Confidence Intervals for Mean Absolute Deviations

Douglas G. BONETT and Edith SEIER

The mean absolute deviation is a simple and informative measure of variability. Approximate confidence intervals for mean absolute deviations in one-group and two-group designs are derived and are shown to have excellent small-sample properties under moderate nonnormality. Sample size planning formulas are derived.

KEY WORDS: Dispersion; Interval estimation; Sample size; Variability.

1. INTRODUCTION

Normal-theory tests and confidence intervals for variances are known to be hypersensitive to minor violations of the normality assumption and are "sensitive to amounts of kurtosis that may pass unnoticed in the handling of the data" (Snedecor and Cochran 1980, p. 81). Miller (1986, p. 264) described the effects of nonnormality on tests of variances as "catastrophic." Some statisticians (Moore 2000, p. 413) now advise researchers to "avoid inference about standard deviations" even though these methods are routinely used to answer important questions in engineering statistics, quality control, genetics, and psychometrics (see Ott and Longnecker 2001, chap. 7) and also provide important diagnostic information regarding homoscedasticity assumptions (see Miller 1986, secs. 2.3, 3.3, and 3.7).

Conover, Johnson, and Johnson (1981) compared 56 homogeneity of variance tests and found that the most satisfactory procedures were those that used mean deviations from the median such as the popular Levene test (Brown and Forsythe 1974). Carroll and Schneider (1985) showed that tests based on mean absolute deviation from the median have the correct asymptotic level in skewed distributions while tests based on the mean absolute deviation from the mean do not. Wilcox (1990) evaluated several robust confidence intervals for a ratio of variances and found none to be satisfactory except possibly the Box–Scheffé method which works reasonably well in equal-sized samples from distributions having identical shapes but gives very wide intervals. A nonparametric confidence interval for a ratio of variances based on ranks is available (Gibbons 1997, p. 244–246) but has little practical use because the two population medians are assumed to be known as well as other restrictive assumptions. Moses (1963) showed that no rank-based method can effectively test for variability around an unknown measure of location unless very strong restrictions are made about the location measure. Hall and Padmanabhan (1997) proposed a bootstrap confidence interval for a ratio of trimmed variances that performs well when samples are drawn from standardized distributions that have similar shapes. Shoemaker (2003) proposed a confidence interval for a ratio of variances that performs well when samples are drawn from distributions that have similar kurtosis values. We found that the Hall–Padmanabhan and Shoemaker methods are not robust to violations of their assumptions. Pan (1999) gave a confidence interval for a ratio of mean absolute deviations from the median that performs well unless the sample sizes are small and highly unbalanced.

The mean absolute deviation from the median (or simply mean absolute deviation) is one of the oldest measures of variability (David 1998). It is simple to compute and easy to understand. Approximate confidence intervals are proposed for a mean absolute deviation in one-group designs and a ratio of mean absolute deviations in two-group designs. Sample size formulas are presented to help the researcher design a study that will yield confidence intervals having the desired precision.

2. CONFIDENCE INTERVALS

Let $Y_{ij} (i = 1, 2, \ldots n_j; j = 1, 2)$ be continuous, independent and identically distributed random variables within group j with $0 < \text{var}(Y_{ij}) = \sigma_j^2 < \infty$ and $E(Y_{ij}) = \mu_j$. The population median of Y_{ij} is denoted as η_j. The j subscript is omitted in one-group designs. Consider a mean absolute deviation from the population median $D = \Sigma|Y_i - \eta|/n$ and let $\tau = E(D)$. A result from Stuart and Ord (1994, p. 361) can be used to show that $\text{var}\,D = \{(\mu-\eta)^2+\sigma^2-\tau^2\}/n$. Application of the delta method gives $\text{var}\ln D \doteq \{(\mu-\eta)^2+\sigma^2-\tau^2\}/\tau^2 n = (\delta^2+\gamma-1)/n$, where $\gamma = \sigma^2/\tau^2$ and $\delta = (\mu-\eta)/\tau$. The parameter γ is a measure of kurtosis similar to G-kurtosis defined by Bonett and Seier (2002) and δ is the skewness measure of Groeneveld and Meeden (1984).

A consistent estimator of the mean absolute deviation τ is

$$\hat{\tau} = \Sigma|Y_i - \hat{\eta}|/n, \tag{1}$$

where $\hat{\eta}$ is the sample median. This estimator is biased and we are unaware of any simple and general way to remove its bias. Our empirical findings suggest that $c\hat{\tau}$ is less biased than (1) where $c = n/(n-1)$.

Because $D = \hat{\tau}\{1 + O(n^{-1/2})\}$, $\text{var}\,\hat{\tau}$ will approximately equal $\text{var}\,D$ in large samples (Stuart and Ord 1994, p. 361). Hence, we propose the following estimator of $\text{var}\ln(\hat{\tau})$

$$\overline{\text{var}}\ln(\hat{\tau}) = (\hat{\delta}^2 + \hat{\gamma} - 1)/n \tag{2}$$

which is a function of consistent estimators $\hat{\delta} = (\hat{\mu} - \hat{\eta})/\hat{\tau}$ and $\hat{\gamma} = \hat{\sigma}^2/\hat{\tau}^2$, where $\hat{\sigma}^2 = \Sigma(Y_i - \hat{\mu})^2/(n-1)$ and $\hat{\mu} = \Sigma Y_i/n$.

The distribution of $\hat{\tau}$ has positive skew and a logarithmic transformation should reduce its skewness. Thus, a confidence interval for $\ln(\tau)$ might perform better in small samples than a confidence interval for τ. Furthermore, exponentiating the endpoints

Douglas G. Bonett is Professor, Departments of Statistics and Psychology at Iowa State University, Ames, IA 50011 (E-mail: dgbonett@iastate.edu). Edith Seier is Associate Professor, Department of Mathematics at East Tennessee State University, Johnson City, TN 37614 (E-mail: seier@etsu.edu). The authors thank the referees and both past and current editors for many helpful comments and suggestions.

of a confidence interval for $\ln(\tau)$ gives a confidence interval for τ that cannot include negative values.

In a one-group design the following $100(1-\alpha)\%$ confidence interval for τ is proposed

$$\exp[\ln(c\hat{\tau}) \pm z_{\alpha/2}\{\overline{\text{var}}\ \ln(\hat{\tau})\}^{1/2}], \quad (3)$$

and in a two-group design the following $100(1-\alpha)\%$ confidence interval for τ_1/τ_2 is proposed

$$\exp[\ln(c_1\hat{\tau}_1/c_2\hat{\tau}_2) \pm z_{\alpha/2}\{\overline{\text{var}}\ \ln(\hat{\tau}_1) + \overline{\text{var}}\ \ln(\hat{\tau}_2)\}^{1/2}], \quad (4)$$

where $c_j = n_j/(n_j - 1)$. The bias adjustments in (3) and (4) help equalize the tail probabilities. The distributions of Y_{i1} and Y_{i2} are not required to be identical, but if the two distributions have similar shapes such that $\tau_1/\tau_2 = \sigma_1/\sigma_2$, then (4) yields a confidence interval for σ_1/σ_2.

One-sided confidence intervals are obtained by replacing $z_{\alpha/2}$ with z_α in (3) and (4). Simultaneous Bonferroni confidence intervals are obtained by replacing $z_{\alpha/2}$ with $z_{\alpha/2v}$ in (3) and (4) where v is the number of confidence intervals of potential interest.

3. SAMPLE SIZE PLANNING FORMULAS

Let w denote the desired ratio of the upper to lower endpoints of (3) or (4). Solving for n and replacing $\hat{\delta}$ and $\hat{\gamma}$ with their respective planning values ($\tilde{\delta}$ and $\tilde{\gamma}$) gives

$$n = 4(\tilde{\delta}^2 + \tilde{\gamma} - 1)\{z_{\alpha/2}/\ln(w)\}^2 \quad (5)$$

the sample size requirement for (3). Let $n_1 = n_2 = n$. Solving for n and replacing $\hat{\delta}_j$ and $\hat{\gamma}_j$ with their respective planning values gives

$$n = 4(\tilde{\delta}_1^2 + \tilde{\delta}_2^2 + \tilde{\gamma}_1 + \tilde{\gamma}_2 - 2)\{z_{\alpha/2}/\ln(w)\}^2 \quad (6)$$

Table 1. Shape Parameters for Distributions Used in Simulations

Distribution	δ	γ	α_3	α_4
		Set A		
Beta(2,2)	0	1.23	0	1.5
Uniform	0	1.33	0	1.8
Tukey(3)	0	1.42	0	2.1
		Set B		
Normal	0	1.57	0	3
Logistic	0	1.71	0	4.2
Laplace	0	2.00	0	6
		Set C		
ScConN(.05,9)	0	1.82	0	7.6
Student's t(5)	0	2.00	0	9
ScConN(.1,25)	0	2.73	0	16.5
		Set D		
Gamma(1,7)	0.158	1.63	0.76	3.9
Half-Normal	0.261	1.62	0.99	3.9
Gumbel(0,1)	0.217	1.76	1.14	5.4
		Set E		
Chi-square(4)	0.305	1.81	1.41	6
Chi-square(3)	0.356	1.89	1.63	7
Exponential	0.443	2.08	2.00	9

the sample size requirement *per group* for (4).

For one-sided intervals, w may be defined as the desired endpoint-to-point estimate ratio in which case the required sample size is given by 1/4 the value of (5) or (6). The values obtained by (5) and (6) should be rounded up to the nearest integer.

Planning values of δ and γ are obtained from previous research or expert opinion. Values of δ and γ for several distributions are given in Table 1 and this may help researchers specify their planning values of δ and γ. Accurate planning values are often difficult to obtain although it may be possible to specify a range of planning values with high confidence. Using the upper planning limits for δ and γ in the sample size formulas will yield a conservatively large sample size requirement.

4. EXAMPLES

Example 1. Ott and Longnecker (2001, p. 346) described a quality control study where the fill weight variability of 500-gram coffee containers is assessed every hour. They provide the fill weights for one sample of $n = 30$ containers. Suppose that management has set the upper specification limit for τ at 2.0 grams. The basic statistics for the sample data are $\hat{\mu} = 500.45$, $\hat{\eta} = 499.75$, $\hat{\sigma} = 3.43$, and $\hat{\tau} = 2.79$. Inserting these statistics into (3) and setting $z_\alpha = 1.645$ give a 95% one-sided lower limit of 2.3 grams which may trigger an investigation to determine the cause of the recent increase in production variability.

Before computing a normal-theory confidence interval for σ, Ott and Longnecker examined a normal probability plot and concluded that the normality assumption "appears to be satisfied." This conclusion may not be justified because the normal-theory confidence interval for σ is not robust to degrees of nonnormality that can be readily detected by even the most powerful tests of normality. The results of Scheffé (1959, p. 336) can be applied to show that the normal-theory 95% confidence interval for σ has an asymptotic coverage probability of about .76, .63, .60, and .51 for the logistic, t_7, Laplace, and t_5 distributions, respectively. This result is disturbing because these distributions are not easily distinguished from normal distributions unless the sample size is large. In fact, to reject the null hypothesis of normality with probability .8, the Shapiro–Wilk test requires a sample size of about 400 when the distribution is logistic and a sample size of about 200 when the distribution is t_5. The current practice of inspecting a normal probability plot prior to computing a test or confidence interval for σ or σ_1/σ_2 in small samples is mostly ceremonial.

Example 2. Shoemaker (2003) analyzed data from a study by Jaffe, Parker, and Wilson (1982) in which measurements of aldrin (in nanograms per liter) were taken downstream of an abandoned dumpsite. The basic statistics for ten surface measurement and ten river-bottom measurements are summarized below.

Depth	$\hat{\mu}$	$\hat{\eta}$	$\hat{\sigma}$	$\hat{\tau}$
Bottom	6.02	5.53	1.58	1.21
Surface	4.18	4.33	0.67	0.50

Inserting the above statistics into (4) gives a 95% confidence interval of $1.06 \leq \tau_1/\tau_2 \leq 5.51$ which suggests that aldrin concentration variability is greater at the river bottom than on the surface. However, the confidence interval is very wide because

Table 2. Coverage Probabilities ($\times$ 100) for (3)

Distributions	n	$1-\alpha$ .90	.95	.99
Set A	10	91-95	95-97	99
	25	91-95	95-97	99
	50	91-93	95-97	99
	200	90-91	95-96	99
Set B	10	88-90	93-95	98-99
	25	89-90	94-95	98-99
	50	90-90	94-95	99
	200	90	95	99
Set C	10	82-89	89-95	96-99
	25	82-90	89-95	96-99
	50	85-90	90-95	96-99
	200	89-90	95-95	98-99
Set D	10	89-90	94-95	98-99
	25	89-90	94-95	98-99
	50	89-90	94-95	98-99
	200	90	95	99
Set E	10	83-87	89-92	95-97
	25	86-89	92-93	97-98
	50	88-89	93-94	98
	200	89	95	99

of the small sample size. Suppose that an upper to lower endpoint ratio of 2.5 is desired. Using the sample estimates of δ_j and γ_j as planning values in (6), we find that about 69 measurements should be taken at each depth in order to obtain a 95% confidence interval with an upper to lower endpoint ratio of 2.5. If the estimates of δ_j and γ_j in the new study are close to the planning values, then the upper to lower endpoint ratio of (4) will be close to 2.5.

5. SIMULATION RESULTS

Assuming that $n\{\widehat{\text{var}}\ln(\hat{\tau})\}$ is a consistent estimator of $n\{\text{var}\ln(\hat{\tau})\}$ and appealing to the central limit theorem with dependent variables (Lehmann 1999, p. 107) for the asymptotic normality of $\ln(\hat{\tau})$, the coverage probabilities of (3) and (4) will approach $1-\alpha$ as the sample size increases, although it is not known how large the sample size must be to obtain a coverage probability close to $1-\alpha$. Estimates of $1-\alpha$ were obtained using 50,000 Monte Carlo random samples of a given sample size from a wide variety of distributions. The distributions used in the simulation are shown in Table 1 along with their shape parameters δ, γ, α_3, and α_4, where α_3 and α_4 are the standardized third and fourth central moments, respectively. The scale-contaminated distributions, denoted as $\text{ScCon}N(p,\sigma^2)$, contaminate a $N(0,1)$ with a $N(0,\sigma^2)$ with probability p. The simulation programs were written in Gauss and executed on a Pentium IV computer.

Coverage probabilities of (3) are summarized in Table 2 with each row giving a range of coverage probabilities for the three distributions within each set. These results suggests that (3) has a coverage probability close to $1-\alpha$ in mild leptokurtic and mild skewed distributions. In small samples, (3) is conservative in platykurtic distributions and liberal in highly leptokurtic and skewed distributions.

A second simulation study examined the small-sample performance of (4) in balanced and unbalanced designs. The results are summarized in Table 3 with each row giving a range of coverage probabilities for nine combinations of distributions. In small samples (4) is conservative in platykurtic distributions and liberal in highly leptokurtic or skewed distributions. The coverage probability decreases slightly if the smaller sample is taken from the distribution having the larger skewness or kurtosis. The $\text{ScCon}N(.05,9)$ distribution is responsible for the liberal coverage probabilities in Tables 2 and 3. Although not shown in Tables 2 and 3, we found that both (3) and (4) have approximately equal tail probabilities except in those few cases where the coverage probability was much smaller than $1-\alpha$.

We also examined a percentile bootstrap confidence interval (Chernick 1999, p. 53) for τ hoping to achieve better coverage probabilities in highly nonnormal distributions. The percentile method performed worse (too liberal) than (3) and (4). For instance, the coverage probability of a 95% confidence interval for $n_1 = n_2 = 25$ was only .90 when sampling from a normal distributions and .86 when sampling from an exponential distribution. The BCa method (Chernick 1999, pp. 55, 59), which has second-order accuracy, performed no better than (3) or (4).

The bootstrap method for a ratio of variances proposed by Hall and Padmanabhan (1997) assumes that the standardized distributions have identical shapes so that the ratio of population trimmed variances will equal the ratio of population variances. If the equal-shape assumption cannot be satisfied, the Hall-Padmanabhan method may perform very poorly. For instance, if $n_1 = 25$ observations are sampled from a normal distribution and $n_2 = 10$ observations are sampled from an exponential distribution, the Hall–Padmanabhan 95% confidence interval for σ_1/σ_2 (Adaptive Method 1) has a coverage probability of about 84.4%. In comparison, Shoemaker's F_1 method which assumes equal kurtosis has a coverage probability of 83.5% while (4) has

Table 3. Coverage Probabilities ($\times$ 100) for (4)

Distributions				$1-\alpha$		
Sample 1	Sample 2	n_1	n_2	.90	.95	.99
Set A	Set B	10	10	90-94	95-97	99
		25	10	92-95	96-97	99
		10	25	88-92	93-96	99
		25	25	91-93	95-97	99
		100	100	90-91	95-96	99
Set B	Set C	10	10	87-91	92-95	98-99
		25	10	89-91	94-95	98-99
		10	25	85-90	91-95	97-99
		25	25	86-90	92-95	98-99
		100	100	88-90	93-95	98-99
Set B	Set D	10	10	89-92	94-96	98-99
		25	10	88-91	94-96	98-99
		10	25	89-91	94-95	98-99
		25	25	90-91	95	99
		100	100	90	95	99
Set D	Set E	10	10	87-90	93-94	98
		25	10	89-90	94-95	98-99
		10	25	86-89	91-93	97-98
		25	25	89-90	94-95	98-99
		100	100	90	95	99

a coverage probability of 91.1% in this setting. Increasing the sample size will bring the a coverage probability of (4) close to $1 - \alpha$ but the same cannot be said for the Hall–Padmanabhan or Shoemaker methods when their assumptions have been violated.

The method proposed by Pan (1999) performs as well as (4) in balanced designs but may have a larger average width and worse tail symmetry in highly unbalanced designs. For instance, with $\alpha = .05$ and samples of $n_1 = 14$ and $n_2 = 6$ from a normal distribution, (4) has lower and upper tail probabilities of about 2.5% and 1.1%, respectively, a coverage probability of about 96.4%, and an average width of about 2.35. Pan's confidence interval has lower and upper tail probabilities of about 4.6% and 1.6%, respectively, a coverage probability of about 93.8%, and an average width of about 2.56.

If accurate planning values of δ or γ are available, the planning values can be used in (2) to improve small-sample performance of (3) and (4) in highly nonnormal distributions. We found that (3) and (4) have coverage probabilities very close to $1 - \alpha$ for many highly nonnormal distributions with sample sizes of about 30 per group if population values of δ and γ are used in (2). The Levene test and the Pan method cannot take advantage of prior shape information in this way.

6. CONCLUDING REMARKS

The normal-theory confidence intervals for σ and σ_1/σ_2 do not have asymptotic coverage probabilities of $1 - \alpha$ in platykurtic or leptokurtic distributions, and in small samples, tests of normality lack the power to detect the degree of nonnormality that would cause problems with these methods. In contrast, tests of normality are more likely to detect the type of nonnormality that would cause problems with (3) and (4).

In small samples, (3) and (4) perform well under moderate departures from normality and are about as robust to nonnormality as the one-sample and two-sample t tests, respectively. As with t tests, a larger sample size provides greater protection against nonnormality. If the distribution is highly skewed, a skewness-reducing transformation will decrease the sample size at which the coverage probability becomes close to $1 - \alpha$. For instance, when sampling from a χ_1^2 distribution, a sample size of about 200 is needed before (3) will have a coverage probability close to $1 - \alpha$. If data from a χ_1^2 distribution are square-root transformed, (3) will have a coverage probability close to $1 - \alpha$ with a sample size of about 30. If extreme nonnormality cannot be ruled out and a data transformation is not possible or desirable, a bootstrap confidence interval for a robust measure of variability, such as the interquartile range or the *median* absolute deviation from median, may be the best option.

In some applications a mean absolute deviation from a target value (h), rather than the median, may be more interesting. In Example 1, for instance, we might want to examine the mean absolute deviation from 500 grams. In applications such as these, simply replace $\hat{\eta}$ with h in (1) and (2). If extreme nonnormality cannot be ruled out, a distribution-free confidence interval for the median absolute deviation from target may be obtained by applying the usual confidence interval for a median (Snedecor and Cochran, 1980 p. 137) to the transformed scores $|Y_i - h|$. A distribution-free confidence interval for a ratio of median absolute deviations from target may be obtained by applying the method of Price and Bonett (2002) to the transformed scores $|Y_{ij} - h|$.

The confidence intervals for τ and τ_1/τ_2 introduced here are simple to compute and are attractive alternatives to the classic confidence intervals for variances. They could be included in introductory statistics courses that typically introduce the mean deviation as an important descriptive statistic but do not follow through with inferential methods for this parameter. The sample size formulas presented here also have pedagogical value in that they clearly show the effects of the confidence level, desired precision, and distribution shape on the sample size requirement.

[Received March 2002. Revised June 2003.]

REFERENCES

Bonett, D. G., and Seier, E. (2002), "A Test of Normality with High Uniform Power," *Computational Statistics and Data Analysis*, 40, 435–445.

Brown, M. B., and Forsythe, A. B. (1974), "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69, 364–367.

Carroll, R. J., and Schneider, H. (1985), "A Note on Levene's Tests for Equality of Variances," *Statistics and Probability Letters*, 3, 191–194.

Chernick, M. R. (1999), *Bootstrap Methods: A Practitioner's Guide*, New York: Wiley.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981), "A Comparative Study of Tests for Homogeneity of Variance, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics*, 23, 351–361.

David, H. A. (1998), "Early Sample Measures of Variability," *Statistical Science*, 13, 368–377.

Gibbons, J. (1997), *Nonparametric Methods for Quantitative Analysis* (3rd ed.), Columbus, OH: American Science Press, Inc.

Groeneveld, R. A., and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *The Statistician*, 33, 391–399.

Hall, P., and Padmanabhan, A. R. (1997), "Adaptive Inference for the Two-Sample Scale Problem," *Technometrics*, 39, 412–422.

Jaffe, P. R., Parker, F. L., and Wilson, D. J. (1982), "Distribution of Toxic Substances in Rivers," *Journal of the Environmental Engineering Division*, 108, 639–649.

Lehmann, E. L. (1999), *Elements of Large-sample Theory*, New York: Springer.

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

Moore, D. S. (2000), *The Basic Practice of Statistics* (2nd ed.), New York: Freeman.

Moses, L. E. (1963), "Rank Tests of Dispersion," *Annals of Mathematical Statistics*, 34, 973–983.

Ott, R. L., and Longnecker, M. (2001), *An Introduction to Statistical Methods and Data Analysis*, Pacific Grove, CA: Duxbury.

Pan, G. (1999), "On a Levene Type Test for Equality of Two Variances," *Journal of Statistical Computation and Simulation*, 63, 59–71.

Price, R. M., and Bonett, D. G. (2002), "Distribution-Free Confidence Intervals for Ratio and Difference of Medians," *Journal of Computational Statistics and Simulation*, 72, 119–124.

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, 105–114.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods* (7th ed.), Ames, IA.: Iowa State University Press.

Stuart, A., and Ord, K. J. (1994), *Kendall's Advanced Theory of Statistics, Vol 1. Distribution Theory* (6th ed.), London: Arnold.

Wilcox, R. (1990), "Comparing Variances and Means when Distributions have Non-identical Shapes," *Communications in Statistics—Simulation and Computation*, 19, 155–173.