1. Rappels de mathématiques

Exercice 1

Transcrire sous une autre forme les expressions suivantes :

1) $a + a + a + a$	2) $a - a - a - a$	3) $\mid a - a - a - a \mid$
4) $a.a.a.a$	5) $\dfrac{1}{a.a.a.a}$	6) $a^{1/2}$
7) $a.\dfrac{1}{a^2}.a^{1/2}.a^{-5}.a^{-5/2}$	8) $a(b+c)$	9) $a.(b.c)$
10) $a + (b.c)$	11) $\dfrac{1}{a}(b+c)$	12) $\dfrac{1}{a}(b.c)$
13) $\dfrac{1}{a}(a+c)$	14) $\dfrac{1}{a}(a.b)$	15) $\dfrac{1}{a}(b+c)$
16) $(a+b).(c+d)$	17) $(a+b)+(c+d)$	18) $(a + b(c+d) + c(f.g)).h$
19) $\left(a^b\right)^c$	20) $(a+b)^2$	21) $(a-b)^2$

Exercice 2

1) Sur un axe muni d'un repère normé, placer les points A d'abscisse 3 et B d'abscisse -2. Calculer la distance $d(A, B)$ et le carré de cette distance.

2) Comment calcule-t-on l'hypoténuse d'un triangle rectangle ? Dans le plan euclidien, calculer la distance $d(A, B)$ sachant que les coordonnées de A et B sont respectivement (3 ; 2) et (5 ; 8).

Réponses : 1) $d(A, B) = |-2-3| = 5$ *;* $d^2(A, B) = 25$. *2)* $d^2(A, B) = (x_B - x_A)^2 + (y_B - y_A)^2 = 40$ *;* $d(A, B) = \sqrt{40} = 6.32$

Exercice 3 *Notations indexées et symbole* $\sum$.

On considère le tableau suivant :

	1	2	3	4	5	6	7	8
a	8	3	2	9	6	2	8	5
b	4	2	9	7	10	1	3	6
c	7	6	1	3	8	2	1	3

1) Donner la valeur des symboles suivants : a_3, b_1.

2) Citer deux éléments du tableau qui ont pour valeur 3.

3) Ecrire à l'aide du symbole $\sum$: $a_1 + a_2 + \cdots + a_8$.

4) Pour chacune des expressions suivantes :

1. La développer de manière symbolique
2. Remplacer chaque symbole par sa valeur
3. Calculer la valeur de l'expression.

a) $\sum_{i=3}^{i=8}(a_i + b_i)$ b) $\sum_{i=2}^{i=5}(a_i b_i c_i)$ c) $\sum_{i=3}^{i=8} a_i^2$

d) $\sum_{i=3}^{i=6} i.a_i$ e) $\sum c_i$

f) $\dfrac{1}{n}\sum_{i=1}^{i=n}(a_i - m)^2$ sachant que $m = \dfrac{\sum a_i}{n}$ et $n = 3$.

Réponses : 4) $\sum_{i=3}^{i=8}(a_i + b_i) = 68$ *;* $\sum_{i=2}^{i=5}(a_i b_i c_i) = 723$ *;* $\sum_{i=3}^{i=8} a_i^2 = 214$ *;* $\sum_{i=3}^{i=6} i.a_i = 84$ *;* $\sum c_i = 31$ *;*

$m = 4.33$; $\dfrac{1}{n}\sum_{i=1}^{i=n}(a_i - m)^2 = 20.67$

Exercice 4

On considère les deux tableaux ci-dessous :

a_{ij} $i \setminus j$	1	2	3	4	5	6	7	8
1	3	1	2	8	3	1	6	7
2	6	3	1	10	7	9	2	4
3	5	8	2	6	9	2	3	8

b_{ij} $i \setminus j$	1	2	3	4	5	6	7	8
1	5	8	2	7	6	3	8	1
2	9	6	5	3	8	6	2	3
3	1	10	9	8	6	3	1	4

1) Donner la valeur des symboles suivants : $a_{3\,2}$, $b_{3\,5}$.
2) Citer deux éléments du tableau des (a_{ij}) qui ont pour valeur 6.
3) Pour chacune des expressions suivantes : – la développer de manière symbolique ; – remplacer chaque symbole par sa valeur ; – calculer la valeur de l'expression.

a) $\sum_{i=1}^{i=3}\sum_{j=1}^{j=6} a_{ij}^2$ b) $\sum_{i=1}^{i=3}\sum_{j=1}^{j=3} a_{ij}b_{ij}$ c) $\sum_{i=1}^{i=3}\sum_{j=5}^{j=8}(a_{ij}+b_{ij})$

d) $\sum_{i=1}^{i=3} a_{i\,4}$ e) $\sum_{i,j} b_{ij}$

Réponses : $\sum_{i=1}^{i=3}\sum_{j=1}^{j=6} a_{ij}^2 = 578$; $\sum_{i=1}^{i=3}\sum_{j=1}^{j=3} a_{ij}b_{ij} = 207$; $\sum_{i=1}^{i=3}\sum_{j=5}^{j=8}(a_{ij}+b_{ij}) = 112$; $\sum_{i=1}^{i=3} a_{i\,4} = 24$; $\sum_{i,j} b_{ij} = 124$.

Exercice 5

On a recensé régulièrement la population d'une agglomération en distinguant trois catégories : âge scolaire, vie active, troisième âge. Les résultats sont consignés dans le tableau suivant :

	1975	1980	1985	1990
Age scolaire	299	271	265	270
Vie active	320	340	310	350
Troisième âge	162	182	212	241

1) Calculer en pourcentages la part de chaque catégorie pour chacun des recensements.
2) Calculer le taux d'augmentation de chaque catégorie entre 1980 et 1985.
3) Calculer les indices des effectifs de chaque catégorie de la population (base 100 en 1980).

Réponses : 1)

	1975	*1980*	*1985*	*1990*
Age scolaire	*38%*	*34%*	*34%*	*31%*
Vie active	*41%*	*43%*	*39%*	*41%*
Troisième âge	*21%*	*23%*	*27%*	*28%*

2)

	Variation
Age scolaire	*-2.21%*
Vie active	*-8.82%*
Troisième âge	*16.5%*

3)

	1975	*1980*	*1985*	*1990*
Age scolaire	*110*	*100*	*98*	*99*
Vie active	*94*	*100*	*91*	*102*
Troisième âge	*89*	*100*	*116*	*132*

Exercice 6

Chômage : + 1,3%. La poussée du mois d'août

(...) au mois d'août, le nombre de demandeurs d'emplois a augmenté de 39500, soit une hausse de 1,3% par rapport à la fin juillet. (...) il y a actuellement 3.085.100 chômeurs. (...) leur nombre à 3.085.100, soit 5,3% de plus qu'un an auparavant.

(...)Le taux de chômage en France au sens du BIT a atteint 12,6% de la population active en août contre 12,5% en juillet.

Le Télégramme de Brest - 28/09/96

Vérifier la cohérence de ces données. Calculer celles qui sont citées sans que leur valeur ne soit indiquée.

Les exercices suivants proposent une série de tableaux. Pour chacun d'eux :

1. *identifier le type de tableau (tableau protocole, tableau d'effectifs, tableau de contingence)*
2. *identifier ses différentes composantes et leurs caractéristiques (les individus statistiques et leur nombre, les variables statistiques, leur nombre, leur nature et leur champ de variation*
3. *si nécessaire, donner au tableau une présentation plus "normale"*
4. *identifier les questions qu'un psychologue pourrait se poser à partir du tableau*

Exercice 7

On réalise une expérience sur des groupes de 3 personnes en vue d'analyser leurs perceptions et réactions face à différents stimuli. On obtient le tableau de données suivant :

Groupe d'âge	Réussite des tests	Classement selon la vitesse de réalisation	Impression de réussite	Composition
jeune	50	5	3	Jacques, Michel, Louis
adulte	25	3	5	Fabienne, Marc, Patrick
vieux	20	10	10	Suzanne, François, Marie
vieux	45	8	4	Daniel, Annie, Jean
adulte	60	6	6	Pascale, Sabine, Louise
vieux	50	2	8	Jean, Philippe, Simone
jeune	70	4	4	Loïc, Maïwen, Sophie
jeune	40	7	3	Pierre, Géraldine, Frédéric
jeune	45	1	10	Estelle, Samuel, Marc
adulte	33	9	2	Henri, Jeanne, Françoise

Réponses : Les individus statistiques sont des groupes de 3 personnes. L'effectif total est de 10. Il s'agit d'un tableau protocole concernant 4 variables : le groupe d'âge (ordinale, 3 modalités), la réussite (numérique), le classement (ordinal, domaine $[0, 10]$*), et l'impression (numérique, domaine* $[0, 10]$*). Parmi les questions que l'on peut se poser : la réussite dépend-elle de l'âge ? l'impression de réussite est-elle liée à la réussite, la vitesse de réalisation dépend-elle de l'âge ? est-elle liée à la réussite ? etc.*

Exercice 8

On dispose des données suivantes concernant huit patients atteints de dépression.

1. mesure du degré de la dépression établi sur la base de différents critères au moment de l'arrivée du patient (échelle de 0 à 10)
2. âge des sujets
3. une mesure du stress lié au travail (échelle allant de 0 à 50)

Les données sont les suivantes :
Données 1 : 5, 8, 10, 3, 7, 9, 4, 6
Données 2 : 30, 25, 50, 50, 28, 42, 60, 35
Données 3 : 6, 20, 48, 6, 8, 32, 5, 10.
Réponses : La population est formée de 8 sujets. Il s'agit d'un tableau protocole. Trois variables ont été observées : une mesure du degré de dépression, l'âge et une mesure du stress lié au travail.

Exercice 9
Le tableau suivant donne la distribution des scores obtenus à un test de développement intellectuel pour trois groupes professionnels.

Notes	Manœuvres	Ouvriers qualifiés	Cadres moyens	Toutes catégories
40				
39		1		1
38				
37				
36			1	1
35			1	1
34		1		1
33			2	2
32		1	2	3
31	1	2	1	4
30	1	2	2	5
29	1	1	3	5
28	2	2	2	6
27	3	3	4	10
26	2	3	4	9
25	3	4	6	13
24	3	6	8	17
23	5	6	7	18
22	5	6	6	17
21	7	7	5	19
20	7	6	4	17
19	6	6	2	14
18	7	7	3	17
17	6	5	3	14
16	7	7	3	17
15	7	7	2	16
14	7	6	2	15
13	5	6	1	12
12	6	6		12
11	5	3		8
10	3	2	1	6
9	2	2	1	5
8	2	2		4
7	2	1		3
6	2	2	1	5
5	1	1		2
4				
3	1			1
2				
1				
0				
N	109	114	77	300

Réponses : La population est formée de 300 sujets. Deux variables ont été observées : la catégorie professionnelle (nominale, 3 modalités) et le score au test (numérique, domaine $[0, 40]$*. Le tableau proposé est un tableau de contingence. La principale question que l'on* *peut se poser est : les scores observés diffèrent-ils de façon significative d'une catégorie à l'autre ?*

Exercice 10
Au cours de certaines expériences, on est amené à mesurer le *temps de réaction* (TR) des sujets. C'est le temps qui s'écoule entre la présentation d'un stimulus (par exemple, une lampe qui s'allume devant le sujet) et la réaction que ce stimulus doit déclencher (par exemple, presser un bouton).
Première expérience. – Le tableau 1 fournit les TR d'une personne qui a réagi 20 fois à l'allumage d'une lampe rouge. On constate que ces 20 TR ne sont pas égaux. Ces variations d'un moment à l'autre sont imprévisibles à partir des informations dont on dispose dans l'expérience.
Deuxième expérience. – Le sujet voit maintenant s'allumer devant lui une lampe qui peut être rouge, verte ou jaune. il doit réagir si la lampe est rouge, mais ne doit pas réagir dans les deux autres cas. Le tableau 1 fournit 20 TR mesurés dans ces conditions. On observe de nouveau des variations imprévisibles d'un moment à l'autre.
Troisième expérience. – Les conditions sont les mêmes que dans la première expérience (une seule lampe) avec une seule différence : au lieu d'être rouge, la lampe donnant le signal de la réaction est verte. La troisième ligne du tableau donne les résultats. Les temps sont de nouveau différents entre eux.

Numéro d'ordre des présentations	20	1	2	3	4	5	6	7	8	9	10
1ère expérience		20	15	18	25	17	32	18	17	19	23
2è expérience		32	40	33	37	35	29	42	62	50	39
3è expérience		16	18	19	18	15	18	17	32	23	19

Numéro d'ordre des présentations	20	11	12	13	14	15	16	17	18	19	20
1ère expérience		19	21	15	22	17	17	21	19	17	23
2è expérience		45	47	52	37	38	39	40	41	42	39
3è expérience		23	20	18	25	15	15	17	23	17	19

Réponses : La population est constituée par les 60 expériences. Trois variables statistiques ont été relevées : la condition expérimentale (nominale, 3 modalités), le numéro d'ordre (ordinale) et le temps de réaction (numérique). Questions à se poser : la condition expérimentale a-t-elle un effet sur le TR ? le TR varie-t-il selon le numéro d'ordre (effet de fatigue ou d'apprentissage) ?

Exercice 11 *Etude du choix professionnel en fonction du sexe.*

Choix professionnel	Sexe								
	Effectifs			Pourcentages par colonne			Pourcentages par ligne		
	G	F	Ensemble	G	F	Ensemble	G	F	Ensemble
Médecins	153	139	292	2.9	2.8	2.8	52.4	47.8	100
Ingénieurs	653	29	682	12.5	0.6	6.8	95.8	4.2	100
Cadres supérieurs	90	23	113	1.7	0.5	1.1	79.6	20.4	100
Instituteurs	327	698	1025	6.3	13.8	10.0	31.8	68.2	100
Techniciens	1013	176	1189	19.5	3.5	11.5	85.2	14.8	100
Cadres moyens	319	649	968	6.1	12.9	9.5	32.9	67.1	100
Employés	269	788	1057	5.2	15.6	10.4	25.5	74.5	100
Ouvriers qualifiés	202	43	245	3.9	0.9	2.4	82.5	17.5	100
Autres choix	2181	2500	4681	41.8	49.6	45.7	46.6	53.4	100
Total	5207	5045	10252	100	100	100	50.6	49.2	100

Réponses : Population formée de 10252 sujets. Tableau de contingence croisant deux variables nominales : le sexe et le choix professionnel.

Exercice 12 *Ordination des sensations*
On peut demander à un sujet d'ordonner les sensations qu'il éprouve lorsqu'il est soumis à un stimulus dont la mesure physique varie. On pourra vérifier la cohérence de ses réponses en observant si elles satisfont aux deux critères d'antisymétrie et de transitivité. Exemple : mesure d'un seuil différentiel kinesthésique.
Le seuil différentiel est la plus petite différence entre deux valeurs d'une même stimulation, suffisante pour permettre une réponse discriminative. Il peut être mesuré dans le domaine kinesthésique à l'aide d'un petit appareil appelé "gravimètre" mis au point part Piéron. Cet appareil permet de comparer deux masses en les soupesant alternativement avec le même index : l'une est l'étalon (soit, par ex. 150 g), l'autre le stimulus variable, que l'expérimentateur fait varier suivant un plan réglé avant l'expérience entre des valeurs inférieures et supérieures à celles de l'étalon (par ex. 138 g. 142 g. 146 g. 150 g. 154 g. 158 g. 162 g. 166 g). Les différentes valeurs du stimulus variable sont présentées dans un ordre fortuit, à plusieurs reprises (le plus souvent, on effectue 10 séries).

Tableau des résultats obtenus avec un sujet

Stimulus variable	138 g	142 g	146 g	150 g	154 g	158 g	162 g	166g
1^{re} série	−	−	=	+	+	+	+	+
2^e série	−	−	+	+	+	+	+	+
3^e série	−	−	−	=	=	−	+	+
4^e série	−	−	−	−	=	+	+	+
5^e série	−	−	+	−	−	+	+	+
6^e série	−	=	=	+	+	+	+	+
7^e série	−	−	−	−	+	=	+	+
8^e série	−	−	=	−	=	−	+	+
9^e série	−	−	−	−	−	+	=	+
10^e série	−	−	−	−	=	+	+	+
Total des R +	0	0	2	3	4	7	9	10
Total des R −	10	9	5	6	2	2	0	0
Total des R =	0	1	3	1	4	1	1	0

Les signes +, −, = indiquent les réponses du sujet (+ signifie que le sujet apprécie le stimulus comme étant "plus lourd" que l'étalon, = signifie "aussi lourd", − signifie "moins lourd").

Réponses : Population de 80 expériences. Tableau protocole donnant les valeurs de trois variables : un numéro d'ordre (de 1 à 10), la masse du stimulus et la réponse du sujet (ordinale, 3 modalités). Les trois dernières lignes du tableau forment un tableau de contingence entre les deux dernières variables.

Exercice 13

Pour un groupe de 30 élèves, on dispose du nombre de redoublements durant leur scolarité et de leur situation de famille. Les redoublements enregistrés vont de 0 à 4. Les situations de famille sont les suivantes :

A. enfant de parents séparés vivant à la maison d'un des deux

B. enfant dont un des parents est décédé vivant à la maison du survivant

C. enfant vivant à la maison avec ses deux parents

D. enfant vivant en internat mais dont les parents ne sont pas séparés

E. enfant vivant en internat et dont les parents sont séparés ou l'un des deux décédé.

Le tableau ci-dessous rassemble les données recueillies.

Enfant	Famille	Redoublement	Enfant	Famille	Redoublement	Enfant	Famille	Redoublement
01	E	0	11	B	1	21	E	0
02	A	2	12	E	0	22	C	0
03	B	1	13	C	2	23	C	1
04	A	0	14	B	0	24	B	1
05	C	1	15	A	1	25	D	1
06	B	3	16	C	4	26	B	0
07	D	0	17	D	3	27	C	0
08	C	2	18	E	2	28	A	1
09	D	1	19	A	0	29	D	0
10	A	0	20	A	2	30	C	2

Réponses : Population de 30 sujets, tableau protocole pour deux variables : le type de famille et le nombre de redoublements.

Exercice 14

Dans une enquête d'opinion, on a posé la question ouverte suivante : "Quelles sont les raisons qui, selon vous, peuvent faire hésiter une femme ou un couple à avoir des enfants ?" A partir des réponses libres, on a retenu un ensemble de 16 mots pertinents (peur, santé, avenir, chômage, etc.) et compté le nombre d'occurrences de ces mots dans divers corpus constitués à partir des catégories sociales. Voici, pour le mot *chômage* et pour l'ensemble des 16 mots retenus, le nombre d'occurrences dans cinq corpus constitués selon le niveau scolaire du répondant : (1) sans diplôme ; (2) CEP ; (3) BEPC ; (4) Baccalauréat ; (5) Université, Grandes écoles, etc.

	(1)	(2)	(3)	(4)	(5)	Total
Chomage	71	111	50	40	11	283
les 16 mots	365	606	378	358	164	1871

Par exemple, dans le corpus des réponses fournies par les personnes interrogées qui sont sans diplôme, on compte 71 occurrences du mot *chômage* parmi 365 occurrences des 16 mots pertinents.

Réponses : Les notions de population et de variable statistique ne s'appliquent pas de façon naturelle ici. On peut cependant choisir comme population les 1871 positions de mots repérées dans chacun des corpus. Les variables sont alors le corpus d'origine (nominale ou ordinale, 5 modalités) et la nature du mot (nominale, 16 modalités).

Exercice 15

Lors d'une enquête portant sur une population de 4830 élèves de 3è, on a noté d'une part la profession du chef de famille et d'autre part le choix professionnel de l'élève. Les résultats sont rassemblés dans le tableau suivant.

	Choix professionnel de l'élève								
Profession du père	A	B	C	D	E	F	G	H	Ensemble
Agriculteurs	4	7	7	8	8	12	12	10	8
Salariés agricoles	0	1	0	4	2	4	3	2	2
Patrons	10	12	13	11	12	8	8	15	12
Pr. lib,cadres sup.	40	18	22	4	5	6	2	4	12
Cadres moyens	24	28	21	23	19	18	9	13	21
Employés	1	4	11	7	6	8	10	5	6
Contremaîtres	7	14	14	18	24	17	19	23	16
O.S., Manœuvres	3	7	8	14	15	13	25	23	12
Personnel de service	3	0	0	3	2	4	3	1	2
Autres	8	9	4	8	7	10	9	4	9
Totaux	100	100	100	100	100	100	100	100	100
Effectifs	143	613	82	306	953	297	242	190	4830

N.B. L'effectif total, 4830, est plus élevé que la somme des effectifs des 8 colonnes parce que l'ensemble des choix professionnels comprend d'autres catégories que les 8 catégories figurant dans le tableau.

Codage des catégories : A : médecins, B : ingénieurs, C : cadres supérieurs, D : instituteurs, E : techniciens, F : cadres moyens, G : ouvriers qualifiés.

1) Préciser la population et les variables statistiques étudiées. Quel est le type (protocole, effectifs, etc) du tableau ci-dessus ?

2) Critiquer le commentaire suivant :

> Sur 100 élèves choisissant la profession de cadres moyens, on constate que 18% ont des pères qui exercent cette profession. Ce pourcentage est élevé : il correspond au mode de la distribution. Il n'y a rien de surprenant à ce que les pères cadres moyens soient fortement représentés dans le groupe d'élèves ayant choisi la même profession. Par contre, on ne trouve dans ce groupe que 8% de pères employés, ce qui est bien peu.

Réponses : 1) Il s'agit d'un tableau de contingence. Les valeurs indiquées sont des fréquences par colonne et non des effectifs.

2) Attention aux bases de calcul des pourcentages. Ce faible pourcentage résulte seulement de ce que les enfants d'employés sont peu représentés dans la population étudiée.

Représentations graphiques

Exercice 16

On a recensé les ménages d'une localité en relevant le nombre d'enfants à charge de chacun d'eux. Les données obtenues sont indiquées ci-dessous :
6 - 3 - 0 - 0 - 1 - 1 - 0 - 1 - 1 - 4 - 5 - 4 - 1 - 4 - 4 - 1 - 6 - 0 - 3 - 0 - 0 - 8 - 3 - 1 - 1 - 0 - 4 - 2 - 1 - 3 - 2 - 2 - 0 - 1 - 1 - 3 - 4 - 2 - 2 - 1 - 1 - 2 - 1 - 0 - 0 - 0 - 4 - 1 - 1 - 1 - 0 - 0 - 1 - 2 - 2 - 1 - 7 - 2 - 1 - 1 - 5 - 1 - 2 - 1 - 0 - 1 - 1 - 0 - 3 - 1 - 4 - 0 - 1 - 0 - 0 - 6 - 0 - 2 - 1 - 4 - 5 - 3 - 1 - 3 - 0 - 1 - 3 - 3 - 2 - 5 - 2 - 0 - 0 - 0 - 3 - 5 - 3 - 3 - 2 - 5 - 2 - 2 - 2 - 2 - 2 - 5 - 0 - 2 - 3 - 1 - 1 - 2 - 4 - 1 - 3 - 5 - 1 - 1 - 0 - 0 - 2 - 6 - 2 - 2 - 1 - 3 - 1 - 2 - 0 - 1 - 1 - 0 - 6 - 2 - 2 - 0 - 0 - 5 - 1 - 0 - 4 - 1 - 2 - 0 - 0 - 4 - 2 - 6 - 3 - 3 - 2 - 2 - 0 - 0 - 4 - 2 - 2 - 3 - 5 - 1 - 0 - 5 - 1 - 1 - 2 - 1 - 0 - 0 - 0 - 2 - 3 - 3 - 1 - 5 - 6 - 6 - 3 - 5 - 1 - 2 - 4 - 0 - 4 - 1 - 2 - 2 - 1 - 1 - 0 - 4 - 2 - 1 - 0 - 1 - 2 - 4 - 1 - 5 - 1 - 0 - 1 - 2 - 4 - 1 - 2 - 1 - 1 - 1 - 1 - 3 - 3 - 1 - 1 - 0 - 1 - 0 - 0 - 1 - 1 - 2 - 0 - 2 - 4 - 0 - 3 - 5 - 3 - 3 - 7
Construire le tableau d'effectifs correspondant et représenter la distribution statistique à l'aide d'un diagramme approprié.
Réponse : Tableau d'effectifs :

Nb enfants	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Nb ménages	*48*	*65*	*44*	*27*	*19*	*15*	*8*	*2*	*1*

Effectif total : 229.

Exercice 17

On reprend les données de l'exercice 14 (enquête d'opinion)
Représenter à l'aide d'un diagramme à bandes le nombre d'occurrences des mots pertinents dans chacun des corpus. Comment faire apparaître sur le même graphique le nombre d'occurrences du mot *chômage*?
Réponse : Le mot "chômage" est l'un des 16 mots retenus. Ses occurrences sont déjà représentées par le graphique. On pourra indiquer à l'aide d'une couleur différente la partie de chaque rectangle correspondant au mot "chômage"

Exercice 18

Dans l'expérience complète de Mendel, deux attributs des pois étaient considérés : la couleur (*jaune* ou *verte*, et la forme (*ronde* ou *ridée*), avec la triple hypothèse suivante : *rond* dominant et *ridé* récessif ; *jaune* dominant et *vert* récessif ; indépendance entre couleur et forme. Le modèle mendélien prédit alors qu'à la deuxième génération, on obtiendra en moyenne la distribution suivante :

Types de pois	*rond jaune*	*rond vert*	*ridé jaune*	*ridé vert*
Fréquences théoriques	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

On réalise l'expérience de Mendel et on obtient les résultats expérimentaux suivants :

Types de pois	*rond jaune*	*rond vert*	*ridé jaune*	*ridé vert*
Effectifs	325	116	101	30

Représenter à l'aide d'un diagramme circulaire les données expérimentales.
Réaliser un diagramme à bandes juxtaposées montrant l'adéquation entre les données expérimentales et le modèle théorique.

Exercice 19

On reprend les données de l'exercice 11.

En ne conservant que les choix professionnels qui ont une fréquence supérieure à 5% dans la population, réaliser deux diagrammes semi-circulaires donnant la distribution des choix dans chacun des deux groupes.

Exercice 20
On reprend les données de l'exercice 15.
Représenter à l'aide d'un diagramme à bandes la distribution des métiers du père pour le groupe d'élèves voulant devenir médecins et pour le groupe d'élèves voulant devenir instituteurs.

Exercice 21
On reprend les données de l'exercice 13 (nombre de redoublements).
Réaliser un diagramme en bâtons pour la variable "nombre de redoublements". Réaliser un diagramme à barres superposées en prenant en compte les deux variables étudiées.

Exercice 22 *Données "Spatial"*
Lors d'une enquête réalisée en 1964 sur 5154 garçons de Troisième, on a relevé les notes (nombre de bonnes réponses à 48 questions) à un test spatial. Il s'agissait de constituer une "distribution de référence" sur les capacités spatiales des garçons de Troisième. Soit X la variable statistique étudiée.

x_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13
n_i	1	2	1	1	2	2	5	12	10	19	29	42	58	65
x_i	14	15	16	17	18	19	20	21	22	23	24	25	26	27
n_i	80	110	162	139	152	185	193	213	270	333	315	310	263	280
x_i	28	29	30	31	32	33	34	35	36	37	38	39	40	41
n_i	312	239	224	215	181	167	153	83	93	68	74	42	20	12
x_i	42	43	44	45	46	47	48							
n_i	8	4	3	1	0	1	0							

Réaliser un diagramme en bâtons de cette distribution.

Découpage en classes

Exercice 23
On reprend les données de l'exercice 9 (test de développement et groupe professionnel).
1) Faire un découpage en 5 classes de même amplitude et réaliser les histogrammes des quatre séries statistiques envisagées.
2) Comparer l'histogramme de la série des notes "toutes catégories" aux histogrammes empilés des trois groupes de sujets.
Réponses : 1)

Classe	*Man.*	*Ouv. Qual.*	*Cadres*	*Total*
$[0, 8[$	*6*	*4*	*1*	*11*
$[8, 16[$	*37*	*34*	*7*	*78*
$[16, 24[$	*50*	*50*	*33*	*133*
$[24, 32[$	*16*	*23*	*30*	*69*
$[32, 40[$	*0*	*3*	*6*	*9*
Total	*109*	*114*	*77*	*300*

Si l'on choisit la même échelle dans les deux cas, les deux graphiques coïncident.

Exercice 24

On reprend les données "Spatial" (exercice 22).
1) Faire un découpage en 8 classes d'égale amplitude ([0; 6[, [6; 12[, ... [42; 48[. Réaliser l'histogramme correspondant.

2) Faire un découpage en 8 classes d'effectifs approximativement égaux (environ 650 individus par classe) et réaliser l'histogramme correspondant.
Réponses : 1)

Classe	$[0,6[$	$[6,12[$	$[12,18[$	$[18,24[$	$[24,30[$	$[30,36[$	$[36,42[$	$[42,48]$
Effectifs	*9*	*117*	*614*	*1346*	*1719*	*1023*	*309*	*17*

2)

Classe	$[0,17[$	$[17,21[$	$[21,24[$	$[24,26[$	$[26,28[$	$[28,31[$	$[31,35[$	$[35,48]$
Effectifs	*601*	*669*	*816*	*625*	*543*	*775*	*716*	*409*
Densité	*35*	*167*	*272*	*312*	*271*	*258*	*179*	*31*

Exercice 25 *Pyramides d'Egypte*
Les données suivantes représentent la structure par classes d'âge et par sexe de la population de l'Egypte en 1947. Représenter ces données à l'aide de deux histogrammes disposés en "pyramide des âges".

Age	[0 ;5[	[5 ;10[	[10 ;15[	[15 ;20[	[20 ;25[	[25 ;30[
Hommes	1280	1209	1142	984	678	686
Femmes	1305	1191	1071	917	706	787
Age	[30 ;35[	[35 ;40[	[40 ;45[	[45 ;50[	[50 ;55[	[55 ;60[
Hommes	620	659	569	429	421	171
Femmes	690	654	566	415	449	173
Age	[60 ;65[	[65 ;70[	[70 ;75[	[75 ;80[	[80 ;85[	[85 ;90[
Hommes	252	84	108	23	35	17
Femmes	299	82	137	24	53	23

Exercice 26
Les notes obtenues par 128 élèves à un test de niveau ont été réparties en 5 classes : [0 ;10[(insuffisant), [10 ;12[(passable), [12 ;14[(assez bien), [14 ;16[(bien) et [16 ;20] (très bien). On a obtenu le tableau suivant :

Classe	[0 ;10[	[10 ;12[	[12 ;14[	[14 ;16[	[16 ;20]
Effectif	35	35	31	19	8

Construire l'histogramme correspondant à ce découpage en classes. Utiliser cet histogramme pour évaluer le nombre d'élèves ayant obtenu une note comprise entre 8 et 11.
Réponse : Les élèves ayant obtenu une note comprise entre 8 et 11 représentent $\frac{2}{10}$ *de la première classe et* $\frac{1}{2}$ *de la seconde. Leur nombre peut être évalué à 24.*

Fonction de répartition

Exercice 27

On reprend les données de l'exercice 13 (Situation de famille et nombre de redoublements). Définir la fonction de répartition de la variable statistique "Nombre de redoublements" et construire sa représentation graphique.

Réponse :

Redoublements	*0*	*1*	*2*	*3*	*4*	
Effectifs	*12*	*9*	*6*	*2*	*1*	
Eff. cum.	*0*	*12*	*21*	*27*	*29*	*30*

Exercice 28

On reprend les données de l'exercice 26 (notes d'élèves).

Définir la fonction de répartition de la variable statistique étudiée et construire sa représentation graphique.

Réponse : La fonction de répartition est une fonction affine par intervalle, continue, croissante, caractérisée par :

x	*0*	*10*	*12*	*14*	*16*	*20*
$F(x)$	*0*	*0.27*	*0.55*	*0.79*	*0.94*	*1*

Exercice 29

On reprend les données "Spatial" (exercice 22) et les découpages en classes définis dans l'exercice 24. Représenter graphiquement les fonctions de répartition correspondantes.

Réponses : Avec le découpage en classes donné en réponse dans l'exercice 24, question 1), on obtient :

x	*0*	*6*	*12*	*18*	*24*	*30*	*36*	*42*	*48*
n_c	*0*	*9*	*126*	*740*	*2086*	*3805*	*4828*	*5137*	*5154*
$F(x)$	*0*	*.0017*	*.024*	*.14*	*.40*	*.74*	*.94*	*.997*	*1*

Avec le découpage en classes de la question 2), on obtient :

x	*0*	*17*	*21*	*24*	*26*	*28*	*31*	*35*	*48*
n_c	*0*	*601*	*1270*	*2086*	*2711*	*3254*	*4029*	*4745*	*5154*
$F(x)$	*0*	*.12*	*.25*	*.40*	*.53*	*.63*	*.78*	*.92*	*1*

Exercice 30

On reprend les données de l'exercice 9 (score à un test et groupe professionnel) et le découpage en 5 classes de même amplitude défini dans l'exercice 23.

1) Définir et représenter graphiquement la fonction de répartition de la variable "Notes", toutes catégories confondues.

2) Représenter graphiquement les fonctions de répartition (fréquences cumulées) des variables "Notes" de chacun des groupes professionnels, sur un même dessin (utiliser trois couleurs différentes).

Réponses : Les fonctions de répartitions sont définies par :

	x	*0*	*8*	*16*	*24*	*32*	*40*
Man.	n_{1c}	*0*	*6*	*43*	*93*	*109*	*109*
	$F_1(x)$	*0*	*.06*	*.39*	*.85*	*1*	*1*
O.Q.	n_{2c}	*0*	*4*	*38*	*88*	*111*	*114*
	$F_2(x)$	*0*	*.04*	*.33*	*.77*	*.97*	*1*
C.M.	n_{3c}	*0*	*1*	*8*	*41*	*71*	*77*
	$F_3(x)$	*0*	*.01*	*.10*	*.53*	*.92*	*1*
Ens.	n_c	*0*	*11*	*89*	*222*	*291*	*300*
	$F(x)$	*0*	*.04*	*.30*	*.74*	*.97*	*1*

Exercice 31

On a fait passer le test Stroop (épreuve de résistance au stress) à 432 étudiants. La distribution des notes (différences de temps par item en centièmes de seconde) est donnée, après regroupement en classes d'étendue 5, par le tableau suivant :

Note	10	15	20	25	30	35	40	45	50	55	60	65	70	75
Effectif	4	9	9	20	45	45	57	49	51	27	25	21	19	11
Note	80	85	90	95	100	105	110	115	120	125	130	135	140	145
Effectif	8	5	6	3	4	1	1	1	0	2	1	1	3	0
Note	150	155	160	165	170	175	180	185	190	195	200	205		
Effectif	0	0	0	0	1	0	0	0	1	1	0	1		

1) Procéder à un nouveau regroupement en 5 classes avec pour limites de classes : -0.5, 12.5, 37.5, 62.5, 87.5, 205.5. Compléter le tableau obtenu en calculant la fréquence et la densité (effectif par unité d'amplitude) de chacune des classes.

2) En utilisant le regroupement en classes établi à la question 1, construire un tableau définissant la fonction de répartition de la série statistique étudiée.

3) En utilisant le regroupement en classes établi à la question 1, construire l'histogramme et la représentation graphique de la fonction de répartition de la série. Evaluer graphiquement la médiane et les quartiles. Calculer la moyenne et l'écart type.

Réponses :

1)

Classes	*Effectif*	*Fréquences*	*Amplitude*	*Densité*
$[-0.5; 12.5[$	*4*	*0,93%*	*13*	*0.31*
$[12.5; 37.5[$	*128*	*29.63%*	*25*	*5.12*
$[37.5; 62.5[$	*209*	*48.38%*	*25*	*8.36*
$[62.5; 87.5[$	*64*	*14.81%*	*25*	*2.56*
$[87.5; 205.5[$	*27*	*6.25%*	*118*	*0.23*

2) Fonction de répartition : fonction affine par intervalles

x_i	*-0.5*	*12.5*	*37.5*	*62.5*	*87.5*	*205.5*
$F(x_i)$	*0*	*0.0093*	*0.3056*	*0.7894*	*0.9375*	*1*

3) $M = 47.5$, $Q_1 = 32.8$, $Q_3 = 60.5$.

Caractéristiques de position

Exercice 32
On reprend les données de l'exercice 13. Quel est le mode de la variable statistique “Situation de famille” ?
Réponse : Le mode est la modalité C “enfant vivant à la maison avec ses deux parents”

Exercice 33
On reprend les données de l'exercice 18 (expérience de Mendel). Quel est le mode de la distribution expérimentale obtenue ?
Réponse : Le mode est la modalité “rond jaune”

Exercice 34
On reprend les données de l'exercice 7, et on s'intéresse à la variable statistique “Classement selon la vitesse de réalisation”. Quelle est la médiane de cette série ? Ce résultat est-il étonnant ? Déterminer la médiane de la variable “Impression de réussite”.
Réponses : La médiane du classement est 5.5, ce qui n'a rien d'étonnant. Celle de l'impression de réussite est 4.5

Exercice 35
On reprend les données de l'exercice 10 (temps de réaction à un stimulus dans trois conditions expérimentales différentes). Déterminer la médiane et la moyenne arithmétique de chacune des trois séries statistiques correspondant aux trois expériences effectuées.
Réponses : Les médianes sont $M_1 = 19,\ M_2 = 39.5\ M_3 = 18$. *Les moyennes sont* $\mu_1 = 19.75,\ \mu_2 = 40.95,\ \mu_3 = 19.35$

Exercice 36
On reprend les données de l'exercice 16 (recensement des ménages). Déterminer le mode, la médiane et la moyenne arithmétique de la série envisagée.
Réponses : Le mode est la modalité 1 ; la médiane, la modalité 2 ; la moyenne arithmétique vaut 1.98

Exercice 37
On reprend les données de l'exercice 22 (test spatial). Déterminer la médiane à partir du tableau des effectifs, puis graphiquement à partir de chacun des histogrammes construits dans l'exercice 24. Calculer la moyenne arithmétique à partir du tableau des effectifs, puis les moyennes arithmétiques relatives aux regroupements en classes définis à l'exercice 24.
Réponses : A partir du tableau des effectifs, on obtient : $M = 25$ *et* $\mu = 25.05$
A partir d'un découpage en 8 classes d'amplitude 6, on obtient : la médiane est dans la classe $[24, 30[$. *Par interpolation linéaire ou évaluation graphique :* $M = 24 + \frac{2577.5-2086}{3805-2085} \times 6 = 25.7$ *et* $\mu = 25.52$
A partir du découpage en 8 classes d'effectifs approximativement égaux indiqué en réponse de l'exercice 24, on obtient : la médiane est dans la classe $[24, 26[$. *Par interpolation linéaire ou lecture graphique :* $M = 24 + \frac{2577.5-2086}{2711-2086} \times 2 = 25.6$ *et* $\mu = 25.21$.
Remarquez que la différence de 0.5 entre la moyenne calculée directement et la moyenne calculée après découpage en classes est due à une erreur de méthode.

Exercice 38
On reprend les données de l'exercice 25 (“Pyramides d'Egypte”). Quelle est la classe modale des deux séries statistiques envisagées. La valeur obtenue permet-elle de prévoir les positions relatives de la médiane et de la moyenne ($M < \mu$ *ou* $M > \mu$). Vérifier par le

calcul.
Réponses : Pour chacune des deux séries, la classe modale est la classe $[0, 5[$. *Le domaine de la variable étudiée est* $[0, 90]$, *mais la moyenne et la médiane seront décalées vers la gauche. Ce décalage sera plus important pour la médiane que pour la moyenne dont le calcul tient compte de modalités "élevées". On obtient ainsi pour la série des hommes :* $M = 20.5$ *et* $\mu = 25.05$.

Exercice 39
On considère une population d'effectif N et, sur cette population, une variable statistique X de modalités $1, 2$ et 3.
1) Lorsque $N = 5$, peut-on contruire un exemple de distribution de la variable X telle que le mode soit égal à 3 et la médiane à 2 ? Et lorsque $N = 7$?
2) Pour $N = 5$, construire un exemple de distribution de la variable X tel que la médiane soit supérieure à la moyenne arithmétique.
Réponses : 1) Impossible avec $N = 5$. *Pour* $N = 7$ *: 1 1 2 2 3 3 3.*
2) Exemple : 1 2 3 3 3. $M = 3$, $\mu = 2.4$

Exercice 40
Sur une population d'effectif $N = 100$, on considère une série statistique de médiane $M = 12$ et de moyenne $\mu = 11$.
En vérifiant les données expérimentales, on s'aperçoit qu'une erreur a été commise et que, pour l'un des individus statistiques, il faut substituer la valeur $x_i = 18$ à la valeur $x_i = 13$. Quelle influence cette substitution a-t-elle sur la médiane ? sur la moyenne ?
Réponses : La médiane est inchangée, la moyenne augmente de 0.05

Exercice 41
On reprend les données de l'exercice 9 (test de développement intellectuel et groupes professionnels). Calculer la moyenne arithmétique des scores pour chacun des groupes professionnels. En déduire la moyenne arithmétique des scores pour l'ensemble des 300 sujets.
Réponses : $\mu_1 = 17.29$, $\mu_2 = 18.76$, $\mu_3 = 22.78$, $\mu = 19.26$

Autres moyennes

La seule notion de moyenne définie en cours est celle de moyenne arithmétique. *Cependant la moyenne arithmétique n'est pas toujours la notion la plus pertinente pour décrire une série de données à l'aide d'un indice de position. On trouvera ci-dessous des exemples d'utilisation de la moyenne* harmonique *et de la moyenne* géométrique.

Exercice 42
Un sportif s'entraîne pour le 100 m. Il court dix fois sur cette distance et obtient les résultats suivants (le temps est exprimé en secondes) :

Essai	1	2	3	4	5	6	7	8	9	10
Temps	12,1	11,8	11,4	10,9	10,5	10,7	10,9	11,2	21,4	21,4

1) Etablir le tableau protocole des vitesses obtenues à chacun des essais (exprimées en mètres par seconde).
2) Calculer le temps moyen obtenu par le sportif sur 100 m et la vitesse moyenne du sportif sur l'ensemble des essais.

Réponses : 1)

Essai	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Vitesse	*8,26*	*8,47*	*8,77*	*9,17*	*9,52*	*9,35*	*9,17*	*8,93*	*4,67*	*4,67*

2) La vitesse moyenne vaut 7.56 alors que la moyenne arithmétique des vitesses vaut 8.10

Exercice 43

Vous faites une étude statistique sur les consommations de carburant d'un ensemble de véhicules automobiles. Vous avez recueilli les données suivantes :

Véhicule	1	2	3
Consommation (l/100km)	5	8	10

1) Quelle est la consommation moyenne des véhicules étudiés ?
2) Un anglo-saxon fait une étude analogue mais, selon l'habitude anglo-saxonne, il mesure la consommation en miles par gallon (que nous avons reconvertis en kilomètres par litre pour simplifier). Il a obtenu le tableau suivant :

Véhicule	1	2	3
Consommation (km/l)	20	12,5	10

Les véhicules étudiés ont-ils un comportement différent des précédents ?
Quelle consommation moyenne obtient-il ?
Convertir cette consommation moyenne en litres aux 100km. Comment peut-on obtenir ce résultat directement à partir du premier tableau ?

Exercice 44

On étudie la croissance d'une population de bactéries en fonction du temps. Avec des unités convenables, on obtient les résultats suivants :

Instant	0	1	2	3	4	5
Nombre	1,0	1,6	2	2,5	3,12	9,8

Quel est le coefficient d'accroissement moyen de cette population par unité de temps ?

Réponses 1)

Période	*1*	*2*	*3*	*4*	*5*
Coeff.	*1,6*	*1,25*	*1,25*	*1,25*	*3,14*

2) Le coefficient moyen d'accroissement par unité de temps est : $\tau = \sqrt[5]{1.6 \times 1.25^3 \times 3.14} = 1.58$.

Caractéristiques de dispersion

Exercice 45

On reprend les données de l'exercice 16 (nombre d'enfants à charge dans une population de ménages). Déterminer les quartiles de la distribution étudiée. Représenter cette distribution à l'aide d'un diagramme de type "boîte à moustaches".
Réponses : $Q_1 = 1,\ M = 2,\ Q_3 = 3$.

Exercice 46

On a mesuré le temps (en secondes) mis par 23 sujets pour effectuer une certaine tâche. Les résultats sont les suivants :

105 107 108 109 110 111 112 113 115 116 117 118
119 121 122 123 124 124 126 127 128 129 130.

Déterminer la médiane et les quartiles, l'étendue et l'écart interquartile. Calculer l'indice de Yule, dont la définition est donnée par $S = \dfrac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$.
Réponses : $Q_1 = 111,\ Q_2 = 118,\ Q_3 = 124$. *L'étendue vaut 25, l'écart interquartile 13. L'indice de Yule vaut -1/13, ce qui montre que la distribution est assez symétrique.*

Exercice 47
On reprend les données "Spatial" (exercice 22). Déterminer les déciles de cette distribution.
Réponse : Les déciles peuvent être déterminés à partir d'un tableau d'effectifs cumulés dans lequel on placera les sujets de rangs 515, 1030, 1546, etc. On obtient ainsi : $D_1 = 16, D_2 = 19, D_3 = 22, D_4 = 23, D_5 = 25, D_6 = 27, D_7 = 29, D_8 = 31, D_9 = 34$

Exercice 48
On reprend les données de l'exercice 9 (test de développement et groupe professionnel). Déterminer la médiane et les quartiles de la distribution globale. Utiliser le découpage en classes ainsi défini pour réaliser des histogrammes des trois distributions correspondant à chacun des groupes professionnels.
Réponse : On obtient : $Q_1 = 15, Q_2 = Md = 19, Q_3 = 24$. *Le découpage en classes obtenu est, par exemple :*

	$[0, 15[$	$[15, 19[$	$[19, 24[$	$[24, 48]$
Man.	*36*	*27*	*30*	*16*
O.Q.	*31*	*26*	*31*	*26*
C.M.	*6*	*11*	*24*	*36*
Ens.	*73*	*64*	*85*	*78*

Exercice 49
On reprend les données de l'exercice 7. Calculer l'écart moyen, la variance et l'écart type de la variable *réussite des tests*.
Réponses : $E_m = 11,44,\ \sigma^2 = 207,94,\ \sigma = 14,42$.

Exercice 50
On reprend les données de l'exercice 13 (nombre de redoublements). Calculer l'écart moyen, puis l'écart type de cette distribution.
Réponses : $\mu = \frac{31}{30},\ E_m = 0.85,\ \sigma = 1.08$

Exercice 51
1) A quelle condition l'écart type d'une distribution est-il nul ?
2) Sur une population d'effectif $N = 100$, on considère une série statistique de moyenne $\mu = 11$.
En vérifiant les données expérimentales, on s'aperçoit que des erreurs ont été commises. Pour deux individus statistiques, il faut substituer respectivement les valeurs $x'_i = 6$ et $x'_j = 16$ aux valeurs $x_i = 9$ et $x_j = 13$.
Quelle influence cette substitution a-t-elle sur la moyenne ? sur l'écart type ? Calculer le nouvel écart type sachant que l'on avait, avec les anciennes valeurs $\sigma = 3,5$.
Réponses : 1) L'écart type n'est nul que si la série est constante.
2) Il n'y a pas de changement pour la moyenne. En revanche, l'écart type est augmenté. Le nouvel écart type vaut $\sigma' = 3.56$

Exercice 52

On reprend les données "Spatial" (exercice 22). Calculer la moyenne m et l'écart type σ de la distribution proposée. Regrouper les données en 7 classes en utilisant comme limites de classes $m - \frac{5\sigma}{2}$, $m - \frac{3\sigma}{2}$, $m - \frac{\sigma}{2}$, $m + \frac{\sigma}{2}$, $m + \frac{3\sigma}{2}$, $m + \frac{5\sigma}{2}$.
Réaliser l'histogramme correspondant.
Réponses : $\mu = 25.04$, $\sigma = 6.79$. *On peut choisir les classes suivantes :*
$[0, 8[$, $[8, 15[$, $[15, 22[$, $[22, 29[$, $[29, 36[$, $[36, 43[$, $[43, 48]$.

Exercice 53
Le tableau suivant donne la distribution des âges dans un groupe de 220 personnes.

classes	[20,24[	[24,28[	[28,32[	[32,36[	[36,40[	[40,44]	Total
effectifs	20	40	60	50	30	20	220

Calculer l'écart moyen, la variance et l'écart type de cette distribution.
Réponses : $\mu = 31.63$; $E_m = 4.69$; $\sigma^2 = 31.5$; $\sigma = 5.61$

Exercice 54
On reprend les données de l'exercice 10 (temps de réaction à un stimulus). Calculer les moyennes μ_1, μ_2, μ_3 et les écarts types σ_1, σ_2, σ_3 des trois distributions correspondant aux trois conditions expérimentales. Calculer les *coefficients de variation* $C_i = \frac{\sigma_i}{\mu_i} \times 100$ des trois distributions.
Réponses :
$\mu_1 = 19.75$; $\sigma_1 = 3.85$; $C_1 = 19.5$
$\mu_2 = 40.95$; $\sigma_2 = 7.34$; $C_2 = 17.9$
$\mu_3 = 19.35$; $\sigma_3 = 4.03$; $C_3 = 20.8$

Exercice 55
On reprend les données de l'exercice 8. Calculer les coefficients de variation des trois séries de données (voir la définition du coefficient de variation dans l'exercice précédent).

Exercice 56
On reprend les données de l'exercice 9 (score obtenu à un test et groupe professionnel) .
1) Calculer l'écart moyen des trois séries correspondant aux trois groupes professionnels. Ces résultats peuvent-ils être utilisés pour calculer l'écart moyen de la distribution des scores sur l'ensemble de la population ?
2) Calculer la variance et l'écart type des trois séries correspondant aux trois groupes professionnels. Utiliser les calculs déjà effectués pour calculer la variance et l'écart type de la distribution des scores sur l'ensemble de la population. Commenter les résultats obtenus.
3) Pour chacune des 4 séries statistiques étudiées, déterminer la proportion d'individus correspondant aux valeurs de l'intervalle $[\mu - \sigma, \mu + \sigma]$, de l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$, de l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$.
Réponses : 2) $\mu_1 = 17.29; \mu_2 = 18.76; \mu_3 = 22.78; \mu = 19.26$
$\sigma_1 = 5.89; \sigma_2 = 6.44; \sigma_3 = 5.80; \sigma = 6.45$
La moyenne μ *peut être obtenue comme moyenne de* μ_1, μ_2, μ_3 *pondérés par les effectifs des trois groupes. Pour la variance, le calcul est plus compliqué, et fait apparaître la moyenne pondérée des variances des trois groupes (variance intra-groupes) et la variance de la série des trois moyennes (variance inter-groupes).*
3) Pour le premier groupe, respectivement 68.8% et 96% des sujets.

Pour le deuxième groupe, respectivement 63% et 96.5% des sujets.
Pour le troisième groupe, respectivement 70% et 94% des sujets.
Pour l'ensemble, respectivement 68.7% et 97% des sujets.

Exercice 57

On reprend les données de l'exercice 14 (nombre d'enfants à charge dans une population de ménages). On rappelle que le tableau des effectifs de la série statistique étudiée est donné par :

Nb enfants	0	1	2	3	4	5	6	7	8	Total
Effectifs	48	65	44	27	19	15	8	2	1	229

1) Quel est le nombre total d'enfants pris en compte par l'étude ?
2) Calculer, pour chaque modalité, le nombre d'enfants pris en compte (*masse de la modalité*) ; procéder ensuite à un cumul des valeurs obtenues (*masses cumulées*).
3) Déterminer les valeurs permettant de compléter les phrases suivantes :
— La moitié des enfants se trouvent dans des foyers comportant ... enfants ou moins (*médiale* de la série étudiée).
— Les 20% de familles nombreuses comportant 4 enfants ou plus rassemblent ...% des enfants.
— 20% des enfants se trouvent dans des foyers comportant ...enfants ou plus.
4) Former un tableau des effectifs pour la variable *nombre de frères et sœurs* étudiée sur l'ensemble des 455 enfants concernés par l'étude.
Déterminer la moyenne, la variance et l'écart type de la distribution ainsi définie.

Réponses : 1) Masse totale de la variable : 455.

2)

a_i	0	1	2	3	4	5	6	7	8	Total
n_i	48	65	44	27	19	15	8	2	1	229
Masse M_i	0	65	88	81	76	75	48	14	8	455
Masse cumulée	0	65	153	234	310	385	433	447	455	
M. cum. relative	0%	14%	34%	51%	68%	85%	95%	98%	100%	

3) La moitié des enfants se trouve dans des familles de 3 enfants ou moins. La médiale vaut 3.
Il y a 45 familles de 4 enfants ou plus ; elles représentent 20% de la population étudiée et rassemblent 49% des enfants.
Les 20% d'enfants membres des familles les plus nombreuses se trouvent dans des familles de 5 enfants ou plus.

4)

Nb frères/sœurs	0	1	2	3	4	5	6	7
Effectif	65	88	81	76	75	48	14	8

Loi binomiale

Exercice 58

1) Calculer : 3! , 4! , 7!.

2) Calculer : $\frac{10!}{7!}, \frac{80!}{77!}, \frac{(n+1)!}{(n-1)!}$.

3) Calculer : $C_3^2, C_5^2, C_5^3, C_{10}^3, C_{10}^7$.

4) Montrer que : $\sum_{p=0}^{p=5} C_5^p = 2^5$.

Réponses : 1) 6 ; 24 ; 5040. 2) 720 ; 492960 ; $n(n+1)$. 3) $C_3^2 = 3$, $C_5^2 = C_5^3 = 10$ (ce n'est pas un hasard), $C_{10}^3 = C_{10}^7 = 120$.

Exercice 59

On considère la loi binomiale de paramètres $n = 4$ et $p = \frac{1}{3}$.
Déterminer la distribution de cette loi. La représenter à l'aide d'un diagramme en bâtons.
Calculer sa moyenne, sa variance et son écart type.

Réponses :

k	0	1	2	3	4
f_k	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$

; $\mu = \frac{4}{3}$; $\sigma^2 = \frac{8}{9}$

Exercice 60

On considère la loi binomiale de paramètres $n = 5$ et $p = 0,4$.
Déterminer la distribution et la fonction de répartition de cette loi. Représenter graphiquement la fonction de répartition.

Réponses.

k	0	1	2	3	4	5
f_k	0.08	0.26	0.35	0.23	0.07	0.01
$F(X < k)$	0	0.08	0.34	0.69	0.92	0.99

Exercice 61

On dispose d'un ensemble de fiches représentant les individus d'une population (supposée très nombreuse) comportant 50% de filles.
1) On tire au hasard une fiche de l'ensemble. Quelle est la probabilité que la fiche tirée soit celle d'une fille ? (Autrement dit, si cette expérience est renouvelée un grand nombre de fois, avec quelle fréquence peut-on s'attendre à voir se réaliser l'événement "la fiche tirée est celle d'une fille" ?)
2) L'expérience consiste maintenant à tirer au hasard 10 fiches dans l'ensemble. Les fiches sont tirées l'une après l'autre et on remet à chaque fois la fiche tirée.
Quelle est la probabilité d'obtenir 4 fiches de sujets féminins sur les 10 fiches tirées ?
Réponses : 1) $p_1 = 0.5$; 2) $p_2 = 0.204$

Exercice 62

On a fait passer à une population nombreuse un QCM comportant 6 questions. Pour chaque question, il y avait trois réponses possibles, dont une seule réponse correcte. On a observé les fréquences de bonnes réponses suivantes :

Réponses correctes	0	1	2	3	4	5	6
Fréquences	1%	6%	15%	29%	30%	16%	3%

1) On désire approcher la distribution expérimentale observée par une distribution binomiale. Quels paramètres proposez-vous pour cette distribution ? Représenter sur un même dessin la distribution expérimentale et la distribution théorique.

2) Quelle distribution devrait-on s'attendre à observer si les sujets répondent complètement au hasard ?

Réponses : 1) On choisit $n = 6$ *(nombre de questions) et* $p = 0.57$ *(variable binomiale de même moyenne que la variable observée).*

2) Si les sujets répondent au hasard on devrait observer une distribution binomiale de paramètres $n = 6$ *et* $p = 0.33$.

Exercice 63

Dans une population d'électeurs (supposée très nombreuse), la proportion d'intentions de votes en faveur du candidat X est de 52%.

1) On interroge des personnes prises isolément. Quelle fréquence d'intentions de votes en faveur du candidat X doit-on s'attendre à observer ?

2) On interroge maintenant des groupes de 5 personnes et on prend comme variable le nombre d'intentions de votes en faveur du candidat X obtenu dans le groupe interrogé.
Quelles sont les modalités de la variable étudiée ?
Quelles sont les fréquences d'apparition théoriques de ces modalités ?

Réponses : 1) 52% de "succès", 48% d'"échecs".

2)

X	*0*	*1*	*2*	*3*	*4*	*5*
f_k	*2.5%*	*14%*	*30%*	*32%*	*18%*	*4%*

Exercice 64

On estime que, parmi les enfants issus d'un certain milieu social, 20% ont des difficultés scolaires. On désire organiser un cours de rattrapage pour les enfants en difficulté parmi un groupe de 20 enfants, dont les performances scolaires n'ont pas encore été testées.
Quelle capacité d'accueil doit-on prévoir pour ce cours de rattrapage si l'on veut être en mesure d'accueillir les enfants en situation d'échec, dans 95% des cas ?

Réponses : Calcul des fréquences cumulées pour une loi binomiale de paramètres $n = 20$ *et* $p = 0.2$.

$F(X < 7) = 0.91$; $F(X < 8) = 0.968$. *Il faut prévoir au moins 7 places.*

Exercice 65

Dans une recherche de psychologie, on a présenté à chaque sujet des photographies de chiens, de diverses races, pour moitié à poils courts et pour moitié à poils longs. Chaque sujet évalue le chien présenté en cochant l'une des 6 cases d'une échelle d'attirance. Les évaluations ont été codées numériquement de 0 (faible attirance) à 5 (forte attirance). Dans le tableau ci-dessous, on donne pour chacun des 16 sujets les deux valeurs moyennes des évaluations concernant les deux types de chiens.

	s1	s2	s3	s4	s5	s6	s7	s8
Poils longs	3,50	2,75	3,50	2,75	1,50	3,00	4,00	3,00
Poils courts	2,50	1,25	2,75	0,50	2,00	3,00	2,00	2,00

	s9	s10	s11	s12	s13	s14	s15	s16
Poils longs	2,75	3,25	4,00	3,00	3,75	4,00	4,00	3,25
Poils courts	2,50	0,75	0,50	1,00	1,50	1,75	2,00	2,00

1) Un *effet individuel* est défini comme la différence entre les deux notes concernant un même sujet. Construire le protocole des effets individuels.

2) On fait l'hypothèse suivante : les sujets ont la même attirance pour les deux types de races de chiens. Dans ces conditions, les effets individuels positifs et négatifs ont la même fréquence d'apparition. Quelle est alors la probabilité d'observer 0 ou 1 effet individuel

négatif sur un groupe de 15 sujets ?
Réponses : 2) Pour une loi binomiale de paramètres $n = 15$ *et* $p = 0.5$*, la fréquence cumulée* $F(X < 2)$ *est égale à 0.5%.*

Exercice 66
Un groupe de 10 sujets entreprend sur une période de six mois un programme d'enrichissement cognitif destiné à améliorer leurs processus de traitement de l'information. Pour évaluer l'effet du programme, on fait passer aux sujets deux tests de niveaux comparables l'un avant, l'autre après la période d'apprentissage.
1) Soit X une variable distribuée selon une loi binomiale de paramètres $n = 10$ et $p = 0.5$. Quelles sont les fréquences des modalités $X = 8$, $X = 9$ et $X = 10$? de l'ensemble de modalités $X \geq 8$?
2) On estime que, si le programme n'a pas d'effet, un sujet a autant de chances de progresser que de régresser entre les deux tests. On constate que, sur les 10 sujets, 8 ont amélioré leur performance. Si le hasard seul est invoqué pour expliquer ce résultat, avec quelle fréquence peut-on observer une progression pour au moins 8 sujets sur 10 ?
Réponses : 1) $b(10, 0.5, 8) = 4.39\%$, $b(10, 0.5, 9) = 0.97\%$, $b(10, 0.5, 8) = 0.097\%$, $F(X \geq 8) = 5.5\%$. *2) 5.5% .*

Exercice 67
Dans le cadre d'une étude sur les familles nombreuses, on veut constituer un échantillon de 100 familles représentatif des familles de 5 enfants.
1) Combien de familles comportant 1 garçon et 4 filles doit-on retenir pour constituer l'échantillon ?
2) Même question pour les familles comportant 0, 2, 3, 4, 5 garçons.
N.B. On fera l'hypothèse que la probabilité d'avoir un garçon ou une fille est égale à 0.5
3) Représenter la distribution obtenue à l'aide d'un diagramme en bâtons.
Réponses : 1) et 2) :

X	*0*	*1*	*2*	*3*	*4*	*5*
P	*3*	*16*	*31*	*31*	*16*	*3*

Loi normale

Exercice 68
Soit Z une variable distribuée selon une loi normale centrée réduite. Etant donné deux nombres réels a et b, on note $P(Z < a)$, $P(Z \geq b)$, $P(a \leq Z < b)$ les fréquences correspondant aux ensembles de modalités $]-\infty, a[$, $[b, +\infty[$, $[a, b[$.
a) Calculer les fréquences suivantes :

$$\begin{array}{llll} P(Z < 1) & P(Z < 1,56) & P(Z < -2) & P(Z < -2,33) \\ P(Z \geq 1) & P(Z \geq 2,575) & P(Z \geq -1,5) & P(Z \geq -0,5) \\ P(-0,5 \leq Z < 0,5) & P(-1 \leq Z < 1) & P(-2 \leq Z < 2) & P(-3 \leq Z < 3) \\ P(-1,5 \leq Z < 1) & P(0,5 \leq Z < 2,4) & P(-2,34 \leq Z < -1,75). & \end{array}$$

b) Dans chacun des cas suivants, déterminer le nombre réel a qui convient :

$$\begin{array}{lll} P(Z < a) = 0,5 & P(Z < a) = 0,90 & P(Z < a) = 0,95 \\ P(Z < a) = 0,98 & P(Z < a) = 0,99 & P(Z < a) = 0,999 \\ P(-a \leq Z < a) = 0,5 & P(-a \leq Z < a) = 0,90 & P(-a \leq Z < a) = 0,95 \\ P(-a \leq Z < a) = 0,98 & P(-a \leq Z < a) = 0,99 & P(-a \leq Z < a) = 0,999 \end{array}$$

c) Dans chacun des cas suivants, déterminer le nombre réel b qui convient :

$$P(Z \geq b) = 0,05 \quad P(Z \geq b) = 0,025 \quad P(Z \geq b) = 0,01 \quad P(Z \geq b) = 0,005$$
$$P(Z \geq b) + P(Z < -b) = 0,05 \quad P(Z \geq b) + P(Z < -b) = 0,01$$

d) Déterminer des couples (a, b) tels que $P(-a \leq Z < b) = 0,90$.

Réponses : a) $P(Z < 1) = 0.8413$ *;* $P(Z < 1,56) = 0.9406$ *;* $P(Z < -2) = 0.0228$ *;* $P(Z < -2,33) = 0.0099$ *;* $P(Z \geq 1) = 0.1587$ *;* $P(Z \geq 2,575) = 0.005$ *;* $P(Z \geq -1,5) = 0.9332$ *;* $P(Z \geq -0,5) = 0.6915$ *;* $P(-0,5 \leq Z < 0,5) = 0.3830$ *;* $P(-1 \leq Z < 1) = 0.6826$ *;* $P(-2 \leq Z < 2) = 0.9544$ *;* $P(-3 \leq Z < 3) = 0.9974$ *;* $P(-1,5 \leq Z < 1) = 0.7745$ *;* $P(0,5 \leq Z < 2,4) = 0.3003$ *;* $P(-2,34 \leq Z < -1,75)0.0305$.

b) $P(Z < a) = 0,5$ *pour* $a = 0$ *;* $P(Z < a) = 0,90$ *pour* $a = 1.28$ *;* $P(Z < a) = 0,95$ *pour* $a = 1.645$ *;* $P(Z < a) = 0,98$ *pour* $a = 2.06$ *;* $P(Z < a) = 0,99$ *pour* $a = 2.33$ *;* $P(Z < a) = 0,999$ *pour* $a = 3.08$ *;* $P(-a \leq Z < a) = 0,5$ *pour* $a = 0.675$ *;* $P(-a \leq Z < a) = 0,90$ *pour* $a = 1.645$ *;* $P(-a \leq Z < a) = 0,95$ *pour* $a = 1.96$ *;* $P(-a \leq Z < a) = 0,98$ *pour* $a = 2.33$ *;* $P(-a \leq Z < a) = 0,99$ *pour* $a = 2.575$ *;* $P(-a \leq Z < a) = 0,999$ *pour* $a = 3.27$.

c) $P(Z \geq b) = 0,05$ *pour* $b = 1.645$ *;* $P(Z \geq b) = 0,025$ *pour* $b = 1.96$ *;* $P(Z \geq b) = 0,01$ *pour* $b = 2.33$ *;* $P(Z \geq b) = 0,005$ *pour* $b = 2.575$ *;* $P(Z \geq b) + P(Z < -b) = 0,05$ *pour* $b = 1.96$ *;* $P(Z \geq b) + P(Z < -b) = 0,01$ *pour* $b = 2.575$.

d) Par exemple, $P(a < Z \leq b) = 0.90$ *pour* $a = 1.29$ *et* $b = 2.97$.

Exercice 69

Soit X une variable distribuée selon une loi normale de paramètres $\mu = 2$ et $\sigma = 5$. Soit Z la variable normale centrée réduite associée.

a) Quelles sont les modalités de Z correspondant aux modalités suivantes de X :

$$X = 2 \quad X = 4,5 \quad X = 0 \quad X = -8$$

b) Quelles sont les modalités de X correspondant aux modalités suivantes de Z :

$$\begin{array}{lll} Z = 0 & Z = 1 & Z = 1,645 \\ Z = 1,96 & Z = 2,33 & Z = -1 \\ Z = -1,645 & Z = -1,96 & Z = -2,33 \end{array}$$

c) Quels sont les intervalles de modalités de Z correspondant aux intervalles suivants de modalités de X :

$$\begin{array}{lll} X < 7,5 & X < -4 & X \geq 7 \\ X \geq -8 & 0 \leq X < 4 & -3 \leq X < 7 \\ 1,2 \leq X < 3,7 & \mu - \sigma \leq X < \mu + \sigma & \mu - 2\sigma \leq X < \mu + 2\sigma \end{array}$$

d) Quels sont les intervalles de modalités de X correspondant aux intervalles suivants de modalités de Z :

$$\begin{array}{lll} Z < 1,5 & Z < -2,33 & Z \geq 1,645 \\ Z \geq -1 & -1 \leq Z < 1 & 1,645 \leq Z < 1,96 \\ -3 \leq Z < -1,75 & & \end{array}$$

Réponses : a) $X = 2$ *correspond à* $Z = 0$ *;* $X = 4,5$ *correspond à* $Z = 0.5$ *;* $X = 0$ *correspond à* $Z = -0.4$ *;* $X = -8$ *correspond à* $Z = -2$.

b) $Z = 0$ *correspond à* $X = 2$ *;* $Z = 1$ *correspond à* $X = 7$ *;* $Z = 1,645$ *correspond à* $X = 10.225$ *;* $Z = 1,96$ *correspond à* $X = 11.8$ *;* $Z = 2,33$ *correspond à* $X = 13.65$ *;* $Z = -1$ *correspond à* $X = -3$ *;* $Z = -1,645$ *correspond à* $X = -6.225$ *;* $Z = -1,96$ *correspond à* $X = -7.8$ *;* $Z = -2,33$ *correspond à* $X = -9.65$.

c) $X < 7,5$ *correspond à* $Z < 0.7$ *;* $X < -4$ *correspond à* $Z < -1.2$ *;* $X \geq 7$ *correspond à* $Z \geq 1$ *;* $X \geq -8$ *correspond à* $Z \geq -2$ *;* $0 \leq X < 4$ *correspond à* $-0.4 \leq Z < 0.4$ *;*

$-3 \le X < 7$ *correspond à* $-1 \le Z < 1$ *;* $1,2 \le X < 3,7$ *correspond à* $-0.16 \le Z < 0.34$ *;* $\mu - \sigma \le X < \mu + \sigma$ *correspond à* $-1 \le Z < 1$ *;* $\mu - 2\sigma \le X < \mu + 2\sigma$ *correspond à* $-2 \le Z < 2$.
d) $Z < 1,5$ *correspond à* $X < 9.5$ *;* $Z < -2,33$ *correspond à* $X < 13.65$ *;* $Z \ge 1,645$ *correspond à* $X \ge 10.225$ *;* $Z \ge -1$ *correspond à* $X \ge -3$ *;* $-1 \le Z < 1$ *correspond à* $-3 \le X < 7$ *;* $1,645 \le Z < 1,96$ *correspond à* $10.225 \le X < 11.8$ *;* $-3 \le Z < -1,75$ *correspond à* $-13 \le X < -6.75$.

Exercice 70
Le tableau suivant donne les valeurs de la densité $f(z)$ de la loi normale centrée réduite pour quelques valeurs de z.

z	$f(z)$
0	0,3989
0,25	0,3867
0,5	0,3521
0,75	0,3011
1	0,2420
1,25	0,1826
1,5	0,1295
1,75	0,0863
2	0,0540
2,25	0,0317
2,5	0,0175
2,75	0,0091
3	0,0044

Utiliser ce tableau pour construire, sur un même dessin, la représentation graphique de la densité de la loi normale de paramètres $\mu = 10$ et $\sigma = 2$ et celle de la loi normale de paramètres $\mu = 10$ et $\sigma = 5$. Quelle remarque peut-on faire ?

Exercice 71
Soit X une variable distribuée selon une loi normale de paramètres μ et σ.
a) On suppose que : $\mu = 10,5$ et $\sigma = 3,2$. Calculer les fréquences suivantes :

$$P(X < 13) \qquad P(X < 7)$$
$$P(X \ge 17) \qquad P(X \ge 3)$$
$$P(7,3 \le X < 13,7) \qquad P(13,7 \le X < 16,9) \qquad P(9 \le X < 12)$$

b) On suppose que : $\mu = 25$ et $\sigma = 6,8$. Dans chacun des cas suivants, déterminer le nombre réel a qui convient :

$$P(X < a) = 0,5 \quad P(X < a) = 0,95 \quad P(X < a) = 0,99$$

c) On suppose que : $\mu = 25$ et $\sigma = 6,8$. Déterminer les réels a et b de façon que le centre de la classe $[a, b[$ soit la moyenne μ et que $P(a \le X < b) = 0.95$.
d) On suppose que : $\mu = 2,42$. Déterminer σ sachant que : $P(1,42 \le X < 3,42) = 0,95$.
Réponses : a) $P(X < 13) = 0.7823$ *;* $P(X < 7) = 0.1379$ *;* $P(X \ge 17) = 0.0212$ *;* $P(X \ge 3) = 0.9904$ *;* $P(7,3 \le X < 13,7) = 0.6826$ *;* $P(13,7 \le X < 16,9) = 0.1359$ *;* $P(9 \le X < 12) = 0.3616$.
b) $P(X < a) = 0,5$ *pour* $a = 25$ *;* $P(X < a) = 0,95$ *pour* $a = 36.2$ *;* $P(X < a) = 0,99$ *pour* $a = 40.8$.
c) $P(a \le X < b) = 0.95$ *pour* $a = 11.7$ *et* $b = 38.3$.
d) $\sigma = 0.51$.

Exercice 72

Les notes obtenues au test de raisonnement Alpha 61 par 300 élèves d'une classe de 6è s'étendent de 0 à 44. Sachant qu'elles constituent une distribution normale de moyenne $m_1 = 21$ et d'écart type $s_1 = 7,11$, on demande :
a) de transformer ces notes pour les exprimer sur une échelle de type QI de moyenne $m_2 = 100$ et d'écart type $s_2 = 15$;
b) de construire, sur cette nouvelle échelle, un étalonnage en 5 classes normalisées (indiquer la valeur numérique des quatre limites nécessaires).
N.B. Bornes des classes à utiliser pour un découpage en classes normalisées d'une variable centrée réduite :

— 5 classes : -3/2 ; -1/2 ; 1/2 ; 3/2
— 7 classes : -5/3 ; -1 ; -1/3 ; 1/3 ; 1 ; 5/3
— 9 classes : -7/4 ; -5/4 ; -3/4 ; -1/4 ; 1/4 ; 3/4 ; 5/4 ; 7/4.

Réponses : a) En désignant par X l'ancienne note et par Y la nouvelle note, la transformation demandée s'écrit : $Y = 2.11X + 55.70$.
b) Les bornes des classes seront : $-\infty, 77.5, 92.5, 107.5, 122.5, +\infty$

Exercice 73

Sur une distribution normale de moyenne $m = 80$ et d'écart type $\sigma = 15$, construire :
a) un étalonnage en 7 classes normalisées ;
b) un étalonnage comportant cinq classes d'effectif égal.
Réponses : a) Bornes : $-\infty, 55, 65, 75, 85, 95, 105, +\infty$
b) Bornes : $-\infty, 67.4, 76.3, 83.8, 92.6, +\infty$.

Exercice 74

Dans un examen, deux correcteurs A et B doivent évaluer chacun 150 copies qui peuvent être considérées comme leur étant attribuées au hasard.
Après correction, on remarque que le correcteur A obtient pour l'ensemble de ses copies une distribution normale $m_A = 11.25$, $\sigma_A = 4.90$ tandis que pour le correcteur B on trouve $m_B = 9.97$, $\sigma_B = 3.37$.
a) A quelles notes chez B correspondent les notes suivantes obtenues chez A :

13, 8, 7, 12, 5, 16 ?

b) A quelles notes chez A correspondent les notes suivantes obtenues chez B :

12, 17, 7, 10, 8, 11 ?

c) Chez quel correcteur l'obtention (ou la perte) d'un point a-t-elle le plus d'importance ?
Réponses : Soit X et Y les notes mises respectivement par les correcteurs A et B. Relations entre X et Y : $X = 1.45Y - 3.25$; $Y = 0.69X + 2.23$.
a) $11.2, 7.7, 7.1, 10.5, 5.7, 13.3$.
b) $14.2, 21.4, 6.9, 11.3, 8.4, 12.7$.
c) Une différence d'un point chez A s'exprime par une différence de moins d'un point chez B. L'obtention d'un point chez B a donc plus d'importance.

Exercice 75

On dispose d'une population scolaire de 7257 enfants du niveau cours moyen. Cette population a été testée avec une épreuve et les résultats obtenus sont :

$$m = 33,72 \ ; \ \sigma = 5,27.$$

On suppose que les scores obtenus à l'épreuve sont distribués suivant une loi normale.
1) Sur quel effectif peut statistiquement compter un chercheur dans chacun des cas suivants :

a) enfants ayant une note comprise entre 25 et 35

b) enfants ayant une note supérieure à 38

c) enfants ayant une note comprise entre 20 et 33.

2) a) Quel est l'intervalle de scores correspondant, d'un point de vue statistique, aux 30% de la population ayant obtenu les meilleurs résultats au test ?
b) Même question pour les 50% de la population dont les résultats sont les plus proches de la moyenne.
Réponses : 1) a) 3955 b) 1517 c) 120
2) a) Scores supérieurs à 36.5. b) Scores compris entre 30.2 et 37.3

Exercice 76
La distribution de la taille des 3000 étudiants d'une université américaine est ajustée par une distribution normale de moyenne $\mu = 175\ cm$ et d'écart type $\sigma = 8\ cm$.
1) Evaluer le nombre d'étudiants sont la taille dépasse 183 cm.
2) Evaluer le nombre d'étudiants dont la taille est inférieure à 165 cm.
3) Evaluer le nombre d'étudiants dont la taille est comprise entre 1m60 et 1m90.
Réponses : 1) 476 ; 2) 317 ; 3) 2818

Exercice 77
On reprend les données "Spatial" (exercice 22) et les résultats obtenus dans l'exercice 52 (moyenne $m = 25,04$, écart type $\sigma = 6,79$).
1) Construire un étalonnage en 7 classes normalisées pour une loi normale de paramètres m et σ.
2) Faire un découpage en classes de la variable expérimentale en utilisant les bornes déterminées à la question 1. Tenir compte d'une correction de continuité en assimilant la note n à la classe $[n - 0.5, n + 0.5[$.
3) Dresser un tableau des fréquences théoriques (pour la loi normale de paramètres m et σ) et des fréquences observées (pour la distribution expérimentale) des 7 classes précédentes.
4) Calculer les densités relatives de chacune des classes pour la distribution expérimentale (fréquence par unité d'amplitude). Construire sur un même dessin l'histogramme de la distribution expérimentale et la densité de la loi normale théorique.
N.B. Utiliser une graduation de l'axe des ordonnées selon les densités relatives de manière à pouvoir comparer les deux représentations obtenues.

Exercice 78
Soit X une variable qui suit une loi binomiale de paramètres $n = 100$ et $p = 0.4$.
A l'aide d'une approximation par une loi normale convenablement choisie, calculer :

— Les fréquences des modalités 25, 30, 40.

— Les fréquences des classes $[10, 30[$, $[30, 50[$, $[50, 70[$.

Réponses : Soit $\tilde{X}$ la variable normale de paramètres $\mu = 40$ et $\sigma = 4.9$ approchant X.
1) $P(X = 25) = P(24.5 \leq \tilde{X} < 25.5) = 0.0007$ *;* $P(X = 30) = P(29.5 \leq \tilde{X} < 30.5) = 0.01$ *;* $P(X = 40) = P(39.5 \leq \tilde{X} < 40.5) = 0.0796$
2) $P(10 \leq X < 30) = P(\tilde{X} < 29.5) = 0.0162$ *;* $P(30 \leq X < 50) = P(29.5 \leq \tilde{X} < 49.5) = 0.9576$ *;* $P(50 \leq X < 70) = P(\tilde{X} \geq 49.5) = 0.0262$*.*

Exercice 79
Au cours d'un test d'alcoolémie effectué auprès d'automobilistes, on a relevé qu'une personne sur 10 avait un test positif.
1) On contrôle 400 personnes par semaine. Quelle est la loi donnant la distribution de la variable X : "nombre de personnes ayant un test positif" ?
Calculer la moyenne et la variance de cette variable.
2) Par quelle loi peut-on approcher la loi de distribution de X ? En déduire les fréquences correspondant à :

$X < 35$

$32 < X < 44$

Réponses : 1) X suit une loi binomiale avec $n = 400$ et $p = 0.1$. $\mu = 40$; $\sigma^2 = 36$.
2) $P(X < 35) = 0.1788$ $P(32 < X < 44) = 0.6134$

Loi et test du χ^2

Exercice 80
Un psychologue fait l'hypothèse que certaines difficultés du langage écrit chez l'enfant sont en relation avec des facteurs dits *instrumentaux*, notamment la latéralisation. Sur un échantillon, ce praticien recueille les données suivantes :

latéralisation	difficultés du langage écrit	
	oui	non
droitier	12	25
ambidextre	14	7
gaucher	17	13

Examiner ces résultats et formuler avec précision une hypothèse statistique. L'éprouver à l'aide du test statistique approprié et indiquer si, au seuil $\alpha = .05$, on peut conclure que la latéralisation a une incidence sur les difficultés du langage écrit.
Réponse : $\chi^2_{obs} = 7.34$ *; Pour ddl = 2 et* $\alpha = 0.05$*,* $\chi^2_c = 5.99$*. Il existe donc une relation entre les deux variables.*

Exercice 81
On veut étudier s'il y a une relation significative entre la couleur des yeux et celle des cheveux dans un groupe déterminé d'individus. Pour ce faire, on fait l'étude dont les résultats sont résumés dans le tableau suivant :

	cheveux				
Yeux	Blonds	Châtains	Noirs	Roux	Total
Bleus	1768	807	189	47	2811
Gris	946	1387	746	53	3132
Bruns	115	438	288	16	857
Total	2829	2632	1223	116	6800

Justifiez le tableau d'effectifs théoriques suivant…

	cheveux				
Yeux	Blonds	Châtains	Noirs	Roux	Total
Bleus	1169	1088	506	48	2811
Gris	1303	1212	563	54	3132
Bruns	357	332	154	14	857
Total	2829	2632	1223	116	6800

…et l'affirmation de l'auteur : *on calcule aisément le* χ^2 *de la répartition qui vaut ici* $\chi^2 = 1073$. Calculer le nombre de degrés de liberté et conclure.
Réponse : ddl = 6, et le χ^2 *trouvé est évidemment significatif d'une liaison entre les variables…*

Exercice 82

Extrait de "Registres mis en jeu par la notion de fonction". I. Guzman-Retamal - Annales de Didactique et de Sciences Cognitives. N.2 (1989) - IREM Strasbourg

Dans une classe expérimentale de 20 élèves et une classe témoin de 24 élèves, on a utilisé deux pédagogies différentes pour enseigner un chapitre du cours de mathématiques. On fait passer aux deux classes un test de connaissances commun, composé d'un certain nombre d'items, et on compare la réussite des deux classes item par item. Commentez et exploitez les trois tableaux suivants :

Item	4.1	6.2	3.2
Expérimentale	.90	.65	0
Témoin	.92	.62	.04

Item	4.2	5.2	10.2
Expérimentale	.65	.85	.50
Témoin	.25	.33	.12

Item	1.2	1.3	8.1
Expérimentale	.25	.25	.20
Témoin	.71	.67	.75

N.B. La correction de Yates pourra être utilisée lors d'éventuels calculs de χ^2.
Réponses : Pour les items 4.1, 6.2 et 3.2, il n'y a pas de différence significative entre les deux groupes (mais seul l'item 6.2 permet de calculer un χ^2*). Pour les autres items, le test du* χ^2 *montre une différence significative entre les deux groupes.*

Exercice 83

Pour 650 enfants en consultation psychiatrique, on a observé d'une part le rang dans la fratrie : *rang 1, rang 2, rang 3* et d'autre part le diagnostic porté sur l'enfant : réaction de type soit dépressif (d), soit anxieux (a), soit schizophrénique (s). Les observations sont résumées dans le tableau suivant :

	Diagnostic			
Rang	d	a	s	Total
rang 1	98	70	57	225
rang 2	108	75	61	244
rang 3	61	68	52	181
Total	267	213	170	650

1) Quelles sont les variables statistiques étudiées. Quelle est la nature de chacune d'elles ? Quelle est la nature (tableau protocole, tableau d'effectifs, tableau de contingence) du tableau ci-dessus ?
2) On se propose d'étudier à l'aide d'un test du χ^2 s'il existe un lien entre le type de maladie et le rang dans la fratrie.
a) Calculer le tableau des effectifs théoriques correspondant au tableau ci-dessus.
b) La valeur χ^2_{obs} observée sur l'échantillon étudié est obtenue comme somme des contributions des différentes cases du tableau.
Détailler le calcul de la contribution de la première case.
c) Le calcul complet donne : $\chi^2_{obs} = 5.75$.
Déterminer le nombre de degrés de liberté. Utiliser la table du χ^2 pour déterminer au seuil de 5% si la liaison est significative.
Réponses : 1) Variables : rang (ordinale) et diagnostic (nominale). Tableau de contingence.

2) a)

	d	*a*	*s*
r1	*92.4*	*73.7*	*58.8*
r2	*100.2*	*80*	*63.8*
r3	*74.3*	*59.3*	*47.3*

2) b) Contribution première case : $\frac{(98-92.4)^2}{92.4} = 0.34$.
2) c) ddl $= (l-1)(c-1) = 4$. *Au seuil de 5%,* $\chi^2_c = 9.49$. *L'hypothèse d'indépendance des deux variables ne peut pas être rejetée.*

Exercice 84
Lors d'une enquête, on a interrogé 150 personnes prises au hasard sur leurs connaissances en langues étrangères. Les résultats obtenus sont les suivants :

	Hommes	Femmes
Anglais	37	24
Allemand	9	19
Espagnol	16	15
Aucune	20	10

Les connaissances en langues étrangères dépendent-elles du sexe dans la population dont est issu l'échantillon étudié ? On répondra à cette question en effectuant un test au seuil de 5%.
Réponses : Le tableau proposé est un tableau de contingence. On va donc procéder à un test du χ^2. *Effectifs théoriques :*

	Hommes	*Femmes*
Anglais	*33.35*	*27.65*
Allemand	*15.31*	*12.69*
Espagnol	*16.95*	*14.05*
Aucune	*16.40*	*13.60*

$\chi^2_{obs} = \frac{(37-33.35)^2}{33.35} + \ldots + \frac{(10-13.60)^2}{13.60} = 8.48$

Pour $\alpha = 5\%$ et ddl $= (2-1)(4-1) = 3$, on a : $\chi^2_c = 7.815$. La différence entre les sexes est donc significative.

Exercice 85

Dans le cadre d'une enquête sur le SIDA réalisée en Allemagne durant l'été 1990 (A. Hahn, W.H. Eirmbter et R. Jacob), on a interrogé 2089 personnes. Le questionnaire comportait notamment l'item suivant : *Le sida est un péril omniprésent. Indiquez si vous êtes : d'accord, indécis, pas d'accord.* Le croisement de la réponse du sujet avec son âge donne le tableau de contingence suivant :

Classe d'âge	d'accord	indécis	pas d'accord	Total
18 à < 30	43	116	365	524
30 à < 40	36	116	273	425
40 à < 50	32	95	217	344
50 à < 60	38	114	167	319
60 et plus	67	160	250	477
Total	216	601	1272	2089

1) Réaliser un test du χ^2 permettant de répondre à la question suivante : "Les réponses des sujets dépendent-elles de leur âge ? "

2) Comparer de même à l'aide d'un test du χ^2 les réponses des deux dernières classes d'âge, puis les réponses des moins de 30 ans à celles des 60 ans et plus.

Réponses : 1) Le tableau des effectifs théoriques est donné par :

Classe d'âge	*d'accord*	*indécis*	*pas d'accord*
18 à < 30	*54.2*	*150.8*	*319.0*
30 à < 40	*43.9*	*122.3*	*258.8*
40 à < 50	*35.6*	*99.0*	*209.4*
50 à < 60	*33.0*	*91.8*	*194.2*
60 et plus	*49.3*	*137.2*	*290.5*

Celui des contributions au χ^2 est donné par :

Classe d'âge	*d'accord*	*indécis*	*pas d'accord*
18 à < 30	*2.31*	*8.01*	*6.61*
30 à < 40	*1.44*	*0.32*	*0.78*
40 à < 50	*0.36*	*0.16*	*0.27*
50 à < 60	*0.76*	*5.58*	*3.82*
60 et plus	*6.33*	*3.78*	*5.63*

On obtient $\chi^2_{obs} = 46$. Or pour un seuil de 1% et ddl $= (5-1)(3-1) = 8$, on a : $\chi^2_c = 20.09$. La réponse du sujet est donc dépendante de son âge.

2) En ne considérant que les deux dernières classes d'âge, on obtient le tableau d'effectifs théoriques suivant :

Classe d'âge	*d'accord*	*indécis*	*pas d'accord*
50 à < 60	*42.1*	*109.8*	*167.1*
60 et plus	*62.9*	*164.2*	*249.9*

On obtient alors $\chi^2_{obs} = 0.92$. Or, pour $\alpha = 5\%$ et ddl $= 2$, $\chi^2_c = 5.99$. Les réponses des sujets sont cette fois indépendantes de leur appartenance à l'une ou l'autre classe d'âge. Dans le dernier cas, $\chi^2_{obs} = 31.62$, ce qui est significatif d'une dépendance.

Exercice 86 *Reprise de l'exercice 18*
Dans l'expérience complète de Mendel, deux attributs des pois étaient considérés : la couleur (*jaune* ou *verte*, et la forme (*ronde* ou *ridée*), avec la triple hypothèse suivante : *rond* dominant et *ridé* récessif ; *jaune* dominant et *vert* récessif ; indépendance entre couleur et forme. Le modèle mendélien prédit alors qu'à la deuxième génération, on obtiendra en moyenne la distribution suivante :

Types de pois	*rond jaune*	*rond vert*	*ridé jaune*	*ridé vert*
Fréquences théoriques	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

On réalise l'expérience de Mendel et on obtient les résultats expérimentaux suivants :

Types de pois	*rond jaune*	*rond vert*	*ridé jaune*	*ridé vert*
Effectifs	325	116	101	30

Comparer à l'aide de la distance du χ^2 et d'un test du χ^2 l'adéquation entre la distribution théorique et la distribution observée (N.B. Le seul paramètre théorique évalué à partir des observations est ici l'effectif total. Le nombre de degrés de liberté est donc égal à $4 - 1 = 3$.)
Réponse : Distribution théorique pour un effectif de 572 individus statistiques :

RoJ	*RoV*	*RiJ*	*RiV*
322	*107*	*107*	*36*

$\chi^2_{obs} = \frac{(325-322)^2}{322} + \ldots + \frac{(30-36)^2}{36} = 2.12$. *Or, pour un seuil de 5% et ddl* $= 3$, *on a* $\chi^2_c = 7.81$. *On accepte donc l'hypothèse selon laquelle l'échantillon observé est issu d'une population distribuée selon la loi théorique indiquée.*

Exercice 87 χ^2 de MacNemar
Soit deux questions d'un test, A et B, auxquelles un groupe de 184 sujets a répondu. On voudrait savoir si la fréquence des réussites et des échecs est la même pour A et B. On a :

A : 99 échecs et 85 réussites
B : 64 échecs et 120 réussites.

On sait que 62 sujets ont répondu avec succès aux deux questions.
Justifier le tableau d'effectifs observés suivant :

	B : réussite	B : échec	Total
A : réussite	62	23	85
A : échec	58	41	99
Total	120	64	184

Pour répondre à cette question, le statisticien forme le tableau d'effectifs attendus suivant :

	B : réussite	B : échec
A : réussite	62	40.5
A : échec	40.5	41

Justifier la construction de ce tableau et comparer les deux distributions à l'aide de la distance du χ^2 et d'un test (le nombre de degrés de liberté est ici *ddl* $= 1$. Pourquoi ?).
Réponse : Les deux cases porteuses d'information sont les deux cases de désaccord (A : échec ; B : réussite) et (A : réussite ; B : échec). Le tableau des effectifs théoriques est construit en répartissant l'effectif total de ces cases pour moitié dans chacune d'elles. En effet, sous l'hypothèse d'équivalence des questions A et B, les désaccords dans un sens

ou dans l'autre ont autant de chances de se produire. On obtient alors $\chi^2_{obs} = \frac{(58-40.5)^2}{40.5} + \frac{(23-40.5)^2}{40.5} = 15.12$. *Or, pour* $ddl = 1$ *et* $\alpha = 1\%$, $\chi^2_c = 6.63$. *L'hypothèse d'équivalence des deux items ne peut donc pas être retenue. Justification du nombre de degrés de liberté : seules deux observations sont utilisées, et le total :* $58 + 23 = 81$ *est utilisé pour construire le tableau des effectifs théoriques. D'où :* $ddl = 2 - 1 = 1$.

Covariance. Coefficient de corrélation

Exercice 88

Soit la série suivante de dix associations de valeurs x_i et y_i :

x_i	12	10	17	18	13	10	9	14	17	12
y_i	42	38	50	49	43	40	37	42	48	40

Représenter graphiquement le nuage de points correspondant.
Calculer la covariance et le coefficient de corrélation des deux séries.
Réponses : $Cov(x_i, y_i) = 12.92$ *;* $\rho = 0.9665$. *Forte corrélation entre les deux séries.*

Exercice 89

Quinze élèves, désignés par les lettres de A à O ont été classés une première fois par une épreuve de français, une seconde fois par une épreuve de mathématiques. Calculer le coefficient décrivant la corrélation entre ces deux classements.

Elèves	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Fran.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Math.	9	3	1	11	2	5	8	13	4	10	7	14	15	6	12

Réponses : $Cov(F, M) = 9.47$ *;* $\rho = 0.51$. *Remarquez qu'il s'agit ici d'un coefficient de corrélation des rangs. On pourra consulter le paragraphe* Corrélation des rangs de Spearman *d'un ouvrage de statistiques.*

Exercice 90

On mène une étude sur les variations circadiennes de la charge mentale induite par une tâche simple et répétitive. (*circadien signifie "sur un cycle de 24 heures"*).
On considère un échantillon homogène de sujets et on relève, à différents moments de la journée :
- la vitesse d'exécution d'une tâche répétitive simple (nombre d'appuis sur un bouton par minute)
- l'indice de charge mentale induite (mesuré à partir du temps de réaction à un stimulus auditif simple).
On obtient les résultats suivants (moyennes obtenues sur l'ensemble des sujets observés).

Moment	8	10	12	14	16	18	20	22	24
Vitesse	64,54	66,61	71,01	70,10	70,08	68,42	66,63	64,12	63,12
Indice	1,117	1,130	1,171	1,140	1,141	1,129	1,107	1,072	1,052

1) Construire un nuage de points en plaçant en abscisse la variable "moment de la journée", en ordonnée la vitesse, et en choisissant judicieusement les unités.
D'après ce graphique :
- semble-t-il exister une relation entre le moment de la journée et la vitesse ?
- serait-il pertinent de calculer un coefficient de corrélation linéaire pour évaluer l'intensité de cette relation ?

2) Mêmes questions pour les variables vitesse et indice de charge mentale.
3) Calculer la covariance et le coefficient de corrélation linéaire entre les variables vitesse (x_i) et indice de charge mentale (y_i). On utilisera les résultats intermédiaires suivants : $\sum x_i = 604,63$; $\sum y_i = 10,059$; $\sum x_i^2 = 40686,3023$, $\sum y_i^2 = 11,253289$; $\sum x_i y_i = 676,53294$.
Commenter les résultats obtenus.
Réponses : 1 et 2) Il semble exister une relation entre le moment de la journée et la vitesse, mais cette relation n'est pas linéaire, et ne peut donc pas être étudiée à l'aide d'un coefficient de corrélation. En revanche, il semble exister une relation linéaire entre la vitesse et l'indice de charge mentale.
3) $Cov(x_i, y_i) = \frac{676.53294}{9} - \frac{604.63}{9} \times \frac{10.059}{9} = 0.084$. $Var(x_i) = \frac{40686.3023}{9} - (\frac{604.63}{9})^2 = 7.39$. σ_x*=2.72. De même,* $\sigma_y = 0.0344$ *et finalement,* $\rho = \frac{0.084}{2.72 \times 0.0344} = 0.899$. *Il existe donc une forte corrélation positive entre ces deux variables.*

Exercice 91

Dans une étude psychométrique, on considère la variable QI, notée X et la variable *Combinatoire*, notée Y pour 20 individus.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
x_i	99	122	108	125	108	113	94	85	112	125
y_i	3.9	5.0	5.3	8.3	5.5	6.6	5.5	2.2	5.3	5.3

	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
x_i	108	91	91	109	125	94	120	112	106	91
y_i	4.6	3.7	4.1	2.7	6.8	2.7	5.4	6.2	2.5	2.4

Aide pour les calculs : $\sum x_i = 2138$, $\sum y_i = 94$, $\sum x_i^2 = 231666$, $\sum y_i^2 = 494.36$, $\sum x_i y_i = 10336.4$.

1) Représenter graphiquement le nuage de points (x_i, y_i).
2) Calculer la moyenne et la variance de chacune des deux séries (x_i) et (y_i) , puis la covariance et le coefficient de corrélation des deux séries.
Réponses : $\overline{x} = 106.9$; $\overline{y} = 4.7$; $\sigma^2(x) = 155.69$; $\sigma^2(y) = 2.628$; $Cov(x,y) = 14.39$; $\rho = 0.7114$.

Distribution d'échantillonnage

Exercice 92
Dans la population, la distribution du QI suit une loi normale de paramètres $\mu = 100$ et $\sigma = 15$. On tire au hasard des échantillons de taille 100 dans cette population.
Quels sont les paramètres de la variable $\overline{X}$, moyenne des QI observés sur un échantillon de taille 100 ?
Donner un intervalle rassemblant 95% des valeurs observées de $\overline{X}$.
Réponses : 1) Moy($\overline{X}$)=100 ; $\sigma(\overline{X}) = 1.5$; intervalle : $[97.06; 102.94]$.

Exercice 93
Dans la population française, la fréquence d'apparition du groupe sanguin A est 45%. On tire au hasard des échantillons de taille 400 dans cette population.
Quels sont les paramètres de la variable F, fréquence du groupe sanguin A observée sur un échantillon de 400 personnes ?
Donner un intervalle rassemblant 99% des valeurs observées de F.
Réponses : Moy(F)=0.45 ; $\sigma(F) = 0.0124$; intervalle : $[0.4181; 0.4819]$

Intervalle de confiance

Exercice 94
Dans un échantillon de 49 sujets prélevé au hasard dans une population parente, on trouve que les scores obtenus ont pour moyenne $\overline{x} = 50$ et pour écart type $s = 9,75$.
Donner des estimations ponctuelles de la moyenne μ et de l'écart type σ de la distribution des scores sur la population.
Estimer μ à l'aide d'un intervalle de confiance, avec un degré de confiance de 95%.
Si l'effectif de l'échantillon était porté à 100 sujets, l'intervalle précédent serait-il plus ou moins étendu ?
Réponses : $\hat{\mu} = 50$; $\hat{\sigma} = \sqrt{\frac{49}{48}} \times 9.75 = 9.85$. Intervalle de confiance pour μ : $[30.69; 69.31]$.
Pour un échantillon de 100 sujets, l'intervalle serait moins étendu.

Exercice 95
Lors d'un tirage au hasard d'un groupe de 100 personnes âgées de 22 ans, on constate que 59 d'entre elles poursuivent des études. A partir de cette valeur observée, calculer au seuil 0.02, les limites de confiance de la proportion des étudiants dans la population générale des personnes âgées de 22 ans.
Réponses : Intervalle : $[0.59 - 2.33 \times \hat{\sigma}, 0.59 + 2.33 \times \hat{\sigma}] = [0.475; 0.705]$.

Tests d'hypothèses

Exercice 96
Dans ce qui suit, on se donne une population de référence dans laquelle la distribution du QI est une loi normale de paramètres $\mu = 100$ et $\sigma = 15$.
Dans une classe de rattrapage, on examine un groupe de 5 élèves et on trouve un QI moyen $\overline{QI} = 84.4$. Peut-on dire que le groupe de 5 élèves est, pour le QI, atypique de la population de référence aux seuils traditionnels ?
La classe de rattrappage compte 25 élèves ; on examine tous les élèves et on trouve un QI moyen $\overline{QI} = 84.4$. Mêmes questions que précédemment.

Réponses : a) $z_{obs} = -2.236$. *Groupe atypique au seuil de 2.5% unilatéral. b)* $z_{obs} = -5.2$. *Groupe atypique au seuil de 1% unilatéral*

Exercice 97

Dans une usine, on cherche à voir si un changement de l'environnement (musique dans les ateliers en particulier) peut modifier le rendement. Ce dernier est mesuré ici par le nombre moyen de pièces produites à l'heure par chaque ouvrier.

On note pour chacun des 33 ouvriers observés, son rendement avant et après l'introduction de ces changements. Les résultats sont les suivants.

Avant	Après	Avant	Après	Avant	Après
45	48	30	35	40	45
36	40	45	50	40	35
47	53	30	40	38	35
40	40	45	50	35	40
45	46	40	45	40	45
35	30	50	50	35	37
36	40	40	40	38	35
50	60	50	45	50	50
50	60	40	35	45	50
40	40	55	50	30	33
40	40	30	35	38	38

Effectuer les calculs permettant de répondre à la question posée et conclure.

Réponses : 1) $\overline{d} = 2.03$, $s_c = 4.41$. $z_{obs} = 2.64$. *La différence est significative pour un test unilatéral au seuil de 1%.*

Exercice 98

Un institut de sondage a observé sur un échantillon de 1600 personnes prises au hasard, 51% d'intentions de vote en faveur du candidat X.

Estimer la proportion d'intentions de vote en faveur du candidat X dans la population à l'aide d'un intervalle de confiance, avec un degré de confiance de 95%.

Peut-on assurer, avec un risque de 5%, que M. X sera élu ?

Réponses : Intervalle de confiance : $[0,485; 0,535]$. *On ne peut donc pas assurer que M. X sera élu*