Journal of Multivariate Analysis 104 (2012) 115–125

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A test of location for exchangeable multivariate normal data with unknown correlation

Dean Follmann *, Michael Proschan

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, 6700B Rockledge Drive MSC 7609, Bethesda, MD 20892, United States

ARTICLE INFO

Article history:
Received 12 February 2009
Available online 28 July 2011

AMS subject classifications:
62F05
62H15
62H10

Keywords:
Confidence interval
Within cluster resampling
Likelihood ratio test

ABSTRACT

We consider the problem of testing whether the common mean of a single n-vector of multivariate normal random variables with known variance and unknown common correlation ρ is zero. We derive the standardized likelihood ratio test for known ρ and explore different ways of proceeding with ρ unknown. We evaluate the performance of the standardized statistic where ρ is replaced with an estimate of ρ and determine the critical value c_n that controls the type I error rate for the least favorable ρ in [0,1]. The constant c_n increases with n and this procedure has pathological behavior if ρ depends on n and ρ_n converges to zero at a certain rate. As an alternate approach, we replace ρ with the upper limit of a $(1 - \beta_n)$ confidence interval chosen so that $c_n = c$ for all n. We determine β_n so that the type I error rate is exactly controlled for all ρ in [0,1]. We also investigate a simpler approach where we bound the type I error rate. The former method performs well for all n while the less powerful bound method may be a useful in some settings as a simple approach. The proposed tests can be used in different applications, including within-cluster resampling and combining exchangeable p-values.

Published by Elsevier Inc.

1. Introduction

Consider the testing problem, where we have a single observation $\boldsymbol{Z} = Z_1, \ldots, Z_n$ from a multivariate normal with common mean μ and unit variance but unknown common correlation $\rho \geq 0$. We are interested in testing whether $E[Z_i] = \mu$ is zero. This differs from more standard multivariate testing paradigms where Σ is completely known, known up to a constant, or completely unknown.

This setting can arise in various situations. First, consider how to combine exchangeable or dependent p-values. Kost & McDermott [5,6] evaluated different methods for combining dependent p-values, but assumed that the covariance matrix was known or known up to a constant. Suppose we have n exchangeable p-values, $U_1, \ldots, U_n$. The transformed values $Z_i = \Phi^{-1}(U_i)$ are, on the null hypothesis, marginal standard normal and jointly exchangeable. If the Z_is are additionally multivariate normal, we have the testing situation described above. This kind of setting can arise, for example in drug licensure settings, where sometimes sponsors synthetically create multiple clinical trials from the same umbrella protocol. Here, some sites are assigned to be Trial *A* and others Trial *B*. One might view the p-values arising from such trials as not being completely independent, but exchangeable, and the effect on inference is of interest. Another setting is in the application of within cluster resampling (WCR) [4,3]. One application of this algorithm is when a large number n of test statistics are created that are exchangeable and standard normal on the null hypothesis. The tests are averaged and standardized, and an asymptotic (on n) standard normal null distribution is used for overall inference. In this paper, we explore the peculiar behavior of this approach under seemingly large n and selected ρ. We also indicate how the methods of this paper can be used for accurate inference for WCR when n is seemingly large.

* Corresponding author.
E-mail address: dfollmann@niaid.nih.gov (D. Follmann).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2011.07.003

In this paper we derive the likelihood ratio test for this setting for known ρ and show that properly standardized, it follows a standard normal distribution on the null that $\mu = 0$. The test statistic depends on the unknown common correlation ρ, but ρ can be consistently estimated by a function of the sample variance of the Z_is, $\hat{\rho} = \max(0, 1 - S_Z^2)$. We explore use of the standardized test statistic based on replacing ρ with either $\hat{\rho}$ or an upper confidence interval for ρ, say $\hat{\rho}_\beta$. We derive the exact distribution of such "plug-in" test statistics, which depends on the unknown ρ, and define testing procedures that control the type I error for the least favorable $\rho \in [0, 1]$. We also derive a simple bound that controls the type I error rate when ρ is replaced with the upper confidence limit. We evaluate these approaches using simulation and limiting behavior and we give some examples to illustrate the method.

2. Proposed tests

Assume that the Z_i are multivariate normal with $E(Z_i) = \mu$, $\mathrm{var}(Z_1) = 1$, and $\mathrm{corr}(Z_i, Z_j) = \rho \geq 0$. Suppose we are interested in testing

$$H_0 : \mu = 0.$$

For fixed ρ, one can show that the likelihood ratio test is a monotone function of the sample average $\overline{Z}$. It is easy to show that $E(\overline{Z}) = \mu$ and $\mathrm{var}(\overline{Z}) = \{1 + (n-1)\rho\}/n$. Thus, a likelihood ratio test can be fashioned by forming

$$\frac{\overline{Z}}{\sqrt{\{1 + (n-1)\rho\}/n}}, \tag{1}$$

which has a standard normal distribution under H_0. Note that (1) can be used to test either a one-sided e.g. $H_0 : \mu \geq 0$, or two sided $H_0 : \mu \neq 0$ alternative.

Since ρ is unknown, we need an estimator. Consider the mle. In the Appendix we show that the mle for μ is $\overline{Z}$ and that the mle for ρ is given by 0, 1 or

$$\tilde{\rho} = \frac{1}{2}(1 - \tilde{S}^2) \pm \frac{1}{2}\sqrt{(1 - \tilde{S}^2)^2 - \frac{4\tilde{S}^2}{n-1}},$$

where $\tilde{S}^2 = \sum(Z_i - \overline{Z})^2/n$. For large n, one of the solutions is close to $1 - \tilde{S}^2$ while the other is close to 0.

If we replace ρ with the mle, the resultant distribution of the standardized test statistic is hard to work with due to the complicated form of $\tilde{\rho}$. Additionally, one can show that $\tilde{\rho}$ is a biased estimate of ρ. This leads us to consider different estimates and approaches. To facilitate our development, think of the Z_i as

$$\begin{aligned} Z_i &= U + V_i \\ U &\sim N(0, \sigma_U^2) \\ V_i &\sim N(\mu, \sigma_V^2), \end{aligned} \tag{2}$$

where $(U, V_1, \ldots, V_n)$ are independent, and

$$\sigma_U^2 + \sigma_V^2 = 1.$$

With this representation,

$$\rho = \mathrm{corr}(Z_i, Z_j) = \frac{\sigma_U^2}{\sigma_U^2 + \sigma_V^2} = \sigma_U^2 = 1 - \sigma_V^2 \geq 0.$$

If we could estimate σ_V^2 then we could estimate ρ. While we do not directly observe the V_is, note that the sample variance of the Zs or S_Z^2 must equal the sample variance of the V_is because U is a constant offset and the sample variance is unchanged if a constant is added to all the data. Thus we can obtain an unbiased estimate of ρ by using simply $1 - S_Z^2$, which is pleasingly similar to the limit of one of the solutions of the mle; $1 - \tilde{S}^2$. Further justification for restricting estimators of ρ to functions of S_Z^2 is given in the Appendix.

2.1. Plug-in approaches to testing

In this subsection we explore use of estimators of ρ that are functions of S_Z^2 and result in tractable distributions for the standardized test statistic (1) with ρ replaced with an estimator.

Using the representation $Z_i = U + V_i$ we have that $S_Z^2 = S_V^2$, from which we deduce that

$$\frac{(n-1)s_Z^2}{1-\rho} \sim \chi^2 \quad \text{with } n-1 \text{ d.f.}$$

This representation will allow us to derive the exact distribution of test statistics that replace ρ in (1) with either an estimate, $\hat{\rho} = \max(0, 1 - S_Z^2)$, or an upper confidence limit. To begin, consider the plug-in estimate where we reject if

$$Z_{\hat{\rho}} = \frac{\overline{Z}}{\sqrt{\{1 + (n-1)\hat{\rho}\}/n}} > c, \tag{3}$$