MAT 2080	**MÉTHODES STATISTIQUES**
EXAMEN FINAL HIVER 2006	
Date : Dimanche 30 avril 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.

Problème 1 [20 *points*]

On tire un échantillon aléatoire simple de n succursales parmi les N succursales d'une banque.

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.

Liste des réponses possibles	*Paramètre à estimer* [*répondre par A, B, …, ou H*]
A: La moyenne μ_y d'une certaine variable Y	1-a) Le nombre total d'employés de plus de 60 ans
B: La moyenne μ_d d'une certaine variable Y dans un domaine D.	1-b) Le nombre de succursales ayant plus de 30 employés
C: Le total τ_d d'une certaine variable Y dans un domaine D.	1-c) La proportion de femmes parmi les employés
D: Le total τ_y d'une certaine variable Y	1-d) L'âge moyen des employés masculins
E: Le nombre N_c d'unités appartenant à une certaine classe C	1-e) Le nombre total de femmes dans les succursales situées en banlieue
F: Le quotient $R = \mu_Y/\mu_X$ de deux variables Y et X	1-f) Le nombre moyen de femmes dans les succursales ayant plus de 30 employés
G: La proportion p d'unités appartenant à une certaine classe	1-g) Le salaire moyen des femmes employées
H: Aucun des paramètres ci-dessus	1-h) Le nombre total de femmes en congé de maternité
	1-i) La proportion de succursales ayant plus de 30 employés
	1-j) La proportion de femmes parmi les employés âgés de plus de 60 ans

Question 2 [40 *points*]
On tire un échantillon aléatoire simple de **30** factures d'une population constituée des **600** factures émises dans un restaurant au courant d'un mois. Pour chaque facture on détermine

Le type de repas (dîner ou souper)
Le montant total de la facture
La quantité de vin servi (en millilitres)

Les données de l'échantillon sont présentées dans le tableau 1 en annexe. [*Veuillez utiliser les écarts-types et les covariance tels que présentés, sans plus de décimales*]
[*Choisissez vos réponses parmi celles proposées au bas de la page*]

2-a) Estimer le *nombre total* de repas servis sans vin

2-b) Estimer le *montant total* des factures

2-c) Estimer la *valeur totale* des *soupers* servis sans vin —estimation par la moyenne

2-d) Estimer la *valeur moyenne* des factures—estimation par le quotient, utilisant le fait, supposé connu, qu'on a servi 180 000 ml de vin en tout durant le mois

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

2-f) Estimer l'écart-type de l'estimateur en 2-c)

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

2-h) Estimer la taille de l'échantillon qu'il aurait fallu prélever pour estimer (par la moyenne) le montant total des factures avec une marge d'erreur de 5 %.

Question 3 [15 *points*]
D'une population constituée de **600** factures de restaurant, dont **200** sont pour des dîners et **400** pour des soupers, on tire un échantillon stratifié selon le type de repas (dîner ou souper) et on note le montant Y de la facture. Les données sont résumées dans le tableau 2 en annexe.
[*Choisissez vos réponses parmi celles proposées au bas de la page*]

3-a) Estimer le montant moyen des factures de la population

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

3-c) Soit n_1 le nombre de factures de dîner qu'on aurait tirées si on avait utilisé l'allocation optimale (d'un échantillon de taille 30). Estimer n_1.

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

A 0,06 – 0,19	**B** 2,2 – 2,4	**C** 2,41 – 2,62	**D** 9,5 – 10,5	**E** 10,6 – 11,0
F 11,10 – 12,80	**G** 14,80 – 15,80	**H** 40,0 – 51.0	**I** 53,6 – 53,9	**J** 60,5 – 61,5
K 72,5 – 73,5	**L** 77,4 – 77,9	**M** 159 - 161	**N** 209 - 211	**O** 259 - 261
P 360 - 370	**Q** 870 - 880	**R** 1820 - 1830	**S** 2910-2930	**T** 3385 - 3395
U 7315 - 7325	**V** 9800 – 12 250	**W** 36 535 – 36 545	**X** 36 550 – 36 650	**Y** Aucune de ces réponses

Question 4 [9 *points*]

Après avoir longtemps engagé des informaticiens, le responsable des ressources humaines d'une compagnie demande à une analyste de comparer les informaticiens de deux *classes* : les « juniors » (avec un DEC) et les « senior » (bacheliers). Entre autres, on effectuera un test du khi-deux pour savoir s'il y a une différence dans la durée de leur service : est-ce que les uns ont tendance à démissionner plus tôt que les autres? On répartit l'ensemble des informaticiens engagés depuis plus de 2 ans selon leur classe et la durée de leur service. Les données accumulées se présentent sous la forme suivante :

	Durée de service (Moment de leur démission)		
Classe	Au courant de leur 1ère année	Au courant de leur 2^{e} année	Après leur 2^{e} année
Juniors	Effectif	Effectif	Effectif
Seniors	Effectif	Effectif	Effectif

[*Pour les questions* 4-a) à 4-c), c*hoisir vos réponses parmi celles proposées au bas de la page*]

4-a) Lequel ou lesquels des énoncés suivants pourrai(en)t servir d'hypothèse nulle ?

- P_1 La classe et la durée de service sont des variables indépendantes
- P_2 : Les seniors démissionnent plus vite que les juniors
- P_3 : La durée de service dépend de la classe.
- P_4 : Il n'y a pas de différence entre les juniors et les seniors quant à la durée de service

4-b) Supposons que la valeur de χ^2 est *supérieure* au point critique 5,9915 (à 5 %). Lequel ou lesquels des énoncés suivants sont justifiés ?

- P_1 : On peut conclure avec confiance qu'il n'y a pas de différence entre les deux classes quant à la durée de service
- P_2 : On peut conclure avec confiance qu'il y a une différence entre les deux classes quant à la durée de service
- P_3 : On peut conclure avec confiance qu'il y a une dépendance entre la classe et la durée de service
- P_4 : La probabilité est de 95 % qu'il y ait une relation entre la classe et la durée de service

4-c) Supposons que la valeur de χ^2 est *inférieure* au point critique 5,9915 (à 5 %). Lequel ou lesquels des énoncés suivants sont justifiés ?

- P_1 : On peut conclure avec 95 % de confiance qu'il n'y a pas de différence entre les deux classes quant à la durée de service
- P_2 On ne peut pas conclure avec confiance qu'il y a une différence entre les deux classes quant à la durée de service
- P_3 : La probabilité est de 5 % qu'il y ait une relation entre la classe et la durée de service
- P_4 : On peut conclure avec confiance qu'il y a une dépendance entre la classe et la durée de service.

Choix de réponse pour les questions 4-a) *à* 4-c)

A Aucun	**B** P_1 seulement	**C** P_2 seulement	**D** P_3 seulement
E P_4 seulement	**F** P_1 et P_2 seulement	**G** P_1 et P_3 seulement	**H** P_1 et P_4 seulement
I P_2 et P_3 seulement	**J** P_2 et P_4 seulement	**K** P_3 et P_4 seulement	**L** P_1 , P_2 et P_3 seulement
M P_1 , P_2 et P_4 seulement	**N** P_1 , P_3 et P_4 seulement	**O** P_2 , P_3 et P_4 seulement	**P** Tous

[*Pour les questions* 5 à 8, c*hoisir vos réponses parmi celles proposées au bas de la page*]

Question 5 [4 *points*]

Dites lequel ou lesquels des énoncés suivants sont vrais. [*Faites vos choix dans la liste au bas de la page*]

P_1 : L'estimateur d'une moyenne dans un échantillon systématique n'a pas de variance.

P_2 : [*Échantillonnage stratifié*] Lorsque les variances (corrigées) des strates sont toutes égales, alors l'allocation proportionnelle est optimale

P_3 : [*Échantillonnage stratifié*] La variance de $\overline{y}_{st}$ (la moyenne d'un échantillon stratifié de taille n) est toujours inférieure ou égale à la variance de $\overline{y}$ dans un échantillon aléatoire simple de même taille tiré dans la même population

Question 6 [4 *points*]

Dites lequel ou lesquels des énoncés suivants sont vrais. [*Faites vos choix dans la liste au bas de la page*]

P_1 : Soit $\hat{\mu}_{opt}$ l'estimateur d'une moyenne μ_y dans un échantillon stratifié de taille n avec allocation *optimale*; et $\hat{\mu}_{prop}$ l'estimateur de μ_y dans un échantillon stratifié de taille n avec allocation *proportionnelle*. Alors $\sigma_{\hat{\mu}_{opt}} \leq \sigma_{\hat{\mu}_{prop}}$

P_2 : Soit $\hat{\mu}_r$ l'estimateur de la moyenne μ par le quotient dans un échantillon aléatoire simple. Alors $\mu_{\hat{\mu}_r} = \mu$.

P_3 : Soit $\overline{y}$ la moyenne d'un échantillon aléatoire simple et $\hat{\mu}_r$ l'estimateur de la moyenne par le quotient. Si la corrélation entre la variable d'intérêt Y et la variable auxiliaire X est positive et très forte, alors on s'attend à ce que $\sigma_{\overline{y}} \geq \sigma_{\hat{\mu}_r}$

Question 7 [4 *points*]

Dites lequel ou lesquels des énoncés suivants sont vrais. [*Faites vos choix dans la liste au bas de la page*]

P_1 : Si p est la proportion de fumeurs dans une population de taille 10 000 et $\hat{p}$ est la proportion de fumeurs dans un échantillon de taille 100 tiré de cette population, alors $\sigma_p \leq \sigma_{\hat{p}}$

P_2 : Soit $\hat{\mu}_{opt}$ l'estimateur d'une moyenne μ_y dans un échantillon stratifié de taille n avec allocation *optimale*; et $\hat{\mu}_{prop}$ l'estimateur de μ_y dans un échantillon stratifié de taille n avec allocation *proportionnelle*. Alors $\hat{\sigma}_{\hat{\mu}_{opt}} \leq \hat{\sigma}_{\hat{\mu}_{prop}}$

P_3 : Si $\hat{p}_1$ est la proportion de fumeurs dans un échantillon de taille 100 tiré sans remise d'une population de 10 000 personnes et $\hat{p}_2$ est la proportion de fumeurs dans un échantillon de taille 100 tiré *avec* remise de la même population, alors $\sigma_{\hat{p}_1} \leq \sigma_{\hat{p}_2}$.

Question 8 [4 *points*]

Deux personnes tirent, indépendamment l'un de l'autre, un échantillon aléatoire simple d'une même très grande population. La taille du premier échantillon est n_1, celle du deuxième n_2.

Dites lequel ou lesquels des énoncés suivants sont vrais. [*Faites vos choix dans la liste au bas de la page*]

P_1 : Si $n_1 = n_2$, $\sigma_{\overline{y}}$ aura la même valeur dans les deux cas

P_2 : Si $n_1 = n_2$, $\hat{\sigma}_{\overline{y}}$ aura la même valeur dans les deux cas

P_3 : Si $n_1 < n_2$, $\sigma_{\overline{y}}$ sera plus petit dans le premier échantillon que dans le deuxième.

Choix de réponse pour les questions 5 *à* 8

A Aucun	**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** Tous

Annexe – Tableau 1

Les données suivantes portent sur un échantillon de **30** factures tirées d'une population des **600** factures émises par un restaurant en un mois donné. Les variables observées sont le type de repas (dîner ou souper), le montant de la facture, et le nombre de millilitres de vin servi.

Échantillon	*Repas* (Dîner ou souper)	*Facture* (Montant de la facture, en dollars)	*Vin* (Nombre de millilitres de vin servi)
1	Dîner	27	0
2	Dîner	23	0
3	Dîner	28	0
4	Dîner	20	0
5	Dîner	30	250
6	Dîner	42	250
7	Dîner	44	250
8	Dîner	31	250
9	Dîner	54	0
10	Dîner	49	250
11	Dîner	48	0
12	Dîner	65	0
13	Dîner	53	500
14	Dîner	56	250
15	Dîner	71	250
16	Dîner	49	0
17	Dîner	41	750
18	Dîner	46	250
19	Dîner	64	375
20	Dîner	44	0
21	Souper	92	0
22	Souper	80	0
23	Souper	109	750
24	Souper	95	0
25	Souper	114	625
26	Souper	99	0
27	Souper	98	625
28	Souper	80	875
29	Souper	86	375
30	Souper	89	625
Sommes		1827	7500
Variances corrigées (S^2)		748	76 509
Covariance		3065	

Annexe – Tableau 2

Échantillon stratifié d'une population de factures pour **200** dîners et **400** soupers

	n	Sommes	Variances corrigées (S^2)
Dîner	20	885	205
Souper	10	942	128

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

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

Taille d'échantillon

Estimation de la moyenne

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

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

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

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 $\hat{\sigma}_{\hat{p}_{st}} = \sqrt{\sum_{h=1}^{L} W_h^2 \sigma_{\hat{p}_h}^2}$, où $\hat{\sigma}_{\hat{p}_{st}}^2 = \left(1-f_h\right)\dfrac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}$.

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

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

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

Brouillon

Brouillon

Brouillon

Feuille-réponses – version jaune

/100

Ne rien écrire ici

MAT 2080 **Examen final hiver 2006** **Méthodes statistiques**
Dimanche 30 avril 2006, de 14h00 à 17h00

Nom :

Prénom :

Code permanent : **Groupe:**

Question	*Réponse*	
1-a)	D	2
1-b)	E	2
1-c)	F	2
1-d)	F	2
1-e)	C	2
1-f)	B	2
1-g)	F	2
1-h)	D	2
1-i)	G	2
1-j)	F	2
2-a)	O	4
2-b)	W	4
2-c)	U	4
2-d)	K	4

Question	*Réponse*	
2-e)	I	6
2-f)	T	6
2-g)	E	6
2-h)	N	6
3-a)	L	5
3-b)	C	5
3-c)	F	5
4-a)	H	3
4-b)	I	3
4-c)	C	3
5	C	4
6	F	4
7	F	4
8	B	4