

Avaliação da concordância de métodos estatísticos para o controle de qualidade de laboratórios de análise do solo

Peter M. Campos^{1†}, Joel A. Muniz², Marcelo S. Oliveira², Eric B. Ferreira³

¹Universidade Federal de São João del-Rei – UFSJ

²Universidade Federal de Lavras – UFLA

³Universidade Federal de Alfenas – UNIFAL-MG

Resumo: *Os métodos estatísticos são parte fundamental de qualquer programa de controle de qualidade, pois fornecem ferramentas para avaliar a ocorrência de erros sistemáticos e aleatórios, aos quais o processo está constantemente sujeito. Com o objetivo de avaliar a concordância entre dois métodos estatísticos utilizados em um programa interlaboratorial para o controle de qualidade de análises de solo, foram gerados quatro cenários, diferenciados pela variabilidade do conjunto de dados, pela prefixação do CV. Baseados no modelo estatístico do delineamento em blocos casualizados com repetição dentro de blocos, simularam-se 120 experimentos, que foram analisados pelos métodos de Waeny (1980a) e do Profert (Oliveira, Corrêa e Campos, 1995). Assim, logo após a verificação da existência de outliers, estimou-se a proporção de coincidências apresentadas em cada cenário na avaliação da exatidão e na atribuição de selos. Os resultados encontrados permitiram verificar que, nos três cenários onde o conjunto de dados possui $CV \leq 30\%$, ambos os métodos apresentaram desempenho equivalente, com proporção média de coincidências acima de 90,44% para a classificação da exatidão e acima de 87,02% para a proporção média de coincidências para atribuição de selos. Portanto, como há uma elevada proporção de coincidências, pode-se recomendar ambos os métodos para avaliar análises efetuadas por laboratórios de solo quando a variabilidade dos dados possuir um $CV \leq 30\%$. No cenário onde os dados foram simulados com $CV > 30\%$, os métodos obtiveram uma proporção de coincidências de 91,18% para a classificação da exatidão, enquanto que para a proporção média de coincidências para atribuição de selos esta concordância foi de apenas 20,70%, isto é, apesar de uma elevada proporção de coincidência na avaliação da exatidão, a proporção de concordância para a atribuição de selos foi muito pequena, por isso, estes métodos não podem ser considerados equivalentes e outros estudos são necessários para saber qual deles seria mais apropriado.*

Palavras-chave: *Delineamento em blocos casualizados, coeficiente de variação, controle estatístico de processo, programa interlaboratorial.*

Abstract: *Statistical methods are a fundamental part of any quality control program because they provide tools to assess the occurrence of systematic and random errors to which the process is constantly subject. In order to compare the concordance between two statistical methods used in an interlaboratory program for the quality control of soil analyzes, four scenarios were generated, differentiated by the variability of the data set, by the CV prefixation. Based on the statistical mathematical model of the randomized block design with in-block repetition, we simulated 120 experiments, which were analyzed by the methods of Waeny*

† Autor correspondente: pmcampos@ufs.br.

(1980a) and Profert (Oliveira, Corrêa and Campos, 1995). Thus, after verification of the existence of outliers, the proportion of coincidences presented by them in each scenario in the assessment of accuracy and the proportion of coincidences for the assignment of stamps were estimated, observing their behavior. The results showed that, in the three scenarios where the data sets have variability measured by $CV \leq 30\%$, both methods presented an equivalent performance, with a mean proportion of coincidences above 90.44% for the classification of accuracy and above 87.02% for the average proportion of coincidences for assignment of postage stamps. Therefore, as there is a high proportion of coincidences, both methods can be used to evaluate the analyzes carried out by soil laboratories when the variability of the data has a $CV \leq 30\%$. In the case of sets of data where the data were simulated with variability measured by $CV > 30\%$, the methods obtained a coincidence ratio of 91.18% for the classification of accuracy, while it is observed that the proportion mean of coincidence for assignment of stamps this agreement was only 20.70%, that is, despite a high proportion of coincidence in the assessment of accuracy, the proportion of agreement for the assignment of stamps was very small, so these methods can not be considered equivalent and other studies are necessary to know which one would be most appropriate. There have been discrepancies in the classifications of the laboratory analyzes and it is not convenient to replace them.

Key words: A randomized block design, coefficient of variation, statistical process control, interlaboratory program.

INTRODUÇÃO

Juran (1993) e Montgomery (2009) conceituam Controle Estatístico do Processo como a adoção de técnicas estatísticas para medir e avaliar as variações dos processos.

De acordo com o INMETRO (2012), uma medição pode ser definida como um conjunto de operações que tem como objetivo determinar o valor de uma quantidade, dimensão ou grandeza. Um valor verdadeiro é um valor consistente com a definição de uma dada grandeza especificada, ou em outras palavras, é um valor que seria obtido por uma medição perfeita e, portanto, de natureza indeterminada.

A confiança dos resultados obtidos é expressa através de uma faixa de valores dentro do qual se situa, com uma certa probabilidade, o valor verdadeiro do parâmetro ou propriedade que está sendo avaliada, tal como o diâmetro de uma peça, a concentração de uma solução, a pureza de um reagente ou uma dada característica de um solo. Esta faixa de valores é denominada incerteza de medição e pode ser definida como o parâmetro associado ao resultado de uma medição, que caracteriza a dispersão dos valores que podem ser fundamentadamente atribuídos a um mensurando¹.

Montgomery (2009) define duas fontes de variação que atuam em qualquer processo: 1) fontes ou causas aleatórias e, 2) fontes ou causas sistemáticas. A primeira se deve a um componente comum inerente ao mesmo que gera a variabilidade natural, que só é eliminada com mudanças básicas.

No caso de análises de solo, podem-se incluir neste grupo os erros devido a variabilidade natural de uma amostra de solo, as diferenças entre operadores, as variações ambientais, entre outros. Já a segunda fonte inclui as causas sistêmicas ou especiais que não fazem parte do processo, podendo ser identificadas num programa de diagnóstico. Neste grupo, pode-se considerar os erros devido a um operador não treinado ou a uma máquina descalibrada, por

¹ Mensurando - é o objeto da medição.

exemplo. Dado este contexto, compete aos pesquisadores, estatísticos e gestores, a identificação e a quantificação dos erros associados às fontes de variação, garantindo o controle assim com a qualidade na mensuração.

Oliveira e Campos (1995), procuram quantificar as referidas fontes de variação, através dos índices de exatidão (E_i) e de precisão (P_i). O primeiro deles é definido de uma forma geral como sendo a diferença existente entre um valor estimado e um valor verdadeiro que se deseja determinar. O índice de precisão (P_i), mede o grau de dispersão entre os valores resultantes de uma série de medições de um parâmetro e a média destas medidas. Dependendo do enfoque dado à precisão dos resultados tem-se a repetitividade que é a concordância entre os resultados de uma série de medidas de uma característica sob as mesmas condições, ou seja, mesmo método, analista, instrumento, local e demais condições operacionais e ambientais. No caso de variar-se premeditadamente uma ou mais das condições que garantem a avaliação da repetitividade (r), a concordância entre os resultados de uma série de medidas de uma característica representaria a reprodutividade (R).

Segundo Waeny (1980b), repetitividade (r) é o valor máximo esperado para a diferença entre pelo menos dois resultados obtidos com a mesma amostra e o mesmo método sob as mesmas condições, tais como: mesmo operador, mesmo equipamento, mesma data, para as mesmas repetições. Já a reprodutividade (R) é o valor máximo esperado para a diferença entre pelo menos dois resultados obtidos com a mesma amostra e o mesmo método, sob condições diferentes, como por exemplo laboratórios diferentes. Este autor afirma ainda que confirmar as medidas de um determinado instrumento é tão importante quanto necessário para a melhoria de qualquer processo produtivo e/ou administrativo.

Olivieri e D'elia (1996) afirmam que um programa interlaboratorial pode ser descrito como uma série de medições de uma ou mais propriedades, realizada independentemente, por um grupo de laboratórios em amostras de um material, com o objetivo de atingir uma ou mais das seguintes finalidades: compatibilização de resultados, precisão de metodologias, certificação de materiais de referência e ensaios de proficiência.

Por meio das comparações interlaboratoriais é possível avaliar o desempenho de laboratórios para ensaios ou medições específicas, identificar problemas analíticos, estabelecer uma comparabilidade de métodos de ensaio ou calibração, prover confiança adicional aos clientes do laboratório, validar a incerteza declarada e atribuir valores para materiais de referência (ABNT, 2011).

Ensaio de Proficiência (EP) são estudos interlaboratoriais utilizados como ferramentas de avaliação externa e demonstração da confiabilidade dos resultados analíticos. Servem também para identificar falhas e possibilitar a tomada de ações corretivas ou preventivas, sendo um dos itens necessários para a acreditação de ensaios pela Norma NBR ISO/IEC 17025:2005.

Considerando o processo de análise laboratorial independentemente do tipo de material analisado, solo, sementes ou óleos lubrificantes, essencialmente o objetivo é o mesmo: realizar uma análise correta ou uma mensuração confiável.

Montgomery (2009) afirma que todo processo, natural ou industrial, está sujeito a variações ou erros derivados de fontes classificadas como sistemáticas ou aleatórias, as quais são respectivamente controláveis ou não-controláveis. Portanto, a identificação e a gestão destas fontes de erro são imprescindíveis para que se garanta o nível de qualidade exigido para as mensurações laboratoriais.

Neste trabalho são avaliados os resultados apresentados por laboratórios de análise de solo por dois métodos estatísticos. Essa escolha se deveu por duas razões: pela disponibilidade de dados e pela importância da análise de fertilidade do solo dentro de uma agricultura moderna, preocupada com sua eficácia e sustentabilidade. Além disso, é importante que os resultados das

análises laboratoriais sejam confiáveis para que os agricultores sintam segurança em investir na análise do solo e desfrutem de seus benefícios, tais como:

- a) Evitar uma correção insuficiente do solo, onde ocorra a deficiência de macro e micronutrientes;
- b) Evitar uma correção excessiva do solo gerando desperdício e risco de contaminação do solo e da água;
- c) Dimensionar adequadamente os gastos com insumos, necessário à uma agricultura que resulte em maior produção, aumentando o lucro do agricultor;
- d) Permitir um melhor aproveitamento de recursos naturais, prevenindo a ocorrência de desastres ecológicos;
- e) Possibilitar maior difusão da análise de fertilidade como importante técnica de auxílio ao produtor.

Além de uma exigência de mercado, a participação dos laboratórios em programas interlaboratoriais possibilita a formação de grupos em torno de um mesmo objetivo, podendo definir qual metodologia de análises seguir, que controles devem ser feitos, quais métodos estatísticos se adequam melhor a sua realidade, permitindo ter parâmetros de comparação para discutir problemas comuns. Deve-se ressaltar que um programa interlaboratorial de controle de qualidade não pressupõe a existência de um laboratório que seja o referencial da verdade, isto é, um laboratório cujas as determinações laboratoriais sejam inquestionavelmente verdadeiras. Assim sendo, um programa interlaboratorial de controle da qualidade não pretende comparar as determinações laboratoriais de cada participante com determinações reconhecidamente verdadeiras, mas sim pretende apenas avaliar a concordância ou discordância entre as determinações laboratoriais dos laboratórios participantes, considerando que os laboratórios que concordam entre si determinam corretamente os atributos medidos, enquanto que os laboratórios discordantes os determinam incorretamente.

Considerando que a Estatística é a ferramenta que instrumentaliza esses programas de controle de qualidade, comparar métodos estatísticos utilizados para esse fim é necessário. Para que os resultados apresentados nas análises laboratoriais sejam confiáveis, os métodos estatísticos devem considerar três aspectos: (i) identificação de outliers², que são valores que podem conferir tendenciosidade aos dados; (ii) exatidão, índice que avalia se em média as análises feitas pelos laboratórios “acertam” o atributo avaliado e (iii) precisão das medidas, índice que quantifica a dispersão das medidas em torno de sua média, evidenciando se a precisão está sob condições de repetitividade e/ou sob condições de reprodutividade. A Figura 1 ilustra as quatro possíveis configurações quando se compara a exatidão com a precisão entre medições.

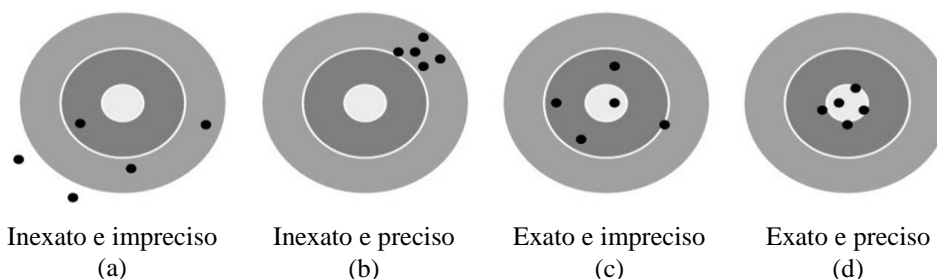


Figura 1: Comparação entre a exatidão e a precisão na determinação do valor de uma quantidade, dimensão ou grandeza.

²Outlier: valor discrepante, muito diferente de outros vindos da mesma população.

O objetivo deste trabalho foi de avaliar os métodos estatísticos Waeny e Profert através de avaliação da habilidade para detectar *outliers*, exatidão e precisão em um programa de controle de qualidade de análise de laboratórios de solo.

METODOLOGIA

Métodos estatísticos

Como são os métodos estatísticos que fornecem subsídios para a avaliação da qualidade dos resultados fornecidos nas análises e nosso objetivo aqui é avaliar dois métodos estatísticos, estes serão apresentados e, logo em seguida, definiremos os critérios adotados neste trabalho para essa avaliação.

Método de Waeny

Waeny (1980a) decompõe a variância total, através de um processo baseado em Youden e Steiner (1975), que utiliza os componentes de variância associados à análise de variância, propondo um índice de repetitividade (r) e outro de reprodutividade (R), ambos ligados a precisão dos resultados. Para conseguir valores confiáveis de r e R deve-se trabalhar com um mínimo de quinze laboratórios. Cada laboratório analisa as diferentes amostras e suas repetições de forma independente, as quais são elaboradas e definidas pelos coordenadores do programa.

Waeny (1980b) considerou a análise de variância de um esquema fatorial, num delineamento inteiramente casualizado (DIC) descrito através do seguinte modelo aleatório:

$$Y_{ijm} = \mu + l_i + s_j + ls_{ij} + e_{ijm} \quad (1)$$

em que Y_{ijm} é a determinação da repetição m de uma amostra j , para um laboratório i ; μ é uma constante associada a todas as observações; l_i é o efeito do i -ésimo laboratório, com $i = 1, 2, \dots, I$; s_j é o efeito da j -ésima amostra, com $j = 1, 2, \dots, J$; ls_{ij} é o efeito da interação do laboratório i com a amostra j ; e_{ijm} é o erro experimental.

Tabela 1: Esquema da análise de variância com as esperanças de Quadrados Médios.

Causas de Variação	GL	QM	E[QM]
Laboratório (L)	$I - 1$	V_1	$\sigma^2 + M\sigma_{ls}^2 + JM\sigma_l^2$
Solo (S)	$J - 1$	V_2	$\sigma^2 + M\sigma_{ls}^2 + IM\sigma_s^2$
L*S	$(I - 1)(J - 1)$	V_3	$\sigma^2 + M\sigma_{ls}^2$
Erro	$IJ(M - 1)$	V_4	σ^2
Total	$IJM - 1$	-	-

Trata-se de um modelo aleatório, com todos os efeitos independentes, então:

$$\begin{aligned} E[l_i] &= 0; E[l_i^2] = Var[l_i] = \sigma_l^2; l_i \sim N(0, \sigma_l^2); \\ E[s_j] &= 0; E[s_j^2] = Var[s_j] = \sigma_s^2; s_j \sim N(0, \sigma_s^2); \\ E[ls_{ij}] &= 0; E[ls_{ij}^2] = Var[ls_{ij}] = \sigma_{ls}^2; ls_{ij} \sim N(0, \sigma_{ls}^2); \end{aligned}$$

$$E[e_{ijm}] = 0; E[e_{ijm}^2] = Var[e_{ijm}] = \sigma^2; e_{ijm} \sim N(0, \sigma^2).$$

O cálculo dos índices de repetitividade (r) e reprodutividade (R) é feito com base nos valores das variâncias relativas a cada índice, $\hat{\sigma}_r^2$ e $\hat{\sigma}_R^2$, que são obtidos através da estimação dos componentes de variância, os quais são obtidos pelas seguintes expressões:

$$\hat{\sigma}^2 = QM_{erro} \quad \hat{\sigma}_{ls}^2 = \frac{QM_{ls} - QM_{erro}}{M} \quad \hat{\sigma}_l^2 = \frac{QM_l - QM_{ls}}{J} \quad (2)$$

Como $\hat{\sigma}_r^2 = \hat{\sigma}^2$ e $\hat{\sigma}_R^2 = \hat{\sigma}^2 + \hat{\sigma}_{ls}^2 + \hat{\sigma}_l^2$, onde a variação que dá origem a reprodutividade $\hat{\sigma}_R^2$ é uma combinação de variâncias entre laboratórios (extrínsecas), dentro de laboratórios (intrínsecas), e da ação da interação.

Os valores dos índices r e R são obtidos pela aplicação de amplitudes padronizadas aos desvios padrões $\hat{\sigma}_r$ e $\hat{\sigma}_R$ e pela escolha do nível de confiança $(1 - \alpha)\%$. Caso o nível de confiança escolhido seja de 95%, com 2 repetições e admitindo-se uma distribuição normal, o fator será de $1,96\sqrt{2} = 2,77$ e as estimativas de r e R são obtidos por: $R = 2,77\hat{\sigma}_R$ e $r = 2,77\hat{\sigma}_r$.

De acordo com Waeny (1980b) de posse das estimativas de r e R , se pode avaliar se estes valores são aceitáveis ou não:

- a) Para verificar se a repetitividade (r) será aceita ou não, aplica-se o seguinte teste:
 - Calcula-se a amplitude (W_n) entre as repetições das determinações de uma característica efetuada por um único laboratório e aplica-se o fator da Tabela 2 no desvio padrão relativo a repetitividade $\hat{\sigma}_r$;
 - Caso $W_n \leq D_n \hat{\sigma}_r$, aceita-se o valor determinado, considerando o laboratório como repetitivo, caso contrário; rejeita-se o valor determinado e o laboratório será considerado como não repetitivo.
- b) Para verificar se a reprodutividade (R) será aceita ou não aplica-se o teste que utiliza a variância relacionada a este índice:
 - Calcula-se a amplitude (W_l) entre as determinações de uma característica efetuada por todos os laboratórios e aplica-se o fator da Tabela 2 no desvio padrão relativo à reprodutividade $\hat{\sigma}_R$;
 - Caso $W_l \leq D_n \hat{\sigma}_R$, os laboratórios são compatíveis entre si, caso contrário, os laboratórios não são compatíveis entre si.

Tabela 2: Tabela de Tukey para $g.l. \rightarrow \infty$ - valores do fator D_n para 95% de confiabilidade.

Repetições	2	3	4	5	6	7	8	9	10	11	12
D_n	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62

Fonte: Waeny (1980b).

Método do Profert

Oliveira, Corrêa e Campos (1995) avaliaram os resultados de um programa de controle de qualidade das análises de laboratórios de solo, utilizando técnicas da estatística básica como medidas de posição e intervalos de confiança. Baseando-se em propriedades da distribuição

normal de probabilidade, são construídas faixas de confiança, fazendo a união de dois intervalos de confiança para a mediana, considerada como valor verdadeiro. Baseados nestes intervalos chega-se a estimativa da exatidão de um laboratório. A precisão é determinada através de um índice de imprecisão, que se baseia num intervalo de confiança para a razão de variâncias.

Identificação de *Outliers*

Para identificar os *outliers*, utilizam-se estatísticas robustas³, especificamente os quartis e o intervalo interquartil. Um quartil x_α ($\alpha = 0,25; 0,50; 0,75$) da distribuição de X é o valor de X que tem uma frequência acima de $(100\alpha\%)$, respectivamente igual a 25%, 50%, ou 75%. A distância interquartil (δ) é definida por: $\delta = x_{0,25} - x_{0,75}$ (BUSSAB; MORETIN, 2013).

As frequências de ocorrência dos valores de X para os intervalos do tipo:

$$\left(x_{0,75} - \frac{B}{2}\delta, x_{0,25} + \frac{B}{2}\delta\right) \quad (3)$$

com B inteiro são dados por:

$$P\left(x_{0,75} - \frac{B}{2}\delta \leq X \leq x_{0,25} + \frac{B}{2}\delta\right) = 2P\left(\mu \leq X \leq x_{0,25} + \frac{B}{2}\delta\right) \quad (4)$$

Definindo $d_1 = B - 1$, para $B > 0$, como uma medida da discrepância que classificaria a determinação X , e fazendo d_1 variar de 0 a 9, considera-se como *outlier* toda determinação que tiver discrepância $d_1 > 2$, isto é, toda determinação X que estiver fora do intervalo:

$$\left(\hat{x}_{0,75} - \frac{3}{2}\hat{\delta}; \hat{x}_{0,25} + \frac{3}{2}\hat{\delta}\right) \quad (5)$$

em que: $\hat{x}_{0,75}$ e $\hat{x}_{0,25}$ são, respectivamente, o estimador do quartil com frequência de 75% acima, ou primeiro quartil e o estimador do quartil com frequência de 25% acima, ou terceiro quartil; $\hat{\delta}$ é o estimador do intervalo interquartil.

Discrepância assim definida para uma determinação laboratorial X , indica que X não é aceito como sendo um valor próximo do verdadeiro valor, que é a mediana md , isto é, X não é um valor aceitável para a característica em questão. A probabilidade de X ser um valor aceitável para o atributo sob análise, e se classificar erroneamente como não aceitável, é menor do que 0,9%.

Cálculo da exatidão

Outra discrepância para a determinação X , é denominada de d_2 , e pode ser definida como sendo:

$$d_2 = \begin{cases} 0, & \text{se } B = 0 \\ B - 1, & \text{se } B > 0 \end{cases}$$

³ - são estatísticas que não sofrem influência (ou são menos influenciadas), quando existem no conjunto de dados valores fora da realidade.

Deve-se salientar que os valores de B são sempre inteiros e que d_2 também varia de 0 a 9. De forma similar ao que foi feito para identificar *outliers*, constroem-se faixas, através da união de vários intervalos de confiança para a mediana de uma distribuição normal de probabilidade. A fórmula geral para construção destas faixas é a seguinte:

$$(md - (B + 1) \hat{\sigma}_{md}; md - B\hat{\sigma}_{md}) \cup (md + B\hat{\sigma}_{md}; (B + 1) \hat{\sigma}_{md}) \quad (6)$$

onde: $\hat{\sigma}_{md}$ é o estimador do desvio padrão da mediana, e pode ser calculado pela expressão (7):

$$\hat{\sigma}_{md} = \sqrt{\frac{\pi}{2} \cdot \frac{\hat{\sigma}^2}{n'}} \quad (7)$$

em que: md é a estimativa da mediana, n' é o novo tamanho da amostra, ($n' \leq n$) e $\hat{\sigma}^2$ é a estimativa da variância para n' .

O cálculo da exatidão é inspirado na teoria clássica de probabilidade, onde cada faixa calculada pela expressão (6) é considerada como um evento (MOOD; GRAYBILL; BOES, 1974).

Faz-se, então, uma frequência relativa entre a soma das discrepâncias, d_1 e d_2 , e o total de discrepância que um laboratório pode assumir. Apresenta-se, então, a fórmula geral que permite, em função do número de amostras, repetições e características analisadas, obter a estimativa da exatidão:

$$E_i = \left(1 - \frac{T_i}{18 \cdot a \cdot c}\right) \cdot 100 \quad (8)$$

sendo: E_i a exatidão do laboratório i , a o número de amostras analisadas, c o número de características analisadas e T_i o total de discrepância para o laboratório i , o qual depende diretamente de a e c .

Análise de precisão

A precisão de um laboratório na determinação de uma característica como pH, Ca (Cálcio), K (Potássio) ou P (Fósforo), é tanto menor quanto maior for a variância entre as repetições das determinações feitas para um mesmo solo. Porém, existe uma discussão em torno das faixas de aceitação consideradas pela estatística e as consideradas na fertilidade. Um exemplo prático desta discordância, é o caso de duas determinações para o Ca (Cálcio) trocável de um solo. Imagine que elas apresentem valores de 0,1 e 0,2 $\text{cmol}_c \text{ kg}^{-1}$ (centimol de carga por kilograma de solo): a segunda tem, em relação à primeira, uma variação de 100%; no entanto, sob o ponto de vista de fertilidade do solo, elas são semelhantes. Nota-se por este exemplo, que é necessário adequar os fundamentos da estatística à realidade da análise do solo, o que não é tarefa fácil.

Para contornar esta divergência, considerou-se que todas as determinações onde a soma das discrepâncias d_1 e d_2 seja zero, serão consideradas como corretas. Calculando-se a variância destes dados, define-se a variância mínima inerente ($\hat{\sigma}_{min}^2$). Calcula-se, também, a

variância de todos os dados da característica envolvida na análise do laboratório durante o ano, a qual é chamada de variância observada ($\hat{\sigma}_{obs}^2$).

A partir daí constrói-se um intervalo de confiança unilateral para a razão das variâncias ($\hat{\sigma}_{min}^2$) e ($\hat{\sigma}_{obs}^2$). Este intervalo compõe o índice de imprecisão (I_i), $I_i > 0$, o qual é calculado a um nível de $\alpha\%$ de confiança, da seguinte forma:

$$I_i = \text{parte inteira} \left(\frac{\hat{\sigma}_{obs}^2}{\hat{\sigma}_{min}^2} \cdot \frac{1}{F_{\alpha(v_2, v_1)}} \right) - 1; \quad 0 \leq I_i \leq 9 \quad (9)$$

sendo: $\hat{\sigma}_{min}^2$ a variância mínima inerente, $\hat{\sigma}_{obs}^2$ a variância observada, $F_{\alpha(v_2, v_1)}$ o valor tabelado da distribuição $F(\alpha = 5\%)$, com v_2 graus de liberdade para o denominador e v_1 graus de liberdade para o numerador.

Partindo também de uma motivação baseada na definição de probabilidade clássica, usando frequências relativas, a precisão do laboratório i (P_i) para o atributo X será:

$$P_i = \text{parte inteira} \left(100 - \frac{I_i \cdot 100}{9} \right) \quad (10)$$

A precisão total para o laboratório i ($PT_i(\%)$) é calculada:

$$PT_i(\%) = \left(\frac{P_i(pH) + P_i(Ca) + P_i(P) + P_i(K)}{4} \right) \quad (11)$$

Critérios de avaliação

Para que fosse viável a avaliação entre os métodos estatísticos de Waeny e Profert, optou-se pela estimação de duas proporções de coincidências. A primeira classifica as análises feitas pelos laboratórios em relação a exatidão, e a outra classifica as análises feitas pelos laboratórios em relação a atribuição de selos feita no final do ano, e que depende do cálculo da precisão.

Uma coincidência de análise ocorre quando os métodos estatísticos em questão classificam de igual forma um laboratório, seja para o caso da exatidão, ou seja, para a atribuição de selos feita no final do ano.

Para o método de Waeny, um laboratório exato é aquele que não é excluído após a análise de exatidão. Ou seja, depois de calculada a amplitude entre os resultados obtidos de uma mesma amostra por todos os laboratórios, se esta amplitude for menor do que o índice de reprodutividade (R) indica que a diferença entre as determinações feita pelos laboratórios remanescentes é considerada aceitável e, por esta razão, os resultados apresentados por estes laboratórios são considerados exatos. Um laboratório preciso é aquele que possui uma amplitude, entre as determinações feitas por ele, menor que o índice de repetitividade (r).

No Profert a exatidão e a precisão são quantificadas, sendo considerados exatos todos os laboratórios que possuem pelo menos 67% de índice de exatidão, para as duas últimas análises feitas no ano. Em outras palavras, significa obter na soma das discrepâncias d_1 e d_2 , valores menores do que 6, em uma única amostra. Os laboratórios considerados precisos são os que apresentam, também, no mínimo, 67% índice de precisão.

A atribuição de selos no final do ano é comum aos dois métodos e se baseia nos seguintes critérios:

- Todos os laboratórios que forem classificados como inexatos não recebem nenhum selo, configuração ilustrada pelos alvos (a) e (b) da Figura 1;
- Todos os laboratórios que forem classificados como exatos, porém imprecisos, recebem um selo prata, configuração ilustrada pelo alvo (c) da Figura 1;
- Todos os laboratórios que forem classificados como exatos e precisos, recebem um selo ouro, vide alvo (d) da Figura 1.

A sistemática de funcionamento do programa interlaboratorial para análise de solos considera a realização de reuniões trimestrais, onde são distribuídas aleatoriamente 3 amostras em cada reunião. Dessa forma, o total de amostras distribuídas ao longo do ano a cada laboratório participante totaliza 12 amostras. Nestas 12 amostras, tem-se 6 tipos de solos com 2 repetições cada. A distribuição destas 12 amostras é feita pela coordenação do programa que sorteia quais amostras serão distribuídas em cada uma das reuniões do ano. Decidiu-se trabalhar com 19 laboratórios que apresentavam todos os resultados para as amostras e características de interesse, número esse que atende as recomendações de Waeny (1980a) (CAMPOS, 1998). Deve-se ressaltar que esses números não constituem perda de generalidade para a avaliação estudada neste artigo.

Simulação de dados

O processo de simulação considerou quatro cenários onde uma variabilidade diferente foi pré-fixada ao conjunto simulado através do coeficiente de variação (CV%), variando-se também o atributo a partir de considerações conhecidas na Ciência do Solo para aquele atributo, a saber, o pH varia pouco, o P (fósforo) varia muito, etc. Para tal, baseou-se num estudo descritivo feito sobre o conjunto de dados fornecidos pelo Profert, que determinou:

CV = 1% : para os dados do pH; CV = 30%: para os dados do K (Potássio);
CV = 10%: para os dados do Ca (Cálcio); CV = 50%: para os dados do P (Fósforo).

Para cada cenário foram simulados 30 experimentos virtuais, entendidos como a determinação de uma característica do solo, por exemplo pH, feita sobre 12 amostras por 19 laboratórios independentes.

O delineamento adotado para estes experimentos virtuais foi o de blocos casualizados com repetição dentro de blocos, desconsiderando-se a interação, portanto, a simulação se baseou no seguinte modelo matemático (BANZATTO; KRONCA, 2006; GOMES, 2009):

$$Y_{ijm} = \mu + l_i + s_j + e_{ijm} \quad (12)$$

em que: Y_{ijm} é a determinação feita pelo laboratório (tratamento) i , de um solo (bloco) j , na repetição m ; μ é uma constante associada a todas as observações; l_i é o efeito do i -ésimo laboratório, com $i = 1, 2, \dots, 19$; s_j é o efeito do j -ésimo solo, com $j = 1, 2, \dots, 6$; e_{ijm} é o erro experimental.

A análise de variância com as esperanças dos quadrados médios está apresentada na Tabela 3. Essa análise de variância é equivalente a apresentada por Waeny (1980b) admitindo que a interação seja nula. Considerando o modelo como aleatório e com todos os efeitos

independentes, pode-se escrever os seguintes resultados através do cálculo da esperança matemática:

$$\begin{aligned} E[l_i] &= 0; E[l_i^2] = Var[l_i] = \sigma_l^2; l_i \sim N(0, \sigma_l^2); \\ E[s_j] &= 0; E[s_j^2] = Var[s_j] = \sigma_s^2; s_j \sim N(0, \sigma_s^2); \\ E[e_{ijm}] &= 0; E[e_{ijm}^2] = Var[e_{ijm}] = \sigma^2; e_{ijm} \sim N(0, \sigma^2). \end{aligned}$$

Os dados simulados para as características do solo, foram gerados mantendo-se uma estrutura mais próxima da realidade de dados reais, fornecidos pelo Programa Interlaboratorial de Controle de Qualidade de Análise de Solo do Estado de Minas Gerais (Profert-MG).

RESULTADOS E DISCUSSÃO

Foram analisados 120 experimentos simulados para cada um dos dois métodos estatísticos, totalizando $2 \times 4 \times 30 = 240$ tabelas de análise de variância do tipo mostrado na Tabela 3. Logo em seguida foram estimadas as duas proporções de coincidências relativas a exatidão e a atribuição de selos feita no final do ano, a qual depende do cálculo da precisão.

Para exemplificar a aplicação dos métodos aqui avaliados apresenta-se a análise do primeiro experimento para a simulação da característica pH, na sequência verificamos o comportamento das proporções de coincidências para os 240 experimentos.

Método de Waeny

Considerando que o cálculo dos índices de repetitividade (r) e de reprodutividade (R) no método de Waeny, dependem dos componentes de variância do modelo derivados da equação (1), a análise de variância para os dados simulados para pH para o experimento 1, é apresentada na tabela 4.

Tabela 4: Análise de variância para os dados simulados para o pH do experimento 1, dados apresentados na tabela 5.

Fontes de Variação	gl	SQ	QM
Laboratório (L)	18	4,6555	0,2586
Solo (S)	5	0,0134	0,0027
Erro	204	0,5398	0,0026
Total	227	5,2087	-

De posse dos valores obtidos para os quadrados médios, estima-se a contribuição de cada efeito aleatório conforme expressões dadas em (2) para a variância da variável pH:

Componentes de Variância	Estimativa
$\hat{\sigma}_l^2$	0,0213
$\hat{\sigma}^2$	0,0026

Assim a estimativa da variância da repetitividade ($\hat{\sigma}_r^2$) é dada por:

$$\hat{\sigma}_r^2 = \hat{\sigma}^2 = QM_{erro} = 0,0026$$

e a da reprodutividade ($\hat{\sigma}_R^2$) por:

$$\hat{\sigma}_R^2 = \hat{\sigma}^2 + \hat{\sigma}_l^2 = 0,0026 + 0,0213 = 0,0240$$

Os índices de repetitividade (r) e reprodutividade (R) são estimados aplicando-se aos respectivos desvios padrões, um fator numérico obtido na tabela 3, tabela de Tukey com graus de liberdade tendendo ao infinito. Como são 2 repetições, este fator será igual a 2,77, assim o índice de repetitividade (r) será:

$$r = D_n \hat{\sigma}_r = 2,77 \times 0,05144 = 0,14$$

enquanto que o índice de reprodutividade (R) é:

$$R = D_l \hat{\sigma}_R = 2,77 \times 0,1548 = 0,43$$

Para classificar os laboratórios como repetitivos calcula-se a amplitude (W_n) obtida entre as determinações efetuadas por um laboratório sobre as 12 amostras, 6 tipos de solo repetidos duas vezes, assim:

- Caso $W_n \leq D_n \hat{\sigma}_r = 0,14$, aceita-se o laboratório como repetitivo;
- Caso $W_n > D_n \hat{\sigma}_r = 0,14$, rejeita-se, considerando o laboratório não repetitivo.

Já para classificar os laboratórios como reprodutivos, calcula-se a amplitude (W_l) entre a determinação de uma característica feita por todos os laboratórios e assim:

- Caso $W_l \leq D_l \hat{\sigma}_R = 0,43$, aceita-se o laboratório como reprodutível;
- Caso $W_l > D_l \hat{\sigma}_R = 0,43$, rejeita-se, considerando o laboratório como não reprodutivo.

Na Tabela 5 estão apresentados os resultados obtidos pelo método de Waeny com os dados simulados para o pH e experimento 1.

Tomando os dados do solo 1 e do solo 6, nas repetições 1 e 2 (Tabela 5), verifica-se que as amplitudes calculadas entre os laboratórios (W_l), nestes casos, superam o índice de reprodutividade que é $R = 0,43$. Portanto, neste caso, existe um ou mais laboratórios que são considerados como não reprodutíveis. Pode-se dizer que possivelmente no momento da análise algum problema existiu dentro dos laboratórios, o que os levou a obter valores extremos que geraram uma amplitude entre os dados maior do que o limite determinado pelo índice de reprodutividade (R).

Para diferenciar quais laboratórios são reprodutíveis e quais não são, toma-se então o laboratório que apresenta resultado com maior dispersão em relação à média, retirando-o do conjunto de dados, calcula-se novamente a amplitude dos dados (W_l), comparando com o índice de reprodutividade (R). Se $W_l < R$ encerra-se o procedimento, caso contrário repete-se o passo anterior até que todos os laboratórios sejam considerados reprodutíveis.

Desta forma, avaliando a reprodutibilidade dos laboratórios na avaliação do pH da amostra do experimento 1, solo 1, nas repetições 1 e 2; concluiu-se que os resultados apresentados pelos laboratórios 9 e 14 são os que geraram um $W_l > R$ e para o solo 6, nas repetições 1 e 2, o laboratório 14 é que inflacionou a amplitude entre as determinações feitas pelos laboratórios sobre uma mesma amostra, portanto, estes laboratórios são classificados como não-reprodutíveis ou inexatos para estes casos.

Para a avaliação da repetitividade, num processo análogo ao anterior, obteve-se a amplitude entre as repetições (W_n) determinadas por único laboratório, comparando-se ao valor do índice de repetitividade que é de $r = 0,14$. Como todas as amplitudes calculadas no exemplo foram menores que o índice de repetitividade (r), então a variação existente dentro dos laboratórios é aceitável, sendo todos os laboratórios considerados como precisos para este caso.

Tabela 5: Resultado obtido pelo método de Waeny, no confronto das determinações de pH do experimento 1, de 6 amostras de solo, repetidas 2 vezes feitas por 19 laboratórios.

Lab	Rep	pH SIMULADO (c molc kg ⁻¹)												W _n	\bar{x}	Selo
		Solo 1		Solo 2		Solo 3		Solo 4		Solo 5		Solo 6				
		1	2	1	2	1	2	1	2	1	2	1	2			
1		5,66	5,69	5,63	5,68	5,63	5,62	5,64	5,70	5,74	5,65	5,63	5,61	0,09	5,66	Ouro
2		5,65	5,67	5,65	5,64	5,53	5,57	5,65	5,64	5,67	5,69	5,75	5,74	0,04	5,65	Ouro
3		5,69	5,69	5,75	5,74	5,75	5,77	5,64	5,62	5,84	5,81	5,83	5,82	0,05	5,75	Ouro
4		5,88	5,86	5,88	5,90	5,90	5,90	5,85	5,87	5,87	5,88	5,85	5,80	0,09	5,87	Ouro
5		5,70	5,69	5,71	5,76	5,67	5,69	5,70	5,71	5,74	5,76	5,70	5,64	0,07	5,71	Ouro
6		5,76	5,76	5,72	5,71	5,79	5,78	5,80	5,76	5,76	5,78	5,78	5,78	0,04	5,77	Ouro
7		5,66	5,63	5,68	5,68	5,74	5,72	5,72	5,74	5,60	5,61	5,63	5,62	0,05	5,67	Ouro
8		5,69	5,67	5,66	5,65	5,68	5,67	5,60	5,63	5,76	5,77	5,65	5,58	0,07	5,67	Ouro
9		5,99	6,02	5,98	6,00	5,90	5,91	5,84	5,84	5,87	5,87	5,92	5,90	0,05	5,92	Ouro
10		5,80	5,71	5,65	5,68	5,64	5,64	5,72	5,73	5,78	5,80	5,79	5,81	0,09	5,73	Ouro
11		5,65	5,60	5,69	5,64	5,79	5,83	5,63	5,66	5,71	5,74	5,64	5,65	0,08	5,69	Ouro
12		5,64	5,60	5,57	5,56	5,58	5,57	5,64	5,62	5,59	5,59	5,60	5,60	0,06	5,60	Ouro
13		5,65	5,64	5,66	5,64	5,65	5,66	5,67	5,68	5,68	5,68	5,67	5,69	0,06	5,66	Ouro
14		5,07	5,07	5,27	5,28	5,20	5,24	5,17	5,15	5,18	5,22	5,16	5,15	0,05	5,18	Sem
15		5,68	5,73	5,71	5,72	5,73	5,74	5,75	5,72	5,64	5,64	5,70	5,69	0,05	5,70	Ouro
16		5,66	5,71	5,68	5,67	5,67	5,63	5,73	5,70	5,71	5,72	5,73	5,69	0,08	5,69	Ouro
17		5,62	5,61	5,69	5,71	5,61	5,61	5,62	5,67	5,59	5,58	5,59	5,63	0,08	5,63	Ouro
18		5,71	5,68	5,79	5,78	5,56	5,54	5,75	5,76	5,71	5,77	5,73	5,71	0,07	5,71	Ouro
19		5,55	5,54	5,58	5,56	5,63	5,59	5,56	5,51	5,48	5,51	5,53	5,52	0,07	5,55	Ouro
W_i		0,92	0,95	0,71	0,72	0,70	0,67	0,68	0,72	0,69	0,66	0,76	0,75			
x		5,67	5,66	5,68	5,68	5,67	5,67	5,67	5,67	5,68	5,69	5,68	5,66			

Método do Profert

Os resultados para a análise dos dados de pH pelo método do Profert estão apresentados na Tabela 6. A análise é feita como se segue.

a) Identificação de *outliers*

Numa pré-análise determinam-se os *outliers* através do intervalo:

$$\left(\hat{x}_{0,75} - \frac{3}{2}\hat{\delta}; \hat{x}_{0,25} + \frac{3}{2}\hat{\delta} \right) = \left(5,65 - \frac{3}{2}0,06; 5,71 + \frac{3}{2}0,06 \right) = (5,56; 5,80)$$

Portanto, todos os valores determinados pelos laboratórios que estiverem fora deste intervalo, inclusive 5,56 e 5,80, são considerados como *outliers*. Observa-se neste exemplo (Tabela 6) que os laboratórios 4, 9, 10, 14 e 19, foram classificados por este critério como *outliers*. Em outras palavras, classificam-se como *outlier* todos os laboratórios que apresentaram um valor para discrepância $d_1 > 2$. Utilizando o mesmo critério para o solo 1, repetição 2, eliminam-se os laboratórios 4, 9, 14, e para o solo 6, tanto na repetição 1 quanto na repetição 2, elimina-se o laboratório 14. Portanto, no cálculo da exatidão estes valores devem ser desconsiderados.

b) Cálculo da exatidão (E_i)

De posse dos valores das discrepâncias atribuídas a cada laboratório fica fácil calcularmos sua exatidão. Por exemplo, tomando-se o solo 6, repetições 1 e 2, que são as duas últimas amostras do ano e os laboratórios 1 e 9, temos respectivamente:

$$E_1 = \left(1 - \frac{T_1}{18 \cdot a \cdot c} \right) \cdot 100 = \left[1 - \left(\frac{0 + 1 + 0 + 1}{18 \cdot 2 \cdot 1} \right) \right] \cdot 100 = 94,4\%$$

$$E_9 = \left(1 - \frac{T_9}{18 \cdot a \cdot c} \right) \cdot 100 = \left[1 - \left(\frac{1 + 6 + 1 + 6}{18 \cdot 2 \cdot 1} \right) \right] \cdot 100 = 61,1\%$$

Sendo que 18 se refere a soma máxima de d_1 e d_2 que variam de 0 a 9, a é o número de amostras envolvidas no cálculo que, no caso são duas, repetição 1 e 2, e c é o número de características, ou seja, apenas o pH.

A coordenação do Profert decidiu estabelecer a regra de que um laboratório exato é aquele que obtém uma exatidão maior ou igual a 67%. Portanto, no exemplo, o laboratório 1, seria exato por apresentar exatidão de 94,4% e o laboratório 9 não seria exato por apresentar um valor de 61,1% para o índice de exatidão. Estes resultados indicam que as determinações feitas pelo laboratório 1 para o solo 6, são aceitáveis e por isto ele já receberia como reconhecimento o selo prata. Já as determinações feitas pelo laboratório 9 para o solo 6, estão acima de um limite aceitável de erro, devendo seus coordenadores se empenhar para identificar as possíveis causas desses erros. A consequência disto é o não recebimento de selo.

- **Cálculo da precisão (P_i)**

O cálculo da precisão é baseado no índice de imprecisão (I), que nada mais é do que uma adaptação de um intervalo de confiança unilateral para a razão entre a variância mínima ($\hat{\sigma}_{min}^2$) e a variância observada ($\hat{\sigma}_{obs}^2$). Para que fique bem claro, ainda utilizando os dados de pH apresentados na Tabela 6, todos os cálculos necessários para se obter a precisão dos laboratórios 1 e 9 são mostrados.

- **Cálculo da variância mínima ($\hat{\sigma}_{min}^2$)**

A variância mínima ($\hat{\sigma}_{min}^2$), como já foi dito, nada mais é do que o cálculo da variância feita com as determinações que possuem soma zero para d_1 e d_2 . O cálculo é feito para cada solo, ou seja, deve-se considerar suas repetições conjuntamente, resultando em $\hat{\sigma}_{min}^2 = 0,0009$.

- **Cálculo da variância observada ($\hat{\sigma}_{obs}^2$)**

A variância observada ($\hat{\sigma}_{obs}^2$) é feita para cada laboratório com os dados de todas as amostras, independente das discrepâncias, o que resulta em $\hat{\sigma}_{obs}^2 = 0,0013$.

Calculadas estas variâncias podemos encontrar o índice de imprecisão (I) que para o laboratório 1 é:

$$I_1 = \text{Valor inteiro} \left(\frac{0,0013}{0,0009} \cdot \frac{1}{1,900466} \right) - 1 = 0$$

onde: $F_{(11,92; 0,05)} = 1,900466$. Finalmente a precisão para o laboratório 1 é:

$$P_1 = 100 - \left(\frac{0 \cdot 100}{9} \right) = 100\%$$

Proporção de coincidências

Para exemplificar a estimativa das proporções de coincidências relacionadas com a classificação da exatidão $\hat{p}_1(\%)$ e a estimativa da proporção na atribuição de selos $\hat{p}_2(\%)$, considerou-se os dados simulados para o pH, que estão apresentados na Tabela 5.

A estimativa para $\hat{p}_1(\%)$ é obtida confrontando-se os resultados do solo 1 e da repetição 1, sendo que o método Waeny (Tabela 5) considerou os laboratórios 9 e 14 como inexatos, enquanto que o método do Profert (Tabela 6) apresentou os laboratórios 4, 9, 10, 14 e 19 como inexatos. Portanto os laboratórios 4, 10 e 19 foram classificados de forma discordante, resultando em coincidência sobre 16 laboratórios.

Assim, a proporção estimada de coincidências $\hat{p}_1(\%)$, é obtida por:

$$\hat{p}_1(\%) = \frac{16}{19} \cdot 100 = 84,2\%$$

Os resultados para os demais solos estão apresentados na tabela 7.

Tabela 6: Resultado obtido pelo método do PROFERT, no confronto das determinações do Ca (Cálcio), de 6 amostras de solo, repetidas 2 vezes feitas por 19 laboratórios.

pH SIMULADO (c mol _c kg ⁻¹)																							
Solo 1						Solo 2						Solo 3											
1			2			1			2			1			2								
Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det
14	9	9	5,07	14	9	9	5,07	14	9	9	5,27	14	7	9	5,3	14	5	9	5,2	14	3	9	5,2
19	3	8	5,55	19	1	6	5,54	12	2	5	5,57	12	1	5	5,6	2	1	3	5,5	18	0	2	5,5
17	1	2	5,62	11	0	3	5,6	19	2	4	5,58	19	1	5	5,6	18	0	2	5,6	2	0	1	5,6
12	0	0	5,64	12	0	3	5,6	1	0	1	5,63	2	0	1	5,6	12	0	1	5,6	12	0	1	5,6
2	0	0	5,65	17	0	2	5,61	2	0	0	5,65	11	0	1	5,6	17	0	0	5,6	19	0	1	5,6
11	0	0	5,65	7	0	1	5,63	10	0	0	5,65	13	0	1	5,6	1	0	0	5,6	17	0	0	5,6
13	0	0	5,65	13	0	1	5,64	8	0	0	5,66	8	0	0	5,7	19	0	0	5,6	1	0	0	5,6
1	0	0	5,66	2	0	0	5,67	13	0	0	5,66	16	0	0	5,7	10	0	0	5,6	16	0	0	5,6
7	0	0	5,66	8	0	0	5,67	7	0	0	5,68	1	0	0	5,7	13	0	0	5,7	10	0	0	5,6
16	0	0	5,66	18	0	0	5,68	16	0	0	5,68	7	0	0	5,7	5	0	0	5,7	13	0	0	5,7
15	0	0	5,68	1	0	0	5,69	11	0	0	5,69	10	0	0	5,7	16	0	0	5,7	8	0	0	5,7
3	0	1	5,69	3	0	0	5,69	17	0	0	5,69	6	0	0	5,7	8	0	0	5,7	5	0	0	5,7
8	0	1	5,69	5	0	0	5,69	5	0	0	5,71	17	0	0	5,7	15	0	0	5,7	7	0	0	5,7
5	0	2	5,7	10	0	0	5,71	15	0	0	5,71	15	0	1	5,7	7	0	1	5,7	15	0	1	5,7
18	0	3	5,71	16	0	0	5,71	6	0	1	5,72	3	0	2	5,7	3	0	1	5,8	3	0	2	5,8
6	1	7	5,76	15	0	1	5,73	3	0	2	5,75	5	0	3	5,8	6	0	2	5,8	6	0	2	5,8
10	3	9	5,8	6	1	3	5,76	18	2	5	5,79	18	0	4	5,8	11	0	2	5,8	11	0	4	5,8
4	5	9	5,88	4	3	9	5,86	4	4	9	5,88	4	3	9	5,9	4	2	6	5,9	4	1	6	5,9
9	9	9	5,99	9	6	9	6,02	9	7	9	5,98	9	5	9	6	9	2	6	5,9	9	1	6	5,9

... continua ...

Tabela 6: Resultado obtido pelo método do PROFERT, no confronto das determinações do Ca (Cálcio), de 6 amostras de solo, repetidas 2 vezes feitas por 19 laboratórios.

pH SIMULADO (c molc kg ⁻¹)																													
Solo 4								Solo 5								Solo 6								Classificação					
1				2				1				2				1				2				Lab	Exat	Vobs	Vmin	Prec	Selo
Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det	Lab	d ₁	d ₂	Det						
14	7	9	5,2	14	0	0	5,2	14	5	9	5,2	14	4	9	5,22	14	6	9	5,16	14	5	9	5,15	1	94	0,0013	0,0009	100	Ouro
19	1	3	5,6	19	2	6	5,5	19	1	6	5,5	19	1	5	5,51	19	1	4	5,53	19	1	4	5,52	2	100	0,0002		100	Ouro
8	0	1	5,6	3	0	2	5,6	12	0	2	5,6	17	0	3	5,58	17	0	2	5,59	8	0	2	5,58	3	83	0,0002		100	Ouro
17	0	0	5,6	12	0	2	5,6	17	0	2	5,6	12	0	3	5,59	12	0	2	5,6	12	0	2	5,6	4	83	0,0003		100	Ouro
11	0	0	5,6	8	0	1	5,6	7	0	2	5,6	7	0	2	5,61	1	0	1	5,63	1	0	1	5,61	5	100	0,0006		100	Ouro
1	0	0	5,6	2	0	1	5,6	15	0	1	5,6	15	0	1	5,64	7	0	1	5,63	7	0	1	5,62	6	92	0,0002		100	Ouro
3	0	0	5,6	11	0	0	5,7	2	0	0	5,7	1	0	1	5,65	11	0	1	5,64	17	0	1	5,63	7	94	0,0002		100	Ouro
12	0	0	5,6	17	0	0	5,7	13	0	0	5,7	13	0	0	5,68	8	0	0	5,65	5	0	0	5,64	8	94	0,0005		100	Ouro
2	0	0	5,7	13	0	0	5,7	11	0	0	5,7	2	0	0	5,69	13	0	0	5,67	11	0	0	5,65	9	61	0,0002		100	Sem
13	0	0	5,7	1	0	0	5,7	16	0	0	5,7	16	0	0	5,72	5	0	0	5,7	13	0	0	5,69	10	86	0,0009		100	Ouro
5	0	0	5,7	16	0	0	5,7	18	0	0	5,7	11	0	0	5,74	15	0	0	5,7	15	0	0	5,69	11	97	0,0007		100	Ouro
7	0	1	5,7	5	0	0	5,7	1	0	0	5,7	5	0	0	5,76	16	0	0	5,73	16	0	0	5,69	12	89	0,0002		100	Ouro
10	0	1	5,7	15	0	0	5,7	5	0	0	5,7	8	0	0	5,77	18	0	0	5,73	18	0	0	5,71	13	100	0,0001		100	Ouro
16	0	1	5,7	10	0	0	5,7	6	0	0	5,8	18	0	0	5,77	2	0	0	5,75	2	0	0	5,74	14	19	0,0003		100	Sem
15	0	2	5,8	7	0	0	5,7	8	0	0	5,8	6	0	0	5,78	6	0	1	5,78	6	0	2	5,78	15	100	0,0003		100	Ouro
18	0	2	5,8	6	0	1	5,8	10	0	1	5,8	10	0	1	5,8	10	0	2	5,79	4	0	2	5,8	16	100	0,0006		100	Ouro
6	0	4	5,8	18	0	1	5,8	3	0	3	5,8	3	0	1	5,81	3	0	3	5,83	10	0	3	5,81	17	92	0,0004		100	Ouro
9	1	6	5,8	9	1	4	5,8	4	1	4	5,9	9	1	3	5,87	4	0	4	5,85	3	0	3	5,82	18	100	0,0005		100	Ouro
4	1	6	5,9	4	2	5	5,9	9	1	4	5,9	4	1	4	5,88	9	1	6	5,92	9	1	6	5,9	19	72	0,0005		100	Ouro

Tabela 7: Estimativas das proporções de coincidências para a classificação da exatidão dos laboratórios $\hat{p}_1(\%)$, feita pelos métodos do Profert e de Waeny, ao longo do ano, para o experimento 1 dos dados simulados para o pH.

Rep	Solo 1		Solo 2		Solo 3		Solo 4		Solo 5		Solo 6	
	1	2	1	2	1	2	1	2	1	2	1	2
$\hat{p}_1(\%)$	84,2	89,5	78,9	94,7	89,5	89,5	89,5	89,5	94,7	100,0	94,7	94,7

Para a estimativa da segunda proporção de coincidências $\hat{p}_2(\%)$, serão confrontados por ambos os métodos o experimento de pH, Ca, K e P, feitas somente no final do ano quando se tem todos os dados em mãos. Para o pH verifica-se que o método de Waeny (Tabela 5) classificou o laboratório 9 como sendo exato e preciso e atribui-se um selo ouro, enquanto que o método do Profert (Tabela 6) classificou o mesmo laboratório 9 como inexato e preciso, não dando nenhum selo ao laboratório. Como ocorreu apenas esta discordância entre os dois e, por diferença 18 coincidências, a proporção de coincidências estimada é:

$$\hat{p}_2(\%) = \frac{18}{19} \cdot 100 = 94,7\%$$

Proporções médias para coincidências

Procurou-se estudar o efeito do aumento da variabilidade do conjunto de dados, lembrando que cada experimento representa um ano de programa, estimando a proporção média de coincidências para a classificação relativa à exatidão $\tilde{p}_1(\%)$ em cada cenário, entre os métodos de Waeny e do Profert. Para os mesmos experimentos estimou-se a proporção média de coincidências devido a atribuição de selos ao final de cada ano $\tilde{p}_2(\%)$. Os resultados estão apresentados na Tabela 8.

Observando a Tabela 8, a proporção de coincidência quanto a classificação da exatidão, ou seja $\tilde{p}_1(\%)$, pode-se inferir que existe uma equivalência nos resultados fornecidos por ambas as metodologias no que diz respeito à análise de exatidão dos laboratórios. De forma ilustrativa verifica-se pela Figura 2 que as concordâncias existentes entre os métodos apresentaram elevadas proporções de coincidências quanto a classificação de exatidão, mesmo aumentando a variabilidade.

Nota-se, também, que apesar de se trabalhar com apenas 4 das 7 variáveis existentes numa análise de fertilidade, as demais poderão ser enquadradas neste estudo naturalmente pela amplitude de variabilidade envolvida.

Tabela 8: Proporção média de coincidências para a classificação da exatidão ao longo do ano $\tilde{p}_1(\%)$ e proporção média de coincidências para a atribuição de selos $\tilde{p}_2(\%)$, feita por 19 laboratórios em 30 experimentos com quatro diferentes valores para CV(%).

Atributo (CV em %)	$\tilde{p}_1(\%)$	$\tilde{p}_2(\%)$
pH (CV = 1%)	91,95	94,91
Ca (CV = 10%)	90,44	93,86
K (CV = 30%)	93,75	87,02
P (CV = 50%)	91,18	20,70

Extrapolando-se esses resultados para um programa de controle de qualidade de laboratórios de solo, no caso da avaliação da exatidão, os métodos seriam equivalentes podendo-se utilizá-los como substitutos um do outro, independentemente da variabilidade intrínseca existente no conjunto de dados.

Retornando a Tabela 8, observando-se $\bar{p}_2(\%)$, a proporção de coincidências na atribuição de selos por ambos os métodos, nota-se que para $CV > 30\%$ na variabilidade do conjunto de dados, caso das análises de P (Fósforo), ocorreu uma diminuição nas proporções de coincidência. Pela Figura 2, observa-se que principalmente no caso do P (Fósforo), que possui um $CV = 50\%$, esta proporção cai drasticamente.

Estes resultados mostram que se deve ser cauteloso no caso de uma avaliação de precisão, pois, se o conjunto de dados possuir $CV > 30\%$ os métodos estatísticos em estudo, não podem ser considerados equivalentes.

Nota-se que os métodos possuem procedimentos diferentes, o Profert por exemplo, foi desenvolvido para avaliar resultados de análises de solo. Observa-se, também que as amostras de solo possuem composição química e física altamente variável, e não se sabe qual é verdadeiramente a composição química de amostra que é distribuída para análise.

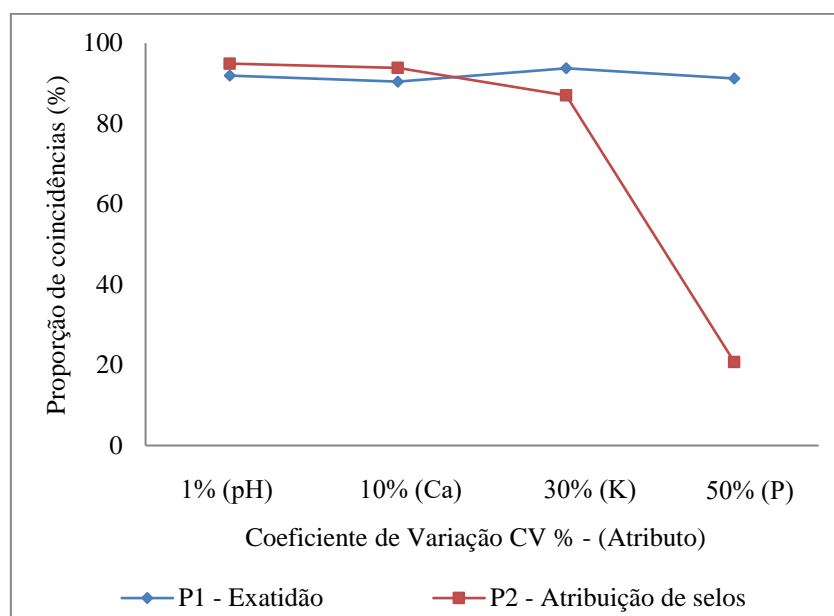


Figura 2: Comportamento da proporção média de coincidências para a classificação da exatidão $\bar{p}_1(\%)$, e a proporção média de coincidências para a atribuição de selos $\bar{p}_2(\%)$ em função do Coeficiente de Variação CV(%).

Isto dificulta qualquer avaliação feita sobre os resultados apresentados pelos laboratórios. Assim, sem o verdadeiro valor ou o valor padrão, não se poderia afirmar que valores são corretos. (Outro agravante é que, na prática, agronomicamente existem faixas de valores onde a classificação do solo é aceitável, ou considerada como boa, enquanto que matematicamente não se tem tanta flexibilidade).

Esta é a justificativa apresentada por Oliveira e Campos (1995) para a construção de um intervalo de confiança da razão entre as variâncias mínima e observada fazendo este cálculo, que na verdade é uma relação entre a variação dentro e a variação entre, procura-se levar em consideração esses aspectos, para que o método se torne o mais próximo possível da realidade.

Por outro lado, o método descrito por Waeny (1980a) foi desenvolvido como uma metodologia mais ampla e de aplicação imediata sendo utilizada para a certificação de análises

feitas essencialmente dentro da indústria. Como os materiais envolvidos possuem provavelmente composição química e física mais bem definida, a transferência dessa técnica para avaliar amostras de solo, poderia levar a uma avaliação inadequada.

Assim, olhando-se com maior atenção para o índice de repetitividade (r), nota-se que ele envolve uma amplitude padronizada e uma componente de variância que é dada pelo quadrado médio do erro. Portanto, ele considera apenas a variação existente dentro de laboratórios. Por estas considerações, supõe-se ser esta a principal causa que resulta na diferença apresentada pelos métodos na avaliação das análises quanto à precisão, já que no caso do Profert ele considera uma razão entre as variações dentro e entre laboratórios.

As análises estatísticas, no caso do método de Waeny, são feitas por algum software estatístico como o software R, por exemplo, com complementação em uma planilha de cálculos como o Excel, exigindo um certo conhecimento destes programas. Leva-se em média 15 minutos por análise, sendo necessário um conhecimento geral sobre análise de variância, componentes de variância e modelos lineares. A análise é simples, bastando comparar as amplitudes calculadas com o valor da repetitividade ou o valor da reprodutividade, dependendo do caso. Eliminam-se os valores que superam estes índices e que ao mesmo tempo possuam o maior desvio em relação à média.

Apesar da análise ser simples, o que é uma vantagem, as informações fornecidas por este método só permitem identificar se a determinação é ou não aceitável, tendo a desvantagem de não quantificar o nível de reprodutividade e de repetitividade de cada laboratório.

Para o Profert, as análises estatísticas são efetuadas inteiramente numa planilha de cálculo, o que é uma vantagem, exigindo, porém, um conhecimento mais aprofundado deste tipo de software. É indispensável um bom conhecimento da estatística básica, como: medidas de posição, medidas de dispersão, além do conhecimento de técnicas de estimação por ponto e por intervalos.

Embora a montagem da planilha do Profert seja bem mais complexa sua análise também é simples, possuindo três etapas: a) determinação dos *outliers*; b) cálculo da exatidão, obtidos com base nas discrepâncias d_1 e d_2 ; c) cálculo do intervalo de confiança para a razão de variâncias, chegando-se à precisão. O tempo gasto é de 32 minutos em média para realizar uma análise completa.

Porém, é necessário observar que o método Profert, apesar de exigir maior tempo para realizar as análises, utiliza uma teoria mais simples, além de fornecer informações mais detalhadas sobre os índices de exatidão e precisão.

CONCLUSÕES

Com base nos resultados apresentados na seção anterior, obtidos pela aplicação do método de Waeny e do método do Profert sobre cada um dos 4 cenários prefixados pelo CV, isto é, CV = 1%, CV = 10%, CV = 30% e CV = 50%, suscita as seguintes conclusões:

- a) Nos três cenários onde o conjunto de dados possuem variabilidade medida por $CV \leq 30\%$, ambos os métodos avaliados apresentaram desempenho equivalente, com proporção média de coincidências acima de 90,44% para a classificação da exatidão e acima de 87,02% para a proporção média de coincidências para atribuição de selos. Portanto, como há uma elevada proporção de coincidências, pode-se dizer que o método de Waeny e o método do Profert são métodos equivalentes quando o CV do atributo é menor ou igual a 30%, para a avaliação das análises efetuadas por laboratórios de solos.

- b) No cenário onde os dados foram simulados com $CV > 30\%$, os métodos obtiveram uma proporção de coincidências de 91,18% para a classificação da exatidão, enquanto que para a proporção média de coincidências para atribuição de selos esta concordância foi de apenas 20,70%, isto é, apesar da proporção de coincidências para a classificação da exatidão ser elevada, a concordância para a atribuição de selos foi muito pequena e, por essa razão, estes métodos não podem ser considerados equivalentes para casos onde os dados possuem $CV > 30\%$.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 17025**. Requisitos gerais para competência de laboratórios de ensaio e calibração. Rio de Janeiro, 2005

BANZATTO, D.A.; KONKRA, S.N. **Experimentação agrícola**. 4.ed. Jaboticabal: FUNEP. 2006. 237 p.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 8. ed. São Paulo: Saraiva, 2013. 548 p.

CAMPOS, P.M. **Comparação de métodos para a avaliação do controle de qualidade de laboratório de análises de solo**. 1998. 99 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

GOMES, Frederico Pimentel. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451 p.

INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA - INMETRO. **Vocabulário Internacional de Metrologia: Conceitos fundamentais e gerais e termos associados** (VIM 2012). Duque de Caxias, RJ : INMETRO, 2012. 94 p.

JURAN, J.M. **Juran na liderança da qualidade**. 2 ed. São Paulo: Pioneira, 1993, 386p. (série: novos umbrais).

MOOD, A. M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3 ed. New York: J. Wiley, 1974. 564 p

MONTGOMERY, D. C. **Introduction to statistical quality control**, 6th ed. Hoboken, NJ: Wiley, 2009. xiv 734 p.

OLIVEIRA, M.S. de; CAMPOS, P.M. Método estatístico para o controle de qualidade interlaboratorial e interlaboratorial de análise do solo. In: CONGRESSO DE INICIAÇÃO CIENTÍFICA DA ESAL/UFLA, 8 E SEMINÁRIO DE AVALIAÇÃO DO PIBIC/CNPq, 3, Lavras, 1995. **Resumos...** Lavras: UFLA, 1995. P.147.

OLIVEIRA, M. S. de; CORRÊA, J.B.; CAMPOS, P.M. Programa interlaboratorial de controle de qualidade de análise de solo do estado de Minas Gerais (PROFERT-MG). **Boletim Informativo da S.B.C.S.**, Campinas, v.20, n.1, p.18-23, jan. 1995.

OLIVIERI, J.C.; D'ELIA, A.G. Programas interlaboratoriais. **Controle da Qualidade**. São Paulo, n.50, p.5, jun. 1996.

R CORE TEAM. **R:A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing. 2017.

WAENY, J.C.C. **Programas interlaboratoriais**: introdução a metodologia e análise de resultados. São Paulo: IPT, 1980a. 24p. (Série Técnicas e Métodos, 7).

WAENY, J.C.C. **Repetitividade e reprodutividade I**. São Paulo: IPT, 1980b. 18p. (Série ACM, 18).