

## Statistiques

### Exposés

A. Les fabricants d'ordinateurs d'un pays produisent essentiellement deux classes d'ordinateurs. Une enquête annuelle sur les prix dans cette branche porte sur deux modèles de chacune de ces deux classes (voir tableau ci-dessous). Le modèle 1a n'étant plus fabriqué après 2002, ce modèle a été remplacé dans l'échantillon de l'enquête par une version légèrement perfectionnée, appelée *Modèle 1a+*. Les résultats de l'enquête de 2000 à 2002 sont les suivants :

*Prix (en dollars des États-Unis) des modèles d'ordinateur inclus dans l'enquête*

	2000	2001	2002
<b>Ordinateurs de la classe 1</b>			
Modèle 1a	1 500	1 450	
Modèle 1a+			1 350
Modèle 1b	1 200	1 180	1 180
<b>Ordinateurs de la classe 2</b>			
Modèle 2a	425	400	410
Modèle 2b	540	550	550

*Nombre estimatif d'ordinateurs vendus (en milliers)*

	2000	2001	2002
<b>Ordinateurs de la classe 1</b>			
Modèle 1a	20	35	
Modèle 1a+			40
Modèle 1b	10	15	15
<b>Ordinateurs de la classe 2</b>			
Modèle 2a	360	350	370
Modèle 2b	140	140	130

*Caractéristiques des modèles de classe 1 inclus dans l'enquête*

	Modèle 1a	Modèle 1a+	Modèle 1b
Disque dur	40 GB	80 GB	80 GB
Millions d'opérations en virgule flottante/seconde	201	201	185

1) La formule de Laspeyres et celle de Paasche pour l'indice-chaîne des prix sont les suivantes (p étant le prix et q la quantité) :

$$P_{t-1,t}(q_{t-1}) = \frac{\sum p_t q_{t-1}}{\sum p_{t-1} q_{t-1}} \quad \text{Laspeyres} \qquad P_{t-1,t}(q_t) = \frac{\sum p_t q_t}{\sum p_{t-1} q_t} \quad \text{Paasche}$$

- (a) Calculer les indices-chaînes de Laspeyres et de Paasche pour chacune des classes d'ordinateur (en utilisant toujours la période précédente comme année de base).
- (b) Expliquer quelles hypothèses vous avez retenues pour obtenir vos résultats, surtout en ce qui concerne le remplacement du *Modèle 1a* par le *Modèle 1a+*.

2) Le bureau national de statistique collecte également en continu des informations sur la production de chacune des classes d'ordinateurs :

<i>Parts de production</i>			
	2000	2001	2002
Ordinateurs de classe 1	0,15	0,25	0,23
Ordinateurs de classe 2	0,85	0,75	0,77

- (a) Donner les indices-chaînes composites de Laspeyres et de Paasche des prix des ordinateurs (i.e. les indices généraux des prix pour l'ensemble de la branche, prenant la période précédente comme année de base pour l'indice courant des prix).
  - (b) D'après chacune des deux formules d'indice-chaîne, quelle a été l'augmentation du niveau général des prix des ordinateurs en 2002 par rapport à 2000 ?
  - (c) Comment peut-on expliquer les éventuelles différences entre les résultats obtenus par la formule de Laspeyres et par celle de Paasche ?
- 3) À supposer qu'on ait à interpréter une série chronologique de chiffres de production ou de ventes d'ordinateurs, pourquoi la série chronologique correspondante d'indices des prix serait-elle utile ? Expliquer.

B. Vous êtes chargé de bâtir un questionnaire pour une enquête sur le logement qui sera réalisée en porte-à-porte. Vos collègues vous rappellent qu'il arrive souvent que les interviewés ne comprennent pas ce que les enquêteurs veulent savoir.

1) Citez cinq raisons courantes (et cinq seulement) pour lesquelles on peut manquer de se faire comprendre. Que pouvez-vous faire pour éviter de commettre ces erreurs?

On vous demande de construire l'échantillon de l'enquête en veillant à ce qu'il soit représentatif de la population. Un de vos collègues vous suggère d'opter pour un échantillonnage par grappes, un autre vous conseille un échantillonnage stratifié.

4) Quels sont les effets du choix d'une de ces deux solutions, par exemple sur l'erreur-type? Opposez ces deux méthodes d'échantillonnage à celle de l'échantillonnage aléatoire simple.

## Questions

1. On pose

$$f_{Y|X}(y|x) = \begin{cases} 1 & \text{si } x < y < x+1 \\ 0 & \text{dans les autres cas} \end{cases}$$

et on suppose que  $X$  suit une distribution uniforme  $(0,1)$

Calculez :

- a)  $E(Y)$
- b)  $f_{Y|X}(x|y)$
- c)  $P(X+Y < 1)$

2. Définissez les termes moyenne, mode et médiane et montrez sur un diagramme comment ils sont normalement liés entre eux dans une distribution dissymétrique à droite.

3. Un récipient contient quatre billets de loterie portant les numéros 111, 221, 212 et 122. On tire un billet au hasard et  $A_i$  est l'événement où 2 est à la  $i$ ème place,  $i = 1, 2, 3$ . Déterminer si  $A_1$ ,  $A_2$  et  $A_3$  sont indépendants.

4. Définissez deux moyens de mesurer les taux de croissance de séries temporelles et analysez leurs avantages et inconvénients respectifs.

5. Définissez avec précision la différence, dans une enquête par sondage, entre erreur systématique et erreur aléatoire.

6. 80 hommes et 40 femmes sont inscrits à un cours d'université. Il y a parmi eux 36 fumeurs et 84 non fumeurs. Quelle est la probabilité qu'une personne prise au hasard dans ce groupe soit un fumeur (et non une fumeuse), si :

- a. Le pourcentage de fumeurs est le même parmi les étudiants et parmi les étudiantes (c'est-à-dire qu'il n'y a aucun lien entre le sexe et le fait de fumer)?
- b. Le pourcentage de fumeurs parmi les étudiants est le double de ce qu'il est parmi les étudiantes (c'est-à-dire que la probabilité d'être fumeur est deux fois plus forte pour un étudiant que pour une étudiante)?

7. Pour être recruté comme militaire, il faut subir un test de dépistage des toxicomanies. Sur un groupe de 25 candidats, il y en a un dont le test est positif. L'intéressé prétend ne jamais s'être drogué. On sait qu'il a été établi empiriquement que le matériel utilisé par le laboratoire est capable de détecter les substances interdites dans 99 % des cas où elles sont présentes. D'autre part, le fabricant reconnaît que dans 5 % des cas, l'appareil détecte par erreur des substances qui ne sont pas présentes dans l'échantillon analysé. On sait enfin qu'au moins 90 % des recrues des forces armées n'ont jamais utilisé de drogues. Que peut-on dire, si on peut dire quelque chose, des dénégations du candidat?

8. Le tableau ci-après donne le nombre d'élèves jugés « excellents » (ayant reçu la note maximum à l'examen) et le nombre de ceux jugés « médiocres » (ayant eu la moyenne ou une note inférieure à la moyenne) par trois professeurs,  $M^{\text{mes}} X, Y$  et  $Z$ , qui ont corrigé la même épreuve.

Testez l'hypothèse selon laquelle le nombre d'élèves « médiocres » est le même pour chacun des trois professeurs (indication : testez l'hypothèse selon laquelle le fait d'être jugé « excellent » ou « médiocre » est indépendant du professeur par lequel on a été noté).

	M <sup>me</sup> X	M <sup>me</sup> Y	M <sup>me</sup> Z	Total
Excellent	50	47	56	153
Médiocre	5	14	8	27
<b>Total</b>	<b>55</b>	<b>61</b>	<b>64</b>	<b>180</b>

9. Sur les 50 passagers d'un autobus, 1/5 est contre l'interdiction de fumer à bord. Si on choisit au hasard 15 de ces passagers, sans remplacement, a) quelle est la probabilité d'en choisir exactement 5 qui sont contre l'interdiction? b) quelle est l'espérance mathématique du nombre de passagers de cet échantillon qui sont contre l'interdiction?

#### Sujets à étudier

Les lectures conseillées sont par exemple des manuels de base de statistique sur les domaines suivants :

- théorie des probabilités
- statistique descriptive
- statistique déductive
- statistiques économiques
- statistiques démographiques et sociales