Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 216 (2008) 545–553
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Analysis of variance and linear contrasts in experimental design with generalized secant hyperbolic distribution

Yildiz E. Yilmaz[a], Aysen D. Akkaya[a, b, *]

[a]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1
[b]Department of Statistics, Middle East Technical University, 06531 Ankara, Turkey

Received 12 December 2006; received in revised form 29 May 2007

Abstract

We consider one-way classification model in experimental design when the errors have generalized secant hyperbolic distribution. We obtain efficient and robust estimators for block effects by using the modified maximum likelihood estimation (MML) methodology. A test statistic analogous to the normal-theory F statistic is defined to test block effects. We also define a test statistic for testing linear contrasts. It is shown that test statistics based on MML estimators are efficient and robust. The methodology readily extends to unbalanced designs.

MSC: 62K10; 62F10; 62F12; 62F03; 62F35

Keywords: Experimental design; Non-normality; Generalized secant hyperbolic; Modified maximum likelihood; Linear contrast; Robustness

1. Introduction

Analysis of variance procedures have traditionally been based on the assumption of normality. In practice, however, non-normal distributions occur so frequently. A number of studies have been made to investigate the effect of non-normality on the test statistics used in analysis of variance. The effect of non-normality on Type I error was studied in [11,7,6,2,9,3,15]. The effect of non-normality on power of the test was studied in [4,13,5,18]. They concluded that the effect of moderate non-normality on Type I error is not very serious but the power is adversely affected; the values of the power are, in fact, considerably smaller than under normality if, particularly, the kurtosis of the underlying distribution is different than 3 (kurtosis of normal). The above investigations for symmetric non-normal distributions have been carried out when the mathematical forms of short-tailed distributions (kurtosis less than 3) and long-tailed distributions (kurtosis greater than 3) are quite distinct from one another, e.g., the former is uniform and the latter is Student's t [22, Chapter 1]. The purpose of this paper is to present unified results by considering the family of generalized secant hyperbolic (GSH) distributions. The properties of GSH distributions have been studied by Vaughan [24]. They represent both short-tailed and long-tailed symmetric distributions with kurtosis ranging from 1.8 to 9 and include logistic as a particular case, uniform as a limiting case, and closely approximate the normal and Student's t distributions. Maximum

* Corresponding author. Department of Statistics, Middle East Technical University, 06531 Ankara, Turkey. Tel.: +90 312 2105329; fax: +90 312 2102959.
E-mail address: akkay@metu.edu.tr (A.D. Akkaya).

doi:10.1016/j.cam.2007.06.001

likelihood estimators being intractable [12,24], we derive modified maximum likelihood (MML) estimators of block effects (and scale parameter) in the framework of one-way classification experimental design and show that they are asymptotically fully efficient. We also study their properties for small sample sizes n and show that they are in general considerably more efficient than the normal theory (i.e., least squares) estimators. In fact the least squares estimators have a disconcerting feature, namely, their efficiencies relative to the MML estimators decrease as the sample size in a block increases. A test statistic analogous to the normal-theory F statistic is defined to test block effects. We also define a test statistic for testing linear contrasts. It is shown that test statistics based on MML estimators are efficient and robust. The methodology obtained readily extends to unbalanced designs.

2. One-way classification model

Consider the one-way classification fixed-effects model

$$y_{ij} = \mu + \tau_i + e_{ij} \quad (i = 1, 2, \ldots, k;\ \ j = 1, 2, \ldots, n), \tag{2.1}$$

having k blocks with n observations in each block; y_{ij} is the jth observation in the ith block, μ is the overall mean, τ_i is the ith block effect, and e_{ij} are iid random errors. In the fixed effects model, τ_i are defined as deviations from their overall mean. Thus, $\sum_{i=1}^{k}\tau_i = 0$.

In (2.1), suppose that e_{ij} are iid and have GSH distribution [24]

$$\text{GSH}(0, \sigma; t)\colon f(e) = \frac{c_1}{\sigma}\frac{\exp(c_2 e/\sigma)}{\exp(2c_2 e/\sigma) + 2a\exp(c_2 e/\sigma) + 1} \quad (-\infty < e < \infty), \tag{2.2}$$

where for $-\pi < t \leqslant 0$:

$$a = \cos(t), \quad c_2 = \sqrt{(\pi^2 - t^2)/3} \quad \text{and} \quad c_1 = \frac{\sin(t)}{t}c_2$$

and for $t > 0$:

$$a = \cosh(t), \quad c_2 = \sqrt{(\pi^2 + t^2)/3} \quad \text{and} \quad c_1 = \frac{\sinh(t)}{t}c_2.$$

For $t > \pi$, $t < \pi$ and $t = \pi$, GSH$(0, \sigma; t)$ represents short-tailed, long-tailed and approximately normal distributions, respectively; σ is the scale parameter and t is a nuisance parameter.

The coefficient of kurtosis, $\beta_2 = \mu_4/\mu_2^2$, is given below for a few representative values of the shape parameter, t:

$t=$	$-\pi\sqrt{2/3}$	$-\pi/2$	0	π	$\pi\sqrt{11}$	∞
$\beta_2=$	9.0	5.0	4.2	3.0	2.0	1.8

The likelihood function is

$$L = \frac{c_1^N}{\sigma^N}\prod_{i=1}^{k}\prod_{j=1}^{n}\frac{\exp(c_2 z_{ij})}{\exp(2c_2 z_{ij}) + 2a\exp(c_2 z_{ij}) + 1}, \tag{2.3}$$

where $N = nk$, $z_{ij} = \frac{y_{ij} - \mu - \tau_i}{\sigma}$ $(1 \leqslant i \leqslant k,\ \ 1 \leqslant j \leqslant n)$.

Let $y_{i(1)} \leqslant y_{i(2)} \leqslant \cdots \leqslant y_{i(n)}$ $(1 \leqslant i \leqslant k)$ be the order statistics of the n observations y_{ij} $(1 \leqslant j \leqslant n)$ in the ith block. Then $z_{i(j)} = (y_{i(j)} - \mu - \tau_i)/\sigma$ $(1 \leqslant i \leqslant k)$ are the ordered z_{ij} $(1 \leqslant j \leqslant n)$ variates. Since complete sums are invariant to ordering, the likelihood equations for estimating μ, τ_i $(1 \leqslant i \leqslant k)$ and σ can be expressed in terms of $z_{i(j)}$ as follows:

$$\frac{\partial \ln L}{\partial \mu} = -N\frac{c_2}{\sigma} + 2\frac{c_2}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n} g(z_{i(j)}) = 0,$$

$$\frac{\partial \ln L}{\partial \tau_i} = -n\frac{c_2}{\sigma} + 2\frac{c_2}{\sigma}\sum_{j=1}^{n} g(z_{i(j)}) = 0 \tag{2.4}$$