REVUE DE STATISTIQUE APPLIQUÉE

L. HENRY

Réflexions d'un « Homme de l'Art »

Revue de statistique appliquée, tome 16, n° 3 (1968), p. 99-111

<http://www.numdam.org/item?id=RSA_1968__16_3_99_0>

RÉFLEXIONS D'UN " HOMME DE L'ART "

par L. HENRY

Institut National d'Etudes Démographiques

L'article de G. Calot "Significatif, non significatif" paru récemment dans cette revue (1) met en scène un personnage dont on ne parle pas souvent dans les études de ce genre, celui qu'il appelle l'homme de l'art ; celui-ci peut être industriel, ingénieur, chercheur dans une discipline ou une autre et il a besoin, dans sa profession, plutôt occasionnellement que de manière habituelle, de recourir à des tests statistiques pour asseoir ses décisions ou ses jugements. Cet homme de l'art connait bien son métier et moins bien, cela va de soi, la statistique mathématique ou la théorie des tests. Comme utilisateur, il aurait peut-être son mot à dire dans ce domaine bien qu'il le connaisse sans doute insuffisamment aux yeux du spécialiste ; encore faut-il qu'on lui donne la parole ou qu'on l'incite à donner son avis

Je profite de l'incitation qu'est l'article de G. Calot pour donner mon point de vue. Démographe de métier et par conséquent statisticien, je me rends bien compte qu'un avis isolé peut ne pas être représentatif ; mais je sais aussi, par expérience, que dans certains domaines on en est réduit à commencer par des études qui ne sont pas forcément représentatives, des monographies par exemple. Ces réflexions jouent en quelque sorte le même rôle ; ainsi dégagé du souci de me vouloir le porte-parole d'une catégorie, je puis donner à ces quelques pages un tour personnel, ce qui permet de bien préciser les choses et d'user d'une franchise qui n'engage que moi.

En démographie courante, l'observation reste très largement exhaustive et l'on peut accumuler les études sans jamais recourir, en raison du nombre très élevé d'observations, à des tests de signification ou de comparaison. Il en va différemment en démographie historique, où l'on n'observe généralement que de petits groupes, et c'est à l'occasion de nombreuses recherches dans ce secteur que j'ai eu à me demander si tel résultat différait significativement de tel autre. Pour répondre à cette question, j'avais à ma disposition ce que j'avais appris à l'ISUP ou, plus exactement, ce qu'il me restait de cet enseignement.

Au début cela suffisait certainement. Mais ensuite, mais actuellement ? La théorie des tests a beaucoup évolué depuis vingt ans, mais, utilisateur épisodique, je n'ai pas suivi cette évolution de près et je me rends mal compte de ce qu'il reste d'utilisable dans ce que j'ai appris en 1947.

On peut dans une certaine optique, celle du professeur dans ses rapports avec ses élèves, considérer l'ignorance comme le fruit de la paresse ou de l'inintelligence ; dans l'optique du chercheur l'ignorance,

---

(1) Revue de Statistique appliquée - 1967. Vol XC. n° 1.

la sienne et celle des autres, est une des données dont il faut tenir compte dans l'organisation de la recherche. En raison de l'accroissement très rapide des connaissances, chacun doit se résoudre à se spécialiser dans un secteur assez étroit et à mal connaitre ou à ignorer ce qui est extérieur, même, si, de temps en temps, il en a besoin. Le problème, fort épineux, est alors de satisfaire ces besoins épisodiques. Dans son énoncé, ce problème est semblable à celui que l'on rencontre dans la vie courante lorsqu'on a besoin d'un nouveau costume ou de faire réparer un tuyau qui fuit. Mais la solution diffère, car s'il existe des tailleurs et des plombiers, il n'y a pas de statisticiens qui fassent métier de soumettre à des tests les résultats obtenus par les chercheurs de diverses disciplines. Restent les collègues et amis, s'ils ont des connaissances en la matières ; mais, si j'en juge par mon expérience, il n'est pas facile, du moins en France où les chercheurs sont couramment surchargés de besogne, de se faire ainsi aider.

Dans une telle situation, quelle solution me reste-t-il ? Je puis évidemment ne pas soumettre les résultats à des tests. Je ne serai pas le seul. Mais c'est une solution à laquelle j'ai de la difficulté à me résoudre.

Une autre solution consiste à appliquer les tests comme je l'ai appris il y a vingt ans. Je l'ai déjà adoptée plusieurs fois, pensant que je risquais plus de paraitre démodé, ce qui est mineur, que de me tromper lourdement.

J'aurais pu enfin me tenir au courant ou m'y remettre. Cette solution serait indiscutablement la bonne si j'avais un besoin fréquent d'utiliser les tests. Je ne suis manifestement pas dans ce cas, puisque je n'ai pas à utiliser ces tests plus d'une ou deux fois par an, et je suis, par suite, enclin à penser qu'il serait peu sage de consentir un investissement intellectuel aussi peu rentable à première vue, alors que j'ai beaucoup à faire d'autre part.

Ce jugement serait toutefois sans fondement s'il était facile d'acquérir le complément de connaissances qui me manque ; ayant essayé, je ne crois pas que ce soit facile.

Il existe peut être un ouvrage en français sur la théorie, et la pratique, des tests qui ait toutes les qualités que je lui demande : présenter les tests tels qu'on les pratique actuellement et être accessible à un homme de l'art habitué à un certain concret et facilement dérouté par une trop grande abstraction. Les ouvrages que j'ai pu consulter, en anglais ou en français, m'ont paru en tout cas éloignés, de cet idéal. Ce sont des ouvrages de professeurs pour des élèves, c'est-à-dire des ouvrages qui permettent à ces derniers de faire un investissement intellectuel très important en vue d'une multitude d'applications ; comme les élèves ne sont en général pas encore entrés dans la vie active ou dans la recherche, le professeur peut limiter au minimum les exemples concrets ; certains d'entre eux sont, après tout, moins familiers aux élèves que le calcul des probabilités qu'ils viennent d'apprendre. En somme, ces ouvrages contiennent un enseignement de caractère général en vue d'applications diverses futures et par conséquent non spécifiées au moment même où cet enseignement se donne.

Cette situation n'est pas du tout la mienne ; démographe, je veux savoir sur quels critères je peux juger que deux petits groupes de femmes mariées de 30-34 ans ont des taux de fécondité légitime substantiellement et significativement différents ; j'ai aussi à me préoccuper du recrute-

ment et de la surveillance de personnes occupées à des relevés de documents (état civil, listes nominatives) et je me demande quelles règles de sélection et de vérification je dois employer pour éviter que les erreurs de relevé dépassent une certaine fréquence, 2 % par exemple.

Chacune de ces questions a un certain nombre de caractéristiques qui permettent de la ranger dans une classe correspondant à un critère ou à plusieurs critères combinés.

Ainsi, la première appartient à la classe des comparaisons de deux moyennes, alors que la seconde appartient à celle des comparaisons d'une moyenne à une valeur bien déterminée.

La comparaison de deux fécondités répond, le plus souvent, à un objectif de recherche pure : il s'agit de savoir, par exemple, si la fécondité des femmes d'un certain âge varie substantiellement d'un groupe à un autre. Dans ce cas je ne sais rien des probabilités a priori c'est-à-dire de la répartition des populations suivant leur fécondité car si je la connaissais je n'aurais pas à me poser une question dont je connais déjà la réponse.

Pour recruter des releveurs, il vaut mieux avoir quelque idée des probabilités a priori, c'est-à-dire de la répartition des candidats suivant leur aptitude à faire des relevés. Sinon, on risque de fixer des normes qu'à peu près personne n'atteint et de ne recruter que des releveurs qui ne conviennent pas, mais qui ont, par hasard, satisfait aux épreuves. Il peut, cependant, être très difficile en pratique de connaitre ces probabilités a priori, ou de les connaitre bien. Le problème du recrutement peut donc se poser dans toute une série de situations au regard de la connaissance des probabilités a priori.

Ces deux exemples font apparaitre un deuxième classement possible suivant que l'on connait, que l'on ne connait pas du tout ou que l'on connait, disons, un peu les probabilités a priori.

En combinant ces critères aux précédents, on arrive à un classement des problèmes en un certain nombre de types, en nombre limité. C'est en partant d'un ou de plusieurs de ces types de problèmes que l'homme de l'art souhaite entrer dans un livre sur les tests et y être conduit, en allant du concret à l'abstrait, à la solution de son ou de ses problèmes. L'ouvrage idéal serait donc pour moi celui qu'un spécialiste des tests rédigerait pour le client que je suis et non pour l'élève que je ne serais plus ; il ressemblerait plus à un catalogue qu'à un livre de classe ou à un cours ; je sais bien que le mot catalogue choquera ceux qui, étant avant tout professeurs, sont habitués à un certain type de rapports, où l'élève ne met en question ni le bien-fondé de ce qu'on lui enseigne, ni la manière d'enseigner ; mais le chercheur qui s'adresse au spécialiste d'une autre discipline n'est pas un élève ; il a de l'expérience, pose des questions bien déterminées et désire que la réponse lui soit donnée sous une certaine forme. Il se présente dans une certaine mesure comme un client dont le spécialiste doit satisfaire les exigences.

Dans l'article de G. Calot certaines exigences de l'homme de l'art apparaissent sous forme de spécifications ; ce seul mot introduit déjà entre l'homme de l'art et le statisticien un rapport de client à producteur ; toutefois les spécifications ne portent que sur les normes de l'outil à fabriquer. G. Calot ne s'occupe pas de la livraison de cet outil, de la manière dont le producteur va initier le client à l'emploi de celui-ci

Pour moi, l'homme de l'art devrait aussi avoir son mot à dire dans la manière dont l'outil sera utilisé, soit qu'il souhaite pouvoir faire résoudre ses problèmes par une sorte de statisticien-conseil, soit qu'il veuille être initié au maniement des tests à sa convenance et non à celle du statisticien.

Les ouvrages courants ne correspondent guère à ces exigences. Certains n'ont à peu près aucune introduction en langage ordinaire et l'on y utilise presque immédiatement le langage mathématique ; il s'agit en somme d'ouvrages consacrés aux solutions mathématiques de problèmes d'énoncé assez général sur lequel on ne s'appesantit pas. Ces mathématiques ne sont pas, le plus souvent, de la compétence de l'homme de l'art ; il n'en conteste pas l'utilité, une fois le problème posé ; mais s'il est chercheur, il sait par expérience que les mathématiques en tant qu'outil ne s'appliquent bien qu'à une manière préalablement façonnée, autrement dit que l'énoncé des problèmes a souvent plus d'importance que les méthodes utilisées pour les résoudre. Comme cet énoncé des problèmes est, au moins en partie, de sa compétence, l'homme de l'art n'a que faire, ou presque, de livres qui finalement ne contiennent, ni exposé détaillé en langage ordinaire des problèmes de jugement sur échantillon et de leurs difficultés, ni enseignement de la pratique des tests.

J'ai l'impression, et je la crois fondée, que celui qui diffuse des connaissances par l'enseignement et le livre se préoccupe trop peu des besoins de ses auditeurs et de ses lecteurs. Passe encore lorsqu'il s'agit de connaissances, littéraires ou scientifiques, auxquelles on peut attacher une valeur propre parce qu'elles répondent à un besoin de culture ; dans ce cas, l'auditeur ou le lecteur paie, pour l'acquisition de nouvelles connaissances, un certain prix (sous forme de temps) sans se demander, s'il est ou non trop cher ; mais quand il s'agit d'acquérir des moyens de connaissance ou de recherche, des outils intellectuels, en quelque sorte, il n'est pas raisonnable d'être indifférent au prix ; un chercheur, par exemple, devrait se demander s'il retrouvera ultérieurement sous forme d'économie de temps et d'efficacité de recherche les jours ou les mois qu'il envisage de consacrer à l'acquisition de tel moyen de recherche, que celui-ci appartienne aux mathématiques pures, à la statistiques mathématique ou à toute autre discipline. En cas de réponse négative, il devrait abandonner son projet, au moins sur le plan professionnel, car il est, bien entendu, libre de consacrer ses loisirs à l'acquisition de connaissances non rentables, s'il leur attribue une valeur en soi.

Ce souci de la rentabilité des investissements intellectuels me semble presque inexistant : tout se passe comme si la connaissance, même celle des méthodes, avait toujours, par elle-même, une valeur assez grande pour que le temps passé à l'acquérir ne soit jamais perdu. Cette mentalité, probablement héritée de l'école, conduit à un gaspillage par suréquipement qui a quelque analogie avec celui qui résulte de l'achat inconsidéré d'ordinateurs prestigieux mais d'utilité incertaine.

Mes réflexions ont pris, au fil de la plume, diverses directions et il me parait indispensable d'en donner un résumé avant de passer à l'examen proprement dit des propositions de G. Calot. Réduites à l'essentiel mes critiques sont les suivantes.

1/ L'homme de l'art a besoin, de temps en temps, d'utiliser les tests, sans être très versé dans leur maniement. Dans ce cas, il ne

trouve aucun guide commode parce que les statisticiens ne se sont guère intéressés à ce mode d'utilisation de leur production.

2/ L'homme de l'art n'est pas un fabricant de tests et de ce fait il ne s'intéresse pas à l'élaboration et à la justification mathématique des tests. Ce qui l'intéresse se situe avant -spécifications, énoncé des problèmes- et après -conditions d'emploi de chaque test.

EQUIVALENCE ET INDECISION

Je reviens maintenant au sujet même de l'article de Calot. Tout y découle de deux idées maitresses, mieux exposées dans la conclusion (p. 65 et 66) que dans l'introduction (p. 10-13).

1/ A toute valeur vraie prise comme référence s'attache un intervalle d'équivalence c'est-à-dire tel que toute valeur vraie comprise dans cet intervalle est considérée comme équivalente à la valeur de référence ; on fera, par exemple, correspondre l'intervalle 495-505 grammes à un poids de 500 g en valeur vraie.

2/ Les limites de l'intervalle d'équivalence ne peuvent être fixées avec précision et il convient de border cet intervalle par deux intervalles -un de chaque côté-, dits d'indécision, qui font la transition entre l'intervalle d'équivalence et la zone de non-équivalence ; on pourra ainsi border l'intervalle d'équivalence de 495-505 grammes par les intervalles d'indécision 490-495 et 505-510.

Pour l'auteur, cet intervalle d'indécision se justifie par le fait qu'on ne saurait passer brutalement de l'équivalence à la non-équivalence ; il faut ménager une zone de transition telle que la tendance à attribuer une valeur de cette zone à l'intervalle d'équivalence diminue à mesure qu'on s'éloigne de cet intervalle pour s'annuler à la frontière de l'intervalle de transition et de la zone de non-équivalence ; dans l'exemple retenu, on a de moins en moins tendance à considérer les poids comme équivalents à 500 gr quand on descend de 495 à 490 ou qu'on monte de 505 à 510 grammes.

Discutons avant d'aller plus loin ces deux idées maitresses.

Equivalence

Je ne pense pas que l'existence d'intervalles d'équivalence puisse être contestée, au moins en principe ; je vois même plusieurs raisons d'introduire ces intervalles :

1/ La nécessité d'apprécier, de juger, de décider nous oblige à simplifier beaucoup et à substituer quelques catégories à la gamme des valeurs possibles d'une grandeur, parfois deux seulement dans les cas extrêmes où l'on ne peut se soustraire à une décision par oui ou non (cas courant dans les examens). Les tests participent de cette attitude naturellement simplificatrice, mais si l'on y gardait toute la richesse de l'information numérique, on conclurait a priori que deux grandeurs sont toujours différentes, ne serait-ce que de fort peu. La notion de différence substantielle, elle-même étroitement liée à celle d'équivalence, est donc inséparable de la pratique des tests.

2/ En dehors même des cas où la nécessité de juger nous porte à des simplifications extrêmes, il est naturel de subdiviser en classes tout ensemble continu et étalé de valeurs vraies ; cette subdivision peut être plus ou moins fine, mais il serait contraire à son objet de la pousser toujours plus loin. On s'arrête quelque part, ce qui revient à con-

sidérer qu'à partir d'une certaine finesse de la subdivision toutes les valeurs d'une classe sont équivalentes.

Ainsi, pour la fécondité légitime, les taux relatifs aux femmes d'un même âge ou groupe d'âges, mettons 20-24 ans, varient dans un intervalle étendu : ils sont donnés, couramment avec trois décimales, mais je pense que les démographes sont généralement enclins à considérer comme équivalents des taux qui ne diffèrent pas de plus de 0,01 ou même de 0,02 ; il suffit d'évoquer nos réactions quand un étudiant ou un débutant insiste trop sur des différences minimes.

3/ Tous les résultats de mesures sont affectés d'erreurs d'observations, différentes des erreurs d'échantillonnage et parfois plus importantes ; pour les populations anciennes, par exemple, l'ensemble des erreurs d'enregistrement, de relevé, de reconstitution des familles font que le résultat d'une mesure sans erreur d'échantillonnage pourrait néanmoins être inférieur de 5 % à la valeur vraie.

Pour les populations modernes des pays développés les erreurs d'observation sont moins élevées et pas seulement par défaut ; le sens dépend, en fait, du mode d'observation ; les erreurs par défaut dominent encore dans le cas de réponses de mémoire à des questionnaires d'enquête ; une immigration continue peut, au contraire, affecter d'erreurs par excès les taux calculés à partir des naissances enregistrées et de la population évaluée.

Dans les pays en voie de développement les erreurs par défaut dominent vraisemblablement ; mais, cette fois, il est très difficile d'en donner une limite supérieure de sorte que l'étendue de l'intervalle d'équivalence reste très vague.

Suivant les cas, et sans doute aussi les tempéraments, on sera porté à mettre plus l'accent sur l'une de ces raisons ; pour G. Calot la première est la plus importante ; pour moi les deux autres viennent en tête, sans doute parce que cantonné dans la recherche, j'échappe le plus souvent à l'obligation de simplifier à l'extrême en vue d'une décision.

Equivalence d'une différence à zéro

L'existence d'intervalles d'équivalence à une valeur vraie entraine l'existence d'équivalence à zéro de la différence entre deux résultats de mesure.

Lorsqu'on ne fait pas intervenir les erreurs d'observation dans la définition des intervalles d'équivalence, on considère comme équivalentes deux valeurs vraies du même intervalle et donc équivalentes à zéro toutes les différences à l'intérieur de cet intervalle, différences qui ont, au plus, l'étendue de l'intervalle comme valeur absolue. Il est alors naturel de considérer aussi comme équivalentes à zéro toutes les différences inférieures en valeur absolue à l'étendue de l'intervalle d'équivalence, qu'elles appartiennent ou non à cet intervalle.

Lorsque les intervalles d'équivalence à une valeur vraie résultent d'erreurs d'observation dans un seul sens, on va encore considérer que deux mesures sont équivalentes si leur différence en valeur absolue ne dépasse pas l'étendue de l'intervalle d'équivalence.

Qu'en est-il lorsque les erreurs d'observation sont dans les deux sens et, pour fixer les idées, aussi fréquentes dans un sens que dans l'autre ? Le résultat est le même que précédemment : l'intervalle d'équivalence est défini de telle sorte que tous les résultats de mesure qui

lui appartiennent puissent correspondre à la même valeur vraie ; dans le cas présent, celle-ci est au centre et l'on ne peut considérer deux résultats comme substantiellement différents tant qu'ils sont moins éloignés l'un de l'autre que ne le sont les deux bornes de l'intervalle d'équivalence.

Bref, l'étendue de l'intervalle d'équivalence à une valeur vraie fournit toujours la limite que la valeur absolue de la différence de deux valeurs vraies ou de deux résultats de mesures sans erreurs d'échantillonnage ne doit pas dépasser pour que cette différence puisse être considérée comme équivalente à zéro.

Zone d'indécision

Je ne conteste pas qu'il y ait un grand arbitraire à fixer des frontières et à passer brutalement de l'équivalence à la non-équivalence et que, de ce fait, il paraisse tout naturel de créer des zones de transition sous forme d'intervalles d'indécision. Un examen plus approfondi et l'usage même que fait G. Calot de ces intervalles d'indécision me fait, cependant, douter qu'il soit toujours indispensable de les introduire.

Dans la zone de transition, j'hésite à conclure entre l'équivalence et la non-équivalence, mais cette indécision ne signifie pas forcément qu'il me soit indifférent de conclure dans un sens ou dans l'autre. Si je ne suis pas tenu de répondre par oui ou non et si je ne puis muer l'indécision en indifférence, je ne veux pas être obligé de répondre oui ou non et il me parait naturel que les tests soient toujours à trois issues ; oui, non et abstention.

G. Calot a une optique différente ; il ne fait pas de distinction explicite entre indécision et indifférence et il assimile, finalement, l'indécision de l'homme de l'art à une indifférence qui laisse les mains libres au statisticien. Comme il s'est fixé, d'entrée de jeu, de répondre, sauf exception, par oui ou non, il a profité de cette indifférence supposée de l'homme de l'art pour couper en deux l'intervalle d'indécision et répondre oui dans une moitié et non dans l'autre. A première vue, cela revient à transporter au milieu des intervalles d'indécision les bornes des intervalles d'équivalence et à avoir, de nouveau, une frontière sans transition. Comme l'intervalle d'équivalence est lui-même défini avec assez peu de précision, on peut se demander s'il était nécessaire d'emprunter le détour de l'intervalle d'indécision pour revenir à peu près au point de départ.

En fait, les erreurs d'échantillonnage créent une zone de flou où l'on risque de répondre oui pour non et non pour oui ; il faut éviter que cette zone de flou tombe mal, c'est-à-dire là où, au contraire, on veut avoir une réponse nette ; on a donc à définir la zone où il y a le moindre inconvénient à savoir des réponses incertaines et la zone d'indécision a comme rôle de recevoir la zone de flou du test. Mais lorsqu'on prévoit trois issues, la zone d'abstention joue ce rôle et il est inutile d'ajouter une zone d'indécision.

En conclusion, G. Calot a donné à l'indécision un sens qui le conduit à la lever ; mais comme il a, au préalable, choisi de s'en tenir à des tests à deux issues et non trois, il a été obligé de maintenir, dans un premier temps, cette indécision qu'il abandonne ensuite.

AUTRES SCHEMAS POSSIBLES

Le schéma proposé par G. Calot n'est qu'un des schémas possibles avec les intervalles d'équivalence. On peut le modifier de deux manières.

1/ On peut tirer argument du fait que les limites de l'intervalle d'équivalence sont elles-mêmes vagues et qu'on est, par suite, conduit à les fixer avec quelque arbitraire pour ne pas introduire, en plus, des intervalles d'indécision qu'on est ensuite conduit à abandonner après les avoir partagés entre les deux zones encadrantes.

2/ On peut aussi considérer que l'indécision est irréductible puisqu'elle n'a rien à voir avec les erreurs d'échantillonnage et qu'on doit en conséquence, opérer de manière qu'elle subsiste avec des observations très nombreuses ; dans ce cas, l'incertitude et par suite l'abstention font partie des jugements à porter sur le vu des mesures ; les tests vont donc être logiquement à trois issues ; oui, abstention et non.

SCHEMA SANS INDECISION

Examinons ce schéma sur les deux tests suivants :

a) test unilatéral de signification d'une moyenne et de comparaison de deux moyennes.

b) test bilatéral de comparaison de deux moyennes.

a) Test de signification d'une moyenne et test unilatéral de comparaison de deux moyennes.

Il s'agit d'abord de savoir si la valeur vraie correspondant à une moyenne est significativement inférieure, équivalente ou supérieure à une certaine valeur, 0,400 par exemple pour un taux de fécondité légitime .

Plaçons-nous dans le cas d'un intervalle d'équivalence à 0,400 allant de 0,390 à 0,410.

Avec un très grand nombre d'observations les résultats se répartissent en trois catégories :

1 - Fécondité plutôt faible (inférieure à 0,390).

2 - Fécondité équivalente à 0,400.

3 - Fécondité plutôt forte (supérieure à 0,410)

Avec un nombre d'observations pas très grand on va avoir à chercher deux valeurs c_1 et c_2 relatives à la borne inférieure m_i de l'intervalle d'équivalence telles que, m étant le taux observé, l'on conclue fécondité plutôt faible pour $m < c_1$ et fécondité non plutôt faible pour $m \geqslant c_2$; à la borne supérieure m_s correspondront aussi deux valeurs c_3 et c_4 telles que l'on conclue fécondité plutôt forte pour $m > c_4$ et fécondité non plutôt forte pour $m \leqslant c_3$. On fixe c_1 de manière que la probabilité de conclure fécondité faible soit inférieure à un seuil δ fixé à l'avance lorsque la valeur vraie m^* est égale à m_i et, a fortiori, supérieure (1).

---

(1) m^* peut être aussi, la valeur, sans erreur d'échantillonnage, qu'on obtiendrait avec un très grand nombre d'observations ; elle peut être encore entachée d'erreurs d'observation et différer par là de la valeur vraie.

On peut, de manière analogue, fixer c_2 de manière que la probabilité de conclure fécondité non faible soit inférieure au seuil δ lorsque m^* est inférieur, d'aussi peu que ce soit, à m_i.

Entre c_1 et c_2 les deux conditions imposées ne sont pas remplies ; on ne peut donc rien conclure. Il en est du même entre c_3 et c_4.

On a donc, de toute manière, deux zones de réponse assurée, l'une à gauche de c_1 l'autre à droite de c_4 et deux intervalles d'abstention (c_1, c_2) et (c_3, c_4).

Deux cas peuvent se produire :

1/ Les deux intervalles d'abstention sont disjoints. Dans ce cas, ils laissent la place entre c_2 et c_3 à une plage d'équivalence, dont l'étendue est inférieure à celle de l'intervalle d'équivalence (m_i, m_s).

2/ Les deux intervalles d'abstention se chevauchent ; il n'y a plus d'intervalle d'équivalence à m^* pour les résultats de l'observation, l'intervalle d'équivalence étant, totalement recouvert par les intervalles d'abstention.

Les résultats du test ainsi conçus sont satisfaisants pour l'esprit ; on ne conclut que lorsque l'on peut vraiment le faire, soit parce que les moyennes observées sont loin, d'un côté ou de l'autre de l'intervalle d'équivalence, soit parce qu'il reste une plage d'équivalence entre les intervalles d'abstention et que la moyenne observée est tombée dedans. Sinon, on s'abstient ce qui est, bien sûr, une invitation à augmenter le nombre d'observations. La situation a, en somme, quelque analogie avec celle qu'on rencontre dans les tests séquentiels.

Le test unilatéral de comparaison de deux moyennes ne diffère du précédent que par la valeur particulière, zéro, de la valeur vraie de référence. On va avoir deux zones de non-équivalence où l'on conclura que les valeurs vraies sont dans le même ordre que les moyennes ; entre les deux zones il y aura au moins une zone d'abstention, où l'on ne pourra conclure ni que les valeurs vraies sont équivalentes ni qu'elles diffèrent substantiellement ; dans certains cas, enfin, on aura aussi, entre les zones d'abstention, une plage d'équivalence où l'on conclura que les deux valeurs vraies sont équivalentes.

Remarque : équivalence ou indécision ?

Dans ce qui précède, on a considéré l'équivalence et l'indécision comme bien distinctes ; mais on pourrait aussi avancer que l'intervalle d'équivalence à une valeur vraie, 0,400 par exemple, contient toutes les valeurs que l'on hésite à juger distinctes de cette valeur vraie ; ou encore, pour une différence, que l'intervalle d'équivalence à zéro, contient toutes les différences que l'on hésite à juger substantielles. Il suffirait donc de changer un peu la présentation des choses pour transformer l'intervalle d'équivalence, à une valeur vraie ou à zéro, en intervalle d'indécision.

C'est ce qu'à fait G. Calot, sans le préciser, dans les tests de signification de moyennes, et comme il assimile ensuite indécision à indifférence, il en arrive, en fin de compte, à décider que la valeur vraie est plutôt faible lorsque la moyenne observée tombe à gauche de $\frac{m_i + m_s}{2}$ et plutôt forte si elle tombe à droite de $\frac{m_i + m_s}{2}$, sauf si le nombre d'observations est vraiment trop petit. Ce glissement de l'équivalence à l'indécision, il ne le maintient pas, toutefois, dans les tests

de comparaison de deux moyennes : dans ce cas, les deux intervalles existent et l'on est conduit, en fin de compte, à décider, par exemple, que la valeur vraie n° 1 est plus grande que la valeur vraie n° 2 si la différence des moyennes $(\bar{x}_1 - \bar{x}_2)$ est supérieure à la moyenne arithmétique $\frac{d_1 + d_2}{2}$ des limites de l'intervalle d'indécision (d_1, d_2) qui borde à droite l'intervalle d'équivalence à zéro (1).

b) Test bilatéral de comparaison de deux moyennes

Dans ce cas, on ne fait intervenir que les valeurs absolues des différences et l'intervalle d'équivalence est limité d'un côté par 0 de l'autre par d. On opère comme précédemment, mais seulement à la limite supérieure d et l'on conclut dans certains cas :

à l'équivalence

à l'abstention

ou à la différence substantielle

Dans d'autres cas, caractérisés par une variance plus forte des moyennes observées,

à l'abstention

ou à la différence substantielle

Formules à utiliser

Elles se déduisent très simplement de celles que donne G. Calot.

Dans tous les cas, il suffit d'appliquer à chaque extrémité de l'intervalle d'équivalence, à m^* ou à zéro, ce que deviennent les règles formulées par G. Calot quand l'intervalle d'indécision est nul c'est-à dire quand m_1 et m_2, ou d_1 et d_2, sont confondus (1).

On a, par exemple, pour les valeurs c_1, c_2, c_3, c_4 correspondant au test de signification d'une moyenne.

$$c_1 = m_i - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

$$c_2 = m_i + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

$$c_3 = m_s - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

$$c_4 = m_s + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

Pour le test unilatéral de comparaison de deux moyennes, les valeurs c_1, c_2, c_3, c_4 auxquelles on doit comparer $\bar{x}_1 - \bar{x}_2$ sont, de manière analogue,

---

(1) Cette différence de traitement se justifie d'ailleurs. Dans le test de signification l'emplacement où la zone de flou est le moins gênante est l'intervalle d'équivalence ; dans les tests de comparaison cet emplacement est, au contraire, aux frontières de l'intervalle d'équivalence à zéro.

(2) m_1, m_2, d_1 et d_2 sont les symboles de l'article de G. Calot.

$$c_1 = -d_i - \sigma_d \; u_{1-\delta}$$
$$c_2 = -d_i + \sigma_d \; u_{1-\delta}$$
$$c_3 = d_s - \sigma_d \; u_{1-\delta}$$
$$c_4 = d_s + \sigma_d \; u_{1-\delta}$$

$-d_i$ et d_s étant les limites, respectivement inférieure et supérieure, de l'intervalle d'équivalence à zéro, σ_d l'écart-type de la différence, c'est-à-dire :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

c_1 et c_4 existent toujours ; c_2 et c_3 ne subsistent, par contre, qu'à la condition $c_2 < c_3$ c'est-à-dire.

$$n > \left(\frac{2\,\sigma_0\,u_{1-\delta}}{m_s - m_i}\right)^2$$

condition qui est la même que celle qui figure à la page 20 de l'article, de G. Calot. Dans ce cas, il introduit un intervalle d'abstention allant de c_2 à c_3, alors qu'ici, dans la même situation, les deux intervalles d'abstention se trouvent fondus en un seul qui s'étend de c_1 à c_4.

Si les écarts-types ne sont pas connus, on remplace $u_{1-\delta}$ par $t_{1-\delta}$ pour le nombre de degrés de liberté convenable ($n - 1$ pour une signification de moyenne, $n_1 + n_2 - 2$ pour une comparaison) et σ_d par

$$s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Dans le cas du test bilatéral de comparaison de moyennes, on n'a plus que deux valeurs à considérer, c_i et c_s.

$$c_i = C_1' \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$c_s = C_2' \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

C_1 et C_2 étant donnés par la table 2 (p. 33) en fonction de

$$L_1 = L_2 = L = \frac{d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Lorsque L est grand, c'est-à-dire lorsque l'intervalle d'équivalence est grand comparé à l'écart-type de la différence, on a à peu près

$$C_1 = L - u_{1-\delta}$$
$$C_2 = L + u_{1-\delta}$$

et par conséquent

$$c_i = d - u_{1-\delta} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = d - u_{1-\delta}\ \sigma_d$$

$$c_s = d + u_{1-\delta} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = d + u_{1-\delta}\ \sigma_d$$

c'est-à-dire les mêmes formules que dans le test unilatéral

c_s existe toujours ; c_i ne subsiste que si l'on a $d > u_{1-\delta}\ \sigma_d$; lorsqu'il n'en est pas ainsi, l'intervalle d'équivalence est tout entier recouvert par un intervalle d'abstention qui s'étend de 0 à c_s.

Indécision irréductible

Dans ce cas on a quatre valeurs, m_{i1}, m_{i2}, m_{s1}, m_{s2}, qui vont devenir chacune le centre d'un intervalle d'abstention. Considérons par exemple, les bornes de l'intervalle d'indécision à droite, m_{s1} et m_{s2}. Posant $\frac{\sigma_0}{\sqrt{n}}\ u_{1-\delta} = k$ on concluera que la valeur observée est plutôt forte pour

$$\bar{x} > m_{s2} + k$$

et équivalente à m^* pour

$$m_{i2} + k < \bar{x} < m_{s1} - k$$

si toutefois,

$$m_{s1} - k > m_{i2} + k$$

Entre $m_{s1} - k$ et $m_{s2} + k$, se situent les intervalles d'abstention et, le cas échéant, la partie de l'intervalle d'indécision qui n'est pas recouverte par ceux-ci. Indécision et abstention étant deux attitudes voisines, parfois prises, d'ailleurs, pour la même raison, l'imprécision des résultats, il paraît naturel de ne pas distinguer ce qui peut rester des intervalles d'indécisions entre les intervalles d'abstention qui le bordent. On a alors un schéma analogue au précédent avec de gauche à droite :

valeur vraie plutôt faible

abstention (ou indécision)

équivalence

abstention (ou indécision)

valeur vraie plutôt forte

La seule différence est que chaque intervalle d'abstention est majoré de l'étendue de l'intervalle d'indécision

Je ne suis pas persuadé que ce schéma vaille mieux que le précédent ; je crains qu'en voulant quantifier le flou des frontières de l'intervalle d'équivalence on ne se donne l'impression d'une précision impossible à atteindre.

UNE DIFFICULTE : L'EQUIVALENCE DE PLUSIEURS RESULTATS

L'équivalence de deux valeurs vraies ou de mesures sans erreur d'échantillonnage a été définie par une distance à ne pas dépasser ; tant que cette condition est remplie, les deux valeurs ou mesures peuvent être équivalentes à la même valeur vraie ; il parait naturel d'opérer de même lorsqu'on a affaire à plus de deux valeurs ou mesures et de les considérer comme équivalentes si elles peuvent correspondre à une même valeur vraie, ce qui implique qu'elles sont toutes à moins d'une certaine distance d'une valeur de référence et, par conséquent, qu'aucune de leurs distances mutuelles ne dépasse l'étendue de l'intervalle d'équivalence.

Cette condition ne fait intervenir aucune moyenne, ni des distances, ni de leurs carrés.

Or G. Calot traite du test d'homogénéité de plusieurs moyennes dans le cas où l'équivalence est définie à partir de la variance des valeurs vraies. Il ne discute pas ce choix et agit, en somme, comme s'il était évident qu'on n'en puisse faire aucun autre.

Cette évidence ne m'apparait pas, de sorte que je me demande si le choix fait est vraiment le seul qu'on puisse faire. Je lui reconnais l'avantage de la commodité puisqu'il permet de soumettre à un seul test autant de moyennes qu'on veut, mais cela ne saurait me suffire.

Me voici donc devant une difficulté : j'ai une certaine conception de l'équivalence de plusieurs valeurs vraies alors que le statisticien a construit ses tests sur une autre. C'est le cas type ou une discussion serait utile.

CONCLUSION

Dans la première partie j'ai essayé de montrer que les rapports entre les spécialistes de la statistique mathématique et l'homme de l'art, utilisateur éventuel des résultats obtenus dans cette discipline, ne satisfont pas toujours ce dernier. La raison principale de cette situation, est, à mes yeux, un manque de réflexion approfondie sur la collaboration entre les créateurs de méthodes et les utilisateurs de celles-ci. On en reste trop, à mes yeux, à des rapports unilatéraux analogues à ceux qui existent entre un professeur et ses élèves alors qu'il serait préférable d'instituer des rapports bilatéraux, du genre de ceux qui existent entre des gens qui sont tour à tour fournisseurs et clients.

Cette critique s'étend à toute la collaboration entre disciplines différentes ; on en parle beaucoup, mais on ne cherche guère à la faciliter.

Dans la seconde partie, j'ai repris les idées de G. Calot et tenté de discuter plus qu'il ne l'a fait, les notions d'équivalence d'indécision et d'indifférence.

Suivant le sens qu'on leur donne, on utilise les tests comme G. Calot le préconise ou d'une autre manière ; celle-ci se déduit d'ailleurs assez facilement des procédés indiqués par G. Calot.

Sur un point, cependant, je rencontre une difficulté assez sérieuse ; j'ai une notion de l'équivalence de plusieurs valeurs vraies différente de celle qui est couramment adoptée.