Psychology Science, Volume 46, 2004 (4), p. 544 - 548

Robustness is parameter-specific
A comment on Rasch and Guiard's robustness study

ALEXANDER VON EYE[1]

Abstract

In this comment on an article by Rasch and Guiard (2004), it is argued that the selection of nonparametric over parametric tests cannot rest solely on the characteristic of tests as robust against non-normality. Other violations exist. It is shown that the t-test, which is known to be virtually immune to violations of the normality assumption, is highly sensitive to violations of the independence assumption. It is concluded that a dismissal of nonparametric tests that is based on the result that tests are robust against normality violations, may be premature. More research is encouraged that addresses issues of violations of assumptions. This research should include both parametric and nonparametric tests.

Key words: robustness, normality assumption, distribution-free, t-test, autocorrelation

[1] Alexander von Eye, Michigan State University, Department of Psychology, 107 D Psychology Building, East Lansing, MI 48824-1116, USA; E-mail: voneye@mu.edu. Thanks go to G. Anne Bogat for helpful comments on an earlier draft of this comment.

Robustness is one of the rather important characteristics of methods of statistical decision-making. Statistical tests are considered robust if they keep their characteristics under adverse conditions, that is, under conditions in which the assumptions are violated on which the derivation of the tests rests. For many tests, distribution-free alternatives have been formulated. For example, the Wilcoxon test is a distribution-free alternative to Student's t-test, or Friedman's method is a distribution-free alternative to one-way analysis of variance.

It is well known that tests can be robust against one violation while being sensitive to others. One of the most powerful tests is Student's t-test. This test is widely known as being robust against violations of the normality assumption. Rasch and Guiard (2004) present impressive results both at the theoretical and the applied levels that show that the t-test is extremely robust against violations of the normality assumption. Indeed, the authors present a very concise and illustrative specification of the well known result (Bartlett, 1935) that the t-test has a nonparametric property over a parametric class consisting of normally distributed variables (Randless & Wolfe, 1979; von Eye, 1988). It has been said that the t-test is asymptotically nonparametric (cf. Lehmann, 1975). The authors therefore conclude that there is practically no need for such alternatives as the Wilcoxon test.

In this brief comment, we attempt to show that alternative tests may be needed after all. We show that the t-test, while almost immune to violations of normality, is highly sensitive to violations of the assumption of independence. Drawings are independent if no case systematically carries information that another case also carries. Specifically, we present the results of a simulation study in which we vary the degree of autocorrelation and show that the distribution of the t-statistic depends systematically on the degree of autocorrelation (cf. von Eye, 1983).

1. A simulation of the t-distribution under various degrees of autocorrelation

Consider the first order stationary autoregressive process

$$x_i = E(x) + \varphi(x_{i-1} - E(x)) + \varepsilon_i$$

for i = 1, ..., n, the sample size, where $E(x)$ is the expectancy of x, φ is the first order autocorrelation, the x_i are random variates, and ε_i is the $N(0; 1)$ random error. Without loss of generality, we set $E(x) = 0$. Because we have stationarity, $|\varphi| < 1$.

In the following illustration, we focus on Rasch and Guiard's experiment 1, case 2, that is, the one-sample t-test. The simulation followed the steps outlined by von Eye (1983):

1) the sample size n varied from 5 to 40, in steps of 5;
2) the autocorrelation varied from -0.8 to +0.8, in steps of 0.2 (note that this selection of values differs from the one in the earlier simulation, because here, we wanted φ = 0 to be part of the autocorrelation values);
3) the autocorrelated series was calculated beginning with $x_1 = \varepsilon_1$. The following scores in the series were $x_i = \varphi x_i + \varepsilon_i$, for i = 2, ..., n. Also in deviation from von Eye's (1983) study in which a random number generator was used that was provided by the system, we here used the generator in the function GASDEV from the Numerical Recipes FORTRAN collection (Press, Flannery, Teukolsky, & Vetterling, 1989).

This generator returns a normally distributed deviate with zero mean and unit variance. It is based on the function RAN1, also provided in the recipe collection. This function uses three congruential generators (cf. the discussion of generators in Rasch & Guiard, 2004). The authors state that this generator's "period is (for all practical purposes) infinite and ought to have no sensible sequential correlations" (Press et al., 1989, p. 196). The latter characteristic is most important here because we hope that the equation given above specifies the only source of autocorrelation in our simulation;

4) for each series of size n and autocorrelation φ, the t-score $t = \frac{\bar{x}}{s}\sqrt{n}$ was calculated. The size of each data set was determined to be 2000 (and thus larger than in von Eye, 1983).

2. Results

We present results in two sections. First, we report ANOVA results; then, we examine the shapes of the t-distributions. For the ANOVAs, we treat the data as collected for a fixed effect 8(sample sizes) x 9(autocorrelation levels) design. Each of the cells of this design contains 2000 cases. The design is therefore orthogonal. For each of the sample sizes, there were thus 18,000 cases, and for each correlation level, there were 16,000 cases. The results of this analysis are summarized in Table 1.

Table 1 shows that the sources of variation explain basically nothing of the variation of the means of the t-distributions created in the simulation. The partial η^2 estimates are close to zero, the largest individual score being no bigger than 2 per mill, for the sample size factor. Accordingly, the estimated R^2 = 0.003 is also very low (adjusted R^2 = 0.003). We conclude that the mean parameter of the t-distribution remained unaffected by the simulation. Please note that we refrain from interpreting the significance of effects because, in this study, the sample sizes are that large that even the most minuscule effects turn out significant. The only exception is the effect for ρ which caused virtually no mean variation at all, hence the low power.

Table 1:
ANOVA of the sample size by autocorrelation level simulation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power
Corrected Model	1509.935	71	21.267	6.570	.000	.003	1.000
Intercept	34.241	1	34.241	10.577	.001	.000	.902
N	812.542	7	116.077	35.858	.000	.002	1.000
RHO	9.554	8	1.194	.369	.937	.000	.179
N * RHO	687.839	56	12.283	3.794	.000	.001	1.000
Error	465917.077	143928	3.237				
Total	467461.252	144000					
Corrected Total	467427.012	143999					

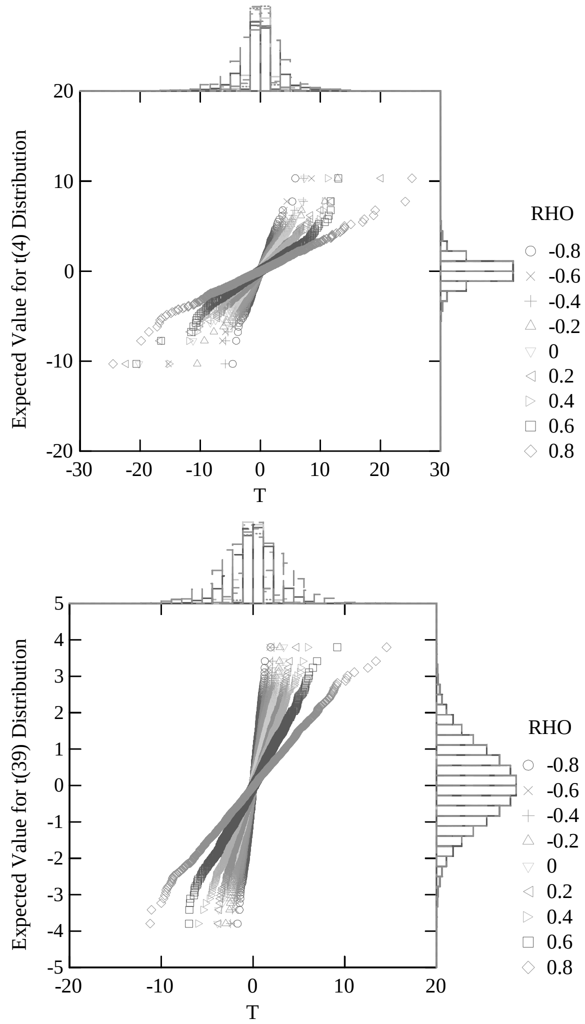

Figure 1:
t-distributions by sample size and autocorrelation level

Without going into more detail, we conclude from this part of the analyses that the means of the statistic of the one-sample t-test are indeed not sensible at all to the variations induced in the present simulations. We now ask, however, whether the shape of the distribution is also unaffected. To answer this question, we produce probability plots. Figure 1 displays two examples. The first example, presented in the left-hand panel, gives the nine curves for the t-statistic under the nine autocorrelation conditions, for $n = 5$. The second example, presented in the right-hand panel, does the same, for $n = 40$.

Figure 1 shows first that each of the various distributions of the t-statistic does indeed have a mean of zero. However, the shapes of the distributions depend on both the sample size and the autocorrelation level with the distortion being larger for smaller sample sizes. As a consequence, the percentiles of the distribution are not independent of the level of autocorrelation, and significance tests will suggest biased decisions. The magnitude of the bias varies with the magnitude of the autocorrelation.

2. Conclusion

While it is well known, and has repeatedly and convincingly been shown that the t-statistic is practically immune to violations of the normality assumption (Rasch & Guiard, 2004), the statistic is also known to be highly sensitive to other violations. In this comment, we showed that one such violation is the one against the independence assumption. Already in 1980, Brillinger had issued warnings against applying models that had been developed under the assumption of independence to data with dependence structures.

Based on these results, we wonder whether the dismissal of nonparametric tests is premature. Yes, these tests may be unnecessary if only violations of normality are suspected. However, other violations are lurking. The degree to which parametric as well as nonparametric tests are sensitive to such violations as dependency structures is largely underresearched. Methodologists should be encouraged to study the behavior of tests under these and many other conditions.

References

1. Bartlett, M.S. (1935). The effect of non-normality on the t-distribution. Proceedings of the Cambridge Philosophical Society, 31, 223 - 231.
2. Brillinger, D.R. (1980). Analysis of variance and problems under time series models. In: Krishnaiah, P.R. (Ed.), Handbook of Statistics, Vol. 1 (pp. 237-278). Amsterdam: North Holland.
3. Lehmann, E.L. (1975). Nonparametrics. San Francisco: Holden Day.
4. Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. (1989). Numerical recipes. The art of scientific programming (FORTRAN version). Cambridge, UK: Cambridge University Press.
5. Randless, R.H., & Wolfe, D.A. (1979). An introduction to the theory of nonparametric statistics. New York: Wiley.
6. Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. Psychology Science, 46, 175 - 208.
7. von Eye, A. (1983). t-tests for single means of autocorrelated data - a simulation study. Biometrical Journal, 25, 801-805.
8. von Eye, A. (1988). Some multivariate developments in nonparametric statistics. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (pp. 367-398). New York: Plenum Press.