

MAT 2080

MÉTHODES STATISTIQUES

EXAMEN FINAL HIVER 2005

Date : Dimanche 1^{er} mai 2005 de 14h00 à 17h00

Nom :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Prénom :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Code permanent :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Groupe:

--	--

INSTRUCTIONS

Prendre grand soin de ne pas désassembler les feuilles du présent cahier (5 pages + annexe + formulaire + papier brouillon), qui doit être remis en entier. Seuls l'annexe, le formulaire et les feuilles brouillon peuvent être détachés du cahier et n'ont pas à être retournés.

Par mesure de précaution, inscrire lisiblement votre nom au haut de chacune des pages 2 à 5.

Suivre minutieusement les consignes sur la façon de répondre [inscription d'une réponse numérique dans une case, une coche dans un carré, encerclement d'un choix, etc.].

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.

L'usage d'une calculatrice est autorisé.

L'étudiant doit présenter sa carte d'étudiant (avec photo) sur la table et signer la feuille de présence lors de la remise de son cahier.

Grille à l'usage du correcteur

1-a)	1-b)	2	3	4	5
/16	/16	/16	/27	/15	/10

Total :

/100

NOM.....

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

On tire un échantillon aléatoire simple de **30** (représentant 30 tables servies) factures d'un restaurant. La population est l'ensemble des **400** factures de la semaine. Pour chacune des factures on note le nombre de personnes servies (la variable X), le montant total de la facture (Y), et le montant dépensé en boisson (Z).

1-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) Le montant total τ_y des recettes de la semaine—estimation par la moyenne	Réponse : (A, B, ..., ou W) ►	V ou W
ii) Les dépenses moyennes par personne	Réponse : (A, B, ..., ou W) ►	H
iii) La <i>proportion</i> des factures destinées à une seule personne.	Réponse : (A, B, ..., ou W) ►	D
iv) La proportion des recettes provenant du service de boissons.	Réponse : (A, B, ..., ou W) ►	E
v) Le montant total τ_z des recettes de boisson—estimation par la différence, utilisant le fait que le nombre de personnes servies (dans la population) est 800 .	Réponse : (A, B, ..., ou W) ►	S
vi) Le total τ_y —estimation par le quotient utilisant le fait que le nombre de personnes servies (dans la population) est 800 .	Réponse : (A, B, ..., ou W) ►	U
vii) Le montant total des factures destinées à une seule personne—estimation par la moyenne.	Réponse : (A, B, ..., ou W) ►	O
viii) Le montant total des factures destinées à une seule personne—estimation par le quotient utilisant le fait que le nombre de personnes servies (dans la population) est 800 .	Réponse : (A, B, ..., ou W) ►	N

1-b) Déterminer l'**écart-type** de chacun des estimateurs suivants :

i) L'estimateur en a)-i).	Réponse : (A, B, ..., ou W) ►	R
ii) L'estimateur en a)-iii)	Réponse : (A, B, ..., ou W) ►	B
iii) L'estimateur en a)-iv)	Réponse : (A, B, ..., ou W) ►	A
i) L'estimateur en a)-vii)	Réponse : (A, B, ..., ou W) ►	L

Choix de réponse pour le problème 1 [Choisir l'intervalle qui contient votre réponse] :

A 0,0200 à 0,0201	B 0,065 à 0,067	C 0,08275 à 0,08275	D 0,166 à 0,167	E 0,3120 à 0,3121	F 4,467 à 4,468	G 7,605 à 7,610	H 33,7 à 33,8	I 89,95 à 89,99	J 134 à 135	K 670 à 671	L 731 à 732
M 842 à 843	N 1339 à 1342	O 1785 à 1789	P 2695 à 2701	Q 2730 à 2734	R 3040 à 3046	S 10 960 à 10 970	T 11 230 à 11 232	U 26 990 à 26 995	V 35 980 à 35 986	W Aucune de ces réponses	

NOM.....

Problème 2 [16 points] *Mêmes contexte qu'au problème 1.*

On compte prélever un échantillon la semaine prochaine afin d'estimer certains paramètres. On estime que le nombre de factures n'aura pas changé—il demeurera 400.

- 2-a) Déterminer la taille de l'échantillon qu'il faudra tirer pour que la marge d'erreur dans l'estimation du des recettes totales soit de **4000 \$**. Vous estimerez les paramètres nécessaires à partir de votre échantillon en supposant que ces paramètres n'auront pas changé d'ici la semaine prochaine

-----Réponse (A, B, ..., P)

E

- 2-b) Déterminer la taille de l'échantillon qu'il faudra tirer pour que la marge d'erreur dans l'estimation des recettes totales soit de **5 %**. Vous estimerez les paramètres nécessaires à partir de votre échantillon en supposant que ces paramètres n'auront pas changé d'ici la semaine prochaine.

-----Réponse (A, B, ..., P)

I

- 2-c) Supposez maintenant que tous les prix seront majorés de **15 %** [sans pour autant affecter la consommation]. Déterminer la taille de l'échantillon qu'il faudra tirer pour que la marge d'erreur dans l'estimation des recettes totales soit de **4000 \$**.

-----Réponse (A, B, ..., P)

H

- 2-d) Déterminer la taille de l'échantillon qu'il faudra tirer pour que la marge d'erreur dans l'estimation de la proportion de factures servies à une seule personne soit de **0,02**. Vous estimerez les paramètres nécessaires à partir de votre échantillon en supposant qu'il n'y aura pas grand changement entre cette semaine et la semaine prochaine.

-----Réponse (A, B, ..., P)

J

Problème 3 [9+9+9 points] *Même contexte qu'aux problèmes 1 et 2.*

Les données présentées aux problèmes 1 et 2 portent sur les factures relatives *au souper*. Mais la population comprend également des factures relatives *au dîner* [le restaurant ouvre à midi]. Le nombre de factures pour le dîner est **600** dans la population. Un échantillon de taille **20** a été tiré. Les données sont présentées à l'annexe 2.

- 3-a) Utiliser les données des deux échantillons pour estimer les recettes totales du dîner et du souper.

-----Réponse (A, B, ..., P)

N

- 3-b) Déterminer l'écart-type de l'estimateur utilisé en a).

-----Réponse (A, B, ..., P)

K

- 3-c) Pour une prochaine fois, estimer l'allocation optimale pour estimer les recettes totales à partir d'un échantillon stratifié de taille **50**.

Nombre de factures à prélever :

Dîner (A, B, ..., P) :

15

Souper (A, B, ..., P)

35

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

A 7 à 9	B 14 à 16	C 34 à 36	D 41 à 43	E 62 à 64	F 67 à 69	G 71 à 72	H 79 à 81
------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------

I 190 à 194	J 310 à 312	K 3470 3480	L 4698 à 4702	M 6690 à 6694	N 48260 à 48264	O 62160 à 62170	P Aucune de ces réponses
-------------------	-------------------	-------------------	---------------------	---------------------	-----------------------	-----------------------	-----------------------------

NOM.....

Problème 4 [7+8 points]

On tire au hasard un échantillon de 400 employés dans une grande compagnie afin de déterminer s'il y a une relation entre le sexe de l'employé et le taux de ses absences. Voici les données :

Sexe de l'employé(e)	Taux d'absentéisme		Total
	Faible	Élevé	
Masculin	20	80	100
Féminin	120	180	300
Total	140	260	400

4-a) Dites lequel ou lesquels des énoncés suivants est (sont) acceptable(s) comme hypothèse nulle:
[Faites votre choix dans la liste présentée plus bas.]

A	Y a-t-il une relation entre le sexe et le taux d'absentéisme?
B	Il n'y a pas de relation entre le sexe et le taux d'absentéisme
C	La proportion de femmes avec un faible taux d'absentéisme est égale à la proportion d'hommes avec un faible taux d'absentéisme
D	La proportion de personnes avec un faible taux d'absentéisme est de 0,5
E	Il y a une relation entre le sexe et le taux d'absentéisme
F	La proportion de femmes avec un faible taux d'absentéisme est différente de la proportion d'hommes avec un faible taux d'absentéisme

Choix de réponse pour le problème 4-a):

Les énoncés acceptables sont [Encercler *un seul* des choix suivants] :

A, B et D A et C A, C et E A, C et D **B et C** C et E Aucune de ces
B et D B et F B, C et D D et E E et F F réponses

4-b) L'hypothèse nulle est : les variables *Sexe de l'employé* et *Taux d'absentéisme* sont [cochez]

indépendantes

dépendantes

4-c) Déterminer les effectifs théoriques [**Attention** : vos points ici ne compteront que si votre réponse en 4-b) est correcte].

Sexe de l'employé(e)	Taux d'absentéisme		Total
	Faible	Élevé	
Masculin	35	65	100
Féminin	105	195	300
Total	140	260	400

NOM.....

Problème 5 [10 points]

On tire un échantillon aléatoire simple de n ménages parmi les N ménages d'une ville.

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>	<i>Réponse</i> (A, ..., G)
<i>A</i> : La moyenne μ_y d'une certaine variable Y	1) La proportion de femmes parmi les membres du ménage	F
<i>B</i> : La moyenne μ_d s'une certaine variable Y dans un domaine \mathcal{D} .	2) Le nombre total de personnes dans la ville	D
<i>C</i> : Le total τ_d d'une certaine variable Y dans un domaine	3) Le nombre de ménages monoparentaux	E
<i>D</i> : Le total τ_y d'une certaine variable Y	4) Le nombre de personnes dans les ménages sans enfants	C
<i>E</i> : Le nombre N_c d'unités appartenant à une certaine classe \mathcal{C}	5) Le nombre moyen de téléphones cellulaires par personne	F
<i>F</i> : Le quotient $R = \mu_Y/\mu_X$ de deux variables Y et X	6) La proportion de personnes actives dans le ménage	F
<i>G</i> : La proportion p d'unités appartenant à une certaine classe	7) Le nombre total de téléphones cellulaires possédés par des enfants	D
	8) La valeur moyenne des téléphones cellulaires appartenant à des enfants	F
	9) Le nombre moyen de personnes dans les ménages situés en banlieue	B
	10) Le nombre moyen de téléviseurs par ménage	A

Annexe 1

Les données suivantes sont issues d'un échantillon aléatoire simple de **30** factures tirées parmi les **400** factures représentant les 400 soupers de la semaine.

	<i>Boisson (Z)</i>	<i>Nombre de personnes (X)</i>	<i>Facture totale (Y)</i>
	6.95	1	30.15
	7.42	1	19.32
	7.12	1	21.04
	5.62	1	31.43
	8.20	1	32.08
	22.14	2	72.44
	23.83	2	74.74
	30.05	2	68.74
	25.58	2	60.71
	0.00	2	57.35
	30.63	2	86.73
	25.21	2	70.57
	24.35	2	75.70
	29.56	2	87.66
	23.37	2	76.10
	28.22	2	84.60
	39.75	3	116.21
	0.00	3	73.64
	15.17	3	66.95
	50.85	3	119.48
	34.96	3	123.90
	59.71	4	179.35
	43.93	4	132.44
	40.12	4	151.33
	25.45	4	107.07
	38.37	4	132.27
	37.19	4	134.24
	34.56	4	154.20
	56.43	5	163.03
	67.56	5	95.64
<i>n</i>	30	30	30
Sommes	842,30	80,00	2699,11
Variances (corrigées)	302,46	1,47	1876,84
Covariances (corrigées)	Entre Z et X	16,36	
	Entre Z et Y	607,70	
	Entre X et Y	45,67	

Annexe 2

Les données suivantes sont issues d'un échantillon aléatoire simple de **20** factures tirées parmi les **600** factures représentant les dîners de la semaine

	<i>Nombre de personnes (X)</i>	<i>Facture totale (Y)</i>
	1	3.61
	1	19.80
	1	10.47
	1	10.71
	1	12.19
	1	5.75
	1	15.71
	1	4.80
	1	8.02
	2	15.24
	2	24.25
	2	27.60
	2	12.35
	3	25.50
	3	24.60
	3	27.20
	4	41.34
	4	34.07
	4	40.48
	4	45.43
<i>n</i>	20	20
Sommes	42	409,12
Variances (corrigées)	1,46	162,02
Covariance (corrigée) Entre Boisson et Facture totale	14,10	

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

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

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 $n = \frac{n_o}{1 + \frac{n_o}{N}}$ 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 $\sigma_{\bar{y}_{st}} =$

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

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

$$\text{estimé par } \hat{\sigma}_{\hat{p}_{st}} = \sqrt{W_h^2 \hat{\sigma}_{\hat{p}_h}^2} \text{ où } \hat{\sigma}_{\hat{p}_h}^2 = (1-f_h) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}$$

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 crtiques ($\alpha = 5\%$) 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

