MAT 2080	**MÉTHODES STATISTIQUES**
	EXAMEN FINAL AUTOMNE 2007
Date : Dimanche 16 décembre 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. Aucun téléphone cellulaire sur la table.
7. Personne ne quitte la salle avant 15h00; personne n'est admis après 15h00.

Problème 1 [10 *points*]

À l'occasion d'une vérification des factures d'un magasin, on tire un échantillon aléatoire simple de 30 jours parmi les 1096 jours d'une période de 3 ans. [*Chaque vente est accompagnée d'une facture et la somme des montants des factures d'une journée constitue les recettes de la journée*].

Dans chacun des cas énumérés dans le cadre de gauche (*Paramètre à estimer,* ci-dessous), identifier le paramètre qu'il s'agit d'estimer. Faites votre choix dans le cadre de droite (Inscrire A, B, …, ou H).

Paramètre à estimer
1-a) Les recettes moyennes par jour
1-b) Les recettes totales de 3 ans
1-c) La proportion de jours sans aucune vente
1-d) Le nombre total de factures émises durant les week-ends
1-e) La proportion de factures de moins de 20 $
1-f) La valeur moyenne d'une facture
1-g) La proportion des recettes réalisées durant les week-ends
1-h) Le nombre total de jours sans aucune vente
1-i) Le nombre de factures de moins de 20 $
1-j) Les recettes moyennes *par jour durant la première année*

Liste des réponses possibles
A : La moyenne μ_y d'une certaine variable Y
B : La moyenne μ_d d'une certaine variable Y dans un domaine $\mathfrak{D}$.
C : Le total τ_d d'une certaine variable Y dans un domaine
D : Le total τ_y d'une certaine variable Y
E : Le nombre N_c d'unités appartenant à une certaine classe $\mathfrak{C}$
F : Le quotient $R = \mu_Y/\mu_X$ de deux variables Y et X
G : La proportion p d'unités appartenant à une certaine classe
H : La taille N de la population

Problèmes relatifs aux annexes 1 et 2 : Utilisez les moyennes, variances et covariances telles que présentées, *sans changer le nombre de décimales*.

Problème 2 [20+16+5 *points*] *Les données pertinentes sont présentées à l'annexe* 1.

Afin de vérifier les comptes d'un certain restaurant, on prélève un échantillon aléatoire simple de **30** jours parmi les **260** *jours de semaine* dans l'année 2006 (les week-ends feront l'objet d'un échantillonnage séparé). Pour chacun des jours sélectionnés, on a noté le nombre de factures ainsi que les recettes totales de la journée. On a également pris note du jour de la semaine sélectionné. [L'année 2006 compte exactement 52 de chaque jour de semaine]

2-a) Déterminer une estimation *ponctuelle* (l'estimation seulement, pas l'écart-type) de chacun des paramètres suivants (choisir vos réponses dans la liste au bas de la page) :
- i) La valeur moyenne d'une facture
- ii) Le nombre total de factures émises les vendredis (*vous* savez *que la population compte* 52 *vendredis. Utilisez ce fait*)
- iii) Le nombre total de factures émises les jours jugés *déficitaires* (ce sont les jours avec moins de 15 ventes)
- iv) Le nombre de jours où moins de 20 factures ont été émises.

2-b) Déterminer l'*écart-type* de chacun des estimateurs suivants :
- i) L'estimateur en a)-i).
- ii) L'estimateur en a)-ii) [*Utilisez le fait que la population compte* 52 v*endredis*]
- iii) L'estimateur en a)-iii)
- iv) L'estimateur en a)-iv)

2-c) On a l'intention de prélever un échantillon semblable l'année prochaine, l'année 2007 [qui compte **261** jours de semaine]. Déterminer la taille de l'échantillon qu'il faudra prélever si on veut pouvoir estimer les recettes totales (durant les 261 jours semaine) avec une marge d'erreur de 10 %.
[*Estimez tous les paramètres inconnus par les estimations que vous avez obtenues pour* 2006]

Problème 3 [6+6+6 *points*] *Suite du problème* 2. *Les données pertinentes sont présentées à l'annexe* 2.

On procède à deux autres échantillonnages indépendants, chacun de taille **20**, l'un tiré dans la population des **52** samedis de l'année et l'autre dans la population des **53** dimanches de l'année.

3-a) Estimer les recettes totales de l'année (jours de semaine et week-ends inclus)
3-b) Déterminer l'écart-type de l'estimateur utilisé en a).
3-c) On a l'intention de tirer l'année prochaine un échantillon stratifié de taille 60 (les trois strates étant : jours de semaine, samedis, dimanches) afin d'estimer les recettes totales en 2007.
- i) Combien de jours de semaine devrait-on sélectionner? (Il y a 53 lundis en 2007). Vous supposerez que les valeurs des paramètres en 2007 sont celles que vous avez estimées pour 2006.
- ii) Combien doit-on sélectionner de dimanches?

Problème 4 [6+4 *points*]

On a l'intention de prélever un échantillon de n employés parmi les 1400 d'une compagnie afin d'estimer la proportion p d'employés qui seraient en faveur d'un régime d'assurance de soins dentaires. On a de bonnes raisons de croire que p se situe entre 0,4 et 0,7. Quelle est la taille de l'échantillon qu'on devrait prélever

4-a) s'il faut pouvoir estimer p avec une marge d'erreur *absolue* de 3 points de pourcentage?
4-b) s'il faut pouvoir estimer p avec une marge d'erreur *relative* de 10 %?

Choix de réponse pour les problèmes 2, 3 et 4 [*Choisir l'intervalle qui contient votre réponse*] :

A	B	C	D	E	F	G	H	I	J	K	L
0,5 à 0,6	14,5 à 16,5	21 à 23	23,5 à 24,5	30 à 33	36 à 38	38,5 à 40	129 à 131	151 à 154	167 à 170	210 à 218	310 à 312

M	N	O	P	Q	R	S	T	U	X
319 à 322	417 à 422	556 à 562	604 à 606	618 à 620	1415 à 1420	12900 à 12990	643 800 à 644 000	714 215 à 714 230	Aucune de ces réponses

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

[*Choisir vos réponses parmi celles proposées après la question* 5-c) *: inscrire A,* B, … , *ou* P)]
Considérer l'échantillon présenté à **l'annexe 1** et décrit au problème 2. Lors d'un litige entre le propriétaire du restaurant et le Ministère du revenu, quelqu'un fait remarquer que certains jours de la semaine ont l'air d'y être fortement représentés alors que d'autres, au contraire, le sont très faiblement. On soupçonne alors que l'échantillon n'a pas été tiré de façon aléatoire, comme il le faudrait, mais qu'on a procédé d'une façon telle que certains jours de la semaine ont été favorisés aux dépends d'autres jours. Vous avez pour mandat de déterminer si ces soupçons sont justifiés. Vous allez donc faire un test du khi-deux.

5-a) Lequel ou lesquels des énoncés suivants pourrai(en)t servir d'hypothèse nulle?
E_1 : L'échantillon est biaisé
E_2 : Chacun des 5 jours de semaine avait la même probabilité de sélection
E_3 Les lundis sont trop fréquents
E_4 : Le tirage a été fait de façon purement aléatoire.

5-b) Supposons que la valeur calculée de khi-deux est supérieure au point critique. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?
E_1 : On peut conclure que l'échantillon n'a pas été tiré de façon purement aléatoire.
E_2 : On peut conclure avec confiance que l'échantillon a été tiré aléatoirement.
E_3 : On peut conclure avec confiance que certains jours ont été favorisés dans le tirage.
E_4 : On ne peut pas conclure avec confiance que l'échantillon n'a pas été tiré aléatoirement.

5-c) Supposons que la valeur calculée de khi-deux est inférieure au point critique. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?
E_1 : On peut conclure avec confiance que l'échantillon a été tiré aléatoirement
E_2 : On ne peut pas conclure avec confiance que l'échantillon n'a pas été tiré aléatoirement
E_3 : On ne peut pas conclure avec confiance que certains jours ont été favorisés par rapport à d'autres
E_4 : On peut conclure avec confiance que l'échantillon n'a pas été tiré aléatoirement.

Choix de réponse pour les questions 5-a) *à* 5-c)

A Aucun	**B** E_1 seulement	**C** E_2 seulement	**D** E_3 seulement
E E_4 seulement	**F** E_1 et E_2 seulement	**G** E_1 et E_3 seulement	**H** E_1 et E_4 seulement
I E_2 et E_3 seulement	**J** E_2 et E_4 seulement	**K** E_3 et E_4 seulement	**L** E_1 , E_2 et E_3 seulement
M E_1 , E_2 et E_4 seulement	**N** E_1 , E_3 et E_4 seulement	**O** E_2 , E_3 et E_4 seulement	**P** Tous

5-d) Déterminer le point critique (selon la table en annexe)

5-e) Déterminer la valeur calculée de χ^2.

Choix de réponse pour les questions 5-d) *et* 5-e)

A 1,90 à 2,00	**B** 2,10 à 2,30	**C** 2,31 à 2,35	**D** 2,37 à 2,39
E 5,5-7,5	**F** 7,5-8,5	**G** 9,4 à 9,6	**H** 11,0 à 11,5
M 12-13	**N** 13,1-14,1	**O** 23-25	**P** Aucun

Problème 6 [1+1 *points*]

On voudrait augmenter la teneur en fer (X) d'un certain alliage, mais on craint que ceci ait un effet corrosif. On veut donc répondre à la question suivante (Q):

Q : Est-ce que la teneur en fer dans cet alliage affecte la corrosion?

À cette fin, on plonge n spécimens, chacun ayant une teneur en fer distincte, dans une solution saline pendant 60 jours, et on observe ensuite le degré de corrosion Y (mesuré par la perte de poids par décimètre cube par jour). On calcule les coefficients a et b de la droite des moindres carrés $Y = a + bX$; le coefficient de corrélation r; et enfin la statistique $Z = \dfrac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}$.

6-a) Dites lesquels des énoncés suivants sont vrais :

E_1 : Si b est très grand, on peut immédiatement répondre OUI à la question Q

E_2 : Si $b = 0$, c'est que *l'échantillon* ne présente aucune relation linéaire entre X et Y

E_3 : Si r est très proche de 1, on répondra OUI à la question Q avec confiance quelle que soit la taille de l'échantillon

E_4 : Si r est négatif, on peut immédiatement répondre NON à la question Q.

6-b) Dites lesquels des énoncés suivants sont vrais :

E_1 : Si $|Z| > 2$, on pourra prédire avec précision le degré de corrosion à partir de la teneur en fer

E_2 : Si r est très petit mais $|Z| > 2$, on peut répondre OUI à la question Q, mais on sait qu'on ne pourra pas prédire le degré de corrosion à partir de la teneur en fer avec beaucoup de précision

E_3 : Si r est grand, mais $|Z| < 2$, on ne répondra ni OUI ni NON à la question Q, en expliquant que la taille de l'échantillon est trop petite pour tirer quelque conclusion que ce soit à propos du phénomène étudié

E_4 : Si r est négatif, et $|Z| > 2$, on peut conclure avec confiance que plus il y a du fer dans l'alliage, moins grave sera la corrosion.

Choix de réponse pour la question 5 : *Les énoncés vrais sont les suivants* :

A Aucune	**B** E_1 seulement	**C** E_2 seulement	**D** E_3 seulement
E E_4 seulement	**F** E_1 et E_2 seulement	**G** E_1 et E_3 seulement	**H** E_1 et E_4 seulement
I E_2 et E_3 seulement	**J** E_2 et E_4 seulement	**K** E_3 et E_4 seulement	**L** E_1, E_2 et E_3 seulement
M E_1, E_2 et E_4 seulement	**N** E_1, E_3 et E_4 seulement	**O** E_2, E_3 et E_4 seulement	**P** Tous

Problème 7 [4 *points*]

On doit échantillonner une population constituée de toutes les factures émises par un commerce durant une période de vérification de 3 ans afin d'estimer le montant moyen des factures. 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.

7-a)	On tire au hasard 20 jours parmi les 1096 de la période. L'échantillon est l'ensemble des factures émises ces jours-là.
7-b)	On tire au hasard 20 factures parmi celles émises le dimanche; 20 parmi celles émises le samedi; et 30 parmi celles émises un jour de semaine.
7-c)	On dresse une liste de toutes les factures, puis on tire au hasard 100 factures dans cette liste.
7-d)	On dresse une liste de toutes les factures, dans l'ordre où elles ont été émises, puis à partir d'un point choisi au hasard, on tire chaque 1000e facture.

Annexe 1

Les données suivantes sont issues d'un échantillon aléatoire simple de **30** jours tirés parmi les **260** jours de semaine de l'année.

	Jour de la semaine	*Nombre de factures* (X)	*Montant total des recettes* (Y)
1	Lundi	12	446
2	Mercredi	12	482
3	Lundi	13	603
4	Lundi	15	659
5	Lundi	15	503
6	Mardi	15	542
7	Mercredi	15	628
8	Lundi	16	666
9	Lundi	17	599
10	Lundi	17	634
11	Mercredi	17	768
12	Jeudi	17	660
13	Mercredi	18	731
14	Lundi	19	798
15	Lundi	19	822
16	Mardi	20	744
17	Mardi	20	719
18	Mercredi	20	845
19	Mercredi	20	852
20	Jeudi	20	846
21	Jeudi	21	721
22	Mardi	22	845
23	Jeudi	23	816
24	Jeudi	23	928
25	Mardi	25	1073
26	Jeudi	25	1087
27	**Vendredi**	26	1007
28	**Vendredi**	27	1045
29	**Vendredi**	27	1125
30	**Vendredi**	29	1116
	n	30	30
	Moyennes	19,5	777
	Variances (corrigées)	21,22	36969
	Covariance (corrigée)	839,3	

***Annexe* 2**

Les données suivantes sont issues d'un échantillon aléatoire simple de 20 jours tirés parmi les 52 samedis et d'un échantillon aléatoire simple de 20 jours tirés parmi les 53 dimanches de l'année.

	Montant total des recettes	
	Samedis	*Dimanches*
	4692	1254
	4893	1625
	4683	2586
	5098	2304
	5467	2348
	4912	1861
	5144	2172
	5511	2624
	5865	2655
	5676	2652
	6062	2845
	5529	2810
	6185	2636
	6131	2779
	6763	2858
	5920	3477
	6439	2480
	6581	3395
	7089	4014
	7920	3045
n	20	20
Moyennes	5828	2621
Variances (corrigées)	722308	392497

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	$\bar{y}$	$\sigma_{\bar{y}} = \sqrt{1-f}\ \dfrac{S}{\sqrt{n}}$	$\hat{\sigma}_{\bar{y}} = \sqrt{1-f}\ \dfrac{s}{\sqrt{n}}$
Proportion p	$\hat{p} = \dfrac{X}{n}$	$\sigma_{\hat{p}} = \sqrt{\dfrac{N-n}{N-1}}\sqrt{\dfrac{p(1-p)}{n}}$	$\hat{\sigma}_{\hat{p}} = \sqrt{1-f}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n-1}}$
Un quotient R = $\dfrac{\mu_y}{\mu_x}$	$\hat{R} = \dfrac{\bar{y}}{\bar{x}}$	$\sigma_{\hat{R}} \approx \dfrac{\sqrt{1-f}}{\mu_x}\dfrac{\sqrt{S_y^2+R^2S_x^2-2RS_{xy}}}{\sqrt{n}}$	$\hat{\sigma}_{\hat{R}} = \dfrac{\sqrt{1-f}}{\bar{x}}\dfrac{\sqrt{s_y^2+\hat{R}^2s_x^2-2\hat{R}s_{xy}}}{\sqrt{n}}$
Moyenne μ Estimation par la différence	$\hat{\mu}_{yd} = \mu_x + (\bar{y}-\bar{x})$	$\sigma_{\hat{\mu}_{yd}} = \sqrt{1-f}\ \dfrac{\sqrt{S_y^2+S_x^2-2S_{xy}}}{\sqrt{n}}$	$\hat{\sigma}_{\hat{\mu}_{yd}} = \sqrt{1-f}\ \dfrac{\sqrt{s_y^2+s_x^2-2s_{xy}}}{\sqrt{n}}$
Moyenne μ Estimation par le quotient	$\hat{\mu}_{yq} = \mu_x\hat{R}$	$\sqrt{1-f}\ \dfrac{\sqrt{S_y^2+R^2S_x^2-2RS_{xy}}}{\sqrt{n}}$	$\sqrt{1-f}\ \dfrac{\sqrt{s_y^2+\hat{R}^2s_x^2-2\hat{R}s_{xy}}}{\sqrt{n}}$
Moyenne μ_d *d'un domaine* $\mathfrak{D}$	$\bar{y}_d$: *Moyenne du domaine dans l'échantillon*		$\sqrt{1-\dfrac{n_d}{N_d}}\ \dfrac{s_d}{\sqrt{n_d}}$ *ou* $\sqrt{1-\dfrac{n}{N}}\ \dfrac{s_d}{\sqrt{n_d}}$ *selon que* N_d *est connu ou pas*
Total $\tau_d = N_d\mu_d$ *d'un domaine* (N_d *connu*)	$T_d = N_d\,\bar{y}_d$		$N_d\sqrt{1-\dfrac{n_d}{N_d}}\ \dfrac{s_d}{\sqrt{n_d}}$
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* $\hat{N}_d = \dfrac{n_d}{n}N$		$N\sqrt{1-f}\ \dfrac{s'}{\sqrt{n}}$

Taille d'échantillon

Estimation de la moyenne

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

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

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

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

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

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

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

Régression

Variance corrigée :

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}$$

Covariance corrigée :

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{n-1}$$

Coefficient de corrélation :

$$r = \frac{s_{xy}}{s_x s_y}$$

Droite des moindres carrés y = a + bx :

$$b = \frac{s_{xy}}{s_x^2} \;;\; a = \bar{y} - b\bar{x}$$

Statistique pour tester l'indépendance : $Z = \dfrac{\sqrt{n-2}\; r}{\sqrt{1-r^2}}$

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

NOM ....

Examen Final A07

Feuille-réponses (version blanche)

/100

Ne rien écrire ici

Nom :

Prénom :

Code permanent :

Groupe:

Question	*Réponse*	
1-a)	A	1
1-b)	D	1
1-c)	E	1
1-d)	C	1
1-e)	F	1
1-f)	F	1
1-g)	F	1
1-h)	E	1
1-i)	D	1
1-j)	B	1
2-a-i)	G	5
2-a-ii)	R	5
2-a-iii)	M	5
2-a-iv)	H	5
2-b)-i)	A	4
2-b)-ii)	E	4
2-b)-iii)	J	4
2-b)-iv)	C	4

Question	*Réponse*	
2-c)	C	5
3-a)	T	6
3-b)	S	6
3-c)-i)	D	3
3-c)-ii)	B	3
4-a)	Q	6
4-b)	N	4
5-a)	J	2
5-b)	H	2
5-c)	I	2
5-d)	G	4
5-e)	C	5
6-a)	C	1
6-b)	O	1
7-a)	D	1
7-b)	B	1
7-c)	A	1
7-d)	C	1