Title
Laboratory Quality Control

Author
Rocke, David

Publication Date
1998

Peer reviewed

ENCYCLOPEDIA OF BIOSTATISTICS

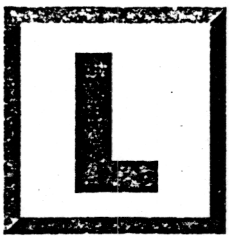
Volume 3
H–MEA

Editors-in-Chief

PETER ARMITAGE
University of Oxford, UK

THEODORE COLTON
Boston University, USA

John Wiley & Sons
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto



L-Estimator *see* Robustness

L_1 Regression *see* Least Squares

L_p Regression *see* Robust Regression

Laboratory Quality Control

Quality control, or more broadly quality assurance, is an essential part of the conduct of an analytic laboratory. This is especially critical for clinical laboratories since errors in measurement can lead to inappropriate treatment of patients. A full treatment of the many management, planning, record keeping, and audit procedures is beyond the scope of this article (see Garfield [1] for these concerns); we concentrate on a few critical issues in statistical quality control as they impact analytical laboratories.

Estimation of Precision

An important issue in an analytic laboratory involves the attachment of **standard errors** to measured values. Although every effort should be made to avoid bias, it may be difficult to estimate the remaining bias, so standard errors are usually based on the precision of measurements; that is, the standard deviation (sd) of repeat measurements made under identical conditions. This is necessarily only a lower bound on the true error standard deviation.

It will hardly ever be the case that the error standard deviation is constant, independent of the level measured. For large concentrations of the analyte, the standard deviation will often be approximately a multiple of the concentration (more generally, this may be a power function). If small concentrations (near the limit of detection) are not of interest, then a linear equation describing the relationship between the precision and the concentration can be estimated from a series of repeat measurements at several concentrations, by regressing the standard deviation of the repeats on the mean of the repeats. This can often be done during the process of producing a **calibration** curve for the instrument.

If values near the limit of detection are of interest, then a more complex method is called for. Rocke & Lorenzato [10] present a model in which an additive error and a multiplicative error are both present, which allows for realistic behavior at both high and low concentrations. This model can also be estimated during the calibration process.

Control Charts

There are many problems, systematic and sporadic, that can interfere with the accurate determination of values in the analytic laboratory. Reagents may lose potency, temperature or humidity may affect the

results, and operators may differ in their technique. There are many methods of statistical quality control that can be used to detect these problems; detailed descriptions may be found in Wadsworth et al. [11] and Montgomery [5]. Two of the most important will be described here. These are Shewhart charts, also known as $\bar{X}$ and R charts, which are used to detect general departure from a state of statistical control, and CUSUM charts, which are used to detect a shift in the mean.

The "state of statistical control" referred to above is, in the context of laboratory errors, one in which measurement errors have zero mean and a standard deviation that does not change with time (although it may differ with the concentration of the analyte), and in which successive measurement errors are not correlated.

Shewhart Charts

Periodically, a small group of repeat measurements is taken on a standard solution. At time t, k replicates are taken, denoted x_{it}, where $1 \leq i \leq k$. The mean $\bar{X}_t$ and the **range** R_t are computed. If the process is in a state of statistical control, and under the additional assumption of normality, each value of the range is an estimate of a multiple of the standard deviation, the value of the multiplier being given in standard tables (see, for example, Wadsworth et al. [11] or Montgomery [5]). After 10 or 20 groups have been accumulated, the mean of the ranges $\bar{R}$, when multiplied by a standardizing constant, is an estimate of the process standard deviation. For example, if $k = 3$, then $\bar{R}/1.693$ is an estimate $\hat{\sigma}$ of the process sd σ.

The group means are then plotted on a chart which has a centerline equal either to the known value for the reference solution or to the average of the group means, and which has two reference lines above and below the centerline at a distance of $\pm 3\hat{\sigma}/\sqrt{k}$ (this whole quantity can be computed as a multiple A_2 of $\bar{R}$). Since the estimate of the process standard deviation is based on the variability of repeats, any factor such as temperature changes, or operator technique, that varies from group to group but not within a group will generate extra variability. If there are no such extra factors (called special causes), and all the variability is caused by the repeat variation (common cause), then the group means will lie outside the two reference lines (called control limits) only very rarely. If, on the other hand, some extra factor is an important cause of variability, then the group mean will more frequently lie outside the control limits. When this event occurs (a signal, in quality control terminology), the cause of the extra variability should be sought, and eliminated or reduced.

Similarly, the group ranges are plotted on a chart that has a centerline at $\bar{R}$ and control limits at multiples D_3 and D_4 of $\bar{R}$, calculated to be three standard errors above and below the centerline. Any values outside these control limits also indicate a probable departure from a state of statistical control.

Over time, this attention to Shewhart charts, and subsequent investigation of problems, should place the measurement process in a state of statistical control. Continued attention is required to detect later departures from this desirable condition.

CUSUM Charts

Shewhart charts are designed to detect many kinds of departure from a state of statistical control, and are therefore not particularly sensitive to any given one of these. A frequent concern is a shift in the process mean, which would involve, in the laboratory context, the development of a bias in the measurements. Cumulative sum (CUSUM) charts are specifically designed to detect this type of departure, which is often of great concern in the analytic laboratory. More details on this methodology may be found in Wadsworth et al. [11] or in Hawkins & Olwell [3]. Briefly, CUSUM charts achieve their superior detection ability by adding up successive deviations of the group mean $\bar{X}$ from the correct value μ. This allows a small difference to accumulate until a strong signal can be observed; however, it would also allow extremely tiny signals to eventually manifest themselves even if the difference were of no practical importance. The user must therefore specify a critical shift Δ that may be of importance to detect. One then defines the upper CUSUM S_t and the lower CUSUM T_t recursively by

$$S_t = \max(0, S_{t-1} + \bar{X}_t - F),$$

$$T_t = \min(0, T_{t-1} + \bar{X}_t + F),$$

where F is usually taken to be approximately $\Delta/2$. If the mean has shifted by as much as Δ, then the upper (respectively lower) CUSUM will exhibit a trend that will eventually pass any fixed control limits.

Control limits for the upper and lower CUSUMs are derivable from somewhat complicated numerical calculations and must be found in tables in one of the cited references. They are designed so that if the process is in control, the renewal process that would then describe the path of the CUSUM variables would have an expected first passage time (called the *average run length*) that would be sufficiently large not to induce many false alarms. In the form in which we have written the CUSUM, the control limits will be at multiples of the process standard deviation. The originally developed form of the CUSUM chart uses a different, but entirely equivalent, method of determining control limits, called a V-mask. This alternative formulation is described in Wadsworth et al. [11] and Hawkins & Olwell [3].

Proficiency Testing

Proficiency testing is used internally by analytic laboratories to evaluate their own performance and externally to develop new analytic methods or to validate the performance of laboratories. This will usually involve the submission of spiked samples to the analytic process which are blind (the amount present is not known to the analyst) or double-blind (the analyst does not know that the sample is a check sample and not a routine sample). Detailed discussion of these issues may be found Garfield [1] and Youden & Steiner [12].

Intralaboratory Studies

These studies are used by laboratories to check themselves or to improve operations. Blind or double-blind samples can be routinely run and used as the input for control charts or for examination of specific results. Another useful method is Youden's ruggedness testing [12], in which a designed experiment is used deliberately to vary the conditions of the analysis in order to find out what factors influence the variability of the results. An example of this type of study for immunoassays may be found in Jones et al. [4].

Interlaboratory Studies

Interlaboratory studies are an essential component in the development of analytic methods. A proposed method is described to a series of laboratories, each of which receive a series of samples to analyze. The variability of a measured result then can be partitioned into within-laboratory variance (repeatability or precision) and between-laboratory variance. The total variability, which is the sum of the within- and between-laboratory variances, is called the reproducibility and is a measure of accuracy.

Ideally, the between-laboratory variance would be small, but in practice it is often considerably larger than the within-laboratory variance. This may be due to inadequately described methods, or to the influence of identifiable factors that can be determined with ruggedness testing and controlled in a revised procedure. Poorly performing laboratories may also be identified in an interlaboratory study; Youden & Steiner [12] gives a rank test for this purpose.

Outliers

Outliers can cause a significant disruption in quality control procedures, as well as inaccurate measurement values. Especially, outliers in the initial samples used to determine the control limits for Shewhart charts or to estimate the process standard deviation for CUSUM charts can reduce the effectiveness of these tools. Outliers can also seriously distort the analysis of an interlaboratory study. **Robust** procedures are available for standard Shewhart charts (see Rocke [7] for the technical details and Rocke [9] for practical implementation), CUSUM charts [2], and interlaboratory studies [6, 8]. If outliers are frequent in the check samples used to produce the control charts, it would be essential to discover and eliminate the source of the outliers, since detection of outliers in routine samples would be difficult.

References

[1] Garfield, F.M. (1991). *Quality Assurance Principles for Analytical Laboratories*, 2nd Ed. Association of Official Analytical Chemists, Washington.

[2] Hawkins, D.M. (1993). Robustification of cumulative sum charts by winsorization, *Journal of Quality Technology* **25**, 248–261.

[3] Hawkins, D.M. & Olwell, D.H. (1997). *Cumulative Sum Charts and Charting*. Springer-Verlag, New York.

[4] Jones, G. Wortberg, M. Kreissig, S.B., Hammock, B.D. & Rocke, D.M. (1995). Sources of experimental variation in calibration curves for enzyme-linked immunosorbent assay, *Analytica Chimica Acta* **313**, 197–207.

[5] Montgomery, D.C. (1996). *Introduction to Statistical Quality Control*, 3rd Ed. Wiley, New York.
[6] Rocke, D.M. (1983). Robust statistical analysis of interlaboratory studies. *Biometrika* **70**, 421-431.
[7] Rocke, D.M. (1989). Robust control charts. *Technometrics* **31**, 173-184.
[8] Rocke, D.M. (1991). Robustness and balance in the mixed model. *Biometrics* **47**, 303-309.
[9] Rocke, D.M. (1992). $\bar{X}_Q$ and R_Q charts: robust control charts. *Statistician* **41**, 97-104.
[10] Rocke, D.M. & Lorenzato, S. (1995). A two-component model for measurement error in analytical chemistry. *Technometrics* **37**, 176-184.
[11] Wadsworth, H.M., Stephens, K.S. & Godfrey, A.B. (1986). *Modern Methods of Quality Control and Improvement*. Wiley, New York.
[12] Youden, W.J. & Steiner, E.H., (1975). *Statistical Manual of the Association of Official Analytical Chemists*. Association of Official Analytical Chemists, Washington.

DAVID M. ROCKE

Lack-of-Fit Sum of Squares *see* Goodness of Fit

Lagged Cumulative Exposure *see* Occupational Health and Medicine

Lagged Dependent Variable

In **longitudinal** studies, several observations are taken from each individual at different time points. Often, an observation depends on previous observations; for example, in a **crossover** clinical trial, observations in one period may depend on the observations in the previous periods. A simple model for this scenario might include a lag-1 dependent variable as an explanatory variable [2]:

$$y_{it} = \gamma y_{i,t-} \quad + \mathbf{x}_{it}\boldsymbol{\beta} + u_i + e_{it} \qquad (1)$$

where y_{it} is the observation from subject i in period t, $\mathbf{x}_{it}$ is a vector of **covariates**, u_i is a subject effect, and e_{it} is an error term. This model can be extended to include multiple lagged variables by replacing $\gamma y_{i,j-1}$ by $\sum_{l=1}^{p} \gamma_l y_{i,t-l}$ in (1). Model (1) is different from a **serially correlated** model with the same covariates. In the latter, $y_{i,t}$ depends on $\mathbf{x}_{it}$ only (not $y_{i,t-1}$), while in the former it depends on all $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}$ [4].

Statistical inference based on model (1) includes model fitting, **model checking** and **hypothesis tests**. In biostatistics, the number of subjects is often large, but the number of observations from each subject is small. In this situation we should be careful when using the asymptotic properties of the estimated parameters. For n subjects and times $1, 2, \ldots, T$, and conditional on u_i, the log **likelihood** function of this model can be written as

$$l(\boldsymbol{\beta}, \gamma, \mathbf{u}) = \sum_{i=1}^{n}\sum_{j=1}^{T} \log[p(y_{it}|y_{i,t-1}, \boldsymbol{\beta}, \gamma, u_i)]. \quad (2)$$

When there are no subject effects ($u_i = 0$), this model can be fitted easily using the lagged dependent variables as covariates [3]. When u_i is fixed and $u_i \neq 0$ the **maximum likelihood** estimates (mle) of γ and $\boldsymbol{\beta}$ are not **consistent** for fixed T when the total sample size $n \to \infty$ [2]. To obtain consistent estimates, the **instrumental variable** procedure can be used either for fixed or random u_i. To illustrate how this procedure works we write model (1) as

$$y_{it} - y_{i,t\cdot} \quad = \gamma(y_{i,t-1} - y_{i,t-2}) + \boldsymbol{\beta}(\mathbf{x}_{it} - \mathbf{x}_{i\ \ 1}) + e_{it} - e_{i,t} \qquad (3)$$

Directly using $(y_{i,t-1} - y_{i,t-2})$ as a covariate may lead to inconsistency, since it and $e_{it} - e_{i,t-1}$ are correlated. However, $(y_{i,t-2} - y_{i,t-3})$ or $y_{i,t-2}$ is independent of $e_{it} - e_{i,t-1}$ and can be used as an instrumental variable. When assuming $u_i \sim \mathrm{N}(0, \sigma_u^2)$ the log likelihood function is more complicated than (2), but the mle can be obtained by the Newton-Raphson method (*see* **Optimization and Nonlinear Equations**). In this case the mle is consistent for fixed T and $n \to \infty$.

Model (1) can be extended to include discrete outcomes. One approach is to discretize y_{ij} by letting $y_{ij}^* = 1$ if $y_{ij} > 0$ and $y_{ij}^* = 0$ otherwise. This approach leads to the autoregressive probit model [1]. A more general approach is to use (1) as the linear predictor in a **generalized linear model**, and a wide