RRPA87041007 (4 .P)

行政院國家科學委員會專題研究計畫成果報告

最大區別卡方檢定在不同情況下之考驗力比較問題
Power Comparisons of Maximally Discriminated Chi Square Approaches under Various Conditions

計畫編號：NSC 87-2118-M-227-001
執行期限：86 年 08 月 01 日至 87 年 07 月 31 日
主持人：陳素秋　　執行機構及單位：國立臺北護理學院醫護管理學系

I. Abstract (中英文摘要)

醫藥衛生領域的研究者，通常希望找出連續性資料分切點以分別兩組資料來作分析，但卻沒有報告其相關考驗力。因此，本研究的目的為有系統的評估兩極化資料中最大區別檢定與原始連續性資料的檢定間之考驗力。結果顯示在長尾 Cauchy 分佈之資料中，若總樣本數達 50，則一般無母數分析法及兩極化資料中最大區別檢定法很明顯地在考驗力上優於原始連續性資料的母數 t 檢定法。此結果有助於研究者選擇最佳資料分切點及其相關假說檢定法。

關鍵詞：Bootstrap，分切點，最大區別統計值，考驗力，模擬研究法

Researchers in health related fields often wish to analyze data by taking a cutpoint of a quantitative factor without reports for the power of this approach. Therefore, the purpose of this study is to evaluate power between maximal approaches of a dichotomous form and related tests of a quantitative form under different conditions of data. Results show that maximal approaches and non-parametric tests have superior power than that of the parametric two-sample t test when total sample size 50 from the heavy-tailed Cauchy distributed data.

Keywords: Bootstrap, Cutpoints, Maximally Discriminated Statistics, Power, Simulations

II. Introduction and Purpose (緣由與目的)

In research of health sciences, investigators often wish to classify a quantitative factor into two groups such as risk and normal groups according to a cutpoint c^*. Subjects with values exceeding c^* will be considered more likely to have a specific event in the sense of prediction. For example, in environmental health, Perera et al. (1992) reported possible molecular and genetic damage from environmental pollution in Poland [1]. They classified quantitative oncoprotein *ras* p21 (a biological marker) into two groups (+, -) to be compared with exposed and control groups of environmental pollution. The + p21 was determined by two standard deviations above the mean p21 in 50 normal, healthy, non-smoking subjects. The p value of the chi square statistic for this 2×2 table was 0.8 [1].

Criteria of choosing a cutpoint are not unique. In the issue of individual's growth, researchers continue to show a significantly increased risk of some chronic diseases in obesity [2][3]. However, there are various indices for the definition of obesity by choosing different cutpoints. Yen et al. (1994) compared three existing body screening indices for obesity: a skinfold exceeding 85th %tile for subjects with the same sex and age, a body mass index [(weight:kg)/(height:m)^2] exceeding 85th %tile for subjects with the same sex, and a weight more than 20% of an average weight for subjects with the same sex [4]. In cancer research, Bochner et al. (1995) explored the relationship between prognostic indicators and tumor angiogenesis (a measurement of tumor microvessel densities) in the patients with bladder cancer. Patients were divided into three groups (low, intermediate, and high microvessel counts) with equal numbers of individual from the quantitative angiogenesis as was planned before the data were analyzed [5]. In addition, an analysis of an optimal cutpoint in microvessel counts related to disease recurrence and overall survival was performed [5].

In the preceding examples, one question emerges: is there a way to choose an optimal cutpoint of a quantitative factor for maximal discrimination and what are associated statistical tests and their power for alternatives? The maximal chi square approach for testing a 2×2 table from an optimal cutpoint was first developed by Miller and Siegmund in 1982 [6]. Halpern (1982) studied small-sample properties of this approach [7]. Lausen and Schumacher (1992) later proposed a maximally selected rank statistic that provides a test and an estimate of a cutpoint as a simple classification rule [8]. Pros and cons of cutpoints not related to maximal chi square or rank approaches were discussed by Goldsmith (1995). Her work mainly compared sample sizes between dichotomized methods exploring cutpoint choice and other methods that take advantage of numerical nature of the underlying data [9]. To my knowledge, these papers do not provide enough information to evaluate the related power. Thus, this project attempts to do such evaluation through bootstrap techniques and simulations [10].

The purpose of this project is as follows. (i) To choose an optimal cutpoint of a quantitative factor based on a maximally selected chi square statistic. (ii) To evaluate power of maximally chi square statistics for various alternatives under various sample sizes. (iii) To compare maximal chi square approaches of a dichotomized form with related tests of the originally quantitative form. (iv) To give suggestions for researchers in choice of appropriate tests between quantitative variables and related dichotomized forms.

III. Research Methods (研究方法)

An ANOVA-type model for generalization of notation in this study was specified as

$$x_{ij} = \mu + \tau_j + e_{ij}; i = 1, \cdots, n; j = 1, \cdots, g. \quad (1)$$

x_{ij} is the i^{th} observation (a continuously ordinal/interval/ratio variable) from the j^{th} group, μ is the true grand mean of all the gn observations, τ_j is the amount by which the j^{th} group mean differs from the true grand mean, and e_{ij} is an error term that the i^{th} observation deviates from the group mean by this amount. Throughout this project, an equal number n of observations in each group was assumed for simplicity, and a case of unequal numbers n_j can be done in a similar way. The assumptions for this fixed effects model which makes inferences only to the g groups are (a) the g sets of observed data constituting g independent samples from the respective populations and (b) $\sum_{j=1}^{g} n\tau_j = 0$.

Suppose x_{ij} in (1) can be dichotomized by a series of cutpoints c'_k, $k = 1, 2, \ldots, v$, and $c'_1 < c'_2 < \ldots < c'_k < \ldots < c'_v$ are the ordered distinct values from a certain range of gn observations and v is the number of these distinct values. Let $g = 2$, then the k^{th} 2×2 contingency table can be constructed as

	$j = 1$	$j = 2$	Total
$X_{ijk} > c'_k$	a_k	b_k	
$X_{ijk} \le c'_k$	c_k	d_k	
Total	n	n	$2n$

(2)

where a_k, b_k, c_k, and d_k are numbers of observations in related cells of this k^{th} 2×2 table.

Related null hypotheses of different methods to analyze data of the model (1) are described from H_{01} to H_{03} as follows. H_{01} for the two-sample t test: Two population means are the same, that is $\tau_1 = \tau_2 = 0$. H_{02} for the Wilcoxon rank sum test (or the Mann-Whitty U test): Two population medians are the same by ranks of the data. H_{03} for the two-sample median test: Two population medians are the same by median cutpoint. For the k sets of 2×2 contingency tables in (2), related null hypotheses of various methods are proposed below. H_{04} for the maximally selected Fisher exact test (when a sample size is very small): The two populations are homogeneous with respect to cutpoint classification. H_{05} for the maximally selected chi square test: The two populations are homogeneous with respect to cutpoint classification.

Miller and Siegmund (1982) developed the maximally selected chi square statistic which a cutpoint c^*_k in (2) is selected so as to maximize the standard chi square statistic χ^2_k,

$$\chi_k^2 = \frac{2n(a_k d_k - b_k c_k)^2}{(a_k + b_k)(c_k + d_k)(n)(n)}. \qquad (3)$$

They showed that the chi square percentile points are inappropriate for this statistic. Thus, they computed actual significance levels for large samples, and tabulated correct percentile points for this approach [6]. Halpern (1982) later simulated the small sample null distributions of the maximum chi square statistic [7]. Similar approaches were explored for choosing optimal cutpoints which maximizing the Fisher exact test statistic and the chi square test statistic computed over the central 80% of data in this project. The null small sample distributions of these maximal statistics will be simulated differently by bootstrap techniques. Bootstrap methods resample a given data set b times and the b bootstrapped maximal specified test statistics generate a null small sample distribution of this maximal statistic. This maximal statistic of the same given data set is then compared with percentile points of the bootstrapped null small sample distribution for recording a p value. We set b to be 500, and evaluate the power of these proposed methods.

Monte Carlo simulations (500 times) were done to evaluate the power of maximally discriminated approaches. Suppose alternative hypotheses of the model (1) are set to be H_A: $\tau_1 = -(g-1)gd = -2d$, $\tau_2 = \ldots = \tau_g = gd = 2d$. Some error ($e_{ij}$) distributions will be generated such as a family of heavier-tailed distributions (h distributions): $h(w) = Z\times\exp(w\times Z^2/2)$, where Z is the standard normal distribution, that is $Z \sim N(0,1)$. $h(0)$ is equivalent to $N(0,1)$, $h(1)$ is similar to t distribution with 1 degree of freedom (t_1), and t_1 is equivalent to the Cauchy distribution.

IV. Results and Discussion (結果與討論)

Figure 1 shows power of maximally discriminated tests, other non-parametric tests,

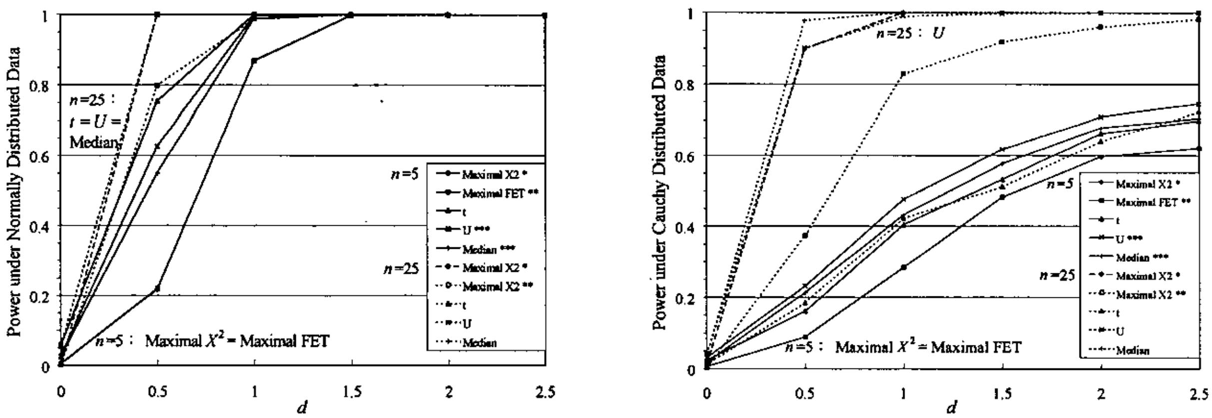

Figure 1: Power under the normally (left figure) and Cauchy (right) distributed data for various tests (α=0.05).

* represents that the maximal chi square test using the criteria of Miller and Siegmund (1982).

** represents that the maximal chi square test or maximal Fisher exact test (FET, for n = 5) using the bootstrap resampling techniques.

*** means that the exact methods of the Mann-Whitty U test or two-sample median test (for n = 5).

and t test under the normally and Cauchy distributed data if the significance level (α) is set to be 0.05. The power is also compared between the total sample size ($2n$) 10 and 50 in Figure 1. Under the Cauchy distributed data, poor power is presented for all tests if the total sample size is very small ($2n$=10) and good power is displayed for all tests but not the parametric t test if $2n$ is equal to 50. As the distance (d) between two group means is increasing, the power is also increasing for all situations. However, it should be noted that the power is approximately equal to α if d is zero, and the power should be renamed as the probability of type I error in this case. The power under the normally distributed data is higher than that of the Cauchy distributed data given the same condition, this is especially obvious when $2n$ is equal to 10. In general, the power under the normally distributed data is good.

Results of this study may give medical researchers suggestions regarding the choice of cutpoints and associate hypotheses tests.

V. Evaluation of Results (計畫成果自評)

Less statistical research on power consideration of a maximally discriminated cutpoint has been compared with other methods that take advantage of numerical nature of the underlying data. However, as it was addressed in Introduction, investigators often wish to analyze their data by taking a cutpoint for the purpose of maximal discrimination in health related research. Thus, it is important to understand which cutpoints maximize discrimination and to evaluate power for alternatives under this situation.

The results of this study may give health-related researchers suggestions regarding their choice of cutpoints and appropriate tests under different conditions. In this project, maximal approaches of a dichotomized form were evaluated on the Fisher exact test and chi square tests. Both parametric and non-parametric/distribution-free tests of a quantitative form were considered, and their powers were compared with those of maximal approaches of a dichotomized form under various conditions.

An approach of maximal logistic regression will be deserved for further evaluation. Logistic regression is commonly used in epidemiological research, and it can take covariates and confounders into accounts. It will be useful to have a maximally discriminated cutpoint by adjusting related variables through logistic regression.

VI. References (參考文獻)

[1] Perera, F. P., Hemminki, K., Gryzbowska, E., Motykiewicz, G., Michalska, J., Santella, R. M., Young, T.-L., Dickey, C., Brandt-Rauf, P., DeVivo, I., Blaner, W., Tsai, W.-Y., and Chorazy, M. (1992), "Molecular and Genetic Damage in Humans from Environmental Pollution in Poland," *Nature*, 360, 256-258.

[2] Dietz, W. H. Jr. (1983), "Childhood Obesity: Susceptibility, Cause, and Management," *The Journal of Pediatrics*, 103, 676-686.

[3] National Task Force on the Prevention and Treatment of Obesity (1994), "Weight Cycling," *Journal of American Medical Association*, 272, 1196-1202.

[4] Yen, L.-L., Pan, W.-H., Chen, C.-H., and Lee, Y. M. (1994), "The Prevalence of Obesity in the Seventh Graders in Taipei City, 1991: A Comparison of Various Body Screening Indices," *Chinese Journal of Public Health (Taipei)*, 13, 11-19.

[5] Bochner, B. H., Cote, R. J., Weidner, N., Groshen, S., Chen, S.-C., Skinner, D. G., and Nichols, P. W. (1995), "Angiogensis in Bladder Cancer: Relationship Between Microvessel Density and Tumor Prognosis," *Journal of National Cancer Institute*, 87, 1603-1612.

[6] Miller, R., and Siegmund, D. (1982), "Maximally Selected Chi Square Statistics," *Biometrics*, 38, 1011-1016.

[7] Halpern, J. (1982), "Maximally Selected Chi Square Statistics for Small Samples," *Biometrics*, 38, 1017-1023.

[8] Lausen, B., and Schumacher, M. (1992), "Maximally Selected Rank Statistics," *Biometrics*, 48, 73-85.

[9] Goldsmith, L. J. (1995), "Pros and Cons of Cutpoints," *American Statistical Association 1995 Proceedings of the Biometrics Section*, 272-276.

[10] Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.