

EXAMEN FINAL AUTOMNE 2006

Date : Dimanche 17 décembre 2006, de 14h00 à 17h00

INSTRUCTIONS

1. Détachez la feuille-réponses à la fin de ce cahier et inscrivez-y *immédiatement* votre nom, votre code permanent et votre numéro de groupe.
2. Seule la feuille-réponses doit être remise. Vous y inscrirez vos réponses sous la forme d'une lettre majuscule correspondant à votre choix.
3. Tout texte de référence (manuel, notes de cours, notes personnelles, etc.) est interdit. **Tout cas de plagiat ou de fraude sera soumis au Comité de discipline.**
4. L'usage d'une calculatrice est autorisé.
5. L'étudiant doit placer sa carte d'étudiant (avec photo) sur la table et signer la feuille de présence lors de la remise de sa feuille-réponses.
6. Pas de téléphone cellulaire sur la table.

Question 1 [10 × 1 points]

On tire un échantillon aléatoire simple de n ménages dans un quartier de N ménages.

Dans chacun des cas énumérés dans la 2^e colonne (*Paramètre à estimer*) du tableau ci-dessous, identifier le paramètre qu'il s'agit d'estimer. Faites votre choix de réponse dans la liste de gauche.

<i>Liste des réponses possibles</i>	<i>Paramètre à estimer</i> [répondre par A, B, ..., ou H]
A: La moyenne μ_y d'une certaine variable Y	1-a) Le nombre moyen de personnes dans les ménages francophones
B: La moyenne μ_d d'une certaine variable Y dans un domaine \mathcal{D} .	1-b) La proportion d'enfants parmi les habitants du quartier
C: Le total τ_d d'une certaine variable Y dans un domaine \mathcal{D} .	1-c) Le nombre de ménages francophones
D: Le total τ_y d'une certaine variable Y	1-d) La proportion de ménages sans enfants
E: Le nombre N_c d'unités appartenant à une certaine classe \mathcal{C}	1-e) Le nombre d'enfants qui possèdent un ordinateur personnel
F: Le quotient $R = \mu_Y/\mu_X$ de deux variables Y et X	1-f) Le nombre d'adultes francophones
G: La proportion p d'unités appartenant à une certaine classe	1-g) Le nombre d'ordinateurs possédés entièrement par des enfants
H: Aucun des paramètres ci-dessus	1-h) Le nombre total d'ordinateurs dans les ménages francophones
	1-i) Le nombre moyen d'ordinateurs par personne
	1-j) Le nombre moyen de chambres à coucher dans les ménages sans enfants

Question 2 [4+4+4+4+5+5 points]

On tire un échantillon de **26** ménages d'un quartier comptant **800** ménages. Pour chaque ménage on prélève l'information suivante

- Le nombre d'enfants dans le ménage
- Le nombre de personnes dans le ménage, incluant les enfants
- Le nombre de voitures possédées par les membres du ménage

On sait que la population du quartier est de **3040** habitants.

Les données de l'échantillon sont présentées en annexe

[Choisissez vos réponses parmi celles proposées au bas de la page]

- 2-a) Estimer la proportion de ménages sans voiture.
- 2-b) Estimer la proportion d'enfants.
- 2-c) Estimer le nombre total d'enfants en utilisant l'estimateur par la différence et le fait que le nombre d'habitants dans le quartier est 3040.
- 2-d) Estimer le nombre total de voitures en utilisant l'estimateur par le quotient et le fait que le nombre d'habitants dans le quartier est 3040.
- 2-e) Estimer l'écart type de l'estimateur en 2-a).
- 2-f) Estimer l'écart type de l'estimateur en 2-b).
- 2-g) Estimer l'écart type de l'estimateur en 2-c).

Question 3 [9+4+4+4+4 points]

Mêmes données qu'au numéro 2.

- 3-a) Estimer la taille de l'échantillon qu'il aurait fallu prélever pour estimer (par la moyenne) le nombre total d'enfants avec une marge d'erreur de 10 %.
- 3-b) Estimer la proportion d'enfants en utilisant le fait que le nombre d'habitants dans la population est 3040.
- 3-c) Estimer l'écart type de l'estimateur en 3-b)
- 3-d) Estimer le nombre total d'enfants dans les ménages sans voiture
- 3-e) Estimer l'écart type de l'estimateur en 3-d)

[Choisissez vos réponses parmi celles proposées au bas de la page]

Choix de réponses [pour les questions 2 et 3. Choisir l'intervalle qui contient votre réponse]

A 0,01 – 0,03	B 0,044 -0,048	C 0,061 -0,075	D 0,08 -0,09	E 0,14 -0,16	F 0,61 – 0,63
G 0,64 – 0,66	H 9 -22	I 25 -28	J 40 -60	K 64 -66	L 90 -110
M 120 - 122	N 131 -134	O 155 -159	P 245 - 248	Q 480 - 486	R 758 -762
S 798 -802	T 1838 -1842	U 1898 - 1902	V 1998 - 2002	W 2100 - 2500	X Aucune des réponses ci-dessus

Question 4 [5×2 points]

Pour chacune des descriptions suivantes, dire de quel mode d'échantillonnage il s'agit. Choisir une réponse parmi les suivantes:

A: aléatoire simple**B:** stratifié**C:** systématique**D:** par grappes avec probabilités de sélection égales **E:** par grappes avec probabilités de sélection inégales.

4-a) Population: l'ensemble des formulaires soumis aux autorités douanières par des passagers arrivant à un aéroport durant l'année 2005. On tire au hasard 20 jours parmi les 365 de l'année. L'échantillon est l'ensemble des formulaires soumis ces jours-là.
4-b) Population: l'ensemble des PME d'un pays. On divise le pays en 15 régions; puis on tire 25 PME dans chaque région.
4-c) Population: l'ensemble des employés d'une compagnie. On tire au hasard 15 des succursales de la compagnie. L'échantillon est l'ensemble des employés des 15 succursales sélectionnées.
4-d) Population: l'ensemble des agents voyageurs d'une compagnie. On tire 10 comptes de voyage parmi tous ceux fournis par les agents. L'échantillon est l'ensemble des agents ayant soumis les comptes sélectionnés.
4-e) Population : l'ensemble des personnes ayant assisté à une pièce de théâtre. À partir d'un moment choisi au hasard peu après la fin de la pièce, on interroge chaque 10 ^e personne sortant du théâtre.

Question 5 [5 points]

On doit tirer un échantillon de taille n d'une population divisée en trois strates. Laquelle ou lesquelles des affirmations suivantes sont vraies?

P_1 : Si les variances (corrigées) S_h^2 des strates sont toutes égales, l'allocation proportionnelle est optimale

P_2 : Si les tailles N_h des strates sont toutes égales, l'allocation proportionnelle est optimale

P_3 : Un échantillon stratifié est toujours plus précis qu'un échantillon aléatoire simple, quelle que soit l'allocation

[Choisir votre réponse parmi celles présentées au bas de la page]

Question 6 [5 points]

On doit tirer un échantillon aléatoire simple de taille n afin d'estimer la moyenne μ de la population d'une variable Y . Laquelle ou lesquelles des affirmations suivantes sont vraies?

P_1 : Plus n est grand, plus l'écart-type de l'estimateur sera grand

P_2 : Une estimation par la différence sera d'autant meilleure que la variable auxiliaire est indépendante de Y .

P_3 : Une estimation par le quotient est sans biais, quelle que soit la taille de l'échantillon.

[Choisir votre réponse parmi celles présentées au bas de la page]

Question 7 [5 points]

Une compagnie de marketing qui a engagé récemment 200 vendeurs se demande quels facteurs déterminent le succès de ces vendeurs. L'un des facteurs considérés est l'âge. Afin de vérifier si l'âge est un facteur déterminant, elle classe les 200 vendeurs selon deux variables : leur catégorie d'âge (30 ans ou moins; entre 30 et 40 ans; 40 ans ou plus) et leur succès (c'est-à-dire, s'ils ont atteint leur quota ou non). Lequel ou lesquels des énoncés ci-dessous pourraient être posés comme hypothèse nulle?

P_1 : Le pourcentage de vendeurs qui réussissent est le même dans les trois groupes d'âge

P_2 : Le pourcentage de vendeurs qui réussissent diffère d'un groupe d'âge à l'autre.

P_3 : Le succès d'un vendeur ne dépend pas de son âge.

[Choisir votre réponse parmi celles présentées au bas de la page]

Choix de réponses [pour les questions 5, 6 et 7]

A : Aucune	B : P_1 seulement	C : P_2 seulement	D : P_3 seulement
E : P_1 et P_2 seulement	F : P_1 et P_3 seulement	G : P_2 et P_3 seulement	H : P_1, P_2 et P_3

Question 8 [5+5 points]

Mêmes données qu'au numéro précédent.

a) Supposons qu'on effectue un test d'indépendance entre les deux variables et qu'on trouve une valeur de χ^2 supérieure au point critique. Laquelle ou lesquelles des conclusions suivantes sont alors justifiées?
 [Choisir votre réponse parmi les combinaisons présentées au bas de la page]

C_1 : On peut conclure avec confiance que l'âge n'influence pas le succès

C_2 : On peut conclure avec confiance que l'âge influence le succès

C_3 : On ne peut pas conclure avec confiance que l'âge influence le succès

C_4 : On peut affirmer avec confiance que le pourcentage de vendeurs qui réussissent est le même dans les trois groupes d'âge

C_5 : On peut affirmer avec confiance que le pourcentage de vendeurs qui réussissent n'est pas le même dans tous les groupes d'âge

b) Supposons qu'on effectue un test d'indépendance entre les deux variables et qu'on trouve une valeur de χ^2 inférieure au point critique. Laquelle ou lesquelles des conclusions suivantes sont alors justifiées?
 [Choisir votre réponse parmi les combinaisons présentées au bas de la page]

C_1 : On peut conclure avec confiance que l'âge n'influence pas le succès

C_2 : On peut conclure avec confiance que l'âge influence le succès

C_3 : On ne peut pas conclure avec confiance que l'âge influence le succès

C_4 : On peut affirmer avec confiance que le pourcentage de vendeurs qui réussissent est le même dans les trois groupes d'âge

C_5 : On peut affirmer avec confiance que le pourcentage de vendeurs qui réussissent n'est pas le même dans tous les groupes

Choix de réponses [pour les questions 8-a) et 8-b)]

Z : Aucune	A : C_1 seulement	B : C_2 seulement	C : C_3 seulement
D : C_4 seulement	E : C_5 seulement	F : C_1 et C_2 seulement	G : C_1 et C_3 seulement
H : C_1 et C_4 seulement	I : C_1 et C_5 seulement	J : C_2 et C_3 seulement	K : C_2 et C_4 seulement
L : C_2 et C_5 seulement	M : C_3 et C_4 seulement	N : C_3 et C_5 seulement	O : C_4 et C_5 seulement

Annexe

Données sur un échantillon de taille **26** tiré d'une population des **800** ménages d'un quartier. Le nombre d'habitants du quartier est **3040**.

#	Nombre de personnes (enfants inclus)	Nombre d'enfants	Nombre de voitures
1	4	3	0
2	4	2	0
3	3	2	0
4	3	1	0
5	9	7	1
6	6	5	1
7	6	4	1
8	4	4	1
9	5	3	1
10	5	2	1
11	4	2	1
12	4	2	1
13	2	1	1
14	2	0	1
15	2	0	1
16	5	4	1
17	4	4	1
18	5	3	1
19	3	3	1
20	5	2	1
21	4	2	1
22	3	1	1
23	2	0	1
24	6	5	2
25	3	3	2
26	1	0	3
Sommes	104	65	26
Variances (corrigées) s^2	2,88	3,06	0,4
Covariance (Personnes enfants)	2,6		

Formulaire MAT2080 Examen final

Résumé des paramètres, leur estimateur, l'écart-type de l'estimateur, et l'estimateur de l'écart-type de l'estimateur.

Paramètre	Estimateur	Écart-type de l'estimateur	Estimateur de l'écart-type de l'estimateur
Moyenne μ	\bar{y}	$\sigma_{\bar{y}} = \sqrt{1-f} \frac{S}{\sqrt{n}}$	$\hat{\sigma}_{\bar{y}} = \sqrt{1-f} \frac{s}{\sqrt{n}}$
Proportion p	$\hat{p} = \frac{X}{n}$	$\sigma_{\hat{p}} = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}}$	$\hat{\sigma}_{\hat{p}} = \sqrt{1-f} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$
Un quotient $R = \frac{\mu_y}{\mu_x}$	$\hat{R} = \frac{\bar{y}}{\bar{x}}$	$\sigma_{\hat{R}} \approx \frac{\sqrt{1-f}}{\mu_x} \frac{\sqrt{S_y^2 + R^2 S_x^2 - 2RS_{xy}}}{\sqrt{n}}$	$\hat{\sigma}_{\hat{R}} = \frac{\sqrt{1-f}}{\bar{x}} \frac{\sqrt{s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}}}{\sqrt{n}}$
Moyenne μ Estimation par la différence	$\hat{\mu}_{yd} = \mu_x + (\bar{y} - \bar{x})$	$\sigma_{\hat{\mu}_{yd}} = \sqrt{1-f} \frac{\sqrt{S_y^2 + S_x^2 - 2S_{xy}}}{\sqrt{n}}$	$\hat{\sigma}_{\hat{\mu}_{yd}} = \sqrt{1-f} \frac{\sqrt{s_y^2 + s_x^2 - 2s_{xy}}}{\sqrt{n}}$
Moyenne μ Estimation par le quotient	$\hat{\mu}_{yq} = \mu_x \hat{R}$	$\sqrt{1-f} \frac{\sqrt{S_y^2 + R^2 S_x^2 - 2RS_{xy}}}{\sqrt{n}}$	$\sqrt{1-f} \frac{\sqrt{s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}}}{\sqrt{n}}$
Moyenne μ_d d'un domaine \mathfrak{D}	\bar{y}_d : Moyenne du domaine dans l'échantillon		$\sqrt{1-\frac{n_d}{N_d}} \frac{s_d}{\sqrt{n_d}}$ ou $\sqrt{1-\frac{n}{N}} \frac{s_d}{\sqrt{n_d}}$ selon que N_d est connu ou pas
Total $\tau_d = N_d \mu_d$ d'un domaine (N_d connu)	$T_d = N_d \bar{y}_d$		$N_d \sqrt{1-\frac{n_d}{N_d}} \frac{s_d}{\sqrt{n_d}}$
Total $\tau_d = N_d \mu_d$ d'un domaine (N_d inconnu)	$\hat{T}_d = \hat{N}_d \bar{y}_d = N \bar{y}'$ où $\hat{N}_d = \frac{n_d}{n} N$		$N \sqrt{1-f} \frac{s'}{\sqrt{n}}$

$$f = \frac{n}{N}$$

Taille d'échantillon

Estimation de la moyenne

La taille d'échantillon nécessaire pour que la marge d'erreur absolue soit égale à E est

$$n = \frac{n_o}{1 + \frac{n_o}{N}} \quad \text{où } n_o = \left(\frac{2S}{E} \right)^2.$$

La taille d'échantillon nécessaire pour que la marge d'erreur relative soit égale à R est

$$n = \frac{n_o}{1 + \frac{n_o}{N}} \quad \text{où } n_o = \left(\frac{2S}{R\mu} \right)^2.$$

Estimation d'une proportion p

Pour estimer une proportion p de telle sorte que la marge d'erreur absolue soit égale à E , la taille approximative de l'échantillon qu'il faut tirer est donnée par $n = \frac{n_o}{1 + \frac{n_o}{N}}$ où $n_o = \frac{4p(1-p)}{E^2}$.

Pour estimer une proportion p de telle sorte que la marge d'erreur *relative* soit égale à R , la taille approximative de l'échantillon qu'il faut tirer est donnée par $n = \frac{n_o}{1 + \frac{n_o}{N}}$ où $n_o = \frac{4(1-p)}{R^2 p}$.

Échantillonnage par strates

L'estimateur de la moyenne dans un échantillon stratifié est $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$.

Son écart type est $\sigma_{\bar{y}_{st}} = \sqrt{\sum_{h=1}^L W_h^2 \sigma_{\bar{y}_h}^2}$ où $\sigma_{\bar{y}_h}^2 = (1-f_h) \frac{S_h^2}{n_h}$ et $f_h = n_h/N_h$.

L'estimateur d'une proportion dans un échantillon stratifié est $\hat{p}_{st} = \sum_{h=1}^L W_h \hat{p}_h$.

Son écart-type est estimé par $\hat{\sigma}_{\hat{p}_{st}} = \sqrt{1-f_h} \sqrt{\frac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}}$.

L'allocation optimale pour l'estimation d'une moyenne dans un échantillon stratifié est donnée par

$$n_h \text{ proportionnels aux } W_h S_h$$

Test du khi-deux

$$\chi^2 = \sum \frac{(O_i - T_i)^2}{T_i}$$

Points critiques ($\alpha = 5\%$) d'une loi khi-deux

v	χ_v^2	v	χ_v^2	v	χ_v^2	v	χ_v^2
1	3,8415	6	12,5916	11	19,6751	16	26,2962
2	5,9915	7	14,0671	12	21,026	17	27,5871
3	7,8147	8	15,5073	13	22,362	18	28,8693
4	9,4877	9	16,919	14	23,6848	19	30,1435
5	11,0705	10	18,307	15	24,9958	20	31,4104

