Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 279–290
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Tests for assessment of agreement using probability criteria

Pankaj K. Choudhary[a,*], H.N. Nagaraja[b]

[a]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083-0688, USA
[b]Department of Statistics, Ohio State University, Columbus, OH 43210-1247, USA

Received 18 May 2004; accepted 11 November 2005
Available online 19 January 2006

Abstract

For the assessment of agreement using probability criteria, we obtain an exact test, and for sample sizes exceeding 30, we give a bootstrap-*t* test that is remarkably accurate. We show that for assessing agreement, the total deviation index approach of Lin [2000. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. Statist. Med. 19, 255–270] is not consistent and may not preserve its asymptotic nominal level, and that the coverage probability approach of Lin et al. [2002. Statistical methods in assessing agreement: models, issues and tools. J. Amer. Statist. Assoc. 97, 257–270] is overly conservative for moderate sample sizes. We also show that the nearly unbiased test of Wang and Hwang [2001. A nearly unbiased test for individual bioequivalence problems using probability criteria. J. Statist. Plann. Inference 99, 41–58] may be liberal for large sample sizes, and suggest a minor modification that gives numerically equivalent approximation to the exact test for sample sizes 30 or less. We present a simple and accurate sample size formula for planning studies on assessing agreement, and illustrate our methodology with a real data set from the literature.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Bootstrap; Concordance correlation; Coverage probability; Limits of agreement; Total deviation index; Tolerance interval

1. Introduction

Suppose a paired sample of reference and test measurements, (X, Y), are available on n randomly chosen subjects from a population of interest. Generally the Y's are cheaper, faster, easier or less invasive to obtain than the X's. The question of our concern is: "Are X and Y close enough so that they can be used interchangeably?" This comparison is the goal of method comparison studies. We focus on a test of hypotheses approach for this problem that infers *satisfactory agreement* between Y and X when the difference $D = Y - X$ lies within an acceptable margin with a sufficiently high probability. We will assume that D follows a $\mathrm{N}(\mu, \sigma^2)$ distribution. Let $F(\cdot)$ and $\Phi(\cdot)$ denote the c.d.f.'s of $|D|$ and a $\mathrm{N}(0, 1)$ distribution, respectively.

There are two measures of agreement based on the probability criteria. The first is the p_0th percentile of $|D|$, say $Q(p_0)$, where p_0 (> 0.5) is a specified large probability (usually $\geqslant 0.80$). It was introduced by Lin (2000) who called it the *total deviation index* (TDI). Its small value indicates a good agreement between (X, Y). The TDI can be

* Corresponding author. Tel.: +1 972 883 4436; fax: +1 972 883 6622.
E-mail addresses: pankaj@utdallas.edu (P.K. Choudhary), hnn@stat.ohio-state.edu (H.N. Nagaraja).

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.11.009

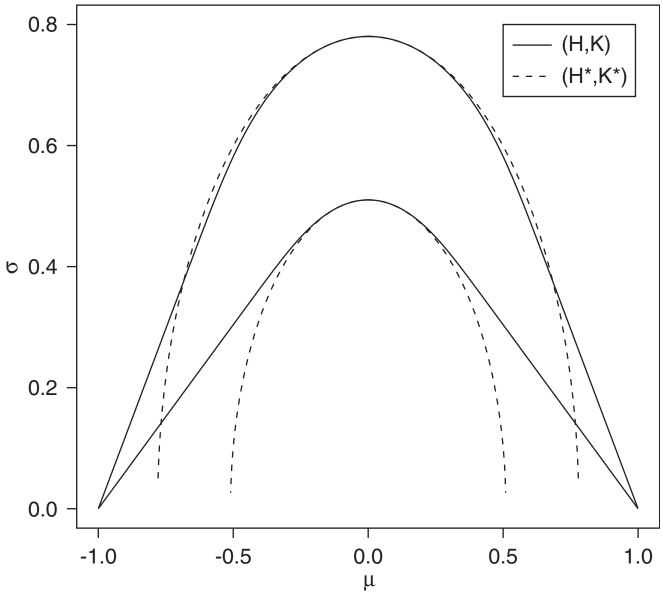

Fig. 1. The regions under the hypotheses (H, K) of our interest, given by (4) or (5), and those under (H*, K*) of Lin (2000), given by (7). The top two curves represent the boundaries of the two null regions for $p_0 = 0.80$, and the bottom two for $p_0 = 0.95$. The alternative regions lie under the respective curves. Here we have taken $\delta_0 = 1.0$, so that the x- and the y-axes actually represent μ/δ_0 and σ/δ_0, respectively.

expressed as,

$$Q(p_0) = F^{-1}(p_0) = \sigma\{\chi_1^2(p_0, \mu^2/\sigma^2)\}^{1/2}, \tag{1}$$

where $\chi_1^2(p_0, \Delta)$ is the p_0th percentile of a χ^2-distribution with single degree of freedom and non-centrality parameter Δ.

The second measure, introduced by Lin et al. (2002), is the *coverage probability* (CP) of the interval $[-\delta_0, \delta_0]$, where a difference under $\pm\delta_0$ is considered practically equivalent to zero. There is no loss of generality in taking this interval to be symmetric around zero as it can be achieved by a location shift. Letting,

$$d_l = (-\delta_0 - \mu)/\sigma, \quad d_u = (\delta_0 - \mu)/\sigma, \tag{2}$$

the CP can be expressed as

$$F(\delta_0) = \Phi(d_u) - \Phi(d_l). \tag{3}$$

A high value of $F(\delta_0)$ implies a good agreement between the methods.

For specified (p_0, δ_0), $F(\delta_0) \leqslant p_0 \iff Q(p_0) \geqslant \delta_0$. Consequently, for assessing agreement one can test either the hypotheses

$$\text{H: } Q(p_0) \geqslant \delta_0 \quad \text{vs.} \quad \text{K: } Q(p_0) < \delta_0, \tag{4}$$

or

$$\text{H: } F(\delta_0) \leqslant p_0 \quad \text{vs.} \quad \text{K: } F(\delta_0) > p_0, \tag{5}$$

and infer satisfactory agreement if H is rejected.

Let $\Theta = \{(\mu, \sigma) : |\mu| < \infty, 0 < \sigma < \infty\}$ be the parameter space. Also let Θ_{H} and Θ_{K} be the subsets of Θ representing the regions under H and K, respectively, and Θ_{B} be the boundary of H. These regions can be visualized through the solid curves in Fig. 1 for $p_0 = 0.80, 0.95$. Note that they are symmetric in μ about zero.

Lin (2000) suggests a large sample test for the hypotheses (4) and we refer to it as the TDI test. Lin et al. (2002) suggest a large sample test for (5), and we call it the CP test. They also conclude that these tests are more powerful for inferring

Download English Version:

https://daneshyari.com/en/article/1148727

Download Persian Version:

https://daneshyari.com/article/1148727

Daneshyari.com