

De la valeur des réponses individuelles dans une enquête populaire

Par D^r M. Ney, Berne

Lorsqu'une statistique se base sur les réponses individuelles d'une enquête populaire, on ne peut pas prétendre atteindre la vérité absolue. Les renseignements obtenus ne sont qu'une probabilité plus ou moins grande de la réalité. Il ne faut pas en déduire que les résultats d'une statistique ne soient qu'une approximation plus ou moins grossière de la réalité. La tâche de la statistique elle-même a pour objet de chercher à déterminer les causes d'erreur pour les éliminer en tout ou en partie dans les enquêtes ultérieures, ou pour les corriger dans l'enquête en cours.

Les causes d'erreur sont de nature très diverse; elles peuvent être volontaires ou involontaires, elles peuvent dépendre de l'organisation de l'enquête elle-même ou des personnes chargées de donner les renseignements.

I. Motifs des causes d'erreur

a) Les motifs qui produisent les *causes d'erreur dépendant de l'organisation de l'enquête*, et qui sont des causes involontaires, proviennent:

- 1^o du manque de clarté, de simplicité et de précision dans les questions contenues dans le formulaire d'enquête, ces questions devant être à la portée de tous;
- 2^o d'une instruction insuffisante concernant soit l'étendue de l'enquête, soit la détermination des personnes qui sont tenues de répondre, ainsi que celle de l'endroit et de la date où les renseignements doivent être fournis;
- 3^o du choix judicieux de personnes capables de recueillir et de compléter les renseignements (agents recenseurs).

Toutes ces erreurs seront réduites à un minimum ensuite d'une préparation et d'une organisation consciencieuses de l'enquête; celle-ci s'inspirera des expériences acquises et devra surtout veiller à ce que les renseignements demandés et reçus ne soient utilisés que dans le but de répondre aux besoins de la statistique et ne puissent en aucun cas être utilisés comme renseignements individuels.

b) Les motifs qui sont à la base des *causes d'erreur provenant des personnes chargées de répondre au questionnaire*, causes d'erreur qui peuvent être volontaires ou involontaires, sont très divers. Nous pouvons citer comme motifs principaux:

- 1^o une instruction individuelle insuffisante, et le manque de compréhension du but de la statistique basée sur le questionnaire;

- 2° la plaisanterie;
- 3° la distraction et la négligence;
- 4° l'indifférence, pouvant aller jusqu'au refus de répondre;
- 5° la méfiance et la crainte du fisc;
- 6° la vanité, ou le désir de pouvoir retirer un avantage personnel d'une réponse fausse;
- 7° la honte, ou plutôt la fausse honte attribuée à une réponse concernant la morale ou la confession.

La plupart des erreurs individuelles peuvent être écartées par la popularisation en faveur de l'enquête projetée, elles peuvent l'être également par l'institution de mesures pénales dans le cas d'erreurs volontaires manifestes. Un certain nombre d'erreurs personnelles peuvent être déterminées et rectifiées si l'on possède des enquêtes partielles parallèles offrant plus de garanties que l'enquête populaire.

II. Erreurs quantitatives et erreurs qualitatives

Nous pouvons également diviser les erreurs contenues dans les réponses à une enquête populaire en erreurs quantitatives et erreurs qualitatives.

Si nous considérons le recensement de la population comme exemple d'une enquête populaire, nous désignerons comme erreurs quantitatives celles qui résident dans le relevé total des habitants; tandis que les erreurs d'ordre qualitatif seront celles qui seront renfermées dans les réponses faites aux diverses questions contenues dans les bulletins individuels.

a) *Erreurs quantitatives.* Cette sorte d'erreurs se décompose à son tour en erreurs négatives et erreurs positives. L'erreur négative provient du fait que certaines personnes tenues de répondre au questionnaire de l'enquête échappent à la distribution des formulaires, par exemple par suite d'une absence momentanée du lieu du domicile, surtout si cette absence résulte d'un séjour à l'étranger. C'est pour éviter cette erreur que le recensement en Suisse prévoit que pour toute personne momentanément absente de la commune de domicile on devra établir une carte (carte a. b.), et que la personne absente, si elle est en Suisse, établira également une carte (carte b. a.). Cette seconde carte est, par la suite, annulée après avoir servi à compléter et à vérifier la première carte. Malgré cette précaution nous devons admettre que quelques personnes, pour des raisons particulières, se soustraient à l'enquête, ce qui nous permet d'affirmer l'existence d'une erreur quantitative négative, sans pouvoir en déterminer l'importance ni les limites. L'erreur quantitative positive est une conséquence de la mesure de précaution prise pour éviter l'erreur négative, elle consiste en des comptages à double. La compensation entre les deux cartes (cartes a. b. et b. a.) ne se fait que sur la base des réponses données aux questions 10 et 11 des cartes individuelles. Si par erreur il n'a pas été répondu exactement à ces deux questions, il n'est pas possible dans la suite de réintégrer la carte b. a. au domicile régulier de la personne, ni de trouver la compensation à la carte a. b.

Les comptages à double seraient éliminés par le procédé d'un pointage qui pourrait se faire à l'aide d'un classement par ordre alphabétique et par date de naissance de toute la population; ce procédé serait trop long et par suite trop coûteux, proportionnellement au résultat obtenu. En effet, un essai a été tenté lors du recensement de la population en 1920, en ce qui concerne la population du district d'Interlaken, dont le lieu de naissance indiqué était l'une des communes du district. Cette population, d'après la méthode utilisée pour l'ensemble de la Suisse, comprenait 20.073 personnes nées dans le district d'Interlaken et habitant l'une des communes de ce district. Le classement de ces 20.073 cartes par années et jours de naissance a permis la vérification, d'après les registres de l'état civil, dans chaque commune du district, des réponses concernant le lieu de naissance et d'origine, ainsi que la date de naissance de chaque personne. Le pointage de chacune des cartes précitées a fait tout d'abord connaître l'erreur quantitative positive qui a permis de constater que 13 personnes avaient été comptées à double. Ces cas étaient les suivants:

- 1° 3 personnes constituant un ménage habitaient une maison objet de litige entre les communes de Därligen et Interlaken. Elles furent recensées par chaque commune et considérées comme habitant régulièrement chacune des deux communes. (Ce cas doit être considéré comme exceptionnel, le litige a été transigé ensuite à l'avantage de la commune de Därligen).
- 2° Les cartes de 3 personnes absentes n'avaient pu être classées dans leurs ménages de domicile respectifs faute d'indication suffisante (absence de la commune du domicile).
- 3° 5 personnes avaient établi deux cartes, dans la même commune — (cartes a. a.) — à leur lieu de domicile et à leur lieu de travail.
- 4° 2 personnes en place ont été régulièrement recensées dans leur commune de domicile, et ont été également indiquées par leurs parents comme régulièrement domiciliées dans une autre commune, celle du domicile des parents.

Dans ces 13 cas, la compensation des cartes n'avait pas pu se faire par le procédé ordinaire, il s'agissait là de véritables erreurs. L'erreur quantitative positive, résultant du comptage à double, a été ainsi pour le district d'Interlaken, en 1920, de $\frac{13}{20.073} = 0,00065$. L'erreur quantitative totale a, dans notre exemple, pour limite positive $0,65 \text{ ‰}$, cette limite devant être diminuée ou compensée, ou devenir négative si l'on tient compte de l'erreur quantitative négative restée inconnue.

b) Erreurs qualitatives. Nous avons défini les erreurs d'ordre qualitatif comme la conséquence des erreurs renfermées dans les réponses faites aux diverses questions contenues dans les bulletins individuels. Ces erreurs seront plus ou moins nombreuses suivant la difficulté ou la simplicité de la réponse à donner. Une réponse faite en soulignant le mot qui seul convient à la personne, ou l'un des mots «oui» ou «non», sera plus sûre et exacte qu'une réponse qui nécessite une indication

laissée à l'appréciation de la personne questionnée. On peut donc déjà affirmer que, dans le questionnaire du recensement, les réponses concernant le sexe, l'état civil et la religion seront plus exactes que celles qui ont trait à la profession principale et aux professions accessoires. Existe-t-il des moyens d'apprécier et de déterminer la grandeur des erreurs qualitatives? Oui, chaque fois qu'il existe un moyen de contrôler, à la date du recensement, la valeur des réponses par une méthode indirecte offrant plus de sûreté. On peut, par exemple, contrôler les réponses relatives à une profession spéciale si l'on connaît toutes les personnes exerçant cette profession, au moyen de la feuille de solde établie au 1^{er} décembre, pour autant que toutes les feuilles de solde connues embrassent l'ensemble des personnes exerçant la profession déterminée. On peut également vérifier les réponses concernant le sexe, l'âge, le lieu d'origine et le lieu de naissance de la population d'origine suisse à l'aide des registres de l'état civil. Cet essai a été fait, comme nous l'avons déjà mentionné, pour la population habitant le district d'Interlaken et native de ce district. Nous en donnons ci-après les résultats:

1^o *Vérification de la commune de naissance.* La vérification de l'indication de la commune de naissance, telle qu'elle figure dans les 20.060 cartes individuelles de la population native du district d'Interlaken, domiciliée dans ce même district au 1^{er} décembre 1920, a fait constater 432 cas d'erreurs, soit 206 cartes masculines et 226 cartes féminines, ce qui représente une valeur d'erreur qualitative de 2,15 % pour les sexes réunis, de 2,10 % pour le sexe masculin et de 2,21 % pour le sexe féminin. La proportion des erreurs est donc sensiblement la même pour chaque sexe. La principale cause d'erreurs paraît provenir du fait qu'un certain nombre de naissances ont lieu dans des cliniques ou à l'hôpital (pour le district d'Interlaken principalement à Unterseen), alors que le lieu de naissance indiqué est celui du domicile des parents. Ceci explique surtout pourquoi le 43,6 % de l'ensemble des erreurs consiste dans la fausse indication de la commune de naissance, alors que la véritable commune se trouve être une autre commune du même district. Pour l'ensemble des deux sexes et pour chaque sexe, les indications fausses de la commune de naissance se répartissaient comme suit:

La véritable commune de naissance se trouvait située	Pour les deux	Pour le sexe	Pour le sexe
	sexes	masculin	féminin
	%	%	%
dans le même district	43,6	48,3	39,4
dans un autre district du canton	13,0	12,2	13,7
dans un autre canton	6,7	6,4	7,1
à l'étranger	2,8	2,9	2,6
Commune inconnue	33,9	30,2	37,2

Le nombre des communes fausses, pour lesquelles nous n'avons pu trouver d'indication correspondante dans les états civils provenait de 147 cartes (63 du sexe masculin et 84 du sexe féminin). Il ne nous a pas été possible non plus de contrôler pour ces cartes les réponses concernant la commune d'origine et la date de naissance.

2^o *Vérification de la commune d'origine.* L'erreur qualitative se rapportant à la commune d'origine se base sur 19.913 cartes, elle comprenait 435 cartes avec erreurs dont 87 relatives au sexe masculin et 348 au sexe féminin, ce qui représente une erreur totale de 2,18 % pour les deux sexes, soit de 0,89 % pour les hommes et 3,42 % pour les femmes. Cette prépondérance d'erreurs relatives au sexe féminin par rapport au sexe masculin provient du fait que beaucoup de femmes mariées ou veuves indiquent encore comme commune d'origine, non pas celle qu'elles ont acquise par mariage mais celle dont elles étaient ressortissantes avant le mariage.

La répartition des indications fausses était la suivante:

La véritable commune d'origine était située	Pour les deux sexes %	Pour le sexe masculin %	Pour le sexe féminin %
dans le même district	61,1	63,2	60,6
dans un autre district du canton	27,4	21,8	28,7
dans un autre canton	9,9	11,5	9,5
à l'étranger	1,6	3,5	1,2

3^o *Vérification de la date de naissance.* On entend fréquemment citer le fait que dans les recensements de la population, le classement de cette population par âge renferme des erreurs manifestes provenant du fait que bon nombre de personnes indiquent leur âge en chiffres arrondis se terminant par un zéro ou par un cinq, pour la population masculine, tandis que la population féminine d'un âge supérieur à la quarantaine indique de préférence l'âge inférieur se terminant par neuf. Cette cause d'erreur ne peut pas être invoquée pour la population de la Suisse, car le bulletin individuel, au lieu de demander l'âge en années de la personne recensée, demande la date de naissance, soit l'année, le mois et le jour de la naissance.

Le contrôle des 19.913 habitants du district d'Interlaken nés dans ce district a fait constater les erreurs suivantes, en ce qui concerne la date de naissance:

242 bulletins individuels contenaient une erreur de date de naissance, dans le millésime de naissance, ce qui représente le 1,21 % de l'ensemble des cartes, savoir 114 cartes masculines (1,17%) et 128 cartes féminines (1,26%). La différence entre les sexes est minime. Sur les 242 cas d'erreurs, 88, soit un peu plus du tiers, se compensaient mutuellement, tandis que 154 cas étaient de nature à modifier la classification par années d'âge. Ces modifications augmentaient la somme des âges du sexe masculin de 84 années et diminuaient celle des âges du sexe féminin de 50 années. Dans leur ensemble, ces erreurs influençaient les deux sexes d'une différence d'âge moyen de + 0,0017 année, tandis que cette différence aurait été de + 0,0085 pour le sexe masculin et de - 0,0049 pour le sexe féminin. Ces différences n'ont ainsi aucune influence sur l'âge moyen de la population totale ou de la population distinguée par sexe.

Les 154 cas ayant réellement une influence sur le classement par années d'âge représentent ainsi le 7,7 % du total des cartes, tandis que si le classement par âge s'était effectué par groupes de 5 en 5 ans, il n'y aurait plus eu que 34 erreurs non compensées, ce qui ne représentait plus que le 1,7 %.

La répartition des erreurs d'après les différences d'années a été la suivante:

	Différence négative de		Différence positive de		
	Sexe masculin	Sexe féminin		Sexe masculin	Sexe féminin
14 années	— cas	1 cas	1 année	59 cas	55 cas
10 »	2 »	1 »	2 années	3 »	5 »
9 »	— »	1 »	3 »	1 »	2 »
6 »	— »	1 »	5 »	1 »	— »
5 »	1 »	2 »	6 »	1 »	1 »
4 »	1 »	3 »	7 »	2 »	— »
3 »	2 »	6 »	10 »	3 »	— »
2 »	6 »	11 »	12 »	— »	1 »
1 année	30 »	38 »	18 »	1 »	— »
			20 »	1 »	— »

Si nous voulons contrôler la valeur de la critique concernant les erreurs d'âge par rapport aux années se terminant par 0 ou 5, nous trouvons que, sur les 242 cas avec erreurs, 51 cas concernent des âges se terminant par 0 ou 5, ce qui représente le 21 % des erreurs. Or nous trouvons que le nombre des classes d'âge entre 0 et 100 ans qui se terminent par 0 ou 5 est également au nombre de 21, ce qui nous permet de conclure qu'il n'y a pas plus d'erreurs parmi les âges multiples de 5 que parmi les âges intermédiaires.

Influence de l'âge comme cause d'erreur

Ensuite de l'introduction de la loi fédérale concernant la tenue des registres de l'état civil, à partir de 1876, nous avons réparti la population du district d'Interlaken en deux groupes d'âge, le premier comprenant la population née à partir du 1^{er} janvier 1876 (âgée de 0 à 44 ans), et le second comprenant la population née avant 1876 (âgée de 45 ans et plus).

Les trois groupes d'erreurs qualitatives que nous avons déterminées se décomposent alors comme suit:

Erreur qualitative concernant	Population	Population âgée de	
	totale	0 à 44 ans	45 ans et plus
	%	%	%
la commune de naissance	2,15	1,86	3,26
la commune d'origine	2,18	1,97	2,97
la date de naissance	1,21	1,08	1,67

On constate ainsi que c'est dans la population la plus âgée que les erreurs sont proportionnellement les plus nombreuses.

Répartition locale des erreurs

Les erreurs constatées dans le district d'Interlaken réparties par arrondissements d'état civil ont donné le résultat suivant:

Arrondissement d'état civil	Nombre des cartes	Dont avec erreur	En %
Beatenberg	935	28	3,0
Brienz	3657	196	5,4
Grindelwald	2228	108	4,8
Habkern	590	37	6,2
Interlaken	6934	419	6,0
Lauterbrunnen	1888	99	5,2
Leissigen	718	32	4,5
Ringgenberg	1324	65	4,9
Unterseen	1799	96	5,3
Total du district	20073	1080	5,4

Les différences d'un état civil à l'autre ne sont pas considérables, à l'exception de l'état civil de Beatenberg qui, avec le 3 %, est bien au-dessous de la moyenne de 5,4 % du district. Il n'est pas possible de tirer des conclusions en faveur d'un groupe de population spécial.

Erreurs multiples

Dans la première partie de ce travail nous avons mentionné que les causes d'erreurs pouvaient être volontaires ou involontaires; nous pouvons reconnaître sûrement comme erreurs volontaires les cas où toutes les réponses à un questionnaire seraient fausses. Nous avons donc recherché parmi les erreurs constatées si, dans certains cas, ces erreurs se trouvaient être réunies dans un même questionnaire. Nous n'avons pas trouvé de formulaire renfermant simultanément une erreur dans la commune de naissance, la commune d'origine et l'année de naissance (les 147 cartes avec erreurs dans la désignation de la commune de naissance, restée inconnue, n'ayant pas pu être contrôlées). Par contre, nous avons trouvé 29 cartes avec deux erreurs (10 cartes du sexe masculin et 19 cartes du sexe féminin), ces erreurs se répartissant comme suit:

Cartes portant deux indications inexactes, savoir	Total	Masc.	Fém.
Commune d'origine et commune de naissance . . .	14	7	7
Commune de naissance et année de naissance . . .	10	3	7
Commune d'origine et année de naissance . . .	5	—	5
	29	10	19

Conclusion

Notre enquête a ainsi démontré la présence d'erreurs dans les résultats d'une statistique basée sur une enquête populaire.

Ces erreurs, pour autant qu'elles concernent les réponses à des questions simples et dont la réponse ne peut être faite que d'une seule façon bien définie, restent dans des limites restreintes lorsqu'il s'agit du recensement de la population en Suisse. La vérification faite en 1920 dans le district d'Interlaken a fourni les résultats suivants:

1^o une erreur quantitative positive de 0,65 ‰, l'erreur quantitative négative restant inconnue;

2^o erreurs qualitatives.

L'erreur reposait sur l'indication inexacte de	Erreurs			En ‰ du total des cartes		
	Total	Masc.	Fém.	Total	Masc.	Fém.
la commune de naissance	432	206	226	2,15	2,10	2,21
la commune d'origine	435	87	348	2,18	0,89	3,42
l'année de naissance	242	114	128	1,21	1,17	1,26

Ces résultats prouvent que l'enquête populaire est bien près d'atteindre l'image de la réalité dans les cas de réponses simples et définies. L'erreur se rapportant à des questions plus compliquées concernant la profession et surtout le classement social sera certainement plus grande; nous espérons que, pour le recensement de 1930, nous pourrons continuer nos recherches à ce sujet.