

Exercice 1

1. Le fichier Excel 'M2 Ex01 données 21.xls' contient une liste d'une population de  $N = 3\,048$  personnes. Le fichier contient une variable d'identification et deux variables d'enquête : l'indice de masse corporelle (IMC) de chaque répondant ainsi qu'une variable binaire (1=oui, 0=non) indiquant si le répondant s'est déjà fait dire par un médecin qu'il avait cholestérol élevé (CHLEV). Tirer un échantillon aléatoire simple de taille  $n = 40$  en utilisant le tableau de nombres aléatoires dans « M2 Random Digits 21.pdf ». Par souci de réplication, sélectionnez les exemples de nombres aléatoires en commençant dans le coin supérieur gauche, ligne 1, colonnes 1-4 (premier nombre, 4901). Continuez à sélectionner des nombres en descendant les colonnes 1-4. Si plus de 50 nombres aléatoires sont nécessaires pour sélectionner l'échantillon (c'est-à-dire que les 50 nombres des colonnes 1 à 4 ont été utilisés), passez à la ligne 1, colonnes 6 à 9 (numéro 3404). Si plus de 100 nombres aléatoires sont nécessaires pour sélectionner l'échantillon, continuez avec la ligne 1, colonnes 11-14 (numéro 7200). Répondez ensuite aux questions suivantes concernant l'échantillon :
  - (a) Énumérez les numéros d'identification (ID) et les deux variables (IMC et CHLEV) des unités échantillonnées.
  - (b) Estimez l'indice de masse corporelle (IMC) moyen et comparez-le à la moyenne de la population pour l'IMC. S'ils diffèrent, expliquez brièvement pourquoi.
  - (c) Estimez la variance d'élément de l'IMC, la variance d'échantillonnage et l'erreur type de l'IMC moyen.
  - (d) Calculez un intervalle de confiance à 95% pour l'IMC moyen.
  - (e) Estimez la proportion de personnes à qui un médecin a déjà dit qu'elles avaient un taux de cholestérol élevé (CHLEV = 1), son erreur standard et son intervalle de confiance à 90 %.
  - (f) Si la taille de l'échantillon d'un SRS de cette population était augmentée de  $n = 40$  à  $n = 100$ , quelle serait l'erreur type estimée de l'IMC moyen ?
  - (g) Quelle taille d'échantillon  $n$  pour un SRS de cette population est nécessaire pour obtenir un coefficient de variation d'au plus 0,025 à la fois pour l'IMC moyen et la proportion de personnes à qui on a déjà dit qu'elles souffraient d'hypertension ?
  - (h) Calculez un intervalle de confiance à 99 % pour le nombre total de personnes à qui un médecin a déjà dit qu'elles avaient un taux de cholestérol élevé.

**Percentiles of the  $t$  distribution**

v	$1 - \alpha/2$			
	0.95	0.975	0.99	0.995
1	6.31	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.05
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.72	2.09	2.53	2.85
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.50	2.81
24	1.71	2.06	2.49	2.80
25	1.71	2.06	2.49	2.79
26	1.71	2.06	2.48	2.78
27	1.70	2.05	2.47	2.77
28	1.70	2.05	2.47	2.76
29	1.70	2.05	2.46	2.76
30	1.70	2.04	2.46	2.75
31	1.70	2.04	2.45	2.74
32	1.69	2.04	2.45	2.74
33	1.69	2.03	2.44	2.73
34	1.69	2.03	2.44	2.73
35	1.69	2.03	2.44	2.72
36	1.69	2.03	2.43	2.72
37	1.69	2.03	2.43	2.72
38	1.69	2.02	2.43	2.71
39	1.68	2.02	2.43	2.71
40	1.68	2.02	2.42	2.70
$\infty(z)$	1.64	1.96	2.33	2.58