

## Comparison of pharmacopeial statistical methods applied in microbiological assay for antibiotics potency determination using parallel lines and three-dose level

Miriam de Fátima Vianna Leonel<sup>\*1</sup>, Cristina Duarte Vianna Soares<sup>1</sup>, Arminda Lucia Siqueira<sup>2</sup>

<sup>1</sup>Department of Pharmaceutical Products, Faculty of Pharmacy, Federal University of Minas Gerais, Brazil; <sup>2</sup>Department of Statistics, Federal University of Minas Gerais

Pharmaceutical equivalence studies, useful for checking the interchangeability of generic medicines and their respective innovator medicines, have been carried out in Brazil since 1999, as a consequence of the establishment of the generic medicine policy. For medicines containing antibiotics, microbiological assays are often the most appropriate method. However, the statistical methods applied in these assays are not widely known due to the difficult access to official codes and/or little knowledge of the statistical tools of analysis. Thus, the aim of this work was to compare the statistical methods for determining the potency of antibiotics through the cylinder-plate method using parallel lines and a three-dose level model, as described in the Brazilian Pharmacopeia (4th edition, 1988), British Pharmacopoeia 2011, European Pharmacopoeia (7th edition), The International Pharmacopoeia (4th edition), and United States Pharmacopoeia (34th edition). The assay is illustrated with the antibiotic ofloxacin, and details on orthogonal coefficients, normality test, homogeneity of variance test, and detection of outliers are discussed. The calculations obtained by statistical analyses from different pharmacopeias lead to the same final interpretation. In practice, with the availability of alternative pharmacopeial methods, the analyst can choose the most appropriate statistical calculation to be used.

**Uniterms:** Antibiotics/potency. Parallel lines model. Pharmacopeial statistical methods.

Os estudos de equivalência farmacêutica, úteis na verificação da intercambialidade entre os medicamentos genéricos e respectivos medicamentos de referência, têm sido realizados no Brasil desde 1999, como consequência do estabelecimento da política de medicamentos genéricos. Para medicamentos contendo antibióticos, os ensaios microbiológicos são, muitas vezes, o método mais adequado. Entretanto, os métodos estatísticos aplicados nesses ensaios não são amplamente conhecidos devido à dificuldade de acesso aos compêndios oficiais e/ou pouca compreensão das ferramentas estatísticas de análises. Portanto, o objetivo desse trabalho foi comparar os métodos estatísticos para determinação de potência de antibióticos pelo delineamento por retas paralelas e três níveis de doses, descritos nas farmacopeias Brasileira 4. ed. (1988), Britânica 2011, Europeia 7. ed., Internacional 4. ed. e na Farmacopeia dos Estados Unidos 34. ed. (2011). O ensaio é exemplificado com o antibiótico ofloxacina e detalhes sobre coeficientes ortogonais, teste de normalidade, teste de homogeneidade de variância e detecção de outliers são discutidos. Os cálculos obtidos pelas análises estatísticas segundo as diferentes farmacopeias resultaram na mesma interpretação final. Na prática, métodos farmacopéicos alternativos permitem ao analista a escolha do cálculo estatístico mais apropriado a ser utilizado.

**Unitermos:** Antibióticos/potência. Delineamento por retas paralelas. Métodos estatísticos farmacopéicos.

**\*Correspondence:** M. F. V. Leonel. Departamento de Produtos Farmacêuticos, Faculdade de Farmácia – UFMG. Av. Pres. Antônio Carlos, 6627 - Pampulha - 31270-901 - Belo Horizonte - MG, Brasil. E-mail: vileonel@yahoo.com.br

## INTRODUCTION

Microbiological assays on antibiotics are used in pharmaceutical preparations and raw materials whose contents are related to their biological activity and usually cannot be determined by chemical analyses. Typically, antibiotic drugs have their potency determined by comparison with a standard substance or a biological reference substance. Such potency is expressed as the ratio of the dose that inhibits the growth of a sensitive microorganism, relative to the standard substance dose that produces similar inhibition. The agar diffusion or cylinder-plate method used consists of a dose-response model in which the antibiotic concentration is proportional to the inhibition zone of microorganism growth.

The statistical analysis of such experiments requires the establishment of a mathematical relationship of the mean responses ( $y$ ) with the crescent concentrations (treatments or doses) of the antibiotic. The most commonly used model is the parallel lines, in which every reference standard (S) and unknown (U) preparation should be tested with the same number of doses. Normally, two or three doses per preparation are used, and the ratio of adjacent doses should be constant for all treatments of the assay.

The potency calculation is based, in this case, on the analysis of two straight lines, one referring to the standard responses and the other to the unknown responses. They are obtained by the relationship between their responses and the decimal logarithm ( $\log_{10}$ ) or natural logarithm ( $\ln$ ) of each of the doses (Hewitt, 2004; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). The least squares method (Farias et al., 2003) is most commonly used to check the linearity of the curves, and, therefore, the same number of reference standard and unknown responses is used ( $S_1, S_2, S_3$ , and  $U_1, U_2, U_3$ ). This method is still based on the assumption that the responses ( $y$ ) follow the normal distribution (Gaddum, 1953; Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011) and have a constant variance (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011), and the errors are independent (random) (Farmacopeia Brasileira, 1988).

Evaluating these experiments involves the application of statistical calculations currently described in pharmacopoeias. However, these calculations are not understood by analysts, primarily due to the unavailability of official codes in the pharmaceutical industries. Moreover, a lack of specialists in the area, together with little knowledge of the statistical tools of analysis, has also been observed. Thus, an investment in qualified personnel is extremely

important for the pharmaceutical industry. The success of the experimental antibiotics assay design requires a knowledge of various pharmacopoeial methods, a familiarity with statistical hypothesis testing, and a correct interpretation of the validity parameters. It is worth emphasizing that microbiological assays of antibiotics are performed, among others, in studies of pharmaceutical equivalence for registration or renewal of generic medicines. In Brazil, with the advent of the generic drug policy established in 1999, the regulatory agency admits many foreign pharmaceutical codes as official references for the quality control of raw materials and pharmaceuticals, in the absence of a monograph in the Brazilian Pharmacopoeia.

### Normality test

A test to check the normality of the response variable ( $y$ ) is recommended in the British Pharmacopoeia (BP) (British Pharmacopoeia, 2011), European Pharmacopoeia (EP) (European Pharmacopoeia, 2011), Brazilian Pharmacopoeia (FB) (Farmacopeia Brasileira, 1988), International Pharmacopoeia (IP) (The International Pharmacopoeia, 2011), and United States Pharmacopoeia (USP 34) (United States Pharmacopoeia, 2011). In BP and EP, the work by Shapiro and Wilk is recommended (Shapiro, Wilk, 1965) for verification of the test and the calculations. Consider  $y_1 < y_2 < \dots < y_n$  the ordered sample responses of size  $n$  in which  $i = 1, 2, \dots, n$  and  $y_1$  is the smallest value. The statistic is defined as

$$W = b^2 / \sum_{i=1}^n (y_i - \bar{y})^2. \quad (1)$$

The constant  $b$  is calculated by

$$b = \sum_{i=1}^n a_{n-i+1} (y_{n-i+1} - y_i), \quad (2)$$

in which  $a_i$  is tabulated (for  $n \leq 50$ ) in the referred citation. For normal distribution, the value of  $W$  should be close to 1.

Another test of normality is known as Anderson-Darling ( $AD$ ) (Minitab, 2010), which is not mentioned in the pharmacopoeias. It is based on the cumulative distribution function (cdf) of the empirical data to be analyzed. In the calculation, the weighted quadratic distance is used between the line adjusted of the probability plot (based on the selected distribution) and nonparametric function. The statistic is defined as

$$AD = -n - (1/n) \sum_{i=1}^n (2i-1) \left[ \text{inv}F(y_i) + \text{inv}(1-F(y_{n-i+1})) \right], \quad (3)$$

in which  $F$  is the cumulative distribution function of the

normal distribution of the response  $y_i$  sorted in ascending order.

In a scale from 0 to 1, a small  $AD$  value indicates that normal distribution fits best the data. In cases in which the normality condition is not met, the transformation of the response ( $y$ ), e.g.,  $\log_{10}y$  or  $\ln y$ ,  $y^{1/2}$  and  $y^2$ , is suggested to achieve compliance with this condition (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011).

Another quantitative measure for reporting the result of the  $AD$  normality test is the  $p$ -value. The significance probability value ( $p$ ) is calculated from the value obtained from  $AD'$  where

$$AD' = AD (1 + 0.75/n + 2.25/n^2). \quad (4)$$

The calculations are implemented in some software packages (Minitab, 2010) according to the following equations:

$$13 > AD' > 0.600, \\ p = \exp [1.2937 - (5.709 \times AD') + (0.0186 \times AD'^2)] \quad (5)$$

$$0.600 > AD' > 0.340, \\ p = \exp [0.9177 - (4.279 \times AD') - (1.38 \times AD'^2)] \quad (6)$$

$$0.340 > AD' > 0.200, \\ p = 1 - \exp [-8.318 - (42.796 \times AD') - (59.938 \times AD'^2)] \quad (7)$$

$$AD' < 0.200, \\ p = 1 - \exp [-13.436 + (101.14 \times AD') - (223.73 \times AD'^2)] \quad (8)$$

## Homogeneity of variance

The homogeneity of the variance (homocedasticity) test is cited in BP and EP. Its implementation is also recommended by Bartlett (Bartlett, 1937) and Cochran (Cochran, 1951). According to Cochran, the test is based on the measure of the variances ( $s_i^2$ ,  $i = 1, \dots, k$ ) between  $k$  treatments considered as different treatments or doses, of size  $n_i$ , and is calculated by

$$C_{\text{Cochran}} = s_{\text{maximum}}^2 / \sum_{i=1}^k s_i^2. \quad (9)$$

The sizes of all  $n_i$  groups are assumed to be equal when the calculated  $C$  is less than the critical value for a given significance level (Fisher, Belle, 1993).

## Detecting outliers

If a response is doubtful or discrepant, it should be

investigated. A good way to detect such observations is by means of graphical inspection, such as a scatter plot and/or box plot (Ludbrook, 2008). In BP and EP the calculation of the variance of responses in each treatment group is suggested for the verification of outliers. FB and USP 34 recommend the following tests.

### Criterion 1 – Based on the variation of group with supposedly equivalent responses

For normally distributed data, different tests are available to identify outliers; these are sometimes called  $Q$ -tests (Burke, 2001). Other tests, such as the Dixon test (Dixon, 1950; Dixon, 1951; Bliss, 1956; Miller, 1993; Massart *et al.*, 1997), are based on the calculation of the experimental  $Q$  (or  $r_{10}$  in Dixon's notation) (Efsthathiou, 2006), defined as the ratio given by the distance of the suspect value from its nearest neighbor divided by the range of the values. These values are compared with tabulated critical values. Similar to the Shapiro-Wilk test, responses should be sorted in ascending order. In FB and USP 34, the statistic is called  $G$ .

$$G_1 = (y_2 - y_1) / (y_n - y_1), \text{ for } n = 3 \text{ to } 7 \quad (10)$$

$$G_2 = (y_3 - y_1) / (y_{n-1} - y_1), \text{ for } n = 8 \text{ to } 13 \quad (11)$$

$$G_3 = (y_3 - y_1) / (y_{n-2} - y_1), \text{ for } n = 14 \text{ to } 24 \quad (12)$$

If  $G_1$ ,  $G_2$ , or  $G_3$  exceeds the critical value, for a given significance level, such response can be eliminated. Otherwise, the suspect value is maintained (Farmacopeia Brasileira, 1988; United States Pharmacopoeia, 2011).

### Criterion 2 – Based on $k$ total groups

This criterion evaluates the variation of responses in each treatment (Bliss *et al.*, 1956). The biggest difference or range should be divided by the sum of all differences and must not exceed the critical value ( $R$ ) (Farmacopeia Brasileira, 1988; United States Pharmacopoeia, 2011) for  $k$  number of doses or treatments and  $n$  replicates.

Given the presence of outliers, the following decisions can be taken: (i) to repeat the experiment, (ii) to reduce the number of observations by eliminating the entire block where the response was lost (Hewitt, 1977), (iii) to replace the lost response, at most one in each treatment, or up to 5% (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011) or 10% (The International Pharmacopoeia, 2011; United States Pharmacopoeia, 2011) of total responses, and recalculate the values. The replacement of missing values is also mentioned by other authors (Bliss, 1956; Finney,

1978; Ludbrook, 2008). Let  $T_r'$  be the sum of all responses to the block (plate) in which the value is lost;  $T_i'$  be the sum of all treatment responses, for which the value is lost;  $T'$  be the total sum of the observations;  $f$  be the number of plates; and  $k$  be the number of treatments or doses. The value to be replaced is calculated by

$$y' = (fT_r' + kT_i' - T') / (f-1)(k-1). \quad (13)$$

### Validity test: analysis of variance

The analysis of variance (ANOVA) performed for calculations of antibiotic potency aims to verify the validity of the assay, to calculate the residual error ( $s^2$ ) (Farmacopeia Brasileira, 1988), and to distinguish the various sources of error (variation) affecting the responses that are not due to the gradual increase of doses. These sources can be deviations from parallelism and linearity (or curvature) that can occur between the reference standard and unknown responses, in addition to differences between preparations, blocks, and doses.

The responses are registered for each dose and for each preparation that comprises the total responses. The variation of responses due to different doses is calculated and subtracted from the sum of squares (SS) for each block to obtain the residual error. Thus, the total sum of squares is equal to the SS of the dose, between blocks and within the dose (error).

To check the significance of the sources of variation, SS is divided by the corresponding degree of freedom (d.f.) to obtain the mean square (MS), which in turn is divided by the residual variance (mean squared error or  $s^2$ ) to obtain the variance ratio (or  $F$  value). The  $F$  value calculated is compared with the critical limit obtained from Fisher's statistical table (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011) for significance levels of 5% ( $\alpha = 0.05$ ) or 1% ( $\alpha = 0.01$ ) and depends on the source of variation. If the calculated  $F$  value is greater than the critical value, the source of variation is considered significant for the significance level adopted.

Additionally to the  $F$  statistic, the probabilities or  $p$ -values are recommended (Hewitt, 2004; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). The procedures for the calculation of  $p$ -values are shown in general method of BP (A632, supplementary chapter IV G) and EP (general method 5.3, statistical analysis of results of biological assays and tests). The significance level ( $\alpha$ ) generally is 0.05 or 0.01, and for a  $p$ -value result less than  $\alpha$ , the null hypothesis must be rejected. One way to evaluate  $p$  is first to calculate the cdf value of the data and then subtract it from one ( $1 - \text{cdf}$ ) (Fisher; Yates, 1971).

To check the validity of the test, the regression is studied through the precision of the slope ( $C$ ), in which a test with a well-defined coefficient should be close to 1 (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). The use of the statistic  $g$  (The International Pharmacopoeia, 2011; United States Pharmacopoeia, 2011) is also described as a positive number and should have a value less than 1 or less than 0.1, indicating highly significant regression. The statistics  $C$  and  $g$  are related by the expression  $g = (C - 1)/C$  or  $C(1 - g) = 1$ . Table I shows the calculations according to different pharmacopeias.

To verify the linearity deviation, the parameter is expressed as two sources of variation in the pharmacopeias: the first, in the same direction (quadratic), and the other, in the opposite direction (quadratic difference). In BP and EP, from year 2000 on, these parameters were grouped into one, under the title non-linearity, requiring two degrees of freedom for its calculation.

The calculations according to the methods of IP and USP 34 are based on the use of orthogonal coefficients (Hewitt, 1977) to obtain the sum of squared deviations, which are attributed to several sources of variation. The coefficients are derived from the table of orthogonal polynomials, e.g., as in Fisher and Yates (1971). Statistical analysis is exemplified in papers published by Bliss (1956) and Hewitt (1977). This method, easily applied, is used in balanced assay, and the coefficients used are applied when the logarithms of the doses are equally spaced. One of the features of such coefficients is that each line represents the contrast between individual responses, e.g., for the regression line, the sum of their coefficients,  $-1 + 0 + 1 - 1 + 0 + 1$ , is zero. Each pair of contrasts is orthogonal. The multiplication of coefficients for the regression and parallelism (1, 0, -1, -1, 0, 1) is -1, 0, -1, 1, 0, 1, whose total is zero. The divisors ( $e_i$ ) are the sums of the squares of the coefficients of individual responses. Thus,  $(-1)^2 + (0)^2 + (1)^2 + (-1)^2 + (0)^2 + (1)^2 = 4$ .

The sums of the products of the coefficient by their corresponding treatment are defined as  $T_i$ . Then, its squared value is divided by the number of replicates, and its respective divisor ( $e_i$ ) for  $a$ ,  $b$ ,  $ab$ ,  $q$ , and  $a$ , where  $a$  is the difference in the response to the reference standard and to the unknown;  $b$ , in this case, is the difference between all highest doses and all lowest doses (slope);  $ab$  is the difference in slope between the reference standard and the unknown (if the curves are parallel);  $q$  is the common curvature of both preparations (if the curvatures have the same direction);  $aq$  is the contrast between curvatures of the two preparations (if the curvatures have opposite direction); and  $f$  is the number of replicates or block (Petri dishes).

**TABLE I** - Formula for calculating the precision of the slope (*C* or *g*) used by the Brazilian Pharmacopeia (FB), The International Pharmacopoeia (IP), United States Pharmacopeia (USP 34), British Pharmacopoeia (BP), and European Pharmacopoeia (EP)

Parameter	Pharmacopeial code	Formula		
<i>C</i>	FB	$C = \frac{E}{E - s^2 t^2} \quad E = \frac{(L_S + L_U)^2}{2nh}$		
	BP and EP	$C = \frac{SQ_{reg}}{SQ_{reg} - s^2 t^2}$	$SQ_{reg} = \frac{1}{h} H_L (L_S + L_U)^2$	$H_L = \frac{12n}{d^3 - d}$
	USP 34	$C = \frac{T_b^2}{T_b^2 - e_b f s^2 t^2}$		
<i>g</i>	IP	$g = \frac{s^2 t^2}{b^2 \sum xx}$	$\sum xx = \sum_{i=1}^n (x_i - \bar{x})^2$	$b = \frac{E}{i} \quad E = 1/4 (U_3 - U_1 + S_3 - S_1)$

$L_S$  and  $L_U$ , linear contrasts to the reference standard ( $S_3 - S_1$ ) and unknown ( $U_3 - U_1$ ) preparation;  $t$ , Student's  $t$  for  $n$  degrees of freedom (d.f.) in  $s^2$ ;  $n$  and  $f$ , number of replicates or number of Petri dishes or blocks;  $d$ , number of doses;  $T_b$ , sums of the products of the coefficient  $e_b$  by their corresponding treatment;  $e_b$ , sums of the squares of the orthogonal coefficients of individual responses in Table VII;  $i$ , interval in logarithms between successive log-doses, the same for both reference standard and unknown;  $h$ , number of preparations including reference standard.

For the validity test, each of the rows  $ab$ ,  $q$ , and  $aq$  values should not exceed three times the variance value ( $s^2$ ). Otherwise, if the ratio of any of the three values is greater than 3, then  $F_3$  should be calculated by

$$F_3 = \sum_{ab,q,aq} (T_i^2 / e_i f) / 3s^2, \quad (14)$$

where

$$s^2 = [ \sum y^2 - (\sum T_r^2 / k) - (\sum T_i^2 / f) + T^2 / N ] / n \quad (15)$$

and  $N$  is the total number of responses;  $n$ , in this case, is the number of degrees of freedom for residual error;  $k$  is the number of treatments;  $T_r$  is the sum of all responses to the block (plate);  $T_i$  is the sum of all treatment responses; and  $T$  is the total sum of the observations.

According to USP 34,  $F_3$  is applicable to a balanced assay with three or four dose levels. To test validity, the calculated  $F_3$  value must not exceed the critical value (United States Pharmacopeia, 2011), considering the number of degrees of freedom for  $s^2$ . If this condition is satisfied, the precision of the slope  $C$  is then calculated, as previously reported (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011).

**Potency**

The ratio of a given reference standard dose to the corresponding unknown dose is designated by  $R$ . The

logarithm of the ratio of potency of the unknown, in quantities assumed to be equal to those of the reference standard, is designated by  $M'$  and the mathematical equations according to pharmacopeias USP 34, FB, IP, BP, and EP are  $(ci/Ta/Tb)$ ,  $(\bar{y}_A - \bar{y}_P / b)$ ,  $(F/b)$ ,  $((A - P)/3b)$ ,  $(ci/Ta/Tb)$ , respectively.

Ideally,  $M'$  should not significantly differ from zero. The logarithm of the ratio of relative potency or estimated potency is obtained by

$$M = M' + \log R \quad (16)$$

or

$$\text{Potency (P)} = \text{antilog } M = \text{antilog } M' \times R \quad (17)$$

**MATERIALS AND METHODS**

The statistical calculations described in many pharmacopeias (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011; The International Pharmacopoeia, 2011; United States Pharmacopeia, 2011) were applied to the results of a microbiological assay of the antibiotic ofloxacin in tablet preparation using the cylinder-plate method (parallel lines with three reference standard and unknown doses). The dilutions of the unknown (U), tablets labeled potency claimed 400 mg (R), and reference standard (S) were prepared at concentrations of (1) 20  $\mu\text{g mL}^{-1}$ , (2) 30  $\mu\text{g mL}^{-1}$ , and (3) 45  $\mu\text{g mL}^{-1}$ . Each of the six doses was placed once in each block (Petri dish), up to a total

**TABLE II** – Diameters of inhibition zone (mm) of the ofloxacin microbiological assay using three-dose level (3x3)

Plate	Doses ( $\mu\text{g mL}^{-1}$ )					
	S <sub>1</sub> (20)	S <sub>2</sub> (30)	S <sub>3</sub> (45)	U <sub>1</sub> (20)	U <sub>2</sub> (30)	U <sub>3</sub> (45)
1	14.80	17.93	19.76	15.39	17.35	20.72
2	15.13	18.33	20.57	14.98	17.72	21.39
3	15.57	18.22	21.55	15.02	18.52	20.67
4	15.61	18.51	20.73	15.12	18.14	20.59
5	15.24	18.47	20.72	15.26	18.23	20.87
6	15.18	18.19	20.63	14.64	18.12	21.23
7	15.14	17.69	20.38	14.80	17.78	20.74
Total	106.67	127.34	144.34	105.21	125.86	146.21
Mean	15.24	18.19	20.62	15.03	17.98	20.89
Range	0.81	0.82	1.79	0.75	1.17	0.80

S, reference standard; U, unknown.

of seven Petri dishes. Table II describes the diameters of inhibition zones ( $y$ ), in millimeter (mm), in addition to the total treatment values (columns), means, and range (difference between maximum and minimum) for each treatment.

To evaluate the results, preliminary calculations are presented for the purpose of preparing the tables required to complete the results. Calculations and graphs were performed using Excel and Minitab (Minitab, 2010).

### CALCULATION AND RESULTS

#### Normality test

First, for the Shapiro-Wilk method, the treatment columns (doses) were sorted in ascending order (Table III). For the low-reference standard dose (S<sub>1</sub>), the sorted responses were 14.80, 15.13, 15.14, 15.18, 15.24, 15.57, and 15.61, and the sum of the squared deviations of the responses was 0.4651. In the third column, the responses were sorted in descending order,  $n-i+1$ , until the fourth observation, and, in the fourth column data, are shown in ascending order until the fourth observation. The sum of the last column represents the constant  $B$ , i.e., the sum of the values obtained by multiplying the coefficients  $a_i$  by the difference between  $y_{n-i+1}$  and  $y_i$ . (0.6523). The coefficient  $b^2$  is 0.4255 and the statistic  $W$ , applying equation (1), is

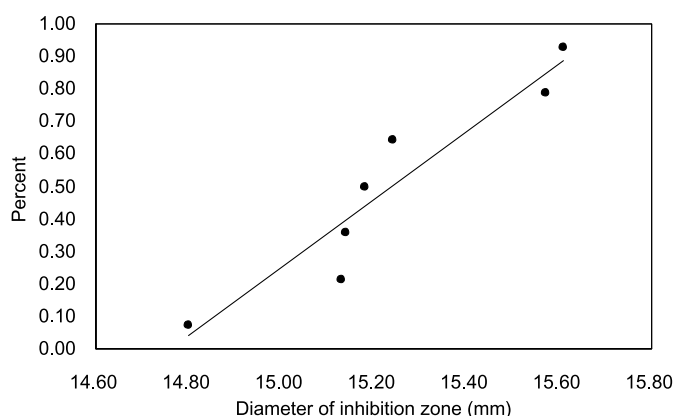
$$W = 0.4255/0.4651 = 0.915.$$

The graph of normal probability for the S<sub>1</sub> dose was obtained and is shown in Figure 1.

**TABLE III** - Sorted results for calculation of coefficient  $b$  of the reference standard dose responses, S<sub>1</sub>, for the ofloxacin microbiological assay

$i$	$n - i + 1$	$y_{n-i+1}$	$y_i$	$a_i$	$a_i(y_{n-i+1} - y_i)$
1	7	15.61	14.80	0.6233	0.5049
2	6	15.57	15.13	0.3031	0.1334
3	5	15.24	15.14	0.1401	0.0140
4	4	15.18	15.18	0.0000	0.0000

$i$ , ordered  $i$ th response;  $n$ , number of observations;  $y$ , responses;  $a_i$ , coefficient used in the  $W$  statistic.



**FIGURE 1** - Chart of normal probability of the S<sub>1</sub> dose (diameter of inhibition zone, mm) at 20  $\mu\text{g mL}^{-1}$ , for the ofloxacin microbiological assay

After the statistical value  $AD$  was calculated, the  $p$ -value was calculated for the other treatment groups. Initially, cdf values were calculated, then, the  $p$ -value was obtained by subtracting  $(1 - \text{cdf})$ .

**TABLE IV** - Statistics obtained for each dose for the ofloxacin microbiological assay, using three-dose level (3x3)

Parameter	Doses ( $\mu\text{g mL}^{-1}$ )					
	S <sub>1</sub> (20)	S <sub>2</sub> (30)	S <sub>3</sub> (45)	U <sub>1</sub> (20)	U <sub>2</sub> (30)	U <sub>3</sub> (45)
$\sum_{i=1}^n (y_i - \bar{y})^2$	0.4651	0.5189	1.6868	0.3982	0.9038	0.5558
$b^2$	0.4254	0.4818	1.5827	0.3723	0.8721	0.4750
$W$	0.915	0.928	0.938	0.935	0.965	0.854
$AD$	0.380	0.271	0.439	0.132	0.225	0.523
$p$	0.295	0.548	0.201	0.960	0.712	0.117
$s^2$	0.0775	0.0865	0.2811	0.0664	0.1506	0.0926
$G_{1a}$	0.407	0.293	0.346	0.213	0.316	0.100
$G_{1b}$	0.049	0.049	0.458	0.173	0.248	0.200

S, reference standard; U, unknown;  $b^2$ , square of the constant used in the Shapiro-Wilk test ( $W$ );  $AD$ , Anderson-Darling test statistic;  $p$ , Anderson-Darling test  $p$ -value;  $s^2$ , variance;  $G_{1a}$ , Dixon's statistic for the first observation;  $G_{1b}$ , Dixon's statistic for the second observation.

$AD'$  was estimated according to equations (4) and (6), taking the value of statistic  $AD$  (0.439), for the S<sub>3</sub> dose.

$$AD' = 0.439 (1 + 0.75/7 + 2.25/49) = 0.5062$$

$$p = \exp [0.9177 - (4.279 \times 0.5062) - (1.38 \times 0.5062^2)] = 0.201$$

### Homogeneity of the variance test

Since each treatment in the assay is a sampling value, the variances for each treatment group were calculated. According to equation (9), the following was obtained:

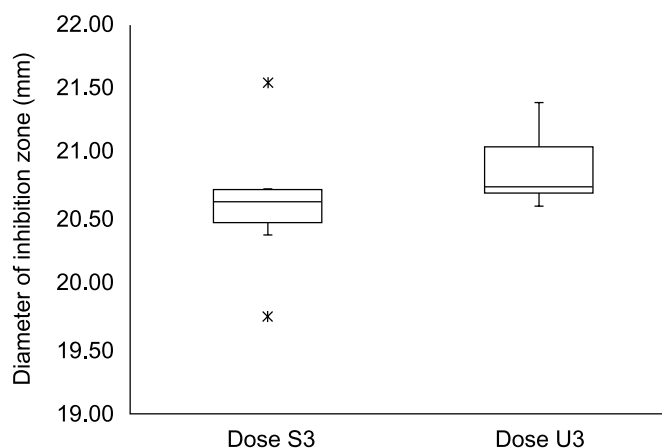
$$C_{\text{Cochran}} = 0.0775/0.7547 = 0.103$$

Table IV summarizes the results of  $\sum_{i=1}^n (y_i - \bar{y})^2$ ; constant  $B^2$ ; statistics  $W$ ,  $AD$ ,  $G$ ; probability  $p$ ; and variance ( $s^2$ ) for all treatment groups in the assay.

### Detecting outliers

To detect outliers by criterion 1, the responses were sorted in ascending order. The maximum difference (range) was then calculated for each treatment. It was observed that for the high reference standard dose (S<sub>3</sub>) there were responses that could be considered possible outliers, e.g., values 19.76 mm and 21.55 mm. In the box plot of Figure 2, one can notice the presence of such two points for the S<sub>3</sub> dose.

Hence, the ordering of data for verification, in mm, is 19.76, 20.38, 20.57, 20.63, 20.72, 20.73, 21.55. Considering the first observation as a suspect value (lower value), 19.76 mm,  $G_1$  ( $n = 7$ ) was estimated, according to equation (10):



**FIGURE 2** - Box plot of the S<sub>3</sub> (reference standard) and U<sub>3</sub> (unknown) doses at 45  $\mu\text{g mL}^{-1}$ , for the ofloxacin microbiological assay.

$$G_1 = (20.38 - 19.76) / (21.55 - 19.76) = 0.346.$$

For the second suspect value (higher value), 21.55,  $G_1$  is

$$G_1 = (21.55 - 20.73) / (21.55 - 19.76) = 0.458.$$

Table IV shows the calculated values of  $G_1$  for high S<sub>3</sub> dose and for the other treatment groups.

For verification by criterion 2, considering that the S<sub>3</sub> dose showed the largest variation, the  $R$ -value was calculated by the ratio of the maximum range and the sum of all ranges.

$$R = 1.79/6.14 = 0.292$$

## Replacing outliers

The calculation for the replacement of the responses  $y = 19.76$  to  $21.55$  of dose  $S_3$  values, identified as outliers in the box plot (Figure 2), was carried out. The responses resulted in the order (---; 20.57; ---; 20.73; 20.72; 20.63; 20.38), excluding the outliers.

The average of the remaining five responses was 20.61, and this value was inserted into the first gap to calculate the third gap. The new order is now (20.61; 20.57; ---; 20.73; 20.72; 20.63; 20.38), and the sum of total responses is 123.64.

Applying equation (13), the value of the third gap was recalculated:

$$y' = [(7 \times 88) + (6 \times 123.64) - 734.93] / [(7 - 1)(6 - 1)] = 20.76,$$

for which  $T_r' = 88$  ( $109.55 - 21.55$ );  $T_i' = 123.64$  ( $145.19 - 21.55$ ); and  $T' = 734.93$ ;  $f = 7$ ;  $k = 6$ .

Eliminating the value added in the first column (average 20.61) in place of the response  $y = 19.76$ , the following sequence (---; 20.57; 20.76; 20.73; 20.72; 20.63; 20.38) was obtained. To calculate the new value of the first gap, and again applying the pharmacopeial calculations, this value became

$$y' = [(7 \times 86.19) + (6 \times 123.79) - 735.08] / [(7 - 1)(6 - 1)] = 20.37.$$

The  $S_3$  treatment values, after replacement of discrepant responses, became (20.37; 20.57; 20.76; 20.73; 20.72; 20.63; 20.38).

However, for calculations of ANOVA and potency of the unknown, the responses 19.76 and 21.55 were not replaced. The two observations have not been pointed out as outliers in the Dixon test (criterion 1). Additionally, the responses were found to be homoscedastic in the test of variance homogeneity.

## Analysis of variance

Tables V and VI show the results of the ANOVA performed according to FB, BP and EP respectively, considering all possible sources of variation.

Table VII shows the results of the sum of squares allocated to sources of variation due to preparation, regression, parallelism, and curvature in the same direction and opposite direction, following calculations using the orthogonal coefficients recommended by IP and USP 34.

The total values ( $T_i^2/e_i f$ ) for  $ab$ ,  $q$ , and  $aq$  were 0.396, 0.188 and 0.135, respectively. The ratios of each total to the variance (0.096), as in equation (15), were 4.125 (0.396/0.096), 1.958 (0.188/0.096), and 1.406 (0.135/0.096). Considering that the ratio of total value of  $ab$  to the variance (4.125) was greater than 3,  $F_3$  was calculated according to equation (14):

$$F_3 = (0.396 + 0.188 + 0.135) / (3 \times 0.096) = 2.50.$$

## DISCUSSION

### Normality test

The values of  $W$  statistics, according to the Shapiro-Wilk test, were greater (values near 1) than the critical

**TABLE V** - Analysis of variance (ANOVA) for the ofloxacin microbiological assay according to the Brazilian Pharmacopeia (FB)

Source of variation	d.f. <sup>a</sup>	MS <sup>b</sup>	$F^c$	Limiting value <sup>d</sup>	Limiting value <sup>e</sup>	$p$ -value <sup>f</sup>
Preparation	1	0.027	0.28	<4.17	-	0.598
Regression	1	221.035	2.300	-	>7.56	0.000
Non-parallelism	1	0.396	4.12	<4.17	-	0.051
Quadratic curvature	1	0.188	1.95	<4.17	-	0.173
Difference of quadratic	1	0.135	1.41	<4.17	-	0.244
Treatments (between doses)	5	44.356	457.70	>2.53	-	0.000
Blocks	6	0.274	2.85	-	<3.47	0.026
Error (within doses)	30	0.096	-	-	-	-
Total	41					

a: d.f., degrees of freedom; b: MS, mean square; c:  $F$ , variance ratio; d: critical  $F$ -value at significance level 0.05; e: critical  $F$ -value at significance level 0.01; f:  $p$ , calculated probability.



**TABLE VI** - Analysis of variance (ANOVA) for the ofloxacin microbiological assay according to the British Pharmacopoeia (BP) and the European Pharmacopoeia (EP)

Source of variation	d.f. <sup>a</sup>	MS <sup>b</sup>	<i>F</i> <sup>c</sup>	Limiting value <sup>d</sup>	Limiting value <sup>e</sup>	<i>p</i> -value <sup>f</sup>
Preparation	1	0.027	0.28	<4.17	-	0.601
Regression	1	221.032	2300.91	-	>7.56	0.000
Non-parallelism	1	0.396	4.12	<4.17	-	0.051
Non-linearity <sup>g</sup>	2	0.161	1.68	<3.32	-	0.203
Treatments (between doses)	5	44.356	457.70	>2.53	-	0.000
Blocks	6	0.275	2.86	-	<3.47	0.025
Error (within doses)	30	0.096	-	-	-	-
Total	41					

a: d.f., degrees of freedom; b: MS, mean square; c: *F*, variance ratio; d: critical *F*-value at significance level 0.05; e: critical *F*-value at significance level 0.01; f: *p*, calculated probability; g: same direction and opposite direction.

**TABLE VII** - Coefficients and orthogonal contrasts for the ofloxacin microbiological assay, according to the International Pharmacopoeia (IP) and the United States Pharmacopoeia (USP 34)

Source of variation	Orthogonal coefficients						<i>e</i> <sub><i>i</i></sub>	<i>T</i> <sub><i>i</i></sub>	<i>T</i> <sub><i>i</i></sub> <sup>2</sup> / <i>e</i> <sub><i>i</i></sub> <sup><i>f</i></sup>
<i>a</i>	-1	-1	-1	1	1	1	6	-1.07	0.027
<i>b</i>	-1	0	1	-1	0	1	4	78.67	221.035
<i>ab</i>	1	0	-1	-1	0	1	4	3.33	0.396
<i>q</i>	1	-2	1	1	-2	1	12	-3.97	0.188
<i>aq</i>	-1	2	-1	1	-2	1	12	3.37	0.135

*a*, difference between reference standard and unknown; *b*, difference between all highest and lowest doses (slope); *ab*, difference between slopes; *q*, common curvature of both preparations; *aq*, contrast between curvature of the two preparations; *e*<sub>*i*</sub>, sum of squares of coefficients in each row; *T*<sub>*i*</sub>, sum of the products of individual coefficients in each row with their corresponding total treatment; *f*, number of replicates or block (Petri dishes).

value (*W* is 0.803 for  $\alpha = 0.05$ ) (Shapiro & Wilk, 1965). The chart of normal probability (Figure 1) represents the responses for dose *S*<sub>1</sub> (*x*-axis) versus percentages (*y*-axis) of the values that are less than or equal to them, along the reference line (central line). The plotted points form a reasonably straight line and the *p*-value for statistics *AD* was 0.380. A normal distribution with a mean of 15.24 and a standard deviation of 0.2784 appeared to fit the data well. The probability *p* calculated for all treatment groups was greater than the significance level  $\alpha = 0.05$ , as shown in Table IV. Thus, the responses were not transformed. One can state that the unknown preparation comes from a normal population ( $p > 0.100$ ), at a significance level of 5%.

### Homogeneity of variance test

The statistic value according to Cochran ( $C = 0.372$ )

was lower than the critical value ( $C$  is 0.418 for d.f. = 6 and  $\alpha = 0.05$ ) (Fisher, Belle, 1993). Although there was a discrepant difference for dose *S*<sub>3</sub>, the null hypothesis of equality of variances was not rejected, i.e., the data did not provide enough evidence that the treatments presented different variances.

### Detection and replacement of outliers

In the test for the detection of outliers, the results led to conflicting interpretations. Calculations made by criterion 1 suggested that the responses listed as discrepant (19.76 and 21.55) were consistent with the rest of the data and should not be eliminated, because the calculated values of *G*<sub>1</sub>, i.e.,  $G_{1a} = 0.346$  and  $G_{1b} = 0.458$  (Table IV), were smaller than the critical value ( $G_1$  is 0.586 for  $n = 7$ ) (Farmacopoeia Brasileira, 1988; United States Pharmacopoeia, 2011).

By criterion 2, however, the  $R$ -value (0.292) was greater than the critical value ( $R$  was 0.288 for  $n = 7$  and  $k = 6$ ) (Farmacopeia Brasileira, 1988; United States Pharmacopeia, 2011), suggesting that the responses are not consistent with other data and should be treated as outliers. The option to null and to replace these values is not recommended, since the data were not fully recognized as outliers (Ludbrook, 2008), i.e., both criteria were contradictory. In the case of loss of some responses, the replacement of missing values is justified to avoid breaking the orthogonality (Finney, 1978). Otherwise, this may hamper the use of orthogonal coefficients in the ANOVA. In an evaluation of the data obtained, it was found that (i) the tests of normality and homogeneity of variance were satisfactory, and (ii) the replacement of two dose responses in the  $S_3$  column contradicted the recommendation of the pharmacopeia that advocates “to replace the lost response, at most one in each treatment or, up to 5% or 10% of total responses”. Thus, an option was made to keep the data originally obtained in the assay by accepting them as a genuine and predictable part of the experimental data.

To check the impact on the replacement of these responses, values were replaced from 19.76 and 21.55 to 20.37 and 20.76, respectively. The replacement of such responses did not significantly alter the potency, but the ANOVA showed a relevant deviation from parallelism ( $F$  calculated was 6.54, greater than the critical value  $F$  critical = 4.17) between the reference standard and unknown regression lines.

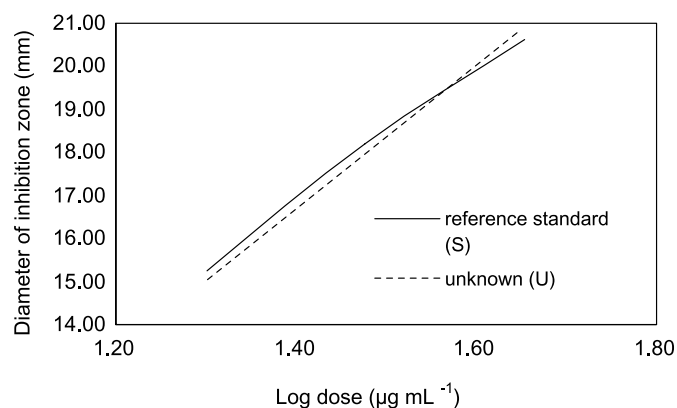
## Analysis of variance

Considering ANOVA in all pharmacopeial methods, observations were made for each source of variation, as follows:

**Preparation:** The calculated  $F$ -value ( $F = 0.28$ ) was lower than the critical value ( $F_{1,30;0.05} = 4.17$ ) (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). This indicates that the preparations are similar or that the unknown has a potency close to that of the reference standard.

**Regression:** The calculated  $F$ -value ( $F = 2300$ ) was higher than the critical value at 1% significance level ( $F_{1,30;0.01} = 7.56$ ) (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011), indicating that the dose-response curve was highly significant. Table VIII shows the values of common slope  $b$ . The calculation was not performed according to USP 34, since this parameter is not a requirement for testing the validity of the model by parallel lines–balanced assay, according to general methods of the pharmacopeia. The significance of the regression was confirmed by statistic  $C$  or  $g$ , i.e.,  $C$  is very close to 1 and  $g$  is less than 1. Considering the average of  $C$  (1.002) and  $g$  (0.002) in Table VIII, the value for expression  $C(1 - g)$  is 0.999996 or  $[1.002(1 - 0.002)] \approx 1$ .

**Deviation from parallelism:** Ideally, lines should be completely parallel, and there should be no deviation from parallelism. Therefore, a difference of at least one unit in the  $F$ -value calculated for the critical value should be established. The calculated  $F$ -value ( $F = 4.12$ ) was lower, but very close to the critical value ( $F_{1,30;0.05} = 4.17$ ) (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). In the graph response versus concentration it was observed that lines cross each other, suggesting a possible deviation of parallelism (Figure 3). The calculated  $F$ -value (4.12) was very close



**FIGURE 3** - Regression curves fitted to the ofloxacin microbiological assay, using three-dose level (3x3) for the reference standard (S) and unknown (U) preparations.

**TABLE VIII** - Comparative results of common slope and precision of the slope ( $g$  or  $C$ ) according to several pharmacopeial codes, applied for the ofloxacin microbiological assay

Parameter	Pharmacopeial code			
	IP	USP 34	BP and EP	FB
Common slope	15.958	NR	15.956	15.956
Precision of the slope	$g = 0.002$	$C = 1.001$	$C = 1.002$	$C = 1.002$

NR, not required.

to the critical value, indicating a deviation of parallelism, i.e., it exists, but is not significant. The  $p$ -value was greater than 0.05, confirming that the deviation from parallelism is not significant.

Deviation from linearity: No deviation in the same direction and opposite direction was observed because the calculated  $F$ -values ( $F = 1.95$  and  $1.41$ ) were smaller than the critical value ( $F_{1,30;0.05} = 4.17$ ) (Farmacopeia, 1988). According to BP and EP, the calculated  $F$ -value ( $1.68$ ) was also smaller than the critical value ( $F_{2,30;0.05} = 3.32$ ) (British Pharmacopoeia, 2011; European Pharmacopoeia, 2011).

Treatments: The calculated  $F$ -value ( $F = 457$ ) was greater than the critical value ( $F_{5,30;0.05} = 2.53$ ) (Farmacopeia, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011), indicating that it is a major influence on the regression curve, as expected.

Blocks (plates): According to FB, BP, and EP, the calculated  $F$ -values ( $F = 2.84$  and  $2.85$ ) were lower than the critical value ( $F_{6,30;0.01} = 3.47$ ) (Farmacopeia Brasileira, 1988; British Pharmacopoeia, 2011; European Pharmacopoeia, 2011). Hence, there is no difference in the experimental model due to the source of variation, blocks. Even if it existed, the assay would not be considered invalid, since each block is an experimental unit that includes all doses, one for each treatment. Therefore, this is an independent assay, the differences of which affect the treatments equally. Considering that every source of variation that affects the response can influence the slope, the differences between blocks are adjusted when using the balanced model.

For the interpretation of the validity of the assay, according to IP and USP 34, only the analysis of sources of variation parallelism, and linearity deviation in the same direction and opposite direction are taken into consideration. The sum of squared deviations, obtained by using orthogonal coefficients (USP 34), is equivalent to all other sources of variation based on the least squares method, denoted by the pharmacopeias BP, EP, and FB. The found  $F_3$ -value ( $2.50$ ) is smaller than the critical value ( $F_3$  is  $2.92$  for d.f. =  $30$ ) (United States Pharmacopeia, 2011), which determines that there are no parameters that invalidate the assay.

### Evaluation of potency and confidence limits

As expected, the results from the application of pharmacopeial statistical methods are quite similar. The expressions for the potency calculations adopted by the pharmacopeias are apparently different, but the variables have the same meaning and are therefore related. Equality demonstrations of these expressions have not been presented in this work.

The ANOVA indicated no difference between the potency of the reference standard and sample preparation of ofloxacin tablets. The results of relative potency (antilog  $M' \times R$ ;  $R = 400$  mg) and their confidence limits (in percentage), according to each pharmaceutical code, are as follows: FB, 99.27 (96.50 to 102.11); BP and EP, 99.27 (96.50 to 102.11); IP, 99.28 (97.22 to 102.86); and USP 34, 99.27 (96.50 to 102.11).

### CONCLUSION

The reliability of the results of antibiotic potency assays depends on the control of the variables that can affect such results. The replacement of outliers should not be a rule, especially when the results of ANOVA are satisfactory and do not invalidate the assay performed. It is important to note that the outliers are random and that replacement does not result in biased results. Therefore, the decision to accept or reject the data, presumably atypical, should be based on statistical analysis, as well as reasonable and appropriate interpretation of results by an experienced analyst.

For the parallel lines and three-dose level model, according to the pharmacopeias cited, one must observe that (i) the statistical calculations for determination of antibiotic potency are based on the regression analysis using the least squares method and on the use of orthogonal coefficients suitable for the balanced model; and (ii) the different statistical methods used are similar regarding the results obtained for potency and confidence interval.

It is important to mention that alternative statistical methods, as long as they are valid, are recognized by pharmacopeias. Finally, this work is a contribution to the choice of the most appropriate statistical calculation to be used.

### REFERENCES

- BARTLETT, M.S. Properties of sufficiency and statistical testes. *Proc. R. Soc. A.*, v.160, n.901, p.268-282, 1937.
- BLISS, C.I. The calculation of microbial assay. *Microbiol. Mol. Biol. Rev.*, v.20, n.4, p.243-258, 1956.
- BLISS, C.I.; COCHRAN, W.G.J.; TUKEY, W. A rejection criterion based upon the range. *Biometrika*, v.43, n.3/4, p.418-422.
- BRITISH PHARMACOPOEIA 2011. London: Her Majesty's Stationery Office, 2011. v.5, suppl. Chapter IV G, A632-A665.

- BURKE, S. Missing values, outliers, robust statistics & nonparametric methods. *LC-GC Eur.*, online suppl.59, p.19-24, 2001.
- COCHRAN, W.G. Testing a linear relation among variances, *Biometrics*, v.7, n.1, p.17-32, 1951.
- DIXON, W.J. Analysis of extreme values. *Ann. Math. Stat.*, v.21, n.4, p.488-506, 1950.
- DIXON, W.J. Ratios involving extreme values. *Ann. Math. Stat.*, v.22, n.1, p.68-78, 1951.
- EFSTATHIOU, C.E. Estimation of type I error probability from experimental Dixon's "Q" parameter on testing for outliers within small size data sets. *Talanta*, v.69, n.5, p.1068-1071, 2006.
- European Pharmacopoeia. 7.ed. Strasbourg: Council of Europe, 2011. p.551-579.
- FARIAS, A.A.; CESAR, C.C.; SOARES, J.F. *Introdução à estatística*. 2.ed. Rio de Janeiro: LTC, 2003. 340 p.
- Farmacopeia Brasileira. 4.ed. São Paulo: Atheneu, 1988. pt.1, p.VI.1-VI.10.
- FINNEY, D.J. *Statistical method in biological assay*. 3.ed. London: C. Griffin, 1978. 508 p.
- FISHER, R.A.; YATES, F. *Tabelas estatísticas para pesquisa em biologia, medicina e agricultura*. São Paulo: Polígono, 1971. 150 p.
- FISHER, L.D.; BELLE, G.V. *Biostatistics: a methodology for the health sciences*. 2.ed. New York: John Wiley-Interscience, 1993. 991 p.
- GADDUM, J.H. Bioassays and mathematics. *Pharmacol. Rev.*, v.5, n.1, p.87-134.
- HEWITT, W. *Microbiological assay for pharmaceutical analysis: a rational approach*. New York: Interpharm/CRC Press, 2004. 244 p.
- HEWITT, W. *Microbiological assay: an introduction to quantitative principles and evaluation*. New York: Academic Press, 1977. 284 p.
- LUDBROOK, J. Outlying observations and missing values: how should they be handled? *Clin. A. Exp. Pharmacol. P.*, v.35, n.5-6, p.670-678, 2008.
- MASSART, D.L.; VANDEGINSTE, B.G.M.; BUYDENS, L.M.C.; DE JONG, S.; LEWI, P.J.; SMEYERS-VERBEKE, J. Some important hypothesis tests. In: \_\_\_\_\_ *Handbook of chemometrics and qualimetrics: Part A*. 1.ed. Amsterdam: Elsevier science BV, 1997. p.93-112.
- MILLER, J.N. Outliers in experimental data and their treatment. *Analyst*, v.118, n.5, p.455-461, 1993.
- MINITAB: software para aprimoramento de qualidade. State College: Minitab, 2010. Available at: <<http://www.minitab.com/pt-BR>>. Accessed on: 10 Feb. 2011.
- SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). *Biometrika*, v.52, n.3/4, p.591-611, 1965.
- THE INTERNATIONAL PHARMACOPOEIA. Suppl.1 Available at: <<http://apps.who.int/phint/en/p/about/>>. Accessed on: 12 Feb. 2011.
- UNITED STATES PHARMACOPEIA. 34.ed. Rockville: The United States Pharmacopoeial Convention, 2011. p.95-106.

Received for publication on 16<sup>th</sup> September 2011

Accepted for publication on 27<sup>th</sup> April 2012