 薪傳天下

JPPS ISSN 1607-7083

統　計　薪　傳

Journal of Propagations in Probability and Statistics

A Comprehensive Journal of Probability and Statistics
for Researchers, Practitioners, Teachers, Students, and Others

Volume 18 Number 1

June 2018

第十八卷　第一期

中華民國 107 年 6 月

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: **June 1, 2018**

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

JOURNAL OF PROPAGATIONS IN PROBABILITY AND STATISTICS

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 18 Number 1 June 2018

第十八卷 第一期 中華民國 107 年 6 月

Table of Contents / 目 次

Part I: English Section (articles written in English)

Estimating the Coefficient of Quartile Deviation with Confidence
--- Shipra Banik and B. M. Golam Kibria 1

Recent Trends in Credit Scoring and Financial Modelling: A Review --------------------- Jubin Antony 13

On Teaching the Method of Post-Stratification in Survey Sampling
-- Ardor Chen and Kuang-Chao Chang 23

Part II: Chinese Section (articles written in Chinese with English abstracts)

On Learning and Using Statistical Tools in Excel -- Bruce Lei 29
(認識 Excel 中的統計工具---- 雷觀籌)

A World-Wide Ranking on Research Productivity Based on Two Leading Statistics Journals During the Recent Century -------- Yi-Lien Hou, Chian-Hua Huang, Jia-Li Zeng, Chun-Han Chang, Han-Hsuan Lin, and Pin-Han Wang 45
(近一世紀以來全球統計學者之兩份頂尖期刊的論文產量排行榜 ----- 侯伊蓮、黃岍樺、曾家莉、張淳涵、林函萱、王品涵)

Journal of Propagations in Probability and Statistics
18(1), 23-28 June 2018

統 計 薪 傳
第十八卷 第一期
民國 107 年 6 月 23-28 頁

On Teaching the Method of Post-Stratification in Survey Sampling

Ardor Chen
Techcom Information Corp.

Kuang-Chao Chang
Fu Jen Catholic University

ABSTRACT In the theory of survey sampling, post-stratification is a remedial method which is used when all strata in a stratified finite population are not identifiable before we take samples. In other words, the population is "implicitly" stratified into a number of strata and there is no way to draw a sample from each individual stratum by the usual method of stratified random sampling. Thus, as a remedy, we draw a simple random sample from the entire population instead of using the impossible method of the usual stratified random sampling. Then, we post-stratify the simple random sample we obtained into a number of the so-called poststrata. Finally, using the sample observations from all poststrata, we may estimate certain population parameter of interest by the usual method of stratified random sampling, assuming that all the stratum weights (or stratum sizes) are known. In this article, we introduce the basic concept of post-stratification (P-St) and illustrate its usage by using an artificial finite population in Chang [1], and we compare the P-St with other sampling plans. The purpose of writing this article is to help students who are beginners in learning P-St in survey research.

Keywords Poststrata; Post-stratification; Simple random sampling; Stratified random sampling; Stratum size; Stratum weight.

1. Introduction

In the theory of survey sampling, **post-stratification**, or **stratification after selection of the sample**, is a sampling procedure which is used in the case that we would like to stratify on a study variable *y*, but are unable to place the sampling units into their correct strata until after the sample is selected. Personal characteristics such as age, sex, race and educational level are common examples (Cochran [3], p. 134). Compared with the usual **stratified random sampling**, post-stratification may be considered as a remedial method such that we obtain observations from **poststrata** for our estimation of unknown population parameters. In this article, we introduce the basic concept/ method of post-stratification and we illustrate the computation of post-stratified estimators by using

□ Received January 2018, revised April 2018, in final form May 2018.
□ Ardor Chen is the founder and CEO of Techcom Information Corp., Taipei, Taiwan, ROC; email: techcom5054@hotmail.com. Kuang-Chao Chang is a retired Professor from 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

an artificial finite population in Chang [1]. In Section 2, a detailed introduction to post-stratification with comparison to the usual stratified random sampling is presented. In addition, we introduce the method of collapsing empty poststrata (if there are any) with neighboring strata for small sample post-stratification. In Section 3, a numerical example with the artificial population in Chang [1] is used for a computational illustration of post-stratification. Section 4 concludes this article.

2. The Method of Post-Stratification

2.1 Post-Stratification

The method of post-stratification can be described in terms of a 3-step sampling procedure as follows:

Step 1 Take a simple random sample of size n from the entire population;

Step 2 Classify the above sample into strata (thus obtaining L poststrata with subsample sizes $n_1, n_2, \cdots, n_L$; $n_1 + n_2 + \cdots + n_L = n$);

Step 3 Use the classified data to estimate the unknown population parameter, say the unknown population mean, by the usual method of stratified random sampling (assuming all the stratum weights are known parameters).

Consider a finite population $\mathcal{F}$ of size N which is divided into L disjoint strata with **stratum sizes** N_h's, $h = 1, 2, \cdots, L$, respectively. Thus $N = N_1 + N_2 + \cdots + N_L$ and the population mean of the study variable y is

$$\mu = \frac{1}{N} \sum_{h=1}^{L} \sum_{j=1}^{N_h} y_{hj}$$

where y_{hj} is the value of the jth element in the hth stratum. The commonly used estimator of μ under the usual stratified random sampling (StRS), assuming each **stratum weight** $W_h = N_h / N$ is known, $h = 1, 2, \cdots, L$, is given by

$$\hat{\mu}_{\text{StRS}} = \sum_{h=1}^{L} W_h \hat{\mu}_h = \sum_{h=1}^{L} W_h \bar{y}_h \tag{1.1}$$

in which

$$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj}$$

is the mean of the subsample drawn from the hth stratum, $h = 1, 2, \cdots, L$.

Meanwhile, if we look at the situation under post-stratification (P-St), the estimator of μ is

$$\hat{\mu}_{\text{P-St}} = \sum_{h=1}^{L} W_h \hat{\mu}_h = \sum_{h=1}^{L} W_h \bar{y}_h \tag{1.2}$$

in which

$$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj} ,$$

$h = 1, 2, \cdots, L$. Eq. (1.2) looks just the same as Eq. (1.1), but strictly speaking they are not exactly the same as the post-stratified subsample size n_h in Eq. (1.2) is a random variable (instead of a fixed number) for each *h*, $h = 1, 2, \cdots, L$.

As a comparison with post-stratification, the usual stratified random sampling can also be described in terms of a 3-step sampling procedure as follows:

Step 1 Determine the total sample size *n* and the subsample sizes $n_1, n_2, \cdots, n_L$ by choosing a suitable allocation (proportional, Neyman, optimum, $\cdots$);

Step 2 Draw a simple random sample of size n_h independently from each stratum *h*, $h = 1, 2, \cdots, L$;

Step 3 Use the data obtained from Step 2 to estimate the unknown population parameter by the weighted type estimator given in Eq. (1.1) (assuming all the stratum weights are known parameters).

In general, the post-stratified estimator given in Eq. (1.2) produces good results (say, if compared with the sample mean under simple random sampling) when the sample size *n* is large and all n_h terms are relatively large as well. A practical consequence of this is that we cannot poststratify too finely (see Scheaffer *et al*. [4], p.146).

On the other hand, if the sample size *n* is small in post-stratification, we may encounter the so-called "empty-poststrata problem", which will be discussed in the next subsection.

2.2 Small Sample Post-Stratification

In the case of small samples in post-stratification, general acceptable theory or technique can hardly be found in textbooks of survey sampling methodology. One of the main difficulties is the possibility of "obtaining zero sample sizes in some strata" (i.e., obtaining empty poststrata) for small samples. A commonly used estimation procedure for μ is to collapse the empty poststrata, if there are any, with neighboring strata. Let $\hat{\mu}_{\mathrm{C}}$ be the estimator of μ under this procedure, where the subscript C in $\hat{\mu}_{\mathrm{C}}$ stands for "collapsing -----". As an illustrative example, we consider a finite population with $L = 4$ strata in which stratum 1 and stratum 2 are neighboring strata. Suppose we take a simple random sample of size *n* from the entire population where *n* is small, and it turns out that the sample contains no elements from stratum 1 (i.e., stratum 1 is an empty poststratum in the sample). Then the estimator $\hat{\mu}_{\mathrm{C}}$ can be written as

$$\hat{\mu}_{\mathrm{C}} = (W_1 + W_2)\bar{y}_2 + W_3\bar{y}_3 + W_4\bar{y}_4$$

where $\bar{y}_h$ is the sample mean of the nonempty poststratum *h*, *h* = 2, 3, 4 and W_h is the known stratum weight of stratum *h* in the population, *h* = 1, 2, 3, 4. It should be noted that two strata in the population are neighboring to each other not because they are geographically close to each other. They are neighboring strata if the values of the parameter of interest in the two strata are close to each other, and this is illustrated in Example 1 below. A drawback of the estimator $\hat{\mu}_{\mathrm{C}}$ is that it is a biased estimator of μ (and a proof of its biasedness in given in Chang *et al*. [2]).

Example 1 Let $\mathcal{F}$ be a finite population which can be implicitly stratified into $L = 5$ strata with known stratum weights. We like to estimate the mean of $\mathcal{F}$, μ, by the method of post-stratification, and we have prior informations about μ_h (the mean of stratum *h*) for *h* = 1, 2, 3, 4, 5 as follows:

Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5
$0 \le \mu_1 \le 10$	$10 \le \mu_2 \le 20$	$75 \le \mu_3 \le 85$	$20 \le \mu_4 \le 30$	$85 \le \mu_5 \le 100$

Suppose we take a simple random sample of size n (where n is small) from $\mathcal{F}$, and it turns out that the sample contains no elements from stratum 3 (i.e., stratum 3 is an empty poststratum in the sample). What will be the estimator $\hat{\mu}_{\mathrm{C}}$ of collapsing empty poststratum with neighboring poststratum in this situation?

Solution Since μ_5 is the stratum mean that is closest to μ_3 among all μ_h's, $h \ne 3$, the neighboring poststratum of the empty poststratum is stratum 5. Thus, the estimator $\hat{\mu}_{\mathrm{C}}$ is given by

$$\hat{\mu}_{\mathrm{C}} = W_1\bar{y}_1 + W_2\bar{y}_2 + W_4\bar{y}_4 + (W_3 + W_5)\bar{y}_5$$

where $\bar{y}_h$ is the sample mean of the nonempty poststratum h, h = 1, 2, 4, 5 and W_h is the known stratum weight of stratum h in the population, h = 1, 2, 3, 4, 5. □

3. A Computational Example

In the following example, we use the artificial population in Chang [1] for computational illustrations.

Example 2 An elementary school in the mountain area of Nantou County in Taiwan has $N = 120$ students, who can be stratified into three strata by their heights as follows:

5^{th} & 6^{th} grades ($N_1 = 50$)				3^{rd} & 4^{th} grades ($N_2 = 40$)				1^{st} & 2^{nd} grades ($N_3 = 30$)	
#	Height y_i (cm)	#	Height y_i (cm)	#	Height y_i (cm)	#	Height y_i (cm)	#	Height y_i (cm)
1	161	31	148	51	150	81	153	91	135
2	158	32	151	52	149	82	147	92	133
3	152	33	162	53	145	83	158	93	138
4	155	34	157	54	152	84	154	94	140
5	159	35	163	55	146	85	155	95	131
6	163	36	166	56	147	86	145	96	129
7	158	37	156	57	151	87	143	97	141
8	166	38	158	58	160	88	151	98	142
9	150	39	160	59	148	89	162	99	139
10	157	40	162	60	155	90	152	100	141
11	162	41	155	61	152			101	133
12	168	42	153	62	150			102	130
13	163	43	151	63	149			103	129
14	159	44	165	64	154			104	144
15	165	45	167	65	156			105	145
16	164	46	154	66	151			106	143
17	156	47	172	67	146			107	140
18	152	48	157	68	148			108	138
19	171	49	159	69	142			109	136
20	154	50	162	70	150			110	137
21	149			71	153			111	147
22	155			72	155			112	146
23	167			73	153			113	139
24	162			74	148			114	126
25	161			75	150			115	132
26	170			76	151			116	128
27	158			77	144			117	138
28	154			78	152			118	143
29	160			79	145			119	146
30	152			80	153			120	142

Table 1 Selected random numbers (from the Appendix in Walpole [5])

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

(a) Draw a simple random sample of size $n=12$ (by choosing 12 non-repeated two-digit integers from Table 1) from the above population of 120 students and estimate the average height of all 120 students (i.e., estimate μ_y, the mean of y) by using the sample mean $\bar{y}$.
Note: Starting from the first row in Table 1, select 12 non-repeated two-digit integers (from left to right). Then, multiply each one of the selected 12 integers by 1.2 and round off.

(b) Using the 12 two-digit integers selected in Part (a), estimate the average height of all 120 students by the method of post-stratification. (It is now assumed that the population of 120 students is implicitly stratified into three strata.)
Note: Again, multiply each one of the selected 12 integers by 1.2 and round off.

Solutions

(a) Starting from the first row in Table 1, we select 12 non-repeated two-digit integers (from left to right) as follows: 17, 14, 35, 01, 18, 41, 68, 72, 24, 75, 67, and 44. Multiplying each one of the 12 integers by 1.2 and rounding off, we obtain another 12 non-repeated integers (which are between 1 and 120 inclusively) as follows: 20, 17, 42, 1, 22, 49, 82, 86, 29, 90, 80, and 53. Thus, we select #20, #17, $\cdots$, and #53 from the population of 120 students and we obtain sample observations as follows:

154, 156, 153, 161, 155, 159, 147, 145, 160, 152, 153, 145.

Therefore, the sample mean is

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{12}(154+156+\cdots+145)=\frac{1840}{12}=153.\bar{3},$$

which is the estimate of μ_y under simple random sampling.

(b) The selected 12 non-repeated two-digit integers in Part (a), after being multiplied by 1.2 and rounded off, become 20, 17, 42, 1, 22, 49, 82, 86, 29, 90, 80, and 53. Thus, we select #20, #17, $\cdots$, and #53 from the population of 120 students. Now, among these 12 selected students, #20, #17, #42, #1, #22, #49, and #29 belong to the first poststratum; #82, #86, #90, #80, and #53 belong to the second poststratum; and the third poststratum is empty. Thus, $n_1=7$, $n_2=5$, and $n_3=0$. Hence, we obtain sample observations in each poststratum as follows:

	Sample observations
Poststratum I ($n_1=7$)	154, 156, 153, 161, 155, 159, 160
Poststratum II ($n_2=5$)	147, 145, 152, 153, 145
Poststratum III ($n_3=0$)	Empty poststratum

Since the third poststratum is empty, we combine it with it's neighboring stratum: the second poststratum, in regard to the estimator $\hat{\mu}_C$ in Example 1. Therefore, the post-stratification estimate of population mean is

$$\hat{\mu}_{\text{P-St}} = \hat{\mu}_{\text{C}} = W_1\bar{y}_1 + (W_2 + W_3)\bar{y}_2 = W_1\left(\frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j}\right) + (W_2 + W_3)\left(\frac{1}{n_2}\sum_{j=1}^{n_2} y_{2j}\right)$$

$$= \frac{5}{12}\cdot\frac{1}{7}(154+156+153+161+155+159+160) + (\frac{4}{12}+\frac{3}{12})\cdot\frac{1}{5}(147+145+152+153+145)$$

$$= \frac{5490}{84} + \frac{5194}{60} = \frac{915}{14} + \frac{2597}{30} \approx 65.3571 + 86.5667 = 151.9238 .$$

Since the true value of μ_y is $\mu_y \approx 150.96$, $\hat{\mu}_{\text{P-St}} \approx 151.9238$ is a better estimate of μ_y than the estimate $\bar{y} = 153.\bar{3}$. □

From Example 2 above, we see that the post-stratified estimator performed well.

4. Conclusions

In this article, we have introduced the basic concept/method of post-stratification and illustrated its usage by a numerical example with the artificial population in Chang [1]. In particular, we have encountered an empty poststratum in the numerical example, and we have combined the empty poststratum with it's neighboring poststratum for our post-stratified estimator. It turned out that our post-stratified estimator performed well and had smaller estimation error as compared with the sample mean under simple random sampling.

References

[1] Chang, K. M. (2015). Teaching "stratified random sampling" in survey research, *Journal of Propagations in Probability and Statistics*, **15**(1), 49-54.

[2] Chang, K. C., Liu, J. F., and Han, C. P. (1998). Multiple inverse sampling in post-stratification, *Journal of Statistical Planning and Inference*, **69**, 209-227.

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

[4] Scheaffer, R. L., Mendenhall, W., Ott, R. L., and Gerow, K. G. (2012). Survey Sampling, 7th ed., BROOKS/COLE, Cengage Learning.

[5] Walpole, R. E. (1982). Introduction to Statistics, 3rd ed., Macmillan Publishing Co., Inc., New York.

 JPPS: ISSN 1607-7083