A Note on the Reliability Tests of Estimates from ARMS Data

C. S. Kim, C. Hallahan, W. Lindamood, G. Schaible, and J. Payne

USDA uses the concept of "publish-ability" rather than statistical reliability of an estimate for quality validation of USDA estimates, which is solely based on the sample size and the coefficient of variation (CV). We demonstrate conceptually how the reliability of the sample mean can be tested by estimating the upper and lower bounds of the confidence interval for an unknown population mean using the CV. However, the reliability test for the sample mean can be made only under the normality assumption. USDA multiple-way Agricultural Resource Management Survey (ARMS) estimates are used to illustrate the relative measure of precision for sample-based estimators.

Key Words: ARMS data, coefficient of variation, publish-ability, reliability

The National Agricultural Statistics Service (NASS) and Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA) are striving to improve the availability and the quality of data on crop production practices as well as farm financial management. The Agricultural Resource Management Survey (ARMS)–Phase II is USDA's primary source of information on farm crop production practices for major crops, including corn, soybeans, wheat, grain sorghum, barley, oats, and cotton. This survey provides annual field-level data by crop on irrigation technology and water use, nutrient use and nutrient management practices, crop residue management practices, pesticide use and pest management practices, and crop seed varieties including genetically modified seeds. These data summaries, currently available for soybeans, wheat, and cotton for the period 1996–2000, and corn for the period 1996–2001, are invaluable to decision makers and analysts within government agencies and the public.[1]

Quality validation of USDA ARMS estimates is based solely on the sample size and the coefficient of variation (CV), which is also called the relative standard error. Some details can be found in Dubman (2000), Kott (1997, 2001), and in Sommer et al. (1998). According to USDA's general guidelines for statistical reporting standards, no estimate should be suppressed simply because it is deemed statistically unreliable. Nevertheless, the presence of such an estimate in a published table should be noted. In particular, an estimator (mean or proportion) in a data summary table of an agency publication should be marked with an asterisk denoting it as potentially unreliable (in a statistical sense) if either the sample size is less than a fixed number of individuals or if the estimate's CV is greater than some designated limit (USDA, 1993). The designated CV can be set at the agency's discretion for an estimator based on commonly occurring events. For the ARMS–Phase II data, each estimator is identified as having a CV less than or equal to 25%, greater than 25% but less than or equal to 50%, greater than 50% but less than or equal to 100%, or greater than 100%.

C. S. Kim is a senior economist, C. Hallahan is a mathematician, W. Lindamood is a programmer, and G. Schaible and J. Payne are agricultural economists, all with the U.S. Department of Agriculture, Economic Research Service, Washington, DC.

The authors thank Phil Kott, Chief Statistician at USDA/NASS, and Robert Dubman, Data and Survey Coordinator at USDA/ERS, for their comments and useful discussions, and two anonymous referees for their helpful comments. The views expressed here are the sole responsibility of the authors and do not necessarily reflect those of the U.S. Department of Agriculture.

[1] Data summaries are available at the USDA-ERS online website, http://www.ers.usda.gov/data/cropproductionpractices/.

The CV is an ideal measure for comparing variation across numerous sets of data expressed in different units, such as corn price per bushel and corn yield per acre. However, the CV is not a very meaningful measure without some assurance that the population mean, μ, lies within a preassigned precision level. For USDA ARMS measures, use of these estimates for policy analyses requires a broader statistically determined measurement of precision.

Therefore, the objectives of this paper are fourfold: first, to inform ARMS data users of the reasons why USDA provides the CV for each estimate; second, to explain conceptually how the CV can be used for testing the reliability of an estimator; third, to address the assumption of normality applied to our reliability tests; and finally, to demonstrate our reliability tests applied to a subset of 2001 ARMS–Phase II estimates. While ARMS–Phase II summary data tables can contain estimates for both means and proportions, we concentrate on mean estimates in this paper.

Reliability versus Publish-ability of an ARMS Estimator

The coefficient of variation associated with each USDA estimate is published for two primary reasons. First, the statistical interpretation of the square root of the expected value of the square of the relative absolute error of an estimator is *roughly* equal to its CV, stated as follows (Kott, 2003):

$$(1)\quad \left[E\left(|(m-\mu)/\mu|\right)^2\right]^{1/2} = \left[(\sigma^2/n)/\mu^2\right]^{1/2} = \left[\left(\sigma/\sqrt{n}\right)/\mu\right] \approx \left[\left(s/\sqrt{n}\right)/m\right] = \text{CV},$$

where E represents the expectation operator, μ and σ^2 represent the population mean and variance, respectively, and *m* and *s* represent the sample mean and sample standard deviation, respectively, of a random sample of size *n* from a normally distributed population such that $E(m) = \mu$, $E(s^2) = \sigma^2$, and $m \sim N(\mu, \sigma^2/n)$. Second, an estimate may be published based on USDA's general guidelines for statistical reporting standards, which states that no estimate should be suppressed even though it is deemed statistically unreliable, as long as the estimate is marked if its CV is greater than some designated limit. Therefore, consistent with USDA statistical guidelines, the estimate *m* may be a publishable rather than a reliable estimator (Kott, 2003).

The CV, $[(s/\sqrt{n})/m]$, is a function of the ratio of two independent random variables, $(s/\sqrt{n})$ and *m*. In this case, however, the standard method for deriving the integrating density function of the ratio of two independent random variables (e.g., Dwass, 1970; Parzen, 1960) fails because the integral corresponding to the mathematical expectation of $\text{CV} = (s/\sqrt{n})/m$ is not solvable. Therefore, while we know that $E[(s/\sqrt{n})/m] \neq [(\sigma/\sqrt{n})/\mu]$, we do not know the magnitude of the bias of the CV. Consequently, the current state-of-affairs with respect to statistical validation procedures based only on a CV statistic is somewhat unsatisfactory as an appropriate statistical assurance that USDA's ARMS estimates are reliable.

The concept of reliability of an estimate we use is based on the precision of an estimator as a measure of reliability [see Barlow and Proschan (1965) for a different view of reliability over time]. To construct a $(1-\alpha)$ confidence interval for μ when σ is unknown, the estimator, $[(m-\mu)\sqrt{n}/s]$, assuming a normal distribution, has a *t*-distribution with $(n-1)$ degrees of freedom. Therefore, the confidence interval for μ for random samples from a normal population with a degree of confidence of $(1-\alpha)$ is represented as follows:

$$(2)\quad \Pr\left[\left(m - t_{1-\alpha/2,n-1}(s/\sqrt{n})\right) \le \mu \le \left(m + t_{1-\alpha/2,n-1}(s/\sqrt{n})\right)\right]$$
$$= \Pr\left[-t_{1-\alpha/2,n-1}(s/\sqrt{n}) \le \mu - m \le t_{1-\alpha/2,n-1}(s/\sqrt{n})\right]$$
$$= \Pr\left[-t_{1-\alpha/2,n-1}\text{CV} \le (\mu - m)/m \le t_{1-\alpha/2,n-1}\text{CV}\right]$$
$$= \Pr\left[|(\mu - m)/m| \le k\right]$$
$$= (1-\alpha), \quad \text{where } k = t_{1-\alpha/2,n-1}\text{CV}.$$

In equation (2), the *reliability* of *m* as an estimate of μ requires that the relative deviation of *m* from μ be less than or equal to a given level of precision *k* at a $(1-\alpha)$ confidence level. For example, if the relative error of *m* is less than or equal to the precision level *k* at a 95% confidence level, then:

$$\Pr\left[|(\mu - m)/m| \le k\right] = 95\%.$$

Equation (2) can now be rewritten as:

$$(3)\quad \Pr\Big[\big(m - t_{1-\alpha/2,n-1}(s/\sqrt{n})\big) \le \mu \le \big(m + t_{1-\alpha/2,n-1}(s/\sqrt{n})\big)\Big]$$
$$= \Pr\Big[m\big(1 - t_{1-\alpha/2,n-1}(s/\sqrt{n})/m\big) \le \mu \le m\big(1 + t_{1-\alpha/2,n-1}(s/\sqrt{n})/m\big)\Big]$$
$$= \Pr\big[m(1-k) \le \mu \le m(1+k)\big]$$
$$= (1-\alpha), \quad \text{where } k = t_{1-\alpha/2,n-1}\text{CV}.$$

Equation (3) shows that the lower and upper bounds of the confidence interval for an unknown μ are represented by $m(1-k)$ and $m(1+k)$, respectively.

Using equations (2) and (3), and USDA ARMS data as an example, we can illustrate a relative measure of precision of an estimator under two reliability perspectives. Given that a predetermined number of sample replications for NASS's ARMS data is $n^* = 15$ for a delete-a-group jackknife variance, $t_{0.975,\,14} = 2.145$, then:

$$(4)\quad k = 2.145\,\text{CV}.$$

As illustrated by equation (4), if the relative error of an estimate is preassigned at a precision level $k = k_0$, the upper bound of the CV must be equal to $100(k_0/2.145)$ percent at the 95% confidence level. Similarly, if the CV is estimated to be $\text{CV} = \text{CV}_0$ at the 95% confidence level, the precision level required is estimated by $k = 2.145\,\text{CV}_0$.

The reliability condition for an estimator, expressed in equations (2) and (3), was derived under the assumption that samples are drawn from a normal distribution. Therefore, it is also appropriate to address the question about how the reliability condition holds up under a nonnormal assumption. We can evaluate whether a reliability condition, similar to that expressed in equations (2) and (3), can be established when there is no knowledge of the distribution, but assuming a random sample. For example, for any preassigned error ε, Chebyshev's inequality may be used:

$$(5)\quad P\big\{|(m-\mu)| \le \varepsilon\big\} \ge 1 - \big[(\sigma^2/n)/\varepsilon^2\big].$$

The absolute deviation in equation (5) can be rewritten in the form of the relative deviation as presented in equation (2):

$$(6)\quad P\big\{|(m-\mu)/m| \le z\big\} \ge 1 - \big[(\sigma^2/n)/(m^2 z^2)\big],$$
$$\text{where } z = \varepsilon/m.$$

From equation (6), one obtains the following:

$$(7)\quad P\big\{|(m-\mu)/m| \le z\big\} \ge 95\%$$
$$\text{if } z \ge 4.4721\big(\sigma/\sqrt{n}\big)/m.$$

Since the population variance is unknown, the reliability for an estimator cannot be tested with equation (7). Therefore, this result implies that the reliability test for the sample mean can be made only under the normality assumption.

The Delete-a-Group versus Delete-a-Sample Jackknife Variance

USDA's NASS adopted a multi-phased stratified sampling procedure for its ARMS data, and therefore the variance of an ARMS estimator for a CV value is estimated with a delete-a-group jackknife method. However, a simple delete-a-sample jackknife method can be illustrated as follows:

$$(8)\quad m_{(i)} = \big(1/(n-1)\big)\sum_{j\neq i}^{n} x_j = (nm - x_i)/(n-1),$$

where $m_{(i)}$ is the sample mean of the data set deleting the ith sample element, x_i, and n remains the sample size. The full sample mean estimate is represented by:

$$(9)\quad m_{(\cdot)} = \sum_{i=1}^{n} m_{(i)}/n.$$

The delete-a-sample jackknife estimate of variance is then denoted by:

$$(10)\quad \text{var}(m_{jack}) = \big[(n-1)/n\big]\sum_{i=1}^{n}\big(m_{(i)} - m_{(\cdot)}\big)^2.$$

It should be noted that n in equations (8)–(10) normally represents the number of observations, based on the literature explaining a delete-a-sample jackknife method (for instance, see Efron, 1982). However, for USDA ARMS data, n represents the number of groups corresponding to a delete-a-group jackknife method. NASS predetermined the number of group replications at $n = 15$.

For the case of a delete-a-sample jackknife method, Miller (1964) demonstrated that the confidence interval for an estimator approaches the confidence interval for an estimator from the normal distribution as the sample size increases, i.e., as $n \to \infty$. However, for a delete-a-group jackknife method, an increase in sample size does not change the number of replications. So, at this time, it remains unclear how much the sample size affects the confidence interval for a jackknife estimator.

Table 1. Reliability Tests for Nutrient Use by Tillage Practice and Irrigation System—Corn (all survey states, 2001)

[1] Description	[2] % Acres Treated [a]	[3] Sample Size	[4] 100 (CV (%)	[5] 100 (k [b] (%)	[6] Lower Bound [c]	[7] Upper Bound [d]
Gravity Irrigation System:						
Nitrogen	95.6	120	2.5	5.3	90.5	100.0
Phosphate	69.7	120	13.7	29.4	49.2	90.2
Potash	11.2*	120	33.9	72.8	3.0	19.4
Pressure Irrigation System:						
Nitrogen	98.4	273	0.5	1.2	97.3	99.5
Phosphate	84.7	273	2.3	4.8	80.6	88.8
Potash	42.9	273	6.2	13.2	37.2	48.6
Description	Pounds per Treated Acre [a]	Sample Size	100 (CV (%)	100 (k [b] (%)	Lower Bound [c]	Upper Bound [d]
Gravity Irrigation System:						
All Acres:						
Nitrogen	141.8	103	7.9	16.9	117.8	165.8
Phosphate	49.7	103	10.4	22.3	38.6	60.8
Potash	53.4*	103	30.6	65.6	18.3	88.5
Non-conservation Tillage:						
Nitrogen	145.5	67	6.2	13.3	126.1	164.9
Phosphate	56.9	67	18.6	39.8	34.2	79.6
Potash	48.5*	67	26.2	56.2	21.2	75.8
Conservation Tillage:						
Nitrogen	139.1	31	13.0	27.8	100.4	177.8
Phosphate	46.8	31	15.7	33.6	31.1	62.5
Potash	59.9**	31	72.1	154.6	0.0	152.5
Pressure Irrigation System:						
All Acres:						
Nitrogen	158.2	261	5.2	11.2	140.5	175.9
Phosphate	41.6	261	5.8	12.5	36.4	46.8
Potash	53.1	261	16.6	35.6	34.2	72.0
Non-conservation Tillage:						
Nitrogen	158.1	162	6.6	14.1	135.8	180.4
Phosphate	44.8	162	7.5	16.2	37.6	52.0
Potash	67.9	162	19.4	41.7	39.6	96.2
Conservation Tillage:						
Nitrogen	158.5	98	7.9	16.9	131.7	185.3
Phosphate	38.5	98	8.9	19.0	31.2	45.8
Potash	38.2	98	21.2	45.4	20.9	55.5

[a] Asterisks (*, **, ***) indicate a CV such that 25% < CV # 50%, 50% < CV # 100%, and CV > 100%, respectively.
[b] $k = 2.145\,CV$.
[c] Lower bound $(L) = m(1 \; ! \; k)$.
[d] Upper bound $(U) = m(1 + k)$.

An Example from USDA-ARMS Estimates

To explain the reliability of estimates based on USDA ARMS data, we base our analysis on a summarized data table from the *multiple*-way ARMS tables posted on the USDA-ERS website under the title "Nutrient use by tillage system and irrigation system" associated with corn for all survey states (refer to website given in footnote 1).

The USDA ARMS information is represented in the first two columns of table 1, where the second column represents the sample means (percent of acres treated and pounds per treated acre) for the year 2001, and each estimator is identified with its CV. Estimates are marked based on a CV of less than or equal to 25%, greater than 25% but less than or equal to 50%, and greater than 50% but less than or equal to 100% (see table 1 footnote).

Column [4] identifies the CV (in percent). Column [5] identifies the precision level (in percent), 100 (*k*, which is estimated using equation (4). For example, the rate of potash application for corn production in 2001 by non-conservation tillage practice using a gravity irrigation system is estimated to be m = 48.5 pounds per treated acre, and the associated precision level and CV are estimated to be 100 (k = 56.2% and 100 (CV = 26.2%, respectively (table 1). The estimated precision level indicates that at the 95% confidence level, the deviation of the estimate m = 48.5 pounds per acre from its unknown μ is less than or equal to 56.2% of the estimate, or 27.3 pounds per acre (i.e., 0.562 (48.5 = 27.3 pounds per acre). The lower and upper bounds of the 95% confidence interval for the estimator's unknown μ are identified in columns [6] and [7], respectively. These confidence bounds are estimated by 48.5 ± 27.3, using the precision level k shown in column [5].

Results of the reliability tests by the precision level reveal that for ARMS 2001 data, for all corn acres, estimates of nutrient application rates are relatively more precise for acres irrigated with a pressure irrigation system than a gravity system. In addition, similar estimates for pressure-irrigated acres using conservation tillage are generally more precise than are estimates of application rates for gravity-irrigated acres using conservation tillage. Finally, while estimates for gravity-irrigated acres using non-conservation tillage are slightly more precise than the estimates for conservation tillage on gravity-irrigated acres, the precision of these estimates is relatively similar for pressure-irrigated acres.

Conclusions

We have reviewed the quality validation procedure associated with USDA ARMS estimates based on the assumption that samples are drawn from a normally distributed population. While the concept of publish-ability of an estimate is used by USDA when providing the CV statistic for each estimate, we demonstrate conceptually how the CV can be used for testing the reliability of an estimator. Specifically, if the relative bias of an estimate is preassigned at a precision level k_0, the upper bound of the CV can be derived from $k_0 = 2.145\text{CV}$ at the 95% confidence level. Similarly, if the CV is estimated to be CV_0, then at a 95% confidence level, a required upper bound on the corresponding precision level is estimated by $k = 2.145\text{CV}_0$. Furthermore, we have also shown that the upper and lower bounds of the confidence interval for an unknown μ can be estimated using the coefficient of variation with $m(1+k)$ and $m(1 ! k)$, respectively. Finally, our conceptual analysis has demonstrated that the reliability of an estimate cannot be tested when there is no knowledge of the distribution, because the population variance is unknown—i.e., the reliability test for the sample mean can be made only under the normality assumption.

For USDA's 2001 ARMS estimates of nutrient application rates by tillage practice and irrigation system, the new reliability tests provide significantly improved information from which ARMS data users may judge estimator precision. This relative test of estimator reliability can help transform ARMS estimates from being just publishable to accommodating a level of confidence or reliability when estimates are used within policy analyses.

References

Barlow, R. E., and F. Proschan. (1965). *Mathematical Theory of Reliability.* New York: John Wiley and Sons, Inc.

Dubman, R. W. (2000). "Variance Estimation with USDA's Farm Costs and Return Surveys and Agricultural Resource Management Study Surveys." Staff Report No. AGES 00-01, USDA/Economic Research Service, Washington, DC.

Dwass, M. (1970). *Probability and Statistics.* New York: W. A. Benjamin, Inc.

Efron, B. (1982). "The Jackknife, the Bootstrap, and Other Resampling Plans." Society for Industrial and Applied Mathematics, Philadelphia, PA.

Kott, P. S. (1997). "Statistical Analysis with a Delete-a-Group Jackknife." Unpublished paper, USDA/National Agricultural Statistics Service, Washington, DC.

———. (2001, July). "Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys." Revised Research Report No. RD-98-01, USDA/National Agricultural Statistics Service, Washington, DC.

———. (2003). Chief Statistician, National Agricultural Statistics Service, USDA, Washington, DC. Personal communication.

Miller, R. G. (1964). "A Trustworthy Jackknife." *Annals of Mathematical Statistics* 39, 1594–1605.

Parzen, E. (1960). *Modern Probability Theory and Its Applications.* New York: John Wiley and Sons, Inc.

Sommer, J. E., R. A. Hoppe, R. C. Green, and P. J. Korb. (1998, December). "Structural and Financial Characteristics of U.S. Farms, 1995: 20th Annual Family Farm Report to Congress." Agriculture Information Bulletin No. 746, USDA/Economic Research Service, Washington, DC. [108 pp.]. Online. Available at http://www.ers.usda.gov/publications/aib746/aib746f.pdf.

U.S. Department of Agriculture. (1993, September 23). *Joint Policy on Variance Estimation and Statistical Reporting Standards on NHANES III and CSFII Reports: HNIS/NCHS Analytic Working Group Recommendations.* Washington, DC.