Small Ruminant Research 83 (2009) 34–41

Contents lists available at ScienceDirect

Small Ruminant Research

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

A bootstrap approach to estimate reference intervals of biochemical variables in sheep using reduced sample sizes

Corrado Dimauro[a,*], Nicolò P.P. Macciotta[a], Salvatore P.G. Rassu[a], Cristiana Patta[b], Giuseppe Pulina[a,c]

[a] Dipartimento di Scienze Zootecniche, Università di Sassari, Italy
[b] Istituto Zooprofilattico della Sardegna "G. Pegreffi", Sassari, Italy
[c] AGRIS Sardegna, Olmedo (SS), Italy

ARTICLE INFO

Article history:
Received 8 December 2008
Received in revised form 6 March 2009
Accepted 9 March 2009
Available online 10 April 2009

Keywords:
Bootstrap
Biochemical variable
Reference interval
Sample size
Sheep

ABSTRACT

Laboratory analyses represent a key element in veterinary medicine diagnosis providing objective information about the health status of a patient. Analytic data are interpreted by comparing them with a specific reference intervals previously determined on a reference population of healthy animals. The International Federation of Clinical Chemistry recommends the use of nonparametric methods and, as a consequence, a reference sample of at least 120 healthy subjects, to obtain reliable reference intervals. Such order of magnitude for the reference sample is not always feasible especially if the laboratory variable under study is affected by several sources of variation, e.g., environmental conditions, physiological status of the animal, age, or gender. A convenient method to estimate reference intervals should be able to avoid assumptions on the probability distribution of the considered variable and produce robust results even with a limited sample size.

This study presents a new statistical approach, based on data bootstrap, to estimate reference intervals for 12 blood biochemical variables in Sarda dairy sheep. The method was applied to real and simulated data from 120 to 40 animals. The reference intervals calculated with the new method remained quite constant as sample size decreased from 120 down to 60 animals, and became wider with fewer individuals. So, a minimum threshold of 60 animals could be considered a good limit to obtain reliable reference intervals for blood biochemical variables in Sarda dairy sheep. Moreover, comparisons between results from real and simulated data suggested that the method could be also applied to other laboratory variables.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Laboratory analyses are a key element in medicine diagnosis, because their results provide objective information about the health status of a patient. This is fundamental in veterinary medicine, where it is particularly difficult to assess the healthy status of an animal, since nobody knows how an animal really feels. In both cases, analytical data are interpreted by comparing them with specific reference intervals (RIs) previously estimated on a reference sample (RS) extracted from a reference population of healthy subjects.

The first drawback in estimating RI of a given variable is the choice of the most suitable statistical method for estimating it (Grossi et al., 2005). There are two possible ways: the parametric and the nonparametric approach. From the theoretical point of view, the parametric approach would be better than the nonparametric one, because it is a well assessed procedure and does not require a large RS (Solberg, 2004). However, it can be applied only if the

* Corresponding author at: Dipartimento di Scienze Zootecniche, facoltà di Agraria, Università di Sassari, Via De Nicola 9, 07100 Sassari, Italy. Tel.: +39 079 229304; fax: +39 079 229302.
E-mail address: dimauro@uniss.it (C. Dimauro).

0921-4488/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.smallrumres.2009.03.004

variable under investigation has a normal distribution and this widely limits its use. Actually, testing a variable for normality is often very difficult because it may have a normal distribution in the reference population, but not in the RS, due to incorrect sampling procedures or to stochastic drift. When the variable is not normally distributed, mathematical transformations of the data can be used to normalize it, but, as demonstrated by Linnet (1987), some types of data distribution cannot be normalized, often due to bi-modal or multi-modal distribution of the variable. As a consequence, the International Federation of Clinical Chemistry (IFCC) recommends the use of nonparametric methods to estimate RIs, since they do not make *a priori* assumptions on the probability distribution of the considered variable. However, the choice of this approach leads to other problems. First of all, even if several nonparametric methodologies are available (Horn and Pesce, 2003), none of them has been considered the most suitable one by the scientific community. Secondly, these nonparametric methods require a large RS of healthy individuals (Horn and Pesce, 2003; Solberg, 2004) and become robust when the sample size is as high as 200 units (Horn et al., 1998). For these reasons, the IFCC recommends the extraction of a RS of at least 120 healthy individuals from the reference population that represents a specific typology of patient (Solberg, 2004). However, if the variable under study is affected by several factors, e.g. environmental conditions, physiological status of the subject, age, or gender, it can be costly and difficult to achieve group sizes of that order of magnitude (Horn and Pesce, 2003). Therefore, new methods to evaluate RIs using small samples are needed.

A convenient method for estimating a RI should be able to avoid assumptions on the probability distribution of the considered variable and produce robust results even with a limited sample size. With this aim, in this work we propose the use of the bootstrap technique, implemented with the parametric approach, to obtain an easy and single method to estimate RIs for blood biochemical variables in Sarda dairy sheep. Bootstrap is a computer intensive resampling technique that bases its rationale on the central limit theorem, thus allowing to make inferences on a given statistic (mean, median, etc.) using the parametric method, without considering the probability distribution of the variable.

With the improvement of the computer performances, the bootstrap technique has been increasingly utilized in several fields of applied biology (e.g. human medicine, molecular biology and genetics), as well as in other sciences (e.g. astronomy, economics and engineering) (Henderson, 2005).

The aims of this work were: (i) to develop a single method to estimate RIs for 12 blood biochemical variables in Sarda dairy sheep using the bootstrap technique implemented with the parametric approach and (ii) to evaluate the effects of a reduction of the required sample size of 120 healthy animals on RI estimates.

2. Materials and methods

2.1. Classic determination of reference intervals

In this study the term "classic RIs" indicates RIs estimated using the parametric or nonparametric approach according to the suggestions of the IFCC. The choice between the two statistical techniques depends on the distribution shape of each parameter, which can be tested by using the normal probability plot and the Anderson–Darling normality test (Linnet, 1988). For normally distributed variables, RIs and their 95% confidence intervals ($CI_{95\%}$) for the lower and upper bounds are estimated using the well-known parametric procedure (Solberg, 2004). If m and s are the mean and the standard deviation of the RS, the parametric RI and the $CI_{95\%}$ for its limits are calculated as:

$$\mathrm{RI} = m \pm 2s \quad \mathrm{CI}_{95\%} = (m \pm 2s) \pm 1.96\sqrt{s^2\left(\frac{1}{n} + \frac{2}{n-1}\right)},$$

where n is the size of the RS and m and s are independent with variances s^2/n and $s^2/2(n-1)$, respectively.

If the probability distribution of the variable is not normal, the RI is estimated using the nonparametric approach. In this study, the chosen nonparametric method was suggested by the IFCC (Solberg, 2004), which calculates the RI using the simple rank-based procedure described by Reed et al. (1971). In brief, values of a variable are ordered according to increasing value and assigned rank numbers. The rank number of the α percentile is calculated as $\alpha(n+1)/100$, where n is the sample size, whereas the $CI_{95\%}$ for the α percentile can be found by using the binomial distribution method (Conover, 1999; Garcia-Perez, 2005).

2.2. The bootstrap method

In this section we examine a new method to estimate RIs that could be used in alternative to the classic methods. This method combines the parametric approach with the bootstrap resampling technique (the bootstrap method).

The sampling distribution of a statistics, such as the mean, is based on many samples being randomly extracted from the population. In the bootstrap approach, different samples are generated by repeatedly sampling with replacement from the same sample. Sampling with replacement means that after an observation is randomly drawn from the RS, it is put back before the next observation is drawn. As a result, any number can be drawn more than once, or not at all, and every number has the same probability of being chosen. In this way, each resample is as large as the original RS and, in practice, hundreds of thousands of resamples can be drawn.

Let $x_1, x_2, \ldots, x_n$ be a set of independent observations measured on n subjects and $\bar{x}$ the sample mean. The estimated standard error (SE) of the mean is

$$\mathrm{SE} = \frac{s}{\sqrt{n}} \tag{1}$$

where the standard deviation is calculated as $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$.

Let $\bar{y}_j$, for $j = 1, 2, \ldots, m$, be the mean of each m-bootstrap sample and $\bar{y}_{\mathrm{boot}}$ the mean of the bootstrap distribution. Then the bootstrap standard error (SE_{boot}), i.e. the standard deviation of the bootstrap distribution, is

$$\mathrm{SE}_{\mathrm{boot}} = \sqrt{\frac{\sum_{j=1}^{m}(\bar{y}_j - \bar{y}_{\mathrm{boot}})^2}{m-1}}.$$

SE_{boot} represents the best estimation of the SE of the mean (or of any other statistics) (DiCiccio and Efron, 1996). Therefore, by replacing SE with SE_{boot} in the formula (1), the bootstrap sampling standard deviation s_{boot} can be estimated as:

$$s_{\mathrm{boot}} = \mathrm{SE}_{\mathrm{boot}}\sqrt{n}$$

The bootstrap accuracy is generally evaluated using the bias, i.e. the difference between the sample mean and the bootstrap estimate of the mean ($\mathrm{bias} = \bar{x} - \bar{y}_{\mathrm{boot}}$). The root mean square error of the mean (RMSE) is calculated as:

$$\mathrm{RMSE} = \sqrt{\mathrm{bias}^2 + \mathrm{SE}_{\mathrm{boot}}^2}$$

As a rule of thumb, Efron and Tibshirani (1993) suggested that if the absolute value of the bias is smaller than $0.25\mathrm{SE}_{\mathrm{boot}}$, the bias can be ignored and therefore $\mathrm{RMSE} \approx \mathrm{SE}_{\mathrm{boot}}$.

Finally, the bootstrap reference interval (RI_{boot}) of a given variable and the $CI_{95\%}$ of its limits can be calculated as:

Download English Version:

https://daneshyari.com/en/article/2457798

Download Persian Version:

https://daneshyari.com/article/2457798

Daneshyari.com