Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: www.tandfonline.com/journals/cjas20

Robust confidence interval for a residual standard deviation

Douglas G. Bonett

To cite this article: Douglas G. Bonett (2005) Robust confidence interval for a residual standard deviation, Journal of Applied Statistics, 32:10, 1089-1094, DOI: 10.1080/02664760500165339

To link to this article: https://doi.org/10.1080/02664760500165339

Published online: 20 Aug 2006.

Journal of Applied Statistics
Vol. 32, No. 10, 1089–1094, December 2005

Routledge
Taylor & Francis Group

Robust Confidence Interval for a Residual Standard Deviation

DOUGLAS G. BONETT

Department of Statistics, Iowa State University, Ames, Iowa USA, 50011

ABSTRACT *The residual standard deviation of a general linear model provides information about predictive accuracy that is not revealed by the multiple correlation or regression coefficients. The classic confidence interval for a residual standard deviation is hypersensitive to minor violations of the normality assumption and its robustness does not improve with increasing sample size. An approximate confidence interval for the residual standard deviation is proposed and shown to be robust to moderate violations of the normality assumption with robustness to extreme non-normality that improves with increasing sample size.*

KEY WORDS: Dispersion, regression, model fit

Introduction

Consider the regression model $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\beta}$ is a vector of $k + 1$ unknown constants and the n elements of the unobservable random vector $\boldsymbol{\varepsilon}$ are assumed to be independent and identically distributed with a common but unknown, finite, and non-zero variance σ_ε^2. The predictor matrix $\mathbf{X}$ is assumed to be full rank and may be fixed or stochastic. The parameter σ_ε describes one aspect of the model that is not revealed by $\boldsymbol{\beta}$ or the coefficient of determination $1 - \sigma_\varepsilon^2/\sigma_{y.}^2$ Graybill & Iyer (1994) emphasize the importance and usefulness of σ_ε throughout their text.

If, in addition to the model assumptions stated above, the distribution of each element of $\boldsymbol{\varepsilon}$ is assumed to be normal, then an exact $100(1 - \alpha)\%$ confidence interval for σ_ε^2 is given by

$$SSE/\chi^2_{n-k-1;1-\alpha/2} < \sigma_\varepsilon^2 < SSE/\chi^2_{n-k-1;\alpha/2} \tag{1}$$

where *SSE* is the *OLS* residual sum of squares and $\chi^2_{n-k\text{-}1;\,p}$ is the pth quantile of the central chi-square distribution with $dfe = n - k - 1$ degrees of freedom (Graybill, 1976, p. 205). This confidence interval is exact for any full-rank $\mathbf{X}$.

The exact confidence interval for σ_y^2 (Tate & Klett, 1959; Kutner *et al.*, 2005, p. 1311) is not robust to subtle violations of the normality assumption. Increasing the sample size does not mitigate the problem and normality tests lack the power to detect the degree

Correspondence Address: Douglas G. Bonett, Department of Statistics, Iowa State University, Ames, Iowa 50011, USA. Email: dgbonett@iastate.edu

0266-4763 Print/1360-0532 Online/05/101089–6
DOI: 10.1080/02664760500165339

of non-normality that would cause the coverage probability to deviate substantially from $1 - \alpha$. Miller (1986, p. 264) characterizes this problem as 'catastrophic'. The problem is more severe for equation (1) because normality tests based on estimated residuals can be much less powerful than the normality tests for observed scores (White & MacDonald, 1980; Weisberg, 1980).

Residual normality should not be automatically assumed (White & MacDonald, 1980; Judge *et al.*, 1980, chapter 7; Greene, 2003, pp. 501–505) and hence the usefulness of equation (1) is severely limited because of its hypersensitivity to minor and often undetectable violations of the normality assumption. The confidence interval proposed here is shown to have good small-sample properties and, unlike equation (1), its robustness improves as the sample size is increased.

Proposed Confidence Interval

Let ε_i denote the ith element of $\boldsymbol{\varepsilon}$. The common variance of var(ε_i) may be expressed as $\sigma^4\{\gamma_4 - (n-3)/(n-1)\}/n$ where $\gamma_4 = \mu_4/\sigma^4$ and μ_4 is the finite population fourth central moment (Mood *et al.*, 1974, p. 229). Application of the delta method gives $\text{var}\{\ln(\hat{\sigma}^2)\} \approx \{\gamma_4 - (n-3)/(n-1)\}/n$. Shoemaker (2003) found that using $(n-3)/n$ rather than $(n-3)/(n-1)$ improved the small-sample performance of his equal-variance test, and this small-sample adjustment will be used here. In addition to the variance-stabilizing property of $\ln(\hat{\sigma}^2)$, Bartlett & Kendall (1946) show that the sampling distribution of $\ln(\hat{\sigma}^2)$ converges to normality much faster than the sampling distribution of $\hat{\sigma}^2$ when $Y_\mathrm{i} \sim N(\mu, \sigma^2)$. Scheffé (1959, p. 84) and Laylard (1973) recommend the logarithmic transformation for non-normal distributions as well.

In practice, γ_4 is unknown and an estimate of $\text{var}\{\ln(\hat{\sigma}^2)\}$ will require an estimate of γ_4. Pearson's estimator $\hat{\gamma}_4 = n\Sigma(Y_\mathrm{i} - \hat{\mu})^4/\{\Sigma(Y_\mathrm{i} - \hat{\mu})^2\}^2$ exhibits large negative bias in leptokurtic ($\gamma_4 > 3$) distributions for small n. The following estimator of γ_4, which is asymptotically equivalent to Pearson's estimator, will be used here to estimate $\text{var}\{\ln(\hat{\sigma}^2)\}$

$$\bar{\gamma}_4 = n\Sigma(Y_\mathrm{i} - m)^4/\{\Sigma(Y_\mathrm{i} - \hat{\mu})^2\}^2 \tag{2}$$

where m is a trimmed mean with trim-proportion equal to $1/\{2(dfe - 4)^{1/2}\}$ so that m converges to μ as n increases without bound. Bonett (2005) shows that this estimator exhibits less negative bias and a smaller coefficient of variation than Pearson's estimator in both skewed and symmetric leptokurtic distributions.

Let $\hat{\boldsymbol{\beta}}$ denote the *OLS* estimate of $\boldsymbol{\beta}$ and let $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ denote the vector of *OLS* residuals. A point estimate of σ_ε^2 is $\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/dfe$. Given the desirable sampling properties of $\ln(\hat{\sigma}^2)$, a large-sample confidence interval for σ_ε^2 may be obtained from a reverse-transformed confidence interval for $\ln(\sigma_\varepsilon^2)$. The following approximate $100(1-\alpha)\%$ confidence interval for σ_ε^2 is proposed

$$\exp\{\ln(c\hat{\sigma}_\varepsilon^2) \pm z_{1-\alpha/2}se\} \tag{3}$$

where $se = c[\{\bar{\gamma}_4 - (n-3)/n\}/(n-k)]^{1/2}$, $c = n/\{n - (n-2)z_{1-\alpha/2}/dfe)\}$ is a small-sample centring adjustment that helps equalize the tail probabilities, $\bar{\gamma}_4$ is computed from the *OLS* residuals, and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. Taking the square root of the endpoints of equation (3) gives a confidence interval for σ_ε. It is easy to show that equation (3) has asymptotic coverage probability of $1 - \alpha$. Its small-sample performance is examined in the following section.

Simulation Results

Estimates of coverage probabilities and average interval widths of equations (1) and (3) were obtained using 100,000 Monte Carlo random samples of a given sample size with various residual distributions. The simulation programs were written in Gauss and executed on a Pentium 4 computer.

The performance of equation (3) for normal residual distributions is examined for several values of k and dfe. Estimated coverage probabilities of equation (3) are given in Table 1. Estimated average confidence interval widths for equations (1) and (3) are given in Table 2. The results in Table 1 suggest that equation (3) has coverage probability close to $1-\alpha$ when the residual distribution is normal and $dfe \geq 10$. Furthermore, the average width of equation (3) is only slightly greater than the average width of equation (1). If equation (3) is used in a sample of size $n+2$ then its expected width will be about the same as the expected width of equation (1) from a sample size of n. The cost of

Table 1. Estimated coverage probabilities of equation (3) for normal residual distribution

		$1-\alpha$		
k	dfe	0.900	0.950	0.990
1	10	0.899	0.950	0.990
	25	0.892	0.943	0.986
	50	0.895	0.945	0.986
5	10	0.898	0.950	0.990
	25	0.893	0.943	0.986
	50	0.895	0.944	0.986
10	10	0.902	0.952	0.992
	25	0.894	0.945	0.988
	50	0.895	0.945	0.988
20	10	0.908	0.955	0.993
	25	0.896	0.945	0.988
	50	0.896	0.946	0.988

Table 2. Estimated interval widths for normal residual distribution ($\alpha = 0.05$)

k	dfe	Eqn (1)	Eqn (3)
1	10	2.59	2.91
	25	1.29	1.34
	50	0.85	0.86
10	10	2.59	2.93
	25	1.29	1.33
	50	0.85	0.86
20	10	2.59	2.90
	25	1.29	1.34
	50	0.85	0.86

Table 3. Estimated coverage probabilities for non-normal residual distributions ($\alpha = 0.05$)

Distribution*	k	dfe	Eqn (1)	Eqn (3)
Uniform	1	10	0.997	0.965
(0, 1.8)		50	0.999	0.951
		200	0.999	0.950
	10	10	0.986	0.966
		50	0.996	0.960
		200	0.998	0.955
	20	10	0.978	0.967
		50	0.993	0.964
		200	0.998	0.959
Beta(3,3)	1	10	0.982	0.957
(0, 2.5)		50	0.985	0.950
		200	0.986	0.950
	10	10	0.971	0.959
		50	0.981	0.953
		200	0.981	0.951
	20	10	0.964	0.962
		50	0.977	0.955
		200	0.982	0.953
Laplace	1	10	0.853	0.905
(0, 6)		50	0.812	0.918
		200	0.797	0.935
	10	10	0.880	0.918
		50	0.833	0.912
		200	0.798	0.933
	20	10	0.899	0.934
		50	0.843	0.910
		200	0.809	0.930
Gamma(6)	1	10	0.907	0.943
(0.82, 4)		50	0.895	0.936
		200	0.893	0.948
	10	10	0.923	0.949
		50	0.902	0.937
		200	0.893	0.949
	20	10	0.930	0.953
		50	0.909	0.934
		200	0.895	0.946
Gamma(3)	1	10	0.882	0.928
(1.15, 5)		50	0.852	0.927
		200	0.839	0.945
	10	10	0.904	0.938
		50	0.861	0.925
		200	0.844	0.943
	20	10	0.913	0.944

(*Table continued*)

Table 3. *Continued*

Distribution*	*k*	*dfe*	Eqn (1)	Eqn (3)
		50	0.872	0.924
		200	0.848	0.941
Exponential (2, 9)	1	10	0.816	0.871
		50	0.741	0.905
		200	0.711	0.940
	10	10	0.850	0.893
		50	0.764	0.898
		200	0.712	0.933
	20	10	0.866	0.910
		50	0.779	0.890
		200	0.720	0.931

*Skewness and kurtosis in parentheses.

sampling two additional units reflects the cost of using equation (3) instead of equation (1) in applications where the residual distribution is known to be closely approximated by the normal. This will be a small price to pay if equation (3) performs substantially better than equation (1) when the residual distribution is non-normal.

With non-normal residuals the performance of equation (3) will depend primarily on k, *dfe* and the shape of the residual distribution. Unlike in the case of normal residuals, the performance of equation (3) also will be affected, but to a much lesser degree, by the elements in $\mathbf{X}$. The results of a preliminary investigation confirm the empirical findings of Weisberg (1980) and suggest that changing the distribution of scores within any column of $\mathbf{X}$ or changing the correlation among columns of $\mathbf{X}$ has only a minor effect on the coverage probability of equation (3). The simulation experiment varies k, *dfe* and the residual distribution with columns of $\mathbf{X}$ sampled from independent normal distributions.

Estimated coverage probabilities of equations (1) and (3) for several non-normal residual distributions are given in Table 3. The skewed residual distributions were centered at their population means so that $E(\boldsymbol{\varepsilon}) = 0$. The results in Table 3 suggest that equation (3) is liberal in leptokurtic and highly skewed residual distributions. As expected, the coverage probability of equation (3) improves as *dfe* increases. In contrast to equation (3), equation (1) is very conservative in platykurtic distributions, very liberal in leptokurtic distributions, and its coverage probability does not improve as *dfe* increases. Clearly equation (3) is superior to equation (1) for all residual distributions considered in Table 3.

Concluding Remarks

The residual standard deviation provides important information that should be reported routinely. However, equation (1) is so fragile that its use is difficult to justify. The new confidence interval presented here performs almost as well as equation (1) with normal residuals and performs substantially better than equation (1) with non-normal residuals.

It is interesting to ask how bootstrap intervals compare with equation (3). A small-scale simulation study conducted by the author found that equation (3) performs better than the percentile method and the BC_a method in small samples. This is remarkable because the BC_a method is known to have second-order accuracy (Davison & Hinkley, 1997, p. 211) which suggests that it may be difficult to develop any method with better overall performance than equation (3).

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 0343552.

References

Bartlet, M. S. & Kendall, D. G. (1946) The statistical analysis of variance – heterogeneity and the logarithmic transformation, *Journal of the Royal Statistical Society, series B*, 8, pp. 128–138.

Bonett, D. G. (2005) Confidence interval for ratio of variances in paired-sample designs, *Journal of Statistical Computation and Simulation* (to appear).

Davison, A. C. & Hinkley, D. V. (1997) *Bootstrap Methods and their Applications* (Cambridge: Cambridge University Press).

Graybill, F. A. (1976) *Theory and Application of the Linear Model* (Belmont, CA: Duxbury Press).

Graybill, F. A. & Iyer, H. K. (1994) *Regression Analysis: Concepts and Applications* (Belmont, CA: Duxbury Press).

Greene, W. H. (2003) *Econometric Analysis*, 5th edn (Upper Saddle River, NJ: Prentice-Hall).

Judge, G. G. *et al.* (1980) *Theory and Practice of Econometrics* (New York: Wiley).

Kutner, M. H. *et al.* (2005) *Applied Linear Statistical Models*, 5th edn (Boston: McGraw-Hill).

Laylard, M. W. J. (1973) Robust large-sample tests for homogeneity of variances, *Journal of the American Statistical Association*, 68, pp. 195–198.

Miller, R. G. (1986) *Beyond ANOVA, Basics of Applied Statistics* (New York: Wiley).

Mood, A. M., Graybill, F. A. & Boes, D. C. (1974) *Introduction to the Theory of Statistics*, 2nd edn (New York: McGraw-Hill).

Scheffé, H. (1959) *The Analysis of Variance* (New York: Wiley).

Shoemaker, L. H. (2003) Fixing the F-test for equal variances, *The American Statistician*, 57, pp. 105–114.

Tate, R. F. & Klett, G. W. (1959) Optimal confidence intervals for the variance of a normal distribution, *Journal of the American Statistical Association*, 54, pp. 674–682.

Weisberg, S. (1980) Comment on 'Some large-sample tests of normality in linear regression models', *Journal of the American Statistical Association*, 75, pp. 28–31.

White, H. & MacDonald, G. M. (1980) Some large-sample tests of normality in linear regression models, *Journal of the American Statistical Association*, 75, pp. 16–28.