

L'enseignement de la statistique appliquée: Différencier échantillon et population

Roch Ouellet et Robert Latour H.E.C.

De nombreuses expériences en psychologie ont montré que les gens identifient un échantillon avec la population dont il a été tiré. De plus, les gens ont tendance à généraliser à la population tout entière les propriétés observées dans l'échantillon: on croit, en effet, que l'échantillon, étant aléatoire et «représentatif», doit refléter fidèlement la population et ne s'en différencier sous aucun aspect essentiel. Ainsi, les journalistes comparent les sondages successifs portant sur les intentions de vote et commentent savamment des écarts de l'ordre de 2%, ou même de 1% parfois.

Cette confusion entre échantillon et population entraîne bien des difficultés pour le professeur de statistique appliquée. Celui-ci décrit, en effet, les liens probabilistes qui relient l'échantillon et la population-mère, puis indique quelles conclusions peuvent être valablement généralisées ou inférées de l'échantillon à la population. Mais comment l'étudiant peut-il comprendre la raison d'être de tous ces raisonnements abstraits et complexes lorsqu'il n'a pas encore saisi clairement le fait qu'un échantillon, même représentatif, peut différer notablement de la population sous bien des aspects ? Il importe donc que le professeur mette en évidence les faits suivants :

- l'échantillon obtenu est l'un des innombrables échantillons qui étaient possibles à priori;
- deux échantillons aléatoires peuvent différer sensiblement, surtout si leurs tailles sont petites;
- et, par conséquent, la population ne saurait ressembler à tous les échantillons.

Nous utilisons divers trucs pour transmettre ces principes. Par exemple, nous énumérons tous les échantillons que l'on peut tirer d'une population de très petite taille. De même, nous fournissons à chaque étudiant un échantillon individualisé. Au début de la session, l'étudiant reçoit une feuille décrivant une population, ainsi que la liste des numéros des individus composant son échantillon (voir figure 2). Il doit d'abord chercher les valeurs x_1, \dots, x_n associées à ces individus, puis calculer diverses statistiques. Ultérieurement, on lui demande de déterminer, à partir de son échantillon, les bornes d'intervalles de confiance pour la moyenne μ de la population. Nous compilons en classe les bornes ainsi obtenues, puis nous donnons la valeur exacte de μ et remarquons que cette valeur n'appartient pas à certains intervalles. (Nous prenons soin de fixer le niveau de confiance à 90% et de choisir les échantillons de façon à ce que «par hasard» un ou plusieurs intervalles ne contiennent pas μ .) Les étudiants constatent alors que les bornes varient d'un échantillon à l'autre, que le fait d'appliquer correctement les formules ne garantit pas que la valeur réelle du paramètre sera comprise entre les bornes. Nous insistons alors sur le fait que tous les échantillons ont été tirés de façon aléatoire. Nous enchaînons avec le commentaire suivant : dans un contexte réel, on dispose d'un seul échantillon et il est impossible de savoir s'il

donne une information correcte ou tendancieuse. Par conséquent, toute décision basée sur un échantillon contient inéluctablement un élément de risque.

La préparation d'échantillons individuels pour chaque étudiant exige un logiciel de simulation. Il en existe beaucoup mais aucun ne présente toutes les caractéristiques que nous recherchions, soit souplesse, facilité d'utilisation pour un professeur néophyte en informatique, qualité dans la présentation des documents remis aux étudiants.

Nous avons donc développé un logiciel original, nommé EchInd, qui fonctionne dans l'environnement PC. EchInd est interactif et se présente comme une suite de menus, l'utilisateur choisissant parmi les options énumérées à l'écran ou fournissant les paramètres demandés. (La figure 1 reproduit le menu principal.) L'utilisateur type est un professeur modérément ou peu versé en informatique, qui n'est pas intéressé à investir beaucoup de temps dans l'apprentissage d'un logiciel sophistiqué de simulation mais qui désire préparer et imprimer des documents individualisés à distribuer aux étudiants.

EchInd permet de créer et d'imprimer une population, des échantillons, ou encore la distribution d'une statistique. En particulier, EchInd génère et imprime des échantillons d'une taille donnée, tirés d'une population déjà existante sous format EchInd. Les sorties produites sont de 6 types :

- échantillons individualisés à l'aveugle (voir figure 2);
- échantillons individualisés complétés (voir figure 3);
- énumération des statistiques des différents échantillons (voir figure 4);
- intervalles de confiance individualisés à l'aveugle (voir figure 5);
- intervalles de confiance individualisés complétés;
- énumération des bornes des intervalles à un niveau de confiance spécifié par l'utilisateur.

Les échantillons et intervalles à l'aveugle contiennent des champs en blanc qui seront complétés par l'étudiant. Les versions complètes permettent à l'étudiant de se corriger lui-même. Quant au professeur, il utilisera généralement les énumérations qui présentent les résultats-clés sous forme compacte.

Nous avons apporté beaucoup de soins à rendre les sorties faciles et agréables à utiliser. Les sorties à l'aveugle que doit compléter l'étudiant sont présentées de façon à laisser suffisamment d'espace pour les réponses. Les listes-résumés ramassent en une même page les résultats associés à une quarantaine d'échantillons et facilitent la correction.

EchInd est distribué par les auteurs à un coût nominal de 10\$. La reproduction du logiciel et de sa documentation est autorisée, en autant que la source soit mentionnée.

Énumération des statistiques des échantillons

Nom du fichier-échantillon: LABRANCH.

La pâtisserie des frères Labranche
Laporte & Ouellet section 5.3 problème # 7
I-604 Automne 1990

No	Semence	Min	Max	X-barre	S-carré	S
1	369220	331	446	380.64	887.06	29.78
2	064694081	333	438	383.48	799.08	28.27
3	476030039	321	475	389.36	1699.67	41.23
4	929826793	317	449	377.32	1183.89	34.41
5	152620959	343	429	385.32	563.39	23.74
6	760429449	322	456	389.12	1396.03	37.36
7	912048263	321	475	387.48	1236.33	35.16
8	864219849	344	420	384.76	508.52	22.55
9	640892223	344	446	388.96	890.96	29.85
10	664448905	349	449	392.52	714.60	26.73
11	268299487	297	475	382.60	1227.50	35.04
12	368493905	329	434	376.96	520.63	22.82
13	120965487	342	429	381.04	654.45	25.58
14	867505905	309	437	389.44	1001.84	31.65
15	616263487	325	475	389.48	1443.42	37.99
16	232583905	333	451	392.84	951.31	30.84
17	314577487	327	456	387.48	1179.42	34.34
18	204008905	294	415	377.24	988.20	31.44
19	372155487	321	439	387.40	740.84	27.22
20	193750905	348	475	386.36	867.75	29.46
21	082696487	304	444	379.52	1095.19	33.09
22	526925905	339	439	390.44	624.17	24.98
23	411951487	341	432	380.88	543.19	23.31
24	799644905	354	445	395.32	585.30	24.19
25	494286487	345	437	391.64	739.48	27.19
26	351485905	309	442	384.52	1151.27	33.93
27	868869487	331	453	388.92	753.48	27.45
28	356297905	342	482	396.80	1033.26	32.14
29	924072487	347	433	399.36	657.67	25.65
30	728860905	340	436	379.64	605.31	24.60
31	688053487	328	412	375.32	655.30	25.60
32	999346905	262	468	393.28	1823.04	42.70
33	182776487	321	427	381.84	901.81	30.03
34	202999905	329	454	385.16	1092.47	33.05
35	640579487	320	450	382.52	994.35	31.53
36	010148905	320	475	388.80	1136.76	33.72
37	014167487	342	446	391.80	747.51	27.34
38	445147905	333	453	389.72	1234.63	35.14
39	864474487	325	429	378.20	637.07	25.24
40	542853905	328	438	384.64	761.98	27.60

Figure 4

Echantillon no 21 Semence 082696487

La pâtisserie Ges frères Labranche
Laporte & Ouellet section 5.3 problème # 7
I-604 Automne 1990

Valeurs X de l'échantillon

413	370	375	371	437	404	444	386	358	357
372	382	373	396	402	304	318	362	352	428
387	401	355	350	391					

Minimum: 304 Maximum: 444
Moyenne: 380
Ecart type: 33 Variance: 1095

INTERVALLE DE CONFIANCE SYMETRIQUE POUR LA MOYENNE DE LA POPULATION

Niveau de confiance	Borne inférieure	Borne supérieure
0.80	_____	_____
0.90	_____	_____
0.95	_____	_____
0.99	_____	_____

INTERVALLE DE CONFIANCE SYMETRIQUE POUR LA VARIANCE DE LA POPULATION

Niveau de confiance	Borne inférieure	Borne supérieure
0.80	_____	_____
0.90	_____	_____
0.95	_____	_____
0.99	_____	_____

Figure 5

Pour toute information, communiquer avec

Roch Ouellet
Méthodes quantitatives
École des H.E.C.
5255, Decelles
Montréal, Qc
H3T 1V6
Tél.: (514) 340-6481

GAGNANTS DES CONCOURS MATHÉMATIQUES

Félicitations à **Christian Dubuc** de la polyvalente J.-H. Leclerc de Granby,
grand gagnant du concours secondaire.
Félicitations également à **Charles Patrick Dugas** du Collège Mérici qui a remporté la palme au collégial.
Les gagnants se méritent un prix de 250\$ de l'AMQ.

VACANCES

Nous profitons de ce dernier numéro avant les vacances pour souhaiter à chacune et chacun une période de ressourcement et de repos bien méritée.
N'oubliez pas cependant de mettre par écrit vos réflexions pour alimenter nos lectures de l'automne.

Date de tombée: 31 août 1991