TEST DE DIXON
RECHERCHE DE VALEURS ABERRANTES

Extrait du référentiel du BTSA ANABIOTEC, module M53 :

<u>Objectif 4.5 : Repérer des valeurs aberrantes, test de Dixon.</u>
Recommandation pédagogique : ce test permet d'écarter des valeurs aberrantes. On traitera le cas d'une valeur aberrante ou de plusieurs.

Préambule

Quiconque voulant découvrir le test de Dixon va vite se trouver confronté à un obstacle : la multiplicité des sources, des méthodes, notations et tables.

L'objectif de cet article est de proposer une méthode simple à comprendre et à utiliser au niveau BTSA, afin d'uniformiser les pratiques pédagogiques à ce niveau.

Un petit peu d'histoire

En 1951, R. B. DEAN, and W. J. DIXON dans leur article *Simplified Statistics for small Numbers of Observations* s'intéressent à ce qu'ils appellent les "extraneous values". Traduisons "extraneous" : "sans grande portée", "superflu", "étranger". Ces "extraneous values" sont ce que nous appelons de nos jours les valeurs aberrantes. Quelques années plus tard (1969), dans les travaux de Grubbs, nous pouvons trouver une définition de cette notion, "outlier " dans le texte :

Valeur aberrante : observation qui semble dévier de façon marquée par rapport à l'ensemble des autres membres de l'échantillon dans lequel elle apparaît.

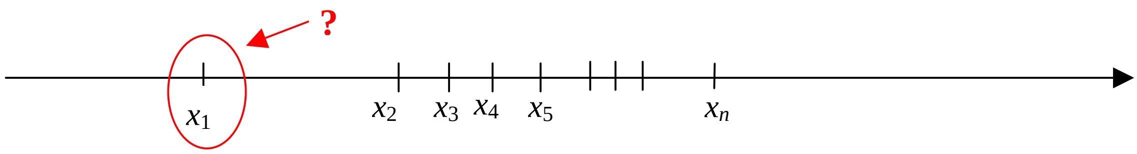
Le contexte

Au cours d'une expérimentation, il peut arriver qu'un des résultats semble s'écarter notablement des autres. Un graphique peut être d'une grande utilité pour s'en apercevoir. Une attitude classique, que l'on rencontre trop souvent, consiste à éliminer cette valeur en la considérant comme aberrante. Une bonne attitude à avoir est d'essayer de trouver la cause de l'écart (erreur de lecture, faute de calcul, etc) ; dans ce cas, il est tout à fait normal de l'éliminer. En revanche, si aucune cause accidentelle n'a pu être détectée, on s'abstiendra d'éliminer brutalement la valeur incriminée. Pour cela, il faut avoir recours à un test statistique permettant de justifier l'élimination de la valeur aberrante avec un risque de se tromper choisi au préalable. Le test de Dixon, que nous allons exposer, permet de réaliser cela, sous condition de normalité du caractère.

Principe du test

Notons tout d'abord qu'il peut s'appliquer aussi bien pour une série statistique à une variable (x_i) que pour une série statistique bivariée (x_i ; y_i).

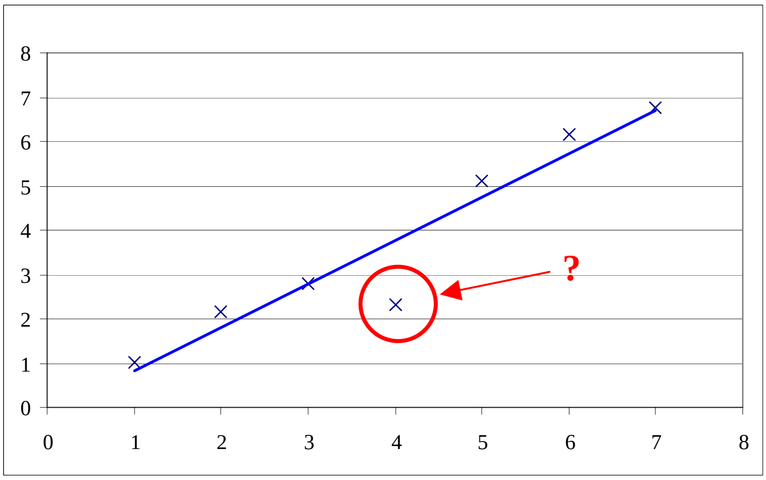
Dans le premier cas, les valeurs x_i étant rangées dans l'ordre croissant, le test de Dixon va détecter la (ou les) valeur(s) aberrante(s), aux extrémités de la distribution.



Si la valeur aberrante suspectée est très supérieure aux autres (à droite du graphique), les valeurs peuvent être alors classées dans l'ordre décroissant.

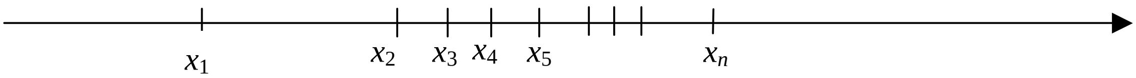
Dans le second cas, les observations sont représentées par un nuage de points dispersés autour de la droite de régression de y en x d'équation $y = a\,x + \mathrm{b}$ (obtenue par la méthode des moindres carrés), le test est basé sur la distribution des résidus.

Ces derniers sont notés, pour tout entier i, $e_i = y_i - \hat{y}_i$, c'est-à-dire $e_i = y_i - (a\,x_i + \mathrm{b})$.



I. Cas d'une seule valeur aberrante

Les valeurs observées sont classées par ordre croissant et notées x_1, x_2, ..., x_n.



Hypothèses

H_0 : "La valeur douteuse n'est pas une valeur aberrante."

H_1 : "La valeur douteuse est une valeur aberrante."

Variable de décision utilisée

Il faut comparer la distance entre la valeur suspectée aberrante et une valeur des plus proches, avec la distance entre la valeur suspectée aberrante et une des valeurs les plus éloignées de l'échantillon.

Notons R la variable aléatoire prenant pour valeur le rapport de ces distances. Sa valeur observée est donnée dans le tableau ci-dessous selon la valeur de n et la position de la valeur suspectée aberrante :

	la valeur suspectée aberrante est x_1	la valeur suspectée aberrante est x_n
$\boldsymbol{n \leq 10}$	$R_{obs} = \dfrac{x_2 - x_1}{x_n - x_1}$	$R_{obs} = \dfrac{x_n - x_{n-1}}{x_n - x_1}$
$\boldsymbol{n > 10}$	$R_{obs} = \dfrac{x_3 - x_1}{x_{n-2} - x_1}$	$R_{obs} = \dfrac{x_n - x_{n-2}}{x_n - x_3}$

Remarque

- Plus la valeur observée de R est élevée, plus la valeur suspectée est aberrante.
- On distingue $n \leq 10$ et $n > 10$ pour détecter les cas où il y a plus d'une valeur aberrante (voir troisième exemple suivant).

Valeur critique

On se fixe un seuil de risque α. La valeur critique est notée $r_{1-\alpha}$, elle est définie par : $P(R \leq r_{1-\alpha}) = 1 - \alpha$ et elle est donnée par la table en fin d'article.

Exemple d'utilisation de la table : $n = 8$ et $\alpha = 0{,}01$.

Dans le cas de la recherche d'une valeur aberrante, la table de Dixon indique que pour $n = 8$ et $\alpha = 0{,}01$, la valeur critique est $r_{0{,}99} = 0{,}59$.

Cela signifie que si l'on prélève aléatoirement un échantillon de taille 8 dans une population dans laquelle les données sont distribuées normalement alors la probabilité que R prenne une valeur inférieure ou égal à 0,59 est 0,99.

Règle de décision

Si $R_{obs} > r_{1-\alpha}$, on rejette H_0, donc la valeur suspectée est aberrante.

Si $R_{obs} \leq r_{1-\alpha}$, on n'est pas en mesure de rejeter H_0.

II. Un peu de pratique

Voici trois exemples d'application.

- Un premier sur une situation classique dans laquelle la valeur la plus élevée apparaît aberrante.
- Un second montrant un point aberrant au sein d'un nuage.
- Puis un troisième exemple dont le but est de montrer une situation dans laquelle on justifie la distinction entre $n \leq 10$ et $n > 10$ et qui montre qu'il peut exister deux valeurs aberrantes (cas traité dans la seconde partie de l'article).

Exemple 1

Dans la fabrication de comprimés effervescents, il est prévu que chaque comprimé doit contenir 1 625 mg de bicarbonate de sodium. Afin de contrôler la fabrication de ces médicaments, on a prélevé un échantillon de 10 comprimés et on a mesuré la quantité de bicarbonate de sodium en mg pour chacun d'eux. Les résultats obtenus sont résumés dans le tableau suivant:

1 620	1 621	1 623	1 628	1 633	1 635	1 637	1 641	1 643	1 659

On peut demander aux étudiants de réaliser un graphique sur un axe gradué pour détecter quelle(s) valeur(s) semble(nt) aberrante(s).

On effectue un test de Dixon au seuil de risque 0,05 pour tester si la valeur supérieure 1 659 est aberrante.

On teste les deux hypothèses :

H_0 : "1 659 n'est pas une valeur aberrante."

H_1 : "1 659 est une valeur aberrante."

$n = 10$ donc on utilise la variable aléatoire R qui prend comme valeur observée $R_{obs} = \dfrac{x_n - x_{n-1}}{x_n - x_1}$, soit $R_{obs} = \dfrac{x_{10} - x_9}{x_{10} - x_1}$ qui est égale à 0,410.

D'après la table, la valeur critique est $r_{0{,}95} = 0{,}412$. Comme $0{,}41 < 0{,}412$: on n'est pas en mesure de rejeter H_0. La valeur 1 659 ne peut pas être considérée comme aberrante, au seuil de 0,05.

Exemple 2

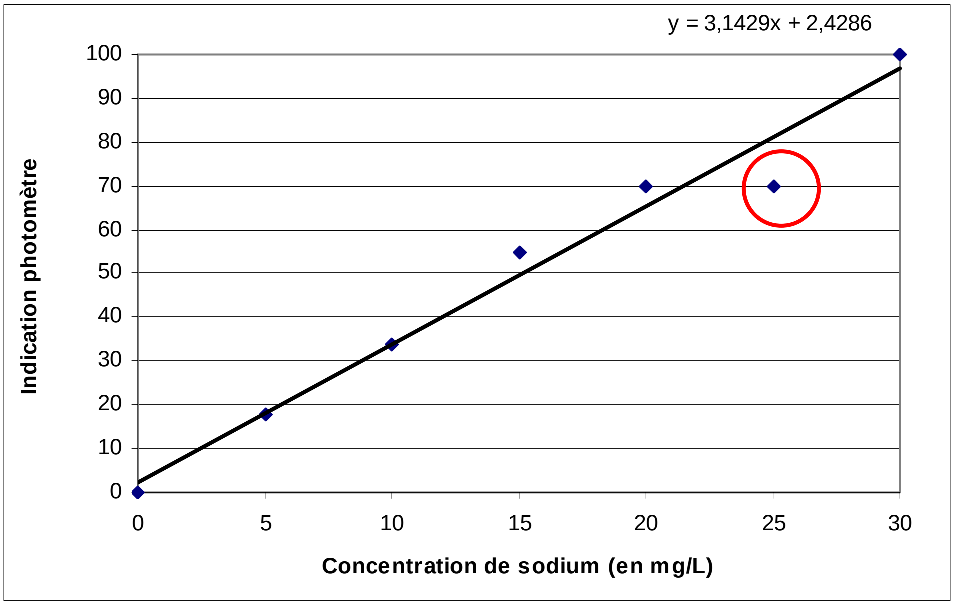
Lors d'un dosage de sodium par photométrie de flamme, on a procédé à un étalonnage (fond de flamme à 0 et solution concentrée à 100).

Les mesures figurent dans le tableau suivant :

Concentration de sodium (en mg/L) : *X*	0	5	10	15	20	25	30
Indication du photomètre : *Y*	0	18	34	55	70	70	100

La valeur observée pour une concentration de 25 mg/L peut-elle être considérée comme aberrante ?

Un petit coup d'œil sur le graphique :



On détermine l'équation de la droite d'ajustement de Y en X par la méthode des moindres carrés : $y = 3{,}1\,x + 2{,}4$.

X	0	5	10	15	20	25	30
Y	0	18	34	55	70	70	100
Estimation : $\hat{Y}$	2,4	17,9	33,4	48,9	64,4	79,9	95,4
Résidus : *e*	– 2,4	0,1	0,6	6,1	5,6	– 9,9	4,6

Classons les résidus par ordre croissant :

i	1	2	3	4	5	6	7
e_i	– 9,9	– 2,4	0,1	0,6	4,6	5,6	6,1

Valeur observée de *R* : $R_{obs} = \dfrac{e_2 - e_1}{e_7 - e_1} \simeq 0{,}75$.

Valeur critique au seuil de 0,05 : $r_{0{,}95} = 0{,}507$.

Décision : $0{,}75 > 0{,}507$, on rejette H_0 au seuil de 0,05 ce qui justifie que la valeur suspectée est aberrante.

Exemple 3

Une entreprise étudie la possibilité de lancer sur le marché un yaourt à la rhubarbe. Elle réalise des mesures de pH sur un échantillon de 11 pots. Les mesures observées sont les suivantes :

5,40 5,70 6,15 6,16 6,18 6,25 6,43 6,45 6,45 6,60 6,75

Existe-il une valeur aberrante ?

Dans un premier temps, nous allons effectuer un test de Dixon au seuil de risque 0,05 sur la valeur $x_1 = 5{,}40$ de manière ensuite à justifier la distinction qui doit être faite entre $n \leq 10$ et $n > 10$ pour la valeur observée de R.

Le nombre d'observations est ici 11 qui est supérieur à 10, que se passerait-il si on utilisait la valeur observée du cas $n \leq 10$?

$\dfrac{x_2 - x_1}{x_{11} - x_1} \simeq 0{,}222$ (à 10^{-3} près).

Bien que nous ne disposions pas de la valeur tabulée pour $n = 11$, il semble évident que la valeur critique $r_{0,95}$ serait largement supérieure à 0,222. Il faudrait donc en conclure que 5,40 n'est pas une valeur aberrante.

Cependant, si on élimine cette valeur de l'échantillon et que l'on effectue un test de Dixon au seuil 0,05 sur la valeur $x_2 = 5{,}70$ en considérant les 10 valeurs restantes, on observe alors que 5,70 est une valeur aberrante ($R_{obs} \simeq 0{,}429$ et $r_{0,95} = 0{,}412$)

Cette situation invite les étudiants à s'interroger sur cette anomalie car il parait évident que si la deuxième valeur est aberrante, la première l'est tout autant. L'erreur de décision qui est faite en utilisant $\dfrac{x_2 - x_1}{x_{11} - x_1}$ se justifie par le fait que les deux premières valeurs sont proches et toutes deux aberrantes.

On vérifie alors que l'utilisation de $\dfrac{x_3 - x_1}{x_{10} - x_1}$ permet de conclure à l'aberration de la première valeur ($R_{obs} \simeq 0{,}714$ et $r_{0,95} = 0{,}637$).

Pour des échantillons de taille strictement supérieure à 10, le calcul de $R_{obs} = \dfrac{x_3 - x_1}{x_{n-2} - x_1}$ prend en compte la possibilité d'avoir deux valeurs aberrantes inférieures (x_1 et x_2).

Cette situation est plus *rare* avec des tailles d'échantillon faibles ($n \leq 10$).

III. Cas de deux valeurs aberrantes

Pour appliquer la méthode, il faut dans ce cas que $n > 10$.

Plusieurs situations sont possibles :

1) Si les résultats douteux sont x_1 et x_n, on applique successivement le test de Dixon aux deux valeurs séparément.

2) Si les deux résultats douteux sont "du même côté", on applique le test à l'avant dernière, après avoir éliminé provisoirement la dernière (comme dans l'exemple 3).

Concrètement, s'il s'agit de x_1 et x_2, après avoir éliminé x_1, on applique le test à x_2 en prenant $R_{obs} = \dfrac{x_4 - x_2}{x_{n-2} - x_2}$

S'il s'agit de x_{n-1} et x_n, après avoir éliminé x_n, on applique le test à x_{n-1}en prenant $R_{obs} = \frac{x_{n-1} - x_{n-3}}{x_{n-1} - x_4}$.

Si le test conduit à considérer x_2 (respectivement x_{n-1}) comme aberrantes, alors x_1 (respectivement x_n) l'est aussi. Sinon on lui applique le test à son tour.

Complément : Test de Grubbs (hors programme)

C'est un test beaucoup plus puissant dans le cas des petits échantillons.

Il permet de rejeter deux valeurs aberrantes dans une série de mesures, ou encore de rejeter une ou deux moyennes par rapport à la moyenne générale.

Il est basé sur le calcul des résidus normalisés : $G = \frac{\overline{x} - x_1}{s}$ ou $G = \frac{x_n - \overline{x}}{s}$.

Mais ceci est une autre histoire…

Une idée, pour finir

On peut proposer ce test dans le cadre de l'objectif 4.1 du module M42 : *Explorer et mettre en œuvre les fonctions avancées du tableur pour résoudre un problème, notamment dans le domaine professionnel de l'option du BTSA.*

Cette séance de TD pourrait être l'occasion d'utiliser les fonctions RECHERCHEV(), NBVAL et SI, ainsi que des commandes de tri.

En guise d'exemple, vous pouvez trouver le fichier nous ayant permis de faire les calculs dans cet article, à l'adresse suivante : http://www.enfa.fr/r2math

Bibliographie

Article de Dean et Dixon :

http://depa.pquim.unam.mx/amyd/archivero/ac1951_23_636_13353.pdf

Table de la loi de Dixon

Valeur de $r_{1-\alpha}$

α / n	**0,01**	**0,05**
3	0,988	0,941
4	0,889	0,765
5	0,780	0,642
6	0,698	0,560
7	0,637	0,507
8	0,590	0,468
9	0,555	0,437
10	0,527	0,412
11	0,745	0,637
12	0,704	0,600
13	0,670	0,570
14	0,641	0,546
15	0,616	0,525
16	0,595	0,507
17	0,577	0,490
18	0,561	0,475
19	0,547	0,462
20	0,535	0,450
21	0,524	0,440
22	0,514	0,430
23	0,505	0,421
24	0,497	0,413
25	0,489	0,406
26	0,486	0,399
27	0,475	0,393
28	0,469	0,387
29	0,463	0,381
30	0,457	0,376

Simplified Statistics for Small Numbers of Observations

R. B. DEAN AND W. J. DIXON
University of Oregon, Eugene, Ore.

Several short-cut statistical methods can be used with considerable saving of time and labor. This paper presents some useful techniques applicable to a small number of observations. Advantages of the median as a substitute for the average are discussed. The median is especially useful when a quick decision must be made and gross errors are suspected. The range may be used to estimate the standard deviation and confidence intervals with little loss in precision for small numbers of observations. A convenient criterion for rejecting gross errors is presented. A table gives efficiencies and conversion factors for two to ten observations. The statistical methods presented are sufficiently simple to find use by those who feel that "statistics are too much trouble." These methods are usually more exact than the statistics of large numbers applied directly to small numbers of observations.

ALTHOUGH all scientific workers are aware that their numerical observations may be analyzed by statistical methods, many have expressed the opinion that the classical methods of statistics are unnecessarily complicated. It is not generally realized that there are available a number of short-cut statistics which will yield a large fraction of the available information with less computation. This paper calls attention to some of these statistics which are applicable to ten or fewer independent observations. These methods are especially valuable when it is necessary to make quick decisions.

SCATTERING OF DATA

In general, repeated observations of the "same" quantity will not be identical. We normally recognize this tendency toward scattering of the data by calculating an average value which is an estimate of the "true" value (the average of the entire population of measurements). For this purpose it is customary to calculate the arithmetic mean, $\bar{x}$. In order to follow general usage, the term "average" is used synonymously with "arithmetic mean" in this paper and is denoted by the symbol $\bar{x}$.

It is also generally recognized that the extent of the scattering is inversely related to the reliability of the measurements. The measures of this scattering which are frequently used are the standard deviation, s, and the average deviation.

Whenever a large number of independent fluctuations contribute to the spread in the values of a population, the values will approximate a normal bell-shaped error curve. This distribution of the population is the one for which most statistical tables have been prepared and is fortunately the one observed frequently in practice. In this paper all conclusions are based on a normally distributed population.

The tabulated values presented apply only if independent random samples are taken from the original population. For example, if five samples of an ore are individually dissolved and made up to volume and the iron in these samples is then determined in duplicate on aliquots of each sample, there would be only five independent observations representing the five averages of the aliquots.

Subject to the conditions that the observations are independent and normally distributed, the best statistics known are the arithmetic mean, $\bar{x}$, for the average, and the standard deviation, s, or the variance, s^2, for the spread.

The term "best" applied to $\bar{x}$ and s^2 is based on the fact that these estimates are the most efficient known estimates for the mean and variance, respectively, of the population when the observations follow the normal (or Gaussian) law of errors. The term "most efficient" indicates a minimum variance or dispersion for the sampling distributions of these statistics—i.e., least amount of spread for repeated estimates of the population mean or population variance.

The classical statistical theory for large numbers of observations is presented, in part, in most textbooks of analytical chemistry (*11, 16*). Rarely do these books recognize the fact that parts of this theory are inaccurate for small numbers of observations. For example, we can no longer expect the interval $\bar{x} \pm 0.6745s/\sqrt{n}$ to include the population mean 50% of the time when the $\bar{x}$ and s are calculated from n observations. Tables for the correct multiplier (Student's t) to replace 0.6745 for various numbers of observations and for various probabilities of covering the population mean are available (*8, 14*). When n is greater than 20 to 30, Student's t is closely approximated by the normal curve of error. To make use of the exact statistical methods, most chemists will have to use tables for the chosen number of observations, for very few analyses are run on as many as 20 samples or replicates. The less efficient statistics presented in this paper also require the use of tables, but the computations are less tedious and the efficiency is adequate for many purposes. Tables of useful functions are presented for two to ten observations (*5*). Accurate values of some of these functions are not yet available for larger numbers of observations.

For convenience a series of observations will be arranged in ascending order of magnitude and assigned the symbols:

$$x_1,\ x_2,\ x_3,\ \ldots\ldots\ x_{n-1},\ x_n$$

The best estimate of the central value is the average, $\bar{x}$

$$\bar{x} = \frac{(x_1 + x_2 + \ldots + \ldots x_{n-1} + x_n)}{n} = \frac{(\Sigma x_i)}{n} \qquad (1)$$

The spread of the values is measured most efficiently by the variance, s^2. The computation is as follows:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots (x_n - \bar{x})^2}{n-1} = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \qquad (2)$$

and

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \qquad (3)$$

It is essential to use $n - 1$ instead of n in the denominator to avoid a biased estimate of σ^2, the variance of the population. Instead of calculating the average, $\bar{x}$, we can use the median value, M, as a measure of the central value. M is the middle value or the average of the two values nearest the middle. If $n = 5$, then $M = x_3$; if $n = 6$, then $M = (x_3 + x_4)/2$. M is an estimate of the central value and has the advantage that it is not influenced markedly by extraneous values. The efficiency (defined here as the ratio of the variances of the sampling distributions of these two estimates of the "true" mean) of M, E_M, is given in column 2 of Table I (*4, 10*). It varies from 1.00 for only two observations (where the median is identical with the average, $\bar{x}$) down to 0.64 for large numbers of observations. It can be shown (*4, 10*) that the numerical value of the efficiency implies, for example, that the median, M, from 100 observations (where the efficiency is 0.64) conveys as much information about

Table I. Efficiencies and Conversion Factors for Two to Ten Observations

No. of Observations, n	Efficiency: Of median, E_M	Efficiency: Of range, E_W	Deviation Factor, K_W	Confidence Factors, Student's: $t_{0.95}$	Confidence Factors, Student's: $t_{0.99}$	Confidence Factors, Range: $t_{w\,0.95}$	Confidence Factors, Range: $t_{w\,0.99}$	Rejection Quotient $Q_{0.90}$
2	1.00	1.00	0.89	12.7	64	6.4	31.83	..
3	0.74	0.99	0.59	4.3	10	1.3	3.01	0.94
4	0.84	0.98	0.49	3.2	5.8	0.72	1.32	0.76
5	0.69	0.96	0.43	2.8	4.6	0.51	0.84	0.64
6	0.78	0.93	0.40	2.6	4.0	0.40	0.63	0.56
7	0.67	0.91	0.37	2.5	3.7	0.33	0.51	0.51
8	0.74	0.89	0.35	2.4	3.5	0.29	0.43	0.47
9	0.65	0.87	0.34	2.3	3.4	0.26	0.37	0.44
10	0.71	0.85	0.33	2.26	3.25	0.23	0.33	0.41
∞	0.64	0.00	0.00	1.960	2.576	0.00	0.00	0.00

the central value of the population as the average, $\bar{x}$, calculated from 64 observations.

The median of a small number of observations can usually be determined by inspection. Although it is less efficient than the average if the population is normally distributed, M may be more efficient than the average, $\bar{x}$, if gross errors are present. A chemist is frequently faced with the problem of deciding whether or not to reject an observation that deviates greatly from the rest of the data. The median is obviously less influenced by a gross error than is the average. It may be desirable to use the median in order to avoid deciding whether a gross error is present. It has been shown that for three observations from a normal population the median is better than the "best two out of three"—i.e., average of the two closest observations (*18*).

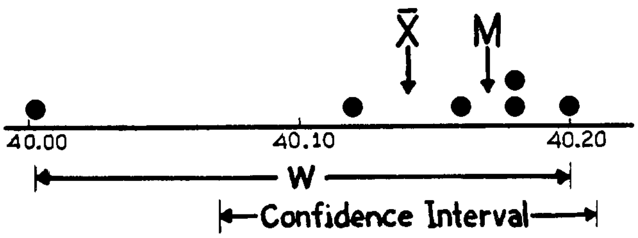


Figure 1. Graphical Presentation of Data Chosen in Example

No attempt is made to give a complete treatment of the problem of gross errors here, but an approach based on a recent fairly complete summary of this problem (*2, 3*) is given in the last section of this paper.

The following series of observations represents calculated percentages of sodium oxide in soda ash. The data have been arranged in order of magnitude, and are presented graphically in Figures 1 and 2.

40.02	x_1
40.12	x_2
40.16	x_3
40.18	x_4
40.18	x_5
40.20	x_6
$\bar{x}$ = 40.14	M = 40.17

The first value may be considered doubtful. The average is 40.14 and the median is 40.17. We may wish to place more confidence in the median, 40.17, than in the average, 40.14. (If the series were obtained in the order given, we might justifiably rule out the first observation as showing lack of experience or other "starting up" errors.)

The range of the observations, w, is the difference between the greatest and least value; $w = x_n - x_1$. The range, w, is a convenient measure of the dispersion. It is highly efficient for ten or fewer observations, as is evident from column 3 of Table I (*5, 17*). This high relative efficiency arises in part from the fact that the standard deviation is a poor estimate of the dispersion for a small number of observations, even though it is the best known estimate for a given set of data. The range is also more efficient than the average deviation for less than eight observations. To convert the range to a measure of dispersion independent of the number of observations we must multiply by the factor, K_w, which is tabulated in column 4 (*5, 17*). This factor adjusts the range, w, so that on the average we estimate the standard deviation of the population. The product $wK_w = s_w$ is therefore an estimate of the standard deviation which can be obtained from the range. In the series presented above, the range is 0.18. From the table we find that K_w for six observations is 0.40, so $s_w = 0.072$. The standard deviation, s, calculated according to Equation 3, equals 0.066.

As n increases the efficiency of the range decreases. If the data are randomly presented, such as in order of production rather than in order of size, the average of the ranges of successive subgroups of 6 or 8 is more efficient than a single range. The same table of multipliers is used, the appropriate K_w being determined by the subgroup size. This factor is the reciprocal of d_2, the divisor used—e.g., in quality control work—to convert the range to an estimate of the standard deviation. Tables of d_2 are given for n up to 100 (*1, 9*), but Lord has shown in a recent publication (*13, 15*) that the most efficient subgroup size for estimating the standard deviation from the average range is from 6 to 8.

CONFIDENCE LIMITS

Although s and s_w are useful measures of the dispersion of the original data, we are usually more interested in the confidence interval or confidence limits. By the confidence interval we mean the distance on either side of $\bar{x}$ in which we would expect to find, with a given probability, the "true" central value. For example, we would expect the true average to be covered by the 95% confidence limits 95% of the time. By taking wider confidence limits, say 99% limits, we can increase our chances of catching the "true" average but the interval will necessarily be longer. The shortest interval for a given probability corresponds to the "t test" of Student (*8, 14*). $\bar{x} \pm ts/\sqrt{n}$ is the confidence interval. t varies with the number of observations and the degree of confidence desired. For convenience t is tabulated for 95 and 99% confidence values (columns 5 and 6).

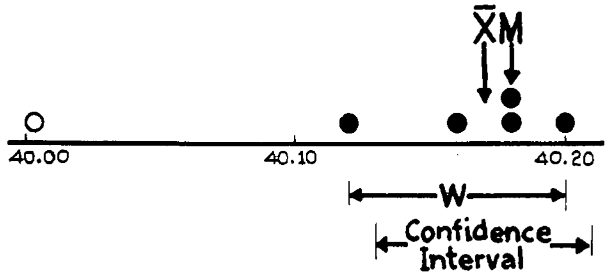


Figure 2. Graphical Presentation of Charges Produced by Rejection of a Questionable Observation

Confidence limits might be calculated in a similar manner, using s_w obtained from the range and a corresponding but different table for t. However, it is more convenient to calculate the limits directly from the range as $\bar{x} \pm wt_w$. The factor for converting w to s_w has been included in the quantity t_w, which is tabulated in column 7 for 95% confidence and in column 8 for 99% confidence (*6, 12*).

For a set of six observations t_w is 0.40 and the range of the set previously listed is 0.18, so wt_w equals 0.072. Hence, we can report as a 95% confidence interval, 40.14 ± 0.072. If we calculate a confidence interval using the tables of t and the calculated value of s we find $t = 2.6$ at the 95% level and $ts\sqrt{6} = 0.078$. We would report 40.14 ± 0.078, a result substantially the same as that obtained from the range of this particular sample.

If the standard deviation of a given population is known or assumed from previous data, we can use the normal curve to calculate the confidence limits. This situation may arise when a given analysis has been used for fifty or so sets of analyses of similar samples. The standard deviation of the population may be estimated by averaging the variance, s^2, of the sets of observations or from K_w times the average range of the sets of observations. The interval $\bar{x} \pm 1.96s/\sqrt{m}$, where $\bar{x}$ is computed from a new set of m observations, can be expected to include the population average 95% of the time.

EXTRANEOUS VALUES

Simplified statistics have been presented, which enable one to obtain estimates of a central value and to set confidence limits on the result. Let us consider the problem of extraneous values. The use of the median eliminates a large part of the effect of extraneous values on the estimate of the central value, but the range obviously gives unnecessary weight to an extraneous value in an estimate of the dispersion. On the other hand, we may wish to eliminate extraneous values which fail to pass a screening test.

One very simple test, the Q test, is as follows:

Calculate the distance of a doubtful observation from its nearest neighbor, then divide this distance by the range. The ratio is Q where

$$Q = (x_2 - x_1)/w$$

or

$$Q = (x_n - x_{n-1})/w$$

If Q exceeds the tabulated values, the questionable observation may be rejected with 90% confidence (*2*, *3*, *7*). In the example cited 40.02 is the questionable value and

$$Q = (40.12 - 40.02)/(40.20 - 40.02) = 0.56$$

This just equals the tabulated value of 0.56 for 90% confidence, so we may wish to reject the value.

If we had decided to reject an extreme low value if Q is as large or larger than would occur 90% of the time in sets of observations from a normal population, we would now reject the observation 40.02. In other words, a deviation this great or greater would occur by chance only 10% of the time at one or the other end of a set of observations from a normally distributed population.

By rejecting the first value we increase the median from 40.17 to 40.18 and the average from 40.14 to 40.17 (see Figure 2). The standard deviation, s, falls from 0.067 to 0.030 (it might be greater than 0.030 if we have erred in rejecting the value 40.02) and s_w falls from 0.072 to 0.034. The 95% confidence interval corresponding to the t test is now 40.17 ± 0.038 and from the median and range is 40.18 ± 0.040, a reduction of about one half in the length of the interval. (The 95% is only approximate, as we have performed an intermediate statistical test.)

LITERATURE CITED

(1) Am. Soc. Testing Materials, "Manual on Presentation of Data," 1945.
(2) Dixon, W. J., *Ann. Math. Stat.*, **21**, 488 (1950).
(3) *Ibid.*, in press.
(4) Dixon, W. J., and Massey, F. J., "Introduction to Statistical Analysis," p. 238, New York, McGraw-Hill Book Co., 1951.
(5) *Ibid.*, p. 239.
(6) *Ibid.*, p. 242.
(7) *Ibid.*, p. 319.
(8) Fisher, R. A., and Yates, Frank, "Statistical Tables for Biological, Agricultural, and Medical Research," Edinburgh, Oliver and Boyd, 1943.
(9) Grant, E. L., "Statistical Quality Control," p. 536, New York, McGraw-Hill Book Co., 1946.
(10) Kendall, M. G., "Advanced Theory of Statistics," Vol. II, p. 6, London, Charles Griffin & Co., 1946.
(11) Kolthoff, I. M., and Sandell, E. B., "Textbook of Quantitative Inorganic Analysis," p. 270, New York, Macmillan Co., 1948.
(12) Lord, E., *Biometrika*, **34**, 41 (1947).
(13) *Ibid.*, **37**, 64 (1950).
(14) Merrington, Maxine, *Ibid.*, **32**, 300 (1942).
(15) Pearson, E. S., *Ibid.*, **37**, 88 (1950).
(16) Pierce, W. C., and Haenisch, E. L., "Quantitative Analysis," p. 48, New York, John Wiley & Sons, 1948.
(17) Tippett, L. H. C., *Biometrika*, **17**, 364 (1925).
(18) Youden, W., unpublished data.

RECEIVED May 27, 1950. Presented before the Division of Analytical Chemistry at the 117th Meeting of the AMERICAN CHEMICAL SOCIETY, Houston, Tex.

Determination of Intrinsic Low Stress Properties of Rubber Compounds

USE OF INCLINED PLANE TESTER

W. B. DUNLAP, JR., C. J. GLASER, JR., AND A. H. NELLEN
Lee Rubber & Tire Corp., Conshohocken, Pa.

PHYSICAL test methods for vulcanized natural and synthetic rubber compounds as commonly employed today do not provide an accurate basis for evaluating service performance, especially in the case of tire compounds. Poor correlation between ordinary laboratory physical tests and observed performance in tire service on the road is the rule rather than the exception.

Substantial improvement is needed in developing tests which are less subject to the human variable, tests in which the criteria examined are more in line with the actual service conditions, and tests which are mechanically simple and easy to perform. One significant advance has been made in this direction by the development of the National Bureau of Standards' comparatively new strain test (*3*). This paper reports progress made by the Lee laboratories in developing one phase of an evaluation system which more accurately predicts compound performance than will the methods previously employed.

As a starting point for this work, it was decided first to determine the important physical characteristics of a tire compound and, secondly, to devise a new test or alter an existing test to give the desired information. It is assumed that an accurate appraisal of the physical characteristics listed in Table I is required.

LES TESTS STATISTIQUES
AVEC UNE CALCULATRICE OU GEOGEBRA
- Partie 1 -

La plupart des calculatrices permettent de traiter des tests statistiques. La dernière version 4.2 de Geogebra permet aussi d'effectuer des tests à partir de données.

Notre intention, avec cet article est, d'une part, de faire le lien entre nos calculs habituels et ceux de la machine ou de GeoGebra, et d'autre part, de réaliser une fiche pratique d'utilisation des calculatrices.

Nous traiterons successivement dans deux articles trois types de tests : tests de conformité (partie 1), tests de comparaison et tests du Khi-deux (partie 2).

Test de conformité d'une moyenne

Exemple :

On souhaite contrôler un lot de pré-emballages de quantité nominale 1 000 mL.
Le caractère *volume de ces pré-emballages* est distribué selon la loi normale de moyenne μ et d'écart-type σ. On cherche à savoir, au vu d'un échantillon aléatoire simple de 19 pré-emballages, si la contenance moyenne du lot n'est pas inférieure à 1 000 mL.
Les contenances, en mL, des pré-emballages de l'échantillon sont :
999,49 – 997,83 – 1 000,42 – 1 002,17 – 1 002,95 – 996,26 – 999,60 – 1 001,86 – 998,15 – 997,13 – 999,83 – 996,86 – 998,34 – 998,69 – 996,40 – 999,03 – 999,31 – 1 000,23 – 999,38.
On veut élaborer un test permettant de conclure si le lot est conforme, au seuil de risque de 0,05.

Travail papier/crayon

Rédigeons *nos* calculs habituels et énonçons *notre* règle de décision

<u>Population</u> : Ensemble des pré-emballages de la fabrication.

<u>Caractère observé</u> : Volume, de moyenne μ et d'écart-type σ dans le population.

On réalise un test unilatéral de conformité d'une moyenne à l'aide d'un échantillon aléatoire simple de taille 19.

<u>Choix des hypothèses</u> : H_0 : "$\mu = 1\,000$ mL" et H_1 : "$\mu < 1\,000$ mL"

<u>Variable de décision</u> :

Le caractère est distribué normalement, σ est inconnu et la taille de l'échantillon, 19, est inférieure à 30 ; alors la variable de décision est $T = \dfrac{\bar{X} - 1\,000}{\dfrac{S}{\sqrt{18}}}$ qui s'écrit aussi $\dfrac{\bar{X} - 1\,000}{\dfrac{\hat{S}}{\sqrt{19}}}$

qui a pour loi de probabilité, sous l'hypothèse H_0, la loi de Student à 18 degrés de liberté.

On désigne par $\bar{X}$, S et $\hat{S}$ les variables aléatoires, qui à chaque échantillon de taille 19, associent respectivement la moyenne, l'écart-type et l'écart-type corrigé des 19 volumes de l'échantillon.

<u>Seuil de risque fixé</u> : $\alpha = 0{,}05$

<u>Détermination de la zone d'acceptation</u> : Soit t_α la valeur critique associée à α.

$$P(T < t_\alpha) = 0{,}05 \iff P(T < -t_\alpha) = 0{,}95 \iff -t_\alpha \simeq 1{,}73 \iff t_\alpha \simeq -1{,}73$$

<u>Règle de décision n° 1</u> : Soit t_{obs} la valeur observée de T sur l'échantillon.

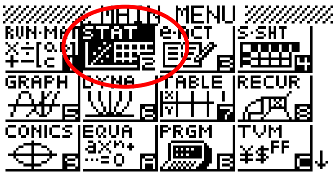
Si $t_{obs} < -1{,}73$, on rejette l'hypothèse H_0.
Si $t_{obs} > -1{,}73$, on n'est pas en mesure de rejeter l'hypothèse H_0.

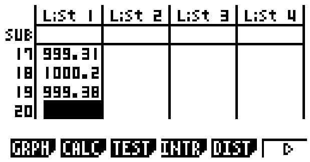
Avec une calculatrice

On sait que l'on va devoir réaliser un test de Student de conformité d'une moyenne, unilatéral à gauche. On a préalablement entré les données de l'échantillon dans une liste statistique (ici List 1).

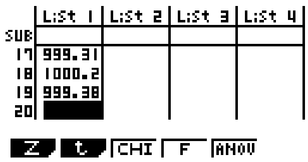
sur CASIO 85

touche MENU, option STAT

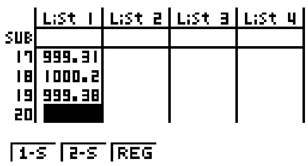




touche F3 : onglet TEST



touche F2 : onglet t

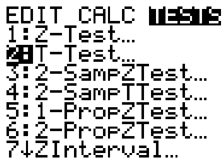


puis touche F1 1-S

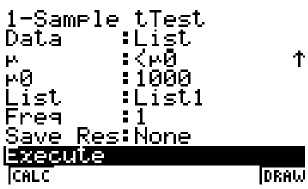
(Test de conformité)

sur TI 83

touche stats, onglet TESTS

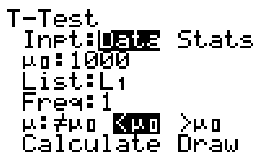


2:T-Test (test de conformité d'une moyenne)



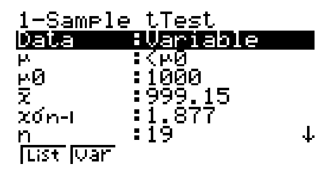
On complète la fenêtre en fonction de :

- l'échantillon (ici, **Data**, données entrées dans la liste 1)
- l'hypothèse H_0 (ici, $\mu_0 = 1\,000$)
- l'hypothèse H_1 (ici, $\mu < \mu_0$)



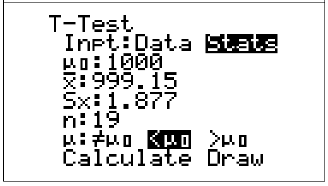
<u>Remarque</u> : Si on n'avait connu de l'échantillon que sa moyenne $\bar{x} \simeq 999{,}15$ et son écart-type $\sigma \simeq 1{,}827$, on aurait modifié la fenêtre précédente ainsi :

sur CASIO 85



Attention ! : $x\sigma_{n-1}$ ou Sx désignent l'écart-type corrigé, à calculer au préalable

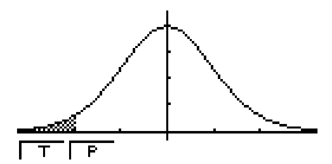
sur TI 83



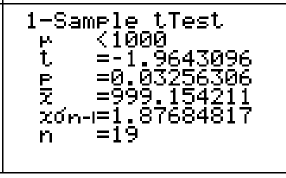
Les résultats obtenus peuvent dans ce cas différer alors un peu des précédents du fait de la saisie arrondie de la moyenne et de l'écart-type corrigé.

Et on obtient les résultats suivants, en sélectionnant DRAW ou CALC :

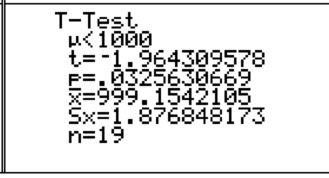
DRAW sur CASIO



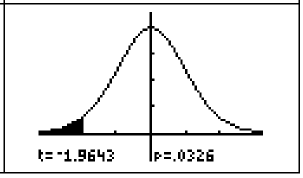
CALC sur CASIO



CALCULATE sur TI



DRAW sur TI



Le **t** affiché est la valeur de T observée sur l'échantillon : $t_{obs} = \dfrac{999{,}154 - 1\,000}{\dfrac{1{,}877}{\sqrt{19}}} \simeq -1{,}96$

Le **p** affiché est la probabilité que, sous l'hypothèse H_0, la variable aléatoire T prenne une valeur inférieure à t_{obs}.

Ici, $\mathbf{p} = P(T < -1{,}96) \simeq 0{,}33$ quand la loi de probabilité de T est la loi de Student à 18 ddl.

Exploitons les résultats donnés par la calculatrice pour rédiger une règle de décision:

<u>Règle de décision n° 2</u> :

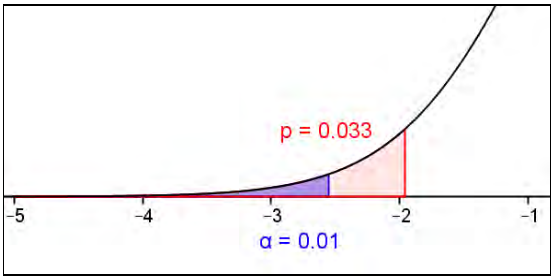
Si $\alpha > 0{,}033$, on refuse l'hypothèse H_0 (alors t_{obs} sera inférieure à la valeur critique associée à α).

Si $\alpha < 0{,}033$, on n'est pas en mesure de rejeter H_0 (alors t_{obs} sera supérieure à la valeur critique associée à α.).

Il ne reste plus qu'à comparer la valeur de **p** avec le seuil de risque α fixé, risque de rejeter H_0 à tort.

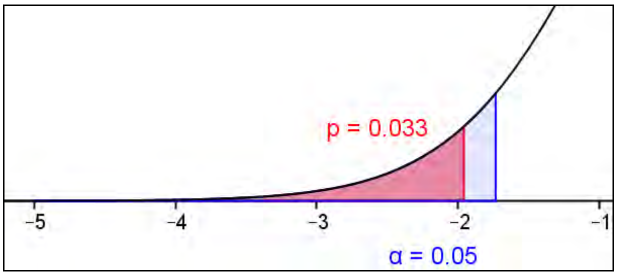
Si $\alpha = 0{,}01$:

$\alpha < \mathbf{p}$, on n'est pas en mesure de refuser l'hypothèse H_0.



Si $\alpha = 0{,}05$:

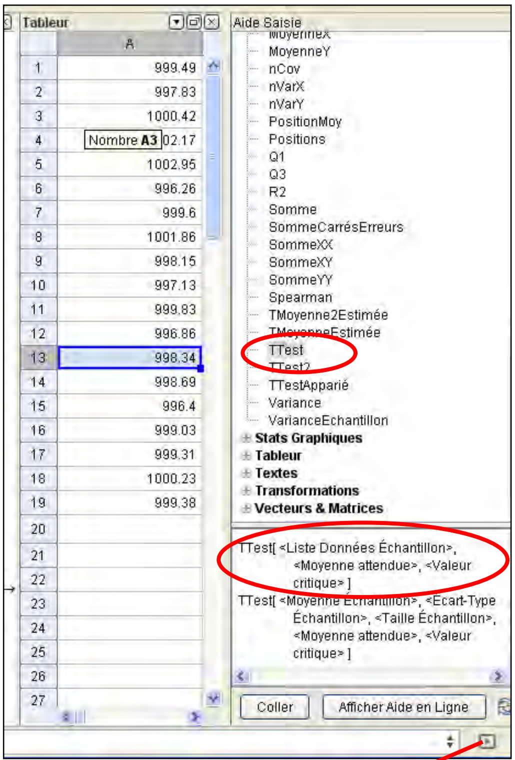
$\alpha > \mathbf{p}$, on refusera l'hypothèse H_0.



Avec Geogebra

On saisit les valeurs dans le tableur de GeoGebra. (Attention à ne pas mettre de séparateur de milliers et de prendre le point comme séparateur décimal !!).

On affiche la liste des commandes et dans la rubrique **Statistiques**, on choisit la commande **TTest**. Ici, on travaille avec les données brutes, on utilise la première option proposée.



Affichage de la liste des commandes

La syntaxe est **TTest[<*Liste Données Échantillon*>,<*Moyenne attendue*>,<*Symbole*>]**

- *Liste Données Échantillon* désigne la liste des n valeurs de l'échantillon qui peut être une liste (entre accolades séparées par des virgule) ou une plage du tableur ;
- *Moyenne attendue* est la valeur m_0 de la moyenne de référence ;
- *Symbole* prend une des trois formes : "<", ">" ou "≠" selon que H_1 est $\mu < m_0$, $\mu > m_0$ ou $\mu \neq m_0$

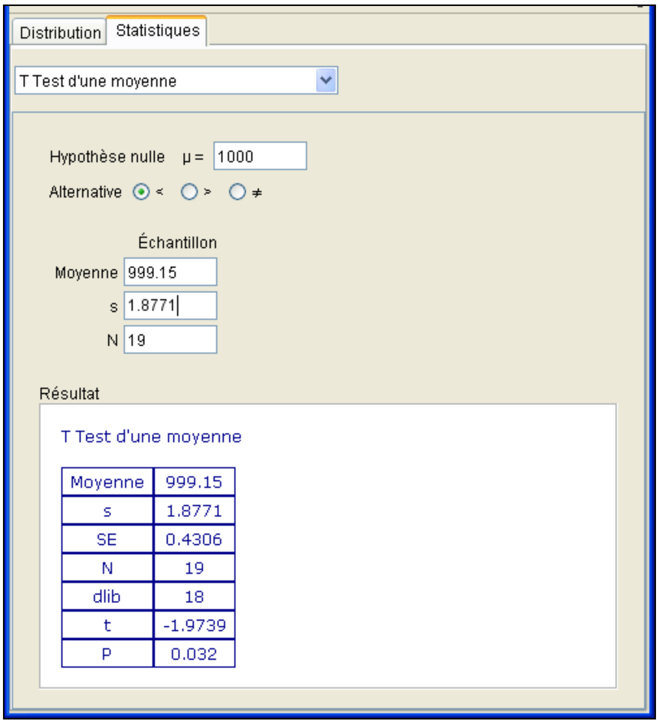
Le résultat est retourné sous la forme d'une liste : $\{p\text{-}value, t_{obs}\}$ où *p-value* est l'une des trois probabilités $P(T \leq t_{obs})$ $P(T \geq t_{obs})$ ou $P(|T| \geq t_{obs})$ selon que H_1 est $\mu < m_0$, $\mu > m_0$ ou $\mu \neq m_0$ où T suit la loi de Student avec $n - 1$ degré de liberté.

Ici, la formule est **TTest[A1:A19,1000,"<"]**. On peut la saisir dans la ligne de saisie ou dans une cellule du tableur. On obtient la liste $\{0.0326\,, -1.9643\}$.
Il reste à comparer α avec *p-value* :

- Si $\alpha > 0{,}0326$, on rejette l'hypothèse H_0 (t_{obs} est inférieure à la valeur critique associée à α).
- Si $\alpha < 0{,}0326$, on n'est pas en mesure de rejeter H_0 (t_{obs} est supérieure à la valeur critique).

Ici, on rejette l'hypothèse H_0 au seuil de 5 %. La contenance moyenne des pré-emballages semble inférieure à 1 000 mL.

Dans le cas où on ne connaît l'échantillon par la moyenne et l'écart-type du caractère, on peut aussi utiliser l'onglet *Statistiques* proposé dans le module *Calculs de probabilités*.



<u>Remarque</u> :

Dans ce cas, on peut utiliser la deuxième option proposée pour la commande TTest, mais elle est moins pratique que le module *Calculs probabilités* !!.

Test de conformité d'une proportion – Cas des grands échantillons

Exemple :

Sur un échantillon aléatoire simple de 200 individus d'une commune, 90 sont favorables à l'implantation d'un centre commercial.
Au vu de cet échantillon et au risque de 0,05, peut-on refuser l'hypothèse que la moitié des habitants de la commune est favorable à l'implantation du centre commercial ?

Et là, on ne panique plus, car on a suivi assidûment les cours et on a déjà réalisé l'étude préalable...

Travail papier/crayon

<u>Population</u> : Ensemble des habitants d'une commune

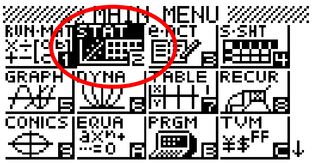
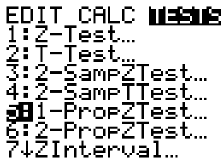
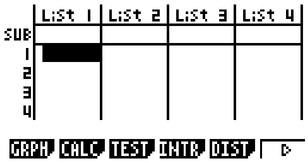
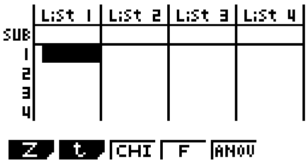
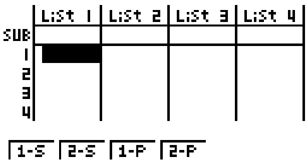
<u>Caractère observé</u> : Le fait d'être favorable à l'implantation d'un centre commercial, de proportion p dans la population.

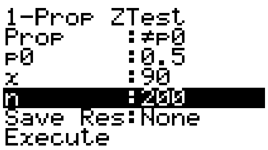
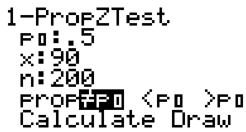
<u>Choix des hypothèses</u> : $H_0 : p = 0,5$ et $H_1 : p \neq 0,5$

<u>Variable de décision</u> : $n = 200$ donc $n > 30$ alors la variable de décision est $Z = \dfrac{F - 0,5}{\sqrt{\dfrac{0,5 \times 0,5}{200}}}$,

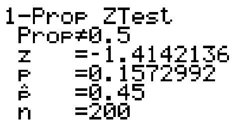
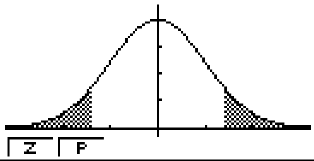
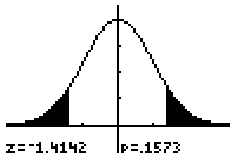
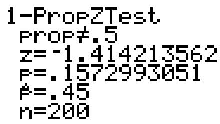
dont la loi de probabilité, sous l'hypothèse H_0, est approchée par la loi normale centrée réduite.

Avec une calculatrice

sur CASIO 85	sur TI 83
touche **MENU**, option **STAT** 	touche **stats**, onglet **TESTS**  **5:1-PropZTest** (test de conformité d'une proportion)
 touche **F3** : onglet **TEST**	
 touche **F1** : onglet **Z**	
 puis touche **F3** **1-P** (Test de conformité)	

	On complète la fenêtre en fonction de : - l'hypothèse H_1 - l'hypothèse H_0 - les données de l'échantillon (en effectifs)	

On obtient les résultats suivants :

DRAW sur CASIO	CALC sur CASIO	CALCULATE sur TI	DRAW sur TI
			

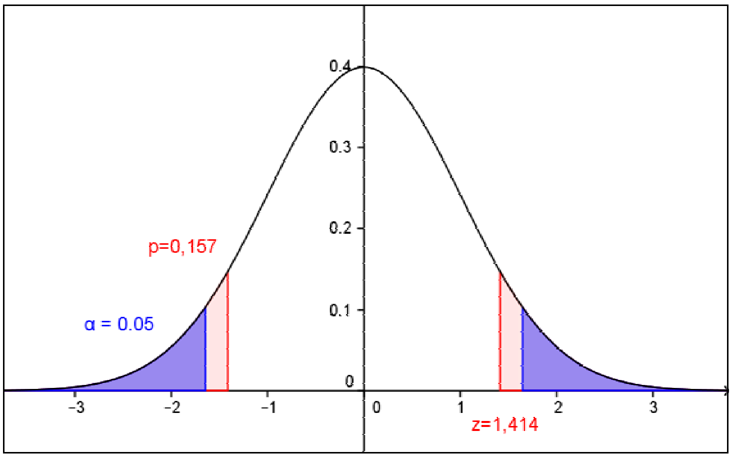
Le **z** affiché est la valeur observée de Z, soit ici $z_{obs} = \dfrac{\dfrac{90}{200} - 0{,}5}{\sqrt{\dfrac{0{,}5 \times 0{,}5}{200}}} \simeq -1{,}414236$.

Le **p** affiché est la probabilité, sous l'hypothèse H_0, que la variable aléatoire Z soit hors de l'intervalle $[-|z_{obs}|\,;\,|z_{obs}|]$.

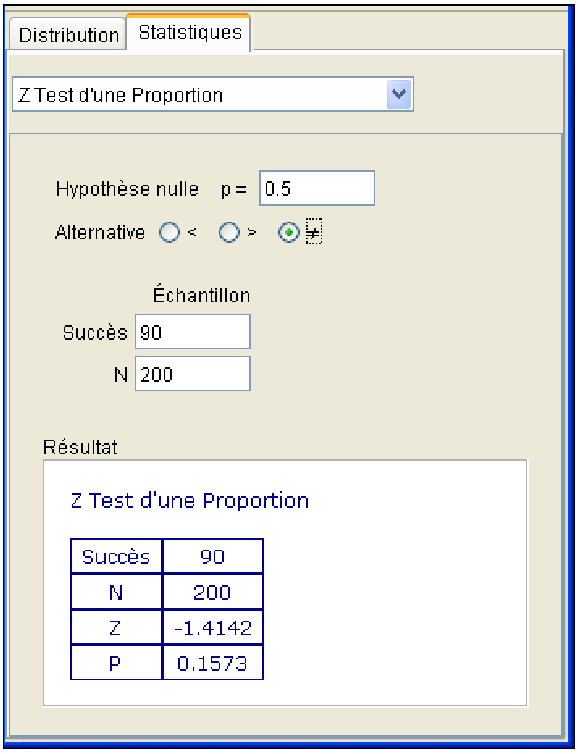
Ici, $\mathbf{p} = 1 - P(-1{,}4142 \leq Z \leq 1{,}4142)$ quand la loi de probabilité de Z est la loi normale centrée réduite. Finalement, $\mathbf{p} = 2 - 2 \times P(Z \leq 1{,}4142)$.
En utilisant la table de la loi normale centrée réduite, on obtient $\mathbf{p} \simeq 0{,}1586$ du fait des arrondis.
Cette valeur **p** est aussi appelée **p-value** ou probabilité critique.

Il ne reste plus qu'à comparer cette valeur de **p** avec le seuil de risque α fixé.

Ici $\alpha = 0{,}05$ et donc $\alpha < \mathbf{p}$: on n'est pas en mesure de rejeter l'hypothèse H_0.



Avec Geogebra



On peut utiliser l'onglet *Statistiques* proposé dans le module *Calculs de probabilités*.

Il est évident que la bonne compréhension des résultats affichés par la calculatrice ou Geogebra (ou même un tableur) ne peut se faire qu'après une bonne compréhension du principe des tests.

Nous en resterons là pour cet article et vous proposerons la suite très bientôt.