

REVUE DE STATISTIQUE APPLIQUÉE

H. PIN

Résolution par abaques des tables de contingence

Revue de statistique appliquée, tome 5, n° 4 (1957), p. 87-100

http://www.numdam.org/item?id=RSA_1957__5_4_87_0

© Société française de statistique, 1957, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

RÉSOLUTION PAR ABAQUES DES TABLES DE CONTINGENCE

par

H. PIN

Dans certaines études, notamment pour l'exploitation d'enquêtes dans les entreprises commerciales, on est amené à dresser de nombreuses tables de contingence pour confronter les situations. On se demande alors dans quelle mesure les différences constatées sont imputables au hasard de l'échantillonnage quand ces enquêtes sont considérées comme effectuées sur un échantillon prélevé au hasard.

Les calculs classiques des tables de contingence permettent de répondre à cette question. Mais ces calculs sont en général assez longs. Or, dans les entreprises, on fait souvent au Service Statistique le reproche d'utiliser des méthodes longues donc coûteuses. La construction d'abaques permettant un calcul rapide est donc souhaitable. On trouvera ci-après l'exposé d'une méthode de construction d'abaques facilitant l'analyse et l'interprétation des tableaux de contingence.

I - TABLEAUX 2×2

Fréquemment, dans les entreprises commerciales les tables de contingence sont représentées en graphiques, plus éloquents. Par exemple la table de contingence 2×2 , figure 1, est représentée graphiquement figure 2, où l'on voit d'un coup d'œil que la catégorie B est proportionnellement plus nombreuse dans la colonne S que dans la colonne T.

	S	T	
A	34	72	R'_1 106
B	108	128	R'_2 236
	C'_1 142	C'_2 200	N 342

Fig. 1

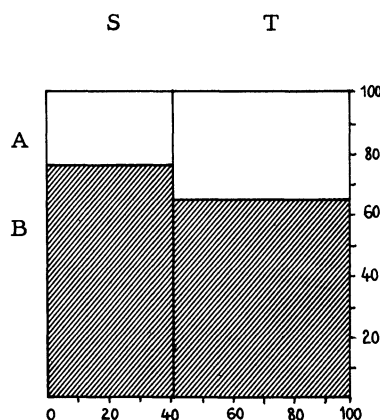


Fig. 2

La table, figure 1, donne le nombre d'individus de chaque case. Dans les cases marginales, à cadre gras, figurent les totaux par rangs et par colonnes. Le graphique figure 2 est établi en pourcentages à l'aide de la table, figure 3,

qui reproduit en pourcentages la table 1. Les pourcentages y sont calculés, pour les cases marginales, par rapport au total N soit ici 342, et, pour les cases de chaque colonne S et T , par rapport au total de la colonne soit ici 142 et 200. Ces calculs effectués à la règle sont rapides.

Le graphique 2, carré, comprend deux échelles, horizontale et verticale, en pourcentages. On a coupé le carré en deux colonnes verticales selon les pourcentages C_1 et C_2 du tableau 3. Puis dans chaque colonne on a reporté respectivement les pourcentages y et z des catégories B .

	S	T	
A	ω 24	x 36	R_1 31
B	y 76	z 64	R_2 69
	C_1 41,5	C_2 58,5	$N=342$ 100

Fig. 3

Il est facile de voir que dans la colonne S le pourcentage d'individus B est supérieur de 12 % à celui de la colonne T . Soit $D = 76 - 64 = 12$.

C'est à ce moment qu'en pratique dans les entreprises commerciales, pour économiser le temps, on a tendance à se contenter de tirer une conclusion sur le seul examen du graphique et de la différence $D = y - z$ (ou $D = x - \omega$) sans s'assurer par calcul de x^2 si cette différence D est ou non significative. Or à partir du tableau 3 il est rapide, par abaque, de calculer x^2 .

En effet le calcul classique de x^2 est donné en utilisant les fréquences absolues de la figure 1 par la relation

$$x^2 = \frac{(\omega' z' - x' y')^2 N}{C_1' C_2' R_1' R_2'} \quad (1)$$

Remplaçons les fréquences absolues par des fréquences relatives. En divisant les fréquences de chaque case par le total de la colonne on obtient :

$$\omega = \frac{\omega'}{C_1'} \quad y = \frac{y'}{C_1'} \quad x = \frac{x'}{C_2'} \quad z = \frac{z'}{C_2'}$$

Pour les cases marginales la fréquence relative est calculée par rapport à N soit :

$$C_1 = \frac{C_1'}{N} \quad C_2 = \frac{C_2'}{N} \quad R_1 = \frac{R_1'}{N} \quad R_2 = \frac{R_2'}{N}$$

On peut donc écrire :

$$C_1' = NC_1 \quad C_2' = NC_2 \quad R_1' = NR_1 \quad R_2' = NR_2 \quad (2)$$

On peut aussi écrire :

$$\begin{aligned} \omega' &= \omega C_1' & \text{ou encore} & \quad \omega' = \omega NC_1 & \text{et de même :} \\ \omega' &= \omega NC_1 & y' &= y NC_1 & x' = x NC_2 & z' = z NC_2 \end{aligned} \quad (3)$$

ABaque POUR LES TABLES DE CONTINGENCE 2 x 2

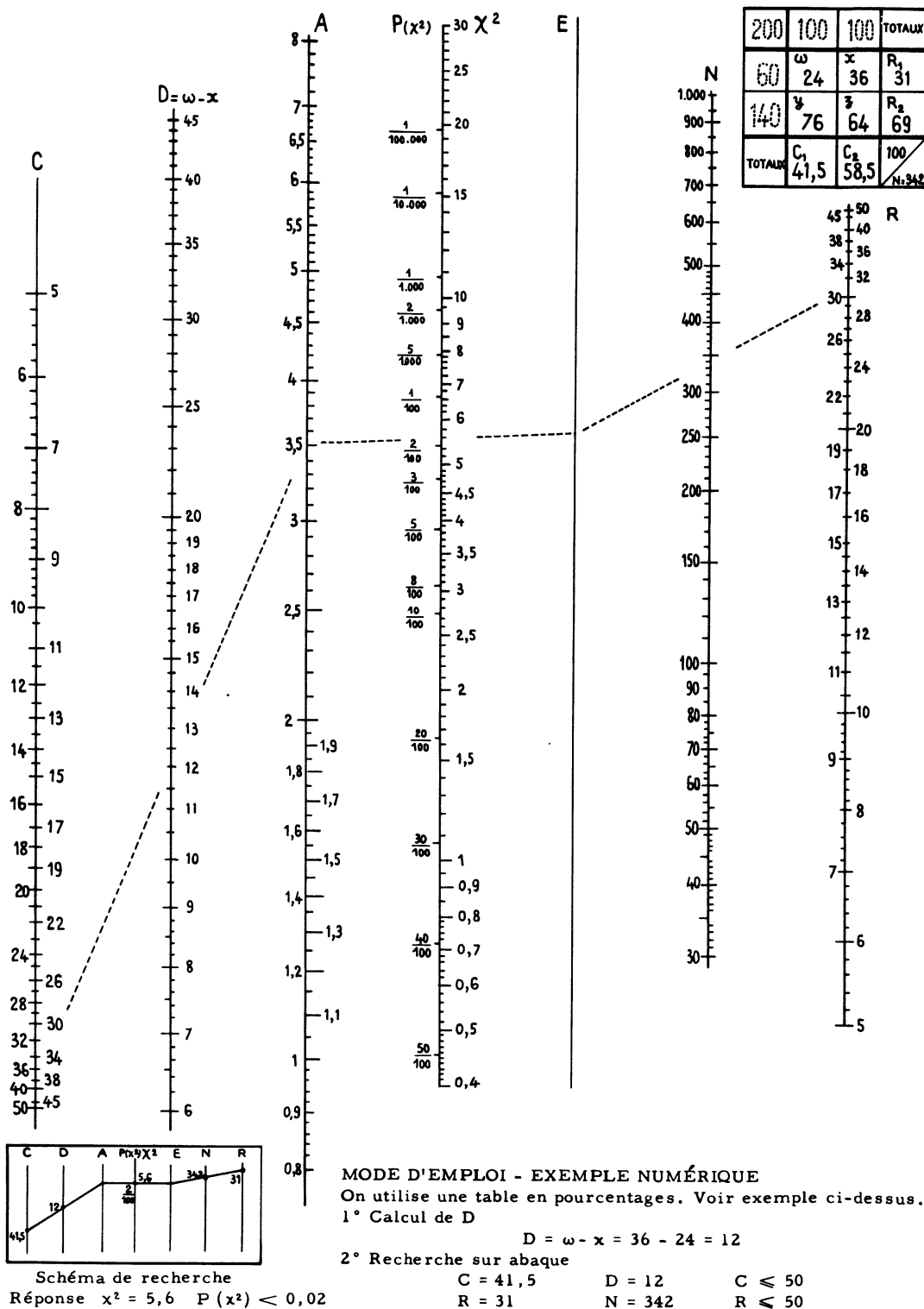


Fig. 4

En remplaçant alors les fréquences absolues de la relation (1) par les valeurs des relations (2) et (3) on obtient

$$x^2 = \frac{(\omega NC_1 z NC_2 - x NC_2 y NC_1)^2 N}{N^4 C_1 C_2 R_1 R_2}$$

Le terme carré du 2ème membre peut s'écrire :

$$\left[(N^2 C_1 C_2) (\omega z - xy) \right]^2 = N^4 (C_1 C_2)^2 (\omega z - xy)^2$$

on obtient donc

$$x^2 = \frac{N^4 (C_1 C_2)^2 (\omega z - xy)^2 N}{N^4 C_1 C_2 R_1 R_2} = \frac{(\omega z - xy)^2 C_1 C_2 N}{R_1 R_2}$$

Or, ω , x , y , z sont des proportions et l'on a $y = 1 - \omega$ et $z = 1 - x$, le terme carré du 2ème membre devient donc

$$\left[\omega(1 - x) - x(1 - \omega) \right]^2 = (\omega - \omega x - x + \omega x)^2 = (\omega - x)^2$$

en posant $\omega - x = D$ on a

$$x^2 = \frac{D^2 C_1 C_2 N}{R_1 R_2} \quad (6)$$

Il est alors facile de dresser un abaque, figure 4. En effet C_1 , C_2 , R_1 , R_2 , sont des proportions si bien que $C_2 = 1 - C_1$ et $R_2 = 1 - R_1$. On peut donc, par anamorphose, dresser une échelle donnant à l'abaque les valeurs de C_1 , C_2 , mais représentant les graduations de C_1 . Le même raisonnement est applicable à R_1 , R_2 .

Par commodité l'abaque est gradué en pourcentages et non en proportions. Le mode d'emploi est donné sur l'abaque figure 4. Dans la table 2 x 2 donnée comme modèle en haut à droite les cases extérieures à gauche, en haut, servent au contrôle. L'abaque donne directement sur l'échelle centrale la probabilité $P(x^2)$ qui n'a donc pas besoin d'être recherchée sur une table de x^2 .

Bien entendu, si le résultat trouvé se situe près du seuil de prudence choisi, il faut toujours craindre l'imprécision de l'abaque. Le calcul classique est nécessaire. Mais dans la plupart des cas, nettement significatifs ou aléatoires, l'abaque rend de grands services.

En choisissant dans la relation (6) une valeur x_0^2 correspondant à un seuil de prudence choisi soit à une valeur $P(x_0^2)$ choisie on a :

$$D^2 = \frac{x_0^2}{N} \frac{R_1 R_2}{C_1 C_2} \quad (7)$$

et

$$D = \sqrt{\frac{R_1 R_2}{C_1 C_2}} \sqrt{\frac{x_0^2}{N}}$$

Comme $R_2 = 1 - R_1$ et $C_2 = 1 - C_1$, il n'y a que 3 paramètres R , C , N ce qui permet de dresser l'abaque, figure 5.

En effet, en pratique, l'emploi de l'abaque 4, si rapide soit-il, est encore trop lent, l'abaque 5, bien qu'incomplet, s'avère le plus souvent très satisfaisant. (1)

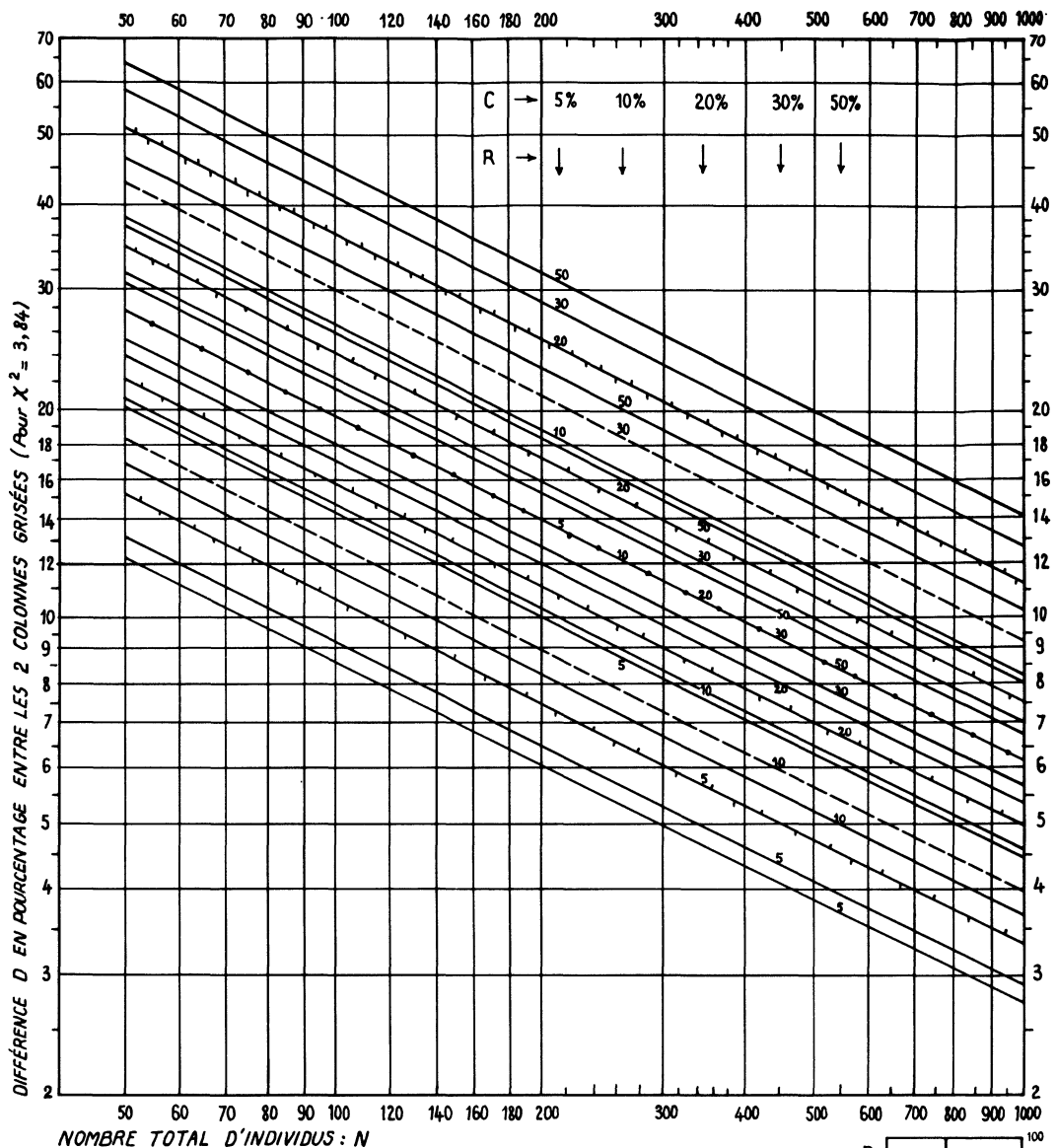
(1) L'abaque multiple de la figure 5 est établi pour les seules valeurs :

C (ou R) = 5 %, 10 %, 20 %, 30 %, 50 %

Il exige donc une double interpolation.

Si l'on veut davantage de précision, on pourra le remplacer par un jeu d'abaques, par exemple celui des figures 5 bis et 5 ter qui correspondent aux combinaisons de valeurs :

C (ou R) = 5, 7, 10, 15, 20, 30, 50 %.



Abaque donnant pour les tables de contingence 2×2 la différence D en pourcentage entre les 2 colonnes grisées, au seuil de prudence de 5 chances sur 100 ($\chi^2 = 3,84$)

R = pourcentage marginal du rang $R \leq 50 \%$
C = pourcentage marginal de la colonne $C \leq 50 \%$
N = nombre total d'individus

Les valeurs de R sont indiquées sur les obliques en colonnes sous les valeurs de C
Exemple: C = 5% R = 50% N = 200 D = 32% oblique 1 en partant du haut
 C = 10% R = 50% N = 200 D = 23% " 4 " " " "

Le résultat est significatif pour une valeur de D supérieure à celle de cet abaque

Fig. 5

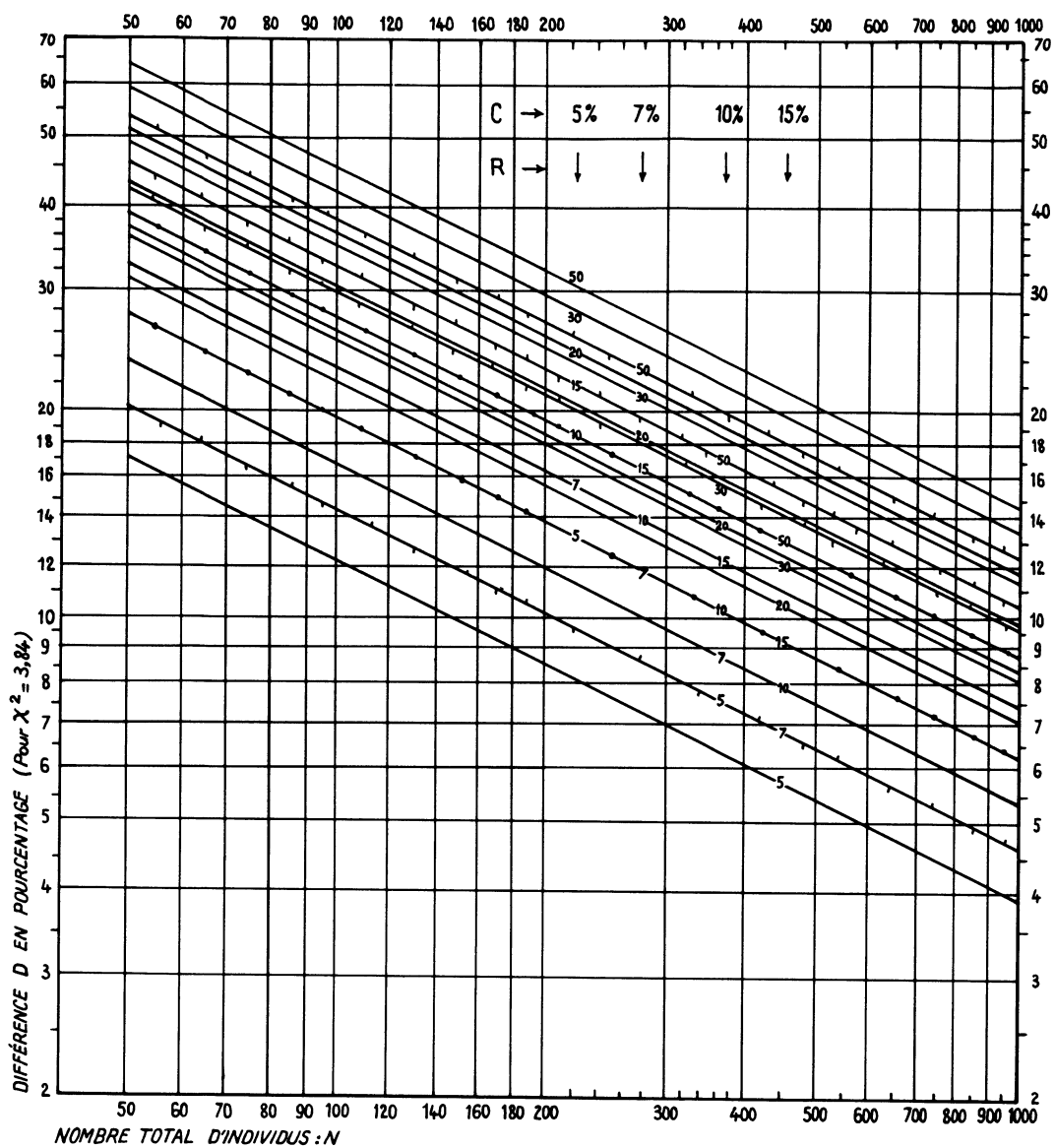


Fig. 5 bis

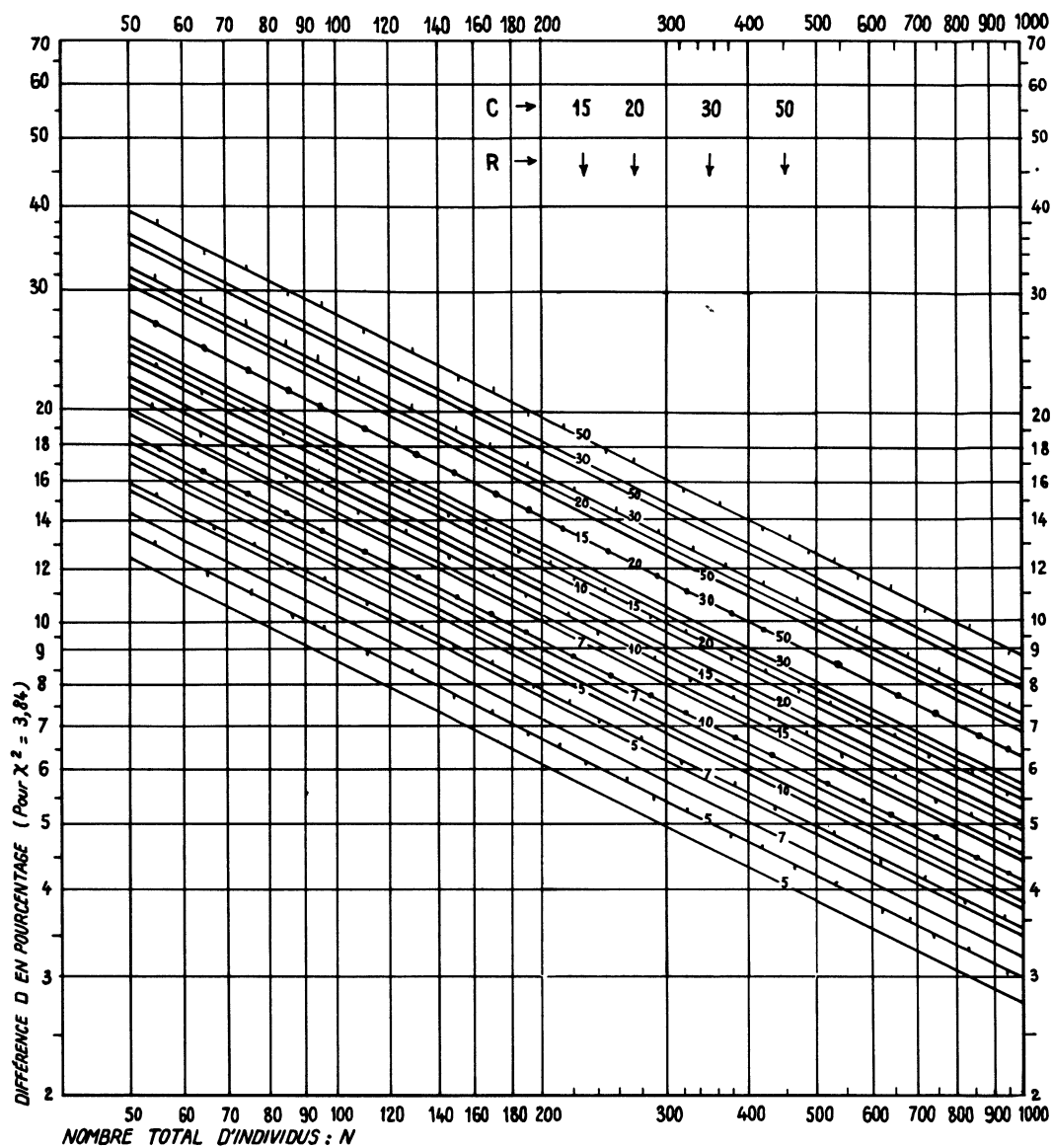


Fig. 5 ter

On voit en examinant la relation (7), ou aussi l'abaque 5, que la différence D d'une table 2×2 est d'autant plus significative que N est plus grand, que les deux colonnes ont des effectifs plus proches de l'égalité ($C = 50 \%$) et que R est plus faible donc qu'on observe dans les colonnes un événement plus rare.

Avant d'entreprendre une enquête il y a intérêt à consulter cet abaque 5 qui peut procurer une sensible économie. Si l'on peut il faut prévoir pour les 2 colonnes des effectifs égaux.

II - TABLEAUX $k \times 2$

Les tables $k \times 2$ comprenant un nombre quelconque k de colonnes et 2 rangées, se représentent aussi en graphique. La table 6, par exemple, est représentée au graphique 7 grâce à la table en pourcentages 8 établie de façon analogue à la table 3.

	S	T	V	
	a'_1	a'_2	a'_3	R'
A	10	24	26	60
	b'_1	b'_2	b'_3	P'
B	22	60	28	110
	c'_1	c'_2	c'_3	N
	32	84	54	170

Fig. 6

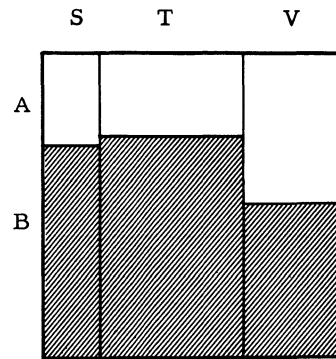


Fig. 7

	S	T	V	
	a_1	a_2	a_3	R
A	31,3	28,6	48,1	35,3
	b_1	b_2	b_3	P
B	68,7	71,4	51,9	64,7
	c_1	c_2	c_3	100
	18,8	49,4	31,8	$N=170$

Fig. 8

On peut encore à partir de la table 8 en pourcentages calculer χ^2 par abaque. Le calcul est plus rapide que par la méthode classique mais bien moins commode cependant que pour les tables 2×2 .

La formule classique qui utilise les fréquences absolues figure 6 est :

$$\chi^2 = \frac{1}{R'P'} \sum_i \frac{(a'_i P' - b'_i R')^2}{C'_i} \quad (8)$$

en utilisant comme précédemment les fréquences relatives on a :

$$C_i' = NC_i \quad R' = NR \quad P' = NP \quad a_i' = a_i NC_i \quad b_i' = b_i NC_i \quad (9)$$

En introduisant les valeurs (9) dans la relation (8) il vient :

$$\chi^2 = \frac{1}{N^2 RP} \sum_i \frac{(a_i NC_i \cdot NP - b_i NC_i \cdot NR)^2}{NC_i}$$

Le terme carré peut s'écrire

$$\left[(N^2 C_i) (a_i P - b_i R) \right]^2 = N^4 C_i^2 (a_i P - b_i R)^2$$

on a alors pour le terme somme

$$\sum \frac{N^4 C_i^2 (a_i P - b_i R)^2}{NC_i}$$

et donc

$$\chi^2 = \frac{N}{RP} \sum_i C_i (a_i P - b_i R)^2$$

En remarquant que $P = 1 - R$ et $b_i = 1 - a_i$ le terme carré devient :

$$a_i (1 - R) - (1 - a_i) R = a_i - R$$

ou enfin

$$\begin{aligned} \chi^2 &= \frac{N}{RP} \sum_i C_i (a_i - R)^2 \quad \text{et avec} \quad a_i - R = D_i \\ \chi^2 &= \frac{N}{RP} \sum_i C_i D_i^2 \end{aligned} \quad (10)$$

Le terme $\frac{N}{RP}$ qui existait déjà dans la relation (6) sous forme $\frac{N}{R_1 R_2}$ est donné à l'échelle E (non graduée) de l'abaque 4. Il est facile par ailleurs d'établir l'abaque 9 qui donne $CD^2 = L$. En faisant l'addition $\sum L = A$ on peut à l'abaque 4 avec A et E trouver aisément χ^2 . Le mode opératoire détaillé est donné à la figure 10. L'abaque 9 est assez délicat à manier à moins d'une bonne habitude.

La valeur $P(\chi^2)$ de l'abaque 4 valable seulement pour 1 degré de liberté n'est pas utilisable. Il faut rechercher $P(\chi^2)$ sur une table de χ^2 à $k - 1$ degré de liberté.

Si l'on veut aller vite on peut, en se fixant un seuil de prudence, donc une valeur de χ_0^2 minimum, calculer directement sur l'abaque 4 la valeur A_0 correspondante (avec R, N, χ_0^2 donnés). On cherche ensuite sur l'abaque 9 les différentes valeurs de L . On cumule les résultats, on abandonne les calculs dès que $\sum L > A_0$.

Ce calcul par abaquages peut être aussi utilisé pour contrôler le calcul classique.

III - TABLEAUX $k \times k'$

Les tables $k \times k'$ comprenant un nombre k quelconque de colonnes et k' quelconque de rangées se représentent aussi en graphique. La table 11 par exemple est représentée au graphique 12 grâce à la table en pourcentages 13 établie comme précédemment.

On peut encore à partir de la table 13 en pourcentage calculer χ^2 par abaque. Le calcul est plus facile que par la formule classique.

ABaque pour les tables de contingence $K \times 2$

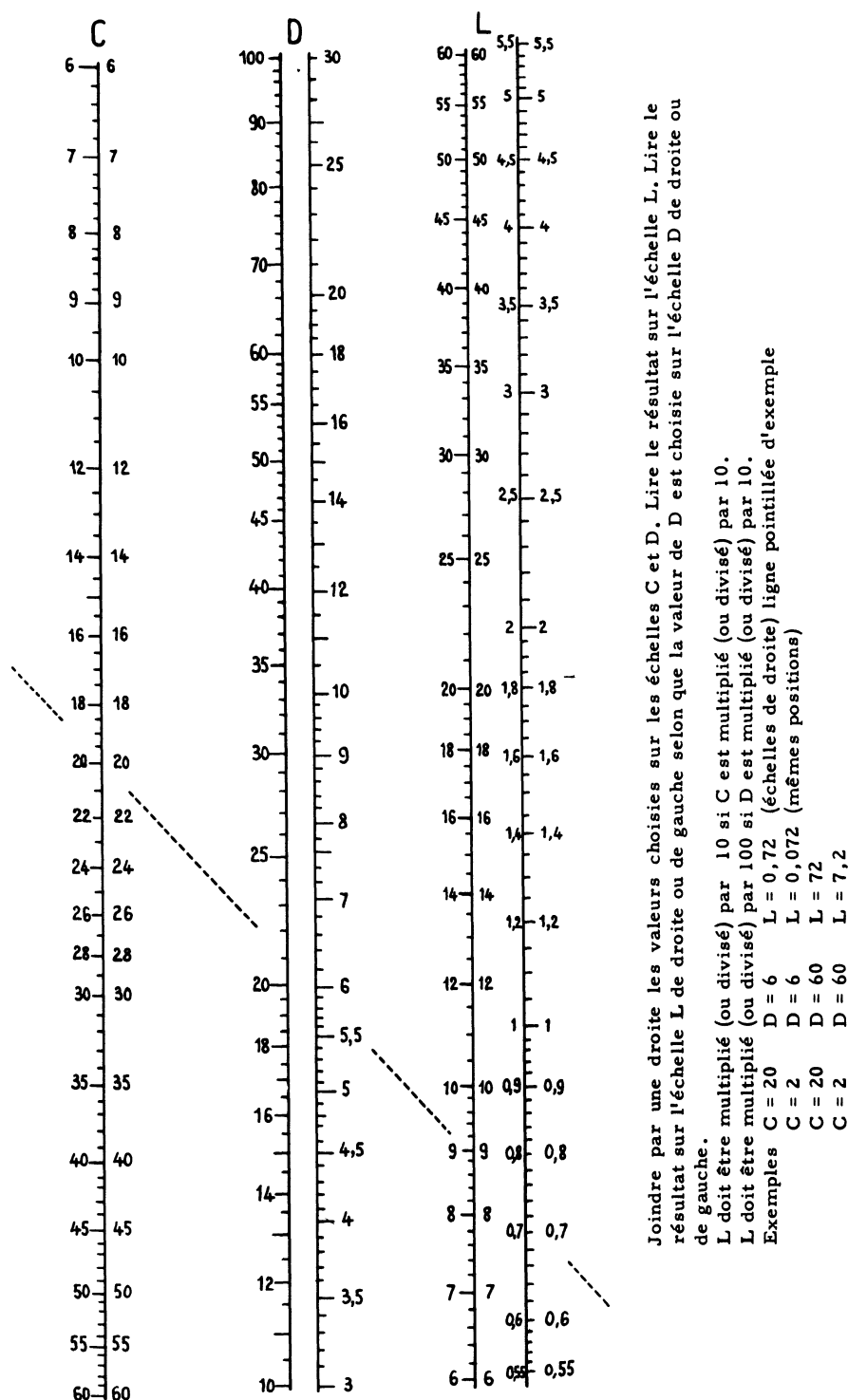


Fig. 9

K COLONNES ET 2 RANGÉES

100 K 300	100	100	100	Totaux
108	a_1 31,3	a_2 28,6	a_3 48,1	R 35,3
192	b_1 68,7	b_2 71,4	b_3 51,9	P 64,7
Totaux	c_1 18,8	c_2 49,4	c_3 31,8	100 N=170

K = Nombre de colonnes

Degré de liberté $V = K-1$

Chaque case doit contenir au moins 5 ou mieux 10 individus

On utilise le tableau en pourcentage

1° CALCUL DE D $D = a - R$

a →	31,3	28,6	48,1
-R →	-35,3	-35,3	-35,3
D →	-4,0	-6,7	+12,8

RECHERCHE DE L SUR ABAQUE 9

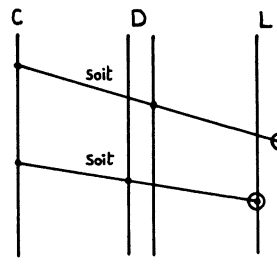
2° Tableau de L

C	D	L
18,8	-4	0,3
49,4	-6,7	2,2
31,8	+12,8	5,2
100	+12,8 -10,7	7,7 = A
TOTAL D A CONTROLER	+2,1	

3° Contrôle du total D
du tableau 2. $D = P-b$

$KP = 3 \times 64,7 = 194,1$
-Somme des b = -192,-
Total D + 2,1

4° Schéma de recherche
de L sur abaque 9



Sur l'abaque 9	
Si on multiplie (ou divise) par 10 l'échelle:	On doit multiplier (ou diviser) l'échelle L par:
C 10	
D 100	

RECHERCHE DE χ^2 SUR ABAQUE 4

N = 170 R = 35,3 A = 7,7

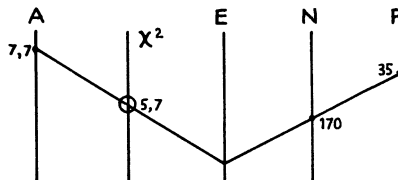
Réponse $\chi^2 = 5,7$

Aléatoire à 2 degrés de liberté

Sur l'abaque 4, R est celle des 2 valeurs R et P ne dépassant pas 50%.

Si l'on multiplie l'échelle A par 10 il faut multiplier par 10 l'échelle χ^2

Schéma de recherche sur abaque 4



FORMULE $\chi^2 = \frac{AN}{RP}$ $A = L_1 + L_2 + \dots + L_K$ $L = CD^2$ $D = a - R$ ou $D = P - b$
--

Figure 10

K COLONNES ET K' RANGÉES

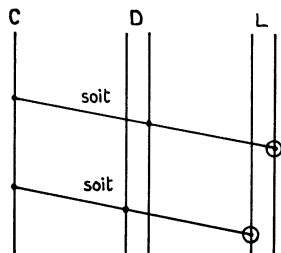
300	100	100	100	Total
H ₁	a ₁	a ₂	a ₃	R ₁
149,9	46,9	50	53	50
H ₂	a ₄	a ₅	a ₆	R ₂
74,6	18,7	26,5	29,4	25
H ₃	a ₇	a ₈	a ₉	R ₃
75,5	34,4	23,5	17,6	25
Totaux	c ₁	c ₂	c ₃	100 N=200
	32	34	34	

DETAIL DES CALCULS

D = a - R et contrôle des totaux

Recherche de L sur abaque 9 à partir de C et D

Schéma de recherche



Si on multiplie (ou divise) par 10 l'échelle C on doit multiplier (ou diviser) par 10 l'échelle L.

Si on multiplie (ou divise) par 10 l'échelle D on doit multiplier (ou diviser) par 100 l'échelle L

Formule $L = CD$;

$\frac{\sum L}{R} = B$ Avec $\sum L$ = somme des L d'une rangée.

$\chi^2 = \frac{N}{10} \sum B$ Avec $\sum B$ = Total des B

On utilise la table en pourcentages.

Chaque case doit comprendre au moins 5 ou mieux 10 individus sur la table en nombres d'individus.

Degré de liberté $V = (K-1)(K'-1)$. Exemple ici $V = 2 \times 2 = 4$.

Le χ^2_0 du seuil 5 chances sur 100 est environ $(1,5 V) + 3,5$. Exemple $V = 4$

$\chi^2_0 = (1,5 \times 4) + 3,5 = 9,5$.

a	C	D a-R=D	L RECHERCHE SUR ABAQUE	B
1ère Rangée	46,9	32 -3,1	0,31	
	50	34 0	0,00	
	53	34 3	0,31	
H ₁	149,9	100 -150	3 -3,1	= 0,012
KR ₁	-150	-0,1	-0,1	
			R ₁ = 50	
2e Rangée	18,7	32 -6,3	1,27	
	26,5	34 1,5	0,07	
	29,4	34 4,4	0,66	
H ₂	74,6	100 -75	5,9 -6,3	= 0,080
KR ₂	-75	-0,4	-0,4	
			R ₂ = 25	
3e Rangée	34,4	32 9,4	2,80	
	23,5	34 -1,5	0,07	
	17,6	34 -7,4	1,85	
H ₃	75,5	100 -75	9,4 -8,9	= 0,189
KR ₃	-75	+0,5	+0,5	
			R ₃ = 25	
N = 200 $\frac{N}{10} = 20$				$\sum B = 0,281$
$\chi^2 = 0,281 \times 20 = 5,62$				
Aléatoire à 4 degrés de liberté				

Figure 14

Les valeurs CD^2 peuvent se rechercher sur l'abaque 9. Le mode opératoire détaillé est donné à la figure 14 sur un exemple.

Sur le tableau de calcul figure 14 les totaux des 3 premières colonnes sont contrôlés. La colonne C (en pointillé) toujours la même pour toutes les rangées est souvent supprimée ou seulement dressée sur feuille mobile. La division par R se fait à la règle à calcul pour chaque rangée.

Ce tableau 14 quand il est utilisé pour vérifier le calcul classique sert non seulement à vérifier le total mais aussi, en cas d'erreur, à vérifier les calculs rangée par rangée. Il suffit de multiplier par $\frac{N}{10}$ la valeur de B de chaque rangée pour retrouver le cumul par rangée du calcul classique.

Quand on choisit $P(\chi^2_0) = 0,05$ on peut éviter de rechercher χ^2_0 sur la table en utilisant la valeur approchée $\chi^2_0 = (1,5 \nu) + 3,5$ où ν est le degré de liberté.

Comme en pratique le calcul est plus commode en pourcentages qu'en proportions la formule (13) devient avec l'abaque 9 :

$$\chi^2 = \frac{N}{10} \sum_{ij} C_i \frac{D_{ij}^2}{R_j}$$