

REVUE DE STATISTIQUE APPLIQUÉE

P. GENNARO

Problèmes de choix dans les recherches de marché nouvelles tables d'intervalles de confiance

Revue de statistique appliquée, tome 5, n° 3 (1957), p. 17-35

http://www.numdam.org/item?id=RSA_1957__5_3_17_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/>

PROBLÈMES DE CHOIX DANS LES RECHERCHES DE MARCHÉ NOUVELLES TABLES D'INTERVALLES DE CONFIANCE ⁽¹⁾

par

P. GENNARO

La comparaison de deux fréquences observées est un problème bien connu de tous les statisticiens.

Particulièrement simple, lorsqu'il s'agit de deux échantillons indépendants et des fréquences relatives à la présence d'un même caractère, le problème se complique un peu lorsqu'il s'agit d'un unique échantillon pouvant donner lieu à plusieurs choix qu'il s'agit de classer compte tenu d'un risque d'erreur admissible.

Les tables données par M. Gennaro permettent de voir immédiatement si l'effectif de l'échantillon permet ce classement.

BUT DE CETTE ÉTUDE

La première analyse des observations dans les enquêtes de marché consiste à étudier les tableaux dans lesquels sont reportées les données brutes importantes, avant de calculer les pourcentages. Déjà, dans cette phase, le statisticien doit décider quels sont les tableaux utiles et quels sont les chiffres significatifs, Ensuite, lorsqu'il s'agit d'une échelle de préférences, il est désirable de disposer d'un critère pratique et rapide pour décider si le classement obtenu des différents choix est significatif.

Nous indiquerons de suite comment ce problème peut être résolu et nous donnerons une table d'emploi pratique pour répondre à cette question (table I ci-après).

D'autre part, le travail de préparation auquel nous conduit ce problème nous permettra ainsi d'obtenir, pour la comparaison des pourcentages provenant du même échantillon, une nouvelle méthode qui se révèle préférable à la méthode traditionnelle que nous avons reprise et développée dans le précédent numéro de cette revue (2). Nous en déduisons une seconde table destinée à faciliter les décisions, dans le cas du choix, entre les pourcentages provenant du même échantillon.

(1) Traduction d'un article publié dans "Studi di mercato" octobre 1956, Revue de l'Association Italienne pour les Etudes de marché.

(2) Cf. "Studi di Mercato" - 2e année n° 1.

De cette analyse, il résulte ensuite que, lorsque les pourcentages proviennent d'échantillons séparés (par exemple dans la comparaison des caractéristiques de zones géographiques différentes, ou bien dans la comparaison des "taux de retour" d'analyse par poste et de propagande directe dans différents groupes d'adresses) on peut avoir un certain avantage de puissance (échantillons mineurs) en utilisant ou la table III (du n° I de l'année 2 de cette revue) ou bien la table II ci-après. Des exemples illustreront les critères d'application.

Enfin, nous examinerons le problème de la comparaison lorsque les choix A et B ne s'excluent pas ; c'est-à-dire lorsque quelques personnes ont choisi à la fois la première et la deuxième alternative à comparer, la somme des pourcentages étant alors supérieure à 100 (cas assez fréquent dans l'exemple du choix entre deux avis publicitaires, entre deux couleurs, entre deux types de confecti-
on) et nous montrerons comment il faut le résoudre.

Nous devons rappeler, bien que cela puisse être superflu, que ces résultats ne s'appliquent qu'à l'étude des résultats fournis par des échantillons probabilistes simples. Pour d'autres types d'échantillons probabilistes, on devra naturellement recourir aux formules adéquates données dans les ouvrages. Pour les échantillons non probabilistes, aucune considération basée sur le calcul des probabilités ne peut être utilisée. Les auteurs spécialisés dans l'étude de ces questions sont d'accord pour estimer que lorsqu'il n'est pas prouvé que l'échantillon non probabiliste (par quotas, par exemple) approche mieux les caractéristiques recherchées, son emploi n'est pas justifié. Les échantillons probabilistes simples sont employés dans beaucoup de cas : dans la propagande par poste, dans la comparaison des avis publicitaires, dans les tests de quantité d'un produit.

COMPARAISON SUR DES DONNÉES BRUTES : PROCÉDÉ PRATIQUE

Considérons un programme concret.

On interroge 144 personnes par exemple sur leur préférence au sujet de la couleur des machines à coudre et on obtient les résultats suivants :

Couleur	préférée par
A	60
B	39
pas de préférence	15
pas d'opinion	<u>30</u>
Total	144

Le problème, dans ce cas, consiste à décider si la différence constatée entre les 60 qui préfèrent A et les 39 qui préfèrent B indique la véritable préférence du public ou, si ce résultat peut s'être manifesté par l'effet de l'inclusion aléatoire dans notre échantillon de proportions non représentatives de certains éléments de la population.

La table I nous permet de résoudre immédiatement le problème : si n_1 est 60, la valeur maximum de n_2 admissible est 40, pour que la différence soit considérée comme significative au seuil 5%.

Puisque dans cet exemple, $n_2 = 39$, la différence est significative ; c'est-à-dire peut être considérée comme étant l'indice d'une véritable différence existant dans la population.

Voici un autre exemple :

En classant les résultats obtenus dans 3505 interviews au sujet de marques de dentifrices employés, on avait obtenu la répartition suivante :

Emploient la marque :

A	803
B	612
C	496
D	269
E	116
F	57
G	33
H	26
I	13
J	10
autres marques	147
Total	<u>2 582</u>

N'utilisent pas de dentifrice	<u>923</u>
	3 505

La question qui se pose immédiatement au statisticien est la suivante: quels sont parmi ces groupes ceux qui diffèrent de façon significative, c'est-à-dire qui indiquent une véritable échelle de classification dans la population?

En prenant n_1 égal successivement à 803, 612, 496 ... et n_2 égal à 612, 496.... et en utilisant notre tableau nous avons immédiatement les réponses ci-après :

Si la fréquence absolue relative à A est 803, la valeur maximum de B qui indique une préférence inférieure est, dans la table, aux environs de 720. Puisque nous avons trouvé 612, nous pouvons affirmer que le dentifrice A est à la première place dans la population. D'une façon analogue, on voit rapidement que la différence entre B et C est significative, et ainsi de suite jusqu'à la différence entre F et G, tandis que la différence entre G et H ne l'est pas, de même que celle entre I et J (13 et 10); au contraire celle qui existe entre H et I (26 et 13) est significative.

Donc dans le tableau final des résultats il apparaît opportun de négliger la classification des dentifrices G, H, I, J en les groupant dans les divers (autres marques).

Après calcul des pourcentages, la classification des dentifrices employés dans les familles italiennes examinées donne alors le résultat suivant :

Utilisent la marque	sur 100
A	31,1 %
B	23,7 %
C	19,2 %
D	10,4 %
E	4,5 %
F	2,2 %
Autres marques	<u>8,9 %</u>
Total	100

(Noter cependant que dans les "autres marques", on a vérifié la prédominance des marques G et H sur I et J).

ÉTUDE THÉORIQUE DE LA COMPARAISON DES RÉSULTATS DONNÉS PAR UNE ÉCHELLE DE CHOIX FOURNIE PAR UN ÉCHANTILLON. Solution théorique

Sans qu'il résulte des échantillons précédents, le problème que nous voulons résoudre consiste à décider si, est significatif ou non, le fait que dans notre échantillon un certain nombre n_1 d'éléments possèdent ou ont choisi la caractéristique A et un autre nombre n_2 la caractéristique B, le nombre total des éléments de l'échantillon étant $N \geq n_1 + n_2$.

Soit N le nombre total des réponses A, B, C...
 n_1 le nombre des réponses A
 n_2 le nombre des réponses B

$$n_1 + n_2 \leq N$$

Si on pose

$$d = n_1 - n_2$$

on a

$$\sigma_d^2 = \sigma_{n_1}^2 + \sigma_{n_2}^2 - 2\rho_{12} \sigma_{n_1} \sigma_{n_2}$$

avec

$$\sigma_{n_1}^2 = N p_1 (1 - p_1)$$

$$\sigma_{n_2}^2 = N p_2 (1 - p_2)$$

Calcul de :

$$\rho_{12} \sigma_{n_1} \sigma_{n_2} = \text{Cor}(n_1, n_2) .$$

Il s'agit de tirages dans une urne à 3 catégories caractérisées par les probabilités

p_1	préférence pour A
p_2	préférence pour B
$p_0 = 1 - p_1 - p_2$	préférence pour les autres

Si l'on considère un tirage particulier, on peut lui associer les deux caractères X, Y tels que

X = 1	Y = 0	avec probabilité	p_1
X = 0	Y = 1		p_2
X = 0	Y = 0		$p_0 = 1 - p_1 - p_2$

d'où :

$$E[(X-p_1)(Y-p_2)] = (1-p_1)(0-p_2)p_1 + (0-p_1)(1-p_2)p_2 + (0-p_1)(0-p_2)(1-p_1-p_2)$$

$$= -p_1 p_2$$

Pour l'ensemble des N tirages on aura

$$\text{Cov}(n_1, n_2) = E[\Sigma (X-p_1)(Y-p_2)] = -N p_1 p_2$$

d'où

$$\sigma_d^2 = N p_1 (1-p_1) + N p_2 (1-p_2) + 2 N p_1 p_2$$

$$= N [p_1 + p_2 - (p_1 - p_2)^2]$$

Si on estime p_1, p_2 à l'aide des résultats obtenus :

$$f_1 = \frac{n_1}{N} \qquad f_2 = \frac{n_2}{N} \quad ,$$

il vient :

$$\begin{aligned} \sigma_d^2 &= N [f_1 + f_2 - (f_1 - f_2)^2] \\ &= n_1 + n_2 - \frac{1}{N} (n_1 - n_2)^2 \quad . \end{aligned}$$

L'hypothèse à tester étant l'égalité de p_1 et p_2 , on a

$$\sigma_d^2 = 2 N p_1 = 2 N p_2$$

p_1 et p_2 étant inconnus, on peut les estimer, si l'hypothèse est vérifiée, par :

$$p_1 = p_2 = \frac{1}{2} \frac{n_1 + n_2}{N}$$

on aura alors

$$\sigma_d^2 = n_1 + n_2$$

Au seuil de signification 95%, par exemple, la différence entre n_1 et n_2 sera considérée comme significative si :

$$\begin{aligned} (n_1 - n_2) &\geq 1.96 \sqrt{n_1 + n_2} \\ (n_1 - n_2)^2 &\geq 3.84 (n_1 + n_2) \end{aligned}$$

Plus généralement, pour un seuil quelconque α auquel correspond la variable normale t , la différence sera significative si

$$(n_1 - n_2)^2 \geq t_\alpha^2 (n_1 + n_2) \quad ,$$

cette relation étant représentée par la région hachurée du graphique ci-contre. Si on désigne par n_1 la plus grande des deux valeurs n_1, n_2 , l'inégalité précédente s'écrit

$$n_2 \leq n_1 + \frac{t^2 - t \sqrt{t^2 + 8 n_1}}{2}$$

La table II ci-après a été calculée pour $t = 1.96$ (probabilité un peu supérieure à 0.95). Si au lieu de raisonner sur les nombres absolus on raisonne sur les pourcentages f_1, f_2 et si on pose :

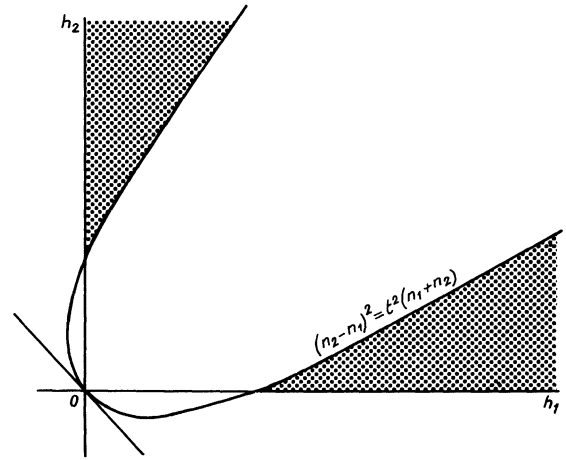
$$\Delta = |f_1 - f_2|$$

on aura, comme ci-dessus

$$\begin{aligned} \sigma_\Delta^2 &= \frac{p_1(1-p_1) + p_2(1-p_2) + 2 p_1 p_2}{N} \\ &= \frac{1}{N} [p_1 + p_2 - (p_1 - p_2)^2] \end{aligned}$$

dont l'estimation sera encore

$$\sigma_\Delta^2 = \frac{f_1 + f_2}{N}$$



Pour $n > 30$, l'approximation normale pouvant être utilisée, la différence $|f_1 - f_2|$ sera considérée comme significative au risque α si

$$|f_1 - f_2| \geq t_\alpha \sqrt{\frac{f_1 + f_2}{N}}$$

(par exemple $t_\alpha = 1.96$ pour $\alpha = 0,05$)

Cette relation exige

$$N \geq t_\alpha^2 \frac{f_1 + f_2}{(f_2 - f_1)^2}$$

Si l'on a à l'avance ou à partir de résultats partiels quelque information sur l'ordre de grandeur de f_1 et f_2 , cette condition peut servir à fixer l'effectif de l'échantillon de façon que le test de signification de la différence $f_1 - f_2$ puisse être satisfait.

(Voir à ce sujet les résultats donnés dans le tableau II).

Pour N inférieur à 30, l'approximation par la loi normale n'est pas valable et il convient d'utiliser la distribution de t de Student Fisher avec un nombre de degrés de liberté

$$v = N - 1$$

Si l'on remarque que la condition ,

$$|f_1 - f_2| \geq t \sqrt{\frac{f_1 + f_2}{N}}$$

peut s'écrire

$$\frac{N}{t^2} \geq \frac{f_1 + f_2}{(f_2 - f_1)^2} ,$$

il convient, par exemple au seuil 5 %, de construire la table suivante

v	$N = v + 1$	N/K^2
19	20	4,565
20	21	4,826
21	22	5,086
22	23	5,348
23	24	5,606
24	25	5,869
25	26	6,126
26	27	6,388
27	28	6,649
28	29	6,915
29	30	7,174

Il suffira de calculer alors les valeurs de la fraction du second membre et de lire dans le petit tableau les valeurs de N correspondantes.

C'est avec cette règle que l'on a complété les valeurs de la table II pour N inférieur à 30.

EMPLOI DE LA TABLE II POUR COMPARER LES POURCENTAGES
PAR L'ÉCHANTILLON MÊME

La table II a été construite de façon à en rendre l'usage très simple, un exemple en éclairera immédiatement l'emploi.

Une enquête conduite pour assurer l'acceptabilité d'un marché pour un colorant textile que l'on craignait d'être capable de provoquer des réactions défavorables avait donné les résultats suivants :

Ont manifesté les réactions ci-après :

	Ménagères	Industries textiles	Marchands de tissus
Favorables	76%	75%	59%
Défavorables	3%	20%	7%
Incertains Indifférents	21%	5%	34%
	100	100	100
	(303)	(20)	(58)

En examinant la première colonne (avis des ménagères) on voit immédiatement dans la table que bien que la valeur 76% n'y figure pas, 20 cas seraient suffisants pour permettre de considérer comme significatives les différences entre 76%, 21% et 3% qui résultent de 303 cas observés.

D'une façon analogue on vérifie d'un seul coup d'oeil que la différence entre 59% et 34% est à peine significative (l'échantillon comprend 58 cas, nombre supérieur de très peu à celui de l'échantillon minimum indiqué par la table (soit 57 cas).

CHOIX NON DISJOINTS (SOMME DES POURCENTAGES SUPÉRIEURE A 100)

La somme des pourcentages peut dépasser 100, par exemple lorsque A et B ont été choisis tous les deux par quelques-uns des éléments interrogés ou lorsqu'ils étaient contenus tous les deux dans quelques éléments examinés, soit f_{12} la fréquence de ces résultats réunis.

Il peut arriver aussi que la somme des pourcentages ne dépasse pas 100 mais que cependant il existe des éléments communs aux différents choix dans le tableau des résultats (les mêmes personnes ont donné plus d'une réponse).

Il est important de se demander si les comparaisons restent significatives et dans quelle mesure.

En général, notre décision devrait être basée uniquement sur les individus ayant réellement préféré A ou B, puisque ceux qui les ont choisis tous les deux sont sans importance pour notre décision. Par exemple, dans la comparaison de deux avis publicitaires, nous devons choisir l'avis le plus efficace en le comparant au jugement de ceux qui les considèrent d'efficacité différente.

Le problème que nous nous posons est donc de savoir si le nombre de ceux qui sont vraiment "décidés" en faveur de A, dans l'échantillon dépasse dans une mesure significative le nombre de ceux qui sont vraiment "décidés" en faveur de B (ce que nous appellerons par commodité les choix purs).

En d'autres termes, dans le cas qui nous intéresse, nous devons comparer $f_1 - f_{12}$ avec $f_2 - f_{12}$. Toutefois en pratique, cette correction est presque toujours supprimée, soit pour celui qui reçoit le rapport final (dans lequel ne sont généralement pas indiqués les cas de double choix) soit aussi pour le chercheur ; car

généralement, la méthode du dépouillement utilisée (à la main ou à la machine) ne permet pas de tenir compte séparément des cas du double choix.

Par bonheur, la distorsion introduite dans la formule, en utilisant les fréquences non corrigées, est en un sens favorable à l'estimation; en fait, si nous utilisions la correction, nous aurions :

$$N \geq 3,84 \frac{f_1 + f_2 - 2 f_{12}}{(f_1 - f_2)^2}$$

qui, évidemment, donne une limite de N inférieure à celle que l'on obtient en négligeant f_{12} (ce qui était prévisible en observant l'allure des valeurs de N dans la diagonale de la table II).

On peut donc conclure avec la règle suivante : même quand la somme des pourcentages dépasse 100, si, selon la table II, la différence est significative, le résultat est acceptable pour la comparaison des choix purs.

Si, au contraire, la dimension de l'échantillon ne donne pas un résultat suffisant pour la signification dans cette première approximation, il est nécessaire d'estimer f_{12} . La grandeur minimum de l'échantillon s'obtiendra en apportant au résultat précédent (tiré de la table) la correction

$$- 3,84 \frac{2 f_{12}}{(f_1 - f_2)^2}$$

COMPARAISON ENTRE LES POURCENTAGES DONNÉS PAR ÉCHANTILLONS ÉGAUX MAIS DISTINCTS

Dans la pratique, ce cas peut aussi se présenter.

Un exemple typique se trouve dans le choix de la méthode, type de lettre ou de présentation qu'il faut retenir comme étant le plus efficace pour une vente ou pour la publicité par correspondance.

Un autre cas est celui de la comparaison entre les avis publicitaires au moyen de la technique du "split-run" (Impression par exemple d'un certain avis sur une moitié des exemplaires distribués et d'un autre sur l'autre moitié).

Dans ces cas, on peut utiliser la table que nous avons déjà calculée déduite par la méthode traditionnelle de comparaison de la différence entre pourcentages (Cf. Studi di mercato, janvier 1956 et annexe ci-joint). Nous en donnons un exemple d'application.

Avant de lancer sur une large échelle une campagne de vente par la poste, une industrie turinoise pensa déterminer au moyen d'un essai quels secteurs de public donneraient les meilleures possibilités de réussite.

Cette société utilisa trois échantillons de 1.000 adresses chacun et deux de 2.000; elle envoya une offre à ces adresses et obtint les résultats suivants (pour 100 noms atteints) :

Catégorie d'adresses	Nombre d'adresses	% de réponses
I	1.000	5
II	1.000	8
III	1.000	6,5
IV	2.000	8
V	2.000	6

La table de la page 15 de l'article cité ci-dessus, nous permet de conclure immédiatement que la catégorie IV est significativement plus efficace que la catégorie V (en fait, pour que la différence entre 8 % et 6 % soit significative, il suffit que chaque échantillon comprenne 1248 éléments et ici, nous en avons 2.000), d'autre part, la catégorie II est significativement supérieure à la catégorie I puisque la comparaison résulte d'échantillons de 1.000 éléments, là où, selon la table, un effectif de 502 serait suffisant.

Au contraire, il n'est pas possible d'après la table, d'établir un classement complet entre les classes I, II, III, car les différences constatées ne seraient toutes significatives que pour des échantillons d'effectif très supérieur à 1.000.

Table I

Signification d'un classement de choix

n_1 effectif du choix A dans l'échantillon

n_2 effectif maximum du choix B pour que la différence soit significative au niveau 5%, A étant préféré à B)

n_3 effectif complémentaire pour observer un échantillon minimum de 30

$$n_2 = n_1 - \frac{1.96 \sqrt{8 n_1 + 3.84} - 3.84}{2}$$

n_1	n_2	n_3
10	3	17
11	4	15
12	4	14
13	5	12
14	5	11
15	6	9
16	7	7
17	7	6
18	8	4
19	9	2
20	9	1

n_1	n_2	n_2	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2	
21	10	41	25	72	50	125	96	225	185	325	277
22	11	42	26	74	52	130	100	230	190	330	282
23	12	43	27	76	54	135	105	235	194	335	286
24	12	44	27	78	55	140	109	240	199	340	291
25	13	45	28	80	57	145	113	245	204	345	295
26	14	46	29			150	118	250	208	350	300
27	14	47	30	82	59						
28	15	48	31	84	61	155	122	255	213	355	305
29	16	49	31	86	62	160	127	260	217	360	309
30	17	50	32	88	64	165	131	265	222	365	314
31	17	52	34			175	140	275	231	375	323
32	18	54	36	92	67	180	145	280	236	380	328
33	19	56	37	94	69	185	149	285	240	385	333
34	20	58	39	96	71	190	154	290	245	390	337
35	20	60	40	98	73	185	158	295	249	395	342
36	21			100	74	200	163	300	254	400	347
37	22	62	42								
38	23	64	44	105	79	205	167	305	259	405	351
39	24	66	45	110	83	210	172	310	263	410	356
40	24	68	47	115	87	215	176	315	268	415	360
		70	49	120	92	220	181	320	272	420	365

n_1	n_2	n_1	n_2	n_1	n_2
425	370	610	544	910	828
430	374	620	553	920	838
435	379	630	562	930	847
440	384	640	572	940	857
445	388	650	581	950	867
450	393	660	591	960	876
		670	600	970	886
455	398	680	610	980	895
460	403	690	619	990	905
465	407	700	629	1000	914
470	412				
475	417	710	638	1050	962
480	421	720	648	1100	1010
485	426	730	657	1150	1058
490	431	740	667	1200	1106
495	435	750	676	1250	1154
500	440	760	686	1300	1202
		770	695	1350	1250
510	449	780	705	1400	1299
520	459	790	714	1450	1346
530	468	800	724	1500	1395
540	478				
550	487	810	733	1550	1443
560	496	820	743	1600	1491
570	506	830	752	1650	1539
580	515	840	762	1700	1588
590	525	850	771	1750	1636
600	534	860	781	1800	1684
		870	790	1850	1733
		880	800	1900	1781
		890	809	1950	1830
		900	819	2000	1878

table I (suite)

EXEMPLES D'EMPLOI DE LA TABLE n° 1

Premier exemple : A a été choisi par 180 personnes interrogées ($n_1 = 180$) B par 150 ($n_2 = 150$). La table montre que la différence n'est pas significative, car si B était à la seconde place, on devrait avoir au maximum $n_2 = 145$.

Second exemple : A a été choisi par 790, B par 712. La table nous permet de considérer que la différence est significative, c'est-à-dire que l'on doit attribuer à B la seconde place puisqu'il a été choisi seulement par 712 tandis que le maximum de n_2 était 714.

Table II

Table de l'effectif minimum de l'échantillon pour obtenir un choix significatif au niveau 5% entre deux pourcentages, dans un même échantillon

Table II de 5 % à 22 %

	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%	21%	22%
1%	144	108	85	71	60	52	46	41	37	34	31	30	30	28	27	25	24	23
2%	299	192	138	107	86	72	62	54	48	43	39	35	32	30	30	29	27	26
3%	768	384	240	169	128	102	84	71	61	54	48	43	39	36	33	31	30	30
4%	3.457	960	469	288	200	149	118	96	81	69	60	53	48	43	39	36	33	31
5%	4.226	1.152	555	335	230	171	133	108	90	77	67	59	52	47	43	39	36	36
6%	4.994	1.345	640	384	261	192	149	120	100	84	73	64	57	51	46	42	42	42
7%	5.762	1.537	726	432	292	213	165	132	109	92	79	69	61	55	50	50	50	50
8%	6.531	1.729	811	480	322	235	180	144	119	100	86	75	66	59	59	66	66	66
9%	7.299	1.921	896	528	354	256	196	156	128	108	92	80	70	70	80	80	80	80
10%	8.067	2.113	982	576	384	277	212	168	137	115	98	85	85	98	98	115	98	85
11%	8.836	2.305	1.067	624	415	299	227	180	147	123	105	105	123	105	147	123	105	105
12%	9.604	2.497	1.152	672	446	320	243	192	156	131	115	105	123	105	147	123	105	105
13%	10.372	2.689	1.238	730	476	341	259	204	166	146	123	105	123	105	147	123	105	105
14%	11.141	2.881	1.323	768	507	363	274	216	175	146	123	105	123	105	147	123	105	105
15%	11.909	3.073	1.408	816	538	384	290	230	188	156	131	115	105	123	105	147	123	105
16%	12.677	3.265	1.494	864	569	405	259	204	166	146	123	105	123	105	147	123	105	105
17%	13.446	3.457	1.579	912	599	426	274	216	175	146	123	105	123	105	147	123	105	105
18%	14.214	3.650	1.665	960	626	446	290	230	188	156	131	115	105	123	105	147	123	105
19%	14.982	3.842	1.750	1.008	656	466	304	243	192	156	131	115	105	123	105	147	123	105
20%	15.751	4.034	1.835	1.056	686	486	318	259	204	166	146	123	105	123	105	147	123	105
21%	16.519	4.226	1.921	1.104	716	506	332	274	216	175	146	123	105	123	105	147	123	105
22%	17.287	4.418	2.007	1.152	746	526	346	289	230	188	156	131	115	105	123	105	147	123

Table II (suite) de 23 % à 39 %

	23%	24%	25%	26%	27%	28%	29%	30%	31%	32%	33%	34%	35%	36%	37%	38%	39%
1%	22	21	20	22	21	20	21	20	21	21	20	21	21	21	20	21	20
2%	25	24	23	24	23	22	23	22	23	22	21	23	22	23	22	23	22
3%	29	27	25	24	23	24	23	22	23	22	22	23	22	23	22	23	22
4%	30	28	27	25	24	24	23	22	23	22	22	23	22	23	22	23	22
5%	33	31	30	30	28	27	26	24	24	22	22	23	22	23	22	23	22
6%	39	36	33	31	30	30	29	27	26	25	24	23	22	21	20	21	20
7%	45	41	38	35	33	31	30	30	29	27	26	25	24	23	22	23	22
8%	53	48	44	40	37	35	32	30	30	29	28	27	26	25	23	24	23
9%	63	56	51	47	43	39	36	34	32	30	30	29	28	27	25	26	25
10%	75	67	60	54	49	45	41	38	36	33	31	30	30	29	28	29	27
11%	91	79	71	63	57	52	48	43	40	37	35	33	31	30	30	30	30
12%	111	96	84	74	67	60	55	50	46	42	39	36	34	32	30	31	30
13%	138	117	104	89	78	70	63	57	52	48	44	41	38	36	34	35	32
14%	175	146	124	107	93	82	74	66	60	55	50	46	43	40	37	39	36
15%	228	185	154	130	112	98	86	77	69	63	57	52	48	44	41	43	40
16%	306	240	194	161	136	117	102	90	80	72	65	60	54	50	46	48	44
17%	427	321	252	204	169	143	123	107	94	84	75	68	62	56	52	54	50
18%	630	448	337	264	213	177	149	128	111	98	87	78	70	64	58	60	56
19%	1.008	661	469	353	276	223	184	155	134	116	102	91	81	73	67	69	63
20%	1.835	1.056	692	491	368	288	232	192	162	139	120	106	94	84	76	78	71
21%	4.226	1.921	1.104	722	512	384	300	242	200	168	144	125	110	97	87	90	81
22%	17.287	4.418	2.006	1.152	753	536	400	312	251	207	174	150	129	114	101	104	93
23%	18.056	4.610	2.091	1.200	783	555	416	324	261	215	181	155	134	118	122	122	108
24%	18.824	4.802	2.177	1.249	815	576	431	336	270	223	187	160	138	138	143	143	125
25%	19.592	4.994	2.262	1.297	845	598	447	348	280	223	187	160	138	138	143	143	125
26%	20.360	5.186	2.347	1.345	876	619	463	360	289	230	193	166	144	144	148	148	148
27%	21.129	5.378	2.433	1.393	907	640	478	372	299	238	200	176	150	150	154	154	148
28%	21.987	5.570	2.518	1.441	938	662	494	384	308	246	206	176	150	150	154	154	148
29%	22.665	5.762	2.603	1.489	968	683	510	396	318	254	214	184	158	158	162	162	148
30%	23.434	5.954	2.689	1.537	999	704	525	408	327	262	222	192	162	162	166	166	148
31%	24.202	6.147	2.774	1.585	1.030	726	537	420	336	270	230	198	168	168	172	172	148
32%	24.970	6.339	2.860	1.633	1.061	747	547	432	345	278	238	206	176	176	180	180	148
33%	25.739	6.531	2.945	1.681	1.091	768	557	444	354	286	246	214	184	184	188	188	148
34%	26.507	6.723	3.030	1.729	1.122	789	567	456	363	294	254	222	192	192	196	196	148
35%	27.275	6.915	3.116	1.777	1.152	810	577	468	372	302	262	230	200	200	204	204	148
36%	28.044	7.107	3.201	1.825	1.182	831	587	480	381	310	270	238	208	208	212	212	148
37%	28.812	7.299	3.286	1.873	1.212	852	597	492	390	318	278	246	216	216	216	216	148
38%	29.580	7.491	3.371	1.921	1.242	873	607	504	400	326	286	254	224	224	224	224	148
39%	30.348	7.683	3.456	1.970	1.272	894	617	516	410	334	294	262	232	232	232	232	148

Table II (suite) de 40 % à 56 %

	40%	41%	42%	43%	44%	45%	46%	47%	48%	49%	50%	51%	52%	53%	54%	55%	56%
8%	21	20	21	20	21	20	21	20	20	21	21	20	21	20	20	21	20
9%	23	22	22	22	22	22	21	21	22	23	23	22	22	21	21	21	20
10%	24	23	24	23	24	23	22	21	22	23	22	21	22	21	20	20	20
11%	26	25	24	25	24	25	24	23	22	23	24	23	22	21	21	21	20
12%	28	27	26	27	26	27	25	24	23	24	25	24	23	22	22	21	21
13%	30	29	28	27	28	27	25	24	24	24	28	27	26	25	24	23	23
14%	31	30	30	29	28	27	25	24	24	24	30	29	28	27	26	25	24
15%	34	32	31	30	30	29	27	26	25	25	30	30	29	28	26	25	25
16%	38	35	33	31	30	30	29	28	27	26	25	24	23	22	22	21	20
17%	42	39	36	34	32	30	30	30	29	28	27	26	25	24	23	22	21
18%	46	43	40	37	35	33	31	30	30	30	28	27	26	25	24	23	23
19%	51	47	45	42	38	36	34	32	31	30	30	29	28	27	26	25	24
20%	58	53	49	46	43	40	38	36	33	32	30	30	30	29	28	26	25
21%	65	60	55	51	47	44	41	39	37	34	33	31	30	30	29	28	27
22%	74	67	61	57	52	49	46	42	40	38	35	34	32	30	30	29	29
23%	84	76	69	63	58	54	50	47	43	41	39	36	35	33	31	30	30
24%	90	86	79	71	65	60	55	52	48	45	42	40	37	35	34	32	30
25%	111	99	89	81	73	67	62	57	53	50	46	43	41	38	36	34	32
26%	129	115	102	92	83	75	69	64	58	55	51	47	44	42	39	37	35
27%	152	133	118	105	94	86	77	71	65	60	56	52	48	46	43	40	38
28%	182	157	137	121	108	97	88	80	73	67	62	58	54	49	47	44	41
29%	219	187	161	141	125	111	100	90	82	75	69	63	59	55	51	48	45
30%	269	225	192	166	145	128	114	102	93	84	77	70	65	61	56	52	49
31%	337	276	231	198	170	149	132	117	105	95	86	79	72	66	62	58	53
32%	432	346	284	238	203	175	153	135	120	108	98	88	81	74	68	64	59
33%	572	444	355	292	244	208	179	157	139	123	110	100	90	83	76	70	65
34%	790	588	456	365	300	250	214	184	161	142	126	113	102	92	84	77	71
35%	1.153	811	604	468	374	307	257	219	188	165	145	129	116	105	94	86	79
36%	1.825	1.183	832	619	480	384	315	263	224	193	169	149	132	118	107	96	88
37%	3.286	1.873	1.214	854	635	492	393	323	269	230	197	172	152	135	121	109	99
38%	7.492	3.372	1.921	1.245	875	651	504	403	330	276	235	202	176	156	138	124	112
39%	30.349	7.683	3.457	1.969	1.275	896	666	516	412	338	282	240	207	180	159	141	126
40%	31.117	7.875	3.542	2.017	1.306	918	682	528	422	346	288	241	211	184	162	144	144
	41%	31.885	8.067	3.628	2.065	1.337	939	698	540	431	353	251	216	188	166	144	144
	42%	32.654	8.259	3.713	2.113	1.368	960	713	552	441	361	256	220	192	166	144	144
	43%	33.422	8.452	3.799	2.161	1.399	982	729	564	450	369	262	225	192	166	144	144
	44%	34.190	8.644	3.884	2.209	1.429	982	729	564	450	369	262	225	192	166	144	144
	45%	34.958	8.836	3.969	2.257	1.460	1.024	760	588	469	376	262	225	192	166	144	144
	46%	35.727	9.028	4.055	2.305	1.491	1.046	776	604	485	392	278	236	202	172	152	136
	47%	36.495	9.220	4.140	2.353	1.522	1.072	792	620	500	408	290	248	214	184	164	148
	48%	37.264	9.412	4.225	2.399	1.553	1.100	808	636	516	424	302	260	220	190	170	154
	49%	38.032	9.604	4.310	2.445	1.584	1.128	824	652	532	440	314	272	224	194	174	158

Table II (suite et fin) de 57 % à 73 %

17%	21	20	59%	60%	61%	62%	63%	64%	65%	66%	67%	68%	69%	70%	71%	72%	73%
18%	22	21	21	20	20	21	20	21	20	20	21	20	21	20	20	21	20
19%	23	22	22	21	20	22	21	23	22	21	22	21	23	22	21	21	20
20%	25	24	23	22	21	24	23	24	23	24	23	22	24	23	21	21	20
21%	26	25	24	23	23	25	24	25	24	25	24	23	25	24	21	21	20
22%	28	27	26	25	24	26	25	26	25	26	25	24	26	25	22	21	20
23%	29	29	27	26	25	27	26	27	26	27	26	25	27	26	23	22	21
24%	30	30	29	28	27	28	27	28	27	28	27	26	28	27	24	23	22
25%	31	30	30	29	29	29	27	26	25	24	23	22	23	22	21	21	20
26%	33	32	30	30	30	29	28	27	26	25	24	23	24	23	22	21	20
27%	36	34	33	31	30	30	30	28	27	26	25	24	25	24	23	22	21
28%	39	37	35	33	31	30	30	30	29	28	27	26	28	27	25	24	23
29%	42	40	38	36	34	32	31	30	30	30	28	27	28	27	25	24	23
30%	46	43	41	39	36	35	33	31	30	30	28	27	28	27	25	24	23
31%	50	47	44	42	40	37	36	33	32	31	30	29	28	27	25	24	23
32%	54	51	48	45	43	40	38	36	34	32	31	30	29	28	26	25	24
33%	60	56	52	49	46	44	41	39	37	35	33	31	30	29	27	26	25
34%	66	62	57	54	50	47	45	42	40	38	36	34	32	30	28	26	25
35%	73	68	63	58	55	51	48	46	44	42	40	38	36	34	32	30	28
36%	81	74	69	64	59	56	52	49	46	43	40	37	34	31	28	25	23
37%	90	83	76	71	66	60	56	52	48	44	40	36	32	28	24	20	18
38%	101	92	84	77	72	64	60	56	52	48	44	40	36	32	28	24	20
39%	114	103	94	86	79	72	66	62	58	54	50	46	42	38	34	30	26
40%	129	117	105	94	86	79	72	66	62	58	54	50	46	42	38	34	30
41%	147	132	117	105	94	86	79	72	66	62	58	54	50	46	42	38	34
42%	169	147	132	117	105	94	86	79	72	66	62	58	54	50	46	42	38

Nota. - Un petit artifice permet d'utiliser la table II pour des valeurs qui n'y figurent pas.

Par exemple, un essai de vente par correspondance a donné les résultats suivants :

Ont choisi le produit A : 4,2%
 " " " " : 3,3%

Pour 42 et 33% la table donne $N \geq 355$. Il suffira de multiplier ce résultat par 10 : l'effectif minimum de l'échantillon d'adresses à utiliser pour qu'un résultat égal à celui qui a été constaté soit significatif (au niveau 5%) est d'au moins 3550.

ANNEXE

Dans l'article cité en référence "Problemi di scelta nelle ricerche di mercato" - Studi di Mercato, Janvier 1956 - Monsieur GENNARO étudie le problème de l'effectif minimum N de deux échantillons de même effectif - nécessaire pour que la différence $p_1 - p_2$ des deux fréquences observées p_1 et p_2 soit significative.

On a dans ce cas :

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1 q_1 + p_2 q_2}{N}}$$

Au niveau 5 %, la différence observée sera significative, si

$$|p_1 - p_2| \geq 1.96 \sqrt{\frac{p_1 q_1 + p_2 q_2}{N}}$$

ce qui exige

$$N \geq 3,84 \frac{p_1 q_1 + p_2 q_2}{(p_1 - p_2)^2}$$

La table ci-après donne l'effectif minimum N pour lequel deux pourcentages observés p_1 et p_2 , dans les conditions ci-dessus, permettent de conclure à une différence significative au niveau 5 %.

Pour les petites valeurs de N, la table a été construite à l'aide de la distribution de Student-Fisher, à partir de

$$\sigma_{p_1-p_2} = \sqrt{\frac{2 p_0 q_0}{N}}$$

avec

$$p_0 = \frac{p_1 + p_2}{2} \qquad q_0 = 1 - p_0$$

soit

$$\sigma_{p_1-p_2} = \sqrt{\frac{(p_1 + p_2)(q_1 + q_2)}{2N}}$$

Exemple : Pour deux tests ayant donné respectivement

$$p_1 = 25\% \qquad p_2 = 16\%$$

la table montre que ces résultats sont significativement différents au niveau 5 %, si leur effectif commun est au moins 153.

Effectif minimum de deux échantillons égaux pour que
la différence entre deux fréquences soit significative
(au niveau 5%)

Table III : de 5 % à 19 %

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1%	138	99	83	65	55	47	41	37	33	30	27	26	25	23	22
2%	282	182	127	103	79	66	56	48	42	37	33	30	29	27	26
3%	735	359	226	154	122	93	76	64	55	47	42	37	33	30	30
4%	3.299	910	435	269	180	141	107	86	72	61	53	46	41	36	33
5%	3.990	3.990	1.081	509	310	206	154	120	96	80	67	57	50	44	39
6%			4.666	1.248	581	351	236	172	133	106	87	73	63	54	48
7%				5.326	1.421	651	394	261	189	145	116	95	79	67	58
8%					5.971	1.571	731	430	286	206	157	125	102	85	72
9%						6.601	1.726	799	468	310	222	169	133	109	91
10%							7.215	1.878	865	505	333	238	181	143	116
11%								7.814	2.026	930	541	355	253	192	151
12%									8.398	2.170	993	576	377	268	203
13%										8.966	2.310	1.054	610	399	283
14%											9.519	2.446	1.114	643	420
15%												10.057	2.579	1.172	675
16%													10.579	2.707	1.228
17%														11.086	2.832
18%															11.578
19%															

Table III (suite) de 20 % à 34 %

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1%	21	24	22	21	20	21	20	21	20	21	20	21	20	21	20
2%	24	23	21	24	23	21	23	21	23	21	22	21	23	21	22
3%	28	26	24	23	21	24	23	24	21	24	22	23	22	23	22
4%	30	30	28	26	24	22	21	22	23	22	20	23	22	25	24
5%	35	32	32	30	27	25	24	22	23	22	20	21	20	23	22
6%	42	38	34	31	31	29	27	25	26	24	23	21	20	23	22
7%	51	45	40	36	33	30	29	28	26	24	23	21	20	23	22
8%	62	54	48	43	38	35	32	32	30	27	25	24	22	25	24
9%	77	66	58	50	45	40	37	33	30	31	28	26	25	23	22
10%	96	81	70	61	53	47	42	38	35	31	29	30	27	26	24
11%	122	101	85	73	64	56	49	44	40	36	33	30	31	28	26
12%	159	129	106	90	77	67	58	51	46	41	38	34	31	32	29
13%	214	167	135	111	94	80	69	60	53	48	43	39	35	32	29
14%	297	224	175	141	116	97	83	72	63	55	50	44	40	36	33
15%	440	311	234	183	147	121	101	86	75	65	57	51	46	41	37
16%	707	459	324	244	190	153	125	105	89	77	67	59	53	47	43
17%	1.283	737	478	337	253	197	158	130	109	92	80	69	61	54	49
18%	2.953	1.336	766	497	350	262	204	163	134	112	95	82	71	63	56
19%	12.054	3.070	1.387	794	515	362	271	211	168	138	115	98	84	73	64
20%	12.515	3.183	1.436	822	532	374	374	280	217	173	142	119	100	87	75
21%	12.960	3.183	1.483	848	548	388	374	280	217	173	142	119	100	87	75
22%	12.960	3.183	1.483	848	548	388	374	280	217	173	142	119	100	87	75
23%	13.390	3.293	1.436	822	532	374	374	280	217	173	142	119	100	87	75
24%	13.805	3.398	1.483	848	548	388	374	280	217	173	142	119	100	87	75
25%	14.204	3.500	1.530	874	574	406	396	296	228	183	152	122	103	89	75
26%	14.588	3.598	1.574	898	579	422	396	296	228	183	152	122	103	89	75
27%	14.957	3.692	1.618	922	594	446	396	296	228	183	152	122	103	89	75
28%	15.310	3.782	1.658	944	594	466	396	296	228	183	152	122	103	89	75
29%	15.648	3.869	1.697	966	594	486	396	296	228	183	152	122	103	89	75
30%	15.971	3.951	1.734	986	594	506	396	296	228	183	152	122	103	89	75
31%	16.278	4.030	1.770	1.006	594	526	396	296	228	183	152	122	103	89	75
32%	16.570	4.105	1.804	1.025	594	546	396	296	228	183	152	122	103	89	75
33%	16.846	4.176	1.836	1.043	594	566	396	296	228	183	152	122	103	89	75
34%	17.107	4.243	1.867	1.061	594	586	396	296	228	183	152	122	103	89	75

Table III (suite) de 35 % à 51 %

7%	20	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
8%	22																	
9%	20																	
10%	22																	
11%	25																	
12%	27																	
13%	30																	
14%	30																	
15%	34																	
16%	38																	
17%	44																	
18%	50																	
19%	57																	
20%	66																	
21%	77																	
22%	91																	
23%	108																	
24%	130																	
25%	159																	
26%	199																	
27%	255																	
28%	336																	
29%	459																	
30%	669																	
31%	1.059																	
32%	1.896																	
33%	4.307																	
34%	17.353																	
35%	17.583																	
	36%																	
	37%																	
	38%																	
	39%																	
	40%																	
	41%																	
	42%																	
	43%																	
	44%																	
	45%																	
	46%																	
	47%																	
	48%																	
	49%																	
	50%																	
	51%																	

Table III suite et fin) de 52 % à 65 %

	52	53	54	55	56	57	58	59	60	61	62	63	64	65
23%	22	21	21	21	21	22	22	22	22	22	21	21	21	21
24%	24	23	23	23	23	23	23	23	22	22	21	21	21	21
25%	26	24	24	24	25	25	25	25	25	24	24	24	24	24
26%	28	26	26	27	27	27	27	27	27	27	27	27	27	27
27%	31	28	29	29	29	29	29	29	29	29	29	29	29	29
28%	30	31	31	31	31	31	31	31	31	31	31	31	31	31
29%	33	30	31	31	31	31	31	31	31	31	31	31	31	31
30%	36	33	33	33	33	33	33	33	33	33	33	33	33	33
31%	40	37	34	34	34	34	34	34	34	34	34	34	34	34
32%	45	41	37	37	37	37	37	37	37	37	37	37	37	37
33%	50	45	41	41	41	41	41	41	41	41	41	41	41	41
34%	56	50	45	45	45	45	45	45	45	45	45	45	45	45
35%	63	57	50	46	46	46	46	46	46	46	46	46	46	46
36%	72	64	57	51	51	51	51	51	51	51	51	51	51	51
37%	82	72	64	57	57	57	57	57	57	57	57	57	57	57
38%	95	82	73	64	64	64	64	64	64	64	64	64	64	64
39%	111	95	83	73	73	73	73	73	73	73	73	73	73	73
40%	130	111	95	83	83	83	83	83	83	83	83	83	83	83
41%	156	131	111	95	95	95	95	95	95	95	95	95	95	95
42%	189	156	131	111	111	111	111	111	111	111	111	111	111	111
43%	234	190	156	131	131	131	131	131	131	131	131	131	131	131
44%	298	235	190	157	157	157	157	157	157	157	157	157	157	157
45%	389	298	235	235	235	235	235	235	235	235	235	235	235	235
46%	528	390	390	390	390	390	390	390	390	390	390	390	390	390
47%	763	763	763	763	763	763	763	763	763	763	763	763	763	763