広島統計談話会

Hiroshima Statistics Study Group

第 233 回談話会を下記にように開催致しますので
御参集下さいますようご案内申し上げます。

You are cordially invited to the 233th meeting as scheduled below.

日　時：　2006 年 12 月 15 日（金）15:00 –

Date ：　December 15, 2006 (Fri) 15:00 –

場　所：　**放射線影響研究所　講堂**

Place ：　RERF Auditorium

演　者：　**カリングス・M・ハリー**
（財団法人　放射線影響研究所統計部　副主任研究員）

Speaker ：　Harry M. Cullings, Ph.D., Associate Senior Scientist,
Department of Statistics, Radiation Effects Research Foundation,
Hiroshima

演　題：　**「抗体価に対する実験結果の解析における有限混合モデルの使用」**

Title ：　Uses of Finite Mixture Models in Analyzing Laboratory Results
on Antibody Titers

要　約：

Abstract ：

Finite mixture models are useful in analyzing laboratory results on a continuous variable used to classify a patient into one of two or more diagnostic categories. The simple and archetypal example is to classify a patient as having or not having a disease based on a laboratory test, when there is no other, more definitive clinical basis for diagnosis. That is, there is no other result available on any subset of subjects for some “gold standard” method such as a direct medical examination of the patient’s body using some invasive procedure. The finite mixture model intends to establish a cutoff value for distinguishing disease from no disease by discerning two distinct empirical distributions in the data, corresponding to those with and without disease, which also then suggests corresponding false negative and false positive error rates for a given cutoff value.

A recent RERF study of gastric cancer provided two interesting cases for application of finite mixture models: 1) antibody titers for *H. pylori* IgG and for a specific type of *H. pylori* called "CagA," and 2) antibody titers for two digestive enzymes, pepsinogen I and II. In each of these cases, a bivariate approach appears useful because two different, correlated test results are relevant to the diagnosis. This talk begins by reviewing some important considerations contained in a recent paper by Baughman *et al.* on fitting multiple antibody-positive groups using a mixture of normal distributions on the logarithms of the data for antibody titers. Their considerations include accounting for the censoring effect of a lower limit of detection of the assay, adjustment of the likelihood for population-based sampling probabilities, and evaluating the adequacy of the logarithmic transformation of the data by a power-transformation method.

Regarding the RERF applications, this talk will introduce the idea of a simple bivariate model using a combined likelihood with a simple linear relationship between the two log-transformed results for *H. pylori IgG* and CagA. This is similar in some ways to an approach suggested by Thompson *et al.* for finite mixture models, which incorporates a linear relationship between the mixture parameters and a vector of separate covariates. The talk concludes by discussing the case of pepsinogen I and II, which many investigators have used in combination to diagnose chronic atrophic gastritis. Various bivariate criteria have been based on medical reasoning without specific statistical assumptions, i.e., pepsinogen I is reduced by chronic atrophic gastritis but pepsinogen II is not, so inclusion of the latter in a bivariate criterion intends to adjust for various sources of variability affecting the overall level of antibody titer in a given sample. The pepsinogen assay is fundamentally different from the *H. pylori IgG* and CagA assays, in that the underlying process does not necessarily suggest distinct distributions of results, but possibly a continuum of gradual decrease in titer related to a corresponding decrease in secretion of the enzyme with progression of atrophy. Our data have the added feature of two sequential results separated by about ten years' time on many of the study participants. The assumption of strict progression in the underlying biological process establishes a useful algebraic relationship among prevalence, probability of progression, false positive error rate, and false negative error rate, based on the observed sequences; i.e., for discrete outcomes a true negative state cannot exist on the second sample if the first sample was true positive.