行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※二維現今資料無母數獨立性之檢定　　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□**個別型計畫**　　　□**整合型計畫**

計畫編號：NSC 89－2118 － M － 009 － 010

執行期間： 89 **年** 8 **月** 1 **日至** 90 **年** 7 **月** 31 **日**

計畫主持人：王維菁

共同主持人：

本成果報告包括以下應繳交之附件：

□**赴國外出差或研習心得報告一份**

□**赴大陸地區出差或研習心得報告一份**

□**出席國際學術會議心得報告及發表之論文各一份**

□**國際合作研究計畫國外研究報告書一份**

執行單位：交通大學統計所

中 華 民 國 90 **年** 10 **月** 30 **日**

行政院國家科學委員會專題研究計畫成果報告
二維現今資料無母數獨立性檢定

計畫編號：NSC 89-2118-M-009-010
執行期限：89 年 8 月 1 日至 90 年 7 月 31 日
主持人：王維菁　交通大學統計所

一、中文摘要

本計劃針對二維現今資料提出檢定獨立性的方法，並討論其大樣本性質，亦將所提方法用來分析中研院生醫所潘文涵教授主持的營養流行病學調查資料。

關鍵詞：區間設限，二項列聯表.

Abstract

This paper develops a nonparametric procedure for testing marginal independence based on bivariate current status data. Asymptotic properties of the proposed tests are derived and their finite sample performance is studied. The method is applied to analyze data from a community based study of cardiovascular epidemiology in Taiwan.

Keywords: Interval Censoring; Mantel Haenszel Test; Two by Two Tables.

II. Motivation and Objective

Current status data commonly arise in epidemiological investigation of the natural history of a disease. Such data structure is also called as interval censoring of case I (Groneboom and Wellner, 1992). In this article, we consider the bivariate case. Our work was motivated by a community-based study of cardiovascular diseases in Taiwan. The purpose of the study was to investigate whether the onset ages of some common cardiovascular diseases, specifically hypertension (HT), diabetes mellitus (DM) and hypercholesterolemia (HC), are correlated with each other. Let (T_1, T_2) be a pair of failure times of interest and C be the monitoring time of T_i $(i = 1,2)$, respectively. Bivariate current status data are of the form, $(C_1, C_2, \delta_1, \delta_2)$, where $\delta_i = I(T_i \le C_i)$. The main objective of this work is to develop an inference procedure for testing independence between two failure time variables given only bivariate current status data. It is important to note that the result obtained by Wang and Ding (2000) can not be applied to test independence. Specifically under independence, the value of the copula association parameter often occurs at the boundary of the parameter space, while Wang and Ding's estimator is only valid in the interior region. Furthermore our goal is to construct a nonparametric method without any model assumption.

III. The Proposed Methods

We consider testing the weaker condition $H_0 : T_1 \mid (C_1, C_2) \perp T_2 \mid (C_1, C_2)$. Given that $(C_1, C_2) = (c_1, c_2)$, one can construct a two-by-two table with margins, $(\delta_1, \delta_2) = (1,1)$, $(\delta_1, \delta_2) = (1,0)$, $(\delta_1, \delta_2) = (0,1)$ and $(\delta_1, \delta_2) = (0,0)$. Then we combine the individual tables to obtain a merged table. The observed counts in the merged table can be

denoted as $N_{ij} = \sum_{k=1}^{k=n} I(\delta_{1k} = i, \delta_{2k} = j)$, for $(i, j) = (0,1)$. Under the null hypothesis, the expected counts in the merged table become

$$E_{11} = \sum_{k=1}^{k=n} F_1(c_{1k}) F_2(c_{2k}),$$

$$E_{10} = \sum_{k=1}^{k=n} F_1(c_{1k}) S_2(c_{2k}),$$

$$E_{01} = \sum_{k=1}^{k=n} S_1(c_{1k}) F_2(c_{2k})$$

$$E_{00} = \sum_{k=1}^{k=n} S_1(c_{1k}) S_2(c_{2k}),$$

where $\Pr(T_j \le t) = F_j(t)$ and $S_j(t) = 1 - F_j(t)$. Since the marginal functions are unknown, we estimate them by the corresponding NP MLEs

The proposed test statistic is of the form,

$$Q = \frac{(N_{00} - \hat{E}_{00})^2}{\hat{A}VAR(N_{00} - \hat{E}_{00})},$$

where

$$\hat{E}_{00} = \sum_{k=1}^{k=n} \hat{S}_1(c_{1k}) \hat{S}_2(c_{2k})$$

and $\hat{A}VAR(N_{00} - \hat{E}_{00})$ is the estimator of the asymptotic variance of the numerator. We show that the test based on one cell is equivalent to that using the rest. We also prove that $n^{-1/2}(N_{00} - \hat{E}_{00})$ converges in distribution to $N(0, \sigma^2)$. We discuss different ways of estimating σ^2. It is easy to see that Q converges in distribution to a Chi -squared distribution with degree of freedom equal to 1.

Power and efficiency of the previous statistic may be improved by including a weight function. We derive the optimal weight function and compute its form when the alternative hypothesis follow some copula models.

We also study finite sample performance of the proposed tests via simulations. The article is under revision by *Biometrika*.

IV Reference

\item {\sc Groeneboom, P. \& Wellner, J. A.} (1992). {\sl Information Bounds and Non-Parametric Maximum Likelihood Estimation.} Boston: Birkh\"{a}user.

Wang, W. and Ding, A. A. (2000). On Assessing the Association for Bivariate Current Status Data. To appear in Biometrika.