SOME THEORETICAL CONSIDERATIONS ON THE USES OF COEFFICIENTS OF VARIATION IN SAMPLE SURVEYS[1]

JOSE S. GUTIERREZ[2]

The more common uses of coefficients of variation ($\frac{\sigma_u}{Eu}$ or simply cv) in sample surveys are in the determination of sample sizes, for specification of the accuracy (precision) of survey results, in the comparisons of efficiencies[3] of two or more survey or experimental designs, and as measures of relative variability. This paper, however, will only consider the uses of coefficients of variation in the comparisons of efficiencies and as measures of relative variability.

Coefficients of Variation As Measures of Relative Variability

On the distribution of coefficient of variation. Mckay (1931) believed that in certain problems arising from the practical application of statistics, coefficients of variation prove to be of as much importance as the absolute values of the means and the standard deviations. For example, when all the observations are by nature positive it is desirable to measure variability relative to the mean rather than in absolute values (Wallis and Roberts, 1963).

---

1 Paper read in the Annual June Conference of the Philippine Statistical Association, June 24, 1965, Manila.

2 Assistant Professor of Statistics, Statistical Center, University of the Philippines.

3 The writer wishes to acknowledge the suggestion of Dr. Alva L. Finkner formerly of North Carolina State College now with the Triangular Institute, North Carolina, as regards this use of the coefficient of variation.

Mckay's (1931) first approximation to the distribution of coefficients of variation was observed to be erroneous by Pearson (1932). Although it is not mathematically accurate, it is adequate for practical purposes and that agreement between the approximation and numerical tests improves as the coefficient of variation decreases and sample size increases (Fieller, 1932). Pearson expressed the opinion that it allows for somewhat too few high values and too many low values but the agreement on the whole is very satisfactory.

Hendricks and Robey (1936) worked out the distribution of the coefficient of variation with the hypothesis that negative and small positive values of the mean occur infrequently. Numerical examples were made to test the validity of this distribution. The agreement between the observed and theoretical on both tests was fairly good graphically, but tests of goodness of fit showed the agreement to be rather poor. The distribution of observed and theoretical values are skewed to the left.

On the relationships of s and $\bar{x}$. Consider the joint distribution of s and X,

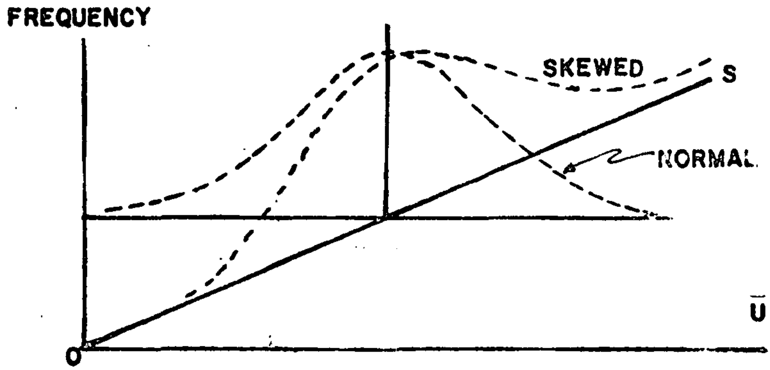
$$f(\bar{x}, s) = Ce^{-n\bar{x}^2/2\sigma^2} s^{n-2} e^{-ns^2/2\sigma^2}$$

as describing a frequency surface (C, a constant). The volume under the surface represents the expected relative frequency of means and standard deviations of all possible samples of size n. In depicting the surface, Deming and Birge ( ) set

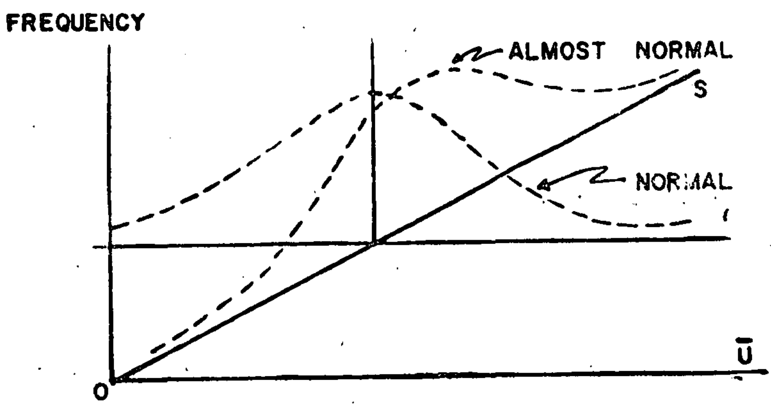
$\bar{u} = \bar{x}-\mu$ so that the origin of u is at $x-\mu$. Also since

$$\int_0^{\infty} \int_0^{\infty} f(x, s)\, d\bar{x}ds = 1$$

the volume under the surface over a closed contour in u, s — plane represents the proportion or percentage of sample means and standard deviations falling simultaneously within the ranges defined by the boundary of the given contour. For two values of n the frequency surfaces are presented by sections in the following figures



N, SMALL, ABOUT 10



N, LARGE, ABOUT 30

These authors also pointed out that the highest point of the surface has the coordinates u = o, $s = \sigma \{\frac{n-2}{n}\}^{\frac{1}{2}}$. If $\bar{x}$ and s are independent, all plane sections with s, constant will be normal curves, with standard deviations, $\sigma/\sqrt{n}$, while, u, constant sections will be skew curves (see figures above) whose equations are given by

$$f(s) = \left(\frac{n}{2\sigma^2}\right)^{\frac{n-1}{2}} \left(\Gamma \frac{n-1}{2}\right)^{-1}$$

$$\exp\{-ns^2/2\sigma^2\}\ s^{n-2}$$

They will have the same mean and mode as n increases, their mean and mode approach coincidence with the value σ, while the curves lose their skewness and become normal with center $s = \sigma$ and standard deviations, $\sigma/\sqrt{2n}$. Also as n increases, the surface becomes more and more concentrated about the point u=o, $s = \sigma$.

A simple relationship between s and x can be established by using Hartley's (1965) concept of the expectation of ratio of two random variables, u and v as follows:

$$E\left(\frac{u}{v}\right) = \frac{E(u)}{E(v)} - \frac{Cov\left(\frac{u}{v} \cdot v\right)}{E(v)}$$

which can be written as

$$E(u) = E\left(\frac{u}{v}\right) E(v) = Cov \left(\frac{u}{v} \, . \, v\right).$$

Suppose s is put in place of u and x instead of v, then a simple relationship between the expectations of $\bar{x}$ and s follows:

$$E(s) = E\left(\frac{s}{\bar{x}}\right) E(x) + Cov\left(\frac{s}{\bar{x}}, \bar{x}\right)$$

$$= E(cv)\, E(\bar{x}) + Cov\,(cv, \bar{x})$$

where the covariance term will take a value zero or not zero depending upon whether the third moment μ_3 is zero or not zero, respectively. μ_3 equals to zero means that the distribution of x's is symmetric and not equal to zero, skewed.

The population standard deviation can be expressed as a linear function[4] of the population coefficient of variation and the population mean as follows:

$$\sigma = a + \beta\mu$$

where

$$\sigma = E(s)$$

$$a = \begin{cases} cov\left(\frac{s}{\bar{x}}, \bar{x}\right), & \mu_3 \neq 0 \\ 0, & \mu_3 = 0 \end{cases}$$

4 Oñate (1965) using experimental data also arrived independently, at the same relationship given above.

$$\beta = E\left(\frac{s}{\bar{x}}\right)$$

$$\mu = E(\bar{x})$$

Illustrative examples of this relationship are

Item	N	s	σ
Net farm	3	2,014.5	658.4
Income	10	1,996.5	
Cash	3	5,568.1	6,976.4
Borrowed	10	6,791.4	

Illustrative examples of the magnitudes of μ, σ^2, μ_3 and $\frac{\sigma}{\mu}$ are as follows:

Item	μ	σ^2	μ_3	$\frac{\sigma}{\mu}$
Close[5] Segment	11.06	115.30	1,312.57	.3070
Farm[5] Headquarters	7.98	84.12	1,365.96	1.1493
Net farm[6] Income	2,118.38	4,334,767.60	36,151,973,656.00	.9827
Cash[6] Borrowed	3,860.44	48,677.321.00	10,251,581,503,036.00	1.8072

<u>Bias of the estimated coefficient of variation.</u> The sources of bias of the estimated coefficient of variation are the bias in s and the bias due to the covariance between $\frac{s}{\bar{x}}$ and $\bar{x}$.

---

5 Source Miraville, Comparison of Two Estimates of Relative Efficiency.

6 Source Gutierrez, Regression Analyses for Evaluation and Planning of Economic Development Programs.

The E(s) is

$$E(s) = b(n)\,\sigma$$

where b(n) is asymptotically equal to (Romanovsky. 1925):

$$1 - \frac{3}{4n} - \frac{7}{32n^2} - \cdots\cdots$$

Hence for large n, the bias in s can be ignored. This implies that for large n and $\mu_3 = 0$, $\frac{s}{\bar{x}}$ can be considered

as an unbiased estimator of $\frac{\sigma_u}{Eu}$ (or $\frac{\sigma}{\mu}$). However, even for large N and $\mu_3 \neq 0$ the bias of $\frac{s}{\bar{x}}$ as an estimator of $\frac{\sigma_\mu}{Eu}$ is

$$-\frac{\operatorname{cov}\left(\frac{s}{\bar{x}}, \bar{x}\right)}{E(\bar{x})} .$$

Variance of the estimated coefficient of variation. By definition the variance of $\frac{s}{\bar{x}}$ is

$$\operatorname{Var}\left(\frac{s}{\bar{x}}\right) = E\left[\frac{s}{\bar{x}} - E\left(\frac{s}{\bar{x}}\right)\right]^2$$

which can be estimated by

$$\operatorname{var}\left(\frac{s}{\bar{x}}\right) = \frac{\sum^{n}\left(\frac{s}{\bar{x}}\right)^2 - \frac{\sum^{n}\left(\frac{s^2}{\bar{x}}\right)}{n}}{n-1}$$

The estimated variance of the coefficient of variation, $\frac{s}{\bar{x}}$

was studied using net farm income and cash borrowed by Iowa farmers as reported in a farm-socio-economic survey conducted in 1960. Two sample sizes were also considered in this study, namely, 3 and 10. One hundred samples of each of these samples sizes were drawn. (The population used in this study is only a sub-sample of an Iowa farm-socio economic survey, hence n largar than 10 was not investigated.)

The results are summarized as follows:

Items	Sample Size	$\frac{s}{\bar{x}}$	$var\left(\frac{s}{\bar{x}}\right)$	$cov\left(\frac{s}{\bar{x}}, \bar{x}\right)$	Bias
Net farm Income	3	1.0473	.1996	-2.5236	+.0646
	10	1.0589	.1910	.0942	+.0762
Cash	3	1.3512	.4149	+2.6772	-.3560
borrowed	10	1.6798	.2097	-1.0442	-.1277

Losses and break-even values (zeros) are included in the item, net farm income, while item cash borrowed includes only zero item (no cash borrowed) values.

The effect of screening zero values on the coefficient of variation. The effect of excluding zero values in the computation of the coefficient of variation was also studied. The results are as follows:

Item	Coefficient of Variation $\frac{\sigma_o}{\mu_o}$
Net farm income	.8673
Cash borrowed	.4275

Screening the zero values will reduce the coefficient of variation if the proportion of zero values is not centrally (or nearly centrally) located. The relation between σ_o^2 (variance with zeros excluded) and σ^2 (variance with zeros included) is given by the following relation:

$$\sigma_o^2 = \frac{\sigma^2}{p} - \frac{q}{p^2}\mu^2$$

where p is the proportion of zero values, q = 1-p and μ the population mean. If the zero values are centrally (or nearly centrally) located, $\sigma_o^2 > \sigma^2$, which implies that $\frac{\sigma_o}{\mu_o} > \frac{\sigma}{\mu}$. If, however, the zero values are on either side of the distribution, $\sigma_o^2 < \sigma^2$ and $\frac{\sigma_o}{\mu_o} < \frac{\sigma}{\mu}$.

As given above there is a considerable reduction in the coefficient of variation for cash borrowed, while for net farm income, the coefficient was not reduced as much as that for cash borrowed due to the presence of a small proportion of negative values.

Approximate variance formula of the estimated coefficient of variation. Since in general only one estimate of the coefficient of variation is available the variance formula discussed in an earlier section is not very useful. However, an approximate formula based on the mean square error can be used. This approximate formula is as follows:

$$\mathrm{Var}\left(\frac{s}{\bar{x}}\right) \doteq \frac{E^2(s)}{E^2(\bar{x})}\left[\frac{\mathrm{Var}(s)}{E^2(s)} - \frac{\mathrm{Var}(\bar{x})}{E^2(\bar{x})} - \frac{2\,\mathrm{Cov}(s,\bar{x})}{E(s)\;E(\bar{x})}\right]$$

Putting the estimator of Var s and Var $\bar{x}$ in the above formula the following approximate variance formula is obtained:

$$\text{Var}\left(\frac{s}{\bar{x}}\right) \doteq \frac{E^2(s)}{E^2(\bar{x})} \left[\frac{s^2}{2nE^2(s)} + \frac{s^2}{nE^2(\bar{x})} - \frac{2\text{Cov}(s,\ \bar{x})}{E(s)\ e(\bar{x})}\right]$$

$$\doteq \frac{E^2(s)}{E^2(\bar{x})} \left[\frac{1}{2n} + \frac{(cv)^2}{n} - \frac{2\text{Cov}(s,\ \bar{x})}{E(s)\ E(\bar{x})}\right]$$

If the covariance term is zero the approximate variance formula can be written as follows:

$$\text{Var}\left(\frac{s}{\bar{x}}\right) \doteq \frac{E^2(s)}{E^2(\bar{x})} \left[\frac{\text{Var}\ (s)}{E^2\ (s)} + \frac{\text{Var}\ (\bar{x})}{E^2\ (\bar{x})}\right]$$

$$\doteq \frac{E^2(s)}{E^2(\bar{x})} \left[\frac{1}{2n} + \frac{(cv)^1}{n}\right]$$

Both can be estimated by using the following formula:

$$\text{Var}\left(\frac{s}{\bar{x}}\right) \doteq \frac{(cv)^2}{2n}$$

Using the approximate formula the estimated variances are:

Items	Sample size	Estimated Variance
Net farm	3	.0619
Income	10	.0056
Cash	3	.3042
Borrowed	10	.1410

The Coefficient of Variation As A Measure of Relative Efficiency

The need for thorough statistical investigation of different sampling designs is great. Quite often such investigations are neglected since once a survey has been conducted, the question of whether it could have been carried out more efficiently is purely historical as far the survey is concerned (Yates, 1953).

The ratio of squared coefficients of variation can be used as a measure or indicator of relative efficiency.

Rel-variance ($\sigma^2_u/\bar{E}^2_u$) of the ratio of squared coefficients of variation.

Comparison of efficiencies of two or more surveys or experimental designs is commonly done with the use of ratio of variances (F-distribution) but the ratio of squared coefficients of variation, especially in survey sampling can also be used. If we let R be the indicator of the true relative efficiency which is equal to the ratio of population variances, then we can write R symbolically as follows:

$$R = \frac{\sigma^2_2}{\mu^2} \Big/ \frac{\sigma^2_1}{\mu^2} = \frac{\sigma^2_2}{\sigma^2_1}$$

where σ^2_i (i = 1, 2) is the variance of each type of sampling approach and μ, the population mean. If instead of the parameter σ^2_i and μ, s^2_i and $\bar{x}_i$ (i = 1, 2) are used, then we can estimate the parameter R by the two estimators, namely:

$$\hat{R}cv = \frac{S^2{}_2}{\bar{x}_2^2} \Big/ \frac{S^2{}_1}{\bar{x}_1^2}$$

$$\frac{\hat{R}cv_1}{\hat{R}cv_2}$$

and

$$\hat{R}v = \frac{S^2{}_2}{S^2{}_1}$$

Using the concepts given by Hansen, Hurwitz and Madow (1956), the rel-variances of the estimators are

$$V^2_{Rc} \doteq V^2_{\hat{R}cv_1} + V^2_{\hat{R}cv_2}$$

$$\doteq \frac{\beta_1 - 1}{n} + \frac{4}{n}\frac{\sigma_1^2}{\mu_1 2} - \frac{4}{n}\frac{\mu_{3(1)}}{\sigma_1^2 \mu_1}$$

$$\frac{\beta_2 - 1}{n} + \frac{4}{n}\frac{\sigma_2^2}{\mu_2^2} - \frac{4}{n}\frac{\mu_{3(2)}}{\sigma_2^2 \mu_2}$$

with $\mu_1 = \mu_2$. $V^2_{\hat{R}cv}$ can be written as

$$V^2_{\hat{R}cv} \doteq \frac{\beta_1 - 1}{n} + \frac{4}{n}\frac{\sigma_1^2}{2} - \frac{\mu_{3(1)}}{\sigma_1^2 \mu}$$

$$+ \frac{\beta_2 - 1}{n} + \frac{4}{n}\frac{\sigma_2^2}{\mu^2} - \frac{4}{n}\frac{\mu_{3(2)}}{\sigma_2^2 \mu} .$$

and

$$V^2_{Rv} \doteq V^2_{Rv_1} + V^2_{Rv_2}$$

$$\doteq \frac{\beta_{1-1}}{n} + \frac{\beta_2 - 1}{n}$$

where $\beta_i = \frac{\mu_{4(i)}}{\sigma_i^{\,4}}$ and with the assumption that $\hat{R}_{cv_1}$ is uncorrelated with $\hat{R}_{cv_2}$ and $\hat{R}_{v_1}$ uncorrelated with $\hat{R}_{v_2}$:

$$V^2_{\hat{R}cv} \quad V^2_{\hat{R}v}$$

if

$$\frac{4}{n}\ \frac{\sigma_i^2}{\mu^2} - \frac{4}{n} - \frac{\mu_{3(i)}}{\mu\ \sigma_i^2} \quad < \quad 0$$

or

$$\frac{4}{n}\frac{\sigma_1^2}{\mu^2}\left[1 - \frac{\mu_{3(i)}}{\sigma_i^4}\right] \quad < \quad 0$$

or

$$\mu_3(i) \quad \frac{\sigma_i^{\,4}}{\mu} \quad < \quad 1$$

In a symmetric population this i's not true since $\mu_3 = 0$. Miravalle (1957) pointed out that when the mean and variance are correlated, $\mu_3 \neq 0$. and if $\mu_3 > \frac{\sigma^4}{\mu}$ then $\overset{2}{\hat{R}cv} < V^2_{\hat{R}v}$.

This appears to be reasonable since in a sample form a skewed population if the mean is overestimated or underestimated the variance is also overestimated or underestimated and the ratio $\frac{S^2}{\bar{x}^2}$ should partially compensate for any poor estimators of the numerator and denominator and thus yielding a better measure of the true population coefficient of variation, and hence of the true R, then the estimated variance would give us of the true σ^2 .

Miravalle (1956) using the results from two sampling systems for estimating cotton acreage have shown that $V^2_{\hat{R}v} = 0.181$ and $V^2_{\hat{R}cv} = 0.137$.

The systems used were closed segment and farm headquarter approach. The results from both systems are expected to be skewed distributions. In this case of skewed populations and with an assumption of independence Miravalle remarked that the ratio of the coefficients of variation seemed to be a better estimate of the relative efficiency than is the ratio of variances.

The writer (1958) in an unpublished research studied the influence of type of agricultural items, the influence of matched and unmatched samples and the influence of sample size on the feasibility of the ratio of squared coefficients of variation as indicator of relative efficiency.

For studying the influence of type of agricultural items the following were used:

Cultivated land and pasture (for brevity, pasture)
Cultivated land and tobacco (for brevity, tobacco)
and for sample sizes:

$$n = 20$$
$$n = 100$$

Two hundred samples of size 20 and 100 were drawn from an assumed population of tobacco and pasture. The mean of R's by item, estimator, type of sample and are summarized as follows:

Type of Sample	Estimator	Pasture		Tobacco	
		n = 20	n = 100	n = 20	n = 100
Matched	$\hat{R}_{cv}$	3.4578	3.8306	1.0701	1.0724
	$\hat{R}_{v}$	4.9026	4.1707	1.1299	1.0873
Unmatched	$\hat{R}_{cv}$	3.5960	3.8811	1.3434	1.1358
	$\hat{R}_{v}$	6.4484	4.3969	1.9025	1.2225
True B		3.9704		1.0818	

The estimated variances of $\hat{R}$'s by item, estimator, type of sample and size of sample are given in the following table:

Type of Sample	Estimator	Pasture		Tobacco	
		n = 20	n = 100	n = 20	n = 100
Matched	$\hat{R}_{cv}$	7.8929	3.3775	0.2198	0.0675
	$\hat{R}_{v}$	70.0134	7.9838	0.3769	0.0911
Unmatched	$\hat{R}_{cv}$	9.2606	4.0821	1.6502	0.2523
	$\hat{R}_{v}$	169.1016	12.4210	11.4144	0.5746

where the variance was estimated by using

$$\mathrm{Var}\ (\hat{R}_j) = \frac{1}{n(n-1)} \sum^{n} (\hat{\bar{R}}_j - E\ \hat{\bar{R}}_j)^2,\ j = cv,\ v.$$

In studying the relative efficiency of the estimator Rcv and Rv, Pitman's "closeness" criterion was used. Geary (1943) in the normal case, evaluated the probability of

$| x - e | < | Y - e |$ greater than ½ by using the equation:

$$p = \frac{1}{\pi} \tan^{-1} (2\sigma_x \sigma_y \sqrt{1-\rho^2} / \sigma_x^2 - \sigma_y^2)$$

Geary added that for all values of ρ the probability is greater than or less than ½ as σ_x is respectively less than or greater than σ_y.

Applying Geary's formula in evaluating Pitman's criterion of Closeness, the probability of $|\hat{R}_{cv} - \hat{R}| < | \hat{R}_v - \hat{R} |$ will be greater than ½ if σ_{Rcv} is ~~less than $\sigma_{\hat{R}cv}$~~ and will be less than ½ if $\sigma_{\hat{R}cv}$ is less than $\sigma_{\hat{R}v}$. Three cases are evident, namely:

Case I. $\sigma_{\hat{R}cv} = \sigma_{\hat{R}v}$. If σ_{Rcv} is equal to $\sigma_{\hat{R}v}$, the probability is equal to ½; hence, it would not make any difference which estimator is to be used. Now if $\sigma_{\hat{R}cv} = \sigma_{\hat{R}v}$ then

$$p = \frac{1}{\pi} \tan^{-1} (2\sigma_{\hat{R}cv}\, \sigma_{\hat{R}v} \sqrt{1-\rho^2} / \sigma^2_{\hat{R}cv} - \sigma^2_{\hat{R}v}) = \tfrac{1}{2}$$

Case II. $\sigma_{\hat{R}cv} > \sigma_{\hat{R}v}$. If $\sigma_{\hat{R}cv}$ is greater than σ_{Rv} such that Rv is definitely better than $\hat{R}_{cv}$, the probability of $|\hat{R}_{cv} - R| < |\hat{R}_v - R|$ will approach zero as $\sigma_{\hat{R}v}$ approaches zero.

i.e.
$$p = \frac{1}{\pi} \tan^{-1} \left(2\, \sigma_{\hat{R}cv}\, \sigma_{\hat{R}v} \sqrt{1-\rho^2} \big/ \sigma^2_{\hat{R}cv} - \sigma^2_{\hat{R}_v}\, \sigma_{\hat{R}_v}\right) \xrightarrow{\sigma_{\hat{R}v} \to 0} 0$$

Case III. $\sigma_{\hat{R}_{cv}} < \sigma_{\hat{R}_v}$. If $\sigma_{\hat{R}cv}$ is less than $\sigma_{\hat{R}v}$ such that the probability of $|\hat{R}_{cv} - R| < |\hat{R}_v - R|$ approaches unity as $\sigma\hat{R}_v$ approaches zero, i.e.

$$P = \frac{1}{\pi} \tan^{-1} \left(2\sigma_{\hat{R}cv}\sigma_{\hat{R}v}\sqrt{1-\rho^2} / \sigma 2_{\hat{R}_{cv}} - \sigma 2_{\hat{R}_v}\, \sigma 2_{\hat{R}v}\right) \xrightarrow{\sigma_{\hat{R}v} \to 0} (-) \cdot 0$$

Diagrammatically, these cases can be shown as follows:

I

$\sigma_{\hat{R}_{cv}} = \sigma_{\hat{R}_v}$

$p = \frac{1}{2}$: $\theta = \frac{\pi}{2}$

III	II
$\sigma_{R_{cv}} < \sigma_{R_v}$	$\sigma_{\hat{R}_{cv}} > \sigma_{\hat{R}_v}$
$\frac{1}{2} < P \leq 1,\ \frac{\pi}{2} < \theta \leq \pi$	$0 \leq P < \frac{1}{2},\ 0 \leq \theta < \frac{\pi}{2}$

0

that is for any value of θ in the first quadrant $\hat{R}_v$ would be a better estimator than $\hat{R}_{cv}$ and for any value of θ in the second quadrant. $\hat{R}_{cv}$ would be a better estimator than $\hat{R}_v$.

The probability values obtained in above analysis are as follows:

Item	Type of Sample	Size of Sample	Probability Value
Pasture	Matched	20	0.6193
		100	0.6580
	Unmatched	20	0.6649
		100	0.8314
Tobacco	Matched	20	0.7649
		100	0.8155
	Unmatched	20	0.7315
		100	0.8728

Evidently, the results of the above analysis indicate that $\hat{R}_{cv}$ is a better estimator of R than $\hat{R}v$. Generalization, however, is difficult since the study was limited to skewed distributions.

Bias in estimates. If the population are normal, then both estimators are unbiased since

$$E\,(\hat{R}v) = E\left(\frac{S_1^{\,2}}{S_2^{\,2}}\right) = \frac{E(S_1^{\,2})}{E(S_2^{\,2})};$$

S_1^2 and S_2^2 being independent. Likewise

$$E(R_{cv}) = E\left(\frac{R_{cv_1}}{R_{cv_2}}\right) = \frac{E(R_{cv_1})}{E(R_{cv_2})};$$

$\hat{R}_{cv_1}$ and $\hat{R}_{cv_2}$ being independent, and also the sample mean and variance for the normal populations are independent (Mood 1963).

Miravalle (1957) on the other hand considered the case when $\bar{x}^2$ is not independent of s^2 . She estimated the bias for $\hat{R}_{cv}$ as

$$\text{Bias} = E(\hat{R}_{cv} - R)$$

$$= \frac{\sigma_1^2 + \sigma_1^2\left[\frac{4\sigma_1^2}{n\mu^2} - \rho \bar{s}_1^2 x_2^2 \frac{2\sigma_1}{n\mu} B_1 - 1\right]}{\sigma_2^2 + \sigma_2^2\left[\frac{4\sigma_2^2}{n\mu^2} - \rho s_2^2 x_2^2 \frac{2\sigma_2}{n\mu} B_2 - 1\right]} - \frac{\sigma_1^2}{\sigma_2^2}$$

where B i (i = 1 2) is the bias in R_{cv_i} (i = 1, 2). Miravalle's results on the value of $\frac{\text{Bias}}{R}$ proportional bias, for different Values of ρ, β i and σ_i^2 are as follows:

ρ	β_1	σ_1^2	β_2	σ_2^2	$\frac{\text{Bias}}{R}$
0	7	100	7	100	0
0	7	225	7	100	.084
0	9	100	6	100	0
0	9	225	6	100	.048
0	10	100	5	100	0
0	10	225	5	100	.048

.5	7	100	7	100	0
.5	7	144	7	100	.013
.5	7	225	7	100	.037
.5	9	100	9	100	—.006
.5	9	144	9	100	.006
.5	9	225	9	100	.029
.5	10	100	10	100	.010
.5	10	144	10	100	.002
.5	10	225	10	100	.025
1.0	7	100	7	100	0
1.0	7	225	7	100	.026
1.0	6	100	9	100	—.012
1.0	6	225	9	100	.010
1.0	5	100	10	100	—.020
1.0	5	225	10	100	0

Using the same populations (skewed) the writer arrived at the following biases for the $\hat{R}_{cv}$'s and $\hat{P}_v$'s by item, type of sample and size of sample:

Type of Sample	Estimator	Pasture		Tobacco	
		n = 20	n = 100	n = 20	n = 100
	$\hat{R}_{cv}$	−0.5216*	− 0.1398	−0.0117	− 0.0094
Matched	$\hat{R}_v$	+0.9322	+ 0.2003	+0.0481	+ 0.0055
	$\hat{R}_{cv}$	−0.3744	− 0.0893	+0.2616*	0.0540
Unmatched	$\hat{R}v$	+=.4780*	− 0.4265	+0.8207*	0.1407
R		3.9704		1.081'	

a/ Significant, 5% level

LITERATURE CITED

Cochran, W.G. 1950. The comparison of percentages in matched samples. Biometrika 37; 256-266

Deming, W.E. and Birge, R.T. — On the Statistical Theory of Errors Review of Modern Physics 6: 119-161 (as cited in Kenney, J.F. and Keeping, E.S. Mathematics of Statistics, Part Two)

Finkner, A.L. 1952, Adjustment for Non-Response Bias in a Rural Mailed Survey, Agricultural Economics Research 4; 77-82.

Geary, R.C. 1943. Comparison of the Concepts of Efficiency and Closeness for Consistent Estimators of a Parameter, Biometrika 33: 123-128.

Gutierrez, J.S. 1958. On the Comparison of the Ratio of Squared Coefficients of Variation and Ratio of Variances as Indicators of Relative Efficiency. Unpublished M.S. thesis, North Carolina State College.

Gutierrez, J.S. 1965. Regression Analysis in the Evaluation and Planning of Economic Development Programs. Unpublished research Iowa State University.

Hamaker, H.C. 1949. Statistika 3: 209 (as cited in N.L. Johnson 1950. On the Comparison of Estimators Biometrika 37:287).

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. 1956. Sample Survey: Methods and Theory Vol. II: Theory John Wiley and Sons, Inc. New York.

Hartley, H.O. 1961. Lectures on Advanced Survey Designs, Iowa State University.

Hendricks, W.A. and Robey, K.W. 1936. The Sampling Distribution of the Coefficients of Variation. The Annals of Mathematical Statistics 7:129-132.

John, N.L. 1950. On the Comparison of Estimators. Biometrika 37: 281-287

Landau, H.G. 1947. University of Pittsburgh Bulletin 43: 443 (as cited in N.L. Johnson, 1950. On the Comparison of Estimators.. Biometrika 37:287).

Mckay, A.T. 1931. The Distribution of the Estimated Coefficients of Variation. Journal of Royal Statistical Society 94: 564-567.

Mckay, A.T., Fieller, E.C. and Pearson, E.S. 1932. Distribution of the Coefficients of Variation and the Extended to Distribution. Journal of Royal Statistical Society. 95: 695-704.

Miravalle, S.J. 1956. A Comparison of Alternative Methods for Defining and Allocating Area Sampling Units for Agricultural Surveys. Unpublished Master's Thesis, North Carolina State College.

Miravalle, S.J. 1957. Comparison of Two Estimates of Relative Efficiency. Unpublished Report, North Carolina State College.

Mood, A.M. and Graybill, F.A. 1963. Introduction to the Theory of Statistics McGraw-Hill Book Company, Inc. New York.

Oñate, B.T. 1965. SQC in Agricultural Research. Paper presented at the Third SQC Seminar, Philippine Statistical Association.

Pitman, E.J.G. 1937. The Closest Estimators of Statistical Parameters. Proceedings of Cambridge Philosophical Society. 33: 212-222.

Romanovsky, V. 1925. On the Moments of the Standard Deviation and of the Correlation Coefficients in Samples from Normal. Metron 5:3-46 (as cited in Kenney, J.F. and Keeping E.S. Mathematical Statistics, Part Two).

Wallis, W.A. and Roberts, H.V. 1963. Statistics: A New approach. The Free Press of Glencoe, Inc. Brooklyn, New York.

Yates, F. 1953. Sampling Methods for Censuses and Surveys. Hafner Publishing Company, New York.

Republic of the Philippines
Department of Public Works and Communications
BUREAU OF POSTS
Manila
SWORN STATEMENT

(Required by Act 2580)

The undersigned, BERNARDINO G. BANTEGUI, editor of THE PHILIPPINE STATISTICIAN, published quarterly, in English at 1046 Vergara, Quiapo, Manila, after having been duly sworn in accordance with law, hereby submits the following statement of ownerships, management, circulation, etc., which is required by Act 2580, as amended by Commonwealth Act No. 201:

	Name	Post Office Address
Editor:	BERNARDINO G. BANTEGUI	OSCAS, NEC, Manila
Business Manager:	MERCEDES B. CONCEPCION	P. O. Box 479, Manila
Owner:	PHIL. STATISTICAL ASS'N.	P. O. Box 3223, Manila
Publisher:	PHIL. STATISTICAL ASS'N.	P. O. Box 3223, Manila
Printer:	VENTURA PRINTING PRESS	2416 Juan Luna, Manila
Office of Publication:		Rizal Hall – Padre Faura, Manila

If publication is owned by a corporation, stockholders owning one per cent or more of the total amount of stocks: None

Bondholders, mortgagees, or other security holders owning one per cent or more of total amount of security None

In case of daily publication, average number of copies printed and circulated of each issue during the preceding month of ...., 19....

In case of publication other than daily, total number of copies printed and circulated of the last issue dated DECEMBER, 1964:

1.	Sent to paid subcribers	477
2.	Sent to others than paid subscribers	48
	Total	525

(Sgd.) BERNARDINO G. BANTEGUI
Editor

SUBSCRIBED AND SWORN to before me this 1st day of April, 1965, at Manila, the affiant exhibiting his Residence Certificate No. A-0290040 issued at Manila on February 2, 1965.

Doc. No. 58
Page No. 71
Book No. I
Series of 1965

(Sgd.) FRANCISCO A. FIDELINO
Notary Public
My commission expires on Dec. 31, 1965

NOTE: This form is exempt from the payment of documentary stamp tax.