REVUE DE STATISTIQUE APPLIQUÉE

J. BAYART

Trois abaques originaux évitant les calculs dans les estimations sur échantillon

Revue de statistique appliquée, tome 8, n° 2 (1960), p. 77-86
<http://www.numdam.org/item?id=RSA_1960__8_2_77_0>

TROIS ABAQUES ORIGINAUX ÉVITANT LES CALCULS DANS LES ESTIMATIONS SUR ÉCHANTILLON

par J. BAYART

Statisticien-Conseil

INTRODUCTION -

Les ABAQUES ou NOMOGRAMMES constituent un outil commode pour les personnes qui, dans leur spécialité, sont appelées à faire fréquemment les mêmes calculs, basés sur une même relation algébrique.

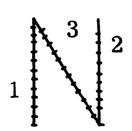
Ils jouent le même rôle que les tables de calculs tout faits - souvent encombrantes et longues à consulter - et s'adaptent à des formules extrêmement variées, des plus simples aux plus complexes. Leur consultation est immédiate et, s'ils ont été bien construits, la précision qu'ils donnent correspond au problème posé.

Les abaques présentés dans ce document matérialisent tous une relation algébrique entre trois variables (ce qui est le cas le plus simple et ... le plus fréquent). Ils sont du type "à points alignés", ce qui signifie que pour les utiliser il suffit de réaliser un simple alignement à l'aide d'une règle plate.

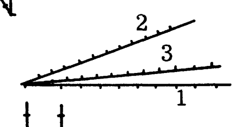
Chaque variable étant "cotée" sur une échelle graduée d'une façon appropriée (centimétrique, logarithmique, fonctionnelle de types divers), ces abaques sont constitués de trois échelles rectilignes. Dans certains cas, il faut faire appel à des échelles curvilignes.

Les abaques à trois échelles rectilignes peuvent avoir trois dispositions :

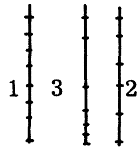
- abaque à échelles en N



- abaque à échelles concourantes



- abaque à échelles parallèles



On verra dans les pages qui suivent des exemples de ces trois dispositions.

L'étude et la mise au point des abaques constituent la NOMOGRAPHIE, branche des Mathématiques Appliquées qui peut rendre de grands services aux

Ingénieurs et Techniciens(1).

I - ABAQUE POUR LA DETERMINATION DE LA MARGE D'ERREUR A PREVOIR DANS L'ESTIMATION D'UNE MOYENNE -

1/ Enoncé du problème.

Je viens de recevoir un lot important de bobines dont le poids moyen annoncé par le fournisseur est de 250 grammes.

Je sais par des réceptions contrôlées antérieurement que les poids de ce genre de bobines sont distribués sensiblement suivant la loi de Gauss (courbe en cloche. (Cette exigence peut être ignorée si l'on fait plus de 30 mesures).

J'ai également vérifié antérieurement que le coefficient de variation des poids de ce genre de bobines est assez stable et de l'ordre de 4%(2).

Je pèse 5 bobines prises au hasard et j'obtiens le poids moyen de 240 grammes : dois-je suspecter le fournisseur de livrer des bobines trop légères ? - La conclusion serait-elle changée si j'avais pesé 20 bobines pour obtenir encore 240 grammes ?

2/ Mode opératoire.

a) Aligner (avec une règle plate) le point 4% de l'échelle "COEFFICIENT DE VARIATION" avec le point 5 de l'échelle "NOMBRE DE MESURES" : on lit sur l'échelle "MARGE D'ERREUR" : 5, 5%.

b) On peut déclarer qu'IL Y A 95 CHANCES SUR 100 QUE LA VRAIE MOYENNE DE L'ENSEMBLE DU LOT RECU EST COMPRISE ENTRE :

250 grammes + 5, 5% = 250 + 13, 75 = 263, 75 grammes

et

250 grammes - 5, 5% = 250 - 13, 75 = 236, 25 grammes

Comme la moyenne constatée (sur 5 mesures) est entre ces limites, il n'y a pas lieu de suspecter le lot reçu.

Mais si l'on a obtenu 240 grammes après 20 pesées, l'abaque montre que la marge d'erreur n'est plus, dans ce cas, que de 2%, d'où les nouvelles limites :

250 grammes + 2% = 250 + 5 = 255 grammes

et

250 grammes - 2% = 250 - 5 = 245 grammes

Cette fois, la moyenne constatée n'étant plus dans les limites, on est en droit de suspecter le trop de légèreté des bobines reçues. Si l'on réclame auprès du fournisseur, on court le risque de réclamer à tort de 5 chances sur 100, ce qui est un risque assez faible pour être négligé en général.

3/ Variante du problème précédent.

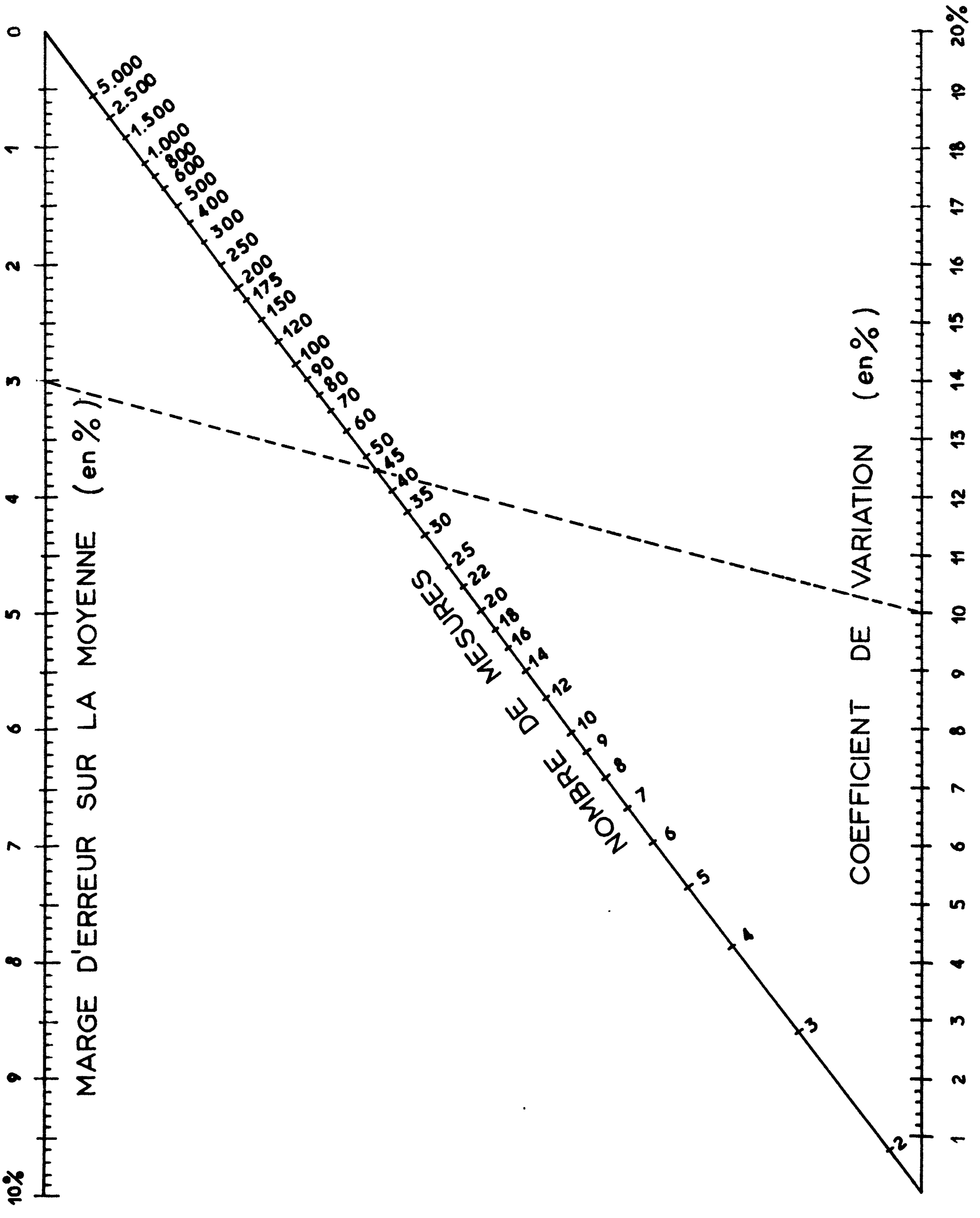
Je désire déterminer à 3% près (marge d'erreur) la résistance d'un certain type de fil dont le coefficient de variation (en résistance) est de l'ordre de

----

(1) Ouvrages recommandés : "Abaques et Nomogrammes" par A. Giet - Dunod
"Aide-mémoire de Géométrie Appliquée" par R. Beurrier - Dunod.

(2) Il faut - pour bien faire - avoir calculé le coefficient de variation sur trois livraisons (ou séries d'essais) antérieures, avec chaque fois une cinquantaine de mesures.

ESTIMATION D'UNE MOYENNE



10% : combien dois-je faire de mesures sur des tronçons de fil pris au hasard ?

En alignant 3% (MARGE D'ERREUR) avec 10% (COEFFICIENT DE VAR.) l'échelle centrale de l'abaque donne : NOMBRE DE MESURES : 45.

II - ABAQUES POUR LA DETERMINATION D'UN SONDAGE A DEUX DEGRES

1/ Enoncé du problème.

Supposons que l'on veuille déterminer le prix moyen de vente au détail du litre de vin rouge ordinaire sur l'ensemble des 63 villes de plus de 50 000 habitants (L = 63).

On sait par des enquêtes antérieures que le coefficient de variation des prix des denrées alimentaires entre ces villes (unités primaires) est de 5% et que le coefficient de variation des prix entre les détaillants (unités secondaires) d'une quelconque de ces villes est de 3%.

On demande dans combien de villes et chez combien de détaillants il y a lieu d'enquêter pour obtenir le prix moyen en question à 1% près (marge d'erreur) et avec un niveau de certitude de 95 chances sur 100 ?

2/ Mode opératoire.

a) Aligner sur la première planche L = 63 villes, avec CV_1 = 5% (coefficient de variation des unités primaires) : on obtient sur l'échelle centrale du faisceau :

Nombre minimum d'unités primaires (villes) à sonder : k_{min} = 38

Ce premier résultat signifie que si l'on a l'intention d'aller enquêter chez TOUS LES DETAILLANTS (unités secondaires) des villes tirées au sort, IL SUFFIRA d'enquêter dans 38 villes (unités primaires).

Cette solution est certainement onéreuse car il n'est vraisemblablement pas nécessaire de voir tous les détaillants d'une même ville. On peut assouplir cette première solution en faisant intervenir maintenant le coefficient de variation intérieur aux villes CV_2.

b) Calculer le rapport CV_2/CV_1 en divisant le coefficient de variation primaire (villes) par le coefficient de variation secondaire (détaillants d'une même ville). On obtient ici :

$$CV_2/CV_1 = 3\% : 5\% = 0,6$$

c) Aligner sur la deuxième planche le rapport ci-dessus (0, 6) lu sur l'échelle de droite avec les diverses valeurs notées n sur l'échelle de gauche et lire sur l'échelle centrale le multiplicateur k/k_{min}.

On obtient par exemple :

n	k/k_{kim}	d'où k	et n.k.
10	1,05	40 villes	400 détaillants
5	1,08	41 -	205 -
2	1,15	44 -	88 -
1	1,40	53 -	53 -

Ce qui montre que si, au lieu de voir tous les détaillants des 38 villes (minimum de villes) on ne visite que 10 détaillants par ville, il faudra visiter 40 villes, soit un total de 400 détaillants.

Si l'on visite 2 détaillants par ville, il faudra aller dans 44 villes et cela fera 88 détaillants à visiter.

En définitive, on optera pour la combinaison la plus économique, compte tenu des distances entre villes, du temps passé, etc.

Autre exemple - Réception de balles de coton ou de laine.

Un filateur reçoit périodiquement des livraisons de balles de coton par exemple. Il sait d'après ses contrôles antérieurs, ou d'après des normes établies à l'échelon professionnel ou national, que pour tel pays d'origine les coefficients de variation moyens entre les balles (unités primaires), CV_1, et entre prises d'une même balle (unités secondaires), CV_2, ont certaines valeurs en ce qui concerne un certain caractère auquel il s'intéresse (finesse, résistance, teneur en gras,...)[1].

Il désire savoir combien il doit sonder de balles et à raison de combien de prises par balle, pour déterminer à 1 % près et avec un niveau de certitude de 95 chances sur 100, la valeur moyenne du caractère auquel il s'intéresse.

Données pour l'exemple : L = 190 balles
CV_1 = 3% (entre les balles)
CV_2 = 3% (à l'intérieur des balles)

a) La première planche donne à partir des deux premiers renseignements ci-dessus :

$$k_{min} = 30 \text{ balles}$$

En analysant intégralement le contenu des balles tirées au hasard, il suffirait de contrôler 30 balles.

b) Le rapport $CV_2/CV_1 = 3\% : 3\% = 1$.

c) La deuxième planche fournit pour le rapport ci-dessus, associé tour à tour avec n = 1, 2, 4, 10 les diverses valeurs du multiplicateur k/k_{min} ci-après, à partir desquelles on déduit le nombre réel de balles à sonder, ainsi que le nombre total de prises à prélever :

Prises par balle	Multiplicateur	Nombre de balles	Nombre total de prises
n	k/k_{min}	k	n.k.
1	2	60	60
2	1,50	45	90
4	1,25	37	148
10	1,10	33	330

Ces diverses combinaisons conduisent au même résultat, pour autant que les hypothèses énoncées au départ soient exactes.

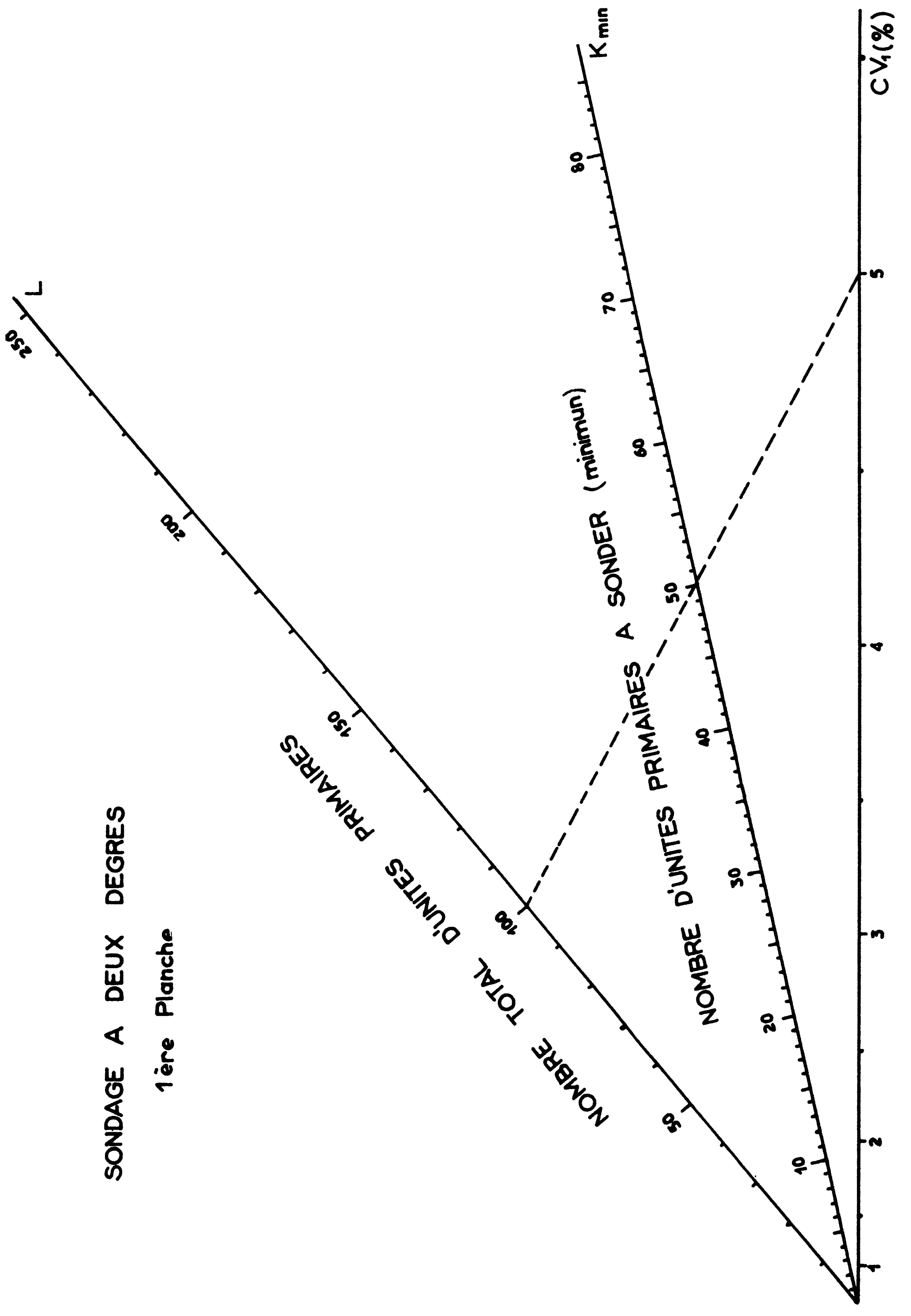
Le service de réception pourra opter par exemple entre l'ouverture de 60 balles avec 1 prise par balle, ou l'ouverture de 33 balles avec 10 prises par balle.

----

(1) Pour les balles de laine, voir par exemple les normes américaines A.S.T.M.

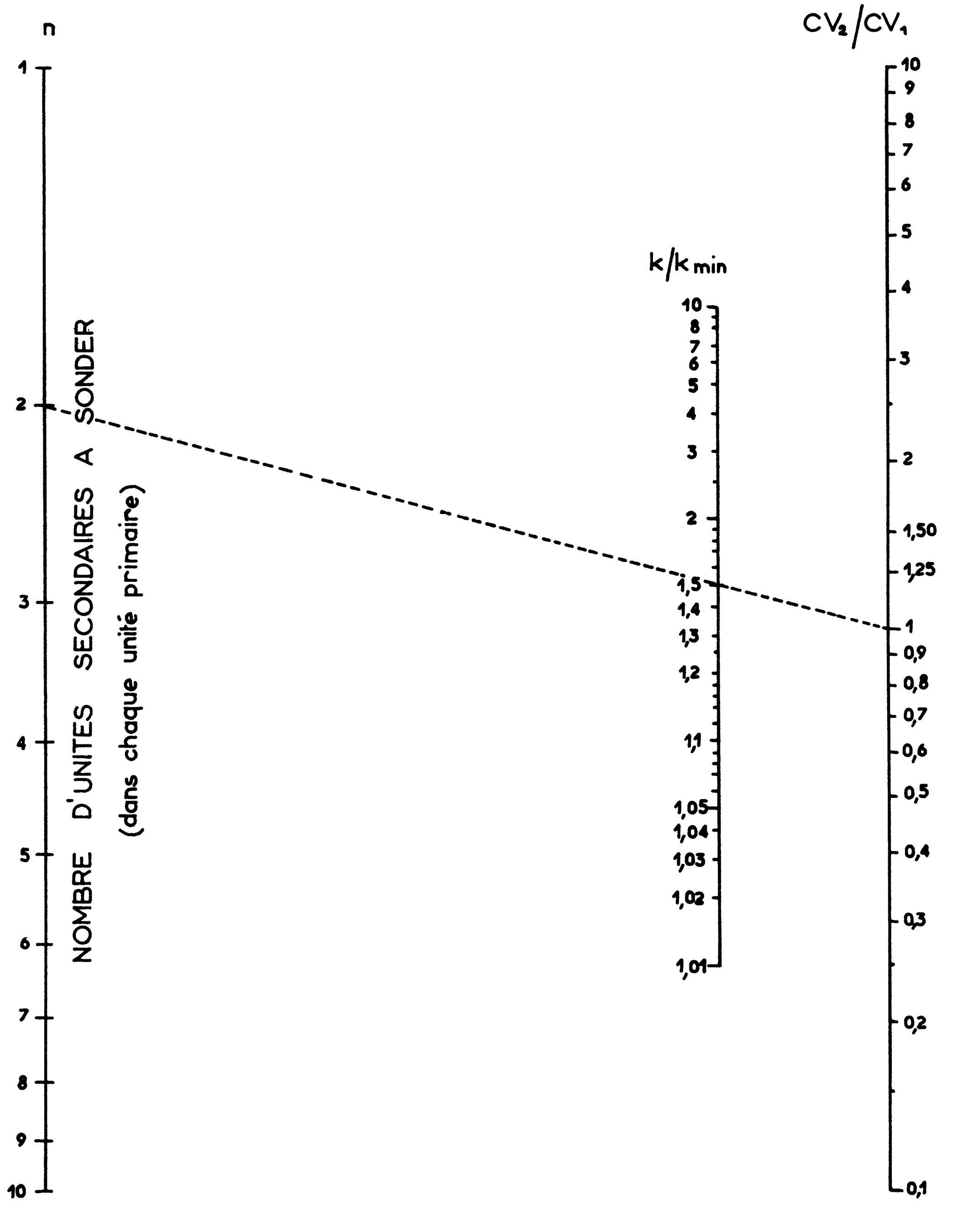
SONDAGE A DEUX DEGRES

1ère Planche



SONDAGE A DEUX DEGRES

2ème Planche



Il est clair que sur le plan de l'économie il y a matière à réflexion : difficultés de déplacer des balles peu accessibles si l'on sonde beaucoup de balles; temps à passer pour l'analyse d'un grand nombre total de prises; éventuellement coût de la matière détruite; etc.

N.B. - Rappelons que les balles à sonder ainsi que les prises à l'intérieur des balles doivent être choisies "au hasard" ce qui ne veut dire ni au "petit bonheur" ni "de la façon qui nous dérange le moins".

III - ABAQUE POUR L'EXAMEN DES RESULTATS OBTENUS DANS UN "SONDAGE D'OPINION" -

1/ Enoncé du problème.

On interroge par l'intermédiaire d'enquêteurs ou par correspondance un nombre total de N personnes, sur leurs préférences à propos de divers produits (pâtes dentifrices de diverses marques, candidats à des élections, horaires divers proposés pour un cours post-scolaire, ...). Le dépouillement donne un tableau du type suivant :

	n_1	personnes préfèrent le genre A
	n_2	personnes préfèrent le genre B
	n_3	personnes préfèrent le genre C
	n_4	personnes voudraient un genre non proposé
	n_5	personnes n'ont pas d'opinion ou n'ont pas répondu.
Total :	N	personnes ont été interrogées.

On se propose par exemple, en supposant que n_1 est plus grand que n_2, de décider si la différence $(n_1 - n_2)$ est significative d'une préférence réelle pour le genre A, ou si cette différence est purement fortuite, c'est-à-dire non probante.

2/ Mode opératoire et exemple.

L'abaque permet de répondre à la question posée avec un niveau de certitude de 95 chances sur 100. Il fait intervenir le nombre total N de personnes interrogées et le total des réponses utiles pour la comparaison effectuée, c'est-à-dire $(n_1 + n_2)$, s'il s'agit de comparer les préférences pour A et pour B.

En alignant la valeur de N - (lue sur l'échelle de droite) - avec la valeur de $(n_1 + n_2)$ - (lue sur l'échelle de gauche) -, on obtient sur l'échelle centrale la valeur minima que doit atteindre la différence $(n_1 - n_2)$ pour que la préférence pour A soit significative.

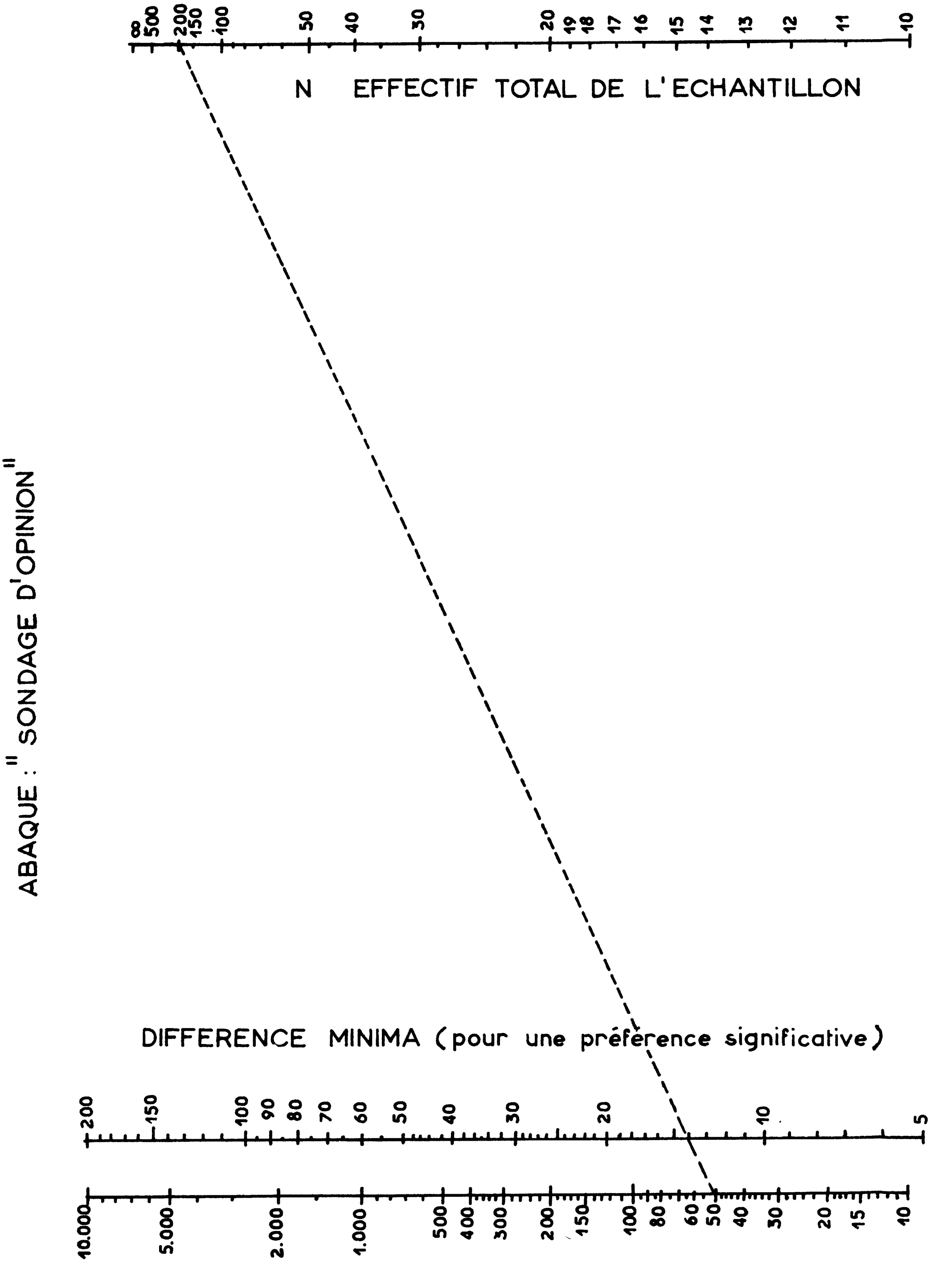
Exemple - On a interrogé par correspondance 200 femmes de standing aisé sur leurs préférences pour la couleur des draps. On a obtenu les résultats :

n_1 = 30 préfèrent les draps roses
n_2 = 20 préfèrent les draps bleus
n_3 = 100 préfèrent les draps blancs
n_4 = 50 n'ont pas d'opinion (ou : pas répondu)

1er test : Entre les draps roses et bleus, la préférence pour les draps roses est-elle significative ?

$N = 200$, $n_1 + n_2 = 50$, l'abaque donne diff. min. : 14

Comme ici, $n_1 - n_2 = 10$ (inférieur à 14), la préférence pour les draps roses plutôt que bleus n'est pas significative.



ABAQUE : " SONDAGE D'OPINION "
N EFFECTIF TOTAL DE L'ECHANTILLON
∞ 500 200 150 100 50 40 30 20 19 18 17 16 15 14 13 12 11 10
DIFFERENCE MINIMA (pour une préférence significative)
200 150 100 90 80 70 60 50 40 30 20 10 5
10.000 5.000 2.000 1.000 500 400 300 200 150 100 80 60 50 40 30 20 15 10
NOMBRE DE REPONSES "UTILES" (n_1+n_2)

2ème test : Ayant réuni les draps roses et bleus on obtient le nouveau score :

n_1 = 100 préfèrent les draps blancs
n_2 = 50 préfèrent les draps de couleurs
n_3 = 50 n'ont pas d'opinion (ou : pas répondu)

Cette fois : N = 200, $n_1 + n_2 = 150$: l'abaque donne diff. min. : 24.

Comme ici, $n_1 - n_2 = 50$ (supérieur à 24), la préférence pour les draps blancs est significative, même contre les deux couleurs réunies.

3/ Remarque.

Si le nombre total de personnes interrogées reste constant, par exemple N = 200, la différence minima significative - donnée par l'abaque - pour que la préférence pour A soit probante, varie avec le nombre de réponses utiles $n_1 + n_2$ (réponses pour A et pour B réunies). On peut voir aisément en faisant pivoter une règle autour du point N = 200 que si

$n_1 + n_2$ =	10	20	30	50	100	150	200
il faut diff. min. =	7	9	11	14	20	24	28

Il est clair que si peu de personnes répondent "utilement" (entre A et B), on obtiendra plus difficilement une conclusion significative, d'où la nécessité d'interroger un échantillon suffisamment important de personnes.

4/ Exemple à traiter.

3 505 interviews au sujet de marques de dentifrices employés ont donné les résultats :

emploient la marque	A	803 personnes
	B	612
	C	496
	D	269
	E	116
	F	57
	G	33
	H	26
	I	13
	J	10
autres marques		147
n'emploient pas de dentifrice		923

Examiner s'il existe dans la population une véritable échelle de classification des préférences pour les diverses marques ?