MAT 2080	**MÉTHODES STATISTIQUES**

EXAMEN FINAL AUTOMNE 2008

Date : Dimanche 14 décembre 2008, 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.
7. Personne n'est admis après 15h00; personne ne quitte la salle avant 15h00.

Question 1 [8 ×1 *points*]

D'une population de N ménages, on tire un échantillon aléatoire simple de n ménages.

Dans chacun des cas énumérés dans la 1$^{\text{ère}}$ colonne (*Paramètre à estimer*) du tableau ci-dessous, identifier le paramètre qu'il s'agit d'estimer. Faites votre choix dans la liste de droite.

	Paramètre à estimer [*répondre par A, B, …, ou H*]	*Liste des réponses possibles*
①	1-a) Le nombre moyen d'ordinateurs par personne dans le ménage.	A : La moyenne $\boldsymbol{\mu_y}$ d'une certaine variable Y
②	1-b) Le revenu moyen par personne	B : La moyenne $\boldsymbol{\mu_d}$ d'une certaine variable Y dans un domaine $\mathfrak{D}$
③	1-c) Le revenu moyen par ménage	C : Le total τ_d d'une certaine variable Y dans un domaine $\mathfrak{D}$
④	1-d) Le nombre total de personnes dans la population	D : Le total τ_y d'une certaine variable Y
⑤	1-e) Le nombre total de femmes dans la population	E : Le nombre $\boldsymbol{N_c}$ d'unités appartenant à une certaine classe $\mathcal{C}$
⑥	1-f) La proportion de femmes dans la population	F : Le quotient $\boldsymbol{R} = \mu_Y/\mu_X$ de deux variables Y et X
⑦	1-g) Le nombre total de ménages monoparentaux	G : La proportion $\boldsymbol{p}$ d'unités appartenant à une certaine classe
⑧	1-h) La proportion de ménages monoparentaux	H : Aucun des paramètres ci-dessus

Question 2 [5+5+5+5+6+6+6+6 *points*]

Afin d'évaluer le potentiel d'un nouveau magasin qu'elle compte construire sur un certain site, une chaîne de magasins d'alimentation décide d'échantillonner les ménages de la région entourant le site. Elle répartit la région en deux zones : la première (la « zone 1 ») est constituée de **900** ménages situées dans le voisinage immédiat du site et la deuxième (la « zone 2 ») est constituée de **2000** ménages plus éloignés. On tire un échantillon aléatoire simple de **36** ménages dans la première zone et de **30** ménages dans la deuxième. On prélève dans chaque ménage : le *nombre de personnes* et le *montant dépensé en épicerie* au courant de la semaine précédant le sondage (annexes 1 et 2).

Les questions suivantes portent sur la zone 1 seulement

Dans tous vos calculs, utilisez les écarts-types et les covariances avec la précision (deux décimales) affichée dans les annexes—ne les recalculez pas. Autrement, retenez le plus grand nombre de décimales possible dans vos calculs intermédiaires.

⑨ 2-a) Estimer les dépenses moyennes par personne.

A	B	C	D	E	F	G	H
18-21	23-25	36-39	44-46	48-51	54-56	62-66	67-69

⑩ 2-b) Estimer le nombre total de personnes qui habitent seules.

A	B	C	D	E	F	G	H
112-115	124-126	132-136	138-141	151-154	161-163	170-173	181-189

⑪ 2-c) Estimer le montant total des dépenses faites par des personnes qui habitent seules.

A	B	C	D	E	F	G	H
6440-6460	10230-10240	15320-15330	22490-22500	31200-31210	33330-33350	41020-41040	46430-46450

⑫ 2-d) Supposons qu'on parvienne à déterminer, à l'aide d'un recensement, que le nombre exact de personnes dans la zone 1 est **3000**. Utiliser cette information pour estimer par le quotient les dépenses moyennes par ménage.

A	B	C	D	E	F	G	H
40,1-40,3	40,4-40,6	41,7-41,9	45,6-45,9	100-102	138-140	149-151	163-165

⑬ 2-e) Estimer l'écart type de l'estimateur en 2-a).

A	B	C	D	E	F	G	H
0,19-0,20	0,21-0,23	0,28-0,31	0,33-0,35	0,37-0,39	0,44-0,48	0,51-0,57	0,59-0,66

⑭ 2-f) Estimer l'écart type de l'estimateur en 2-b).

A	B	C	D	E	F	G	H
44,6-44,9	45,1-45,8	45,9-46,7	47,2-47,9	48,8-49,9	51,1-51,8	53,9-55,2	56,7-58,2

⑮ 2-g) Estimer l'écart type de l'estimateur en 2-c).

A	B	C	D	E	F	G	H
2674-2678	3021-3098	5142-5222	8322-8324	8928-8930	9145-9148	9356-9367	9867-9883

⑯ 2-h) Estimer l'écart type de l'estimateur en 2-d).

A	B	C	D	E	F	G	H
0,86-0,89	1,10-1,13	1,24-1,27	2,18-2,31	2,92-2,95	3,16-3,19	3,24-3,27	3,43-3,47

Question 3 [4+5 *points*]

Mêmes données qu'au numéro 2. Cette question porte sur la région entière (constituée des deux zones.)

⑰ 3-a) Estimer les dépenses moyennes par ménage.

A	B	C	D	E	F	G	H
110-111	112-113	114-115	116-117	118-119	120-121	122-123	123-124

⑱ 3-b) Estimer l'écart-type de l'estimateur en 3-a).

A	B	C	D	E	F	G	H
6,19-6,22	6,23-6,26	6,33-6,41	6,55-6,62	6,66-6,78	7,19-7,26	7,28-7,31	7,33-7,35

⑲ **Question 4** [5 *points*]

Mêmes données qu'au numéro 2. Supposons que les données des échantillons doivent servir à planifier un nouvel échantillonnage, stratifié encore selon la zone, et que les ressources disponibles pour ce faire permettent de tirer un échantillon total de **150** ménages. On voudrait répartir ces **150** ménages entre les deux strates afin d'estimer (par la moyenne) le montant total des dépenses en épicerie. Combien devrait-on tirer de ménages *dans la zone 1* [Vous *arrondirez* le nombre]?

A	B	C	D	E	F	G	H
30	38	54	58	65	67	70	72

Question 5 [5+5 *points*]

On doit déterminer la taille n d'un échantillon à prélever pour estimer une certaine proportion p. On sait que la population est très grande et on a assez d'information pour *savoir* que p se situe quelque part entre **0,4** et **0,6**. Déterminer la taille n de l'échantillon qu'on devrait prélever

⑳ 5-a) pour s'assurer d'une marge d'erreur *relative* de 20 %.

A	B	C	D	E	F	G	H
80	90	95	120	135	140	145	150

㉑ 5-b) pour s'assurer d'une marge d'erreur *absolue* de 5 points de pourcentage.

A	B	C	D	E	F	G	H
250	275	300	325	350	400	425	450

[*Dans les deux cas, n ne doit pas être plus grand que nécessaire*]

Question 6 [4+4 *points*]

On tire un échantillon aléatoire simple de **50** ménages dans un quartier de N = **300** ménages afin d'estimer le nombre moyen d'ordinateurs *par personne*. L'échantillon donne une moyenne (nombre d'ordinateurs par ménage) de $\bar{y}$ = **0,8** avec un écart-type (corrigé) s_y = **0,2**. Par ailleurs, on *sait* que le nombre de personnes dans la population est **690**.

㉒ 6-a) Estimer le nombre moyen d'ordinateurs par personne.

A	B	C	D	E	F	G	H
0,21-0,24	0,28-0,31	0,33-0,35	0,41-0,44	0,47-0,51	0,55-0,59	0,61-0,64	0,67-0,71

㉓ 6-b) Estimer l'écart-type de l'estimateur.

A	B	C	D	E	F	G	H
0,01-0,02	0,04-0,05	0,06-0,07	0,08-0,09	0,10-0,11	0,12-0,13	0,14-0,15	0,16-0,17

Question 7 [5×1 *points*]

Choisir dans la liste que voici l'expression mathématique qui complète correctement les phrases ci-dessous :

A : $\mu_{\bar{y}} = \mu$	**B :** $\sigma_{\hat{\sigma}_1} < \mu_{\hat{\sigma}_2}$	**C :** $\mu_{\hat{\sigma}^2_{\bar{y}}} = \sigma^2_{\bar{y}}$	**D :** $\mu_{\hat{\sigma}} = \sigma$
E : $\mu_{\hat{\sigma}_1} < \mu_{\hat{\sigma}_2}$	**F :** $\sigma_{\hat{\mu}_{yd}} < \sigma_{\bar{y}}$	**G :** $\mu_{\hat{\sigma}} = \mu$	**H :** $\sigma_{\hat{\sigma}_1} < \sigma_{\hat{\sigma}_2}$

㉔ 7-a) Dire que $\hat{\mu}_{y_d}$, l'estimateur de μ par la différence, est meilleur que $\bar{y}$, c'est dire que…

㉕ 7-b) On a deux estimateurs sans biais, $\hat{\sigma}_1$ et $\hat{\sigma}_2$, de l'écart-type σ de la population. Dire que $\hat{\sigma}_1$ est le meilleur des deux c'est dire que…

㉖ 7-c) Dire qu'un estimateur $\hat{\sigma}$ de σ (l'écart-type de la population) est sans biais, c'est dire que…

㉗ 7-d) Dire que $\bar{y}$ est sans biais pour la moyenne μ, c'est dire que…

㉘ 7-e) Dire que $\hat{\sigma}^2_{\bar{y}}$ est un estimateur sans biais de la variance de $\bar{y}$ c'est dire que…

Question 8 [5 ×1 *points*]

Dans le cadre d'une évaluation d'un nouveau procédé de fabrication de lampes à incandescence, on teste 1000 lampes : 600 sont de type A (procédé habituel) et 400 de type B (nouveau procédé, dont on souhaite que la durée soit plus longue). Les résultats sont les suivants (le « succès » est défini par une durée supérieure à 900 heures) :

	Type A (technologie traditionnelle)	Type B (nouvelle technologie)	Total
Succès (durée plus de 900 heures)	450	350	800
Échec (durée de moins de 900 heures)	150	50	200
Total	600	400	1000

On constate que dans cet échantillon, les lampes de type B ont une durée supérieure, et on se demande s'il s'agit d'un phénomène réel ou d'un simple hasard. On effectuera un test statistique pour le savoir. Soient p_A et p_B les probabilités de succès des lampes de type A et B, respectivement.

(29) 8-a) Si l'hypothèse nulle est vraie, le nombre attendu de succès avec les lampes de type A devrait être égal à…

A	B	C	D	E	F	G	H
300	450	480	500	525	600	800	850

Pour les questions 8-b) à 8-e), *choisir vos réponses parmi celles proposées au bas de la page : inscrire A, B, …, ou H.*

(30) 8-b) Lequel ou lesquels des énoncés suivants pourrai(en)t servir d'hypothèse nulle?

C_1 : $p_A = p_B$.
C_2 : Les deux types de lampe ont des durées différentes.
C_3 : Les lampes de type *B* durent plus longtemps.

(31) 8-c) Supposons que la valeur calculée de khi-deux est supérieure au point critique. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?

C_1 : On peut conclure avec confiance que la durée d'une lampe dépend du procédé de fabrication
C_2 : On peut conclure avec confiance que la durée des lampes est la même avec les deux procédés.
C_3 : On peut conclure avec confiance que la durée des deux types de lampe n'est pas la même.

(32) 8-d) Supposons que la valeur calculée de khi-deux est inférieure au point critique. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?

C_1 : On peut conclure avec confiance que les variables « durée » et « procédé de fabrication » sont dépendantes.
C_2 : On peut conclure avec confiance que $p_A = p_B$.
C_3 : On ne peut pas conclure avec confiance que $p_A \neq p_B$.

(33) 8-e) Le niveau du test est de 5 %, ce qui signifie que

C_1 : Si l'hypothèse nulle est vraie, on a 5 % de chance de la rejeter.
C_2 : On a une probabilité de 95 % d'affirmer correctement qu'il n'y a pas de différence entre les deux types de lampes.
C_3 : Avec une probabilité de 5 %, on affirmera qu'il y a une différence entre les deux types de lampes si en fait il n'y en a pas.

Choix de réponse pour les questions 8-b) *à* 8-e) : Les conclusions suivantes sont justifiées :

A : Aucune	**B :** C_1 seulement	**C :** C_2 seulement	**D :** C_3 seulement
E : C_1 et C_2 seulement	**F :** C_1 et C_3 seulement	**G :** C_2 et C_3 seulement	**H :** Toutes

34 **Question 9** [2 *points*]

Indiquer le(s)quel(s) des énoncés suivants est (sont) vrais. Faire votre choix parmi ceux présentés au bas de la page.

E_1 : Plus n est grand, plus l'écart-type de l'estimateur sera petit.
E_2 : Une estimation par le quotient sera d'autant meilleure que la variable auxiliaire est indépendante de Y.
E_3 : Une estimation par la différence est sans biais, quelle que soit la taille de l'échantillon.

35 **Question 10** [2 *points*]

Indiquer le(s)quel(s) des énoncés suivants est (sont) vrais. Faire votre choix parmi ceux présentés au bas de la page.

E_1 : L'allocation optimale dans un échantillon stratifié est parfois moins précise que l'allocation proportionnelle.
E_2 : L'estimateur par le quotient n'est pas sans biais.
E_3 : L'estimateur par la différence n'est pas sans biais.

36 **Question 11** [2 *points*]

Indiquer le(s)quel(s) des énoncés suivants est (sont) vrais. Faire votre choix parmi ceux présentés au bas de la page.

E_1 : L'estimateur de la moyenne dans un échantillon systématique n'a pas de variance.
E_2 : Un estimateur sans biais est un estimateur qui est égal au paramètre.
E_3 : Un échantillon stratifié est toujours plus précis qu'un échantillon aléatoire simple, quelle que soit l'allocation.

Choix de réponses pour les questions 9 *à* 11*. Les énoncés vrais sont les suivants* :

A : Aucune	**B :** E_1 seulement	**C :** E_2 seulement	**D :** E_3 seulement
E : E_1 et E_2 seulement	**F :** E_1 et E_3 seulement	**G :** E_2 et E_3 seulement	**H :** Toutes

Annexe 1 (Données sur l'échantillon de la zone 1)

Données sur un échantillon de **36** ménages tiré d'une population de **900** ménages dans le voisinage immédiat du site. La variable « Dépenses » est le montant dépensé en épicerie au courant de la semaine précédant l'échantillonnage.

#	Nombre de personnes (*X*)	Dépenses (*Y*)
1	1	45
2	1	48
3	1	50
4	1	55
5	1	60
6	2	89
7	2	85
8	2	75
9	2	78
10	2	87
11	2	77
12	2	95
13	2	98
14	2	94
15	2	101
16	2	86
17	3	140
18	3	139
19	3	136
20	3	132
21	3	134
22	3	139
23	3	141
24	3	129
25	4	184
26	4	165
27	4	177
28	4	173
29	4	178
30	5	224
31	5	226
32	5	227
33	5	220
34	5	226
35	6	264
36	6	283
Moyennes	3	135
Écarts-types corrigés	1,45	64,91
Covariance corrigée (entre *X* et *Y*)	93,80	

Annexe 2 (Données sur l'échantillon de la zone 2)

Données sur un échantillon de **30** ménages tiré d'une population de **2000** ménages dans un voisinage éloigné du site. La variable « Dépenses » est le montant dépensé en épicerie au courant de la semaine précédant l'échantillonnage.

#	Nombre de personnes (*X*)	Dépenses (*Y*)
1	1	31
2	1	28
3	1	32
4	1	47
5	1	27
6	2	69
7	2	63
8	2	49
9	2	54
10	2	66
11	2	52
12	2	77
13	3	109
14	3	112
15	3	124
16	3	99
17	3	113
18	3	112
19	3	106
20	3	101
21	4	138
22	4	148
23	4	153
24	4	128
25	4	148
26	5	142
27	5	169
28	5	160
39	5	170
30	7	233
Moyennes	3	102
Écarts-types corrigés	1,49	51,57
Covariance corrigée (entre *X* et *Y*))	74,79	

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 ($f = n/N$)

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 $R = \frac{\mu_y}{\mu_x}$	$\hat{R} = \frac{\bar{y}}{\bar{x}}$	$\sigma_{\hat{R}} \approx \frac{\sqrt{1-f}}{\mu_x}\frac{\sqrt{S_y^2+R^2S_x^2-2RS_{xy}}}{\sqrt{n}}$	$\hat{\sigma}_{\hat{R}} = \frac{\sqrt{1-f}}{\bar{x}}\frac{\sqrt{s_y^2+\hat{R}^2s_x^2-2\hat{R}s_{xy}}}{\sqrt{n}}$
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	$\hat{\mu}_{yq} = \mu_x\hat{R}$	$\sqrt{1-f}\,\frac{\sqrt{S_y^2+R^2S_x^2-2RS_{xy}}}{\sqrt{n}}$	$\sqrt{1-f}\,\frac{\sqrt{s_y^2+\hat{R}^2s_x^2-2\hat{R}s_{xy}}}{\sqrt{n}}$
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* $\hat{N}_d = \frac{n_d}{n}N$		$N\sqrt{1-f}\,\frac{s'}{\sqrt{n}}$

Taille d'échantillon

Estimation de la moyenne

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

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

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

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

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 $n = \dfrac{n_o}{1+\dfrac{n_o}{N}}$ où $n_o = \dfrac{4(1-p)}{R^2 p}$.

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 $\sigma_{\bar{y}_h}^2 = (1 - f_h)\dfrac{S_h^2}{n_h}$ 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 $\hat{\sigma}_{\hat{p}_{st}} = \sqrt{\sum_{h=1}^{L} W_h^2 (1-f_h)\dfrac{\hat{p}_h(1-\hat{p}_h)}{n_h - 1}}$.

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

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

Test du khi-deux

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

Points critiques **(α = 5 %)** ***d'une loi khi-deux***

ν	χ^2_ν	ν	χ^2_ν	ν	χ^2_ν	ν	χ^2_ν
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

Brouillons

Brouillons

MAT 2080 **Examen final automne 2008** **Méthodes statistiques**
Dimanche 14 décembre 2008, de 14h00 à 17h00

Nom : ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Prénom : ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Code permanent : ☐☐☐☐☐☐☐☐☐☐☐☐ **Groupe:** ☐☐

Question		*Réponse*	
1	1-a)	F	1
2	1-b)	F	1
3	1-c)	A	1
4	1-d)	D	1
5	1-e)	D	1
6	1-f)	F	1
7	1-g)	E	1
8	1-h)	G	1
9	2-a)	D	5
10	2-b)	B	5
11	2-c)	A	5
12	2-d)	G	5
13	2-e)	C	6
14	2-f)	F	6
15	2-g)	A	6
16	2-h)	A	6
17	3-a)	B	4
18	3-b)	F	5

Question		*Réponse*	
19	4	C	5
20	5-a)	H	5
21	5-b)	F	5
22	6-a)	C	4
23	6-b)	A	4
24	7-a)	F	1
25	7-b)	H	1
26	7-c)	D	1
27	7-d)	A	1
28	7-e)	C	1
29	8-a)	C	1
30	8-b)	B	1
31	8-c)	F	1
32	8-d)	D	1
33	8-e)	H	1
34	9	F	2
35	10	C	2
36	11	A	2

Blanc