Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 86 (2007) 82–85
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

To triplicate or not to triplicate?

Julio M. Singer [a,*], Antonio C. Pedroso-de-Lima [a], Nelson I. Tanaka [a], Verónica A. González-López [b]

[a] *Departamento de Estatística, Universidade de São Paulo, Brazil*
[b] *Departamento de Estatística, Universidade Estadual de Campinas, Brazil*

Received 25 May 2006; received in revised form 22 August 2006; accepted 23 August 2006
Available online 29 September 2006

Abstract

A common practice in scientific experimentation in areas such as Medicine, Pharmacy, Nutrition, among others, is to measure each sample unit three times (in triplicate) or more generally, m times (in m–plicate) and take the average of such measurements as the response variable. This is generally done to improve the precision of model parameter estimates. When the objective is to estimate the population mean, we use a random effects model to show that the efficiency of working with m-plicates is related to the magnitude of the intraclass correlation coefficient, which essentially measures the contribution of the variance between sample units to the total variance. We show that above certain values of this parameter, the use of m-plicates does not bring significant improvement (say, of 10% or more) to the precision of the estimates. Additionally, taking the costs of sampling units and making measurements into account, we compare sampling schemes with and without m-plicates designed to obtain fixed width confidence intervals for the mean. We illustrate the results through a practical example.

Keywords: Cost based optimization; Intraclass correlation; Precision of estimates

1. Introduction

In many experimental studies, observations are obtained in triplicate (or more generally, in m-plicate) and their average is taken as the response variable. This is a common practice in areas like Medicine, Pharmacy, Nutrition etc., as evidenced in Vaughan and Oram [1], Paquet et al. [2] or Viyoch et al. [3], among others. In many cases, like in Nutter et al. [4] or Thuresson et al. [5], the objective is to evaluate intraobserver variability. Also, there are instances where the procedure is adopted simply by tradition. We focus on situations where the use of m-plicates is intended to improve the precision of model parameter estimates, in particular, the mean.

A practical example involves the estimation of the average amount of oil contained in lemon juice, an important feature for the decision about the destination of this commodity (plain consumption, as a cosmetic ingredient etc) and hence, about its trade price. Each of 60 samples was obtained from a batch of lemon juice and divided into three portions (haphazardly labeled A, B and C) each of which was analyzed with respect to the amount of oil (kg/ton). The data for the 60 triplicates are displayed in Table 1.

We are interested in evaluating the effect of using triplicates in the precision of the estimate of the mean amount of oil per ton of lemon juice.

Assuming a Gaussian model we obtained four 95% confidence intervals for the mean: the first three ones are based on the observations (A, B or C) considered separately; the fourth is based on the average of the three within sample units observations. The results are presented in Table 2 and show that the precision of the four intervals is practically the same, suggesting that the use of triplicates is unnecessary, in the sense that, a single observation per sample unit would generate confidence intervals with similar widths and consequently reduce costs.

A similar problem was considered in Fagan et al. [6] to evaluate the need for triplicate blood pressure measurements. Considering Analysis of Variance for repeated measurements and correlation methods, the authors conclude that averaging the triplicate within sample units observations or simply using the first of the three observations produces similar results. Shapiro et al. [7], on the other hand, argue that more replicates are better,

* Corresponding author.
E-mail address: jmsinger@ime.usp.br (J.M. Singer).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2006.08.008

Table 1
Amount of oil in lemon juice (kg of oil/ton of juice)

Sample	A	B	C	Sample	A	B	C	Sample	A	B	C
1	5.29	5.10	5.13	21	5.66	5.64	5.46	41	4.90	4.75	4.84
2	5.34	5.34	5.27	22	5.62	5.49	5.73	42	4.88	4.57	4.54
3	5.20	5.07	5.08	23	5.36	5.33	5.46	43	4.80	4.82	4.94
4	5.43	5.38	5.36	24	4.91	5.01	4.86	44	5.29	5.29	5.10
5	5.18	5.03	5.02	25	5.28	5.35	5.14	45	4.53	4.66	4.63
6	5.33	5.07	5.07	26	5.02	4.80	4.64	46	4.39	4.49	4.39
7	5.16	5.40	5.23	27	5.57	5.54	5.29	47	4.50	4.51	4.52
8	4.91	5.10	4.84	28	5.09	5.22	4.95	48	4.82	4.80	4.66
9	5.07	5.01	4.87	29	5.58	5.45	5.32	49	5.06	4.96	4.94
10	4.85	4.76	4.54	30	5.04	4.90	4.94	50	5.20	4.97	5.11
11	5.31	5.42	5.52	31	5.79	5.65	5.58	51	5.63	5.75	5.63
12	5.12	5.40	5.27	32	5.46	5.38	5.36	52	5.38	5.51	5.14
13	5.29	5.47	5.13	33	5.21	5.20	5.07	53	5.37	5.06	5.13
14	5.04	5.09	4.98	34	4.84	4.98	4.91	54	5.06	5.20	5.07
15	5.11	5.11	5.11	35	5.27	5.11	5.25	55	5.15	5.32	4.99
16	4.96	5.07	4.94	36	5.06	5.08	4.89	56	4.74	4.74	4.64
17	5.36	5.06	5.10	37	5.10	5.24	5.05	57	4.48	4.4	4.37
18	5.36	5.40	5.33	38	5.32	5.51	5.22	58	4.26	4.12	4.37
19	5.39	5.13	5.34	39	4.80	4.70	4.58	59	4.46	4.37	4.62
20	5.49	5.60	5.28	40	5.18	4.83	4.80	60	5.20	4.93	5.07

since blood pressure varies considerably from beat to beat. They seem to misunderstand the role of within sample unit variability in the evaluation of total variability as pointed by Fagan et al., who show that the magnitude of intraunit short term measurements is not as large as to produce more precise measurements. We try to clarify such issues in a broader context by attacking the problem from two perspectives. The first consists of knowing under what circumstances the precision of the estimate of the mean of a Gaussian distribution based on a sample of n units is affected by considering m within sample units measurements; this is the object of Section 2. The second, refers to the choice between two experimental designs (with and without repeated within sample unit measurements) when the costs of obtaining sample units and performing measurements are taken into account; this is discussed in Section 3. The conclusion with a brief discussion is presented in Section 4.

2. Using *m*-plicates to reduce the width of confidence intervals

Assuming a Gaussian distribution, the data (collected in m-plicates) can be represented by the random effects model

$$y_{ij} = \mu + a_i + e_{ij}, \tag{1}$$

where $a_i \sim N(0,\sigma_a^2)$ and $e_{ij} \sim N\ (0,\ \sigma_e^2)$, are independent, $i=1,...,n$ and $j=1,...,m$. In the example shown in Table 1, we have $n=60$ and $m=3$. This model induces a dependence in the within sample units observations that may be quantified by the intraclass correlation coefficient,

$$\rho = \sigma_a^2/(\sigma_a^2 + \sigma_e^2). \tag{2}$$

When the within sample units observations are independent, it follows that $\rho=0$; otherwise, when the dependence between the within sample units observations is large, we have ρ close to 1.

Under Eq. (1), a 95% confidence interval for μ based on just one of the within sample units observations (the first one, for instance) is given by

$$\bar{y}_{+1} \pm 1.96\ \sigma/\sqrt{n}, \tag{3}$$

where $\bar{y}_{+1} = n^{-1}\sum_{i=1}^{n} y_{i1}$ and $\hat{\sigma}^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_{i1}-\bar{y}_{+1})^2$ is an estimate of

$$\sigma^2 = \sigma_a^2 + \sigma_e^2. \tag{4}$$

Alternatively, under the same model (1), a 95% confidence interval for μ based on the average of the m within sample units observations is

$$\bar{y}_{++} \pm\ 1.96\sqrt{\frac{\hat{\sigma}_a^2 + \hat{\sigma}_e^2/m}{n}}, \tag{5}$$

where $\bar{y}_{++} = n^{-1}\sum_{i=1}^{n}\ \bar{y}_{i+}$, $\bar{y}_{i+} = m^{-1}\sum_{j=1}^{m} y_{ij}$, $\hat{\sigma}_e^2 = [n(m-1)]^{-1}$ $\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i+})^2$ and $\hat{\sigma}_a^2 + \hat{\sigma}_e^2/m = (n-1)^{-1}\sum_{i=1}^{n}(\bar{y}_{i+}-\bar{y}_{++})^2$.

Using Eqs. (2) and (4), we can show that the width of the confidence interval Eq. (5) is

$$2 \times \frac{1.96}{\sqrt{n}}\sqrt{\hat{\sigma}_a^2 + \hat{\sigma}_e^2/m} = 2 \times 1.96\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\hat{\rho} + (1-\hat{\rho})/m}. \tag{6}$$

Table 2
95% confidence intervals for the mean amount of oil in lemon juice

Data	Lower bound	Upper bound	Width
First observations (A)	5.04	5.21	0.17
Second observations (B)	5.00	5.18	0.18
Third observations (C)	4.94	5.11	0.17
Average	5.00	5.16	0.16

Table 3
Percent reduction (for $m=3$) in the width of the confidence interval based on a single within sample unit observation

ρ	0.05	0.10	0.25	0.50	0.75	0.90
Reduction	39%	37%	29%	18%	8%	3%

Since $\rho<1$, when $m\geq 2$ the width of the confidence interval Eq. (5) is always smaller than the width of Eq. (3) by the factor $\sqrt{\hat{\rho}+(1-\hat{\rho})/m}$.

Given the possibility of obtaining m-plicate measurements, the problem is to know for which values of ρ the width of the confidence interval based on their average reduces to $(1-r)$ 100%, $(0<r<1)$ or less of the length of the interval based on a single within sample unit observation. In this direction, let $\sqrt{\rho+(1-\rho)/m}=1(-r)$, and note that $\sqrt{\rho+(1-\rho)/m}$ is an increasing function of ρ so that the desired result follows when

$$\rho<[m(1-r)^2-1]/(m-1).$$

The maximum reduction of the confidence interval width based on a single within sample unit observation is $r=1-\sqrt{1/m}$ and occurs when $\rho=0$. For the lemon juice data, $m=3$ and the maximum reduction is $r=1-\sqrt{1/3}=0.42$. In Table 3 we present percent reductions in the confidence interval widths for different values of ρ.

For the data in Table 1, $\hat{\sigma}_a^2=0.0992$, $\hat{\sigma}_e^2=0.0131$ and $\hat{\rho}=0.8830$ which implies a reduction of only $100r\%=4\%$ in the width of the confidence interval. Thus, for this example, having a simple observation per sample would seem more adequate, unless the cost of making an observation is negligible. In cases where the intraclass correlation coefficient (ρ) is small (i.e., $\rho=10\%$), we could have considerable gains (i.e., 37%) and working with triplicates could be worthwhile.

3. Using *m*-plicates to reduce costs

In this section we focus our attention on the estimation of the mean of a Gaussian distribution with a fixed precision taking into account that the cost of sampling a unit is A and the cost of obtaining a within sample units measurement is R. More precisely, we want to know which of the following experimental designs has a smaller cost:

- Obtain n_s independent sample units and perform a single measurement in each.
- Obtain $n_c<n_s$ independent sample units and perform $m\geq 2$ measurements in each.

Under model (1), the sample size required to generate a 95% confidence interval based on a single within sample units observation with width equal to d is

$$n_s\left[2\times 1.96\sqrt{\sigma^2}/d\right]^2 \tag{7}$$

and the corresponding cost is $C_s=n_s\ (A+R)$. To achieve confidence intervals with the same width under both experimental designs, first set

$$2\times 1.96\sqrt{\frac{\sigma_a^2+\sigma_e^2/m}{n_c}}=d.$$

Then, let

$$\sqrt{\frac{\sigma^2}{n_s}}=\sqrt{\frac{\sigma_a^2+\sigma_e^2/m}{n_c}}$$

so that

$$\sqrt{n_c/n_s}=\sqrt{\rho+(1-\rho)/m}. \tag{8}$$

Since $n_c\leq n_s$, we can evaluate m for $n_c=2, 3,\ldots, n_s-1$ and compute the cost $C_c=n_c\ (A+mR)$. The solution is the sample size corresponding to $\min(C_c, C_s)$.

Consider, for example that the cost of sampling a unit is $A=10$, the cost of obtaining a within sample units measurement is $R=2$ and that the error standard deviation is $\sigma=3$. Furthermore, assume that the desired length of the 95% confidence interval is $\sigma/2=1.5$. From Eq. (7), it follows that when no replicates are considered, the required sample size is $n_s=62$ and that the corresponding cost is 744. When $\rho=0.2$ the minimum cost (352) is attained by taking $n_c=16$ sample units measured 6 times (6-plicates). When $\rho=0.5$ the minimum cost (658) is reached by considering $n_c=47$ sample units measured in duplicate. If $\rho=0.8$ no cost reduction is possible.

4. Discussion

Although the use of m-plicates is considered in many experimental studies, a careful analysis shows that the benefits of such practice in terms of improving the precision of estimates may not be worthwhile without an analysis of its benefits. We consider two approaches to this problem and in either case, we conclude that the knowledge of the magnitude of the intraclass correlation coefficient (ρ) is fundamental to reach a decision. In general, inference about such coefficient should be based on a small sized pilot sample that excludes classical inference techniques, since under such circumstances, both confidence intervals and hypotheses tests about ρ are based on asymptotic arguments (see Bickel and Doksum [8], for example). The use of Bayesian techniques for such purposes may be more attractive and is under investigation by the authors.

Since the proposed model assumes that the m-plicates are exchangeable it may not be applicable in situations where the measurements are taken sequentially such as those usually known as repeated measurements. In such cases, a different covariance structure must be considered. The reader is referred to Diggle et al. [9], for example.

Although, cost considerations look interesting for the decision about which experimental design to choose, in practice, not only obtaining an analytical solution, but also determining exact costs can be difficult. With this in mind, we proposed an

algorithm that may be easily implemented in a spreadsheet and may help in the decision making process, since the user may simulate different scenarios (with different relationships between costs). An example of such spreadsheet may be obtained from www.ime.usp.br/~jmsinger/triplicate.

Acknowledgments

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for providing partial financial support during the realization of this work.

References

[1] A.M. Vaughan, J.F. Oram, ABCA1 redistributes membrane cholesterol independent of apolipoprotein interactions, Journal of Lipid Research 44 (2003) 1373–1380.

[2] J. Paquet, C. Lacroix, P. Audet, J. Thibault, Electrical conductivity as a tool for analysing fermentation processes for production of cheese starters, International Dairy Journal 10 (2000) 391–399.

[3] J. Viyoch, P. Patcharaworakulchai, R. Songmek, V. Pimsan, S. Wittaya-Areekul, Formulation and development of a patch containing tamarind fruit extract by using the blended chitosan-starch as a rate-controlling matrix, International Journal of Cosmetic Science 25 (2003) 113.

[4] F.W. Nutter Jr., M.L. Gleason, J.H. Jenco, N.C. Christians, Assessing the accuracy, intrarater repeatability, and inter-rater reliability of disease assessment systems, Phytopathology 83 (1993) 806–812.

[5] M. Thuresson, B. Ang, J. Linder, K. Harms-Ringdahl, Intrarater reliability of electromyographic recordings and subjective evaluation of neck muscle fatigue among helicopter pilots, Journal of Electromyography and Kinesiology 15 (2005) 323–331.

[6] T.C. Fagan, K.A. Conrad, P.V. Mayshar, M.J. Mackie, R.M. Hagaman, Single versus triplicate measurements of blood pressure and heart rate, Hypertension 11 (1988) 282–284.

[7] D. Shapiro, L.D. Jamner, I.B. Goldstein, D. Guthrie, Single versus triplicate measurements: a commentary on Fagan et al. (with response), Hypertension 16 (1990) 103–106.

[8] P.J. Bickel, K.A. Doksum, Mathematical Statistics, Holden-Day, San Francisco, 1977.

[9] P.J. Diggle, P. Heagerty, K.Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, 2nd ed.Oxford University Press, Oxford, 2002.