Document généré le 2 juin 2025 21:58

L'Actualité économique

Exposé d'un plan d'échantillonnage des usagers d'un réseau de transport

Robert Stock

Volume 33, numéro 1, avril–juin 1957

URI : https://id.erudit.org/iderudit/1001214ar
DOI : https://doi.org/10.7202/1001214ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)
1710-3991 (numérique)

Découvrir la revue

Citer cet article

Stock, R. (1957). Exposé d'un plan d'échantillonnage des usagers d'un réseau de transport. *L'Actualité économique*, *33*(1), 51–110. https://doi.org/10.7202/1001214ar

Tous droits réservés © HEC Montréal, 1957

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Exposé d'un plan d'échantillonnage des usagers d'un réseau de transport

Depuis quelques années, le public s'est de plus en plus intéressé aux sondages d'opinions, et par là au procédé statistique dit d'échantillonnage. Il se passionne pour les résultats que publient certaines organisations spécialisées dans ce genre de travail et consulte leurs bulletins avec une curiosité mêlée de scepticisme et de confiance tout à la fois; un peu, en somme, comme il s'intéresse aux conseils des astrologues.

Par ailleurs, il se fait une idée assez grossière des précautions qu'il a été nécessaire de prendre pour que l'échantillonnage apporte des résultats qui puissent être considérés comme représentatifs de la réalité. Il croit volontiers qu'il suffit de questionner un certain nombre de personnes, choisies un peu n'importe comment; comme, par exemple, de s'installer à l'angle des rues Sainte-Catherine et Saint-Denis pour y questionner, disons, les 100 premières personnes qui passeront.

En fait, un bon échantillonnage n'est rien d'aussi simple. Car si les mathématiciens et les statisticiens peuvent affirmer que la méthode conduit à des résultats satisfaisants, c'est qu'ils sont en mesure de le démontrer mathématiquement par des théories passablement rigoureuses, en même temps que fort ingénieuses et complexes. Mais alors, justement, il faut, pour que les conclusions de ces théories s'appliquent à des échantillonnages concrets, respecter les conditions à l'intérieur desquelles elles sont seules valables. Autrement, ces théories elles-mêmes montrent que l'échantillonnage risque hautement d'être sans valeur.

L'article qui suit, rédigé par Monsieur Robert Stock, professeur de statistique à l'École des Hautes Études commerciales, apporte un exemple fort intéressant de mise sur pied et de réalisation d'un échantillonnage conduit selon les règles de l'art. Et l'exemple n'est pas purement imaginaire. Il correspond à une opération réelle que l'auteur lui-même a dû réaliser, à Montréal, dans des conditions particulièrement difficiles du point de vue statistique. En pleine possession des théories mathématiques sous-jacentes, et face à une donnée complexe, il a donc dû inventer au fur et à mesure des besoins les procédés nécessaires pour rendre l'échantillonnage valide, mathématiquement parlant.

C'est à l'opération de réalisation de ces moyens que l'article va nous faire assister, dans l'ordre même de leur apparition au cours de la mise en oeuvre; et non pas à un exposé ou à une analyse des résultats, qu'il n'était pas au pouvoir de l'auteur de dévoiler. Ce n'est pas au spectacle monté que nous sommes conviés, mais bien au montage même du spectacle; nous ne serons pas dans la salle, mais dans les coulisses. En un sens, cela apparaîtra sans doute plus passionnant encore pour tous ceux qui s'interrogent sur ce nouveau moyen de pratiquer la vérité . . . ou le mensonge (!) statistique qu'est l'échantillonnage. Et les spécialistes eux-mêmes, croyons-nous, trouveront profit et intérêt à l'exposé.

Qui prend le tramway? Que tout le monde prenne le tramway, comme on l'affirme parfois, ou qu'une classe particulière de la population forme à elle seule le groupe des usagers, comme on le soutient aussi, on demande en somme ici d'explorer les faits par delà les opinions, et il appartient en définitive à la statistique de le faire. Et ce sont les moyens qui ont été effectivement mis en œuvre en ce sens qu'il s'agit maintenant d'exposer.

D'ailleurs, il s'agirait même de situer cet exposé dans une ambiance laissant place à quelques généralités sur les méthodes de la statistique. La chose se comprend, car il est certain que la statistique, après des débuts mal compris, s'implante rapidement et fortement dans la vie de tous les jours du monde des affaires. Grâce à des développements relativement récents dirigés spécifiquement vers la rationalisation des décisions administratives, on peut dire que sous cette forme la statistique se trouve en quelque sorte dans la situation qu'a connue la comptabilité il n'y a pas si longtemps, alors que cette dernière venait à peine de s'imposer comme un moyen de contrôle inéluctable, sans se douter qu'elle n'était que le prélude d'une analyse appelée à devenir scientifique. En fait, il a fallu se borner ici aux aperçus les plus élémentaires touchant aux calculs pour s'en tenir, si l'on peut dire, au côté pittoresque du tableau, risquant de perdre, si l'on n'y prenait garde, la notion même des méthodes statistiques, lesquelles ne permettent qu'on fasse un seul pas sans payer tribut au calcul des probabilités.

Malgré les apparences, l'énoncé de la question posée exigera une réponse essentiellement quantitative, et la décision administrative en jeu ne saurait être prise autrement que sur la recommandation d'un ou plusieurs résultats chiffrés pertinemment interprétés. Quant à ces résultats, ils seront vraisemblablement à attendre d'un échantillonnage ou d'un recensement complet. D'ailleurs, le cas à l'étude revêt sans aucun doute un caractère analytique, en ce sens que l'on ne saurait se désintéresser des causes de variabilité des diverses caractéristiques de la population des usagers du réseau, et l'étude de ces causes est elle-même justifiable, non plus d'un échantillonnage unique, ou d'un seul recensement, mais bien plutôt d'un système d'échantillonnages ou de recensements ordonnés dans le temps. Ce qui revient à dire que si rien n'empêche qu'une

décision administrative soit prise dès le premier échelon du système, rien n'autorise par contre à se montrer satisfait par la suite d'une connaissance figée dans le temps des caractéristiques de la population, et sans exercer un contrôle permanent sur les conséquences de telle décision ou sur toute évolution subséquente. Pour ne citer qu'un exemple, il pourrait arriver qu'une variation dans la qualité des usagers vienne à être masquée par un accroissement de la population en général, et par conséquent non apparente sur les états comptables, alors qu'elle serait due à une désaffection d'une partie des usagers pour ses transports en commun, et cette partie des usagers pourrait être précisément celle qui aurait bénéficié de la décision administrative à l'essai. Mettons que ce soit là un cas extrême, mais les cas extrêmes se produisent, et il en est d'autres qui ne sont pas sans importance.

Nous avons bien dit que la réponse à la question posée devait s'appuyer sur un résultat numérique, ou plutôt même sur une suite de tels résultats ordonnés dans le temps, et cependant ne doit-on pas l'attendre décisive et sans appel si l'on prend en considération la nature des moyens disponibles. Pas même un recensement complet, nonobstant sa nature rassurante et son coût élevé, ne pourrait prétendre lui conférer cette qualité, car, tout recensement qu'il est, il ne sera encore qu'un échantillon dans le processus analytique, et rien ne pourra lui ôter l'erreur d'estimation dont il sera de ce fait entaché.

De tout ceci, faut-il conclure que l'on doive se laisser gagner par le doute sur l'efficacité des méthodes statistiques appliquées au domaine qui nous occupe présentement? Certainement non, car dans l'ordre des décisions administratives, quelles qu'elles soient, entre une part de risque dont il faut apprendre à s'accommoder; et le plus sûr moyen de le faire avec sagesse est bien de savoir mesurer le dit risque. Et c'est ce que nous nous proposons de montrer ici, dans le cadre d'une situation particulière.

* * *

Il semble manifeste que la question à débattre maintenant est celle-ci: recensement ou échantillonnage? D'une façon générale, le recensement, isolé de toute étude antérieure ou subséquente,

a l'avantage de ne poser aucune question relevant des probabilités: tout ce qui est à compter ou à mesurer y passe, et il n'y a aucune erreur d'estimation — le résultat est net. Par contre, l'échantillonnage s'appuie entièrement sur le calcul des probabilités, et ses méthodes sont telles que l'on doit s'attendre à connaître l'état d'une caractéristique de la population totale d'une façon quelque peu inaccoutumée. Par exemple, le résultat d'un échantillonnage peut nous mettre en mesure de dire que tel pourcentage exact, relatif à la population totale, se situe entre 23 et 25 p.c., sans manquer d'ajouter qu'en fixant ces deux bornes à l'intervalle réputé contenir le vrai pourcentage, on court un risque de se tromper dans 3 p.c. des cas, par exemple. On pourrait d'ailleurs augmenter l'intervalle en question, agissant par suite sur la dite probabilité de se tromper, au point de pouvoir rendre cette dernière pratiquement négligeable. Du point de vue pratique, sachant donc que le vrai pourcentage n'est pas supérieur à la plus grande des deux bornes et qu'il n'est pas inférieur à la plus petite, on pourra prendre une décision administrative relative à la population totale avec un risque calculé. Pour être utilisables, ces bornes devront être plus ou moins rapprochées l'une de l'autre selon la nature ou la gravité de la décision à prendre, et ce sera en agissant sur le montant des dépenses allouées que l'on pourra en fait les tenir dans un intervalle susceptible de concilier à la fois le coût de l'investigation et le risque de la décision.

Ce qui vient d'être dit sur le recensement, celui-ci étant toujours considéré dans un état isolé, c'est-à-dire avec un but énumératif seulement, ne signifie pas qu'il soit appelé à fournir un résultat exempt de toute erreur. S'il est vrai que le recensement complet d'une population ne saurait souffrir du même mal que l'échantillonnage, en ce sens que ce dernier doit passer à la population totale sur la foi d'un échantillon courant le risque d'être plus ou moins représentatif, il est par contre sujet, et il l'est à un haut degré, à ces sortes d'erreurs qui possèdent le grave inconvénient de ne pouvoir être contrôlées. Nous voulons parler plus particulièrement de ces multiples interférences avec le vrai que commettent sciemment, ou le plus souvent à leur insu, interviewer comme interviewé. Sans doute l'échantillonnage ne se trouve pas à l'abri de ces sortes d'erreurs; mais il y est sujet à un degré bien moindre, vu

que son effectif d'interviewers étant très réduit par rapport à celui que demande le recensement de la même population, le soin que l'on peut donner dans un cas est pratiquement impossible à apporter dans l'autre. D'ailleurs, la mesure dans laquelle un recensement peut ainsi s'écarter de la réalité, ce qui se contrôle ordinairement par un échantillonnage subséquent, peut aller jusqu'au point où la prudence exige de retenir la publication du résultat obtenu.

Indépendamment de tout ceci, le coût élevé d'un recensement, surtout s'il devait être répété, comme il a été suggéré dans le cas actuel, constituerait à lui seul un obstacle insurmontable, et bien d'autres viendraient s'y ajouter. Aussi, pour éviter de pousser plus avant un parallèle inutile dans les circonstances, entre recensement et échantillonnage, convenons tout de suite d'en rester à ce dernier mode d'investigation. D'ailleurs, les obstacles signalés deviendront évidents par l'exposé qui va suivre si l'on veut bien se représenter un recensement comme étant un cas limite de l'échantillonnage, alors que le nombre des éléments de la population à prendre dans celui-ci va jusqu'au point d'englober la totalité de la population.

Nous ne nous attarderons pas pour cela au cas de l'échantillon dit «représentatif», bien qu'il se trouve toujours en tête de liste des solutions gratuitement suggérées. Non pas que l'échantillon représentatif ne puisse exister, et avec grand mérite d'ailleurs. Un couturier, par exemple, réussira très bien à former un excellent échantillon représentatif de sa collection, pourvu seulement qu'il puisse trouver un commun diviseur aux nombres d'articles exécutés dans les différents modèles, couleurs, tailles, etc. Mais évidemment s'agit-il ici d'un échantillon constitué à posteriori, tandis que celui que nous attendons est destiné au contraire à estimer les caractéristiques d'une population inconnue. Pris dans ces dernières conditions, l'échantillon «représentatif» n'est autre que le fruit d'un jugement à priori sur la population, portant sur toute caractéristique censément mise à l'étude, et le jugement étant tout arrêté, l'échantillon devient un leurre, sinon un subterfuge dont on se sert pour se justifier. On serait fort embarrassé, d'ailleurs, s'il fallait assigner une erreur d'estimation à un tel échan-

tillon, puisqu'il échappe complètement aux probabilités, à moins qu'on ne le considère représentatif au point de n'avoir pas d'erreur du tout.

* * *

Et maintenant, pour revenir et rester dans le domaine de l'échantillonnage basé sur les probabilités, disons que la condition première à devoir être satisfaite est d'être en mesure de se donner une définition sans équivoque de la population à échantillonner ou, autrement dit, d'être en mesure de se procurer une liste complète, sous une forme ou sous une autre, de tous les éléments de la dite population. Ensuite, ayant présentes à l'esprit la nature et surtout la gravité de la décision ou des décisions administratives à prendre, on devra se fixer une limite pour l'erreur dont pourra être entachée l'estimation tirée de l'échantillon et à partir de laquelle telle décision sera prise, et, par dessus tout, on fixera une limite au risque que l'on peut accepter de prendre une mauvaise décision. Viendra ensuite un inventaire complet des divers procédés, classiques ou inédits, qui peuvent s'offrir pour échantillonner la population à l'étude, compte tenu de toutes ses particularités, certains de ces procédés pouvant d'ailleurs se trouver éliminés par suite des effets manifestes des restrictions administratives qu'on ne manquera pas de rencontrer. Puis, à partir des limites qu'on vient à l'instant de se fixer, à partir de certaines caractéristiques de la population sur lesquelles on devra faire des estimations, le calcul donnera le contenu que devra avoir l'échantillon ou, si celui-ci doit se faire en plusieurs stades, le contenu de chacun des échantillons successifs.

En fait, une autre variable, et non la moindre, vient encore prendre sa place parmi les autres variables du problème: le coût de l'échantillonnage. Celui-ci peut comprendre des dépenses fixes, mais il est évidemment lié d'une manière ou d'une autre au contenu de l'échantillon, si bien que les diverses variables dont il vient d'être question se trouvent réagir les unes sur les autres, tandis que, par surcroît, le coût est ordinairement plafonné par des considérations purement économiques. C'est en faisant une étude analytique de la fonction liant toutes ces variables pour un plan d'échantillonnage donné que l'on arrivera à une solution, et cette solution en sera ordinairement une de compromis, où l'on

aura consenti à diminuer le coût au détriment du risque ou à augmenter le coût pour ne pas accroître le risque. De tout ceci sortira, et le procédé d'échantillonnage et le contenu de l'échantillon, ou le contenu des échantillons successifs suivant le cas.

En guise d'étude des diverses méthodes d'échantillonnage que l'on aurait lieu de considérer dans le cas qui nous occupe, disons deux mots de celui-ci, que l'on ne saurait classer sans l'avoir examiné. Imaginons de nous fixer une circonscription territoriale intéressant la question et bien délimitée, telle la ville de Montréal, et de procéder à un échantillonnage de sa population relativement à quelques caractéristiques en relation directe avec le sujet: usager des transports en commun ou non, usager régulier ou occasionnel, avec ou sans auto, employeur ou employé, etc. La population à échantillonner ayant été définie de cette façon, nombreux sont pratiquement les plans susceptibles d'être mis en œuvre pour fournir un échantillon statistiquement valide. Sans compter que pour chacun d'eux, la façon de calculer la valeur estimée d'une caractéristique n'est pas nécessairement unique. Si notre choix devait se porter sur l'un de ces plans et, avec lui, sur l'un des procédés applicables au calcul de la valeur estimée, nous obtiendrions d'un coup, et les mesures des caractéristiques de la population en général, et les mesures des caractéristiques de la population des usagers du réseau. Et l'avantage de posséder de telles données relatives aux deux populations, et obtenues à une même époque, est manifeste quand on songe à l'objet de cette étude dont le but ultime doit être de mesurer des rapports de toutes sortes entre caractéristiques des deux populations.

Pour valoir, le plan d'échantillonnage ainsi conçu devrait être étendu à toutes les circonscriptions, telles que Montréal, placées dans le cas de fournir des usagers au réseau; à moins de ne se limiter systématiquement à une seule circonscription ou à un groupe de celles-ci, quitte à modifier la question initialement posée pour la restreindre, par exemple, à cette autre: «qui, résidant à Montréal, prend le tramway?» Rien n'autorisant la chose, et bien au contraire, on se trouve donc ramené au point d'englober toutes les circonscriptions fournissant des usagers.

Or, on imagine bien que les usagers viennent en fait d'un peu partout, ne serait-ce qu'en comptant avec les touristes, si bien que

le plan devient manifestement impraticable, à moins qu'on ne songe à le modifier quelque peu. Par exemple, on pourrait l'appliquer tel quel à l'Île de Montréal — ce qui ne serait d'ailleurs pas une mince besogne — puis il serait à prolonger par un échantillonnage de cette partie de la population des usagers en provenance de toute circonscription exclue par le plan précédent. Ceci entraînerait un contrôle de toutes les voies d'accès à l'Île — gares, ponts — sans compter que le réseau, par quelques-unes de ses lignes, sort de l'Île de Montréal. En outre, cette population qui passe ainsi sur l'Île, et qui paraît à première vue formée de non-résidents, pourrait très bien être en partie constituée d'éléments déjà englobés dans la première partie du plan. Toutes ces considérations expliquent pourquoi il a pu paraître peu praticable de donner suite à un tel plan, dont le moindre désavantage est d'être composite, tandis que le plus grave est la disproportion des moyens à mettre en œuvre pour le réaliser, avec le coût prohibitif qui en est la conséquence immédiate.

On imagine facilement que d'autres plans ont dû se présenter comme possibles, cependant qu'il sera coupé court à l'exposé de ceux qui ont dû être rejetés après avoir été mis aux prises avec les difficultés d'ordre administratif et économique, pour passer tout de suite à celui qui a été, de fait, mis en œuvre. Ce plan, avec ses détails d'exécution, n'est pas mis de l'avant avec la prétention qu'il soit le meilleur, encore moins est-il présenté comme un prototype du genre: il est simplement celui qui a paru pouvoir résister le mieux à l'épreuve des circonstances, sans plus. Comme on doit s'y attendre, ses points faibles ont eu l'occasion de se déceler en cours d'exécution, et s'il était à reprendre, il reviendrait avec les améliorations ou modifications dictées par son usage.

* * *

À commencer par là, la population totale dont les caractéristiques sont à estimer par voie d'échantillonnage est celle des usagers du réseau de la Commission de Transport de Montréal — de tous les usagers, et des usagers seulement. Si facile d'application que paraisse cette définition, elle appelle cependant des précisions. Tout d'abord, la population en question est arrêtée, par

définition, à l'état où elle se trouvait le jour même où a été prélevé l'échantillon. Ensuite, qu'entendrons-nous exactement par un usager? Sous réserve de ce qui suit, deviendra un usager tout individu qui sera monté à bord d'un véhicule de la Commission de Transport de Montréal le jour de l'enquête — autobus et trolleybus aussi bien que tramway proprement dit. La dite réserve a pour objet d'exclure sciemment les enfants voyageant à titre gratuit en vertu de leur âge, les personnes transportées uniquement à bord des véhicules circulant en vertu d'un contrat de louage, les personnes en uniforme voyageant régulièrement sans produire de titre de transport, les personnes munies d'un permis de circulation et les personnes voyageant frauduleusement. Et le jour de l'enquête doit être entendu en tenant compte qu'il existe chaque jour une courte durée pendant laquelle tous les véhicules sont retirés du service régulier, rentrant au garage ou au dépôt, et le jour de l'enquête commencera à l'arrivée sur place du premier véhicule à prendre son service sur le réseau, au jour dit, soit à un instant situé quelque part entre quatre et cinq heures du matin, et se terminera au départ du dernier véhicule à être retiré du réseau, c'est-à-dire vers deux heures le lendemain matin.

Les conditions d'application du plan nous amèneront, non seulement à faire l'énumération des personnes montées à bord de certains véhicules, mais aussi à considérer un ordre parmi ces personnes, à savoir l'ordre suivant lequel elles auront eu accès au véhicule. Et la question à régler est de définir quand et comment tel fait, celui d'être monté à bord du véhicule, devra être réputé accompli. L'idée première avait été de se fixer comme norme le fait de poser le pied sur le marchepied. Or, dans certaines circonstances bien connues du public voyageur, et c'est le cas du tramway que nous avons en vue, ce n'est pas un pied qui se pose sur le dit marchepied, ce sont vingt pieds qui se posent à la fois, et d'ailleurs serait-on bien embarrassé pour dire ce qui se passe, alors qu'une partie de la grappe humaine est portée directement à l'intérieur du véhicule sans avoir utilisé de marchepied du tout. Une observation attentive du phénomène, laquelle a précédé bien entendu la mise en route du plan, a eu comme conséquence de faire rejeter la première définition, et la suivante lui a été substituée. Ayant toujours en vue qu'il s'agit d'établir un ordre parmi les

voyageurs par l'application d'une règle certaine, un individu sera réputé avoir eu accès au véhicule à l'instant même où il se mettra en règle avec le préposé à la perception, quelle que soit la façon de se mettre en règle: ticket, comptant, remise du titre de correspondance ou présentation de celui-ci.

Mais ce n'est pas tout. Sans doute, aucun usager n'échappera, comme tel, à la classification; mais, par contre, un individu donné, un usager, peut n'emprunter qu'un seul véhicule au cours de la journée, tandis qu'un autre, un usager encore, peut en emprunter plusieurs, et rien ne les distingue l'un de l'autre alors qu'ils se mettent respectivement en règle avec un préposé à la perception. Et puisque c'est au moment même où il a accès à un véhicule qu'un individu quelconque court le risque d'être pris dans l'échantillon, le second usager a manifestement une probabilité plus grande que le premier de tomber dans le dit échantillon.

Tout ceci nous impose de formuler de nouvelles précisions. Nous dirons que tout individu qui se met en règle avec un préposé, et au moment même où il le fait, devient de ce fait même un *voyageur*, et tout individu distinct qui est devenu un voyageur au cours de la journée, et quel que soit le nombre de fois qu'il ait pu le devenir, est un *usager*. Dans ces conditions, la population que nous nous proposons d'échantillonner n'est plus celle des usagers, comme nous voulions paraître nous y attendre dès le début; mais elle sera bien plutôt celle des voyageurs, lesquels remplissent seuls cette condition essentielle pour qu'un échantillon soit statistiquement valide, à savoir que tout élément de la population doit avoir la même probabilité que tout autre de se trouver pris dans l'échantillon. Il nous restera ensuite à passer de cette population des voyageurs, adoptée à titre transitoire, à celle des usagers, qui se trouve être en définitive la véritable population à l'étude. C'est ainsi que la première population peut compter quelque chose comme un million et demi de voyageurs, tandis que la seconde, d'après les résultats de l'enquête, semble compter à peine cinq cent mille usagers.

* * *

Les éléments de la population des voyageurs étant donc définis sans équivoque, il s'agit maintenant de tirer un échantillon des dits

éléments, nous proposant de passer par la suite à une estimation des caractéristiques de la population des usagers. Avant que de nous mettre en peine du nombre de ces éléments qu'il faudra dans l'échantillon — lequel nombre devant assurer un compromis entre le coût de l'opération qu'il faudra mettre en œuvre pour l'obtenir, les moyens disponibles et une précision des estimations compatible avec un risque administratif acceptable—, passons à la façon pratique de prélever le dit échantillon. Il sera obtenu en deux stades. On tirera d'abord un certain nombre de véhicules, pris dans une population de véhicules qu'il faudra définir, et ceux-là formeront ensemble un premier échantillon: l'échantillon des véhicules. Puis on prélèvera à l'intérieur de chacun de ces derniers véhicules un sous-échantillon de voyageurs, et ces sous-échantillons formeront ensemble un second échantillon: ce sera l'échantillon définitif, celui des voyageurs.

Pratiquement, imaginons donc de placer un interviewer à l'intérieur de l'un de ces véhicules désignés dans l'échantillon, et il s'agira pour lui de prélever à même l'ensemble des voyageurs devant avoir accès au dit véhicule un sous-échantillon de voyageurs qu'il faudra interviewer sur place. Et puisque les voyageurs se succèdent dans un même véhicule suivant un ordre bien déterminé qui découle de nos définitions, il sera facile de mettre à profit cet ordre pour faire jouer le hasard dans l'opération de prélèvement du sous-échantillon. Pour cela, l'interviewer sera mis en possession d'une liste de numéros, indiquant chacun un rang bien déterminé dans une suite, et ce seront les voyageurs qui occuperont ce rang en défilant à leur tour devant le préposé qui seront pris dans le sous-échantillon. Pour chaque véhicule, les dits numéros d'ordre appartiennent à une suite prédéterminée, et ils forment une progression arithmétique ou, comme l'on dit en langage courant, ils se succèdent à intervalle régulier. Admettons à titre d'exemple, que cet intervalle ait été fixé à 12 sur le véhicule en question, et qu'il ait d'ailleurs été fixé de cette façon pour des raisons qu'il s'agirait de valider. Cet intervalle étant donné, admettons encore que nous ayons fait choix, à partir de là, d'un entier compris entre 1 et 12, et soit 7 l'entier ainsi choisi. Nous retiendrons dans le sous-échantillon de ce véhicule les voyageurs qui, s'étant mis en règle avec le préposé, auront reçu de ce fait les numéros d'ordre 7, 19, 31, 43, etc.

Retenons encore un instant ce dernier exemple. Lorsque nous disons que l'entier 7 — qui se trouve être le numéro du premier voyageur du véhicule à être interviewé — a été «choisi» parmi les entiers compris entre 1 et 12, il s'agit de s'entendre sur l'idée de choix qui se trouve impliquée en pareille circonstance. En fait, point n'est question d'en décider soi-même, c'est-à-dire de se substituer au hasard, en ce sens que l'on n'a pas qualité pour répondre par la pensée à cette condition primordiale ici, à savoir que tout entier compris entre 1 et 12 doit avoir la même probabilité que tout autre d'être choisi. Pour arriver pratiquement à ce résultat, le plus sûr moyen est d'utiliser une de ces tables où les nombres, et parties de nombres, se succèdent tout comme s'ils avaient été produits par un jeu de hasard.

Maintenant, sur quoi se basera-t-on pour fixer l'intervalle que nous avons laissé entendre devoir être attaché à chaque véhicule en particulier, tel celui de 12 que nous avons à l'instant rencontré? Bien entendu rentre-t-il dans un système de probabilités avec lequel il doit, de toute nécessité, se maintenir en règle; cependant qu'un autre point de vue ne manque pas de se manifester: à savoir que les conditions de son application à bord du véhicule dépendent étroitement des circonstances particulières de la circulation rencontrées par celui-ci. Imaginons tout de suite le cas de deux véhicules distincts tombés l'un et l'autre dans l'échantillon des véhicules et se trouvant dans cette situation pour une durée d'une heure, l'un donnant accès à 40 voyageurs pendant ce temps, tandis que l'autre en recevrait 400 dans les mêmes conditions — ce qui correspond à une situation très vraisemblable en pratique, puisque le nombre annoncé, tout en étant lié à la capacité du véhicule en voyageurs, tient compte du renouvellement de ceux-ci. On saisit sans peine qu'il serait facile d'observer un intervalle de 4 dans le premier véhicule, ce qui donnerait un sous-échantillon de 10 voyageurs à interroger, laissant tout compte fait une moyenne de six minutes pour chacun d'eux (on remarquera d'ailleurs qu'il est bon de ne pas exagérer la valeur de la dite moyenne, ne serait-ce que pour tenir compte du voyageur qui monte et descend respectivement à deux points d'arrêt consécutifs). Manifestement, le même intervalle de 4 ne saurait s'appliquer au second véhicule sous peine de conduire à une situation inextricable, en particulier

celle d'avoir à aborder 100 voyageurs, disposant de moins d'une minute pour chacun d'eux, et dans des conditions d'encombrement et de va-et-vient que l'on imagine.

Visiblement, il faudra donc renoncer à cette ligne de conduite qui paraît de prime abord être la plus logique, et peut-être même la seule normale, et qui consisterait à échantillonner dans les divers véhicules proportionnellement à leur contenu en voyageurs: un échantillon de 4 voyageurs pris parmi 40 voyageurs, un échantillon de 40 voyageurs pris parmi 400 voyageurs. Tout au contraire, voit-on l'avantage qu'il y aurait si l'on pouvait créer une situation amenant à interroger le même nombre de voyageurs à bord de tous les véhicules, quels qu'ils soient, quelle que soit leur condition d'encombrement. Pour reprendre le dernier exemple à la lumière de ce qui vient d'être dit, imaginons maintenant de fixer à 10 le nombre des voyageurs à interroger aussi bien sur le premier véhicule que sur le second. Pour le premier, rien ne sera changé sur les conditions rencontrées tout à l'heure — intervalle et contenu de l'échantillon—, et elles semblaient normales alors. Pour le second véhicule, l'intervalle se trouve porté de ce fait à 40: c'est-à-dire qu'après avoir interrogé un voyageur, on aura donc le loisir d'en laisser se succéder quarante autres avant d'avoir à interroger le suivant. Il est manifeste que les conditions à bord de ce véhicule, d'impraticables qu'elles étaient tout à l'heure, sont devenues des plus raisonnables. Tandis que l'intervalle était fixé, et servait à déterminer le contenu de l'échantillon, c'est le contraire qui arrive maintenant. Et c'est à ce mode d'échantillonnage à bord des véhicules que nous nous arrêterons en définitive.

Évidemment, quelque chose devra être fait du côté des probabilités, car le voyageur du véhicule relativement peu occupé aura une plus grande «chance» d'être pris dans l'échantillon de son véhicule que n'en aura le voyageur de l'autre véhicule d'être pris dans l'échantillon du sien. Et cette inégalité dans les probabilités à l'intérieur des véhicules se propagera à l'échantillon définitif des voyageurs si les véhicules, de leur côté, doivent être appelés dans leur échantillon avec des probabilités égales. La conséquence immédiate est que les véhicules, de toute nécessité, devront être tirés dans leur échantillon avec des probabilités inégales. Mais, que devront être celles-ci?

Imaginons donc que chaque véhicule se voit attribuer une probabilité propre d'être tiré dans l'échantillon des véhicules, et cherchons ce que devra être cette dernière. Revenons encore au cas des deux véhicules qui viennent encore à l'instant de nous servir d'exemple: ils sont donc réputés devoir renfermer 40 et 400 voyageurs respectivement, et l'on doit prélever sur chacun d'eux un échantillon de voyageurs de même contenu, soit 10 voyageurs. D'une part, décidons de choisir arbitrairement un entier tel que 5,000, et convenons ensuite d'attribuer à nos deux véhicules les probabilités respectives $\frac{40}{5,000}$ et $\frac{400}{5,000}$ d'être tirés dans l'échantillon des véhicules. D'une part, il est certain que la probabilité que possède un voyageur quelconque du premier véhicule de se trouver pris dans l'échantillon des voyageurs de son véhicule est égale à $\frac{1}{4}$ ou $\frac{10}{40}$ (puisque l'intervalle est de 4 et que le sort d'un voyageur de numéro quelconque est lié à l'entier compris entre 1 et 4 que l'on tire comme nous l'avons dit), tandis que la probabilité d'un voyageur quelconque du deuxième véhicule de se trouver pris dans l'échantillon de son véhicule est égale à $\frac{1}{40}$ ou $\frac{10}{400}$ (l'intervalle est ici de 40). D'une façon générale, pour qu'un voyageur quelconque se trouve pris dans l'échantillon des voyageurs, il faut nécessairement que soient remplies les deux conditions suivantes: il faut tout d'abord que son véhicule sorte dans l'échantillon des véhicules puis, cela s'étant produit, il faut encore que lui-même sorte dans l'échantillon des voyageurs de son véhicule. Ce qui revient à dire, avec les probabilités en jeu dans notre exemple, et d'après un théorème très connu, que la probabilité d'un voyageur du premier véhicule et la probabilité d'un voyageur du second sont respectivement égales aux produits

$$(1) \qquad \frac{40}{5,000} \times \frac{10}{40} = \frac{10}{5,000} \quad \text{et} \quad \frac{400}{5,000} \times \frac{10}{400} = \frac{10}{5,000}.$$

Et ces deux probabilités sont égales. L'entier 10 est égal au nombre constant de voyageurs pris dans l'échantillon à bord de chaque véhicule, les entiers 40 et 400 sont respectivement égaux aux contenus en voyageurs des deux véhicules et l'entier 5,000 est une constante. On généralisera facilement.

Ainsi donc, si l'on veut, comme il se doit, que tous les voyageurs aient la même probabilité d'être pris dans l'échantillon définitif, et si l'on veut, pour la commodité, que le nombre des voyageurs à interroger à bord des divers véhicules soit constant, il faudra attribuer aux divers véhicules des probabilités *proportionnelles à leur contenu en voyageurs*.

Tout ceci est fort bien, pourvu que nous ayons la possibilité de prévoir le nombre des voyageurs que recevra chaque véhicule au jour de l'enquête: de là, il n'y aura qu'un pas pour être en mesure d'attribuer aux véhicules leurs probabilités respectives. Mais qu'adviendra-t-il si nous nous trompons dans nos prévisions, comme il ne manquera certainement pas d'arriver? Qu'adviendra-t-il des relations entre probabilités, telles que nous en avons un exemple en (1) et alors qu'elles nous donnaient pleine satisfaction, si quelque chose vient à y être changé?

Pour examiner la question, revenons une fois de plus à l'exemple des deux véhicules devant recevoir 40 et 400 voyageurs respectivement. On devait prélever un échantillon de même contenu dans chacun d'eux, soit 10 voyageurs, et les deux relations (1) montraient que tous les voyageurs, peu importe le véhicule, avaient la même probabilité d'être pris dans l'échantillon. Supposons maintenant que le premier véhicule, au lieu de recevoir 40 voyageurs comme il était prévu, vienne à en recevoir effectivement 80 au jour de l'enquête. Or, il est entendu que l'intervalle attaché à ce véhicule lui avait été attribué à l'avance et en nous basant sur nos prévisions quant à ses conditions d'occupation: nous escomptions 40 voyageurs, nous devions en prélever 10, donc l'intervalle se trouvait fixé à 4. Et au jour de l'enquête, en dépit des conditions qui ne sont pas celles que nous avions prévues, c'est bien avec cet intervalle de 4 que nous aurons à travailler. Autrement, comment pourrions-nous faire, puisque rien ne nous avertit de ce qui va se passer: dès le début interrogeant les voyageurs de quatre en quatre comme ils se présentent, sauf à constater, à partir d'un certain moment, et au fur et à mesure de la marche des événements que ceux-ci ne continuent pas à se produire comme on les avait prévus. Si bien qu'en appliquant l'intervalle prescrit tout au long de l'opération à bord du véhicule, nous nous trouvons en fin de compte avec 20 voyageurs dans l'échantillon au lieu de 10 (bien entendu,

l'idée insoutenable de couper court aux opérations, sitôt obtenu le contingent de 10 voyageurs, ne nous est-elle jamais venue). Si maintenant nous reprenons la première des deux relations (1) pour la modifier et la mettre en accord avec les circonstances nouvelles, cette relation qui nous donnait la probabilité prévue pour un voyageur quelconque du premier véhicule de tomber dans l'échantillon définitif devra nous donner maintenant la probabilité effectivement encourue par le voyageur au jour de l'enquête, soit à trouver la relation

$$\frac{40}{5{,}000} \times \frac{20}{80} = \frac{10}{5{,}000}.$$

Et celle-ci nous montre que la probabilité n'a pas changé. On généralisera facilement.

Ainsi, la probabilité de tomber dans l'échantillon pour un voyageur quelconque restera constante et la même pour tous, en dépit des erreurs que nous pourrons commettre dans nos estimations du contenu des véhicules: en d'autres termes, l'échantillon se *rectifiera* de lui-même — et la remarque est importante.

Bien entendu, cela ne veut pas dire que nous pourrons prendre prétexte de ce fait pour nous permettre d'attribuer inconsidérément leurs probabilités respectives aux divers véhicules. Il demeure que la probabilité attachée à un véhicule commande l'intervalle qui sera utilisé dans les opérations sur le terrain, et ces opérations se rapprocheront d'autant plus des conditions idéales qui nous ont amené à bâtir notre système que nos estimations seront plus près des conditions d'occupation que l'on rencontrera. D'ailleurs, et pour des raisons bien différentes de ces dernières, puisqu'elles nous sont dictées par le calcul, devons-nous encore abonder dans le même sens en faveur de la précision dans les prévisions relatives au degré d'occupation des véhicules. Et ce sera cette fois dans le but de diminuer autant que possible les erreurs des estimations qui seront tirées de l'échantillon à l'égard des caractéristiques de la population des usagers.

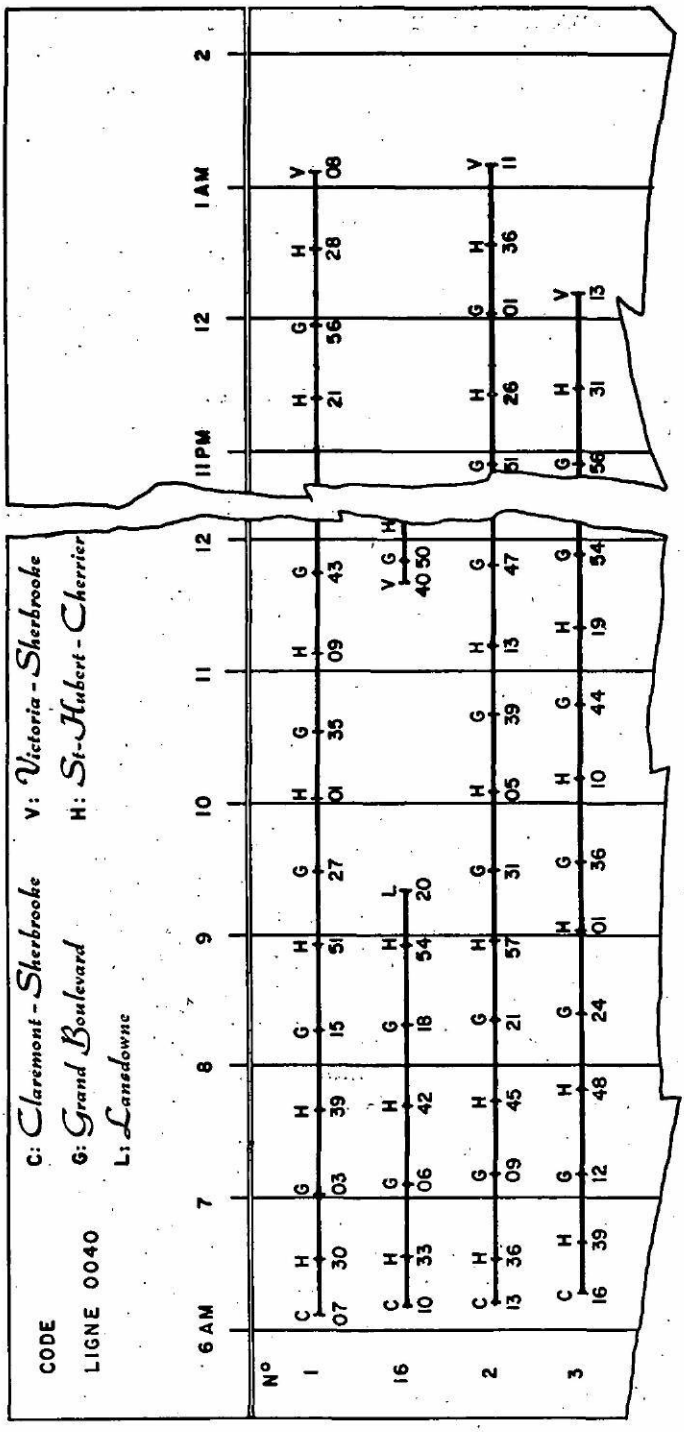
* * *

Dans tout ceci, nous avons passablement tiré parti de la notion de «véhicule», et peut-être l'avons-nous fait trop longtemps, puisqu'elle n'a pas encore reçu de définition. Comme il se doit,

nous allons préciser ce que nous entendions et ce qui sera entendu par là dans la suite, tout comme nous avons dû définir notre voyageur, en temps et lieu et sans équivoque, en tant qu'élément de la population à échantillonner. Certaines voitures de la Commission de Transport prennent la route de grand matin et rentrent au dépôt ou au garage tard le soir, ou même vers 1 h. ou 2 h. du matin. D'autres ne sont en service qu'aux heures d'affluence, soit deux heures consécutives le matin, ou guère plus, puis elles sont retirées jusqu'au moment de revenir prendre un court service pendant les heures d'affluence de la soirée. Or, un tel véhicule, considéré depuis son heure de départ pour le service jusqu'à l'heure où il le quitte, ne répondrait ordinairement pas à notre attente, en ce sens que nous ne serions pas maître du nombre des voyageurs qu'il nous accommoderait d'y trouver pour les fins de notre échantillonnage.

Avec chacun de ces véhicules, indivisibles en tant que matériel roulant, nous en ferons plusieurs que nous imaginerons se succéder l'un à l'autre, du moins aurons-nous à le faire ordinairement. Pour éviter toute confusion, ces véhicules que nous aurons détachés par l'effet d'une définition, nous les appellerons dorénavant des *véhicules-unités*. Par définition, un véhicule-unité prendra naissance d'un véhicule circulant régulièrement sur le réseau au moment où le dit véhicule passera en un point déterminé de son parcours, et il prendra fin lors du passage, à un moment subséquent, au même point ou en un autre point également bien déterminé; le tout, bien entendu, considéré au cours de la même journée. Les dits points du parcours seront choisis de façon à accorder une plus ou moins longue durée au véhicule-unité, tenant compte des conditions de la circulation du moment et se faisant une image des circonstances dans lesquelles devrait s'effectuer le travail de l'interviewer susceptible de se trouver posté à bord du véhicule. Et le même soin sera apporté à la définition de chacun des véhicules-unités, chacun d'eux courant le risque d'être pris dans l'échantillon. Ainsi sectionné en véhicules-unités, l'ensemble du matériel roulant en service régulier sur le réseau de la C. T. M. s'est trouvé réparti en quelque 15,000 véhicules-unités ainsi définis, et aucun véhicule à bord duquel un usager était susceptible d'avoir accès n'a dû échapper à la nomenclature. Ceci entendu avec

Figure 1


CODE
LIGNE 0040
C: Claremont - Sherbrooke
G: Grand Boulevard
L: Lanedowne
V: Victoria - Sherbrooke
H: St-Hubert - Cherrier
6 AM
7
8
9
10
11
12
11 PM
12
1 AM
2
N°
1
16
2
3

la restriction déjà signalée, à savoir que les véhicules loués en vertu d'un contrat permanent ou occasionnel se trouvent exclus tant qu'ils circulent dans ces conditions, et sont également exclus les véhicules spéciaux effectuant spécifiquement leur circuit pour l'usage des touristes et promeneurs.

Nous aurons donc une population de véhicules-unités, une fois que chacun aura été défini, et chacun d'eux recevra un certain groupe de voyageurs, et tout voyageur montera à bord de l'un de ces véhicules-unités. Quitte à nous répéter, disons que l'échantillon des véhicules doit être tiré de cette population d'après un système de probabilités inégales prévoyant pour chacun des véhicules-unités l'attribution d'une probabilité propre suivant laquelle il sera susceptible de tomber dans l'échantillon. De plus, il a été précisé que ces probabilités doivent, autant que possible, être respectivement proportionnelles aux contenus en voyageurs. En fait, nous aurons donc à obtenir par le calcul l'estimation du contenu de chacun de nos 15,000 véhicules-unités. Et c'est l'opération, ou plutôt la suite des opérations qu'il va falloir effectuer en ce sens, que nous nous proposons d'examiner présentement.

* * *

Tout d'abord, nous allons transformer l'ensemble des tableaux horaires de la Commission de Transport (quelque 120 lignes, et jusqu'à 1,500 véhicules distincts en service simultanément) en une représentation graphique. Ceci, dans le but d'absorber dans un dispositif commode des calculs très abondants, et dans le but aussi de faciliter par une perception visuelle d'un ensemble le choix et la détermination des véhicules-unités. La figure 1 est un fac-similé en raccourci de cette représentation, juste avant que ne viennent prendre place les données numériques et les calculs devant aboutir à la détermination des véhicules-unités et à l'estimation de leur contenu.

La figure montre ce qu'est devenu, dans ces conditions, le tout début du tableau horaire relatif à la ligne numéro 4 (N.D.G.). On voit que le premier véhicule mentionné à l'horaire porte le numéro 1 de service interne, qu'il vient prendre son service sur la ligne à 6.07 heures A.M., au point d'arrêt situé au coin des rues

Claremont/Sherbrooke, qu'il passe immédiatement après au point d'arrêt Saint-Hubert/Cherrier, où il est réputé passer à 6.30 heures, puis encore qu'il atteint consécutivement le point d'arrêt Grand-Boulevard, où il est réputé passer à 7.03 heures A.M.; et ainsi de suite jusqu'à ce qu'il soit retiré du service. En fait, il se trouve effectuer ainsi seize circuits complets en repassant pour la dernière fois au point d'arrêt Grand-Boulevard, puis il quitte définitivement son service quelques minutes plus tard, alors qu'il atteint le point d'arrêt Victoria/Sherbrooke. Le second véhicule inscrit au tableau horaire porte le numéro 16 pour le service interne: il prend son service à 6.10 heures A.M., au point d'arrêt Claremont/Sherbrooke, il passe immédiatement après au point d'arrêt Saint-Hubert/Cherrier à 6.33 heures et, après avoir effectué deux circuits complets en repassant par ce même point d'arrêt, il se dirige encore une fois vers l'ouest de la ville pour atteindre le point d'arrêt Lansdowne, où il quitte son service sur la ligne à 9.20 A.M. Le même véhicule revient prendre du service sur la ligne à 11.40 heures A.M., au point d'arrêt Victoria/Sherbrooke, atteignant immédiatement après le point d'arrêt Grand-Boulevard, où il est réputé passer à 11.50 heures et, après avoir effectué sept circuits complets, repassant en ce même point à 8.34 heures P.M., il est retiré définitivement du service à 8.43 heures P.M., lorsqu'il repasse au point d'arrêt Victoria/Sherbrooke. Et ainsi de suite pour les autres véhicules en service régulier sur la ligne numéro 4 et, d'une façon analogue pour les autres lignes du réseau, jusqu'à ce que l'on ait atteint le total déjà signalé de quelque 1,500 véhicules distincts.

Maintenant que les horaires viennent de recevoir leur forme graphique, l'opération venant à la suite est relative à l'estimation du contenu en voyageurs des diverses portions de véhicules appelées à constituer nos véhicules-unités, et c'est sur le graphique même que les calculs y relatifs recevront leur traitement matériel. Voici, dans cet ordre d'idées, les renseignements que possède la Commission de Transport et qui ont été mis en œuvre pour nous conduire plus ou moins directement à notre but. Elle connaît pour chaque ligne, et au jour le jour, d'une part la recette ou, plus exactement, le nombre des perceptions effectuées à bord de l'ensemble des véhicules de la ligne, d'autre part elle connaît le nombre des titres de correspondance remis aux préposés et couvrant le même ensemble

de véhicules. C'est donc un total unique de perceptions et un total unique de titres de correspondance que l'on connaît pour chaque ligne, sans qu'il soit aucunement gardé de traces de l'apport individuel de chaque véhicule composant la ligne. Et ces renseignements, essentiellement comptables, sont, avec les tableaux horaires, les seuls n'ayant pas un caractère d'estimation et sur lesquels on pouvait s'appuyer.

Tout d'abord, les états en question ne donnent qu'une partie du compte des voyageurs, car on a omis d'une façon constante cette catégorie de voyageurs qui trouve accès aux véhicules sur simple présentation du titre de correspondance, c'est-à-dire sans remise de celui-ci. Et ici on devra avoir recours aux estimations pour faire entrer en ligne de compte cet apport en voyageurs.

À diverses reprises ont été effectués quelques comptages complets à bord du matériel roulant de certaines lignes afin d'exercer un contrôle sur la dite catégorie de voyageurs, et on en a retenu une estimation de l'importance de cette dernière par rapport à la catégorie des voyageurs faisant remise du titre de correspondance. La ligne numéro 27, par exemple, a été choisie pour être placée dans cet état de contrôle en raison de l'importance qu'y prend la catégorie de voyageurs qui nous occupe — importance que justifie d'ailleurs le tracé même de la ligne, qui lui en fait relier un grand nombre d'autres ayant une direction commune, tandis qu'elle-même ne dessert guère de points de conséquence. Pour les autres lignes, où le comptage direct n'a pas eu lieu, il a fallu s'en remettre au personnel du réseau compétent en la matière et apte à fournir une évaluation du rapport qui nous intéresse, c'est-à-dire du rapport entre les nombres des voyageurs des deux catégories. Ce rapport peut être très variable: tandis qu'on a le cas extrême qu'on vient de citer, il se trouve certaines lignes n'ayant que peu ou pas de raison de transporter des voyageurs «en transit», et l'on a d'autre part de nombreux cas intermédiaires.

Prenons un exemple et, pour cela, retournons à la ligne qui nous a servi à illustrer notre représentation graphique, soit la ligne numéro 4. Pour une certaine date, on nous fournira, par exemple, l'état suivant (en fait l'état couvrira les lignes numéros 4 et 4A, sans les dissocier).

Tableau I

Voyageurs ayant donné lieu à perception directe	27,584 (1)
Voyageurs ayant remis un titre de correspondance	9,760 (2)
Voyageurs ayant présenté un titre de correspondance (45% catégorie précédente)	4,392 (3)
Total sur la ligne 4-4A	41,736

Les résultats (1) et (2) provenant de comptages varient d'un jour à l'autre et sont connus pour tous les jours de l'année; le pourcentage appliqué en (3) provenant d'une estimation ou de comptages occasionnels est une constante pour une certaine durée.

Le total est relativement constant d'un jour ouvrable à l'autre au cours de plusieurs semaines consécutives — sauf le cas d'intempéries peu communes —, et les variations ne se font sentir que lentement au cours de l'année. En prenant simplement la moyenne pour quelques jours ouvrables précédant immédiatement le jour fixé pour l'enquête, celui-ci se plaçant à une époque où aucun changement brusque n'a lieu d'être prévu, on se trouvera en présence d'une estimation raisonnable du nombre des voyageurs, sur la ligne et pour le dit jour. Estimation raisonnable en ce sens que rien ne nous oblige à exagérer dans le sens d'une précision trop scrupuleuse, puisque l'échantillon se rectifie de lui-même, comme nous l'avons montré, alors même qu'une erreur d'estimation grave vient à se produire. Il suffit en somme de tenter un honnête effort visant à se rapprocher du système de probabilités que le calcul montre devoir diminuer autant que possible les erreurs d'estimation des caractéristiques de la population, sans plus.

Une autre estimation devant concourir au même but, celui de détacher les véhicules-unités et d'en estimer le contenu, a encore été fournie par le personnel supérieur de la Commission de Transport de Montréal. Cette estimation devait répondre à la question suivante: en gardant en vue la vitesse de renouvellement des voyageurs à bord des véhicules, sans confondre celle-ci avec la notion d'encombrement ou de vacuité apparente du véhicule, estimer, à l'aide d'un cœfficient, le nombre relatif des voyageurs se succédant à l'intérieur de véhicules types circulant sur chacune des lignes du réseau. Il était entendu que ce cœfficient, dit *coefficient d'occupation* en la circonstance, devait varier entre une borne inférieure de 1 et une borne supérieure de 10 pour chaque ligne,

et qu'une valeur devait lui être attribuée d'heure en heure tout au cours de la journée.

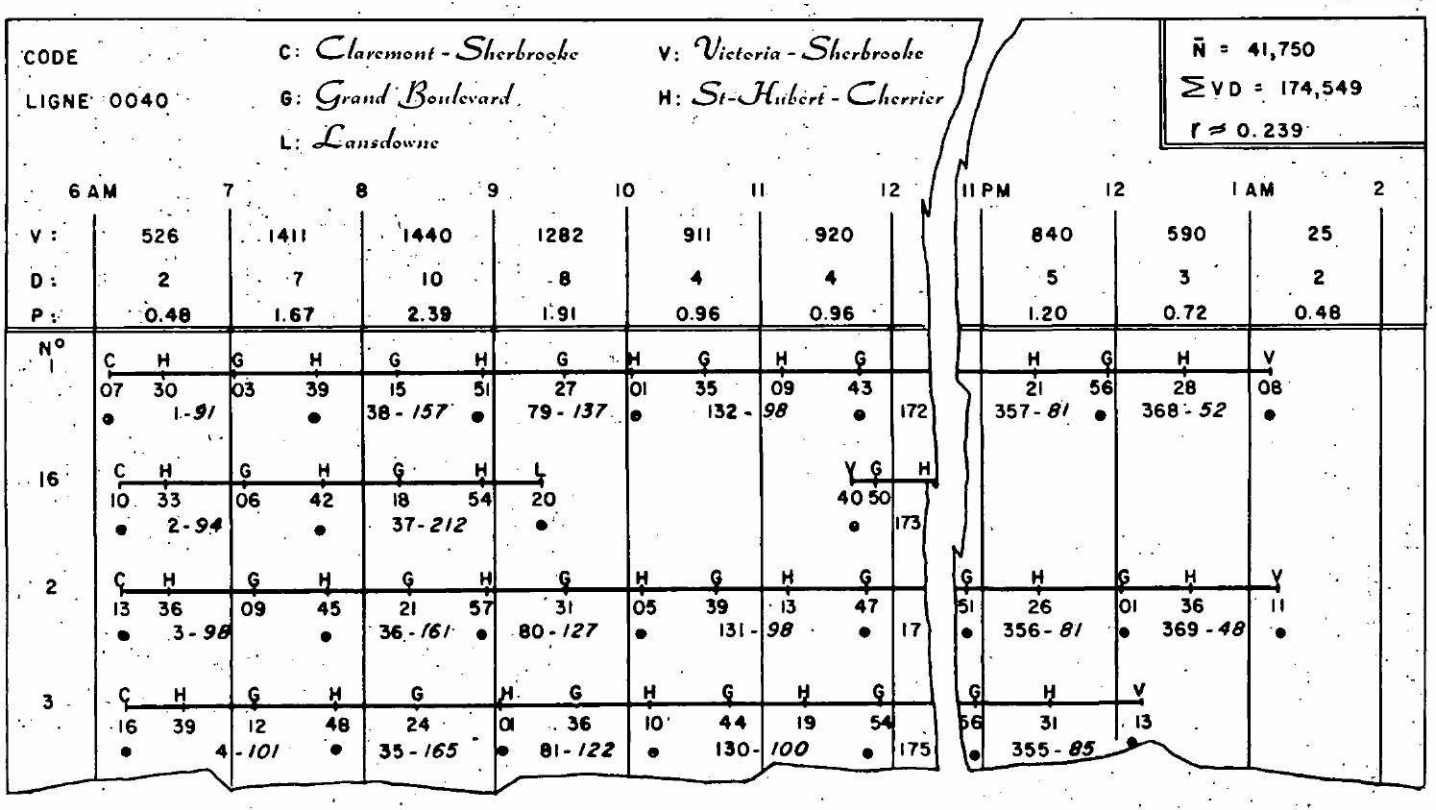
On peut remarquer en passant que le coefficient en question, du fait qu'il se trouve détaché de tout système d'unités, n'est pas susceptible d'être transporté tel quel d'une ligne sur une autre ligne, tout au moins dans le sens que voici. Par exemple, un coefficient de 10 attribué pour une certaine heure aux véhicules de la ligne 27 et un coefficient de 10 attribué, à la même heure ou à une autre heure, aux véhicules de la ligne 65 ne permettent pas un rapprochement immédiat quant au nombre des voyageurs, par véhicule, de l'une et l'autre lignes. Les deux coefficients signifient seulement que les voyageurs se sont renouvelés à la vitesse maximum dont sont respectivement capables les véhicules circulant sur les deux lignes, compte tenu, non pas de leur capacité, mais bien plutôt des habitudes du public voyageur. Et le nombre des voyageurs montés à bord du véhicule de la ligne 27 est probablement très supérieur au nombre des voyageurs du véhicule de la ligne 65, alors qu'ici le voyageur type effectue ordinairement un long parcours sans descendre, par opposition au voyageur de l'autre véhicule qui n'y séjourne souvent qu'entre deux points d'intersection assez rapprochés.

* * *

Nous allons revenir à notre représentation graphique des tableaux horaires et l'utiliser comme nous l'avons dit pour mettre commodément en œuvre le système des données comptables et des données estimées qu'on vient d'examiner, et duquel nous voulons faire sortir les définitions de nos 15,000 véhicules-unités ainsi que l'estimation de leurs contenus en voyageurs. La figure 1 nous donnait tout à l'heure un aperçu des horaires sur la ligne numéro 4, et nous allons reprendre le même graphique laissé dans cet état pour l'amener à l'état de la figure 2, alors qu'y ont pris place les calculs relatifs à la suite des opérations.

C'est ainsi que l'on trouve tout d'abord (figure 2,) une série de valeurs de la variable V. Ces valeurs sont respectivement 526, 1411, 1440, etc., et correspondent de proche en proche aux différentes heures où les véhicules sont en service sur la ligne. Elles sont obtenues en calculant pour chaque heure la somme des

Figure 2


CODE
LIGNE 0040
C: Claremont - Sherbrooke
G: Grand Boulevard
L: Lansdowne
V: Victoria - Sherbrooke
H: St-Hubert - Cherrier
N̄ = 41,750
ΣVD = 174,549
r ≈ 0.239
6 AM 7 8 9 10 11 12 11 PM 12 1 AM 2
V : 526 1411 1440 1282 911 920 840 590 25
D : 2 7 10 8 4 4 5 3 2
P : 0.48 1.67 2.39 1.91 0.96 0.96 1.20 0.72 0.48
N°
1
16
2
3

durées, en minutes, pendant lesquelles les diverses voitures de la ligne sont maintenues en service. Par exemple, entre 6 heures et 7 heures A.M., le véhicule numéro 1 étant en service pendant 53 minutes, le véhicule numéro 16 l'étant pendant 50 minutes, et ainsi de suite jusqu'au dernier véhicule en service à cette heure, on a trouvé 526 en calculant la somme dont les termes sont 53, 50, . . . et ainsi de suite jusqu'au terme correspondant au dernier véhicule. Si l'on veut s'exprimer ainsi, on dira qu'il s'agit là de «voitures-minutes».

On peut encore voir la suite des valeurs d'une autre variable, celle-ci représentée par la lettre D, et dont les termes correspondent encore aux différentes heures de la journée. Ce sont les valeurs du coefficient d'occupation, ainsi que nous avons eu l'occasion de l'appeler, et tel que nous l'avons défini.

La notation $\overline{N}$ que l'on trouve encore, et qui indique évidemment une valeur moyenne, représente la valeur probable du nombre des voyageurs qui prendront place à bord des véhicules de la ligne au cours de la journée. Elle tient compte, comme il a été dit, de plusieurs états quotidiens relatifs à la ligne ou groupe de lignes, c'est-à-dire de plusieurs états tels que celui qui nous a servi d'exemple au Tableau I. Ici, la valeur 41,750 s'applique aux deux lignes 4 et 4A, comme d'ailleurs s'y appliquent les résultats qui suivent immédiatement, car les états comptables groupent dans un même total les données relatives aux deux lignes réunies.

La notation ΣVD a un sens usuel, c'est-à-dire qu'elle représente la somme des produits des valeurs correspondantes des deux variables V et D. En d'autres termes, il s'agit ici de la somme $526\times2 + 1411\times7 + \ldots . + 25\times2$; ceci pour la ligne numéro 4. On a une somme analogue pour la ligne numéro 4A, et c'est ainsi que l'on arrive au total de 174,549 pour les deux lignes, obtenant ainsi un résultat susceptible d'être mis en rapport avec la valeur de $\overline{N}$ qui s'étend, elle aussi, aux deux lignes 4 et 4A.

La lettre r représente ici une constante, tout au moins pour une ligne donnée ou un groupe de lignes donnant lieu à un relevé quotidien unique du nombre des voyageurs. Par définition, cette constante est égale au rapport $\frac{N}{\Sigma VD}$. Dans le cas présent, c'est-à-dire dans le cas des lignes 4 et 4A réunies, on trouve donc le

rapport $\frac{41,750}{174,549}$, et on sera satisfait, comme on pourra le vérifier par la suite, d'une valeur approchée à 0.001 près, soit 0.239.

Enfin la variable P est égale au produit $D \times r$, et on voit ici (figure 2) la suite des valeurs approchées que l'on a calculées pour les différentes heures de la journée, soit la suite 0.48, 1.67, 2.39, ..., 0.48. Si l'on veut bien se reporter à la suite des opérations conduisant de proche en proche à une valeur particulière de P, on verra sans peine que cette valeur est en définitive une mesure en «passagers-minutes» applicable à tout véhicule considéré durant cette heure de la journée pour laquelle la valeur particulière de P est valide. Si bien qu'il suffira de calculer le produit de cette dernière par la mesure du temps, en minutes, pendant lequel un véhicule quelconque est en service à cette heure de la journée pour obtenir une estimation du nombre des voyageurs devant prendre place dans le dit véhicule.

Par exemple, soit à donner une telle estimation dans le cas du véhicule numéro 1 de la ligne numéro 4, alors qu'il exécute cette partie de son circuit comprise entre les points d'arrêt Grand-Boulevard et Cherrier/Saint-Hubert, et réputée devoir s'effectuer entre 8.15 A.M. et 8.51 heures A.M. La durée théorique du parcours est de 36 minutes, comme on peut le vérifier, et cette durée s'écoulant entre 8 heures et 9 heures A.M., alors que la variable P prend la valeur 2.39, on arrive au produit $36 \times 2.39 = 86.04$, soit à arrêter notre estimation à 86 voyageurs pour cette partie du circuit. On voit comment généraliser pour le cas où le parcours s'effectuerait entre des heures susceptibles d'affecter plusieurs valeurs distinctes de la variable P.

Évidemment, notre façon de faire suppose, ligne par ligne, une vitesse de renouvellement en voyageurs constante pour une durée de 60 minutes, à laquelle succède brusquement une autre vitesse de renouvellement pour les 60 minutes qui suivent, et ainsi de suite. Or, on pourrait être tenté de vouloir introduire plus de souplesse dans ces évaluations afin de satisfaire à l'idée reçue que les phénomènes observés varient sans à-coups, et la solution consisterait vraisemblablement à donner au coefficient d'occupation un traitement plus raffiné, en ce sens que l'on devrait diminuer la durée de 60 minutes à laquelle il s'applique.

D'une part, il serait difficile de nuancer au delà de ce point un coefficient dont on n'est déjà pas si sûr, étant données les conditions extrêmement flottantes de la circulation à certaines heures du jour. D'autre part, il faut reconnaître que la variation sans à-coups n'est précisément pas celle qui se rencontre ici, où l'affluence se fait brusquement sentir à certains instants de la journée, comme par exemple à la sortie des usines ou des bureaux. Sans prétendre que ces instants coïncident exactement avec les changements de valeur du coefficient, il est néanmoins certain que la valeur qui s'applique à cet instant, comme celle qui précède et celle qui suit, tiennent compte dans l'esprit de celui qui les a données de cet état de choses. D'ailleurs, en fait de données ayant une part d'estimation, le cas présent, celui du coefficient d'occupation, rejoint celui du nombre des voyageurs par ligne, à propos duquel nous avons convenu de ne pas rechercher une précision exagérée, puisque l'échantillon se corrige de lui-même pour compenser tout écart, même grave, dans les prévisions.

Il reste le point de vue des opérations sur le terrain, là où les erreurs dans les prévisions risquent de réserver des surprises désagréables, comme par exemple celle de trouver 400 voyageurs à bord d'un véhicule ayant un personnel prévu pour n'en trouver qu'une trentaine. Nous pouvons être assurés que notre façon d'attribuer ces valeurs au coefficient d'occupation ne sera pas responsable de difficultés d'exécution de cette sorte.

Quant aux conditions que le calcul prescrit pour que soient diminuées autant que possible les erreurs d'estimation qui affecteront notre échantillon, elles seront raisonnablement satisfaites par l'attribution des probabilités qui résulteront des prévisions ainsi faites.

* * *

La valeur de p étant calculée pour toutes les heures de la journée, et pour chaque ligne, nous sommes maintenant en mesure de procéder au détachement systématique de nos véhicules-unités. Ceux-ci ont été obtenus en mettant à contribution notre représentation graphique et ils y ont été indiqués (voir figure 2) à l'aide de points servant à marquer le début et la fin de chacun d'eux. On a déterminé ces bornes en se guidant sur une estimation sommaire des

contenus, fournie à vue d'œil en comparant la valeur ou les valeurs de *P* avec la longueur du tracé. Ensuite on a procédé à l'estimation définitive en s'y prenant comme on l'a déjà indiqué. Entre les deux points consécutifs délimitant graphiquement un même véhicule-unité se trouvent inscrits deux entiers: le premier est le numéro attribué sur sa ligne au véhicule-unité, le second est la valeur probable calculée de son contenu en voyageurs. Bien entendu, chaque véhicule-unité devant être identifié sans équivoque, il devra porter le numéro de sa ligne en plus de celui qu'on vient de lui attribuer sur celle-ci. En fait, ne serait-ce que pour faciliter les recherches, il a été convenu d'ajouter le numéro de service interne du véhicule qui lui a donné naissance. Par exemple, et toujours dans le cas de la figure, le premier véhicule-unité à être détaché portera le numéro 0040-01-001, dont le sens est évident à partir de ce qui précède (le nombre constant des chiffres vient de ce que le tout devra être reporté sur cartes perforées). Comme on peut le voir, le dit véhicule-unité commence au point d'arrêt Claremont/Sherbrooke, où il est réputé passer à 6.07 heures A.M. et se termine au point d'arrêt Cherrier/Saint-Hubert, où il est réputé passer à 7.39 heures A.M. On vérifiera que son contenu estimé est bien celui qui a été inscrit sur le graphique à la suite de son numéro d'ordre sur la ligne, soit 91 voyageurs.

Une question se pose immédiatement. Étant donné que c'est à la même heure et en un point d'arrêt commun que se terminent et que commencent deux véhicules-unités détachés consécutivement d'un même véhicule proprement dit, on doit trancher la question de savoir auquel de ces deux véhicules-unités doit être attribué le voyageur qui monte à bord du véhicule à un tel point d'arrêt. Par définition, il appartiendra au deuxième véhicule-unité, c'est-à-dire à celui qui débute en ce point.

Sans doute observera-t-on encore qu'il arrive de temps à autre que les horaires, sur lesquels nous semblons si fortement appuyés, ne soient pas rigoureusement suivis. Qu'à cela ne tienne, car nous rattachons nos définitions des divers véhicules-unités aux heures théoriques du passage des véhicules aux points d'arrêt, et nous nous y retrouvons avec ces heures tout aussi bien que le font constamment receveurs, wattmen et chauffeurs en dépit du rôle incertain des inspecteurs — privés que sont ces derniers des renseignements

intéressant un ensemble de positions du matériel roulant. L'essentiel pour nous est d'assurer une identité à chacune des transformations successives d'une même voiture proprement dite en véhicules-unités, et, en ceci, le facteur temps venant seulement comme un moyen de contrôle de l'ordre dans la série des transformations, l'importance de la précision pour lui s'en amoindrit d'autant.

Le cas de la suppression pure et simple d'un véhicule-unité est une autre éventualité qui mérite considération lors de l'établissement du plan, bien qu'il ne doive être sans doute qu'un cas d'exception. Que va-t-il advenir ici du système de probabilités sur lequel repose tout le plan d'échantillonnage lorsqu'une probabilité a été attribuée à un véhicule-unité qui n'existe pas et que, par surcroît, cette même unité vient à tomber dans l'échantillon? Il ne se passera rien qui n'appartienne d'emblée aux prévisions générales du plan. Comme on le sait, le véhicule-unité est tiré dans l'échantillon des véhicules avec une probabilité qui lui est propre et qui détermine par contre-coup l'intervalle à appliquer pour le prélèvement de son sous-échantillon en voyageurs, ce qui entraîne encore de fixer le rang du premier voyageur à y inclure. Or, dans le cas du véhicule supprimé, ce voyageur ne se présente pas à coup sûr, et il s'agit simplement d'un passage à la limite, si l'on tient compte que l'échantillon se rectifie de lui-même, où le contenu du sous-échantillon devient 0. D'ailleurs, on peut établir directement par le calcul que l'attribution d'une probabilité à un véhicule n'existant pas ne modifie pas la valeur probable de la variable d'estimation. La seule répercussion pourrait être de diminuer le contenu de l'échantillon définitif en voyageurs.

Plus sérieux est le cas où un trop grand retard va jusqu'à entraîner le déroutement du véhicule, sans aller cependant jusqu'à sa suppression. Il est bien prévu que notre enquêteur est tenu de prendre son poste à bord du véhicule-unité qui lui a été assigné bien avant l'heure et le point d'arrêt qui marquent le début de celui-ci. Cependant, si l'on est sans inquiétude du côté de la présence de l'enquêteur, il reste que le passage du véhicule au point d'arrêt prévu ne s'effectuant pas régulièrement, l'enquêteur n'est pas sans apporter une part de jugement dans la définition, à posteriori, des véhicules-unités, dont la théorie n'aurait pas lieu de se féliciter. En fait, un tel cas ne peut-il être que très rare, d'autant

plus que le personnel du service du mouvement de la C.T.M. avait été invité par l'administration supérieure à se montrer favorable à notre enquête, et les perturbations de ce genre ont pû être pratiquement évitées.

Maintenant que les véhicules-unités ont tous été définis, et le contenu de chacun étant estimé, il y a lieu de passer tout de suite au prélèvement de l'échantillon à tirer de cette population de véhicules. On se rappelle qu'il s'agit là d'un premier stade d'échantillonnage et qu'il restera, au deuxième stade, à tirer un certain nombre de voyageurs de chacun des véhicules-unités sortis dans le premier échantillon.

* * *

Imaginons les données suivantes, qui se rapprochent d'ailleurs assez bien des faits: soit un total de 1,500,000 voyageurs pour la journée, soit à tirer 150 véhicules-unités au premier stade d'échantillonnage, et soit à tirer un nombre probable de 12 voyageurs de chacun de ces cent cinquante véhicules-unités, au deuxième stade d'échantillonnage. Tel qu'a été établi notre plan d'échantillonnage, il est bien entendu que ce nombre probable de voyageurs doit être égal à une constante: et celle-ci a été fixée à 12 pour des raisons d'ordre exclusivement pratique. Il s'agissait d'accommoder des situations extrêmes, d'une part lorsque le véhicule est très peu occupé, d'autre part lorsqu'il doit y avoir foule, et en s'aidant de la possibilité d'agir sur la définition du véhicule-unité pour en allonger le parcours ou le diminuer de façon à prévoir des conditions d'enquête réalisables dans chaque cas particulier. Quant au nombre des véhicules-unités à tirer dans l'échantillon, il est à déterminer par le calcul, partant de la constante qu'on vient de fixer à l'instant, partant encore d'une évaluation plus ou moins risquée de quelques paramètres appartenant à la population des voyageurs, et cherchant sur une base de calcul à réaliser un compromis entre le coût de l'enquête et la précision attendue des variables d'estimation pour les caractéristiques essentielles de la population à l'étude. De plus, il est compris que les véhicules-unités doivent être tirés dans leur échantillon avec des probabilités proportionnelles à leurs contenus estimés.

Voici le moyen mis en oeuvre pour tirer l'échantillon des véhicules-unités, en accord avec ce dernier point. Nous com-

mençons par construire le tableau suivant (voir tableau II). La première colonne reçoit les numéros d'identité de nos 15,000 véhicules-unités, dans la deuxième colonne sont inscrits respectivement les contenus estimés, la troisième et dernière colonne porte le total cumulatif des termes de la précédente.

Tableau II

0010-01-001	30	30
. . .	100	130
. . .	150	280
. . .	120	400
. . .	200	600
. . .	150	750
. . .	300	1,050
. . .	250	1,300
. . .	100	1,400
0010-12-010	200	1,600
. . .	100	1,700
. . .	. . .	.
. . .	100	10,920
. . .	200	11,120
. . .	300	11,420
0020-09-123	180	11,600
. . .	250	11,850
. . .	.	.
. . .	.	.
. . .	150	1,499,800
1950-07-118	200	1,500,000
	1,500,000	

La somme des entiers de la deuxième colonne étant 1,500,000, faisons la répartition des entiers de 1 à 1,500,000 de la façon suivante: au premier véhicule-unité, de contenu 30, attribuons les entiers de 1 à 30; au deuxième, de contenu 100, attribuons à la suite les entiers de 31 à 130; au troisième, de contenu 150, attribuons de la même façon les entiers de 131 à 280, et ainsi de suite, jusqu'au dernier véhicule-unité, de contenu 200, auquel seront attribués les entiers de 1,499,801 à 1,500,000. À chaque véhicule-unité correspond donc un groupe d'entiers consécutifs bien définis, ou groupe de numéros consécutifs, en nombre égal au contenu en voyageurs du dit véhicule; et ce sont ces groupes d'entiers attribués respectivement aux divers véhicules-unités que l'on retrouve tout au long de la troisième colonne de notre tableau.

Cette répartition des numéros étant faite, nous allons maintenant désigner les véhicules-unités à prendre dans l'échantillon en les prélevant à intervalle régulier, tout comme il nous est arrivé de faire pour prélever le sous-échantillon de voyageurs à l'intérieur d'un véhicule-unité donné, la seule différence étant qu'ici chaque voyageur n'avait que son seul numéro d'ordre, tandis que là, chaque véhicule-unité en possède autant qu'il lui est attribué de voyageurs. Comme nous devons tirer 150 véhicules-unités dans l'échantillon et que nos numéros d'ordre vont jusqu'à 1,500,000, calculons le rapport $\frac{1,500,000}{150}$, lequel étant égal à l'entier 10,000 nous donne immédiatement l'intervalle qu'il va falloir appliquer. Ensuite, et toujours comme nous avons fait dans le cas du sous-échantillon de voyageurs, tirons au hasard un entier compris entre 1 et 10,000. Soit, par exemple, 1,462 l'entier ainsi tiré, que nous aurons d'ailleurs obtenu en faisant usage d'une table, tout comme dans le cas du sous-échantillon de voyageurs. Finalement, et toujours en poursuivant l'analogie avec le cas du sous-échantillon de voyageurs, nous formons la suite des entiers 1,462, 11,462, 21,462, ... 141,462, se succédant à intervalle régulier de 10,000, et au nombre de 150, et nous prélevons dans l'échantillon les 150 véhicules-unités qui possèdent ces numéros parmi les numéros d'ordre qui leur sont respectivement attribués.

Si l'on veut bien rapprocher ce dernier exemple de celui du tableau II, on trouvera que le premier véhicule-unité à tomber dans l'échantillon porterait le numéro 0010-12-010, le deuxième serait celui qui porte le numéro 0020-09-123, et ainsi de suite jusqu'au cent cinquantième et dernier véhicule-unité.

Il est encore facile de vérifier ceci. Nous commençons donc par tirer au hasard un entier qui doit être compris entre 1 et 10,000. Et cet entier détermine une suite de 150 termes qui décideront des 150 véhicules-unités tombant dans l'échantillon. Imaginons le cas d'un véhicule-unité quelconque auquel on aurait attribué, par exemple, 280 numéros d'ordre, et que ceux-ci soient les entiers compris entre 23,201 et 23,480. Il est manifeste que le véhicule sortira chaque fois que l'entier tiré au hasard sera l'un des 280 entiers compris entre 3,201 et 3,480, et l'un de ces entiers seulement. Et la probabilité de chacun d'eux d'être tiré étant égale à $\frac{1}{10,000}$, la

probabilité du véhicule de sortir dans l'échantillon est égale à $\frac{280}{10,000}$. Un autre véhicule devant recevoir 300 voyageurs se verra attribuer une probabilité égale au rapport $\frac{300}{10,000}$. On généralisera facilement. Ainsi donc les diverses unités de notre population de véhicules-unités auront reçu, d'après ce système, des probabilités proportionnelles à leur contenu en voyageurs, et ceci est entièrement conforme aux conditions d'application de notre plan d'échantillonnage.

Les véhicules-unités de l'échantillon étant maintenant connus, c'est donc le premier stade de l'échantillonnage qui vient d'être achevé. Au deuxième stade de l'échantillonnage, il reste à tirer un échantillon de voyageurs de chacun de ces véhicules-ci. Or, nous savons comment calculer l'intervalle à appliquer à l'intérieur d'un véhicule donné pour arriver à cette fin, et nous savons comment procéder pour désigner le premier voyageur à passer dans l'échantillon. Le tout est déterminé en partant de la probabilité qui a été attribuée au dit véhicule ou, en fait, à son contenu estimé, si bien que nous sommes prêts à construire le tableau que voici (tableau III) et dans lequel vont être consignées toutes les données dont nous aurons besoin pour effectuer les opérations sur le terrain.

Comme on le voit, toutes les indications portées au tableau se suivent d'elles-mêmes d'après ce qui a été dit jusqu'ici, exception faite pour le numéro d'ordre qui vient seulement d'être attribué à chaque véhicule-unité à l'intérieur de l'échantillon.

* * *

À partir de cet état, il reste à établir un plan pour arriver à une répartition des tâches entre les différents interviewers, sans compter le cas du personnel qui sera chargé de la direction et de la surveillance de ceux-ci sur le terrain. En pratique, on s'apercevra qu'il devient assez laborieux de concilier pour l'équipe des interviewers, et les heures de service, et la durée des parcours que ces derniers auront à effectuer pour prendre de nouveau du service, ces parcours se passant entre des points d'arrêt que l'on trouvera souvent fort éparpillés. Sans compter que le chevauchement des tâches devient encore plus difficile à éviter quand

Tableau III

ÉCHANTILLON DES VÉHICULES						ÉCHANTILLON DES VOYAGEURS		
Numéro d'ordre dans l'échantillon	Numéro de la ligne	Numéro de service du véhicule	Numéro du véhicule-unité	Circuit du véhicule-unité		Contenu estimé du véhicule	Intervalle	Rang du premier voyageur
				Point d'arrêt	Heure du passage au point d'arrêt			
26	4A	32	58	Dép. Victoria–Sherbrooke via Gd Boulevard, St–Hubert Arr. Girouard–Sherbrooke	2.31 P.M. 3.48 P.M.	92	8	5
27	5	24	39	Dép. Delorimier–Ontario via Aylmer Loop, Viau–N.–D. Arr. Aylmer Loop	6.29 A.M. 7.48 A.M.	238	20	11
88	29	14	9	Dép. Garland via Youville Loop Arr. Park Avenue–Bernard	8.14 A.M. 9.38 A.M.	182	15	2
89	31	3	116	Dép. Terminus Craig via Belgrave, Terminus Craig Arr. Belgrave	1.44 A.M. 2.55 A.M.	28	2	1

on doit s'accorder une marge de sécurité prévoyant les retards toujours possibles sur les horaires. En plus, faut-il encore tenir compte de la nécessité de prévoir deux hommes à bord de certains véhicules où un encombrement brusque est prévu, ou susceptible de se produire, l'un des deux hommes étant plus particulièrement affecté au comptage — du moins pour une partie du parcours, après quoi il pourrait devenir disponible — tandis que l'autre devra faire les interviews. Enfin citons cette restriction, qui s'applique à l'équipe des interviewers et qui vient encore diminuer ses possibilités, et qui oblige à tenir compte dans une certaine mesure de la personnalité d'un interviewer, en particulier vis-à-vis les réactions possibles des voyageurs de telle ou telle ligne.

La répartition des tâches entre les divers interviewers étant faite, en d'autres termes, chacun de ceux-ci s'étant vu attribuer un certain nombre de véhicules-unités de l'échantillon, on fera correspondre à chacun de ces véhicules une carte de travail décrivant, pour le bénéfice de l'interviewer, les différentes phases de son travail à bord du dit véhicule. On peut voir sur la figure 3 un fac-similé de l'une de ces cartes. La légende et les inscriptions rapportées s'interprètent d'elles-mêmes.

Si l'on veut bien comparer la carte représentée ici aux données relatives au véhicule-unité qui lui a donné naissance, données qui figurent précisément au tableau III, on pourra se figurer facilement une marche des événements justifiant les écritures qu'on y relève.

On aura remarqué que le nombre des voyageurs ou, en d'autres termes, le rang du dernier voyageur à monter effectivement à bord du véhicule, est égal à 112, différant ainsi du nombre prévu, et qui était 92. On constatera encore que le nombre des voyageurs interrogés a été de 14, et s'écarte ainsi du nombre probable, fixé à 12 pour tous les véhicules-unités, ce qui n'est d'ailleurs qu'une conséquence de ce qui précède. D'une façon générale, il est bien certain que le nombre des voyageurs effectivement rencontrés est susceptible de s'écarter plus ou moins du contenu estimé pour le véhicule et, une fois de plus, sommes-nous au courant que l'inégalité de ces deux nombres, l'un estimé l'autre observé, n'entraîne aucune conséquence du point de vue de la validité de l'échantillon, celui-ci, encore une fois, se rectifiant de lui-même.

Figure 3

ÉNUMÉRATION DES USAGERS DU RÉSEAU DE LA COMMISSION DE TRANSPORT DE MONTRÉAL

Echantillon No : 26 | Nom de l'interviewer : Georges Trudeau N°17 | Ligne No : 4 A, Véhicule No : 32

Point de départ : Victoria - Sherbrooke | Point d'arrivée : Girouard - Sherbrooke
Heure de départ : 2.31 PM | Heure d'arrivée : 3.48 PM

via Grand Boulevard, St. Hubert

Numéro d'ordre des usagers tombant dans l'échantillon — Compte rendu de l'interview (*)

5	13	21	29	37	45	53	61	69	77	85	93	101	109	117	125	133	141
1	2	2	1	1	1	1	2	1	3	1	1	2	2				

TOTAL DES USAGERS : 112

(*) CODE : 1 — Renseignements complets;
2 — Identité obtenue;
3 — Refus ou fausse identité.

Georges Trudeau
Signature de l'interviewer :

Il est entendu que c'est le contenu estimé du véhicule-unité qui sert à déterminer la suite des numéros d'ordre des voyageurs qui devront être pris dans l'échantillon — soit, ici, la suite 5, 13, 21, 29, . . . 141 — et que celle-ci a dû être inscrite sur la carte avant que cette dernière ne soit remise à l'interviewer. Celui-ci inscrivant seulement de son propre chef le numéro d'ordre du dernier voyageur monté à bord du véhicule. Toujours dans le même ordre d'idées, on voit que la suite a été prolongée bien au delà de ses 12 premiers termes: c'est afin de se trouver prêt à répondre à des circonstances analogues à celles que l'on a rencontrées ici, c'est-à-dire lorsque le contenu observé du véhicule dépasse le contenu estimé.

La figure 4 est un fac-similé du questionnaire qui devait être rempli pour chaque voyageur désigné par son rang sur une carte de travail telle que celle illustrée par la figure 3. En le comparant précisément à cette dernière carte, on vérifiera par les écritures qui y sont portées qu'il concerne le premier voyageur désigné sur la dite carte de travail, les numéros d'identification se correspondant: numéro de l'échantillon et rang du voyageur. Citons à la suite quelques remarques s'appliquant au questionnaire d'une façon générale ou à la façon de le remplir.

* * *

On ne sera pas sans remarquer que la question qui vient en tête de la formule est peut-être celle que le voyageur serait le moins disposé à accepter comme entrée en matière: rares sont ceux qui produisent leur nom à tout venant. Il faut tout d'abord observer que le questionnaire n'étant pas destiné à être placé entre les mains du public, l'ordre des articles imprimés y importe peu, puisqu'il est entendu que l'interviewer, instruit de son rôle, saura adopter pour ses questions un ordre qu'il fera varier selon l'attitude de son interlocuteur. En définitive, l'ordre qui a été adopté pour la rédaction de la formule est plutôt celui qui semblait se prêter le mieux au dépouillement des documents.

D'ailleurs, dans les cas où le temps est limité à l'extrême, comme par exemple lorsqu'un voyageur ne doit séjourner dans le véhicule qu'entre deux points d'arrêt consécutifs, il est entendu

Figure 4

1 09-02-0 04-03-1 12-18 05-10

ENUMÉRATION DES USAGERS DU RÉSEAU DE LA COMMISSION DE TRANSPORT DE MONTRÉAL

ENUMERATION OF TRAVELLERS ABOARD VEHICLES OF THE MONTREAL TRANSPORTATION COMMISSION

Nom / *Name*	Adresse / *Address*	Tél.	Profession
Roland Desrosiers	2240 Des Érables	LAG-9809	Chauffeur camion

Place d'affaire ou Employeur : *Business Address or Employer* Brown & Mc Dougal 6824 St Patrick

Véhicules empruntés au cours de la journée — *Buses or Tramways boarded during the day* (1)	Condition d'utilisation du réseau — *Use made of public transportation* (2)	Avez-vous une auto ? — *Do you have a car ?* (3)	Statut occupationnel — *Occupational status* (4)	Personnes à charge — *Dependants* (5)	ECHANTILLON No : 26 — USAGER No : 5
6	1	N	2	3	

N°17

(CODE AU VERSO)

Georges Trudeau

Signature de l'interviewer / *Signature of the interviewer*

CODE:

(2): 1 – Régulièrement pour le travail;
2 – Régulièrement pour sorties et magasins;
3 – À l'occasion seulement.

(4): 1 – Employeur;
2 – Employé;
3 – Travail individuel.

que l'interviewer doit s'attacher avant tout à obtenir n'importe quelle bribe de renseignement susceptible d'être utilisée par la suite pour retrouver le dit voyageur; et les renseignements demandés en tête de la formule sont de cet ordre. Le numéro de téléphone, par exemple, est celui qui se prête le mieux à cet objet, se trouvant être le plus directement et le plus économiquement utilisable.

Le cas qui vient d'être cité rejoint d'ailleurs, pour la suite des opérations, celui des questionnaires incomplètement remplis, puisque l'intéressé doit être de nouveau atteint pour continuer l'interview, que ce soit à son domicile ou au lieu de son travail, par téléphone ou par visite.

Dans le cas d'une ménagère, on demandait par surcroît la profession du mari: car on se doute qu'une rubrique générale comprenant toutes les ménagères, sans distinction, ne manquerait pas d'appeler une classification subséquente, et le renseignement supplémentaire obtenu pourra vraisemblablement être utilisé à cette fin.

On a omis systématiquement de questionner ouvertement sur le montant du revenu ou, plutôt, sur le montant du salaire. S'il est certain que des renseignements de cet ordre méritent une place de choix dans cette étude, il est non moins certain qu'une question directe posée dans les circonstances actuelles n'aurait pas manqué d'aboutir à une collection de données arbitrairement faussées. Il a été préféré d'employer l'alternative, détournée sans doute mais discrète, consistant à se montrer satisfait de connaître la nature de l'occupation rétribuée et le lieu de son exercice, quitte à passer au salaire après coup et dans des conditions d'enquête plus favorables. D'ailleurs, avec ces renseignements en main, était-il légitime d'escompter régler un bon nombre de cas en se reportant à des relevés d'intérêt général sur les salaires. Quant à la question posée relativement au lieu où s'exerce l'occupation, elle était justifiée bien autrement que comme expédient, puisqu'elle était de toute façon nécessaire pour répondre à la partie de cette étude traitant du trafic.

Pour ce qui est de la façon de remplir la colonne (1) du questionnaire, disons bien qu'il s'agissait ici de donner le nombre des véhicules distincts que l'usager était présumé devoir emprunter tout au cours de la journée d'enquête, et quel que soit

le moyen d'accès à ces véhicules: perception directe sous n'importe quelle forme, remise du titre de correspondance ou simple présentation de celui-ci. Non sans peine a-t-on pu inculquer aux interviewers une idée nette relativement à la portée de cette question. Comment faire admettre que la question elle-même n'était pas limitée au nombre des perceptions au cours de la journée, ou au nombre des trajets dans un sens ou dans l'autre, pas plus qu'elle ne s'appliquait à un nombre quotidien relatif à une période arbitraire? Et l'attachement à de telles idées préconçues n'a pas été la moindre difficulté rencontrée à l'occasion du cours d'instruction donné à tous les interviewers, préalablement à l'enquête.

On se doute bien qu'un voyageur qui, tôt dans la journée, aura fait une déclaration relative à la colonne (1) du questionnaire n'ira pas s'astreindre, pour le restant de la journée, à prendre exactement le nombre de véhicules qui aura été inscrit, tout juste pour faire honneur à sa déclaration. Malgré tout, on a dû considérer que l'erreur qui peut résulter des écarts possibles dans ce sens doit être pratiquement négligeable, qu'elle vienne de l'ensemble des usagers réguliers ou qu'elle vienne de l'ensemble des usagers occasionnels.

Quant à l'importance du nombre que l'on s'attend de voir à la colonne (1) du questionnaire, elle est extrême. C'est ce nombre qui va rendre possible le passage des caractéristiques mesurées sur les voyageurs — ceux-ci formant la population effectivement échantillonnée — aux caractéristiques telles qu'elles existent dans la population des usagers. Sans doute, nous trouvons-nous porter intérêt au voyageur, et, de là, à certaines questions relatives au trafic, mais il en est seulement ainsi parce que le voyageur nous est devenu un intermédiaire indispensable pour atteindre l'usager, celui-ci étant bien en définitive l'individu distinct utilisant le réseau et qui nous occupe comme tel. Or, le système de probabilités dans lequel nous évoluons présentement est parfaitement en règle vis-à-vis des voyageurs, tandis qu'il ne l'est pas pour les usagers: un usager se trouvant d'autant plus «exposé» à être pris dans l'échantillon qu'il change plus fréquemment de véhicule; et en conséquence, nous ne pouvons mesurer les caractéristiques de la population des usagers dans l'état actuel des choses. Pour

nous rattraper sur ce point, c'est-à-dire pour remettre les usagers sur un pied d'égalité entre eux, voici que pour la troisième fois nous aurons à faire appel à un système de probabilités dans notre plan d'échantillonnage. Déjà, nous sommes parvenus avec les deux systèmes précédents à donner aux voyageurs la même probabilité d'être pris dans l'échantillon, et c'est au tour des usagers maintenant. Pour ce faire, nous aurons à attribuer un coefficient bien déterminé à chaque voyageur, et qu'il soit pris ou non dans l'échantillon tout voyageur aura son coefficient. Celui-ci dépend du nombre de véhicules empruntés par le voyageur, et ce sera à partir du nombre porté à la colonne (1) du questionnaire que nous pourrons l'attribuer en particulier à chaque voyageur tombé dans l'échantillon.

Nous avons bien dit que chaque voyageur se voyait attribuer un coefficient — et c'est sous cette forme que le troisième système de probabilités entre en jeu —, et que ce coefficient existe indépendamment du fait qu'il nous est explicitement connu pour tel ou tel voyageur, ou bien qu'il ne l'est pas; en d'autres termes, il existe indépendamment du fait que le voyageur est passé dans l'échantillon, ou non. À ce sujet, remarquons en passant que l'on ne saurait éluder le nouveau problème sur les probabilités, né du fait que les usagers sont inégalement «exposés» à tomber dans l'échantillon, en s'assurant tout simplement que l'échantillon des voyageurs ne renferme pas deux voyageurs de la même identité. Se montrer satisfait de cette façon de faire serait commettre une erreur grave, issue d'un préjugé fortement enraciné sans doute, mais qui n'en reste pas moins une faute démesurée.

Les contenus des colonnes suivantes du questionnaire s'expliquent d'eux-mêmes, et dans le cas des colonnes (2) et (4), il suffit d'inscrire un seul chiffre, en accord avec le code prévu à cet effet.

* * *

Les notations portées au questionnaire (voir figure 4) et qui ne sont pas des réponses directes aux questions explicitement énoncées dans la formule sont placées là par la suite, c'est-à-dire après dépouillement des documents et, s'il y a lieu, après que ceux-ci ont été complétés. Elles sont ici inscrites en tête de la formule

et sont le résultat d'une refonte d'une partie des renseignements fournis par l'interviewé. Elles répondent à un code établi en vue du transfert du contenu du questionnaire sur carte perforée. Examinons-les l'une après l'autre, et en leur conservant leur ordre d'inscription.

L'entier d'un seul chiffre inscrit en haut et au coin gauche, «1» dans le cas présent, répond au code suivant:

Tableau IV

Ménagère	0
Homme au travail	1
Femme travaillant au-dehors	2
Écolier	3, etc.

Le groupe de cinq chiffres venant à la suite doit être compris de la façon suivante. Il est entièrement relatif au domicile du passager, et se compose de deux entiers de deux chiffres chacun, suivis d'un troisième et dernier entier d'un seul chiffre. Les deux premiers entiers forment ensemble les coordonnées géographiques du lieu du domicile, établies comme il va être dit. Le troisième entier fait connaître la localité administrative de laquelle dépend le même lieu, et répond au code suivant:

Tableau V

Montréal	0
Lachine, Ville St-Pierre	1
Montréal Ouest, Côte St-Luc, Hampstead	2
Côte de Liesse, Ville St-Laurent	3
Verdun, Ville Lasalle	4
Outremont	5, etc.

Les coordonnées géographiques présentement utilisées doivent être entendues dans un système d'axes rectangulaires, l'axe des abscisses ayant la direction Nord-Nord-est, et l'unité de longueur ayant elle-même pour mesure environ 1 mille. Afin de situer quelque peu les axes, disons que, dans ce système, la cathédrale de Montréal se trouve presque exactement au point de coordonnées (07,02).

Le groupe suivant de cinq chiffres est à interpréter tout à fait de la même manière que le précédent, sauf qu'il s'applique au lieu où s'exerce l'occupation rétribuée, en place du domicile. Le système d'axes de coordonnées est le même, et le même code sert à indiquer la localité administrative.

Le groupe de chiffres suivant est à interpréter de la façon suivante. Il procède à la fois des données qui nous ont amené déjà aux deux groupes qui précèdent, c'est-à-dire qu'il est relatif à la fois au domicile et au lieu où s'exerce l'occupation. Il est composé de deux entiers de deux chiffres chacun. Le premier de ces entiers-ci est la mesure de la distance, à vol d'oiseau, qui sépare le domicile du lieu du travail, l'unité étant le quart de mille. Le second est la mesure de l'angle formé par la direction domicile-lieu du travail avec la direction Ouest-Est, ou plutôt la direction réputée telle par l'orientation générale des rues dans la région intéressée. Le sommet de l'angle est pris au domicile, et l'angle est mesuré dans le sens trigonométrique à partir de la direction Ouest-Est qu'on vient de définir, l'unité étant choisie à raison de 9 unités à l'angle droit.

Les données qui viennent d'être mises en code à l'instant même vont permettre deux répartitions différentes des usagers, en effectuant pour cela les tabulations appropriées. Tout d'abord, utilisant la mesure de l'angle, et tenant compte que celui-ci est un angle orienté, nous ferons une répartition des usagers faisant valoir non seulement la direction de leur déplacement domicile-lieu du travail mais aussi le sens de celui-ci. Ensuite, nous aurons une répartition des usagers selon la distance qui sépare domicile et lieu du travail. Évidemment, peut-on avoir cette distance à vol d'oiseau, puisqu'il suffit pour cela d'utiliser directement le renseignement que nous possédons; mais on préférera probablement la répartition en fonction de la distance parcourue selon le tracé général des rues de la région. On fera le calcul à partir des deux mesures que l'on possède: celle de la distance et celle de l'angle. Pour cela, on imaginera que les rues se coupent constamment à angle droit, ce qui ne saurait s'éloigner beaucoup de la réalité, et quant à la distance ainsi calculée, ce sera celle qui est effectivement parcourue par l'usager, que ce soit avec le concours des moyens de transport en commun ou par ses propres moyens. Quant aux tabulations à effectuer, il est entendu qu'elles se feront en accord avec le troisième système de probabilités que nous sommes convenus d'appliquer à la population des voyageurs sous forme d'un coefficient à attribuer à chacun d'eux.

Enfin, nous arrivons au dernier groupe de chiffres appartenant à la série des données rapportées après coup sur le questionnaire, soit 05-10 dans le présent exemple. Il est relatif à la nature de l'occupation ainsi qu'au salaire retiré de l'exercice de celle-ci. Il se compose de deux entiers, de deux chiffres chacun. Le premier fait connaître la nature de l'occupation suivant une classification répondant au code suivant:

Tableau VI

Présidents, gérants, etc.	1
Fonctions administratives	2
Comptables, caissiers, secrétaires, etc.	3
Vendeurs	4
Transport	5
Administrations publiques	6
Services personnels	7
Services professionnels	8
Ouvriers: mécanique et construction	9
Ouvriers: alimentation, textiles, bois ne servant pas à la construction	10
Contremaîtres et inspecteurs	11
Opérateurs	12
Journaliers, servantes	13
Ecoliers, étudiants	14
Ménagères	15
Sans profession, rentiers	16
Profession inconnue	17

Le second entier donne le salaire, ou plus exactement donne-t-il le numéro d'ordre de la classe de salaires à laquelle le présent salaire appartient. De telles classes ont été prévues nombreuses. Par suite, chacune se trouvait être de petite étendue: trop petites et trop nombreuses en fait pour figurer telles quelles dans les répartitions définitives. De cette façon, on se réservait la possibilité de remanier les classes au gré des circonstances, c'est-à-dire de les remanier à la lumière des résultats bruts fournis par une première tabulation. Bien entendu, il s'agissait d'opérer le fusionnement des classes voisines, et de le faire par le meilleur arrangement susceptible de sauvegarder, sinon de provoquer, l'égalité des intervalles entre les classes, éliminant à coup sûr par ce moyen celles dont le contenu trop faible n'aurait pu garantir une estimation raisonnable pour la population. Ainsi traitée, la répartition des salaires de la population des usagers était susceptible d'être mieux adaptée à la situation particulière qu'elle embrasse, com-

parativement à une répartition en classes qui aurait été décidée à priori.

Au point où nous avons avancé les choses, il est compris que les questionnaires sont maintenant prêts à être reportés sur cartes perforées. À chaque questionnaire correspondra une carte et la totalité des renseignements que l'on voit sur ceux-ci sera ainsi reportée, y compris le numéro de l'interviewer. Et le transfert se fera d'une manière immédiate, puisque les renseignements non chiffrés ont reçu le bénéfice d'un code établi, de proche en proche, en vue de cette transformation; sauf dans le cas de la colonne (3), qui n'a pas encore été traitée de ce point de vue, et pour laquelle on emploiera le code suivant: 0 pour non, 1 pour oui.

* * *

En résumé, depuis l'opération de leur dépouillement jusqu'à la confection des cartes perforées, les questionnaires se trouvent avoir donné lieu à la suite des opérations que voici.

Un classement évident les a tout d'abord fait répartir en trois groupes: celui des questionnaires complètement remplis, celui des questionnaires qui, tout en n'étant pas complètement remplis, portent des traces suffisantes de l'identité du voyageur, enfin, le dernier groupe est formé du rebut.

L'opération prochaine était destinée à amener le second goupe à l'état du premier. Et cela ne représentait pas une mince besogne que de rétablir le contact avec tous les usagers touchant à ce groupe. Que l'on songe seulement qu'un bon quart de l'échantillon était dans ce cas, sans compter qu'il ne pouvait être question d'utiliser un quelconque personnel, alors qu'il fallait approcher l'usager avec assez de tact pour qu'il ne vienne pas à se sentir poursuivi. Le téléphone, évidemment, a été en la circonstance le moyen de choix; mais il a fallu parfois recourir à la visite à domicile. D'ailleurs, si l'on veut plus de détails, on apprendra qu'un patient travail préliminaire s'était imposé avant que de pouvoir traiter comme il vient d'être dit les questionnaires de cette catégorie. C'était d'abord à retrouver le vrai nom propre, celui de la rue ou celui du voyageur, au travers des barbarismes curieusement osés et fréquents imputables aux interviewers. D'une façon analogue, il fallait retrouver le vrai numéro, celui du télé-

phone ou celui du domicile, au travers des interversions de chiffres revenant d'une façon non moins persistante. Les principaux instruments pour effectuer ce travail de redressement ont été l'annuaire du téléphone, et son contraire dit «livre rouge», avec l'annuaire dit *Lovell*, donnant la liste des noms et des rues pour la région métropolitaine.

Pour ce qui est du troisième groupe, les versements ne s'y sont fait qu'à regret, comme on l'imagine, puisque chacun de ceux-ci doit être considéré comme une source d'erreur plus ou moins confuse. De ce point de vue, il s'est trouvé recueillir des cas d'inégale importance, depuis le refus de répondre pur et simple, en passant par le refus plus ou moins voilé, jusqu'au cas d'abandon du questionnaire tiré du groupe précédent et jugé trop onéreux à redresser. Numériquement, le groupe s'est tenu dans des limites rassurantes quant à la qualité de l'échantillonnage, d'autant plus qu'une étude statistique destinée à vérifier certaines hypothèses mettant en cause l'interviewer, la ligne ou les lignes sur lesquelles s'enregistraient ses insuccès et celles sur lesquelles il était assigné a démontré presque invariablement que les refus et les abandons se trouvaient étroitement liés à la personnalité de l'interviewer.

Les questionnaires se trouvant désormais ne plus former que deux groupes, on a fait la revue de ceux qui étaient complètement remplis pour y porter, à commencer par là, la première des inscriptions venant après coup, c'est-à-dire l'inscription de la classe revenant au voyageur d'après la classification et en accord avec le code établis au tableau IV.

L'opération prochaine était destinée à transformer, selon le code que l'on connaît, les adresses du domicile et du lieu de travail en coordonnées géographiques, avec numéro de la circonscription administrative, et à inscrire le résultat de la transformation sur le questionnaire même. Ce qui s'est fait en utilisant, d'une part une carte de la région métropolitaine, ayant subi une préparation à cet effet, et d'autre part le code du tableau V. Le calcul des coordonnées polaires combinant domicile et lieu du travail venait à la suite, réclamant évidemment l'emploi d'un dispositif permettant de mesurer sur la carte, à la fois les distances et les angles, et l'inscription du résultat se faisait encore immédiatement sur le questionnaire même.

Ensuite est venu le problème de la détermination de l'exacte catégorie à laquelle revenait la profession déclarée par le voyageur. Chaque questionnaire a été examiné au mérite, les catégories étant bien définies et le code étant conforme au tableau VI.

Enfin, il a fallu faire une estimation du salaire dans chacun des cas. Puis celle-ci a été indiquée sur le questionnaire même, sous la forme du numéro d'ordre de la classe de salaire dans laquelle elle tombait. Ayant renoncé à demander ouvertement une évaluation du salaire, deux procédés s'offraient subséquemment pour se rattrapper. On pouvait s'adresser directement à l'employeur: et plutôt que d'embarrasser ce dernier en lui demandant de préciser le montant du salaire en question, il a semblé préférable de lui laisser la latitude de ne pas aller au-delà de l'indication d'une classe de salaires. De cette façon, il n'était guère à craindre de se voir refuser une réponse. Dans d'autres cas, l'occupation était suffisamment bien définie pour qu'il soit possible de verser l'intéressé dans une classe également bien définie en s'appuyant sur les documents officiels d'intérêt général. Les quelques cas de professions libérales, tout ensemble avec ceux des commerçants, ont donné lieu à une classe à part, sans plus, car il aurait été futile de chercher à obtenir des précisions au delà de la notion d'effectif de la classe.

* * *

Après ce rappel sommaire des opérations auxquelles ont donné lieu les questionnaires, imaginons ceux-ci maintenant reportés sur cartes perforées. D'ailleurs, le questionnaire sera devenu de ce fait pratiquement inutile. Quant aux cartes perforées, elles vont servir dorénavant à effectuer mécaniquement la suite des opérations. On les trouvera pour effectuer les calculs destinés à tenir compte, dans chaque cas, de l'inégalité des probabilités qui reviennent aux divers usagers dans la population des voyageurs, pour effectuer les calculs ayant pour objet de passer des coordonnées polaires enregistrées — distance à vol d'oiseau et mesure de l'angle — aux distances effectivement parcourues en suivant le tracé des rues. Enfin doit venir avec les cartes perforées un travail de tabulations multiples aboutissant finalement aux répartitions de toutes sortes que l'on attendait.

* * *

À un autre stade du travail, et qui sera le dernier mais non le moindre pour ses conséquences, se présente la tâche d'évaluer les erreurs pouvant affecter les diverses estimations fournies par l'échantillon, estimations qui forment la substance des répartitions auxquelles on vient d'aboutir.

On avait bien reconnu, au tout début de cet exposé, la nécessité d'une mesure de la précision à laquelle on pouvait s'attendre de la part d'un échantillon prélevé dans telles ou telles conditions. Sans une telle mesure, irait-on se baser sur un certain résultat tiré de l'échantillon pour prendre une décision administrative? Ce serait tout simplement courir le risque de se déterminer dans le sens opposé à celui qui se doit. Pour prendre un exemple, disons que l'on n'hésiterait guère à prendre une décision dans un sens bien défini — et qui pourrait consister à faire quelque chose tout aussi bien qu'à s'abstenir d'agir — si l'on pouvait nous garantir un résultat prenant la forme suivante: 75 p.c. de la population des usagers possède telle caractéristique, que les autres usagers ne possèdent pas, et le pourcentage donné peut être affecté d'une erreur qui ne dépasse pas 0.50 p.c., dans un sens ou dans l'autre, et cette garantie est donnée avec une probabilité de 95 p.c. de ne pas se tromper. D'ailleurs, si l'on devait avoir quelque inquiétude à ne se retrancher qu'à l'abri d'une garantie de 95 p.c., on pourrait porter celle-ci jusqu'à une quasi certitude en reculant en conséquence, et toujours d'après le calcul, les bornes assignées précédemment à l'erreur. Par contre, se déciderait-on difficilement, et même ne se déciderait-on probablement pas du tout, si l'on ne pouvait faire mieux que nous donner le résultat de 75 p.c. de tout à l'heure avec une erreur possible de 20 p.c. en plus ou en moins. Et, s'il était possible, se déciderait-on encore moins s'il arrivait qu'il ne soit plus même question d'assigner une limite à l'erreur.

Cependant, avant d'aller plus loin dans l'exposé de ce dernier stade du travail, il devient à-propos de faire une mise au point qui devra s'appuyer sur quelques généralités.

* * *

L'expression «population» a été utilisée à diverses reprises et, chaque fois il faut l'admettre, en lui laissant la possibilité de

prendre un sens quelque peu relâché. Et le moment est ici venu de nous montrer plus strict à cet égard. Qu'il soit donc entendu, contrairement à ce qui pouvait être facilement supposé, que ce ne sont pas du tout les individus eux-mêmes que nous avons en vue lorsqu'il est question d'une population telle que celle des usagers par exemple, pas plus que ce ne sont les individus eux-mêmes que nous considérons avoir recueillis dans le cas d'un échantillon relatif à ceux-ci. Représentons-nous alors les choses de la façon que voici. À un moment donné, c'est-à-dire au jour de l'enquête si nous voulons rester dans le cadre de notre étude, il est toujours possible de faire correspondre un nombre à chaque usager et de le lui faire correspondre d'une manière univoque, pourvu que l'on se place à un point de vue bien défini: et ce point de vue pourrait être, par exemple, celui du nombre de personnes à la charge de l'usager. Il est manifeste que le nombre en question sera nécessairement un entier si l'on s'arrête à cette dernière supposition, puisqu'il s'agit d'un nombre de personnes; c'est-à-dire qu'il reviendra à chaque usager un entier et un seul pris nécessairement dans la suite 0, 1, 2, 3, . . . (0 correspondant au cas de l'usager sans aucune personne à charge).

Toujours avec la même supposition, celle de mettre à l'étude le nombre de personnes à charge, qu'on se représente avoir une liste de tous les usagers, établissant pour chacun sa propre identité, et permettant d'assigner un rang à chacun, allant depuis un premier usager jusqu'au dernier, et soit le $500{,}000^{\text{ème}}$ le rang de celui-ci. À cette liste correspondra d'emblée une suite d'entiers bien déterminée — les nombres respectifs de personnes à la charge de chacun — laquelle suite se présentera, par exemple, de la façon suivante:

(2) 0, 0, 3, 1, 0, 2, 2, 0, 0, 6, ... , 4.

Et c'est cette suite de 500,000 nombres, ou plus précisément est-ce cette suite de 500,000 mesures que nous avons effectivement en vue lorsque nous parlons de la population à l'étude, c'est-à-dire au moment où la caractéristique concernée se trouve être le nombre de personnes à charge. Évidemment, la caractéristique étudiée sur les usagers venant à changer, on aura chaque fois une autre population de mesures en rapport avec la nouvelle caractéristique (salaire, classe de profession, etc.).

Prélevons maintenant un échantillon à même la population qui vient d'être définie à l'instant, c'est-à-dire à même la population que forme la suite (2), et décidons que le contenu de cet échantillon sera de 3,000 par exemple, et nous obtiendrons, en fait d'échantillon, non pas un groupe d'individus, mais une suite de mesures, soit, par exemple, la suite de 3,000 entiers

(3) 2, 0, 0, 6, 1, 1, 0, 5, 2, ... , 0.

Et cet échantillon, tiré d'une population bien définie, on fera bien de se le représenter dans la perspective que voici.

Bien d'autres échantillons que celui-ci, tout aussi vraisemblables que lui, auraient pu être tirés de la même population totale, tout en prenant soin d'ailleurs d'appliquer chaque fois le même plan d'échantillonnage. Et tous les échantillons distincts et possibles réalisables dans ces conditions forment une population d'échantillons, tous de contenu 3,000, et notre échantillon, celui-là même que nous avons effectivement tiré, devra être regardé comme n'étant ni plus ni moins que l'un quelconque de ceux-ci. On imagine facilement que cette population d'échantillons doit être très nombreuse, et si nous employons le mot «population» pour parler de cette collection d'échantillons c'est encore avec l'idée que cette collection est au complet et que chacun de ses éléments, c'est-à-dire chaque échantillon la composant, est lui-même formé d'une suite de mesures et non pas d'individus.

Dans le but de simplifier la mise au point que nous comptons faire, nous allons toutefois supposer que les 3,000 éléments de chacun des échantillons possibles doivent être tirés directement de la population totale et par un dispositif assimilable à celui d'un jeu de hasard, c'est-à-dire que l'échantillonnage doit être effectué en un seul stade et les éléments ne doivent pas être tirés systématiquement — toutes dispositions contraires à notre plan d'échantillonnage. Empressons-nous de dire que ces divergences de procédés n'entraînent pas de différences fondamentales du point de vue qui nous occupe, tandis que nous y gagnons en simplicité pour l'exposé.

Imaginons que l'on ait la possibilité de calculer la somme des termes de la suite (2), laquelle se trouve être pour l'instant notre population totale. Il est certain que cette somme existe, indépendamment de toute tentative directe pour l'obtenir, et elle a

une interprétation manifeste, puisqu'elle est égale au nombre total des personnes à charge pour l'ensemble des usagers, comme il est facile de le justifier. D'ailleurs, on passerait volontiers de cette somme à son rapport au nombre total des usagers, obtenant ainsi la *moyenne* pour la population, prise dans son état présent. Sans dire que cette moyenne règle à elle seule tous les problèmes que présente la caractéristique à l'étude, il n'en reste pas moins qu'elle a une interprétation aisée. Imaginons que cette moyenne, dans le cas actuel, soit effectivement égale à 2.3; et ce n'est pas le fait que nous ne puissions guère entreprendre de la calculer directement qui l'empêche d'exister tout aussi bien, et d'être parfaitement définie. Dans le sens mathématique du mot, c'est une constante, à ne pas s'y tromper. Or, ne la connaissant pas, nous voulons l'estimer, et pour ce faire nous avons notre échantillon, celui que nous avons effectivement tiré, c'est-à-dire la suite (3) elle-même. À coup sûr, cet échantillon-ci a une moyenne que nous pouvons calculer — le rapport de la somme des termes, tous connus par les questionnaires, au nombre des termes, soit 3,000 — et nous trouvons pour moyenne, et par exemple, 2.1. C'est une estimation, la seule que nous aurons, de la vraie moyenne 2.3. Mais il est certain que nous aurions pu tout aussi bien tomber sur un autre échantillon que ce dernier, qui n'a d'autre mérite que de nous être échu, alors qu'ils sont des milliers d'échantillons tous également possibles. C'est ainsi qu'un autre échantillon que celui-ci nous étant échu, nous aurions pu trouver une moyenne de 2.4, et telle aurait été alors notre estimation de la vraie moyenne. Avec un autre, nous aurions pu trouver 2.3, et ainsi de suite.

Il existe donc en définitive toute une population de moyennes susceptibles d'avoir été calculées à partir des échantillons, qui ne sont pas nécessairement différentes entre elles, mais qui sont toutes également possibles, tout comme le sont les échantillons qui les fournissent. Et dans un cas pratique d'échantillonnage nous ne faisons que tomber sur l'une d'elles et sans l'avoir pu choisir, et c'est cette moyenne particulière qui nous sert d'estimation pour celle de la population totale.

Or, si nous voulons savoir à quoi nous en tenir lorsque nous avons en main l'une de ces moyennes, savoir à quoi nous en tenir quant à sa valeur en tant qu'estimation d'une constante appar-

tenant à la population totale, alors que nous sommes exposés à trouver n'importe laquelle de la collection, il est nécessaire que nous soyons pleinement renseignés sur la répartition de toutes ces moyennes. Car ce sont celles-ci qui posent un problème, et non la moyenne de la population totale, qui est fixe et sur laquelle il n'y a donc pas de question à se poser, autre que de reconnaître qu'elle ne nous est pas connue. Et si l'on veut continuer avec le même exemple, on pourra se représenter notre population de moyennes sous la forme de la suite que voici:

(4) 2.1, 2.4, 2.3, 2.4, ... , 2.08.

Étant entendu, quitte à nous répéter, qu'en tirant un échantillon de la population totale, nous ne faisons, en fait, que nous ramener à tirer au hasard l'une des moyennes de cette suite.

Disons que ce sont là, dans cette dernière suite, toutes les valeurs d'une certaine *variable*, et comme nous l'utilisons en prenant l'une de ses valeurs pour estimer une constante de la population totale, disons que c'est notre *variable d'estimation* en rapport avec cette constante. On saura tout d'abord que la moyenne des valeurs de la suite (4) est très précisément égale à la moyenne de la population totale, c'est-à-dire qu'elle est égale au nombre même que nous cherchons à estimer. Ce qui est un point important. Quant aux termes eux-mêmes, dont un seul nous servira d'estimation, ils s'écartent plus ou moins de cette valeur centrale qu'est la moyenne, et qui est celle que nous voulons atteindre — ou, comme l'on dit, elles sont plus ou moins *dispersées* par rapport à cette valeur. Et c'est la connaissance que nous pouvons obtenir, par le calcul, relativement à la dispersion des valeurs de la suite (4) qui fait que nous pouvons parler de cette notion utile entre toutes d'une erreur d'estimation.

Il est entendu déjà que la population définie par la suite (2) va changer aussitôt que nous passerons à l'étude d'une autre caractéristique des usagers, en dépit du fait que ceux-ci restent les mêmes — puisque, pour chacun d'eux, la mesure de la caractéristique à l'étude change avec celle-ci. Par exemple, imaginons qu'il soit proposé de s'occuper de la caractéristique intéressant la colonne (3) du questionnaire: avez-vous une auto? Avec cette nouvelle caractéristique, une autre population totale va être

substituée à celle de la suite (2), et elle sera formée, par exemple, de la suite de 500,000 nombres que voici:

(2′) 0, 0, 1, 0, 1, 1, 0, 0, ... , 1;

dans laquelle les entiers 0 et 1 pourront seuls figurer: 0 pour *non*, 1 pour *oui*. Et si l'on tire un échantillon de cette population, on aura la suite (3′) par exemple:

(3′) 1, 1, 0, 0, 0, 0, 1, ... , 0,

comparable à la suite (3) sauf que, eu égard à la caractéristique en question, les deux entiers 0 et 1 y figurent seuls.

La somme des termes de la suite (2′) a un sens immédiat: elle se trouve être égale au nombre total des autos que possède tout le groupe des usagers; de même que pour cette population de valeurs, la moyenne aura-t-elle un sens bien défini. Cette somme et cette moyenne existent à coup sûr, et sont l'une et l'autre parfaitement définies, en dépit du fait qu'il serait plus ou moins praticable de les calculer. Et voilà que l'échantillon (3′) pourra servir à les estimer l'une et l'autre dans des conditions tout à fait analogues à celles que l'on a rencontrées dans le cas précédent. D'ailleurs, cette suite (3′), puisqu'elle est supposée être celle qui a été effectivement tirée comme échantillon, est à se représenter comme provenant directement du dépouillement des questionnaires.

Jusqu'ici, nous n'avons considéré chaque usager qu'en fonction d'une seule des caractéristiques mises en jeu par le questionnaire, quitte à mettre celles-ci successivement à l'étude, comme on vient de le faire; mais on pourra s'intéresser tout aussi bien à une répartition des usagers lorsque chacun d'eux sera considéré dans les perspectives de plusieurs de ces caractéristiques prises simultanément et, évidemment, les mêmes pour tous. Pour prendre un exemple simple, imaginons d'étudier nos usagers du point de vue de la classification suivante: quel est le nombre des usagers qui, possédant une auto, ont en même temps deux personnes ou plus de deux personnes à charge? D'ailleurs, plutôt que de demander le nombre de ces usagers, on serait peut-être satisfait de connaître leur pourcentage par rapport à tous les usagers.

Si telle est la question à l'étude, notre population totale, formée comme il se doit de 500,000 nombres bien déterminés correspondant respectivement aux 500,000 usagers, se présentera sous l'aspect de la suite (2″), soit

(2″) 1, 1, 0, 0, 0, 1, 0, 1, ... , 1,

comparable à la suite (2′) en ce sens que les deux entiers 0 et 1 se trouvent y figurer seuls: l'entier 1 correspondant aux usagers se trouvant appartenir à la catégorie spécifiée, c'est-à-dire aux usagers ayant une auto et deux dépendants au moins, tandis qu'il reviendra un 0 à tout autre passager sans distinction de catégorie, du seul fait qu'il n'appartient pas à la catégorie précitée — tout aussi bien qu'il ait une auto avec un seul dépendant ou qu'il n'ait pas d'auto tout en ayant 5 dépendants. Pour la population totale, formée uniquement d'entiers 0 et 1, la somme de tous les termes a un sens évident, de même que la moyenne (obtenue en calculant le rapport de la somme au nombre de tous les termes): la première est sans nul doute égale au nombre de tous les usagers appartenant à la catégorie en question, et il est tout aussi manifeste que la seconde est précisément égale au pourcentage demandé. Cependant, ni l'un ni l'autre de ces deux nombres ne nous est connu ou ne serait susceptible de l'être par un comptage direct de la population. Par contre, un échantillon de cette même population est à notre portée, puisqu'on l'obtiendra tout simplement par un classement des questionnaires — ou plutôt classera-t-on les cartes perforées sur lesquelles ils ont été respectivement reportés — en tenant compte des deux caractéristiques déterminantes de la catégorie, et cet échantillon nous permettra de faire une estimation du pourcentage cherché, par exemple, tout comme dans les cas précédents. Et comme là, devra-t-on encore juger de la valeur de l'estimation à partir d'une connaissance de la répartition des pourcentages que pourraient fournir tous les échantillons possibles.

* * *

En tenant compte de ce qui vient d'être dit, on voit que les résultats de l'enquête pourront affecter des formes variées, pour autant que seront concernées les répartitions que l'on peut faire

des usagers en fonction des diverses caractéristiques. On pourra prendre une à une les caractéristiques couvertes par le questionnaire, et on pourra grouper celles-ci de diverses manières, ainsi qu'on a imaginé de le faire en tout dernier lieu. C'est ainsi que l'on saura interpréter celles de ces répartitions qui sont présentées ici à titre d'exemple (voir tableaux VII à XII), et dans lesquelles les titres, sous-titres et renvois indiquent suffisamment quelle est la nature de chacune des caractéristiques mises successivement à l'étude. Cependant, pour s'en tenir à une réserve judicieusement dictée, les résultats effectivement obtenus ont été dissimulés en

Tableau VII

Répartition des usagers du réseau de la C.T.M. selon le domicile

	Domiciliés à Montréal	Domiciliés hors Montréal	Total
Nombre des usagers....	—	—	—[1]
Pourcentages.....	50%	50%	100%

1. Avec une erreur d'estimation ne dépassant pas 1.5 p.c.

Tableau VIII

Répartition des usagers selon la condition d'utilisation du réseau

Condition d'utilisation du réseau	Domiciliés à Montréal	Domiciliés hors Montréal	Total
Régulièrement pour le travail[1].....	50%	50%	100%
Régulièrement pour sorties et magasins.....	50%	50%	100%
Occasionnellement[2].....	50%	50%	100%

1. Comprend les usagers qui utilisent le réseau pour sorties et magasins en plus de l'utiliser pour leur travail.

2. Comprend les touristes.

Tableau IX

Répartition des usagers selon la condition du travail

Condition du travail	Domiciliés à Montréal	Domiciliés hors Montréal	Ensemble des usagers
Femme mariée, ménagère[1]	20%	20%	20%
Femme exerçant une profession	20%	20%	20%
Homme exerçant une profession	20%	20%	20%
Étudiant, écolier	20%	20%	20%
Autres catégories[2]	20%	20%	20%
Total	100%	100%	100%

1. Comprend toute femme dont la condition est d'avoir soin du ménage pour soi-même.
2. Comprend les rentiers, pensionnés, chômeurs, touristes.

Tableau X

Répartition des usagers avec et sans auto et selon la condition d'utilisation du réseau

Condition d'utilisation du réseau	Domiciliés à Montréal			Ensemble des usagers		
	Sans auto	Avec auto	Total	Sans auto	Avec auto	Total
Régulièrement pour le travail[1]	50%	50%	100%	50%	50%	100%
Régulièrement pour sorties et magasins	50%	50%	100%	50%	50%	100%
Occasionnellement[2]	50%	50%	100%	50%	50%	100%

1. Comprend les usagers qui utilisent le réseau pour sorties et magasins en plus de l'utiliser pour leur travail.
2. Comprend les touristes.

Tableau XI

Répartition des usagers utilisant régulièrement le réseau pour leur travail, selon le nombre des personnes à charge, le domicile et lieu où s'exerce l'occupation rétribuée[1]

Nombre de personnes à charge	Domicile / lieu de l'occupation			Ensemble des usagers du réseau
	Montréal/ Montréal	Montréal/ hors Montréal	hors Montréal /Montréal	
0	14.3%	14.3%	14.3%	14.3%
1	14.3%	14.3%	14.3%	14.3%
2	14.3%	14.3%	14.3%	14.3%
3	14.3%	14.3%	14.3%	14.3%
4	14.3%	14.3%	14.3%	14.3%
5	14.3%	14.3%	14.3%	14.3%
6 et plus	14.2%	14.2%	14.2%	14.2%
Total	100%	100%	100%	100%

1. À l'exclusion des écoliers et étudiants.

attribuant arbitrairement un égal contenu à toutes les cellules disponibles.

* * *

Revenons maintenant à cette partie de l'enquête proprement dite ayant pour objet d'évaluer les erreurs qui peuvent affecter respectivement les diverses estimations fournies par l'échantillon — à défaut de quoi, avons-nous reconnu, aucune possibilité d'utilisation n'existe pour les dites estimations.

Non seulement cette question du degré d'approximation des résultats attendus de l'échantillon a-t-elle pris place dès le début de notre étude, mais encore devait-elle être à la base d'une décision immédiate et fondamentale: celle de fixer le contenu à donner à l'échantillon. On ne s'est en effet décidé en ce sens qu'après avoir pris pleine conscience de la marge d'erreur qui pouvait être compatible avec une responsabilité mesurée, advenant le cas où une décision administrative grave viendrait à s'appuyer sur un

résultat de l'enquête. Il serait cependant inexact de penser que l'on puisse compter entièrement sur la seule garantie qu'on s'était alors donnée. Certes les calculs qui ont abouti à se fixer un contenu d'échantillon tenaient compte des différents stades d'échantillonnage par lesquels il fallait passer, tenaient compte encore du fait que les éléments étaient tirés dans chaque échantillon successifs d'une manière «systématique»; mais cela n'empêche qu'ils étaient basés aussi sur des estimations à priori, telles celles qui

Tableau XII

Répartition des usagers selon la direction et le sens domicile/lieu où s'exerce l'occupation rétribuée, et distance moyenne parcourue[1]

Direction et sens domicile/lieu de l'occupation[2]	Domiciliés à Montréal		Ensemble des usagers	
	Pourcentage des usagers	Distance moyenne parcourue	Pourcentage des usagers	Distance moyenne parcourue
Est	12.5%	—	12.5%	—
Ouest	12.5%	—	12.5%	—
Nord	12.5%	—	12.5%	—
Sud	12.5%	—	12.5%	—
N-E	12.5%	—	12.5%	—
N-O	12.5%	—	12.5%	—
S-O	12.5%	—	12.5%	—
S-E	12.5%	—	12.5%	—
Total	100%	—[3]	100%	—[4]

1. La distance parcourue par un usager est celle qui sépare, dans un seul sens, son domicile du lieu de son travail, mesurée en tenant compte du tracé général en damier des rues de la ville de Montréal et autres localités desservies par le réseau. La direction domicile-lieu du travail est prise par rapport à l'orientation ouest-est des rues de la ville et le sens est celui du domicile vers le lieu du travail sur cette même direction. Les distances sont mesurées en milles et les directions données sont comprises, pour les points cardinaux, dans des angles de 20° de part et d'autre de la direction réputée vraie et pour les autres directions, dans des angles de 25° de part et d'autre de la direction vraie. Pour les usagers dont le domicile est situé hors d'une localité desservie par le réseau, il est réputé se trouver au point le plus rapproché que l'usager peut atteindre sur le réseau.

2. Partant de son domicile, l'usager se dirige vers le point de l'horizon qui est indiqué.

3. Moyenne pour les usagers domiciliés à Montréal.

4. Moyenne pour l'ensemble des usagers.

portaient notamment sur diverses *dispersions* et diverses *corrélations* possibles entre les éléments de la population à l'étude.

Et c'est cette incertitude qui fait que la question ne se trouve encore pas réglée: comme il a été établi, par l'exemple simplifié qui nous a servi à développer cette façon de se représenter les choses, rien de bon ne peut sortir d'un échantillon tant qu'une connaissance précise de la population d'échantillons de laquelle il est sorti n'a pas été acquise. Et c'est justement ce même échantillon que nous avons en main, et que nous allons pouvoir examiner à loisir, qui va nous servir à réviser notre connaissance de la dite population d'échantillons, basée jusqu'ici sur les estimations plus ou moins arbitraires dont il vient d'être parlé. Quant aux opérations que cela entraîne pour arriver à cette fin, elles appartiennent à l'enquête, logiquement et pécuniairement, au même titre que toutes les opérations décrites jusqu'ici. D'ailleurs, la population d'échantillons devra être examinée au mérite et indépendamment pour chacune des caractéristiques mises à l'étude (autos, personnes à charge, etc.), puisqu'à chacune de ces caractéristiques correspond une population totale qui lui est propre, et à chacune de ces dernières populations correspond une population particulière d'échantillons.

Comme on le voit, les calculs à effectuer en ce sens vont devoir prendre leur matière première à même l'échantillon qui a été tiré. De plus, disons tout de suite qu'afin de nous réserver la possibilité d'effectuer les dits calculs, ce n'est pas un échantillon unique qui a été prélevé, — unique comme nous l'avions laissé entendre au stade où la chose s'était passée, et cela pour fins de simplification de l'exposé — mais que ce sont dix échantillons en place d'un seul qui ont été prélevés. Toutefois, faut-il reconnaître que chacun de ceux-ci a été tiré exactement de la façon qui a été décrite pour celui qui avait retenu seul notre attention. Ces dix échantillons, ou plutôt ces dix sous-échantillons, ont conservé leur individualité tout au cours des tabulations, cependant que les estimations relatives aux caractéristiques de la population des usagers ont été tirées de l'échantillon unique formé par leur réunion en un seul. Quant au calcul même de l'erreur d'estimation, pour une caractéristique donnée, il a été mené de la façon suivante. Il s'agissait d'abord d'obtenir une estimation,

par le calcul, de la *dispersion* de la répartition de la variable d'estimation, alors que cette dernière est basée sur la population des sous-échantillons. L'estimation était faite à partir des dix observations fournies par les dix sous-échantillons. De là, il fallait passer à la *dispersion* de la répartition de la variable d'estimation, celle qui se trouve utilisée en définitive, c'est-à-dire lorsque cette dernière est basée sur la population des échantillons tels que notre échantillon unique. Les mêmes calculs étaient à reprendre relativement à chacune des caractéristiques essentielles de la population des usagers, — nombre des usagers, répartition des usagers selon le nombre de dépendants, etc. — pour lesquelles une estimation était donnée.

Les erreurs d'estimation, calculées comme il vient d'être dit, ont été produites à part des tableaux de répartitions.

Pour terminer, disons qu'en esquivant les difficultés fondamentales au cours de cet exposé point n'était en vue de créer une impression de fausse facilité. S'il est un domaine où les arguments spécieux jouissent d'un crédit que les arguments justes sont loin de connaître, c'est bien celui de la statistique dans ses applications aux choses de la vie courante. Et c'est pourquoi il n'est peut-être pas hors de propos, à la suite de cet exposé forcément simplifié, de parler d'une mise en garde — certes qui n'a pas la prétention d'être à l'adresse de tous — d'une mise en garde contre les procédés dénués de bases bien établies. Et ces bases sont les mathématiques, en général, et le calcul des probabilités, en particulier. À défaut de quoi, pour ce qui est de trouver un nombre, on en trouvera un à coup sûr; mais pour s'en montrer satisfait quel que soit le procédé qui l'a engendré, il faudrait se ranger à l'avis de celui qui disait «qu'il vaut mieux un nombre, si faux soit-il, que pas de nombre du tout».

Robert STOCK,
professeur à l'École des Hautes Études commerciales (Montréal).