The *Journal of Cotton Science* 5:137-141 (2001)
http://journal.cotton.org, © The Cotton Foundation 2001

CONTEMPORARY ISSUES

Common Use of the CV: A Statistical Aberration in Crop Performance Trials

Daryl T. Bowman *

INTERPRETIVE SUMMARY

Statistics are used to interpret data and sometimes to determine whether data are suspect. Most researchers are familiar with the use of the least significant difference (LSD) needed to separate two or more means. Another statistical measurement familiar to many scientists and taught in most basic statistics course is the coefficient of variation or the CV. This familiar measurement was created in the late 1800s as a measure of population variability. However, ever since it was tacitly promoted as a measure of experimental validity by Snedecor and Cochran, its original purpose has been largely ignored.

Many scientists use the CV to accept or reject the validity of trials. The CV is based on the assumption that the mean and error variance change together at a constant rate such that the natural log of the error variance is twice the natural log of the mean. An examination of 22 sets of yield trial data showed that this relationship did not exist in real data. Thus, there is no basis for using the CV to discard trials for the seven crops examined in this study. These crops included barley, corn, cotton, flax, oat, soybean, and wheat. If heterogeneity of error variances is a concern, the data can be transformed prior to analyses. Of more significance than heterogeneous errors is the size of the means; higher means have the largest impact on overall means. The practice of using the CV to discard questionable trials should be abandoned. Alternative statistical measures to examine data validity have been published.

D.T. Bowman, Crop Science Dep., North Carolina State Univ., Raleigh, NC 27695-8604. Received 27 Nov. 2000. *Corresponding author (daryl_bowman@ncsu.edu).

Abbreviations: EMS, error mean squares; β, estimated regression coefficient.

ABSTRACT

The coefficient of variation (CV) was created to measure population variability. However, its most common use is to measure validity of field experiments. The CV can be used to measure variability in genetic populations, to determine the best plot size in uniformity trials, to measure stability of phenotypes, or measure variation in other individual or population attributes. The CV is based on the assumption that the mean and the error variance change together such that regressing the natural log of error variance on the natural log of the mean produces a $\beta = 2.0$. Twenty-two sets of crop performance data revealed no relationship approaching $\beta = 2.0$. Thus, there is no basis for using the CV for crop performance trials. If concern exists about heterogeneity of error variances, the data could be transformed on the basis of the relationship between the error and mean for that crop. Locations with higher yield have a larger impact on overall means than locations with lower yield, but one could minimize that impact by calculating relative yield. The CV should no longer be used to indicate validity in most field trials, particularly crop performance trials.

Agricultural research involves formulation of a hypothesis, experimentation, and acceptance or rejection of the hypothesis based upon statistical procedures. The hypothesis may be associated with the appropriate use of various fertilizer treatments, numerous cultivars, or other factors. Experiments sometimes are designed to test a null hypothesis. In cultivar performance trials, one likely null hypothesis is that there are no differences among the various cultivars. An alternative null hypothesis—that there are differences—has been advocated by Lund et al. (1991). In either case, experiments often are designed to minimize experimental error so true hypotheses would be rejected only occasionally and false hypotheses would be rejected nearly every time (Cochran and Cox, 1957).

Designing field experiments requires a knowledge of spatial variation within the specific field. One of the principles of experimental design is blocking of replicates. Improper blocking usually inflates experimental error (Warren and Mendez, 1982), which may result in acceptance of false hypotheses or rejection of true hypotheses. The best way to avoid improper blocking is to examine past performance in the field because visual examination of the field may not reveal underlying spatial variability (other than slope). Researchers often develop their field plans under circumstances in which field history is generally not known. It is generally believed that soil at the top of a slope is not as deep as the soil at the bottom (Pearce, 1995), thus blocks should be laid out accordingly. However, blocking up and down the slope was very effective for yield in a flue-cured tobacco (*Nicotiana tabacum* L.) trial (Bowman, 1990).

Spatial variation may negate any possible blocking arrangement. Bhatti et al. (1991) showed tremendous spatial variability in cotton (*Gossypium hirsutum* L.) and wheat (*Triticum aestivum* L.) trials, to the point that blocking was essentially useless. Although normally practiced, contiguous blocks in an experimental design are not necessary (Wilcox and Zhang, 1999).

Experimental error may become inflated by improper blocking, mistakes made in the field, errors associated with all field operations including planting and harvest, soil variability, and other unknown and known sources of variation. Various techniques (such as spatial analyses) often can be used to reduce experimental error once the trials are harvested. These techniques include the nearest-neighbor analysis, trend analysis, and covariant analyses (Brownie et al., 1993). Even when the proper experimental design, the most efficacious blocking, and error-reducing analyses are used, some experiments may still result in accepting a false hypothesis or rejecting a true hypothesis. For this reason, researchers use statistical measures to determine whether an experiment is generally valid.

One frequently used measure has been the coefficient of variation or CV. In fact, a measurement of experimental validity is the most frequent use of a CV. The basis for use of the CV to measure experimental validity may not have originated with, but certainly was promoted by, Snedecor and Cochran (1967). They stated, "a knowledge of relative variation is valuable in evaluating experiments." They go on to say "… the coefficient of variation of the yield of hay is comparable to that of the yield of corn." Consequently, the CV has been used as a measure of validity for many types of agricultural experiments, from fertilizer experiments to crop performance trials.

The CV was created as a measure of relative variability by Karl Pearson in 1895 (Kendall and Stuart, 1977). With the CV, a measure of variation, the standard deviation, is expressed as a fraction of the mean. As such, the CV can be used to indicate variability among populations, stability of phenotypes, variability of individual plot sizes in uniformity trials, or similar situations in which individual variability is measured. Aflakpui (1995) states that "… there is no such thing as a CV for an experiment, only for individual variates." However, a CV of 10 to 15% for yield in experiments is generally expected by most field crop researchers. This expectation probably originated from Cochran and Cox's statement that "… the coefficient of variation is often between 5% and 15%." The authors were referring to corn (*Zea mays* L.) yield trials in the Midwest. Soil variability is often much greater in other parts of the country than in the Midwest. In any case, the CV continues to be used by many researchers as a measure of experimental validity. This article reviews the basis of the CV and the possibility of its use in crop performance trials.

The basis for the CV is the assumption that the variance increases as the size of the mean increases. Allen et al. (1978) reported a positive relationship for several crops, and Gotoh and Osanai (1959) reported a positive relationship for wheat. Although a positive relationship may be found in several crops, use of the CV is valid only when the b value of the regression of the natural log of the error variance on the natural log of the mean equals 2.0 (Bowman and Rawlings, 1995). This relationship can be shown algebraically, as reported by Bowman and Watson (1997):

$$CV = s/x = (EMS)^{1/2}/x$$
$$s^2 = CV^2x^2 \text{ or } EMS = CV^2x^2$$
$$\ln(s^2) = 2\ln(CV) + 2\ln(x)$$

EMS is error mean squares, x is arithmetic mean, and s^2 is variance. If this relationship between error variance and mean is less than $\beta = 2.0$, then lower CVs would be achieved with higher means.

Table 1. Relationship for a constant CV using hypothetical data.

CV	Mean	SD	EMS	Ln(mean)	Ln(EMS)
10%	500	50	2500	6.21	7.82
10%	1000	100	10 000	6.91	9.21
10%	1500	150	22 500	7.31	10.02
10%	2000	200	40 000	7.60	10.60
				Δ = 1.39	Δ = 2.78

Table 1 demonstrates this relationship for a constant CV using hypothetical data. A simple way to look at the relationship is to compare the differences in the change of ln(mean) with ln(EMS). The range in ln(EMS) is twice the size of the change of ln(mean) (2.78 vs. 1.39), thus resulting in a β value of 2.0.

Table 2. Regression coefficient data of the natural log of the error variance on the natural log of the mean for five agronomic crops in North Carolina crop performance trials.†

Crop	Maturity	β
Barley		0.58*
Corn	Early	0.12
	Mid	-0.05
	Full	-0.11
Oat		1.17*
Soybean	V	0.73*
	VI	0.05
Wheat		0.60*

*** Significantly ($P < 0.05$) different from zero.**
† Data from Bowman and Rawlings (1995) used with permission.

One must examine historical data from the specific crop to determine whether the relationship between the error variance and mean matches the requirement for a valid use of the CV. Bowman and Rawlings (1995) looked at five agronomic crops grown in North Carolina (Table 2, data reproduced with permission). Of eight data sets, only four showed a positive relationship between error variance and mean. All eight were significantly less than 2.0. Data from Mississippi crop performance trials (Table 3) also are presented with permission (Bowman and Watson, 1997). Of nine crops, seven showed a positive relationship between error variance and mean, although all nine were significantly lower than 2.0. Bowman and Watson (1997) also showed data from a study by Allen et al. (1978), which is reproduced here by permission (Table 4).

Table 3. Regression coefficient data of the natural log of the error variance on the natural log of the mean from Mississippi crop performance trials.†

Crop	Maturity	β
Corn	Early	0.57*
	Late	0.40
Cotton		0.77*
Soybean	IV	0.96*
	V	0.43*
	VI	0.27
	VII	0.78*
Oat		1.11*
Wheat		0.78*

*** Significantly ($P < 0.05$) different from zero.**
† Data used with permission from Joe Askew, former manager of MAFES Variety Testing, and Ted Wallace for the cotton data.

Table 4. Regression coefficients of the natural log of the error variance on the natural log of the mean calculated from published yield data for five crops.†

Crop	β
Barley	0.78
Flax	0.79
Oat	1.31
Soybean	0.44
Wheat	0.50

† Values were calculated by regressing the ln (EMS) on the ln ($\bar{x}$) yield for the tabulated data presented by Allen et al. (1978). Permission to use the data is acknowledged.

Positive relationships exist for many crops, although none approach β = 2.0. For some crops, there appeared to be consistency in the relationship. Oat (*Avena sativa* L.) data ranged from 1.11 for Mississippi to 1.17 for North Carolina to 1.31 for Allen et al. (1978). Wheat ranged from 0.50 for Allen et al. (1978) to 0.60 for North Carolina and 0.78 for Mississippi.

These crop performance data reveal error/mean relationships that do not approach β = 2.0, bringing the use of the CV for checking validity of crop performance trials into question. However, some rational method must be applied for accepting or rejecting data. Unfortunately, there are no hard and fast rules to follow in making this decision. Some researchers arbitrarily use a CV >15% to decide when to throw out a crop performance trial, as implied by Snedecor and Cochran (1967). Because the CV is highly influenced by the mean, throwing out an experiment based on a CV >15% would be foolish without examining the mean. Snedecor and Cochran (1967) emphasize that the CV is informative, but not if used without a working knowledge of typical means and error variances.

Alternative statistical measures of validity were studied by Bowman and Watson (1997). A decision to discard a trial or experiment should be based on the error variance typical for that size and type of experiment (Bowman and Rawlings, 1995). One could weight the data (individual entry data by location) by the inverse of the error variance at each location when averaging data over environments (Yates and Cochran, 1938). Highly heterogeneous error variances definitely would affect ranking of overall means by using this technique. Heterogeneous error variances could invalidate tests of significance (Cochran, 1947). Each individual location would have less impact when a large number of locations are used in calculating overall means.

If the relationship between error variance and mean has been established, like that established above, the error variance can be stabilized using the method of Curtiss (1943). Hinz and Eagles (1976) used the method of Curtiss (1943) to reduce heterogeneity of error variances in oat data. As noted previously, the strongest relationship between error variance and mean was in oat data.

Magnitude of individual location means may be as important or even more important as error variances on overall means. Locations with high yields have a larger impact on overall means than do locations with low yields. Brennan and Byth (1979) proposed using relative yields to overcome this problem. Yau and Hamblin (1994) used relative yield not only to give equal weight to each location, but also to provide a measure of stability.

The CV continues to be used primarily for decisions regarding experimental validity in crop performance trials. Although positive relationships between error and mean do exist in many crops, none approach $\beta = 2.0$. Thus, there is no basis for using the CV in reporting crop performance data. Where positive relationships do exist, larger means translate to lower CVs. A researcher would assume that he or she was reducing the chances of rejecting a true hypothesis simply by irrigating the experiment to boost yield when that probability may not have changed at all. The use of the CV should be abandoned, particularly when it is used to discard trials that may or may not yield data rejecting a true hypothesis.

REFERENCES

Aflakpui, G.K.S. 1995. Some uses/abuses of statistics in crop experimentation. Trop. Sci. 35:347–353.

Allen, F.L., R.E. Comstock, and D.C. Rasmusson. 1978. Optimal environments for yield testing. Crop Sci. 18:747–751.

Bhatti, A.U., D.J. Mulla, F.E. Koehler, and A.H. Gurmani. 1991. Identifying and removing spatial correlation from yield experiments. Soil Sci. Am. J. 55:1523–1528.

Bowman, D.T. 1990. Trend analysis to improve efficiency of agronomic trials in flue-cured tobacco. Agron. J. 82:499–501.

Bowman, D.T., and J.O. Rawlings. 1995. Establishing a rejection procedure for crop performance data. Agron. J. 87:147–151.

Bowman, D.T., and C.E. Watson. 1997. Measures of validity in cultivar performance trials. Agron. J. 89:860–866.

Brennan, P.S., and D.E. Byth. 1979. Genotype x environmental interactions for wheat yields and selection for widely adapted wheat genotypes. Aust. J. Agric. Res. 30:221–232.

Brownie, C., D.T. Bowman, and J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: A comparison of methods. Agron. J. 85:1244–1253.

Cochran, W.G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 3:22–38.

Cochran, W.G., and G.M. Cox. 1957. Experimental designs. John Wiley & Sons, New York.

Curtiss, J.H. 1943. On transformations used in analysis of variance. Ann. Math. Stat. 14:107–122.

Gotoh, J., and S. Osanai. 1959. Efficiency of selection for yield under different fertilizer levels in a wheat cross. Jpn. J. Breed. 9:173–178.

Hinz, P.N., and H.A. Eagles. 1976. Estimation of a transformation for the analysis of some agronomic and genetic experiments. Crop Sci. 16:280–283.

Kendall, M., and A. Stuart. 1977. The advanced theory of statistics. Vol. 1. Distribution theory. 4th ed. MacMillan, New York.

Lund, R.E., J.S. Cherry, and J.M. Martin. 1991. Choosing the best by ranking and selection. Agron. J. 83:637–643.

Pearce, S.C. 1995. Some design problems in crop experimentation. I. The use of blocks. Exp. Agric. 31:191–203.

Snedecor, G.N., and W.G. Cochran. 1967. Statistical methods. 6th ed. Iowa State Univ. Press, Ames, IA.

Warren, J.A., and I. Mendez. 1982. Methods for estimating background variation in field experiments. Agron. J. 74:1004–1009.

Wilcox, J.R., and G. Zhang. 1999. Noncontiguous replications in soybean performance tests. Agron. J. 91:72–75.

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci. 28:556–580.

Yau, S.K., and J. Hamblin. 1994. Relative yield as a measure of entry performance in variable environments. Crop Sci. 34:813–817.