Journal of the Rubber Research Institute of Sri Lanka, (1999) **82**, 31-46

EVALUATION OF ISO RECOMMENDATION ON ANALYSIS OF PRECISION EXPERIMENTS FOR STATISTICAL QUALITY CONTROL

Wasana Wijesuriya and R O Thattil[1]

(Accepted 1 September 2000)

ABSTRACT

Application of appropriate statistical methods is an important task in any quality control process as various statistical approaches such as; sampling, experimentation and process control are involved in quality control processes. This has become an essential practice in many areas since the introduction of ISO quality assurance scheme. This article focuses on precision experiments where several laboratories are compared for a standard measurement method. The whole process of analysis and interpretation suggested by ISO is evaluated and documented in a sequential order. In addition to the ISO recommendation, additional graphical methods are suggested for better judgement. The statisticians role as a member of the panel is also presented at each step of the quality control process.

Key words: ISO recommendation, precision experiments, quality control

INTRODUCTION

Precision experiments are carried out to compare different laboratories against a standard measurement method. Precision is the general term for variability between repeated measurements. It refers to the closeness of test results in an experiment.

For many practical reasons, the two conditions of precision; namely, repeatability and reproducibility, have been found necessary and useful for describing the variability of a measurement method. The factors, which can cause variability in test results, are; a) the operator, b) the equipment used, c) calibration of the equipment, d) environmental conditions, and e) time elapsed between measurements. Under repeatability conditions, all the above factors are considered constants and do not contribute to variability, while under reproducibility conditions they vary and do contribute to the variability of test results. Thus, repeatability and the reproducibility

[1] Department of Crop Science, Faculty of Agriculture, University of Peradeniya

are the two extremes of precision, the first describing the minimum and the second the maximum variability in results. Other intermediate conditions between these two extreme conditions of precision are also likely, when one or more of above mentioned factors, a) to e) are allowed to vary.

In general, the need to consider "precision" arises because tests performed on presumably identical materials in presumably identical circumstances do not produce identical results. This is due to inherent variability (random error) in every measurement procedure. In practical interpretation of measurement data, this variability has to be taken into account. The statistical procedures should be able to test whether the deviation between the measurement and the specified value is merely the random error component. In which case, a real deviation from such a specified value has to be established. Similarly, comparing test results from two batches of material will not indicate a fundamental quality difference if the observed difference between them can be attributed to the inherent variation in the measurement procedure.

A panel of experts in the specified profession with a statistical expert usually carries out precision experiments. This article comprises of appropriate guidelines on statistical analysis and interpretation of precision experiments and, also presents the responsibilities of the statistician as a member of the panel at each step of the quality control process.

METHODOLOGY

Data

The sample data set comprised of test results of iron content in soil measured in parts per million (ppm) carried out in 6 laboratories with 4 different levels and replicated 6 times (Appendix 1) [Source : Sample data set in Genstat for Windows 4.0.].

Statistical methods and interpretation

The principal steps in statistical analysis involved in precision experiments are documented in this section together with decisions to be taken by the statistician depending on the observations and conclusions of different steps in the process of statistical analysis.

Step 1: Tabulation of test results

A typical layout for collecting test results is given in Appendix 1(Form A), arranged into p rows, indexed by Laboratory (i) = 1,2,3,....., p and q columns indexed by levels (j) = 1,2, ..., q.

Step 2: Inspection of data

Step 2, involves inspection of Form A and omission of any obviously erroneous data *viz.* measurements falling outside the range of the instrument and data which are impossible for technical reasons. At this stage, the statistical expert involved in the quality control process is supposed to inform about the discordant data to the other members of the panel.

Step 3: Computation of cell means and within cell standard deviations

This step involves preparation of Form B with cell means and Form C with within cell standard deviations. The mathematical notations for the cell mean and the within cell standard deviation, are as follows;

(a) Cell mean

$$\overline{Y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

(b) Within cell standard deviation

$$S_{ij} = \sqrt{\frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} \left(Y_{ijk} - \overline{Y}_{ij.} \right)^2}$$

Where $\overline{Y}_{ij.}$ = cell mean for each laboratory level combination,
n_{ij} = number of test results in the cell for laboratory i at level j,
Y_{ijk} = any one of these test results ($k = 1, 2, \ldots, n_{ij}$, and
S_{ij} = within cell standard deviation.

It is generally recommended that the cell means and the standard deviations should be recorded to one more significant figures than the observed results (Tables 1 and 2).

Step 4: Scrutiny of results for consistency and outliers

Two approaches introduced by the International Organization for Standardization (ISO) are presented in this article, namely: a) graphical consistency techniques and b) numerical outlier tests to find out individual laboratories, which produce discordant test results.

Table 1. *Cell means for iron content of soil in ppm for different laboratories under different levels*

Form B

Laboratory (*i*)	Level (*j*) 1	2	3	4
1	283.80	333.60	397.00	444.00
2	272.60	324.20	377.90	423.30
3	211.90	261.20	313.80	366.90
4	237.00	287.90	343.50	391.30
5	233.90	280.80	334.30	390.50
6	229.00	277.80	324.30	370.80

Table 2. *Within cell standard deviations for iron content of soil in ppm for different laboratories under different levels*

Form C

Laboratory (*i*)	Level (*j*) 1	2	3	4
1	9.72	6.76	10.62	7.24
2	6.91	7.42	4.66	5.71
3	5.54	7.64	4.78	11.51
4	2.39	1.42	8.71	7.33
5	3.98	5.49	6.44	8.00
6	5.07	6.64	6.06	4.60

Step 4a: Graphical consistency techniques

Mandel's h and k plots produced by h and k statistics can be used to identify laboratories with discordant data. The mathematical expressions of h and k statistics are as follows.

(i) Mandel's h statistic (between laboratory consistency statistic)

$$h_{ij} = \frac{\bar{Y}_{ij.} - \bar{Y}_{.j.}}{\sqrt{\frac{1}{(p_j - 1)} \sum_{i=1}^{p_j} \left(\bar{Y}_{ij.} - \bar{Y}_{.j.}\right)^2}}$$

Where, $\bar{Y}_{ij.}$ = cell mean, $\bar{Y}_{.j}$= mean for the level j, and p_j is the number of laboratories for level j. The denominator is the standard deviation of measurements of laboratory means at level j.

(ii) Mandel's k statistic (within laboratory consistency statistic)

$$k_{ij} = \frac{s_{ij}}{\sqrt{\sum s_{ij}^2 / p_j}}$$

or,

$$k_{ij} = \frac{\text{within cell standard deviation}}{\text{pooled within cell standard deviation}}$$

The h and k plots are drawn by plotting the h_{ij} and k_{ij} values for each cell in order of laboratory, in groups for each level tested in the experiment. Inspection of these plots may indicate that specific laboratories exhibit patterns of results that are markedly different from the others in the study. Some of these indications are, consistently high or low within cell variation and/or extreme cell means across many levels. The indicator lines drawn on the h plots serve as guides when examining patterns in the data. ISO Standards Handbook provides necessary tables for drawing the indicator lines (ISO Standards handbook, 1995). If one laboratory stands out on the k plot with many large values, it indicates that the corresponding laboratory has a poorer repeatability than the others. A laboratory with consistently small k values is an indication of excessive rounding of its data or an insensitive measurement scale. Critical levels for k values are given in ISO Standards handbook (ISO Standards handbook, 1995).

(iii) Use of box plots and histograms

Box plots and histograms are effective ways of displaying the distribution and detecting outliers in one or more sets of data. Box plots for different laboratories can be put together on the same scale to inspect the variability within and between laboratories. Producing such plots is not a difficult task with the availability of statistical software that produces high resolution graphics. Moreover, histograms of cell means and cell ranges can reveal the presence of outlying laboratories.

Illustration of Step 4a for iron data

Mandel's h and k plots for the sample data set are depicted in figures 1 and 2.

The *h* and *k* values for each level in each laboratory have never reached the 1% or 5% critical levels. However, when examining the box plots (Fig. 3) Laboratory 1, always showed higher values and laboratory 3 always showed lower values. This phenomenon is also clearly seen in Fig. 1, with spikes nearing the 5% critical line in Laboratory 1. The *k* plot gives consistent results indicating that the variability between replicate test results is lower.

Compared to *h* and *k* plots, the box plot provides a better judgement on differences of both location and dispersion of the test estimates. Moreover, less effort is involved in preparing the final out come of the results.

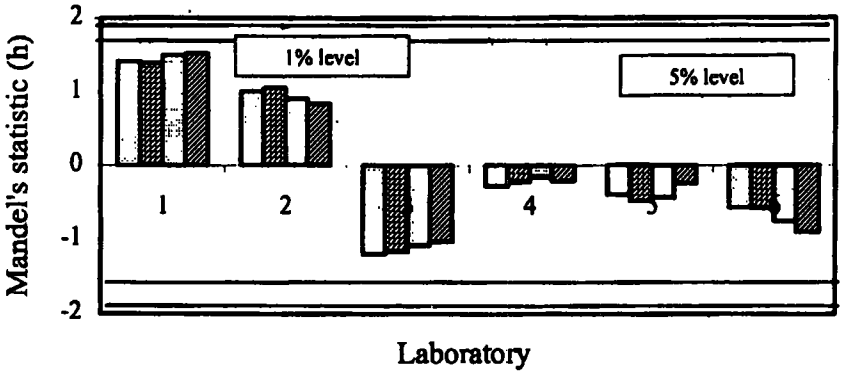

Fig. 1. Mandel's between laboratory consistency statistic (*h*) grouped by laboratories

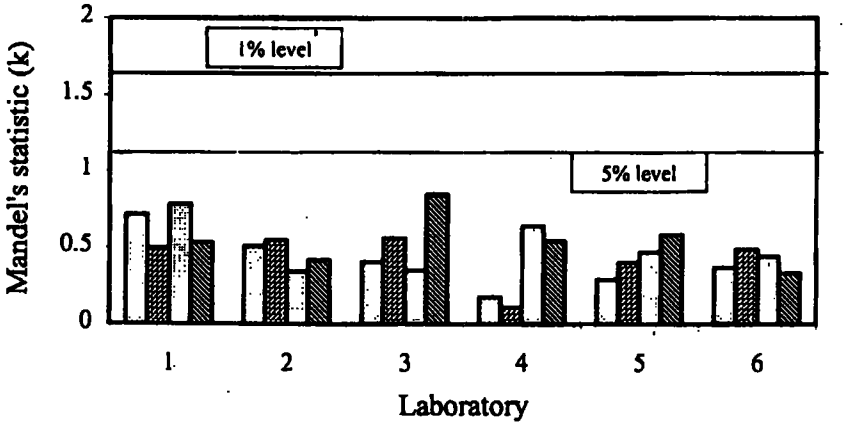

Fig. 2. Mandel's within laboratory consistency statistic (*k*) grouped by laboratories

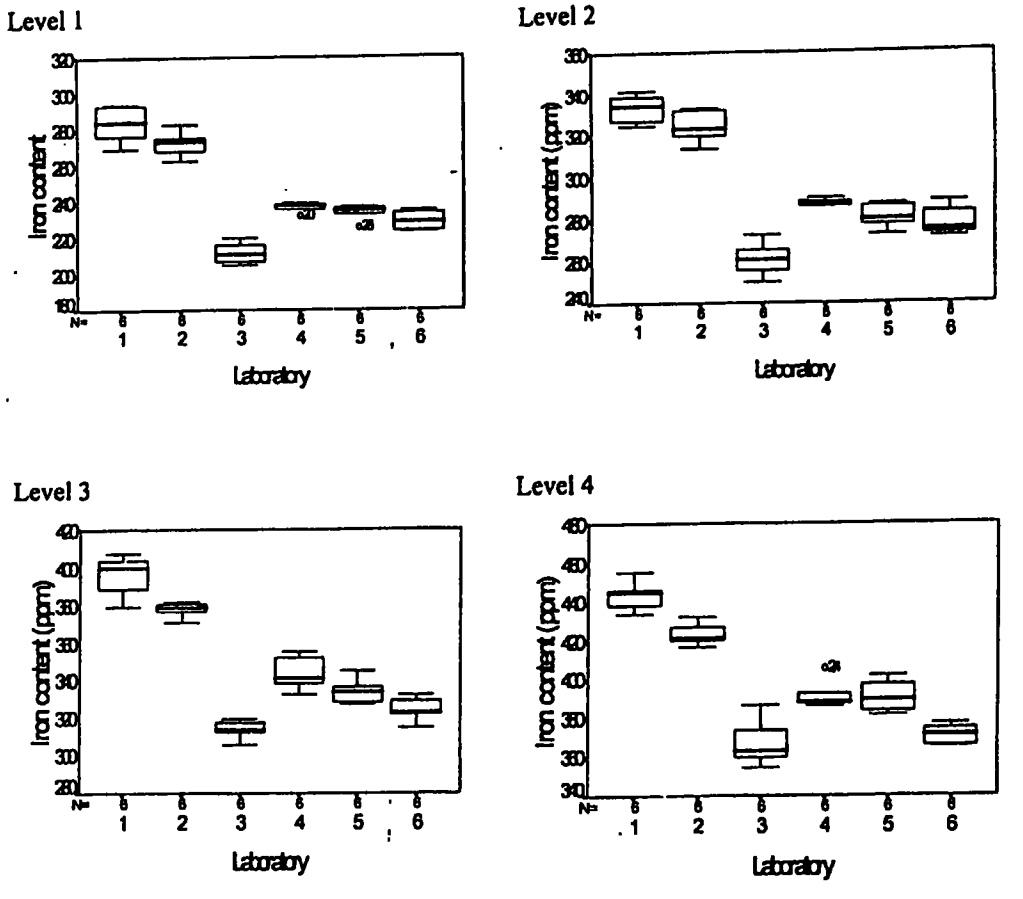

Fig. 3. Box plots indicating the variability of test results under different levels in different laboratories

Step 4b: Numerical outlier technique

The Cochran's and the Grubbs' tests are recommended by the ISO to identify stragglers or outliers. Once, they are identified, an investigation is usually carried out whether the straglers and/or outliers can be explained by some technical error due to a mistake in performing the measurement, an error in computation, a clerical error in transcribing a test result, or analysis of a wrong sample. If there is a technical explanation for the outlying results, a decision can be made whether to discard or correct these data.

i) Cochran's test

The Cochran's test is indicative of the within laboratory variability and should be applied first. Based on that, any necessary action be taken, whether to repeat the tests. Given a set of p standard deviations S_i, all computed from the same number *(n)* of replicates, Cochran's test statistic, C, is

$$C = \frac{S_{max}^2}{\sum_{i=1}^{p} s_i^2}$$

Where, S_{max} is the highest standard deviation in the set. The 1% and 5% critical values for this test are given in the ISO Standards Handbook (1995). If an item has a C value greater than the 1% critical value, it is termed an outlier. If the C value is greater than the 5% critical value and less than the 1% critical value, then the item is called a straggler. If the test statistic is less than or equal to its 5% critical value, the item tested is accepted as correct.

Cochran's test has to be applied to each level separately. However, one of the limitations in Cochran's test is that, it applies strictly only when all the standard deviations are derived from the same number *(n)* of test results obtained under repeatability conditions. In actual cases, this number may vary owing to missing or discarded data. An assumption is made here, and n is taken as the number of test results occurring in the majority of cells. Cochran's criterion tests only the highest value in a set of standard deviations and is therefore, a one-sided outlier test.

ii) Grubbs' test

Grubbs' test is primarily a test of between laboratory variability, and can also be used where Cochran's test has raised suspicions due to the high within laboratory variation attributable to a single test result in the cell. Given a set of data x_i for i = 1,2, 3,4, ..., p, arranged in ascending order, then to determine whether the largest observation is an outlier, Grubbs' test can be used.

Grubbs' statistic, *Gp* is;

$$G_p = (x_p - \bar{x})/s$$

where, and

$$\bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_i \qquad s = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(x_i - \bar{x})^2}$$

The test statistic, which tests the significance of the smallest observation is as follows.

$$G_1 = (\bar{x} - x_1)/s$$

Illustration of Step 4b (Cochran's test)

The Cochran's test statistics computed for each level and the 1% and 5% critical values are given in Table 3. The largest within cell standard deviations in each level are indicated in bold face. However, none of the C values have reached even the 5% level, indicating that the data set analyzed had no straggles or outliers to be removed.

Table 3. *Application of Cochran's test to within cell standard deviations*

		Level			
Laboratory		1	2	3	4
	1	**9.724**	6.756	**10.624**	7.235
	2	6.907	7.425	4.664	5.715
	3	5.538	**7.643**	4.779	**11.51**
	4	2.385	1.421	8.710	7.334
	5	3.976	5.489	6.437	7.998
	6	5.072	6.643	6.059	4.599
C		*0.429*	*0.248*	*0.362*	*0.372*

Critical Levels	
1 %	0.520
5 %	0.445

Illustration of Step 4b (Grubbs' Test)

Application for cell means:

The cell means for a given level j, in which case

$x_i = \bar{y}_{ij.}$

and $p = p_j$, where j is fixed.

The lowest and the highest cell means in this example were tested and accepted as correct, since the test statistics computed for each level had never reached the critical values (Table 4). These finding are in conformation with the findings of Cochran's test.

Table 4. *Application of Grubbs' test to cell means*

Level	Outlier check		
	Single low	Single high	
1	1.19	1.42	Grubbs' test statistics
2	1.16	1.39	
3	1.07	1.50	
4	0.50	0.75	
1%	1.973	1.973	Grubbs' critical values
5%	1.887	1.887	

Further to the above mentioned tests, a critical examination of Form C (Table 2) may sometimes reveal that the standard deviations for a particular laboratory are at all levels or at most levels lower than those for other laboratories. This may indicate that this particular laboratory works with a lower repeatability standard deviation than the other laboratories. This in turn may be caused either by a better technique and equipment or by a modified or incorrect application of the standard measurement method.

If this occurs it should be reported to the panel by the statistician at this stage, to decide whether to proceed with more detailed investigation. Moreover, the specific laboratory should be contacted and possible measures taken to ascertain the cause of such behaviour. The statistical expert could attend to any of the following at this stage, *viz.*

a) retain the laboratory's data for the moment;
b) ask the laboratory to repeat the experiment if feasible, or
c) remove the laboratory's data from the study on the basis of the findings.

Step 5: Computation of general mean and variances

This step involves estimation of general mean, m and the precision for each level separately. In this step means and variances are computed as described below.

Calculation of general mean

For level j, the general mean is

$$\hat{m}_j = \frac{\sum_{i=1}^{p} n_{ij} \bar{y}_{ij.}}{\sum_{i=1}^{p} n_{ij}}$$

Calculation of variances

a) Repeatability variance

$$s_{rj}^2 = \frac{\sum_{i=1}^{p}(n_{ij}-1)s_{ij}^2}{\sum_{i=1}^{p}(n_{ij}-1)}$$

b) Between laboratory variance

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{p-1}\left[\frac{\sum_{i=1}^{p} n_{ij}}{(\sum_{i=1}^{p} n_{ij})^2 - \sum_{i=1}^{p} n_{ij}^2}\right]$$

where,

$$s_{dj}^2 = \frac{1}{p-1}\sum_{i=1}^{p} n_{ij}(\bar{y}_{ij.} - \bar{y}_{.j.})^2$$

c) Reproducibility variance

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2$$

d) Coefficient of Variation *(CV)* for repeatability and reproducibility
CV for repeatability *(CV_{rj})* reproducibility *(CV_{Rj})* are computed as follows.

$$CV_{rj} = \frac{S_{rj}}{m_j} \qquad CV_{Rj} = \frac{S_{Rj}}{m_j}$$

Illustration of step 5 for Iron data

The statistics computed are given in Table 5. The repeatability and reproducibility standard deviations are comparatively low at all levels, justified by low coefficients of variation (Fig. 4). This indicated that the sample data set comprised of sufficiently accurate results.

Table 5. *Computed values of means and variances for each level for iron data*

Level j	P_j	m_j	s^2_{rj}	s^2_{Lj}	s^2_{Rj}	s_{rj}	s_{Rj}
1	6	244.70	36.690	27.477	64.167	6.057	8.010
2	6	294.25	39.247	28.273	67.520	6.265	8.217
3	6	348.47	51.913	32.217	84.130	7.205	9.172
4	6	397.80	59.403	30.048	89.452	7.707	9.458

Note:
p_j = replicates for j^{th} level, m_j = mean for j^{th} level, s^2_{rj} = Repeatability variance, s^2_{Lj} =Between laboratory variance, s^2_{Rj} = Reproducibility variance, s_{rj} = Repeatability standard deviation and s_{Rj} = Reproducibility standard deviation.

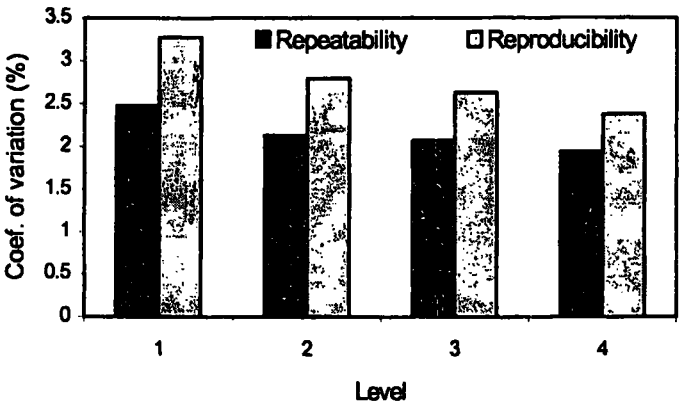

Fig. 4. Coefficients of variation for repeatability and reproducibility for iron data

Step 6: Examining the dependence of the variances upon mean

Once the variances are computed, their relationship with m should be determined.

Illustration of Step 6 for Iron data

It was found that the repeatability and reproducibility standard deviations had significant positive linear relationships with the mean *($p<0.05$)* as depicted in Fig. 5. Thus, it was not necessary to search for other relationships.

The functional forms of the relationships are;

$$s^2_{rj} = 3.1108 + 0.0115\,m \qquad (R^2 = 0.957^*)$$
$$(0.0115) \quad (0.0017)$$

$$s^2_{Rj} = 5.3845 + 0.0104\,m \qquad (R^2 = 0.941^*)$$
$$(0.0104) \quad (0.0018)$$

The standard errors are given in parentheses under respective coefficients. Hence, the estimated values for s_{rj} and s_{Rj} are in the range of 5.927 to 7.689 and 7.920 to 9.507, respectively for the mean values ranging from 244.7 to 397.8.

Step 7: Statement of final results

The responsibility of the statistician at this stage is to write a report to be submitted to the panel. This report should include the following information according to the ISO Handbook (1995).

i) The observations received from the operators and/or supervisors;

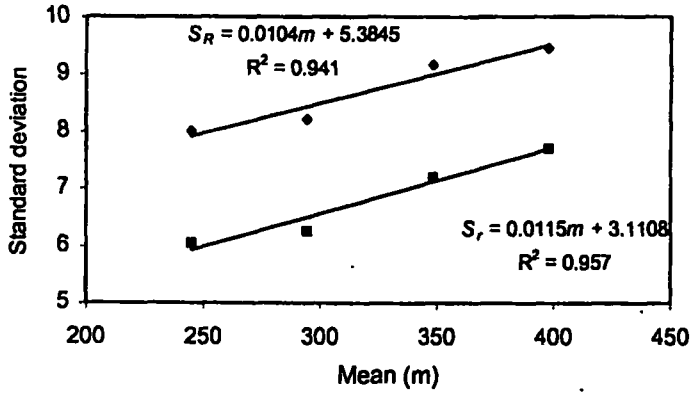

Fig. 5. The relationships between means and standard deviations

ii) a full account of the laboratories which were rejected as outlying laboratories together with reasons for rejection;

iii) a full account of the stragglers and/or statistical outliers which were discarded, stating whether they were explained and corrected or discarded;

iv) a table giving the final results of mean, repeatability and reproducibility standard deviations for the levels tested accompanied with appropriate graphical displays;

v) with an annex comprising of Forms A, B and C.

Step 8: Decisions taken by the panel

Based on the information presented in the statistician's report, the panel will take decisions in consultation of the statistician concerning the following.

i) Whether the discordant observations are due to defects in the description of the standard for the measurement method;

ii) What action to be taken with respect to the rejected outlying laboratories;

iii) If the results of the outlying laboratories and/or the comments received from the operators and supervisors indicate the need to improve the standard for the measurement method, what should the required improvements be;

iv) If the results of the precision experiment justify the establishment of values for repeatability and reproducibility standard deviations, the panel should taken steps to publish them accompanied with the region of measurement in which the precision results are confined to.

Step 9: Preparation of full report

Documentation of the final report is the responsibility of the executive officer, who is in-charge of the central laboratory and responsible for preparation and dispatching of samples to other laboratories. This report should set out the problem, organization of work, statistical expert's report and the conclusions of the panel. The ISO Handbook (1995) suggests the importance of accompanying appropriate graphical presentations in denoting the consistency or variability of test results.

DISCUSSION

Precision experiments are conducted by the Department of Raw Rubber Process and Chemical Analysis of the Rubber Research Institute (RRI). About 12 laboratories are involved in this process to evaluate the standard analytical methods for total solids, dry rubber content, alkalinity, Mg content, mechanical stability, viscosity and volatile fatty acids of rubber latex.

Among the graphical methods for outlier detection, the procedure involved in producing Mandel's plots is tedious when compared to box plots. Moreover, box plots for different laboratories provide information on differences of both location

and dispersion of the test results. The test data sets collected by RRI are generally subjected to both Cochran's and Grubbs' tests to identify variability, within and between laboratories, respectively. Therefore, these two numerical outlier techniques together with box plots seem to be sufficient for the scrutiny of results for consistency and outliers.

CONCLUSIONS

Most of the tests documented in this paper are recommended by the ISO. Moreover, since the ISO standards handbook suggests that the final report should include appropriate graphical presentations, the authors suggest the use of histograms and box plots in addition to the recommended plots. Use of box plots instead of Mandel's *h* and *k* plots accompanied with numerical outlier detection techniques will provide sufficient information on the consistency and accuracy of the results.

REFERENCES

ISO Standards Handbook (1995). 'Statistical methods for quality control. Vol 1& 2, International Organization for Standardization, Switzerland.

Genstat for Windows (1997). An Introductory Course. Statictics Department, IACR-Rothamsted, England.

(Received 26 July 2000)

Appendix 1.

Original data : Iron content of soil in ppm obtained in different laboratories.

Form A

Laboratory *(i)*	Replicate	Level *(j)* 1	2	3	4
1	1	293.3	332.0	408.2	445.0
	2	287.2	324.8	397.7	446.3
	3	294.3	341.5	404.1	455.3
	4	282.1	339.0	402.9	444.6
	5	276.5	337.4	389.2	438.4
	6	269.5	327.1	379.9	434.2
2	1	275.2	323.6	382.3	427.5
	2	267.8	313.6	380.1	420.0
	3	282.9	333.1	377.0	420.6
	4	274.2	322.8	381.0	417.1
	5	272.8	332.1	377.6	432.6
	6	262.6	319.8	371.5	421.7
3	1	205.3	261.1	317.5	363.8
	2	207.4	261.3	313.0	362.5
	3	210.0	255.9	319.4	386.6
	4	220.0	250.6	312.5	373.8
	5	216.4	272.7	314.9	359.9
	6	212.2	265.6	305.7	354.7
4	1	237.7	290.3	342.5	387.7
	2	232.5	288.1	339.4	387.7
	3	238.9	286.4	338.8	388.1
	4	236.2	287.7	332.7	386.4
	5	238.1	288.1	355.7	392.5
	6	238.5	286.5	352.1	405.7
5	1	234.1	277.9	335.2	401.6
	2	226.3	272.7	327.1	397.5
	3	236.4	286.1	333.2	392.2
	4	233.6	281.8	328.7	387.0
	5	237.2	287.4	345.1	383.2
	6	235.9	278.6	336.6	381.6
6	1	235.6	288.6	329.4	374.5
	2	224.0	283.2	332.6	371.5
	3	228.3	274.9	323.3	377.1
	4	223.3	273.6	322.2	370.0
	5	234.2	275.0	322.8	366.1
	6	228.8	271.4	315.3	365.4