MODELLING THE COEFFICIENT OF VARIATION IN FACTORIAL EXPERIMENTS

Craig A. Wilson

Mark E. Payton

MODELLING THE COEFFICIENT OF VARIATION IN FACTORIAL EXPERIMENTS

Craig A. Wilson

Department of Biometrics and Data Management
Hoechst Marion Roussel, Inc.
10236 Marion Park Drive
Kansas City, MO 64137

and

Mark E. Payton

Department of Statistics
Oklahoma State University
Stillwater, OK 74078

ABSTRACT

The coefficient of variation (CV) has long been used as a measure of the relative consistency of sample data. However, little attention has been paid to using the CV to make conclusions about the relative consistency of the population(s) from which the data are drawn, particularly when the data are observed in the context of a designed factorial experiment. This research focused on using three approximations to the exact distribution of the sample CV of normally distributed data (McKay's, David's, and Iglewicz and Myers') in the context of the generalized linear model to develop a method for detecting main effects and interactions among factors when the population characteristic of interest is the CV.

1. INTRODUCTION

As the ratio of the sample standard deviation to the sample mean, the sample coefficient of variation (CV) provides a useful and unitless measure of relative variability. As Ahmed (1994) notes, the CV can sometimes be more relevant than the standard deviation alone, such as when the precision of measuring instruments or the volatility of stocks is considered. Hurlimann (1995) points out that the CV is useful in insurance risk assessment as a measure of the heterogeneity of insurance portfolios. Williams (1991) cites the importance of the CV in the determination of detection limits in instrumental analysis. Feltz and Miller (1996) notes that in medical studies, the CV often determines the feasibility of combining results from separate clinical trials.

Payton (1997) suggests that the CV has relevance only for ratio-level populations. In such populations, an observation equal to zero represents the absence of the measured characteristic, such as with populations of volumes, yields, or weights, since only in this context does the CV ratio itself have meaning. Negative observations are not possible.

Although theoretically not ratio level, normal populations have long been considered in connection with the behavior of sample CVs. In such cases, negative sample means are assumed to be highly improbable. However, in contrast with the mean of the normal distribution, comparatively little work has been done in connection with hypothesis tests and confidence intervals for unknown population CVs based on observed data. Papers that have addressed these subjects for a single population CV include Koopmans, et al. (1964), Vangel (1996), and Payton (1997), which utilize exact and approximate distributions of the sample CV from a normal population. Tests for the equality of k normal population CVs that employ approximate distributions and the normal density include Bennett (1976), Doornbos and Dijkstra (1983), and Shafer and Sullivan (1986). Gupta and Ma (1996) extends a Wald test developed by Rao and Vidya (1992) for two populations based on the normal density to k populations and introduces a score test which also utilizes the actual density of the observations. Feltz and Miller (1996) provides a test based on the asymptotic moments of the CV.

Less work has addressed the analysis of population CVs in the context of designed factorial experiments. Taguchi (1992) discusses a well-known approach to the analysis of product quality using fractional factorial designs that often models a log-transformed CV. However, his approach has yielded recent criticisms (see, for example, Box, 1988) and corrections because of biased tests of factor effects. Absent from the literature is a technique for constructing factorial models of the CVs of normal populations that makes use of known approximate distributions and asymptotic moments of the sample CV.

2. TERMINOLOGY AND DEFINITIONS

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with $E(X_i) = \mu > 0$ and $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$, and let $R = \sigma / \mu$ be the population CV. Define $\overline{X} = \sum_{i=1}^{n} X_i / n$ to be the sample mean and assume that $P(\overline{X} < 0)$ is negligible. Let $S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 / (n-1)$ and $S_n^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 / n$ be the unbiased and maximum-likelihood estimates of σ^2, respectively, and let $r = S / \overline{X}$ and $r_n = S_n / \overline{X}$ be the corresponding point estimates of R. Note that r_n is the maximum-likelihood estimate of R and that $r_n = ((n-1)/n)^{1/2} r$. Although neither r nor r_n is an unbiased estimate of R, both are strongly consistent (Serfling, 1980, pp. 24-26, 136-137). Hence, both are reasonable estimators of R, particularly when computed from large samples.

For later convenience, define the h-function $h(x) = x^2 / (1 + x^2)$ for $x > 0$. Then h has an inverse, and $h^{-1}(x) = (x/(1-x))^{1/2}$ for $0 < x < 1$. Additionally, define a random variable Y to have the gamma distribution with parameters λ and ν if and only if its density is given by

$$f(y) = \frac{1}{y\Gamma(\nu)}\left(\frac{\nu y}{\lambda}\right)^{\nu} \exp\left(-\frac{\nu y}{\lambda}\right), \quad y \geq 0$$

$$= 0, \quad y < 0,$$

where $\lambda > 0$, $\nu > 0$, and $\Gamma(\bullet)$ is the gamma function. It follows that $E(Y) = \lambda$ and $Var(Y) = \lambda^2 / \nu$. The parameter ν is sometimes called the index.

3. APPROXIMATE DISTRIBUTIONS OF THE SAMPLE CV

Under normal theory, the exact distribution of r is a multiple ($\sqrt{n}$) of the inverse of a non-central t distribution having $(n-1)$ degrees of freedom and non-centrality parameter $\sqrt{n}/R$. The density of the non-central t for degrees of freedom p and non-centrality parameter q is given by Lehmann (1959, p. 200) as

$$f(t) = \left(2^{(p+1)/2}\Gamma(p/2)(\pi p)^{1/2}\right)^{-1}\int_0^\infty y^{(p-1)/2} \exp\left[-\frac{y}{2}-\frac{1}{2}\left(t\sqrt{\frac{y}{p}}-q\right)^2\right]dy,$$

for $-\infty < t < \infty$. Given the density of r, the density of r_n can be obtained, in turn, by transforming r according to $r_n = \left((n-1)/n\right)^{1/2} r$. Difficulties associated with direct application of the non-central t distribution itself have prompted the study of several approximations to the exact distributions of r and r_n.

3.1 McKay's and David's Approximations

McKay (1932) gives the earliest approximation to the distribution of r_n when samples are drawn from a normal population. By utilizing a contour-integral expression of the density of r_n, he is able to show that $nh(r_n) / h(R)$ has an approximate χ^2 distribution with (n - 1) degrees of freedom, provided that $R \in (0, 1/3)$. This requirement on R is consistent with the added assumption that negative observations also are highly improbable, in addition to a negative sample mean. Equivalently, $(n / (n - 1))h(r_n)$ has an approximate gamma distribution with expectation h(R) and index (n - 1) / 2.

David (1949) obtains an approximation to the distribution of r by reexpressing McKay's approximation in terms of r and deleting a negligible term. Beginning with $nh(r_n) / h(R)$, she writes

$$\frac{nh(r_n)}{h(R)} = \frac{n}{h(R)}\frac{r_n^2}{1+r_n^2} = \frac{n}{h(R)}\frac{\left(\frac{n-1}{n}\right)r^2}{1+\left(\frac{n-1}{n}\right)r^2}$$

$$= \frac{n-1}{h(R)}\frac{r^2}{1+r^2-\frac{r^2}{n}} \approx \frac{n-1}{h(R)}\frac{r^2}{1+r^2} = \frac{(n-1)h(r)}{h(R)},$$

since r^2 / n is typically close to zero for large n. She thus obtains that $(n - 1)h(r) / h(R)$ also has an approximate χ^2 distribution with (n - 1) degrees of freedom, or, equivalently, that h(r) is distributed approximately gamma with expectation h(R) and index (n - 1) / 2.

3.2 Iglewicz and Myers' Approximation

A third approximation for consideration is discussed by Iglewicz and Myers (1970). They derive asymptotic expansions for the moments of the exact distribution of r under normal theory and conclude that an adequate approximation for even relatively small n can be obtained by assuming that r itself is normally distributed with mean R and variance $\left(\frac{R^2}{n}\right)\left(R^2+\frac{1}{2}\right)$. This variance was apparently given originally by Pearson (David, 1949). Both Serfling (1980, pp. 136-137) and Feltz and Miller (1996) note that r is, in fact, asymptotically normal with these same moments. Hence, an application of Slutsky's Theorem gives that r_n likewise possesses these asymptotic properties (Serfling, p. 19). Simulation results reported by Iglewicz and Myers suggest that this approximation is superior to other normal approximations with higher-order expansions for the mean and variance.

4. THE MODELLING APPROACH

Take a collection of CVs $R_1, R_2, ..., R_k$ of normal populations, where, for convenience, a single subscript is used, but where any number of associated fixed factors may be supposed. Assume that the i^{th} population has mean $\mu_i > 0$ and variance σ_i^2, so that $R_i = \sigma_i / \mu_i$, i = 1, 2, ..., k. One possible model structure for the μ_i is then

$$\mu_i = \exp(\mathbf{x}_i'\alpha),\ i = 1, 2, ..., k,$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ is a parameter vector of fixed factor effects ($p \le k$) and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is the i^{th} set of covariate values. For a factorial model, these covariates are properly assigned values of zero or positive or negative one under some identifiability constraint; for example, that the associated parameters summed across any single subscript must equal zero. Similarly, a model for the σ_i may be written as

$$\sigma_i = \exp(\mathbf{x}_i'\gamma), \quad i = 1, 2, \ldots, k,$$

where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$ is the corresponding parameter vector. Combining these models gives a multiplicative model for the R_i:

$$\begin{aligned} R_i &= \frac{\sigma_i}{\mu_i} = \frac{\exp(\mathbf{x}_i'\gamma)}{\exp(\mathbf{x}_i'\boldsymbol{\alpha})} = \exp(\mathbf{x}_i'(\boldsymbol{\gamma} - \boldsymbol{\alpha})) \\ &= \exp(\mathbf{x}_i'\delta), \quad i = 1, 2, \ldots, k, \end{aligned} \tag{4.1}$$

where $\delta = \gamma - \alpha$.

4.1 The Generalized Linear Model

Model (4.1) can be estimated using any one of the approximate distributions outlined above. In the context of a generalized linear model, maximum- and/or quasi-likelihood estimates are available using iteratively reweighted least squares (Wedderburn, 1974). The algorithm is summarized in the following theorem:

Theorem Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)'$ be a vector of independent observations with expectation $\psi = (\psi_1, \psi_2, \ldots, \psi_N)'$ and covariance matrix $\phi\mathbf{V}(\psi) = \phi \text{diag}\{V_1(\psi_1), V_2(\psi_2), \ldots, V_N(\psi_N)\}$. Suppose that there exists a monotone, differentiable function $g(\bullet)$ such that $g(\psi_i) = \mathbf{x}_i'\boldsymbol{\beta}$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ is the i^{th} set of covariates, i = 1, 2, ..., N, and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a vector of parameters. Then a quasi-likelihood estimate of β may be obtained by repeatedly calculating a weighted linear regression of

$$z_i = g(\psi_i) + g'(\psi_i)(y_i - \psi_i)$$

on $\mathbf{x}_i$ using weight

$$w_i = [g'(\psi_i)]^{-2}[V_i(\psi_i)]^{-1},$$

where the current estimates of the ψ_i are computed from the current estimates of $\beta_1, \ldots, \beta_p$.

It follows in general for large samples that the quasi-likelihood estimate $\hat{\beta} \stackrel{\cdot}{\sim} N_p(\beta, \phi(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1})$, where **X** is an N x p matrix of full rank having elements x_{ij} and where $\mathbf{W} = \text{diag}\{w_1, ..., w_N\}$. If the Y_i belong to a distribution in the exponential family, then $\hat{\beta}$ is also a maximum-likelihood estimate. In any event, the maximum-/quasi-likelihood-ratio test is available to test hypotheses involving the β_i (McCullagh, 1983).

4.2 McKay's and David's Approximations

Assume that independent random samples of size n_i are drawn from each of the k normal populations, and that the sample CVs $r_i = S_i / \overline{X}_i$ and $r_{n,i} = S_{n,i} / \overline{X}_i$ are computed. According to David's approximation, $h(r_i)$ is distributed approximately gamma with expectation $h(R_i)$ and index $(n_i - 1) / 2$, so that $\text{Var}(h(r_i)) \approx \frac{2[h(R_i)]^2}{n_i - 1} = V_i(h(R_i))$ (taking $\phi = 1$). Supposing the model (4.1) for the R_i gives, as a model for the $h(R_i)$,

$$h(R_i) = h(\exp(\mathbf{x}_i'\delta)), \; i = 1, 2, ..., k,$$

for which a linearizing transformation is

$$\log h^{-1}(h(R_i)) = \mathbf{x}_i'\delta . \tag{4.2}$$

Model (4.2) is a generalized linear model of the $h(R_i)$ with link function $\log h^{-1}(\bullet)$, but in the parameters of the original model of the R_i, so that estimating (4.2) simultaneously estimates (4.1). Iteratively reweighted least squares may be employed to fit (4.2). Letting $R_i^* = h(R_i)$ and $r_i^* = h(r_i)$, it follows that

$$\begin{aligned} z_i &= \log h^{-1}(R_i^*) + \frac{d(\log h^{-1}(R_i^*))}{dR_i^*}(r_i^* - R_i^*) \\ &= \log h^{-1}(R_i^*) + \frac{r_i^* - R_i^*}{2R_i^*(1 - R_i^*)} \quad , \end{aligned} \tag{4.3}$$

and

$$w_i = \left[\frac{d(\log h^{-1}(R_i^*))}{dR_i^*}\right]^{-2} [V_i(R_i^*)]^{-1} = \left[\frac{1}{2R_i^*(1 - R_i^*)}\right]^{-2} \left[\frac{2(R_i^*)^2}{n_i - 1}\right]^{-1}$$

$$= \quad 2(n_i - 1)(1 - R_i^*)^2 . \tag{4.4}$$

Appropriate starting values for z_i and w_i may be obtained by substituting r_i^* for R_i^* in (4.3) and (4.4). Given that the t^{th} iteration has been made and that the t^{th} estimate $\delta^{(t)}$ has been obtained, the $(t + 1)^{th}$ estimate of δ can be computed after the substitution of $(R_i^*)^{(t)} = h(\exp(\mathbf{x}_i'\delta^{(t)}))$ into (4.3) and (4.4). When changes become acceptably small, the resulting estimate $\hat{\delta}$ maximizes the approximate likelihood of the r_i^*.

If the alternative approximation of McKay is used, then $(n_i / (n_i - 1))h(r_{n,i})$ is supposed to be distributed approximately gamma with expectation $h(R_i)$ and index $(n_i - 1) / 2$. Hence, $r_{n,i}^* = (n_i / (n_i - 1))h(r_{n,i})$ may be substituted in z_i and w_i in place of r_i^*.

4.3 Iglewicz and Myers' Approximation

According to Iglewicz and Myers' approximation, r_i is distributed approximately normal with mean R_i and variance $\left(\frac{R_i^2}{n_i}\right)\left(R_i^2 + \frac{1}{2}\right) = V_i(R_i)$ (taking $\phi = 1$). This distribution is not in the exponential family. However, quasi-likelihood estimates of the parameters of (4.1) are available from the iterative algorithm outlined above using these moments and the generalized linear model $\log R_i = \mathbf{x}_i'\delta$, with the r_i as the responses.

4.4 Likelihood-Ratio Tests

Let $\delta = (\delta_1', \delta_2')'$ be a p x 1 parameter vector associated with model (4.1), where δ_1 and δ_2 are of dimension q x 1 and (p - q) x 1, respectively. Using either McKay's or David's approximation, let $\hat{\delta}$ be the unrestricted maximum-likelihood estimate of δ and $\tilde{\delta}$ the restricted maximum-likelihood estimate of δ under the null hypothesis H_o: $\delta_2 = \mathbf{0}$. If $\hat{R}_i^* = h(\exp(\mathbf{x}_i'\hat{\delta}))$ and $\tilde{R}_i^* = h(\exp(\mathbf{x}_i'\tilde{\delta}))$ are the corresponding estimates of the R_i^*, i = 1, 2, …, k, then an asymptotic level-α test of H_o versus H_1: $\delta_2 \neq \mathbf{0}$ is to reject H_o if and only if $-\sum_{i=1}^{k}(n_i - 1)\log\left(\frac{\hat{R}_i^*}{\tilde{R}_i^*}\right) > \chi^2_{\alpha,p-q}$, where $\chi^2_{\alpha,p-q}$ is the $(1 - \alpha)$ quantile of a chi-square distribution with (p - q) degrees of freedom. Alternatively, if $\hat{R}_i = \exp(\mathbf{x}_i'\hat{\delta})$ and $\tilde{R}_i = \exp(\mathbf{x}_i'\tilde{\delta})$ are the corresponding quasi-likelihood estimates of the R_i using Iglewicz and Myers' approximation, then an asymptotic level-α test of H_o versus H_1 is to reject H_o if and only if

$$\sum_{i=1}^{k}\left(\sqrt{8}n_i r_i\left[\tan^{-1}\left(\sqrt{2}\hat{R}_i\right)-\tan^{-1}\left(\sqrt{2}\tilde{R}_i\right)\right]+n_i\log\left(\frac{\hat{R}_i^2\left(\tilde{R}_i^2+\frac{1}{2}\right)}{\tilde{R}_i^2\left(\hat{R}_i^2+\frac{1}{2}\right)}\right)\right)>\chi^2_{\alpha,p-q}.$$

Details of these tests are provided in Wilson (1998).

4.5 Confidence Intervals for Fitted Models

Once the significant interactions and main effects in a fitted factorial model have been determined, confidence intervals for estimated contrasts may be desired. For the multiplicative model (4.1), such contrasts estimate ratios of unknown population CVs rather than differences, as in normal-theory analysis of variance.

For simplicity, suppose that two population CVs, R_1 and R_2, are to be contrasted, and assume that the multiplicative model (4.1) has been fitted. Note that although a single subscript is used, these CVs may be associated with either main or simple effects of factors. In this context, the unknown ratio of R_1 and R_2 may be expressed as

$$\log\left(\frac{R_1}{R_2}\right) = \log R_1 - \log R_2 = \mathbf{x}_1'\delta - \mathbf{x}_2'\delta = \left(\mathbf{x}_1' - \mathbf{x}_2'\right)\delta = \mathbf{x}_{12}'\delta.$$

Once the maximum- and/or quasi-likelihood estimate of δ is obtained via one of the three approximations under consideration, an asymptotic 100(1 - α)% confidence interval for the log-ratio is then

$$\widehat{\log\left(\frac{R_1}{R_2}\right)} = \mathbf{x}_{12}'\hat{\boldsymbol{\delta}} \pm z_{\alpha/2}\sqrt{\mathbf{x}_{12}'\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{x}_{12}},$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, and where $\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}$ is the appropriate estimated asymptotic covariance matrix of $\hat{\delta}$. Denoting the lower and upper endpoints of this interval by $\hat{L}$ and $\hat{U}$, respectively, a corresponding 100(1 - α)% confidence interval for R_1 / R_2 is then given by $\left(\exp(\hat{L}), \exp(\hat{U})\right)$.

5. EXAMPLE

Ott (1993, pp. 916, 919) lists the observed pH levels of 2-mL vials of a drug product stored at each of two temperatures (30°C and 40°C) in two labs (#1 and #2). Twelve vials were examined from each temperature-lab combination. The data, along with the sample means, standard deviations, CVs, and Shapiro-Wilk statistics for testing normality are given in Table 1.

The objective in this example is to estimate a factorial model that describes how each factor influences the relative variability of the pH. Model estimation was performed using code written in PROC IML with PC SAS version 6.11.

The saturated model has the form

$$R_{ij} = R\exp(\alpha_i + \beta_j + (\alpha\beta)_{ij}), \ i = 1, 2, \ j = 1, 2,$$

where R is the overall population CV, $\exp(\alpha_i)$ is the effect of the i^{th} temperature, $\exp(\beta_j)$ is the effect of the j^{th} lab, and the terms $\exp((\alpha\beta)_{ij})$ describe the interaction between temperature and lab. The identifiability constraint $\sum_{i=1}^{2}\alpha_i = \sum_{j=1}^{2}\beta_j = \sum_{i=1}^{2}(\alpha\beta)_{ij} = \sum_{j=1}^{2}(\alpha\beta)_{ij} = 0$ was used. McKay's approximation was applied to fit the model. The corresponding test for interaction, based on one degree of freedom, is summarized in Table 2. Note that there is clearly no evidence of interaction, so that a reduced model with only main effects was considered.

Conditional likelihood-ratio χ^2 statistics for assessing the significance of temperature and lab, each based on one degree of freedom, also are given in Table 2. Apparently, temperature can be removed from the model. The estimated parameters of the resulting "lab" model are appended to the table. Based on this model, the estimated log-ratio and ratio of lab CVs (#1 to #2), irrespective of storage temperature, are 0.6350 and 1.8871, respectively, while the asymptotic 95% confidence intervals are (0.2168, 1.0532) and (1.2421, 2.8668). It appears that vials stored in lab #1 have a significantly higher relative variability than those stored in lab #2.

6. CONCLUSION

The modelling approach developed in this paper is significant because it expands the settings in which the normal population CV may be analyzed to include designed factorial experiments. In particular, the use of approximations of the distribution of the sample CV provides a context well suited to the application of the generalized linear model and its iterative algorithms for model estimation. When the CV is the population characteristic of interest, the approach is apparently superior to the modelling efforts associated with Taguchi because it incorporates estimable model and covariance structures for the observed sample CVs rather than use transformed CVs that are assumed to have constant variance. As a result, estimated model parameters are easily interpreted, tests of all effects in a fitted factorial model are available, and asymptotic confidence intervals for ratios of contrasted population CVs are readily obtained.

REFERENCES

Ahmed, S. E. (1994), "Improved Estimation of the Coefficient of Variation," *Journal of Applied Statistics*, 21(6), pp. 565-573.

Bennett, B. M. (1976), "On an Approximate Test for Homogeneity of Coefficents of Variation," *Contributions to Applied Statistics*, ed. W. J. Ziegler, pp. 169-171.

Box, G. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations," *Technometrics*, 30(1), pp. 1-17.

David, F. N. (1949), "Note on the Application of Fisher's k-statistics," *Biometrika*, 36, pp. 383-393.

Doornbos, R. and Dijkstra, J. B. (1983), "A Multi Sample Test for the Equality of Coefficients of Variation in Normal Populations," *Communications in Statistics - Simulation and Computation*, 12(2), pp. 147-158.

Feltz, J. F. and Miller, G. E. (1996), "An Asymptotic Test for the Equality of Coefficients of Variation from k Populations," *Statistics in Medicine*, 15, pp. 647-658.

Gupta, R. C. and Ma, S. (1996), "Testing the Equality of Coefficients of Variation in k Normal Populations," *Communications in Statistics - Theory and Methods*, 25(1), pp. 115-132.

Hurlimann, W. (1995), "A Uniform Approximation to the Sampling Distribution of the Coefficient of Variation," *Statistics and Probability Letters*, 24, pp. 263-268.

Iglewicz, B. and Myers, R. H. (1970), "Comparisons of Approximations to the Percentage Points of the Sample Coefficient of Variation," *Technometrics*, 12(1), pp. 166-169.

Koopmans, L. H., Owen, D. B., and Rosenblatt, J. I. (1964), "Confidence Intervals for the Coefficient of Variation for the Normal and Log-Normal Distributions," *Biometrika*, 51, pp. 25-32.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley and Sons, Inc.

McCullagh, P. (1983), "Quasi-Likelihood Functions," *The Annals of Statistics*, 11(1), pp. 59-67.

McKay, A. T. (1932), "Distribution of the Coefficient of Variation and the Extended 't' Distribution," *Journal of the Royal Statistical Society*, 95, pp. 695-698.

Ott, R. L. (1993), *An Introduction to Statistical Methods and Data Analysis, 4th Ed.*, Belmont, CA: Duxbury Press.

Payton, M. E. (1997), "Confidence Intervals for the Coefficient of Variation," *Applied Statistics in Agriculture*, pp. 82-89.

Rao, K. A. and Vidya, R. (1992), "On the Performance of a Test for Coefficients of Variation," *Calcutta Statistical Association Bulletin*, 42, March & June, pp. 87-95.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley and Sons, Inc.

Shafer, N. J. and Sullivan, J. A. (1986), "A Simulation Study of a Test for the Equality of the Coefficients of Variaton," *Communications in Statistics - Simulation*, 15(3), pp. 681-695.

Taguchi, G. (1992), *Taguchi Methods: Research and Development*, Dearborn, MI: American Supplier Institute.

Vangel, M. G. (1996), "Confidence Intervals for a Normal Coefficient of Variation," *The American Statistician*, 15(1), pp. 21-26.

Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61(3), pp. 439-447.

Williams, R. R. (1991), "Fundamental Limitations on the Use and Comparison of Signal-to-Noise Ratios," *Analytical Chemistry*, 63(15), pp. 1638-1643.

Wilson, C.A. (1998), "An Approach to Modelling the Coefficient of Variation in Factorial Experiments," Unpublished Ph.D. Dissertation, Department of Statistics, Oklahoma State University.

Table 1

Observed means, standard deviations, and CVs of pH levels by temperature and lab

Temperature	Lab	pH data	$\overline{x}$	s_n	r_n
30°C (W = 0.905, p = 0.173)	#1	3.45, 3.48, 3.50, 3.55 3.56, 3.57, 3.59, 3.60 3.60, 3.61, 3.74, 3.81	3.5883	0.0975	0.0272
30°C (W = 0.921, p = 0.277)	#2	3.70, 3.74, 3.75, 3.76 3.77, 3.80, 3.80, 3.84 3.87, 3.90, 3.90, 3.90	3.8108	0.0669	0.0176
40°C (W = 0.931, p = 0.367)	#1	3.29, 3.32, 3.38, 3.39 3.45, 3.51, 3.59, 3.60 3.61, 3.63, 3.65, 3.71	3.5108	0.1348	0.0384
40°C (W = 0.906, p = 0.179)	#2	3.60, 3.64, 3.68, 3.70 3.70, 3.70, 3.70, 3.75 3.80, 3.80, 3.80, 3.81	3.7233	0.0659	0.0177

Note: Values given in parentheses are the Shapiro-Wilk statistics and p-values for testing the null hypotheses that the samples were drawn from normal distributions.

Table 2

Likelihood-ratio test results

Effect	χ^2	df	P-value
Temperature x Lab	0.621	1	0.431
Temperature | Lab	0.661	1	0.416
Lab | Temperature	7.437	1	0.006
Lab	8.323	1	0.004

Note: Fitted CV model gave parameter estimates of $\log \hat{R} = -3.678$ ($\pm$ 0.107) and $\hat{\beta}_1 = 0.318$ ($\pm$ 0.107).