Université du Québec à Montréal
Département de mathématiques

Corrigé

MAT 2080 **STATISTIQUE EN GESTION 1**

EXAMEN FINAL HIVER 04

Date : Vendredi 30 avril 2004, de 18h00 à 21h00

Nom :

Prénom :

Code permanent : **Groupe:**

INSTRUCTIONS

1. Prendre grand soin de ne pas désassembler les feuilles du présent cahier (6 pages annexe + formulaire + tables), qui doit être remis en entier. Seuls l'annexe, le formulaire et les tables peuvent être détachés du cahier et n'ont pas à être retournés.
2. Par mesure de précaution, inscrire lisiblement votre nom au haut de chacune des pages 2 à 6.
3. Les solutions doivent être rédigées dans les espaces prévus. Ne pas négliger d'expliquer clairement votre démarche, de donner les détails de vos calculs et d'identifier clairement les variables considérées.
4. Si l'espace est insuffisant, indiquer clairement que la solution se poursuit au verso de la page.
5. 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.***
6. Vous trouverez à la fin de ce cahier deux feuilles blanches, pour fins de calcul-brouillon.
7. L'usage d'une calculatrice est autorisé.
8. L'étudiant doit présenter sa carte d'étudiant (avec photo) lors de la remise de son cahier et signer la feuille de présence.

Grille à l'usage du correcteur

1-a)	1-b)	1-c)	1-d)	1-e)	2-a)	2-b)
/11	/11	/11	/11	/11	/6	/9
2-c)	**3-a)**	**3-b)**		*Total* :		
/10	/10	/10			/100	

Question 1 [11+11+11+11+11 *points*]
On tire un échantillon aléatoire simple de taille **30** dans une population de **1 250** commerces de détail. On retient l'information suivante sur chacun d'eux :

Nombre d'employés
Masse salariale (la somme des salaires payés aux employés)

Vous trouverez les données en annexe.

1-a) Estimer μ, la masse salariale moyenne par commerce, et déterminer un intervalle de confiance pour μ.

$$\bar{y} = \frac{3842}{30} = 128{,}067. \quad \hat{\sigma}_{\bar{y}} = \sqrt{1-\frac{30}{1250}}\frac{49{,}6553}{\sqrt{30}} = 8{,}956.$$

Intervalle de confiance : $128{,}067-2(8{,}956) \le \mu \le 128{,}067+2(8{,}956)$

$110{,}088 \le \mu \le 145{,}979$

Masse salariale moyenne : $\boxed{128{,}067}$; *Écart-type de l'estimateur* $\boxed{8{,}956}$

Intervalle de confiance : $\boxed{110{,}088} \le \boxed{\mu} \le \boxed{145{,}979}$

1-b) Utiliser le fait que le nombre total d'employés dans la population est de **6 277** pour estimer *par le quotient* la masse salariale moyenne μ des commerces de la population, et déterminer un intervalle de confiance pour μ

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum y_i}{\sum x_i} = \frac{3842}{152} = 25{,}2763158;$$

$$\mu_X = \frac{6277}{1250} = 5{,}0216$$

Estimation : $\hat{R}\mu_X = 25{,}2763158(5{,}0216) = 126{,}928$

$$\hat{\sigma}_{\hat{R}\mu_X} = \frac{\sqrt{1-30/1250}}{\sqrt{30}}\sqrt{(49{,}6553)^2 + \hat{R}^2(1{,}99885)^2 - 2\hat{R}(98{,}1678)} = 1{,}3454$$

Masse salariale moyenne : $\boxed{126{,}927}$; *Écart-type de l'estimateur* $\boxed{1{,}3454}$

Intervalle de confiance : $\boxed{124{,}24} \le \boxed{\mu} \le \boxed{129{,}62}$

1-c) Estimer la *proportion p* de commerces ayant 6 employés ou plus et déterminer un intervalle de confiance pour *p*

$\hat{p} = \frac{10}{30}$; $\hat{\sigma}_{\hat{p}} = \sqrt{1 - \frac{30}{1250}}\sqrt{\frac{(1/3)(2/3)}{30-1}} = 0{,}08648$

Intervalle de confiance :

$$\frac{10}{30} - 2(0{,}08648) \le p \le \frac{10}{30} + 2(0{,}08648)$$

$0{,}160 \le p \le 0{,}506$

Proportion : $\boxed{1/3}$; *Écart-type de l'estimateur* $\boxed{0{,}08648}$

Intervalle de confiance : $\boxed{0{,}160} \le \boxed{p} \le \boxed{0{,}506}$

1-d) Estimer la masse salariale *totale* des commerces ayant **9** employés et déterminer un intervalle de confiance pour la masse salariale totale.

$\bar{y}' = \frac{217 + 214 + 232}{30} = \frac{663}{30} = 22{,}1$;

$$s'^2 = \frac{(217 - 22{,}1)^2 + (214 - 22{,}1)^2 + (232 - 22{,}1)^2}{29}$$

$T = 1250(22{,}1) = 27625$

$\hat{\sigma}_{N\bar{y}'} = 1250\sqrt{1 - \frac{30}{1250}}\frac{s'}{\sqrt{30}} = 15\,214$

Intervalle de confiance : $27\,625 - 2(15\,214) \le \tau_d \le 27\,625 + 2(15\,214)$

$-2803 \le \tau \le 58083$

Masse salariale totale : $\boxed{27\,625}$; *Écart-type de l'estimateur* $\boxed{15\,214}$

Intervalle de confiance : $\boxed{0} \le \boxed{\tau_d} \le \boxed{58\,053}$

1-e) On compte reprendre cet échantillonnage afin d'estimer (par la moyenne) la masse salariale totale des commerces avec une marge d'erreur relative de 10 %. Utiliser les données de cet échantillon pour estimer la taille de l'échantillon qu'il faudra prélever.

$n_o = \left(\frac{2S}{R\mu}\right)^2$. On estime S par s= 49,6553 et μ par $\bar{y} = \frac{3842}{30} = 128{,}067$.

On a donc $n_o = \left[\frac{2(49{,}6553)}{(0{,}1)(128{,}067)}\right]^2 = (7{,}75458)^2 = 60{,}13$.

$$n = \frac{60{,}13}{1 + \frac{60{,}13}{1250}} = 57$$

n = $\boxed{57 \text{ ou } 58}$

Question 2 [6 + 9 +10 *points*]

Des militants antiféministes veulent démontrer par un test d'hypothèse qu'une certaine compagnie discrimine contre les hommes dans l'engagement du personnel saisonnier. Ils prélèvent des données sur les demandes d'emploi temporaire faites par **1 000** étudiantes et étudiants un certain été et obtiennent les résultats suivants :

		Résultat de la demande		
		Engagement	Rejet	
Sexe	Masculin	110	190	300
	Féminin	290	410	700
	Tous	400	600	1000

2-a) [Bonne réponse :1 point ; mauvaise réponse : -1 point; pas de réponse : 0; note minimale : 0]

Vrai ou faux? L'hypothèse nulle est…

(i)	La probabilité qu'une femme soit engagée est égale à la probabilité qu'un homme soit engagé	*vrai*: ☒	*faux*: ☐
(ii)	Il n'y a pas de discrimination contre les hommes	*vrai*: ☐	*faux*: ☒
(iii)	La probabilité qu'une femme soit engagée est de 0,7	*vrai*: ☐	*faux*: ☒
(iv)	Les variables « Masculin » et « Féminin » sont indépendantes	*vrai*: ☐	*faux*: ☒
(v)	Les variables « Sexe de l'étudiant(e) » et « Résultat de la demande » sont indépendantes	*vrai*: ☒	*faux*: ☐
(vi)	La probabilité qu'une femme soit engagée est de 50 % ou moins	*vrai*: ☐	*faux*: ☒

2-b) Inscrire dans le tableau les effectifs théoriques et calculer la valeur de χ^2

		Résultat de la demande		
		Engagement	Rejet	
Sexe	Masculin	120	180	300
	Féminin	280	420	700
	Tous	400	600	1000

$$\chi^2 = \frac{(110-120)^2}{120} + \frac{(190-180)^2}{180} + \frac{(290-280)^2}{280} + \frac{(410-420)^2}{420} = 1{,}98$$

$\chi^2 = \boxed{1{,}98}$

2-c) [Bonne réponse :1 point; mauvaise réponse : -1 point; pas de réponse : 0; note minimale : 0]

Vrai ou faux? La valeur de χ^2 obtenue en b) nous fait dire que…

(i)	Il y a une dépendance entre les variables « Sexe » et « Résultat de la demande »	*vrai*: ☐	*faux*: ☒
(ii)	La probabilité qu'une femme soit engagée est de 70 %	*vrai*: ☐	*faux*: ☒
(iii)	Le hasard tout seul ne suffit pas à expliquer les différences de pourcentages	*vrai*: ☐	*faux*: ☒
(iv)	On peut affirmer avec 95 % de confiance que le résultat de la demande ne dépend pas du sexe	*vrai*: ☐	*faux*: ☒
(v)	Les différences entre les valeurs observées et les valeurs théoriques peuvent être attribuées au hasard	*vrai*: ☒	*faux*: ☐
(vi)	La probabilité qu'une femme soit engagée est inférieure à 50 %	*vrai*: ☐	*faux*: ☒
(vii)	Les variables « Engagée » et « Rejetée » sont indépendantes	*vrai*: ☐	*faux*: ☒
(viii)	Il y a 5 % de chance que l'hypothèse d'indépendance soit fausse	*vrai*: ☐	*faux*: ☒
(ix)	On ne peut pas affirmer avec confiance qu'il y a discrimination	*vrai*: ☒	*faux*: ☐
(x)	On peut affirmer avec confiance qu'il y a discrimination *en faveur* des hommes	*vrai*: ☐	*faux*: ☒

Question 3 [10+10 *points*]

Dans une compagnie, on classe l'ensemble des comptes de dépenses des employés en trois strates selon la classe d'employés. On tire un échantillon stratifié afin d'estimer la valeur moyenne μ des frais de voyage. Voici les données :

Strate 1 Techniciens et professionnels	$N_1 = 7\ 000$; $n_1 = 70$; $\bar{y}_1 = 345$ \$; $s_1 = 40$ \$
Strate 2 Cadres	$N_2 = 2\ 000$; $n_2 = 20$; $\bar{y}_2 = 680$ \$; $s_2 = 75$ \$
Strate 3 Cadres supérieurs	$N_3 = 1\ 000$; $n_3 = 10$; $\bar{y}_3 = 1500$ \$; $s_3 = 300$ \$

3-a) Estimer les dépenses moyennes μ par compte et déterminer un intervalle de confiance pour μ

$$\bar{y}_{st} = 0,7(345) + 0,2(680) + 0,1(1500) = 527,5$$

$$\hat{\sigma}^2_{\bar{y}_1} = \left(1 - \frac{70}{7000}\right)\frac{(40)^2}{70} = 22,6285 \;;$$

$$\hat{\sigma}^2_{\bar{y}_2} = \left(1 - \frac{20}{2000}\right)\frac{(75)^2}{20} = 278,4375 \;;$$

$$\hat{\sigma}^2_{\bar{y}_3} = \left(1 - \frac{10}{1000}\right)\frac{(300)^2}{10} = 8910$$

$$\hat{\sigma}^2_{\bar{y}_{st}} = (0,7)^2(22,62857)+(0,2)^2(278,4375)+(0,2)^2(8910) = 111,3255$$

$\hat{\sigma}_{\bar{y}_{st}} = 10,55$. Intervalle de confiance : $506,4 \le \mu \le 548,6$

Estimation ponctuelle : $\boxed{527,5}$ *Écart-type de l'estimateur* $\boxed{10,55}$

Intervalle de confiance : $\boxed{506,4} \le \boxed{\mu} \le \boxed{548,6}$

3-b) L'échantillon présenté au numéro précédent est un échantillon pilote destiné uniquement à préparer un nouvel échantillonnage stratifié plus important, de taille **500**. Déterminer la répartition de cet échantillon de taille 500 qui minimisera la variance de l'estimateur de μ.

On estime les S_h par les s_h

$W_1s_1 = 28$; $W_2s_2 = 15$; $W_3s_3 = 30$.

$\Sigma W_h s_h = 73$.

$$n_1 = \frac{28}{73}500 = 192 \;;\quad n_2 = \frac{15}{73}500 = 103 \;;\quad n_3 = \frac{30}{73}500 = 205$$

$n_1 = \boxed{192}$; $n_2 = \boxed{103}$; $n_3 = \boxed{205}$

Annexe

Données sur un échantillon aléatoire simple de taille **30** tiré
d'une population de **1 250** petits commerces de détail

Voici les données :

	Nombre d'employés	Masse salariale (en milliers de $)
	9	217
	9	214
	9	232
	7	162
	7	183
	7	169
	7	185
	6	157
	6	150
	6	148
	5	126
	5	138
	5	135
	5	128
	5	140
	5	128
	5	121
	5	123
	4	101
	4	99
	4	109
	4	109
	4	107
	4	103
	4	86
	3	72
	3	78
	2	49
	2	51
	1	22
Somme	152	3842
Écart-type (corrigé) s	1,99885	49,6553
Covariance (corrigée) s_{xy}	98,1678	

Formulaire MAT2080 Examen final

<u>Résumé des paramètres, leur estimateur, l'écart-type de l'estimateur, et l'estimateur de l'écart-type de l'estimateur.</u>

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* $\hat{N}_d = \frac{n_d}{n}N$		$N\sqrt{1-f}\,\frac{s'}{\sqrt{n}}$

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

<u>Taille d'échantillon</u>

Estimation de la moyenne La taille d'échantillon nécessaire pour que la marge d'erreur absolue soit égale à E est $n = \frac{n_o}{1+\frac{n_o}{N}}$ 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 $n = \frac{n_o}{1+\frac{n_o}{N}}$

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 $n_o = \frac{4(1-p)}{R^2 p}$.

<u>Échantillonnage par strates</u>

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 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

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

<u>Test du khi-deux</u>

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

<u>Points crtiques (α = 5 %) d'une loi khi-deux</u>

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