Université du Québec à Montréal
Département de mathématiques

Corrigé

MAT 2080	STATISTIQUE EN GESTION 1
	EXAMEN FINAL AUTOMNE 2003
Date : Mercredi 17 décembre 2003, de 18h00 à 21h00	

Nom :

Prénom :

Code permanent : **Groupe:**

INSTRUCTIONS

1. Prendre grand soin de ne pas désassembler les feuilles du présent cahier (7 pages + tables + formulaire), 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 7.
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 sévèrement sanctionné par les hautes instances universitaires.***
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	2	3	4	5	6	7
/32	/18	/12	/10	/10	/8	/10

Total : /100

Problème 1 [7+7+4+7+7 points]

Les données en annexe portent sur un échantillon aléatoire simple de **34** hôpitaux d'une population de **700** hôpitaux. Pour chaque hôpital de l'échantillon, on note les valeurs des variables suivantes:

`lits:`	le nombre de lits dans l'hôpital
`chambres:`	le nombre de chambres
`univ:`	précise si l'hôpital est un hôpital universitaire ou non

1-a) Déterminer un intervalle de confiance à 95 % pour le *nombre moyen de lits par hôpital*

$n = 34,\ N = 700,\ \bar{y} = \dfrac{14433}{34} = 424{,}50\,.\ \ \hat{\sigma}_{\bar{y}} = \sqrt{1 - \dfrac{34}{700}}\,\dfrac{\sqrt{159951}}{\sqrt{34}} = 66{,}9025\,.$

Intervalle de confiance : $424{,}5-2(66{,}9025) \le \mu_y \le 424{,}5+2(66{,}9025)$

$290{,}695 \le \mu_y \le 558{,}3050$

Nombre moyen de lits par hôpital : $\boxed{424{,}50}$; *Écart-type de l'estimateur* : $\boxed{66{,}9025}$

Intervalle de confiance : $\boxed{290{,}605} \le \boxed{\mu_y} \le \boxed{558{,}3050}$

1-b) Déterminer un intervalle de confiance à 95 % pour le *nombre moyen de lits par chambre*.

$\hat{R} = \dfrac{\bar{y}}{\bar{x}} = \dfrac{14433}{3971} = 3{,}6346$

$\hat{\sigma}_{\hat{R}} = \dfrac{\sqrt{1-n/N}}{\bar{x}\sqrt{n}}\sqrt{s_y^2 + \hat{R}s_x^2 - 2\hat{R}s_{xy}} = \dfrac{\sqrt{1-34/700}}{(3971/34)\sqrt{34}}\sqrt{159951 + \hat{R}^2(15713) - 2\hat{R}(49968)} = 0{,}0953$

Intervalle de confiance : $3{,}6346-2(0{,}0953) \le R \le 3{,}6346+2(0{,}0953)$

Nombre moyen de lits par chambre : $\boxed{3{,}6346}$; *écart-type de l'estimateur* $\boxed{0{,}0953}$

Intervalle de confiance : $\boxed{3{,}444} \le \boxed{R} \le \boxed{3{,}825}$

1-c) Sachant qu'il y a en tout **86 200** chambres dans les hôpitaux de la population, estimer *par le quotient* le nombre total de lits (*estimation ponctuelle seulement*).

$T_{y_q} \;=\; \dfrac{\bar{y}}{\bar{x}}\tau_x = \dfrac{14\,433}{3971}86\,200 = 313303$

Nombre total de lits : $\boxed{313\ 303}$

1-d) Déterminer un intervalle de confiance pour *le nombre total de lits dans les hôpitaux non universitaires* ayant moins de 10 chambres.

Nous définissons la variable y' dont les valeurs sont toutes nulles sauf les quatre qui correspondent à des hôpitaux non universitaires ayant moins de 10 chambres. Ces valeurs non nulles sont 31 ; 40 ; 35; et 63.

On a donc $\bar{y}' = \dfrac{31+40+35+63}{34} = \dfrac{169}{34}$.

$\hat{T}_d = N\bar{y}' = 700\dfrac{169}{34} = 3479$;

$$s'^2 = \frac{(31-169/34)^2+(40-169/34)^2+(35-169/34)^2+(63-169/34)^2+(0-169/34)^2\times 30}{34-1}$$

$= \dfrac{6914{,}9706}{33} = 209{,}5446.$

$$\hat{\sigma}_{\hat{T}_d} = N\sqrt{1-\frac{n}{N}}\frac{s'}{\sqrt{n}} = 700\sqrt{1-\frac{34}{700}}\frac{\sqrt{209{,}5446}}{\sqrt{34}} = 1695$$

Intervalle de confiance : $3479-2(1695) \le \tau_d \le 3479+2(1695)$

$89 \le \tau_d \le 6869$

Estimation ponctuelle : 3479

Intervalle de confiance : 89 $\le \tau_d \le$ 6869

1-e) Vous avez l'intention de refaire ce sondage afin d'estimer (par la moyenne) *le nombre moyen de lits* par hôpital. Vous voulez que votre marge d'erreur relative soit de **10 %**. Estimer *la taille de l'échantillon* (aléatoire simple) que vous devrez tirer.

$n_o = \dfrac{(2S)^2}{(0{,}1\mu)^2}$. Utilisant les données de l'échantillon, on obtient

$$n_o = \frac{4(159951)}{[(0{,}1)(14433/34)]^2} = 355{,}051672$$

$$n = \frac{355{,}051672}{1+\dfrac{355{,}051672}{700}} = 235{,}5677$$

n : 236

Problème 2 [9+9 points]

Mêmes données qu'au problème 1, sauf que vous avez d'un coup la surprise d'apprendre, par hasard, que l'échantillon actuel n'est pas, comme vous le croyiez, un échantillon aléatoire simple: c'est un échantillon *stratifié*. Vous apprenez, en outre, que les effectifs des strates sont connus: la strate 1, constituée des hôpitaux universitaires, contient **140** hôpitaux; et la strate 2, comprenant tous les autre hôpitaux, en contient **560**.

2-a) Déterminer un intervalle de confiance pour le *nombre moyen de lits par hôpital*

Les tailles des strates sont $N_1 = 140$; $N_2 = 560$; $W_1 = 0{,}2$; $W_2 = 0{,}8$

Les moyennes étant échantillonnales sont $\bar{y}_1 = \dfrac{10777}{12} = 898{,}0833$ et $\bar{y}_2 = \dfrac{3656}{22} = 166{,}1818$;

$s_1^2 = 84582$; $s_2^2 = 8981$; $\bar{y}_{st} = (0{,}2)\dfrac{10777}{12} + (0{,}8)\dfrac{3656}{22} = 145{,}8955$

$\hat{\sigma}^2_{\bar{y}_1} = \left(1 - \dfrac{12}{140}\right)\dfrac{84582}{12} = 6444$; $\hat{\sigma}^2_{\bar{y}_2} = \left(1 - \dfrac{22}{560}\right)\dfrac{8981}{22} = 392{,}1898$

$\hat{\sigma}_{\bar{y}_{st}} = \sqrt{(0{,}2)^2(6444) + (0{,}8)^2(392{,}1898)} = 22{,}56$

Intervalle de confiance : $145{,}8955 - 2(22{,}56) \le \mu_y \le 145{,}8955 + 2(22{,}56)$

Nombre moyen de lits par hôpital : $\boxed{145{,}8955}$; *Écart-type de l'estimateur* $\boxed{22{,}56}$

Intervalle de confiance : $\boxed{100{,}78} \le \boxed{\mu_y} \le \boxed{191{,}02}$

2-b) Vous comptez tirer un échantillon stratifié de taille **40** pour estimer le *nombre moyen de lits par hôpital*. Comment répartir ces 40 observations entre les deux strates si vous voulez minimiser la variance de l'estimateur? (La strate 1, constituée des hôpitaux universitaires, contient **140** hôpitaux; et la strate 2, comprenant tous les autre hôpitaux, en contient **560**.)

N_1S_1 et N_2S_2 sont estimés par $N_1s_1 = 140\sqrt{84582} = 40716$; $N_2s_2 = 560\sqrt{8981} = 53070$.

On devrait donc avoir

$n_1 = \dfrac{40716}{40716 + 53070} \times 40 = 17$ et $n_2 = 23$

n_1 : $\boxed{17}$; n_2 : $\boxed{23}$

Problème 3 [5+2+5 points]

Une compagnie qui fait souvent des sondages d'opinion s'interroge sur l'utilité d'envoyer des cadeaux pour encourager les gens à répondre. On décide d'envoyer des questionnaires à **300** personnes : **100** personnes ne reçoivent aucun cadeau; à **100** autres, on promet un abonnement de six mois à un magazine; et à **100** autres enfin on promet un portefeuille en cuir de crocodile. On note alors le nombre de répondants dans chaque groupe. Voici les résultats

	Cadeau promis			
A répondu	*Aucun*	*Abonnement de 6 mois*	*Portefeuille en cuir de crocodile*	*Total*
Oui	30	40	50	120
Non	70	60	50	180
Total	100	100	100	300

3-a) Énoncer l'hypothèse nulle dans le langage du contexte

Il est inutile d'envoyer des cadeaux ; ou
La probabilité de réponse est la même, quelle que soit la promesse; ou
Le fait de répondre ou pas est indépendant de l'offre faite.

3-b) Déterminer les effectifs théoriques :

	Cadeau promis			
A répondu	***Aucun***	***Abonnement de 6 mois***	***Portefeuille en cuir de crocodile***	*Total*
Oui	40	40	40	120
Non	60	60	60	180
	100	100	100	300

3-c) Compléter le test et exprimer votre conclusion dans le langage du contexte

$$\chi^2 = \frac{(30-40)^2}{40} + \frac{(40-40)^2}{40} + \frac{(50-40)^2}{40} + \frac{(70-60)^2}{60} + \frac{(60-60)^2}{60} + \frac{(50-60)^2}{60}$$

$$= \frac{100}{40} + 0 + \frac{100}{40} + \frac{100}{60} + 0 + \frac{100}{60} = 8{,}33$$

Cette valeur étant supérieure 5,9915, on rejette l'hypothèse H_o et on conclut que la promesse faite (ou l'absence de promesse) influence la volonté de répondre.

Problème 4 [10 points]

Afin de déterminer si dans un sondage téléphonique on peut faire augmenter le taux de réponse en prenant un ton doucereux, un enquêteur fait 120 appels sur un ton doucereux et 120 autres sur un ton ferme. Il compte ensuite le nombre de personnes qui acceptent de répondre dans chaque groupe. Enfin, il effectue un test du khi-deux.

4-a) Vrai ou faux? Si la valeur calculée de khi-deux est 15 …

(i)	on conclut que les différences entre les deux taux de réponse sont entièrement dus au hasard	*vrai*: ☐ *faux*: X
(ii)	on peut conclure avec 95% de confiance que le ton n'a pas d'effet sur le taux de réponse	*vrai*: ☐ *faux*: X
(iii)	on conclut que la différence entre les deux groupes est trop importante pour être attribuée au hasard	*vrai*: X *faux*: ☐
(iv)	il y a 5 % de chance que l'hypothèse soit fausse	*vrai*: ☐ *faux*: X
(v)	on peut affirmer avec 95 % de confiance que le ton affecte la probabilité de réponse	*vrai*: X *faux*: ☐

4-b) Vrai ou faux? Si la valeur calculée de khi-deux est 1,4 …

(i)	on peut conclure avec 95% de confiance que le ton n'a pas d'effet sur le taux de réponse	*vrai*: ☐ *faux*: X
(ii)	on conclut que la différence entre les deux groupes est trop importante pour être attribuée au hasard	*vrai*: ☐ *faux*: X
(iii)	il y a 5 % de chance que l'hypothèse soit fausse	*vrai*: ☐ *faux*: X
(iv)	on ne peut pas conclure que le ton a un effet sur le taux de réponse	*vrai*: X *faux*: ☐
(v)	on peut affirmer avec 5 % de confiance que le ton a un effet	*vrai*: ☐ *faux*: X

Problème 5 [10 points]

On tire un échantillon aléatoire simple de n familles dans un quartier habité par N familles. On observe pour chaque famille la valeur des variables suivantes:

Nombre d'adultes dans la famille
Nombre de personnes dans la famille
Revenu annuel de la famille

et on note aussi si la famille est *monoparentale ou pas*.
Dans chacune des situations suivantes, dire précisément quel paramètre il s'agit d'estimer. Faire votre choix parmi les suivants (le numéro 0 ci-dessous illustre le type de réponse attendu):

A. La moyenne μ_y d'une variable Y. Identifier Y.
B. Le total τ_y d'une variable Y. *Identifier* Y
C. La moyenne μ_d d'un domaine. *Identifier la variable et le domaine.*
D. Le total τ_d d'une variable Y dans un domaine $\mathfrak{D}$. *Identifier* Y *et* $\mathfrak{D}$.
E. La proportion p d'unités appartenant à une certaine classe. *Identifier la classe*.
F. Le nombre N_c d'unités appartenant à une certaine classe. *Identifier la classe.*
G. Le quotient de deux moyennes $R = \dfrac{\mu_y}{\mu_x}$. *Identifier* Y *et* X.

	Réponse (A, B , … , ou F)
0. Le nombre moyen de personnes par famille *Il s'agit d'une moyenne* μ_y *où* Y *est le nombre de personnes dans la famille*	*A*
1. Le nombre moyen d'adultes par famille μ_y *où* Y *est le nombre d'adultes dans la famille*	*A*
2. La proportion de familles monoparentales *p, la proportion de familles dans la classe des familles monoparentales*	*E*
3. La proportion d'adultes dans le quartier $R = \bar{y}/\bar{x}$ *où* Y *est le nombre d'adultes et* X *le nombre de personnes dans la famille*	*G*
4. Le nombre de personnes dans le quartier τ_y *où* Y *est le nombre de personnes dans la familles*	*B*
5. Le nombre de personnes dans le quartier appartenant à des familles monoparentales τ_d *où* Y*, le nombre de personnes dans la famille et* $\mathcal{D}$ *est le domaine des familles mono-parentales*	*D*
6. Le nombre moyen d'adultes dans les familles monoparentales μ_d *où* Y *est le nombre d'adultes dans la famille et* $\mathcal{D}$ *est le domaine des familles mono-parentales*	*C*
7. Le nombre de familles monoparentales dans le quartier N_c *le nombre de familles dans la classe des familles monoparentales*	*F*
8. Le revenu annuel par personne $R = \bar{y}/\bar{x}$ *,* Y *= revenu annuel de la famille,* X *= nombre de familles*	*G*
9. Le revenu annuel par famille μ_y *où* Y *est le revenu annuel de la familles*	*A*
10. Le revenu total des familles monoparentales τ_d où Y *est le revenu de la famille et* $\mathfrak{D}$ *est le domaine des familles monoparentales*	*D*

Problème 6 [8 points]

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.

	Réponse (*A, B, C, D ou E*)
1) Population: l'ensemble des ménages d'une petite ville. On divise la ville en 5 secteurs, puis on tire au hasard 15 maisons dans chaque secteur.	B
2) Population: l'ensemble des ménages d'une petite ville. On divise la ville en 35 secteurs; puis on tire 5 secteurs au hasard. L'échantillon comprend tous les ménages des 5 secteurs choisis.	D
3) Population: les 150 succursales d'une banque. L'échantillon est formé en tirant au hasard 25 succursales dans le centre-ville; et 32 succursales en banlieue.	B
4) Population: l'ensemble des auteurs des livres d'une bibliothèque municipale. On tire au hasard 100 *titres* dans la liste de tous les titres. L'échantillon est l'ensemble des auteurs des titres choisis.	E
5) Population: l'ensemble de toutes les factures envoyées par une compagnie à ses clients le mois dernier. On tire au hasard 50 factures. Pour chaque facture tirée, on ajoute à l'échantillon toutes les factures envoyées au même client.	E
6) Population: l'ensemble des étudiants de l'UQAM. On tire au hasard trois programmes, et on inclut dans l'échantillon tous les étudiants inscrits dans ces trois programmes	D
7) Population: l'ensemble des étudiants de l'UQAM. On dresse une liste de tous les étudiants, puis on tire successivement et sans remise 300 étudiants.	A
8) Population: l'ensemble des clients à qui on a envoyé une facture le mois dernier. On tire au hasard 200 factures. Les clients à qui ces factures ont été envoyées constituent l'échantillon.	E

Problème 7 [10 points]

Répondre par *vrai* (V) ou *faux* (F)

	Réponse (*V ou F*)
1) L'estimateur par le quotient peut être beaucoup plus efficace que l'estimateur par la moyenne si la corrélation entre la variable d'intérêt (Y) et la variable auxiliaire (X) est forte	V
2) L'estimateur par la différence peut être beaucoup plus efficace que l'estimateur par la moyenne si la corrélation entre la variable d'intérêt (Y) et la variable auxiliaire (X) est forte	V
3) Pour estimer une moyenne, un échantillon de **50** familles ayant en tout **250** personnes est généralement plus précis qu'un échantillon de **250** personnes tirées au hasard dans la population.	F
4) Un estimateur *sans biais* est d'autant meilleur que son écart-type est grand	F
5) L'estimateur de la moyenne par le quotient n'est pas sans biais	V
6) L'estimateur de la moyenne par la différence est sans biais	V
7) Un estimateur est dit sans biais s'il est égal au paramètre	F
8) Dans l'échantillonnage systématique, l'estimateur de la moyenne n'a pas de variance	F
9) L'estimateur par le quotient est *toujours* plus efficace que l'estimateur par la moyenne	F
10) Un échantillon tiré par grappes contient nécessairement des éléments de chaque grappe	F

Annexe

Nombre de lits et nombre de chambres dans un échantillon de **34** *hôpitaux*
Taille de la population : **700** *hôpitaux*

Lits (y)	*Chambres* (x)	*Univ*	*Lits* (y)	*Chambres* (x)	*Univ*	*Lits* (y)	*Chambres* (x)	*Univ*
823	245	OUI	216	44	NON	121	27	NON
1343	395	OUI	170	31	NON	280	60	NON
908	274	OUI	252	53	NON	188	39	NON
648	202	OUI	38	11	NON	35	6	NON
1043	321	OUI	170	30	NON	134	23	NON
1325	407	OUI	179	38	NON	63	9	NON
1123	327	OUI	31	9	NON	280	57	NON
690	205	OUI	40	9	NON	75	18	NON
519	163	OUI	295	61	NON	142	25	NON
1118	328	OUI	336	70	NON	293	59	NON
522	152	OUI	166	37	NON	152	31	NON
715	205	OUI						

Données pour l'ensemble de l'échantillon (n=34)

	Lits (y)	*Chambres* (x)
Sommes	$\sum y_i$ = 14 433	$\sum x_i$ = 3 971
Variances corrigées	s_y^2 = 159 951	s_x^2 = 15 713
Covariance s_{xy}	s_{xy} = 49 968	

Données pour les 12 hôpitaux universitaires

	Lits (y)	*Chambres* (x)
Sommes	$\sum y_i$ = 10 777	$\sum x_i$ = 3 224
Variances corrigées	s_y^2 = 84 582	s_x^2 = 7 516
Covariance s_{xy}	s_{xy} = 25 123	

Données pour les 22 hôpitaux non universitaires

	Lits (y)	*Chambres* (x)
Sommes	$\sum y_i$ = 3 656	$\sum x_i$ = 747
Variances corrigées	s_y^2 = 8 981	s_x^2 = 385
Covariance s_{xy}	s_{xy} = 1 844	

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*)	$T_d = N_d\,\bar{y}_d$		$Nd\sqrt{1-\frac{n_d}{N_d}}\ \frac{s_d}{\sqrt{n_d}}$
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* $\hat{N}_d = \frac{n_d}{n}N$		$N\sqrt{1-f}\ \frac{s'}{\sqrt{n}}$

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

<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 $n = \dfrac{n_o}{1+\dfrac{n_o}{N}}$ 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'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

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

Points crtiques (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