Incidentally it was found that benzycin appears to be strongly diuretic; in all of the subjects the quantity of urine was more than doubled. With benzyl benzoate the increase was only 40 per cent. The fairly large doses taken in all cases were found to have no apparent ill-effects.

In conclusion, it might be noted that the benzyl esters are probably hydrolyzed before they have proceeded very far after ingestion. The part of the canal in which absorption takes place will undoubtedly influence the physiological action very profoundly. If hydroylsis occurs in the intestinal tract, one might expect muscular relaxation only at some point above this and certainly not after absorption of the products, unless the latter have such an action. Otherwise, a simple mixture of benzoates, phosphates or succinates with benzyl alcohol would be as efficient.

REFERENCES.

(1) Macht, *J. Pharmacol. Exper. Therap.*, 11 (1918), 419; *J. A. M. A.*, 75 (1920), 567; 75 (1920), 769.

(2) Among others: Litzenberg, *Ibid.*, 73 (1919), 601; Haugwout, *Ibid.*, 73 (1919), 1310; McMurray, *Ibid.*, 75 (1920), 433; Hirschfelder, *Ibid.*, 75 (1920), 634; Laubry, *Ibid.*, 77 (1921), 157.

(3) Stater, *Ibid.*, 75 (1920), 463; 79 (1922), 1362; Nielsen, *J. Lab. Clin. Med.*, 7 (1922), 579; Mason, *Ibid.*, 6 (1920), 62; Gruber, *Ibid.*, 9 (1923), 15, 92 and 685; 10 (1924), 284.

(4) Snapper, *et al.*, *Nederland. Tijdschr. Geneeskunde*, 68 (1924), 3125; *Chem. Abst.*, 19 (1925), 1598.

University of Washington, Seattle.

THE APPLICATION OF STATISTICAL METHODS TO PHARMACEUTICAL RESEARCH. I. MEASURES OF ACCURACY.*

BY JAMES C. MUNCH.

The key-note of quantitative investigations is accuracy. Many investigations have been incomplete, and faulty conclusions have been drawn from an inadequate number of results, because of the neglect of this important factor. The variability of animals has been used as an excuse for divergent results from time immemorial. On the other hand, the accidental coincidence of two successive tests has been accepted as conclusive proof of correctness of results, irrespective of the nature or number of variables involved.

Simple mathematical procedures have been developed for measuring the accuracy of results. This type of mathematics has been employed so extensively in connection with the investigations of biometricians, physicists and actuaries that the nomenclature of this field follows their terminology (1, 2, 3, 4, 5, 6, 7, 8, 9, 13). This may explain the reluctance of workers in other fields to apply this type of mathematics to the interpretation of research results (11).

Two types of variations must be differentiated: (*1*) constant errors and (*2*) variable errors. Constant errors result from defects in apparatus, incorrect graduation of equipment, erroneous calibration of weights, etc. A constant error will be produced in all measurements with faulty equipment. Increasing the number of observations will have no effect in correction of a constant error. The presence

* Scientific Section, A. Ph. A., Baltimore meeting, 1930.

of a constant error may often be detected by repeating observations under different conditions, using different equipment. So long as relative measurements are desired in terms of a common standard measured with the same equipment, the disturbance produced by the presence of a constant error is negligible. However, comparison of absolute values is unsafe unless the magnitude of the constant error is determined.

Variable errors are considered to be the resultant effects produced by the concomitant action of all variations, other than constant errors. Some variations will tend to increase, others to decrease the observed results. When all variable errors act in the same direction the maximum or the minimum value is obtained. When some errors tend to compensate for others an intermediate result is observed. By increasing the number of observations, the general tendency is found for a neutralization of the divergent effects and the number of results unaffected by variable errors increases. When the experimental values are plotted graphically, it is noted that the largest number of results occur in the vicinity of the mean and that the results which differ from the mean decrease in number as one proceeds on either side of the mean value. If the tops of the lines representing individual values are connected by a continuous line, a curve is produced which tends to resemble a cocked hat, or in more modern styles a "tin" army helmet. When a sufficiently large number of experimental results have been obtained, a symmetrical curve is usually produced which is called the "normal frequency curve." The mean, the mode and the median tend to coincide. If the curve be folded upon the axis of the mean, the two halves coincide. An infinitely large number of results are usually required to obtain an exact fit; with 25 to 100 observations, a trend toward this type of curve is often evident.

This type of curve, the "normal frequency curve," results when there is an even chance of an error being greater or less than the mean. Under certain conditions, this probability does not hold, and asymmetric curves are produced. The mathematics of these types of frequency curves are somewhat complicated.

Three measures of accuracy have been developed which are related to the normal frequency curve: (1) h, the modulus of precision; (2) σ, the standard deviation, and (3) PE, the probable error.

(1) The modulus of precision, h, is obtained from the formula of the normal frequency curve

$$y = ke^{-h^2x^2}.$$

As the accuracy of results increases, the value of h increases (8). The method of calculation is somewhat involved and is more difficult to interpret than with the other measures of accuracy.

(2) The standard deviation, σ, is the inflexion point of the normal frequency curve. Between the limit of 1 σ above and 1 σ below the mean, approximately 68 per cent of the observations are expected to occur (1, 5, 12). The method of calculation is given in Table I, which contains results obtained by experienced analysts in the assay of tincture of digitalis by the One-Hour Frog Method (10).

A deviation is the difference between an individual observation and the mean of all the observations in a given series. It will have a positive sign if the observed value is numerically greater than the mean, and a negative sign if it is smaller.

For example, the mean of the differences in the second column of Table I is 1.2; the deviation of the first entry (10) is plus 9 from an arbitrary mean of 1.0, or 8.8 from the mean of 1.2. The deviation of the last entry in the same column (−8) is minus 9 from an arbitrary mean of 1.0, or minus 9.2 from the mean of 1.2.

TABLE I.—ACCURACY OF ASSAYS OF TINCTURE OF DIGITALIS BY 1-HOUR FROG METHOD EXPERIENCED ANALYSTS.

Analyst no.	Difference from true strength.	Deviation from 1.0.		d^2.	Deviation from 1.2.		d^2.	Deviation from 2.5.		d^2.
		Plus.	Minus.		Plus.	Minus.		Plus.	Minus.	
2	10	9	...	81	8.8	...	77.44	7.5	...	56.25
2	10	9	...	81	8.8	...	77.44	7.5	...	56.25
2	10	9	...	81	8.8	...	77.44	7.5	...	56.25
5	29	28	...	784	27.8	...	772.84	26.5	...	702.25
5	0	...	1	1	....	1.2	1.44	....	2.5	6.25
5	15	14	...	196	13.8	...	190.44	12.5	...	156.25
6	10	9	...	81	8.8	...	77.44	7.5	...	56.25
6	5	4	...	16	3.8	...	14.44	2.5	...	6.25
6	0	...	1	1	....	1.2	1.44	....	2.5	6.25
7	10	9	...	81	8.8	...	77.44	7.5	...	56.25
7	0	...	1	1	....	1.2	1.44	....	2.5	6.25
7	5	4	...	16	3.8	...	14.44	2.5	...	6.25
10	− 4	...	5	25	....	5.2	27.04	....	6.5	42.25
10	0	...	1	1	....	1.2	1.44	....	2.5	6.25
10	0	...	1	1	....	1.2	1.44	....	2.5	6.25
12	− 4	...	5	25	....	5.2	27.04	....	6.5	42.25
12	−14	...	15	225	....	15.2	231.04	....	16.5	272.25
12	−16	...	17	289	....	17.2	295.84	....	18.5	342.25
17	29	28	...	784	27.8	...	772.84	26.5	...	702.25
17	0	...	1	1	....	1.2	1.44	....	2.5	6.25
17	−40	...	41	1681	....	41.2	1697.44	Omit	...	.....
18	−10	...	11	121	....	11.2	125.44	....	12.5	156.25
18	− 2	...	3	9	....	3.2	10.24	....	4.5	20.25
18	−10	...	11	121	....	11.2	125.44	....	12.5	156.25
19	−10	...	11	121	....	11.2	125.44	....	12.5	156.25
19	10	9	...	81	8.8	...	77.44	7.5	...	56.25
19	15	14	...	196	13.8	...	190.44	12.5	...	156.25
21	1	0	0	0	....	0.2	0.04	....	1.5	2.25
21	− 5	...	6	36	....	6.2	38.44	....	7.5	56.25
21	− 8	...	9	81	....	9.2	84.64	....	10.5	110.25
Sum	36	146	140	5218	143.6	143.6	5216.80	128.0	124.5	3461.25
Mean	1.2	...	...	173.93	....	...	173.89	....	...	119.35
σ	....	...	...	13.18	....	...	13.18	....	...	10.92
AD	....	9.53	...	.....	9.57	...	.....	8.71	...	.....
PE	....	8.06	...	.....	8.09	...	.....	7.36	...	.....
PE_{Mean}	....	1.46	...	.....	1.47	...	.....	1.36	...	.....

In order to determine σ, the deviation of each individual value from the mean is squared. The square of either a positive or a negative value will give a positive value. The sum of these squares is divided by the total number of observations, giving a quotient which is the average of the squares. The square root of this quotient is σ. In Table I, the sum of the squares of the individual deviations is

5218. Dividing this value by the number of observations (30), the average value of the squares is found to be 173.93. The square root of 173.93 is 13.18, which is the standard deviation about the arbitrary mean of 1.0. When calculated about the true mean of 1.2, the sum of the squares is found to be 5216.80, the average square 173.89 and σ is 13.18. This may be interpreted to mean that 68 per cent of the observed values should occur between Mean $\pm$ σ, (1.2 $\pm$ 13.18), or to range from plus 14.38 to minus 11.98. As a matter of fact, 23 of the 30 values (77 per cent) fell within these limits.

If the individual deviations are represented by "*d*," the sum of the deviations by Σ, and the number of observations by N, the equation for the determination of the standard deviation of an individual observation may be written:

$$\sigma = \sqrt{\frac{\Sigma d^2}{N}}.$$

The standard deviation of the mean of a series of observations is obtained by dividing the standard deviation of an individual observation by the square root of the number of observations in the series. In this instance, the standard deviation of the mean is $\frac{13.18}{\sqrt{30}}$ or 2.4.

The standard deviation has been employed in a number of reports. It is particularly useful in the calculation of correlation coefficients and certain other constants.

(3). The probable error (*PE*) is a more understandable measure of accuracy. It is that particular value such that within the limits of Mean $\pm$ *PE*, one-half of the total observations will be included. Another interpretation is that if one more observation is made, it is just as likely (the chances are even), that it will fall within the range between Mean plus *PE* and Mean minus *PE*, as that it will fall without this range of values.

PE may be calculated approximately as two-thirds (more exactly, 0.6745) times the standard deviation. This involves the labor of squaring the deviations, which is often a tedious procedure. As a slightly less accurate but much more rapid method, the sum of the first powers of the deviations is obtained without regard to their arithmetic sign (whether they are positive or minus) and divided by the number of individual observations to obtain the average deviation, *AD*. The quotient is multiplied by the constant factor 0.8453 to obtain the probable error. In Table I, the sum of the positive deviations from the arbitrary mean of 1.0 is 146, and the sum of the negative deviations is 140. The sum of these two values, representing the total deviations from the mean, is 286. The *AD* of the 30 results is 286/30 or 9.53; the *PE* is 0.8453 $\times$ 9.53, or 8.06. Similarly, the total deviations about the mean of 1.2 are 287.2, the average deviation is 9.57 and the *PE* 8.09.

PE of the Mean is obtained by dividing the *PE* of the individual observations by the square root of the number in the series; in this instance, PE_{Mean} equals $\frac{8.06}{\sqrt{30}}$ or 1.46.

When *PE* is calculated from σ, the value obtained is 8.9, which does not agree very closely with the result obtained from *AD*, namely, 8.06. This suggests that one or more observations of the series differ widely from those expected in a normal frequency distribution. Inspection of the table reveals an observation of minus 40, which differs widely from the remaining members of the series. The *PE* is serviceable in determining whether this experimental result should be considered in conjunction with the other values, or whether some special variables have influenced it which did not appear to affect the remainder of the series. The result in question, minus 40, has a deviation from the mean of minus 41.2. This deviation is $\frac{41.2}{8.06}$ or slightly more than five times as large as the *PE*. Tables of probability showing the chance of occurrence of random variations as large as various multiples of the *PE* (11, 12) have been recalculated in Table II; a deviation 5 times *PE* has 1 chance of occurrence in 1350 trials. These odds are so large as to justify the conclusion that this particular result resulted from conditions not affecting the remaining reports, and that it may properly be eliminated from the calculations. When this is done, the mean becomes 2.5, σ 10.92 and *PE* from σ, 7.37.

TABLE II.—PROBABILITY OF OCCURRENCE OF DEVIATIONS LARGER THAN THE PROBABLE ERROR (*PE*).

$\frac{\text{Deviation}}{PE.}$	One chance of occurrence in ... trials.	$\frac{\text{Deviation}}{PE.}$	One chance of occurrence in ... trials.
1.0	2	4.0	143
1.4	3	4.2	216
1.7	4	4.4	333
1.9	5	4.6	520
2.25	8	4.7	660
2.5	10	4.8	825
2.75	15	4.9	1050
3.0	23	5.0	1350
3.25	35	6.0	19,230
3.5	55	7.0	435,000
3.82	100	8.0	1,500,000,000

AD is now 8.71, which yields a value of 7.36 for *PE*, substantially the same result as that obtained in calculations based upon σ.

The proper method of reporting the results of these assays would be, that 29 reports showed a mean of plus 2.5 ± 7.37, which would imply that half of the results fell between plus 9.87 and minus 4.87 (in fact 16 of 29 results, or 55 per cent, fell within these limits), and that PE_{Mean} is 1.36.

When the same samples of tincture of digitalis were assayed by inexperienced analysts (10), the average of 27 reports was minus 15.88 ± 10.08 and PE_{Mean} was 1.94. Half of the reports would be expected to lie between minus 5.80 and minus 25.96 (in fact, 17 of 27 results, or 63 per cent, fell within these limits). In comparing the results of analyses by the experienced and the inexperienced analysts, certain differences are noted. The means, plus 2.5 and minus 15.88, differ by a total of 18.38. To determine whether this difference is significant, its *PE* is determined. The *PE* of the sum or difference of two quantities is the square root of the sum of the squares of the *PEs* of the values compared (9). In this case, the *PE* of 18.38 would be $\sqrt{(1.36)^2 + (1.94),^2}$ or 2.37. The difference between the means is 18.38/2.37, or 7.75 times the *PE*. Reference to Table II shows that the chance of occurrence of a deviation 8 times its *PE* is approximately 1 in 1,500,000,000. Accordingly, it may be concluded that the difference between these means is significant.

These *PE* values may be interpreted as one expression of the accuracy of results attained. The experienced analysts should agree within 7.37 per cent at least half of the time, and should not differ by more than 25 per cent more often than once in fifty times. The inexperienced analysts, on the other hand, tend to obtain results which are 15 per cent too low, should agree within 10 per cent at least half the time, and should not differ by more than 25 per cent more often than once in ten times.

PE is more generally useful and more readily understood. It may be considered as a suitable, readily determinable measure of the accuracy of the results of quantitative assays.

CONCLUSIONS.

1. The standard deviation (σ) and the probable error (*PE*) are measures of the accuracy of quantitative measurements.
2. *PE* is more convenient to determine and apply.
3. The probability of occurrence of deviations greater than *PE* may be applied as a criterion for the rejection of dubious observations.

BIBLIOGRAPHY.

(1) D. Brunt, the combination of observations (1917).
(2) J. H. Burn, the errors of biological assay, *Physiol. Rev.*, 10 (1930), 146–169.
(3) W. Chauvenet, a treatise on the method of least squares (1868).
(4) C. B. Davenport, statistical methods, with special reference to biological variation, 3rd Edition (1914).
(5) W. P. Elderton, frequency curves and correlation.
(6) H. M. Goodwin, precision of measurements and graphical methods (1919).
(7) J. P. Jacobsen, einige grundbergriffe und sätze betreffs beobachtungen, die als wiederholungen betrachtet werden konnen, sowie eine anleitung zur behandlung solcher beobachtungen, *Dansk Tids. Farm.*, 4 (1930), 49–112.
(8) J. W. Mellor, higher mathematics for students of chemistry and physics (1919).
(9) M. Merriman, a textbook on the method of least squares (1913).
(10) James C. Munch, collaborative assay of tinctures of digitalis by the U. S. P. X 1-Hour Frog Method, *J. A. O. A. C.*, 11 (1928), 365–378.
(11) James C. Munch, "Bioassays," 1931.
(12) R. Pearl and J. R. Miner, a table for estimating the probable significance of statistical constants, *Maine Agr. Exp. Sta. Bull.* (1914), 226.
(13) G. U. Yule, an introduction to the theory of statistics, 6th Edition (1922).

BUCHU PRODUCTION IN THE UNION OF SOUTH AFRICA.

The production of buchu leaves in the Union of South Africa for the season of 1930 was expected to be somewhat smaller than that of the year 1929. The yield from the Government Forest Reserves was estimated at 25,000 pounds, as against 35,000 pounds for 1929; 30,000 for 1928; and 20,000 for 1927. This indicated a total yield from all sources for the 1930 season of about 200,000 pounds.

The bulk of the buchu, usually completely harvested before the end of March, is obtained from plants growing on private farms, while some grows wild and in a semicultivated state, but the yield of the Government Forest Reserves is usually a reliable barometer as to the quantity available. (Consul Cecil M. P. Cross, Cape Town.)