DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 20(5), 815-839 (1994)

DETECTION OF DISCORDANT CONTENT UNIFORMITY OBSERVATIONS AND COMPENDIAL COMPLIANCE

N. R. Bohidar
Philadelphia College of Pharmacy and Science
Villanova University

Norman R. Bohidar
University of Washington, Seattle, Washington

Nicholas R. Bohidar
National Computer Systems
Villanova University

ABSTRACT

A discordant observation is a data point whose value is drastically different from that of the rest of the members in the data set. In the context of content uniformity experiments, however, a discordant observation arises in two ways: (i) when the value of an observation is markedly distant from that of the other data points even though it is within the required compendial range, and (ii) when the value of an observation is outside the permissible compendial range. Several statistical tests for detecting one or more discordant observations are presented. Since discordancy distorts the symmetricity of the data, several tests of symmetricity are provided. Tests for detection of group discordancy induced by discordant samples are also included. The compendial requirements are explained in statistical terms. The impact of discordant observations on compendial compliance

requirements is assessed. The statistical basis of the construction of compendial limits as well as the assumptions implicit in the construction is elaborated. The results of the statistical analysis of three content uniformity studies are appropriately interpreted.

INTRODUCTION

In a set of experimental data, it is not uncommon to encounter a limited number of observations (one or two) whose magnitudes are far separated (either very low or very high) from the remainder of the data set. Such observations are called outlying, discordant, spurious, aberrant or maverick observations in the literature. In this paper, these spurious observations, however, will be referred to as discordant values (DVs). Most DVs are primarily detected by the subjective judgement of the scientist. As such, it would be necessary to conduct for objective confirmation appropriate statistical tests of discordancy so that relevant corrective actions are initiated in order to address appropriately the compendial and regulatory requirements. Based on the statistical results, several courses of action must be considered for the resolution of the situation. Investigative actions must be instituted to determine if the occurrence is indeed an outcome of an accidental deviation from a well documented prescribed procedure (such as high temperature, inaccurate solvent concentration or wrong mixing speed) or, the outcome of extreme random fluctuations (variabilities) inherently associated with the process itself.

The primary purpose of this paper is to (a) depict the appropriate statistical procedures for detecting discordant observations, (b) describe the various statistical tests for detecting asymmetricity in the data induced by the Dvs and for detecting group discordancy induced by discordant samples, (c) recommend necessary actions based on the statistical interpretations of the results and (d) review various compendial compliance requirements(1), (e) depict explicitly the statistical basis of the compendial limit formulation as well as the associated assumptions implicit in the formulation, and (f) present appropriate interpretation of the statistical results of three content uniformity studies associated with Product-C, Product-K and Product-Q.

THEORY

In this section, the basic theory associated with (a) the derivation of the cumulative distribution function and the probability density function of a DV, (b) the construction of the test statistic for testing a DV and (c) the determination of the critical value associated with a given significance probability level for accepting or rejecting the test statistic are presented in a succinct manner so as to generate a general appreciation of the various developments involved.

The cumulative distribution function is defined as the probability that a random variable X will assume a value less than or equal to a specified number X_0, denoted by $P(X < X_0)$. For a continuous distribution possessing a continuous derivative at all points, one obtains $dP(X)/dX = g(X)$ which is the probability distribution function of X. The integration of the function leads to $P(X) = \int_{-\infty}^{X} g(X)dX$ which is represented by the area enclosed by the curve g(X) with the property, $\int_{-\infty}^{+\infty} g(X)dX = 1.0$.

With this background consider the following: Let X_1, X_2, ---- X_n denote n random independent observations each emanating from the same probability distribution function and let X(1), X(2),--- X(n) denote the same observations ranked according to the order of their magnitudes. The necessary and sufficient condition that the largest observed value, X(n), is less than or equal to X is that all the observations are less than or equal to X. Since the observations are independent, according to the probability multiplication formula, one has, $P(X(n) < X) = P(X_1 < X) . P(X_2 < X) . ---- P(X_n < X) = [P(X)]^n$. The first derivative of the expression with respect to X, yields the probability distribution function as, $f(X)= n[P(X)]^{n-1} g(X)$, since $dP(X)/dX = g(X)$.

For a test of significance, however, one needs the percentage point of the rejection region (tail area) using the cumulative distribution function from which the percentage points of $X(n)_a$, the critical value, are calculated by solving the equation, $P[X(n) < X(n)_a] = a$ where, (1-a) is the level of significance (for instance, for a 5% test, 1-a = 0.05 and a = .95). Now, from above, $[P(X)]^n = a$ for $X = X(n)_a$. This leads to $P(X) = a^{1/n} = a^*$ for $X = X(n)_a$. The equation, $P(X) = a^*$, however, has the solution, $X = X_a^*$ and therefore, $X(n)_a = X_a^*$ for $a^* = a^{1/n}$. In other words, this essentially implies that the 95% percentage point for the largest observation, X(20), of 20 observations is equal to the 99.7% percentage point of the distribution of all the observations (See table below), because $(.95)^{1/20} = 0.997 = a^*$.

Now consider n observations from a Gaussian distributed population with parameters μ as the population mean and σ^2 the population variance. Then $X(n)_a = \mu + \sigma Z(n)_a = \mu + \sigma Z_a^*$, for $a^* = a^{1/n}$ symbolically then,

$$a^* = \left[\int_{-\infty}^{X} 1/(2\pi)^{\frac{1}{2}} \exp(-(t - \mu)^2/2\sigma^2 \, dt \right]^{1/n}$$

From a standard normal ($\mu = 0$, $\sigma = 1$) table, one finds that, X(20)(.95) (one-sided) = Z(.997)(one-sided) = 2.80, which implies that the probability that the largest of the 20 normally distributed observations is less than $\mu + 2.80\sigma$ is 95%. The following table is constructed to show the critical values of the 95% percentage point of the distribution of the discordant value (the largest observation) as a function of the percentage points of the distribution of all the observations for a selected set of sample sizes, assuming strict normality.

Sample Size (n)	Level of Sig. for DV Test (1-a)	a	a^*	Critical Value for DV Test(one sided) Za^*
1	0.05	0.95	0.95000	1.65
2	0.05	0.95	0.97468	1.96
5	0.05	0.95	0.98979	2.32
10	0.05	0.95	0.99488	2.57
20	0.05	0.95	0.99744	2.80
30	0.05	0.95	0.99824	2.93
60	0.05	0.95	0.99915	3.14

Note that, for n = 5, $a^* = (0.95)^{0.20} = 0.989794$ and now one enters the widely available standard normal statistical table and locates the a^* value and reads off the critical value from the first column of the standard normal table. Consider some examples. A sample of 10 tablets are randomly selected from a large batch whose content uniformity mean and standard deviation are known to be 100.0 and 6.0 respectively (population RSD = 6%). If the mean of the sample is 116.0, then the upper 95% critical value is X(10)(.95) = 100 + (6)(2.57) = 115.42. Since the largest value (116.0) exceeds 115.42, it should be considered as a significant discordant observation. This particular example has some significance in that the current compendial limit for range is 85%-115% and that for RSD is 6%. For a sample of size 30, however, the 95% upper critical value is 100 + (7.8)(2.80) = 121.84(approx. 122.0). (Note that the compendial limit for n=30 is 75%-125%(RSD = 7.8%)).

The test procedures depicted in this paper pertain to test statistics involving the sample estimates of the mean and variance. Consequently, the probability distribution of the test criterion is no longer Gaussian. It is indeed a variation of the Student's t-distribution. The probability distribution function is derived from the following recurrence relationship, (2,3)

$$f_n(t) = K_1 . K_2 . K_3 . K_4 . F_{n-1}(N/D)$$

where $K_1 = (n/n-1)$, $K_2 = (n/\pi)^{\frac{1}{2}}$, $K_3 = \Gamma(n-1/2)/\Gamma(n-2/2)$, $K_4 = [1 - (nt^2/(n-1)^2]^{(n-4)/2}$, $N = n^2(n-2)t^2$
$D = (n-1)[(n-1)^2 - nt^2]$, Limits of $t = (n^{-\frac{1}{2}}, (n-1)n^{-\frac{1}{2}})$, $F_2(t) = 0$ when $t < 2^{-\frac{1}{2}}$ and $F_2(t) = 1$ when $t > 2^{-\frac{1}{2}}$.
The level of significance probability is calculated as,

$$a(t) \leq nP(t_{n-2}) > (N^*/D^*)^{\frac{1}{2}})$$ where $N^* = n(n-2)t^2$, $D^* = (n-1)^2 - nt^2$, and $t_{(n-2)}$ = tabular Student's t value with (n-2) degrees of freedom, and P stands for Probability. For example, the critical value for the one-sided 5% T_n-test with n=10 is 2.176 (See Column I in Table A). The value for $(N^*/D^*)^{\frac{1}{2}}$ with n=10 and t=2.176, is 3.355. By entering the Student's t-table, one finds that for 8 degrees of freedom (n-2), the critical t-value is indeed 3.355 for the significance probability = 0.005. Now by multiplying the value by 10, one has a(2.176) = 0.05 or 5%. For T_{10} = 2.3, $(N^*/D^*)^{\frac{1}{2}}$ = 3.88 and a(3.88) << 0.05 meaning that T_{10}= 2.3 is significant at a level much lower than 5%. For each test statistic presented, appropriate and elaborate tables of critical values are provided for their immediate use.

STATISTICAL TESTS OF DISCORDANCY

In this section, the formula for each test statistic, the application of the test to a set of content uniformity data and the appropriate interpretation of the results, are presented, in detail. The critical values associated with each test statistic, necessary for deriving proper conclusions are presented in Table-A.

I. Test for a Single Discordant Value:

(i) Test for High DV:

(A) T-G Test(4,5): After an examination of the following ten content uniformity values arranged from low to high (99.7, 100.5, 101.1, 101.2, 101.2, 101.6, 101.6, 102.4, 102.8 and 105.7) it is contended by the scientist that the highest (H) value is a DV. A test is carried out to confirm or reject this contention. The test statistic for a single high DV is $T(n,H) = (X_n-X^*)/S$, where X_n is the highest value in a sample of size n, $X^* = \Sigma X/n$ = mean of n values, n = total number of observations in the sample, $S^2 = \Sigma(X-X^*)^2/(n-1)$, and S = positive square-root of S^2 (standard deviation). For the present data set, $T(10,H) = (105.7-101.68)/1.674 = 2.401$. The one-sided 5% critical value of $T(10,.05) = 2.176$ (see n=10 in column T of Table-A). Since the sample T(10,H) is larger than the critical T(10,.05), it is concluded that the highest value is indeed a DV ($p < 0.05$).

(B) D-Test (6): The test statistic for high DV is $R_{11} = (X_n-X_{n-1})/\ (X_n - X_2)$, where, as before, X_n is the highest value, X_{n-1} is the next highest value and X_2 is the second lowest value. Using the same data, $R_{11} = (105.7 - 102.8)/(105.7 - 100.5) = 0.5577$. The critical value for R_{11} (5%, one-sided) is 0.477, (see n=10, in column D of Table-A). Since the sample R_{11} exceeds the critical R_{11}(5%), the DV is considered a significant DV ($p < 0.05$).

It should be noted here that, the TG-test statistic measures the distance of the DV from the mean whereas, TABLE-A

CRITICAL VALUES ASSOCIATED WITH THE DISCORDANCY TESTS

n	T	D*	TML-2	TME-2	TML-3	TME-3	W	Ai	SK	KT
1	----	----	----	----	----	----	----	-0.4254	----	----
2	----	----	----	----	----	----	----	-0.2944	----	----
3	1.153	0.941	----	----	----	----	0.767	-0.2487	----	----
4	1.463	0.765	0.0008	0.001	----	----	0.748	-0.2148	----	----
5	1.672	0.642	0.0183	0.010	----	----	0.762	-0.1870	1.058	----
6	1.822	0.56	0.0564	0.034	0.010	0.004	0.788	-0.1630	1.034	----
7	1.938	0.507	0.1020	0.065	0.032	0.016	0.803	-0.1415	1.008	3.85
8	2.032	0.554	0.1478	0.099	0.064	0.034	0.818	-0.1219	0.991	4.09
9	2.110	0.512	0.1909	0.137	0.099	0.057	0.829	-0.1036	0.977	4.28
10	2.176	0.477	0.2305	0.172	0.129	0.083	0.842	-0.0862	0.950	4.40
11	2.234	0.576	0.2667	0.204	0.162	0.107	0.850	-0.0697	0.929	----
12	2.285	0.546	0.2996	0.234	0.196	0.133	0.859	-0.0537	----	4.56
13	2.331	0.521	0.3295	0.262	0.224	0.156	0.866	-0.0381	0.902	----
14	2.371	0.546	0.3568	0.293	0.250	0.179	0.874	-0.0227	----	----
15	2.409	0.525	0.3818	0.317	0.276	0.206	0.881	-0.0076	0.862	4.66
16	2.443	0.507	0.4048	0.340	0.300	0.227	0.887	0.0076	----	----
17	2.475	0.490	0.4259	0.362	0.322	0.248	0.892	0.0227	0.820	----
18	2.504	0.475	0.4455	0.382	0.337	0.267	0.897	0.0381	----	----
19	2.532	0.462	0.4636	0.398	0.354	0.287	0.901	0.0537	----	----
20	2.557	0.450	0.4804	0.416	0.377	0.302	0.905	0.0697	0.777	4.68

21	2.580	0.440	0.4961	----	----	0.381	0.908	0.0862	----	----
22	2.603	0.430	0.5107	----	----	----	0.911	0.1036	----	----
23	2.624	0.421	0.5244	----	----	----	0.914	0.1219	0.743	----
24	2.644	0.413	0.5373	----	----	----	0.916	0.1415	----	----
25	2.663	0.406	0.5495	0.493	0.450	0.381	0.918	0.1630	0.714	4.63
26	2.681	0.399	0.5609	----	----	----	0.920	0.1870	----	----
27	2.698	0.393	0.5717	----	----	----	0.923	0.2148	----	----
28	2.714	0.387	0.5819	----	----	----	0.924	0.2487	----	----
29	2.730	0.381	0.5916	----	----	----	0.926	0.2944	----	----
30	2.745	0.376	0.6008	0.549	0.506	0.443	0.927	0.4254	0.664	4.57

* For D-test note that,

(a) When n=3 to 7 use R10 = (X2 - X1)/(Xn- X1) if smallest value is suspected
and, R10 = (Xn - Xn-1)/(Xn - X1) if largest value is suspected

(b) When n=8 to 10 use R11 = (X2 - X1)/(Xn-1 - X1) if the smallest value is suspected
and use R11 = (Xn - Xn-1)/(Xn - X2) if largest value is suspected

(c) When n=11 to 13, use R21 = (X3 - X1)/(Xn-1 - X1) if smallest value is suspected
and use R21 = (Xn - Xn-2)/(Xn - X2) if largest value is suspected

(d) When n=14 to 30, use R22 = (X3 - X1)/(Xn-2 - X1) if smallest value is suspected
and use R22 = (Xn - Xn-2)/(Xn - X3) if largest value is suspected

Note: Each X has been identified by a subscript in the text.

the distance between the DV and the next highest value is measured by the D-test statistic. Since each test procedure defines discordancy in a different and unique way, the results of both tests are complimentary and informative and as such it is recommended that both tests should be carried out, and if any one of these tests is significant, then it should be declared that the suspected DV is indeed a significant DV ($p < 0.05$).

(ii) Test for Low DV:

(A) TG-Test (5): Consider the following data set of 10 values ordered from low to high, 90.1, 95.7, 97.5, 98.3, 98.4, 99.7, 100.0, 101.0, 101.5 and 101.7. The lowest (L) value is suspected by the scientist to be a discordant value. The test statistic for single low DV is $T(n,L) = (X^* - X_1)/S$, where X_1 is the lowest value of the sample of size n, and the symbols X^* and S have been defined previously. For this data set, $T(10,L) = (98.39-90.10)/3.4723 = 2.3875$. The one-sided 5 % critical value of $T(10,.05) = 2.176$ (see n=10 in column T of Table-A). Since the sample T(10,L) exceeds T(10,.05), it is concluded that the suspected DV is indeed a significant DV ($p < 0.05$).

(B) D-Test (6): The test statistic for the low DV is $R_{11} = (X_2 - X_1)/ (X_{n-1} - X_1)$, where, X_{n-1} is the next to the highest value and X_1 and X_2 are the two lowest values in the sample. Using the same data, $R_{11} = (95.7 - 90.1)/(101.5 - 90.1) = 0.4912$. The one-sided 5% critical value of R_{11} is 0.477, (see n=10, in column D of Table-A). Since the sample R_{11} exceeds the critical R_{11}(5%), the DV is considered a significant DV ($p < 0.05$).

It should be noted here that, even though the individual content uniformity values in each of the two data sets are within the compendial range of 85%-115% for n=10, and also the RSD values of each set, 1.65% and

3.53%, are below the required limit of 6% for n=10, yet one detects statistically significant DV in each set. For the purpose of quality control, the existence of a DV indicates that the process may not necessarily be in control. Even though only a single DV has been detected, however, in a random representative sampling scheme such a finding means that there is a high probability that other such units whose values may be outside the compendial limits do exist in the lot. This fact must be given serious consideration.

II. Test For Two Discordant Values:

(i) Test For Two High DVs.

TML-2 Test(7): Consider the following data set of 10 values ordered from low to high, 100.0, 100.5, 101.1, 101.2, 101.2, 101.6, 101.6, 102.4, 104.9 and 105.7. The two highest values are suspected to be the DVs. The test statistic for two high DVs is $L(d,n) = SS_d/SS_n$ where, d denotes the number of discordant observations, n denotes the sample size, SS_d denotes the sum of squares $\Sigma(X-X^*_d)^2$ without the d discordant values, SS_n denotes the sum of squares $\Sigma(X-X^*_n)^2$ for all n values, and X^*_d = mean of X without the d DVs, and X^*_n = mean of all n values.

For this data set, $L(2,10) = 3.70/30.916 = 0.11967 \approx 0.12$ where $SS_2 = 3.70$ and $SS_{10} = 30.916$. The one-sided 5% critical value for $L(2,10,.05) = 0.2305$ (see n=10 in column TML-2 of Table-A). Since the sample L(2,10) is less than the critical L(2,10,.05), the two suspected DVs are indeed significant DVs ($p < 0.05$).

(ii) Test for Two Low DVs:

TML-2 Test (7): Consider the following data set of 10 values arranged from low to high, 89.2, 90.1, 95.7, 97.5, 98.3, 99.7, 100.0, 101.0, 101.5 and 101.7. The two lowest values are suspected to be the DVs. The test statistic for two low DVs is $L(d,n) = SS_d/SS_n$ (the symbols have been defined above). For this data set, $L(2,10) = 31.215/184.501 = 0.1692$. As before, the one-sided 5% critical value for $L(2,10,.05) = 0.2305$ (see n=10 in column TML-2 of Table-A). Since the sample $L(2,10)$ is less than the critical $L(2,10,.05)$, the two suspected DVs are indeed significant DVs ($p < 0.05$).

(iii) <u>Test For Two DVs (One Low and One High)</u>

<u>TME-2 Test</u>(7): Consider the following data set of 10 values arranged according to the ascending order of their magnitudes, 97.1, 99.7, 100.5, 101.1, 101.2, 101.2, 101.6, 101.6, 102.4, and 104.7. The first step of the procedure is to compute for each observation, the absolute mean deviation ($ABS(X_i - X^*)$). The second step of the procedure is to arrange the absolute deviations from low to high and then denote the absolute deviations as, Z_1, Z_2---Z_n. Now the two discordant values will be on the high side of the Z-array, Z_{n-1} and Z_n. The test statistic for this is $E(d,n) = SS(Z)_d/SS(Z)_n$, where $SS(Z)_d$ stands for the sum of squares $\Sigma(Z-Z^*_d)^2$ <u>without</u> the d discordant values, $SS(Z)_n$ denotes the sum of squares $\Sigma(Z-Z^*_n)^2$ for all n values, and Z^*_d = mean of Z without the d DVs, and Z^*_n = mean of all nZ's. For this data set, the following are the absolute deviations arranged from low to high, 0.01, 0.09, 0.09, 0.49, 0.49, 0.61, 1.29, 1.41, 3.59 and 4.01. Now, $E(2,10) = 2.012/18.896 = 0.1065$ and the critical $E(2,10,.05) = 0.172$ (see n=10 in column TME-2 of Table-A). Since the sample $E(2,10)$ is <u>less</u> than $E(2,10,.05)$, it is concluded that the two suspected discordant values are both significantly ($p < 0.05$) discordant.

III. <u>Test For Three DVs</u>:

(i) <u>Test For Three High DVs</u>:

<u>TML-3 Test</u>(7): The data set of 10 values arranged from low to high consist of, 100.0, 100.5, 101.1, 101.2, 101.2, 101.6, 101.6, 103.5, 104.9 and 105.7. The test

statistic $L(d,n) = SS_d/SS_n = L(3,10) = 2.054/32.841 = 0.0626$ with a critical value $L(3,10,.05) = 0.129$ (see n=10 in column TML-3 of Table-A), indicating that all three high DVs are significantly ($p < 0.05$) discordant.

(ii) Test For Three Low DVs:

TML-3 Test(7): The data set of 10 values arranged from low to high consists of the following: 89.2, 90.1, 90.5, 95.7, 97.5, 98.3, 99.7, 100.0, 101.0, 101.5, for which $L(3,10) = 25.30/202.65 = 0.125$ with the critical value $L(3,10,.05) = 0.129$, indicating that all three low DVs are significantly ($p < 0.05$) discordant. (see n=10 in column TML-3 of Table-A).

(iii) Test For Three Dvs (one low and two high):

TME-3 Test(7): The data set of 10 values arranged from low to high consist of the following 95.1, 100.1, 100.5, 101.1, 101.2, 101.3, 101.6, 102.1, 104.5 and 105.7 and their corresponding absolute mean deviations arranged from low to high are, 0.02, 0.12, 0.22, 0.28, 0.78, 0.82, 1.22, 3.18, 4.38 and 6.22. The sample test statistic $E(3,10) = 1.201/41.174 = 0.0292$ with the critical value $E(3,10,.05) = 0.083$, indicating that all the three DVs are significantly discordant ($p < 0.05$) (see n=10 in column TME-3 of Table-A).

The foregoing statistical procedures can also be used for other response measurements such as, dissolution, disintegration, tablet strength or friability, and also for other sample sizes depicted in Table-A.

STATISTICAL TESTS OF SYMMETRICITY

The foregoing statistical tests of discordancy are designed to test only for one or two targeted DVs selected by the experimenter. It is however not always possible to detect the DVs in a random set of data by visual inspection in a finite amount of time. Since a

DV would tend to distort the symmetrical (around the mean) nature of the distribution of a set of experimental data, generally statistical tests of symmetricity (normality) are carried out to detect if the degree of distortion is indeed statistically significant and to determine the direction of such distortion. If the tests are significant, it would clearly signal the existence of the DVs in the sample. Besides, these tests should be conducted routinely to determine if the sample data meets the tacit assumptions implicit in the construction of the compendial limits. Three important tests are considered here, (a) Wilk-Shapiro W-Test, (b) Ferguson-Pearson skewness test (SK-Test) and (c) Ferguson-Pearson Kurtosis test (KT-Test). W-test is considered to be the most powerful general purpose (omnibus) test, SK-test is used to detect skewness (right or left) of the distribution and the KT-test is used to detect narrow peakedness of the distribution.

I. <u>W-Test</u>(8): The test statistic of the W-test is $W = B^2/\Sigma(X - X^*)^2$, where $B = \Sigma A_i X_i$ and the magnitudes of the A_i coefficients for n = 10 are provided in the example given below (the magnitude of the A_i coefficients for n = 30 are listed in A_i column of Table-A). The following data set of 10 values arranged from low to high is considered for the test, 99.7, 100.5, 101.1, 101.2, 101.2, 101.6, 101.6, 102.4, 102.8 and 105.7 and the magnitudes of their corresponding A_i coefficients are, -0.5739, -0.3291, -0.2141, -0.1224, -0.0399, +0.0399, +0.1224, +0.2141, +0.3291 and +0.5739. The quantities needed for the test are the sum of the cross product of the X_i-values and A_i-values, $\Sigma A_i X_i = B = 4.5436$, $\Sigma(X-X^*)^2 = 25.2205$, $W(10) = (4.5436)^2/25.2205 = 0.8185$ and the critical W(10,.05) = 0.842 (see n=10 in W column of Table-A). Since the sample W(10) is <u>less</u> than the critical W(10,.05), it is concluded that there is a significant ($p < 0.05$) departure from symmetricity (normality) for this sample.

II. SK-Test (9,10): The test statistic for the SK-test is, $SK = (n^{0.5})(M_3)/(M_2)^{1.5}$, where, $M_3 = \Sigma(X - X^*)^3$ and $M_2 = \Sigma(X - X^*)^2$. For the same data set considered in the W-test, given above, SK(10) = $(3.162)(49.7235)/(23.956)^{1.5} = 1.341$ and the critical SK(10,5%) one-sided value = 0.950 (see n=10, SK column in Table-A), indicating that the distribution is significantly ($p < 0.05$) skewed to the right (note: SK(10) is higher than critical SK(10,.05)).

III. KT-Test (9,10): The test statistic for the KT-test is $KT = n(M_4)/(M_2)^2$, where $M_4 = \Sigma(X - X^*)^4$ and M_2 retains the same definition as above. For the same data set considered in the W-test, KT(10) = $(10)(259.201)/(23.956)^2 = 4.5166$ and the critical value for a two-sided KT (10,.05) = 4.40 (see n=10 in KT column of Table-A), indicating that the distribution is subjected to a significant ($p < 0.05$) peakedness (flat peak, here)(note: KT(10) is higher than KT(10,.05)).

It is interesting to note that, when the same data set is subjected to all the three tests without the highest value(105.7), with n=9, none of the tests are significant, since, W(9) = 0.9659 (critical W(9,.05)= 0.829), SK(9) = -0.129 (critical SK(9,.05) = 0.977) and KT(9) = 2.577 (critical KT(9,.05) = 4.28).

Since each test procedure defines asymmetricity (non-normality) in a different and unique way, the results of all three tests (W, SK, KT) are complementary and informative and as such, it is recommended that all three tests, should be carried out, and if any one of the tests is significant, the sample distribution should be considered non-normal.

STATISTICAL TESTS OF GROUP DISCORDANCY

Consider that there are three or more independent laboratories testing the same lot for content uniformity. It is possible to delineate any one or two of the laboratories which are significantly discordant. Here the entire group (laboratory) is tested for discordancy and hence the name group discordancy. The method can also be used for testing different chemists, different days or different assay methods.

I. P-Test(11): The test statistic for the test is, $P = n(X^{**} - X^{*})/[\Sigma\Sigma(X - X^{**})^2]^{0.5}$, where X^{**} is the overall mean of all the groups, X^{*} is the mean of the group with the lowest mean and $\Sigma\Sigma(X - X^{**})^2$ is the sum of squares of the deviations from the overall mean for all observations (also called "total sum of squares"). The adjusted P here is $AP = (G)[(tn - 1)/n]^{0.5}$, where t is the number of groups and n is the number of observations per group. The means of the three laboratories which participated in a content uniformity test, are, $L_1 = 102.44$, $L_2 = 98.99$and $L_3 = 94.56$, the total sum of squares = 1718.53, t=3, n=30 and $X^{**} = 98.63$. Here P = (30)(98.63-94.56)/41.455 = 2.945 and AP = (2.945)(90-1)/30) = 5.073 and the critical AP(3,30,.05) = 1.75 indicating that there is at least one laboratory which is statistically significantly discordant.

II. DH-Test (12,13:) This test is used to determine which of the t groups are significantly discordant. The test statistic is, $DH = (X^{**} - X^{*}_{d})/S^{*}_{p}$, where X^{**} is the overall mean, X^{*}_{d} is the mean of the presumed discordant group and S^{*}_{p} is the standard error of the mean, calculated as follows: $S^2_p = (S_1^2 + S_2^2 + S_3^2)/3$ and $S^{*}_{p} = [S_p^2/30]^{0.5}$ (Note: in general terms, 3 is replaced by t. The formula for S_p^2 above is applied only when the group sample sizes are equal as in this case). For the above example, DH(3,30) = (98.63-94.56/0.6658 = 6.113 and the critical DH(3,30,.05) = 1.77, indicating that, LAB-3 (L_3) is indeed significantly ($p < 0.05$) discordant. Now when the same test is carried out for LAB-2 (L_2), DH(LAB-2) = 0.541 which is not significant ($p > 0.05$). Note that, for this study, it is decided that only the laboratory means which are below the value of 100.0 are to be tested for discordancy.

INTERPRETATION OF STATISTICALLY SIGNIFICANT DISCORDANT OBSERVATION AND RELATED COMPENDIAL CONSIDERATIONS

This section is devoted to an examination of the role of compendial limit requirements(1) and the detected discordant values in content uniformity studies. At the outset, however, it must be stated that, in the context of content uniformity experiments there are two ways in which a DV is defined, (i) when an observation is far separated from the remainder of the data set even though the unit is within the required compendial limits and (ii) when an observation is outside the prescribed lower or upper compendial limits. In either case, one must resort to statistical tests of discordancy, described above, for the purpose of confirmation and for initiating further action.

If an observation is established as a discordant value by the appropriate statistical discordancy tests (T,D,TML) as well as by the tests of non-normality (W,SK,KT), the predominant interpretation of the result is that the DV, in question, is not a member of the population distribution (batch) under consideration. However, the retention of the DV in the sample primarily depends upon its manner of origin. There are two possibilities(14), P-I: It is merely an extreme manifestation of random fluctuation (variability) inherent in the data associated with the batch, or P-II: It is an outcome of an accidental deviation from a well documented experimental procedure. To determine the extent of P-1 involvement, one is encouraged to take multiple samples (3 or more) of 30 units each from different locations of the container in full confirmation with the randomization procedure described in the product protocol. A statistical comparison of the sample standard deviations as well as of the sample relative standard deviations is conducted with and without the DV, after completing the tests of discordancy, the tests of symmetricity and the tests of group discordancy. If the tests without the DV turn out to be non-significant ($p > 0.05$), then P-I is considered to be the prime cause of the occurrence of the single DV. However, if the tests without the DV are significant, then it will be surmised that the process is out-of-control and heterogeneity among the units is a real possibility. To test for the P-II condition, one should institute an investigation of the conduct of the experiment and of the total process, by re-examining step-by-step, the existing (a) protocol (experimental steps), (b) statistical design of experiment (assignment of chemists, days, randomization), (c) process validation methods (analytical as well as production) and (d) the quality control procedures implemented. If an explanation can be found as to the cause of the incident, then it should be properly documented. If no specific cause can be assigned, the retention of the DV is recommended and as such, it can not be excluded from participating in the required statistical analyses. It must be noted that in a random representative sample, the presence of one non-conforming unit may be symptomatic to the presence of other undetected units lurking in the lot. The results of the tests with the

DV in the data, must carefully be examined and appropriately interpreted. The multiple sampling scheme is primarily directed towards (i) acquiring a deep insight into the statistical characteristics of the sample for making statistical inference about the lot, (ii) discovering if indeed the situation re-occurs in the future samples and (iii) reassuring that the single DV so far discovered is due to a subtle accident impossible to detect and document at the time of the incident.

STATISTICAL BASIS FOR CONSTRUCTION OF COMPENDIAL LIMITS ON SAMPLE RELATIVE STANDARD DEVIATION (RSD)

This section is devoted to a succinct description of the derivation of the statistical formulations associated with the construction of the required compendial limits (1) and to an explanation of the various statistical assumptions (limitations) implicit in the derivation of the formulations. Briefly, the compendial requirements for tablets are as follows: (i) The value of each of the 10 units in the sample must not only be within the range of 85%-115% but also the RSD of the sample must not exceed 6% and (ii) if one unit is outside the range of 85%-115% but no unit is outside the range of 75%-125% in (i), the RSD of the sample of 30(10+20) must not exceed 7.8%. The same requirements for the range and RSD are applied to capsules, as well.

Since the compendial limits on the sample RSD (RSD = $100S/X^*$, where X^* is the sample mean and S is the sample standard deviation) are based on the construction of the one-sided 95% statistical confidence limits, one needs the exact probability distribution of the sample RSD to obtain the necessary 5% percentage-point. A comprehensive coverage of the probability distributions of the sample RSD can be found in a paper by the first

TABLE-B

PERCENTAGE POINTS OF SAMPLE RSD AND ONE-SIDED UPPER 95% CONFIDENCE LIMIT EXPRESSED IN %.

METHOD-M(15)				METHOD-EX(15)			
N = 10		N = 30		N = 10		N = 30	
RSD	C*	RSD	C*	RSD	C*	RSD	C*
0.7	1.152	0.9	1.152	1.216	2.0	1.563	2.0
1.3	2.139	1.6	2.048	2.430	4.0	3.125	4.0
1.9	3.127	2.4	3.072	3.644	6.0	4.686	6.0
2.5	4.115	3.2	4.096	4.855	8.0	6.244	8.0
3.1	5.105	4.0	5.122	<u>6.064</u>	<u>10.0</u>	<u>7.800</u>	<u>10.0</u>
3.7	6.095	4.7	6.019	7.269	12.0	9.535	12.0
4.3	7.086	5.5	7.046	8.470	14.0	10.902	14.0
4.9	8.079	6.3	8.073				
5.5	9.074	7.1	9.101				
<u>6.1</u>	<u>10.070</u>	<u>7.8</u>	<u>10.002</u>				
6.7	11.068	8.6	11.033				
7.3	12.068	9.4	12.065				

RSD = One-sided upper 95% confidence limit, C* = Population (Lot) RSD, Underscored RSD = Current Compendial Limits

two authors(15). Primarily, the paper presents six different methods of obtaining the two-sided upper 95% confidence limits. Since an extensive groundwork has already been done in that paper (15), only the mechanism for obtaining the required compendial limits from the M and EX tables (15) are presented here. The procedures for the M-method and the EX-method are as follows: (i) construct one-sided upper 95% confidence limit for a given RSD (Note: the tables in (15) contain only the two-sided upper 95% confidence limit), (ii) locate the assumed value of the population (lot) RSD in the right-hand column and read off the corresponding RSD in the left-hand column for a given sample size (n=10 or n=30) and for a given method (M or EX) (See TABLE-B). The left-hand column RSDs constitute the 95% upper limit (or 5% percentage point) of sample RSDs. Consider, for instance, if the presumed value of the population (lot) RSD = 10% (as has been assumed for constructing the current compendial limits) then the upper limit on sample RSD for each of the following cases is (from TABLE-B):

METHOD	SAMPLE SIZE	
	n = 10	n = 30
M	6.100	7.8
EX	6.064	7.8

This derivation totally confirms the statistical basis of the compendial limits. Note that in this derivation no assumptions are made with respect to the presumed value of the mean (μ = 100) or that of the standard deviation (σ = 10)(which is erroneously done by using the chi-square distribution).

Assumptions and Limitations: It is true that no mathematical/statistical derivation is free of simplifying assumptions. However, in practice, one should be aware of the limitations of the formulations. One of the cardinal limitations here is the assumption of strict normality (symmetricity). Because of the large range of values allowed, it has been the experience that the distribution is generally skewed (right or left). Only in a symmetrical distribution the mean and standard deviation are meaningful, however, this may not be the case for asymmetrical skewed distributions which are more prevalent in samples of 30 with one or more discordant observations. All derivations are also based on the presumption that the lot RSD is 10% or less. This is a cardinal limitation. An RSD limit can be achieved even with a large standard deviation (magnitude of 7-10) because RSD is a ratio which is necessarily not unique. All these situations will be addressed in a forthcoming publication in which the formulations with the least number of limitations will be the prime consideration.

CONTENT UNIFORMITY STUDIES: RESULTS AND DISCUSSION

For each of the three tablet products denoted by C,K and Q, a content uniformity test is conducted based on the compendial guidelines as depicted on page 1618, USPXXII(1). In this section, the results of the statistical analyses of the three content uniformity studies are presented.

Product-C: The mean, standard deviation, range and RSD of the sample of 10 units are 104.3, 3.338, 98.7-108.6 and 3.20% respectively. Note that the individual units are within the required compendial range of 85%-115% and the two-sided upper 95% confidence limit (TSUCL) for RSD of 3.20% is 5.850% for the M-method and 5.848% for the EX-method [in reference (15), see on page 34, TABLE-A, and locate the RSD in column-1 and the TSUCL in column-6 for the M-method, and see on page 35, TABLE-B, and find for the EX-method, TSUCL = 5.848% by interpolation between 0.04 and 0.06]. Since the TSUCL is below the required compendial limit 6.0% for n = 10, it can be safely surmised that the batch conforms to the compendial specifications. Even though this type of procedure is not in general practice yet, several companies in the industry have started to adopt this new procedure after the publication cited in (15). This type of limit will provide a greater degree of assurance of quality and product integrity.

Product-K: It is proposed here to present only the salient aspects of the statistical analysis associated with the product. A statistically significant (p <

.05) discordant value (85.7) is detected in the sample of 10 units and so twenty more units are tested, and the analysis of the 30(20+10) units indicates that the same value is still significantly discordant ($p < .05$). Thirty additional units are resampled and a test of discordancy indicates that one of the 30 values (112.8) is found to be statistically significantly ($p < .05$) discordant. Both samples failed the symmetricity tests as well as the tests for homogeneity of the two variances. For a resolution of the situation, the principles, procedures and recommendations provided in a previous section must be the prime consideration.

Product-Q: Again, only the salient aspects of the statistical analysis associated with the product are depicted here. In this experiment, one of the values of the sample of 10 units is found to be outside the upper limit of the permissible compendial range. Even though the lot already failed to meet the compendial specifications, the experiment is allowed to continue (a positive thing to do) by testing twenty additional units (SAMP-I, n=20+10), by retesting (SAMP-II, n=30) and by resampling (SAMP-III, n=30) for the purpose of (a) discovering if indeed the incident re-occurs in the future samples, (b) acquiring a deep insight into the statistical characteristics of the samples to make statistical inference about the lot and (c) examining if the DV indeed pertains primarily to a single unit and that it is not a reflection of the entire lot. However, the samples failed the tests of symmetricity, the tests of group discordancy (one of the sample is discordant) and the tests of homogeneity of sample variances. For a resolution of such situations, the statistical principles, actions, procedures and recommendations provided in a previous section must be the prime consideration.

ACKNOWLEDGEMENTS

Grateful thanks are due to Mrs. Banita Mukerji for carrying out some of the early computations for skewness and kurtosis, promptly and efficiently, and to Mrs. Barbara J. Tomlinson for her meticulous efforts in transcribing the intricate statistical formulas, organizing the tables in a perceptible format and carrying out the typing aspects of the manuscript on a wordprocessor with utmost rapidity.

REFERENCES

1. USP-NF United States Pharmacopeia-National Formulary. USP-22nd Revision, NF-17th Revision. USP Convention, Inc. Rockville, Md. Pg. 1618-1619 (1989).
2. E.S. Pearson and C. Chandra Sekar. The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations. Biometrika, Vol. 28, 308-320 (1932).
3. F. E. Grubbs, Sample Criteria for Testing Outlying Observations. Annals of Mathematical Statistics, Vol.21 27-58 (1950).
4. W. R. Thompson, On a Criteria for the Rejection of Observations and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation. Annals of Mathematical Statistics, Vol.6, 214-219 (1935).
5. F. E. Grubbs and G. Beck, Extension of Sample Sizes and Percentage Points of Significance Tests of Outlying Observations. Technometrics. Vol.14, No.4, 847-854 (1972).
6. W. J. Dixon, Processing Data for Outliers Biometrics, Vol. 9, No.1, 74-89 (1953).
7. G.L. Tietjen and R. H. Moore, Some Grubbs-Type Statistics for Detection of Several Outliers. Technometrics, Vol.14, No.3, 583-597 (1972).

8. S.S. Shapiro and M.B. Wilk, An Analysis of Variance Test for Non-normality (Complete Samples). Biometrika, Vol.52, 591-611 (1965)

9. T. S. Ferguson, On the Rejection of Outliers. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Edited by J. Neyman, Univ. of Calif. Press, pp. 253-287 (1961).

10. R. B. D'Agostino and M. A. Stephens, "Goodness-of-Fit Techniques" Marcel Dekker Inc., N.Y. (1986)

11. E. Paulson, An Optimum Solution to the K-Sample Slippage Problem. Annals of Mathematical Statistics, Vol. 23, 610-616 (1925).

12. H. A. David, "Order Statistics", Wiley, N.Y.(1970).

13. C. P. Quesenberry and H. A. David, Some Tests for Outliers, Biometrika, Vol. 48, 379-387 (1961).

14. F. E. Grubbs, Procedures for Detecting Outlying Observations in Samples, Technometrics. Vol.11, No.4, 1-21 (1969).

15. N. R. Bohidar and N. R. Bohidar, Construction of Upper Confidence Limit of Coefficient of Variation for Content Uniformity. Drug Development and Industrial Pharmacy, Vol. 18, No.1, 21-37 (1992).