Chemometrics and Intelligent Laboratory Systems 130 (2014) 98–108

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

On robust testing for normality in chemometrics

M. Stehlík [a,*], L. Střelec [b], M. Thulin [c]

[a] Department of Applied Statistics, Johannes Kepler University in Linz, Austria
[b] Department of Statistics and Operation Analysis, Mendel University in Brno, Czech Republic
[c] Department of Mathematics, Uppsala University, Sweden

ARTICLE INFO

Article history:
Received 12 July 2013
Received in revised form 23 September 2013
Accepted 19 October 2013
Available online 31 October 2013

Keywords:
Trimming
Lehmann–Bickel functional
Model diagnostics
Monte Carlo simulations
Power comparison
Robust tests for normality

ABSTRACT

The assumption that the data has been generated by a normal distribution underlies many statistical methods used in chemometrics. While such methods can be quite robust to small deviations from normality, for instance caused by a small number of outliers, common tests for normality are not and will often needlessly reject normality. It is therefore better to use tests from the little-known class of robust tests for normality. We illustrate the need for robust normality testing in chemometrics with several examples, review a class of robustified omnibus Jarque–Bera tests and propose a new class of robustified directed Lin–Mudholkar tests. The robustness and power of several tests for normality are compared in a large simulation study. The new tests are robust and have high power in comparison with both classic tests and other robust tests. A new graphical method for assessing normality is also introduced.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Classic parametric statistical significance tests, such as analysis of variance and least squares regression, are widely used by researchers in many disciplines of chemistry. For classic parametric tests to produce accurate results, the assumptions underlying them (e.g., normality and homoscedasticity) must be satisfied. These assumptions are rarely met when analyzing real data. The use of classic parametric methods with violated assumptions can result in inaccurate computations of p-values, effect sizes, and confidence intervals. This may lead to substantial errors in the interpretation of data.

For this reason, model diagnostics in general and testing for normality in particular are very important issues in chemometrics. But it is often the case that it is not necessary for the underlying distribution to be exactly normal for a statistical method to be valid. Except for situations where the sample size is extremely small, the question that really is of interest is whether the distribution is approximately normal. For an overview and discussion of robust chemometrical statistical methods, both parametric and non-parametric, see Refs. [30] and [14] and the references therein.

A test for normality that is less sensitive to small deviations from normality, particularly in the form of a few "bad" observations, is called robust. A drawback of virtually all common tests for normality is that they lack robustness and are far too sensitive to outliers, rejecting normality even when methods that require normality would be applicable. In recent years, several studies on robustification of tests for normality, aiming to correct such drawbacks, have appeared in the literature; see Refs. [12,7,11,33].

While books on, e.g., analytical chemistry generally contain a section devoted to normality testing (see e.g., Refs. [25,5]), there are still several open problems related to robust testing for normality, both in theory and in practice. The aim of this paper is to contribute to this discussion. The next section illustrates the necessity of robust testing for normality in chemometrics. Therein we discuss, using real data examples, the importance of robust testing for normality in measuring of mycolic acid and troposphere methane modeling and methane emissions from sedge-grass marsh. In Section 3 we review the recently introduced class of (omnibus) robustified Jarque–Bera tests and introduce a class of (directed) robustified Lin–Mudholkar tests. In Section 4 we conduct an extensive comparison of the robustness and power of several tests for normality. Section 5 addresses robust graphical methods for assessing normality. The paper concludes with a discussion, in which practical guidelines for robust testing of normality are given and the merits of omnibus and directed tests are compared. To maintain the continuity of explanation, proofs and technicalities are deferred to an Appendix A.

We emphasize that in this paper we assume that the possible contamination of the sample is due to outliers. Thus we use two techniques for outlier filtering: trimming and the functional approach introduced by Bickel and Lehmann (see Ref. [4]). For normality testing when the whole distribution may be contaminated, i.e., when any quantile level of the distribution may be contaminated see e.g., Ref. [1].

* Corresponding author at: Department of Applied Statistics, Johannes Kepler University in Linz, Altenberger Strasse 69, 4040 Linz a. D., Austria. Tel.: +43 732 2468 6806; fax: +43 732 2468 9846.
E-mail addresses: Milan.Stehlik@jku.at (M. Stehlík), lubos.strelec@mendelu.cz (L. Střelec), mans.thulin@math.uu.se (M. Thulin).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.10.010

2. Motivation of robust testing for normality in chemometrics

In this section we illustrate some problems that arise when classic tests for normality, such as the Shapiro–Wilk and Jarque–Bera tests, are used in chemometrics. This shows the need for robust testing for normality in chemometrics.

2.1. Mycolic acid: A data set contaminated with a single outlier

Mycolic acids are the major components of the cell walls of *Mycobacterium tuberculosis* and their biochemical properties are paramount to the pathogenesis and survival of these bacteria. For this reason, attempts have been made to create drugs that inhibit mycolic acid synthesis. A data set consisting of 26 measurements of mycolic acids in *M. tuberculosis* is visualized in Fig. 1. The sample is contaminated by a single outlier. For this data set, the popular Shapiro–Wilk test for normality gives a p-value of 0.006. However, if the outlier is removed the p-value is 0.41. Clearly a single outlier can have a huge effect on the result of the Shapiro–Wilk test. In many situations, it would be of greater interest to have a test that looks at the overall shape of the empirical distribution rather than a test that rejects normality because of a small deviation in the form of an outlier. An example of such a test is the RLM_γ test presented in Section 3, that has higher power than the Shapiro–Wilk test in many settings, but yields a p-value of 0.98 for the 26 mycolic acid measurements.

2.2. Methane modeling in the troposphere and on Earth

Statistical modeling plays a central role in studies of methane emission and methane absorption, both in the atmosphere and on the ground. For examples, see e.g., Ref. [40] for methane in the troposphere, Ref. [21] regarding the methane emissions from natural wetlands or Ref. [19] for modeling of the methane emission from a sedge-grass marsh in South Bohemia. Such a modeling is of tremendous complexity and typically requires several distributional assumptions when statistical learning is desired.

It is understandable that outliers from any reasonable distribution are expected and thus the need for robust testing arises. Even for constructing optimal sampling plans we need distributional assumptions, e.g., on error structure (see Ref. [29]). Thus robust testing for normality can be a very useful tool for model diagnostics in this setup.

For illustratory purposes, we consider four methane data sets. The first is the flux rate of methane k_1, taken from Table 1 in Ref. [40]. For this data, both the popular Shapiro–Wilk and Jarque–Bera tests for normality reject normality (the p-values are <0.001 and 0.002, respectively), as does the new robust RLM_γ test (its p-value being <0.001). The non-normality is likely not due to outliers, but to the dependence structure, which was modeled in Ref. [29].

The remaining data sets are from a study of methane emission from a sedge-grass marsh in South Bohemia. The residuals Z,Z−,Z+ of methane emissions were taken from the infinite moving average model (8) in Ref. [19], where only time is taken as a regressor. The Shapiro–Wilk, Jarque–Bera and RLM_γ tests reject normality for all three sets of residuals, with p-values < 0.001. The reason for this non-normality of data is not outliers, but a heavy-tailed pattern, described in Ref. [19].

2.3. Asking the right question

Summarizing our observations from the previous examples, we conclude that it is rare to handle exact normality in experiments. For virtually all statistical methods the inference will however also be valid for approximately normal random variables. The question that we should ask is therefore not "are these random variables normal?" but "are these random variables normal enough?" What can be considered "normal enough" depends on the statistical method and the sample size.

Consider, for instance, Student's t-test. Assume that we have a sample from a random variable X. As it turns out, the non-normality of X can be quantified using the concepts of asymmetry and peakedness. The asymmetry of X is usually measured by its skewness $\gamma = \frac{E(X-\mu)^3}{\sigma^3}$. If X is symmetric about its mean, $\gamma = 0$. If X "leans to the right" then $\gamma > 0$, and we say that X is right-skew. Similarly, X is left-skew if $\gamma < 0$. The peakedness is measured by the (excess) kurtosis $\kappa = \frac{E(X-\mu)^4}{\sigma^4} - 3$. If X is normal, $\kappa = 0$, whereas short-tailed distributions tend to have $\kappa < 0$ and heavy-tailed distributions tend to have $\kappa > 0$.

To understand how skewness and kurtosis can be used to measure non-normality in the context of Student's t-test, we need some tools from theoretical statistics. If T_n is the test statistic of Student's t-test, we can, using a so-called Edgeworth expansion (Ref. [16]), obtain the following approximation of the null distribution of T_n:

$$P(T_n \le x) \approx \Phi(x) + n^{-1/2}\frac{1}{6}\gamma\left(2x^2+1\right)\phi(x) - n^{-1}x\left(\frac{1}{12}\kappa\left(x^2-3\right) - \frac{1}{18}\gamma^2\left(x^4+2x^2-3\right) - \frac{1}{4}\left(x^2+3\right)\right)\phi(x),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $\phi(\cdot)$ is its density function. We can therefore see how skewness and kurtosis affect the null distribution of the test statistic.

When X truly is normal, $\gamma = \kappa = 0$ and we see that $P(T_n \le x) \approx \Phi(x)$. If X is non-normal and γ or κ are nonzero, however, $P(T_n \le x)$ is perturbed by skewness or kurtosis and the size and p-values of the test will no longer behave as desired. The performance of T_n is therefore sensitive to deviations from normality in the form of skewness and kurtosis.

If we wish to investigate whether X is "normal enough" it makes sense to measure non-normality in terms of skewness and kurtosis, as these quantities directly determine how good an approximation for the null distribution of the test statistic is. This is equally true for many other statistical procedures — in general methods based on the normality assumptions work very well for distributions with low γ and κ even if there are a small number of outliers. It thus seems desirable to have tests for normality that are based on estimates of skewness and kurtosis. On a side note, we mention that the influence of different shapes of distribution with the same first 4 moments on robustness has

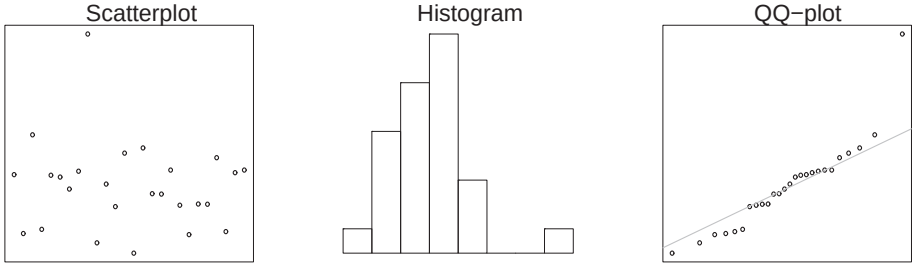

Fig. 1. Mycolic acid measurements with an outlier.

Download English Version:

https://daneshyari.com/en/article/1180673

Download Persian Version:

https://daneshyari.com/article/1180673

Daneshyari.com