

EXERCICE I :

Voici le début d'un questionnaire d'enquête.

1. VILLE D'ECHIROLLES

RECENSEMENT DES DIRIGEANTS SPORTIFS D'ECHIROLLES

1. AGE 2. SEXE : Masculin ☐ Féminin ☐ 3. Nationalité
4. Situation familiale : Marié(e) ☐ Célibataire ☐ Veuf(ve) ☐ Séparé(e) ☐
5. Avez-vous des enfants ? oui ☐ non ☐ 6. Si oui, combien ?
7. Quel est leur âge ?
8. Si vous habitez à ECHIROLLES, dans quel quartier habitez-vous ?
Si non, dans quelle commune habitez-vous ?
9. Depuis combien de temps ?
10. Quelle est votre formation générale ?
Primaire ☐ Secondaire ☐ Supérieure ☐ Sans formation. ☐
11. Quelle est votre formation professionnelle ?
C.A.P. ☐ B.T. ☐ B.T.S. ☐ Ingénieur ☐ Autres : Sans formation ☐

- 1) Quelle est la population statistique étudiée ? Est-ce une population de taille finie ou de taille infinie? Quelles sont les caractéristiques en commun délimitant la population ?
- 2) Quel est l'individu statistique (unité sur laquelle l'observation a été effectuée) ?
- 3) Quelles sont les variables associées à ces questions ? Déterminer le type de chacune de ces variables. Dans quelle échelle de mesure sont-elles observées ?

EXERCICE II :

A) Une échelle de Guttman a pour objet soit d'ordonner les différentes réponses d'un sujet (Attitude, Sensations, Préférences ...) soit d'ordonner plusieurs sujets les uns par rapport aux autres d'après leurs réponses dans une situation déterminée identique pour tous (tests). Considérons l'exemple d'un test composé de 20 questions, chaque sujet doit cocher l'une des deux réponses proposées par question (Q.C.M). Toute réponse juste vaut "1" ("0" sinon). La note finale obtenue par chaque sujet est le total des "1" obtenus.

- 1) Quelles sont les notes possibles que peut avoir un sujet?
- 2) De combien de façons possibles un sujet peut obtenir une note valant "0", valant "1", valant "20"?

3) Pouvez-vous donner une idée (ou encore mieux la réponse exacte) du nombre de façons possibles dont un sujet peut obtenir la note "4"?

B) Ayant fait passer ce test à un certain nombre de sujets on peut :

- considérer les questions comme équivalentes et ordonner les sujets d'après leurs notes.

Comment dans ce cas ordonner les sujets en particulier ceux ayant obtenu une même note? ("4" par exemple)

- ordonner les questions (par ordre de difficulté par exemple) et selon cette classification il faut ordonner les sujets. Dans la plupart des tests (Psychologie, Sociologie ..) on considère que les sujets sachant répondre aux questions d'un niveau donné de difficulté sachent aussi répondre sans erreur à toutes les questions d'un niveau de difficulté inférieur. Telle est la propriété caractéristique des échelles de Guttman.

- 1) Donner quelques exemples d'une telle situation

2) En notant "+" à l'intersection d'une ligne et d'une colonne pour indiquer une réponse juste à une question donnée, compléter le tableau suivant pour illustrer une échelle de **Guttman**. (a, b, c, ... sont les questions placées dans le "bon" ordre; 0,1, 2, ...sont les notes finales)

	a	b	c	d	e	f	...
0							
1							
2							
3							
4							
..							

C) Application : Le Consensus social

Dans l'exemple précédent on avait une hypothèse a priori sur l'ordre des stades opératoires, mais on peut utiliser les échelles de Guttman même dans le cas d'absence d'une telle hypothèse a priori. C'est le cas par exemple (exemple étudié par la suite) de la construction d'une échelle d'opinion et le classement des personnes en fonction du pattern de réponses qu'elles fournissent. Une échelle d'opinion est présentée à un groupe de 10 sujets qui doivent dire s'ils sont d'accord ou non avec les propositions suivantes :

I	Je souhaite vivement la suppression du baccalauréat dans sa forme actuelle	oui	non
II	Il paraît préférable que le baccalauréat dans sa forme actuelle soit supprimé	oui	non
III	Je suis prêt à organiser des manifestations sur la voie publique pour la suppression du baccalauréat dans sa forme actuelle	oui	non

On obtient les résultats suivants :

sujets	I	II	III
1	-	+	+
2	+	+	-
3	+	-	-
4	+	+	-
5	-	+	-
6	-	-	-
7	+	+	+
8	-	-	-
9	-	-	-
10	-	-	-

- 1) Totaliser le nombre de « + » par colonne et en déduire une échelle de Guttman reflétant le consensus du groupe.
- 2) Totaliser le nombre de « + » par ligne et reconstruire le tableau en tenant compte de l'ordre de l'opinion et de l'ordre sur les sujets.
- 3) Relever tous les cas de patterns non prévisibles d'après le modèle.
- 4) Commenter.

Travail Personnel N°1

Sauf exception ces exercices ne seront pas traités en séances normales de travaux dirigés.

PROBLEME : étude d'une échelle ordinale

Partie A : Généralités.

Contrairement aux échelles nominales, les expressions telles que « plus grand que », « précède », « se place après » etc.... prennent un sens dans une échelle ordinale.

1) Pourquoi ?

2) Ce sens leur est-il conféré par le statisticien ou par l'utilisateur de l'outil statistique (psychologue, sociologue, biologiste, économiste etc. ...) ?

Pour construire une échelle ordinale le statisticien (le mathématicien !) considère une relation $\mathcal{R}$ entre les observations (entre les individus) qui vérifie formellement les deux propriétés suivantes :

(a) Si cette relation $\mathcal{R}$ permet de dire que « A est plus grand que B » (on note $A\mathcal{R}B$ ou $A \geq B$) elle ne doit pas permettre de dire que « B est plus grand que A » (sauf cas d'égalité !). On dit que la relation est antisymétrique.

Exemple 1 : Dans un tournoi (foot, tennis, échec etc. ...) si une équipe A est « plus forte » qu'une équipe B (par exemple au niveau du score) alors on ne peut avoir l'équipe B est « plus forte » que l'équipe A (sauf cas d'égalité !).

(b) Si cette relation permet de dire que « A est plus grand que B » et que « B est plus grand que C » alors nécessairement on doit avoir « A est plus grand que C ». On dit que la relation est transitive.

3) Dans l'exemple 1 a-t-on une relation transitive ?

Exemple 2 : Mesure de la douleur

La « pompe à morphine » est un robot distributeur de morphine utilisable directement par le patient pour adapter le traitement anti-douleur à son besoin. Un réglage préalable du distributeur doit être fait par le personnel soignant. Pour effectuer ce réglage propre à chaque malade l'infirmière (ou le médecin ...) demande au patient de positionner un curseur sur une « échelle de douleur » pour représenter ce qu'il ressent, et ceci un certain nombre de fois à différentes heures de la journée. Cette échelle est formée d'une réglette munie d'un curseur coulissant sur l'une des deux faces. Sur cette face un triangle (rouge !) indique une augmentation de la douleur dans le sens de la flèche représentée par l'angle aigu.

A chaque patient on associe quatre relevés (mesures !) de la douleur à 8h, 9h, 10h et 11h.



4) La relation « plus douloureux que » est-elle antisymétrique ? Transitive ?

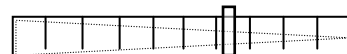
5) Si deux patients donnent une même position du curseur ressentent-ils la même douleur ? S'ils indiquent deux positions différentes peut-on comparer leurs douleurs ?

6) Et si quelqu'un prétend que le curseur dépasse la réglette ?

La deuxième face de la réglette est graduée de façon régulière (c'est à dire ?) de 0 à 10 ainsi le curseur indique un nombre.

Pour un patient on obtient les relevés suivants :

8h	9h	10h	11h
4	5	8	9



7) Peut-on dire qu'à 10h il a eu plus mal qu'à 8h ?

8) Peut-on dire qu'à 10h il a eu deux fois plus mal qu'à 8h ?

9) Peut-on affirmer que l'augmentation de la douleur entre 10h et 11h est la même que celle entre 8h et 9h ?

10) Donner quelques exemples d'échelles ordinales dont au moins un exemple tiré de votre discipline principale (psychologie, sociologie).

Partie B : (Echelle de Thurstone, Méthode des comparaisons par paires)

On désire placer 5 métiers (A, B, C, D et E) sur une échelle ordinale qui rende compte de l'attrait que ces métiers ont pour des adolescents. Pour cela on considère un groupe de collégiens (en pratique on considère un échantillon représentatif de la population des adolescents). A chaque collégien on présente deux métiers X et Y et on lui demande lequel il préfère. On

présente ainsi toutes les paires possibles (il y en a combien ?) à tous les sujets du groupe. On compte le nombre de sujets qui préfèrent **X** à **Y** et on calcule le pourcentage des ces sujets. Cette méthode est connue sous l'appellation de « comparaison par paires ». Elle est très utilisée dans les enquêtes d'opinion et principalement, en esthétique expérimentale.

On présente les résultats de cette étude sous la forme d'un tableau « à deux entrées » qui donne le pourcentage dans lequel le métier placé en haut de colonne a été préféré au métier indiqué sur la ligne à gauche.

	A	B	C	D	E
A	0	60	30	20	20
B	40	0	0	0	0
C	70	100	0	60	30
D	80	100	40	0	70
E	80	100	70	30	0

- 1) Dans combien de cas (en pourcentage) le métier **C** a été préféré au métier **D** ?
- 2) Pourquoi sur la diagonale principale ne figurent que des zéros ?
- 3) En faisant la somme des valeurs symétriques par rapport à la diagonale principale que trouve-t-on ? Expliquer.
- 4) a) Compléter la dernière ligne du tableau par le total des pourcentages par colonne. Quelle interprétation peut-on en donner ?
b) Dédurre de ce qui précède une échelle de préférence des métiers.
- 5) Refaire le tableau précédent en tenant compte de l'ordre dans cette échelle de préférence.
- 6) Si cette échelle répondait exactement au « modèle » d'une échelle ordinale que devrait-on observer ?
- 7) Relever sur le tableau que vous avez dressé (question 5) des cas où :
a) l'antisymétrie n'est pas respectée b) La transitivité n'est pas respectée.
- 8) Conclure.

Travaux Dirigés N°2

Fréquences, pourcentages, proportions, taux,

EL METHNI M.

EXERCICE I :

Afin d'établir un rapport éventuel entre l'âge et les loisirs un psychosociologue enquête auprès d'une population de 20 personnes et obtient les informations suivantes :

Notations S : Sport C : Cinéma T : Théâtre L : Lecture

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X: Age	12	14	40	35	26	30	30	50	75	50	30	45	25	55	28	25	50	40	25	35
Y: Loisir	S	S	C	C	S	T	T	L	L	L	T	C	C	C	S	L	L	C	T	T

- 1) Que représente la première ligne de ce tableau ? Sur combien de sujets l'enquête a-t-elle porté?
- 2) Combien a-t-on de variables? Quel est le type de chaque variable et dans quelle échelle est-elle mesurée?
- 3) Réaliser le tri à plat de chaque série de données et dresser le tableau statistique standard : (distribution des fréquences et des pourcentages, fréquences cumulées et fonction de répartition s'il y a lieu)
- 4) Quel est le pourcentage des sujets :
_a) âgés de moins de 30 ans? b) âgés d'au plus 30 ans? c) âgés d'au moins 30 ans?
_d) qui préfèrent le cinéma? e) qui ne préfèrent pas la lecture?
- 5) La différence des deux pourcentages calculés en (b) et (a) est-elle aussi un pourcentage ? Comment l'interpréter ?
- 6) Les tableaux que vous avez dressés permettent-ils de répondre à la question : "Quel est le nombre de sujets âgés d'au moins 30 ans et qui préfèrent le cinéma?". Si oui donner ce nombre sinon pouvez-vous proposer une méthode pour calculer ce nombre ?

EXERCICE II :

Dans une émission télévisée « la marche du siècle » M^r Jospin déclara « 10% des femmes sont des hauts fonctionnaires ». L'animateur, J-M Cavada ne remarqua rien !

Par contre vous, vous allez immédiatement relever « la confusion » faite par le 1^{er} ministre.

EXERCICE III : D'après examen septembre 00

« Evénement du Jeudi » N°446 Mai 1993

Dans une étude sur l'homogamie (mariage d'un homme et d'une femme de la même catégorie socioprofessionnelle) et selon l'INSEE, si plus de la moitié des femmes cadres (52%) épousent un conjoint également cadre, en revanche 25% seulement des hommes cadres épousent une « cadresse ». Cette différence ne manqua pas de susciter l'étonnement d'une journaliste :

Un quart des cadres épousent une cadresse de la même eau. Mais poussons du côté de ces dames : elles sont plus de 50% à épouser leurs égaux socialement parlant. Où trouvent-elles donc le quart manquant ? A moins qu'eux en épousent deux. Successivement donc. Mais si je me trompe, vous me le dites, parce que l'INSEE elle, motus. Mystère et statistique.

Expliquer à cette journaliste son erreur et dévoiler le « mystère ».

EXERCICE IV : D'après examen Juin 02

Sur une planète (très lointaine) cohabitent deux populations. Les bleus et les verts. Chaque habitant est soit riche soit pauvre. 95% des bleus sont pauvres et 95% des pauvres sont bleus.

Y a-t-il sur cette planète une inégalité sociale due à la couleur ?

Etayer votre réponse (de nature purement statistique) par des exemples numériques.

Travail Personnel N°2

Sauf exception ces exercices ne seront pas traités en séances normales de travaux dirigés.

EXERCICE I :

Lors d'une enquête sur la publicité télévisée on a demandé à chaque sujet d'évaluer, en minutes, le temps des coupures publicitaires pendant le film du dimanche soir sur deux chaînes.

Les résultats obtenus sont récapitulés dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chaînes 1	35	29	30	31	33	32	33	34	32	33	32	30	28	34	33	33
chaînes 2	35	36	34	34	36	37	33	35	35	35	35	33	36	37	34	35

- 1) Que représente la première ligne de ce tableau ? Combien de sujets ont participé à l'expérience?
- 2) Combien a-t-on de variables? Quel est le type et l'échelle de mesure?
- 3) Réaliser le tri à plat de chaque série de données et dresser le tableau statistique standard : (distribution des fréquences et des pourcentages, fréquences cumulées et fonction de répartition)
- 4) Quel est le pourcentage des sujets dont l'évaluation du temps de coupure sur la chaîne 1 :
 _ est strictement inférieur à 33 minutes? _ dépasse 31 minutes?
 _ est compris entre 30 et 34 minutes?
- 5) Les tableaux que vous avez dressés permettent-ils de répondre à la question : "Quel est le nombre de sujets ayant une évaluation du temps de coupure sur la chaîne 1 inférieur à 30 minutes et un temps de coupure sur la chaîne 2 supérieur à 35 minutes?". Si oui donner ce nombre sinon pouvez-vous proposer une méthode pour calculer ce nombre?

EXERCICE II :

On considère les quatre types de comparaisons :

Comparaison	Additive (ou absolue) par la différence	Multiplicative (ou relative) par le rapport
Instantanée	Il gagne 200 € de plus que sa femme	Il gagne 3 fois plus que sa femme
Entre deux dates	Il gagne 300 € de plus que l'année dernière	Il gagne 2 fois plus que l'année dernière

- 1) Donner au moins un exemple pour chacun des quatre cas du tableau.
- 2) Dans la comparaison « il gagne 2000 F de plus que sa femme » imaginons les deux situations suivantes :
 1^{ère} situation : le mari gagne 8000 F et sa femme 6000 F
 2^{ème} situation : le mari gagne 51000 F et sa femme 49000 F
 Traduire chacune des deux situations en exprimant le salaire de la femme comme un pourcentage de celui de son mari et commenter.
- 3) Lors d'un débat entre deux hommes politiques (adversaires bien sûr !) :
 1^{ère} homme politique : grâce à notre action le budget de l'éducation nationale a augmenté de tant cette année. Les chiffres officiels le prouvent.
 2^{ème} homme politique : désolé mais ces mêmes chiffres montrent que sa part dans le budget total a baissé.
 Pourquoi ce débat n'est « contradictoire » qu'en apparence ? Commenter.
- 4) Avez-vous des connaissances en économie ?
 Considérons les inflations annuelles de deux pays d'économies comparables (France et Allemagne par exemple). Sur une période de deux années successives on obtient :

	F	A
1985	13,5%	5,5%
1986	5,8%	2,2%

- a) Pour chaque année faire une comparaison absolue des deux inflations
- b) Pour chaque année faire une comparaison relative des deux inflations
- c) Comparer les deux résultats. Que peut-on en conclure ? Que pensez-vous du cas où l'un des deux pays a une inflation nulle ? Demander à un économiste de vous expliquer tout ça !

EXERCICE III :

1) Une étude portant sur les trois dernières années a montré que 40% des étudiants de 1^{ère} année de sciences humaines ne maîtrisent pas complètement l'utilisation d'une calculatrice et commettent des erreurs de calcul lors de l'examen de statistique. La même étude établit aussi que 25% de ces étudiants qui ne maîtrisent pas complètement l'utilisation d'une calculatrice échouent l'examen de statistique.

Quel est le pourcentage des étudiants qui échouent leur examen de statistique à cause de leur mauvaise utilisation d'une calculatrice ?

2) Vous placez une certaine somme d'argent qui vous rapporte 10% la première année. Vous continuez à placer cette somme avec les intérêts acquis et vous perdez 10% l'année suivante. Est-ce que vous récupérez votre somme initiale ?

3) Dans un village de 200 habitants 99% des habitants sont des fumeurs et le reste (combien?) ne sont pas des fumeurs. Combien de fumeurs faut-il retirer de cette population pour qu'ils ne représentent plus que 98%? Donner une première réponse rapide et approximative (intuitive!) puis calculer exactement ce nombre et comparer avec votre première réponse. Que conclure?

EXERCICE IV :

L'indice de la production industrielle est un nombre calculé « provisoirement » chaque mois et publié sous la forme d'un nombre entier arrondi à l'unité près. Un journal (s'adressant aux entreprises) publie deux indices mensuels successifs : 127 et 128 et conclue que la production industrielle a augmenté de 0,8% en un mois. $\left(\frac{128-127}{127} \approx 0,78\% \right)$.

Réagissez à cette « erreur ».

EXERCICE V :

D'après « le Monde » du 22/23 septembre 1991 sous le titre « Volkswagen assigne Renault en justice »

« Renault vend deux fois plus de voitures en Allemagne que Volkswagen en France cela vous étonne? Pas nous! »

Cette publicité signée Renault et Publicis n'est pas du goût du constructeur allemand qui assignait, vendredi 20 septembre, son concurrent en justice. Pour étayer leur slogan, Renault et son agence Publicis avaient choisi d'avancer des chiffres : « le constructeur allemand a vendu en France 57000 voitures au premier semestre 1991. Une belle performance!(...) Renault a vendu en Allemagne 137000 voitures pendant la même période (...). Une très belle performance. »

Volkswagen ne conteste pas les chiffres choisis mais les juge « non significatifs ».

Sur quel(s) argument(s) statistique(s) se base Volkswagen pour plaider sa cause?

Travaux Dirigés N°3

Tableaux, diagrammes, mode, médiane, quartiles ...

EL METHNI M.

EXERCICE I : (On reprend l'exercice I du TD2)

Afin d'établir un rapport éventuel entre l'âge et les loisirs un psychosociologue enquête auprès d'une population de 20 personnes et obtient les informations suivantes :

Notations S : Sport C : Cinéma T : Théâtre L : Lecture

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X: Age	12	14	40	35	26	30	30	50	75	50	30	45	25	55	28	25	50	40	25	35
Y: Loisir	S	S	C	C	S	T	T	L	L	L	T	C	C	C	S	L	L	C	T	T

1) Reprendre le tri à plat de la variable Y réalisé au TD2 et tracer un diagramme en bâtons le représentant. Proposer d'autres diagrammes représentatifs. Peut-on parler de courbe cumulative ?

2) Donner les valeurs extrêmes ainsi que le milieu de X. Cette question a-t-elle un sens pour Y?

3) Déterminer le(s) mode(s) de Y. Comparer avec la représentation graphique. Déterminer le(s) mode(s) de Y.

4) Calculer la médiane ainsi que les quartiles de X. Cette question a-t-elle un sens pour Y?

5) Dresser un tableau (à 2 dimensions) donnant la distribution de l'âge par type de loisir.

6) Calculer l'âge médian (la médiane des âges!) par type de loisir. Calculer la médiane de ces 4 médianes par type de loisir. Que conclure?

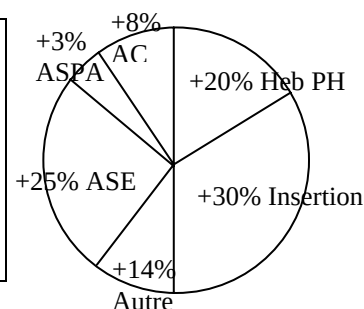
	S	C	T	L
12				
14				
26				
28				

EXERCICE II :

Les dépenses d'aide sociale ont augmenté entre 1992 et 1994. Le journal « Le Monde » du 6 juillet 1995 nous le détaille sur un graphique donnant le pourcentage des augmentations par secteur accompagné d'un commentaire (à titre d'exemple) :

Pour 100 € d'augmentation des dépenses nettes d'aide sociale départementale :

- 30 € sont imputables à l'insertion,
- 25 € à l'aide sociale à l'enfance (ASE),
- 20 € à l'hébergement des personnes handicapées,
- 8 € à l'allocation compensatrice,
- 3 € à l'aide sociale aux personnes âgées,
- 14 € aux autres dépenses d'aide sociale.



Ces informations encadrées sont justes et semblent cohérents avec le diagramme (on a bien, par exemple, 30 € parmi les 100 € d'augmentation sont imputables à l'insertion d'où le 30%). Mais à bien y réfléchir ce graphique est faux !

Pour vous aider à l'analyser et rectifier les erreurs : En cours on vous a bien expliqué que les différents pourcentages totalisent 100%, ce qui est le cas ici, et pourtant c'est là que réside le « hic » !!!

Travail Personnel N°3

Sauf exception ces exercices ne seront pas traités en séances normales de travaux dirigés.

EXERCICE I : (On reprend l'exercice I du TP2)

Lors d'une enquête sur la publicité télévisée on a demandé à chaque sujet d'évaluer, en minutes, le temps des coupures publicitaires pendant le film du dimanche soir sur deux chaînes.

Les résultats obtenus sont récapitulés dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chaînes 1	35	29	30	31	33	32	33	34	32	33	32	30	28	34	33	33
chaînes 2	35	36	34	34	36	37	33	35	35	35	35	33	36	37	34	35

On désigne par X la variable : temps de coupure sur la chaînes 1 et par Y la variable : temps de coupure sur la chaînes 2.

- 1) Reprendre les tris à plat réalisés dans le TP2 et donner les représentations graphiques des distributions ainsi que celle des fonctions de répartition.
- 2) Donner les valeurs extrêmes ainsi que le milieu de X et de Y .
- 3) Calculer la médiane ainsi que les quartiles de X et de Y .
- 4) On considère le tri à plat relatif à X . (voir exercice I TP2 question 3).

a) Retrouver à partir de ce tableau la médiane ainsi que les quartiles de X .

x_i	28	29	30	31	32	33	34	35	
n_i	1	1	2	1	3	5	2	1	16
N_i	1	2	4	5	8	13	15	16	

b) Déterminer le(s) mode(s) de X .

Comparer avec la représentation graphique.

- 4) Reprendre la question (2) pour la variable Y .

EXERCICE II :

On fait passer à 100 sujets une épreuve notée sur une échelle continue de 0 à 20. On obtient la distribution suivante :

0	0,5	1	1	1,5	2	2	2	3	3	4	4	4	4	4	4	4
4	4	4	4	4	4	4,5	4,5	4,5	4,5	4,5	4,5	4,5	5	5	5	5
5	5	5	5	5	5	5	5	5	5,5	5,5	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6,5	6,5	6,5
6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	7	7	7	7
7	7	7	7	7	8	8	8	8	8	8,5	8,5	8,5	9	10		

- 1) Organiser cette distribution en 5 classes de même amplitude. La borne inférieure de la 1^{ère} classe est 0 et la borne supérieure de la dernière classe est 10.

- 2) Dresser le tableau statistique standard : (distribution des fréquences et des pourcentages, fréquences cumulées et fonction de répartition, densité, centre des classes etc.)

- 3) Donner la représentation graphique de la distribution ainsi que celle de la fonction de répartition.

- 4) Quelle est la classe de plus grande densité? Comment l'appelle-t-on?

- 5) Pourquoi le calcul de la densité est-il inutile dans ce cas ? Dans quel cas ce calcul est-il nécessaire ?

- 6) Donner le(s) mode(s), les valeurs extrêmes et l'étendue de cette série statistique.

- 7) Calculer la médiane ainsi que les quartiles

- 8) On considère le tableau regroupant en 5 classes d'égales amplitudes les valeurs de X .

a) Retrouver à partir de ce tableau la médiane ainsi que les quartiles de X .

Note	[0 2]	[2 4]	[4 6]	[6 8]	[8 10]	
n_i	8	15	42	30	5	100
N_i	8	23	65	95	100	

Comparer avec la question 2 et commenter.

b) Reporter ces valeurs sur la représentation graphique de la fonction de répartition et commenter.

c) Déterminer la (les) classe(s) modale(s). Comparer avec la représentation graphique.

- 9) Calculer le 3^{ème} et le 7^{ème} décile. Que représente l'intervalle $[d_3 \quad d_7]$?

Travaux Dirigés N°4

Moyennes

EL METHNI M

EXERCICE I : (On reprend l'exercice I du TD2, du TD3 et du TD4)

Afin d'établir un rapport éventuel entre l'âge et les loisirs un psychosociologue enquête auprès d'une population de 20 personnes et obtient les informations suivantes :

Notations S : Sport C : Cinéma T : Théâtre L : Lecture

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X: Age	12	14	40	35	26	30	30	50	75	50	30	45	25	55	28	25	50	40	25	35
Y: Loisir	S	S	C	C	S	T	T	L	L	L	T	C	C	C	S	L	L	C	T	T

1) Calculer la moyenne des âges $\bar{X}$. Ce calcul est à faire en utilisant les fonctions statistiques d'une calculatrice.

2) On considère le tri à plat (voir exercice I TD2).

a) Calculer la moyenne de X en utilisant les deux premières lignes du tableau.

x_i	12	14	25	26	28	30	35	40	45	50	55	75	
n_i	1	1	3	1	1	3	2	2	1	3	1	1	20
f_i	0,05	0,05	0,15	0,05	0,05	0,15	0,1	0,1	0,05	0,15	0,05	0,05	

b) Calculer la moyenne de X en utilisant la première et dernière ligne du tableau. Comparer et commenter

3) Dresser un tableau (à 2 dimensions) donnant la distribution de l'âge par type de loisir.

	S	C	T	L
12				
14				
26				
28				

4) Calculer l'âge moyen (la moyenne des âges!) par type de loisir. Retrouver l'âge moyen global à partir de ces 4 moyennes par type de loisir.

EXERCICE II : (On reprend l'exercice II du TD3 et du TD4)

Le tableau suivant fournit la mesure, en dixième de mm, de la finesse de discrimination sensorielle de 105 sujets. (On pose sur la peau du sujet les deux pointes d'un compas plus ou moins rapprochés et on mesure la distance minimale perçue par le sujet).

Finesse	n_i	Finesse	n_i	Finesse	n_i	Finesse	n_i	Finesse	n_i
0,1	1	3,0	1	5,6	2	8,6	3	13,5	1
0,2	2	3,6	1	5,7	2	8,7	1	14,0	1
0,4	2	3,7	1	5,8	2	8,8	3	14,2	1
0,6	3	3,9	1	5,9	1	9,5	1	15,9	1
0,7	1	4,0	1	6,3	1	10,0	1	16,2	1
0,8	1	4,1	1	6,4	1	10,1	3	18,0	2
1,3	2	4,2	1	6,6	1	10,3	1	18,3	1
1,5	3	4,3	3	6,7	2	10,6	1	18,6	1
1,7	1	4,4	1	6,8	1	10,9	1	19,1	1
1,8	1	4,6	2	7,1	3	11,5	1	19,8	1
1,9	1	4,7	3	7,7	1	11,6	1	20,0	1
2,4	1	5,0	1	8,0	2	11,9	1	21,0	1
2,5	1	5,2	1	8,2	2	12,4	1	35,0	1
2,6	2	5,3	1	8,3	1	12,8	1	49,9	1
2,7	1	5,4	1	8,5	2	13,1	2	67,8	1

- 1) Si ça vous amuse calculer la moyenne de la finesse de discrimination sensorielle des 105 sujets en utilisant les fonctions statistiques d'une calculatrice ?
 2) On considère le découpage en 7 classes de même intervalle.

Finesse	[0,1 9,8]	]9,8 19,5]	]19,5 29,2]	]29,2 38,9]	]38,9 48,6]	]48,6 58,3]	]58,3 68]	
n_i	75	24	3	1	0	1	1	105
N_i	75	99	102	103	103	104	105	

Calculer à partir de ce tableau la moyenne de X . Comparer et commenter.

EXERCICE III : (D'après examen Juin 2002)

A l'issue de chaque contrôle, un enseignant attribue une note à chaque copie. Pour mieux informer les élèves et leurs parents il reporte en plus sur chaque copie la moyenne (arithmétique) de l'ensemble des notes de la classe. Toto (un élève modèle) ramène ses copies pour être vues et signées par ses parents.

1) 1^{er} contrôle :

Toto : « Regardez, j'ai eu plus que la moyenne de la classe. Donc je suis dans la première moitié de la classe »

Les parents (pas vraiment convaincus) : « Ah ! Vraiment... »

a) En vous appuyant sur des exemples numériques (par exemple une classe de 6 élèves) montrer que toto peut tout aussi bien avoir tort que raison.

b) Quel est le résumé statistique (l'indice) le mieux adapté pour ce genre de discussion ?

2) 2^{ème} contrôle :

Toto (fou de joie) : « Hé ! J'ai eu la meilleure note de la classe. Regardez »

La mère (un peu sceptique) : « Je vois plutôt que ta note est la même que la moyenne de la classe »

Le père (un peu médisant) : « Dans ce cas tu aurais pu tout aussi bien dire que tu as eu la plus mauvaise note de la classe ».

Ramenez la sérénité dans cette famille en donnant raison à tout le monde (par un exemple numérique).

Travail Personnel N°4

Sauf exception ces exercices ne seront pas traités en séances normales de travaux dirigés.

EXERCICE I : (On reprend l'exercice I du TP2, du TP3 et du TP4)

Lors d'une enquête sur la publicité télévisée on a demandé à chaque sujet d'évaluer, en minutes, le temps des coupures publicitaires pendant le film du dimanche soir sur deux chaînes.

Les résultats obtenus sont récapitulés dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chaîne 1	35	29	30	31	33	32	33	34	32	33	32	30	28	34	33	33
chaîne 2	35	36	34	34	36	37	33	35	35	35	35	33	36	37	34	35

On désigne par X la variable : temps de coupure sur la chaîne 1 et par Y la variable : temps de coupure sur la chaîne 2.

1) Calculer la moyenne de X et de Y . Ce calcul est à faire en utilisant les fonctions statistiques d'une calculatrice.

2) On considère le tri à plat relatif à X .

a) Calculer la moyenne de X en

x_i	28	29	30	31	32	33	34	35	
n_i	1	1	2	1	3	5	2	1	16
f_i	0,0625	0,0625	0,125	0,0625	0,1875	0,3125	0,125	0,0625	

utilisant les deux premières lignes de ce tableau.

b) Calculer la moyenne de X en utilisant la première et la troisième ligne de ce tableau.

3) Reprendre la question (2) pour la variable Y .

EXERCICE II : (On reprend l'exercice II du TP3 et du TP4)

On fait passer à 100 sujets une épreuve notée sur une échelle continue de 0 à 20. On obtient la distribution suivante :

0	0,5	1	1	1,5	2	2	2	3	3	4	4	4	4	4	4	4
4	4	4	4	4	4	4,5	4,5	4,5	4,5	4,5	4,5	4,5	5	5	5	5
5	5	5	5	5	5	5	5	5	5,5	5,5	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6,5	6,5	6,5
6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	7	7	7	7	7
7	7	7	7	7	8	8	8	8	8	8,5	8,5	8,5	9	10		

On désigne par X la variable : note obtenue à l'épreuve.

1) Calculer la moyenne de X en utilisant une calculatrice !

2) On considère le tableau regroupant en 5 classes d'égales amplitudes les valeurs de X . (voir exercice II TP 2 question 1).

Note	[0 2]	[2 4]	[4 6]	[6 8]	[8 10]	
n_i	8	15	42	30	5	100
f_i	0,08	0,15	0,42	0,3	0,05	

1).

a) Calculer la moyenne de X .

b) Comparer avec (1) et commenter.

EXERCICE III : (D'après examen septembre 99)

Pour une étude de la transmission de la pensée un parapsychologue réalise l'expérience suivante : Il place deux personnes dans deux pièces séparées, l'une des deux personnes écrit une lettre (choisie parmi les 26 lettres de l'alphabet latin), la deuxième personne écrit alors une lettre (qui devra être la même si effectivement il y a transmission de la pensée et dans ce cas on dit qu'il y a réussite). Pour chaque couple il répète l'expérience dix fois, et il note le nombre de réussites. Il teste 24 couples classés en quatre groupes :

G_1 : Le groupe des couples homme-homme

G_2 : Le groupe des couples homme-femme

G_3 : Le groupe des couples femme-homme

G_4 : Le groupe des couples femme-femme

Il obtient les résultats ci-contre :

Groupes	G_1	G_2	G_3	G_4
N o t e s	7	3	7	5
	5	7	7	4
	2	7	5	8
	2	1	5	4
	2	7	6	4
	8		6	
	2		6	

1) Calculer la moyenne des notes par groupe.

2) En déduire la moyenne globale.

EXERCICE IV : Cyclisme et reportage sportif!

« C'est fantastique! » s'écria le journaliste, « non seulement notre champion a réalisé une performance en bouclant le 1^{er} tour de piste avec une vitesse de 60 Km/h mais il vient de pulvériser tous les records avec un 2^{ème} tour de piste à 80 Km/h! C'est fantastique! Le voilà lancé à l'assaut du 3^{ème} tour » En attendant la fin de ce 3^{ème} tour, le journaliste ne peut pas s'empêcher de commenter « Comme vous l'avez sans doute calculé vous-même, notre champion a bouclé deux tours avec une moyenne fulgurante de 70 Km/h c'est »

Vous qui êtes, maintenant, un vrai spécialiste de la moyenne vous ne pouvez pas vous retenir de sourire tout en pensant « Ah! Ces journalistes toujours champions pour les gaffes! »

Pourquoi ?

Travaux Dirigés N°5

Dispersion

EL METHNI M.

EXERCICE I : (On reprend l'exercice I du TD2, du TD3 et du TD4)

Afin d'établir un rapport éventuel entre l'âge et les loisirs un psychosociologue enquête auprès d'une population de 20 personnes et obtient les informations suivantes :

Notations S : Sport C : Cinéma T : Théâtre L : Lecture

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X: Age	12	14	40	35	26	30	30	50	75	50	30	45	25	55	28	25	50	40	25	35
Y: Loisir	S	S	C	C	S	T	T	L	L	L	T	C	C	C	S	L	L	C	T	T

1) Calculer la variance et l'écart-type des âges X . Ce calcul est à faire en utilisant les fonctions statistiques d'une calculatrice.

2) On considère le tri à plat (voir exercice I TD2).

x_i	12	14	25	26	28	30	35	40	45	50	55	75	
n_i	1	1	3	1	1	3	2	2	1	3	1	1	20
f_i	0,05	0,05	0,15	0,05	0,05	0,15	0,1	0,1	0,05	0,15	0,05	0,05	

a) Calculer la variance et l'écart-

type de X en utilisant les deux premières lignes du tableau.

b) Calculer la variance et l'écart-type de X en utilisant la première et dernière ligne du tableau. Comparer et commenter

3) On reprend le tableau (à 2 dimensions) donnant la distribution de l'âge par type de loisir.

Calculer la variance des âges par type de loisir. Peut-on retrouver la variance et l'écart-type des âges des 20 sujets à partir de ces 4 variances par type de loisir ? (Question ouverte dont la réponse sera traitée ultérieurement)

	S	C	T	L
12				
14				
26				
28				

EXERCICE III :

Cinq collèges A, B, C, D et E participent à une rencontre sportive scolaire. Chaque collège présente un groupe (que l'on désigne par la même lettre que son collège) d'élèves sportifs. Ces cinq groupes sont formés par un même nombre de participants. Une note finale sur 100 est attribuée à chaque participant. Cette note est synthétisée à partir des différentes épreuves de la rencontre. Elle mesure la performance globale du participant.

Groupes	A	B	C	D	E
Moyenne	87	75	86	87	75
Ecart-type	12	4	11	22	20

1) Si on vous demande de désigner les groupes qui se ressemblent le plus du point de vue de la performance lesquels choisiriez-vous ? Et pour quelle(s) raison(s) ?

2) Si on vous demande de désigner le groupe contenant le plus de participants performants lesquels choisiriez-vous ? Et pour quelle(s) raison(s) ?

3) Si on vous demande de désigner le groupe contenant les participants les plus performants lequel choisiriez-vous ? Et pour quelle(s) raison(s) ?

4) (facultative +1 point). Expliquer pourquoi on peut calculer la moyenne des performances des cinq groupes réunis et calculer cette moyenne globale.

5) (facultative +1 point). Peut-on calculer la variance des performances des cinq groupes réunis ? Si oui calculer cette variance globale sinon expliquer.

Travail Personnel N°5

Sauf exception ces exercices ne seront pas traités en séances normales de travaux dirigés.

EXERCICE I : (On reprend l'exercice I du TP2, du TP3 et du TP4)

Lors d'une enquête sur la publicité télévisée on a demandé à chaque sujet d'évaluer, en minutes, le temps des coupures publicitaires pendant le film du dimanche soir sur deux chaînes.

Les résultats obtenus sont récapitulés dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chaîne 1	35	29	30	31	33	32	33	34	32	33	32	30	28	34	33	33
chaîne 2	35	36	34	34	36	37	33	35	35	35	35	33	36	37	34	35

On désigne par X la variable : temps de coupure sur la chaîne 1 et par Y la variable : temps de coupure sur la chaîne 2.

1) Calculer la variance et l'écart-type de X et de Y . Ce calcul est à faire en utilisant les fonctions statistiques d'une calculatrice.

2) On considère le tri à plat relatif à X . (voir exercice I TP2).

x_i	28	29	30	31	32	33	34	35	
n_i	1	1	2	1	3	5	2	1	16
f_i	0,0625	0,0625	0,125	0,0625	0,1875	0,3125	0,125	0,0625	

a) Calculer la variance et l'écart-type de X en utilisant les deux premières lignes de ce tableau.

b) Calculer la variance et l'écart-type de X en utilisant la première et la troisième ligne de ce tableau.

3) Reprendre la question (2) pour la variable Y .

EXERCICE II : (On reprend l'exercice II du TD 3, du TD 4 et du TD 5)

On fait passer à 100 sujets une épreuve notée sur une échelle continue de 0 à 20. On obtient la distribution suivante :

0	0,5	1	1	1,5	2	2	2	3	3	4	4	4	4	4	4	4
4	4	4	4	4	4	4,5	4,5	4,5	4,5	4,5	4,5	4,5	5	5	5	5
5	5	5	5	5	5	5	5	5	5,5	5,5	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6,5	6,5	6,5
6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	7	7	7	7	7
7	7	7	7	7	8	8	8	8	8	8,5	8,5	8,5	9	10		

On désigne par X la variable : note obtenue à l'épreuve.

1) Calculer la variance et l'écart-type de X en utilisant une calculatrice !

2) On considère le tableau regroupant en 5 classes d'égales amplitudes les valeurs de X . (voir exercice II du TD 2).

Note	[0 2]	[2 4]	[4 6]	[6 8]	[8 10]	
n_i	8	15	42	30	5	100
f_i	0,08	0,15	0,42	0,3	0,05	

a) Calculer la variance et l'écart-type de X .

b) Comparer avec (1) et commenter.

EXERCICE III : (D'après examen septembre 99) (voir exercice III TD 5).

Pour une étude de la transmission de la pensée un parapsychologue réalise l'expérience suivante : Il place deux personnes dans deux pièces séparées, l'une des deux personnes écrit une

lettre (choisie parmi les 26 lettres de l'alphabet latin), la deuxième personne écrit alors une lettre (qui devra être la même si effectivement il y a transmission de la pensée et dans ce cas on dit qu'il y a réussite). Pour chaque couple il répète l'expérience dix fois, et il note le nombre de réussites. Il teste 24 couples classés en quatre groupes :

G_1 : Le groupe des couples homme-homme

G_2 : Le groupe des couples homme-femme

G_3 : Le groupe des couples femme-homme

Groupes	G_1	G_2	G_3	G_4
N o t e s	7	3	7	5
	5	7	7	4
	2	7	5	8
	2	1	5	4
	2	7	6	4
	8		6	
	2		6	

G₄ : Le groupe des couples femme-femme

Il obtient les résultats ci-contre :

- 1) Calculer la variance et l'écart-type des notes par groupe.
- 2) Calculer la variance inter et la variance intra. En déduire la variance globale.
- 3) Calculer directement la variance globale et comparer.

EXERCICE IV :

Quelle n'a pas été la surprise de cette institutrice qui retrouva 4 ans plus tard exactement les mêmes élèves qu'elle avaient au CP! Mais sa surprise fût encore plus grande quand elle remarqua qu'ils ont tous grandi en doublant de taille!(Est-ce bien réaliste ?). Le soir ce sujet ne manqua pas de revenir dans la discussion, et voilà que sa fille (étudiante brillante en DEUG) lui rétorqua : « mais maman c'est super! Comme ça je n'aurai pas à recalculer la moyenne et la variance des âges et des tailles cette année puisqu'on a celles de l'année du CP »

Comment va-t-elle s'y prendre ?

EXERCICE II : (On reprend l'exercice II du TD3, du TD4 et du TD5)

Le tableau suivant fournit la mesure, en dixième de mm, de la finesse de discrimination sensorielle de 105 sujets. (On pose sur la peau du sujet les deux pointes d'un compas plus ou moins rapprochés et on mesure la distance minimale perçue par le sujet).

Finesse	n_i	Finesse	n_i	Finesse	n_i	Finesse	n_i	Finesse	n_i
0,1	1	3,0	1	5,6	2	8,6	3	13,5	1
0,2	2	3,6	1	5,7	2	8,7	1	14,0	1
0,4	2	3,7	1	5,8	2	8,8	3	14,2	1
0,6	3	3,9	1	5,9	1	9,5	1	15,9	1
0,7	1	4,0	1	6,3	1	10,0	1	16,2	1
0,8	1	4,1	1	6,4	1	10,1	3	18,0	2
1,3	2	4,2	1	6,6	1	10,3	1	18,3	1
1,5	3	4,3	3	6,7	2	10,6	1	18,6	1
1,7	1	4,4	1	6,8	1	10,9	1	19,1	1
1,8	1	4,6	2	7,1	3	11,5	1	19,8	1
1,9	1	4,7	3	7,7	1	11,6	1	20,0	1
2,4	1	5,0	1	8,0	2	11,9	1	21,0	1
2,5	1	5,2	1	8,2	2	12,4	1	35,0	1
2,6	2	5,3	1	8,3	1	12,8	1	49,9	1
2,7	1	5,4	1	8,5	2	13,1	2	67,8	1

1) Si ça vous amuse calculer la variance et l'écart-type de la finesse de discrimination sensorielle des 105 sujets en utilisant les fonctions statistiques d'une calculatrice ?

2) On considère le découpage en 7 classes de même intervalle. (Voir exercice II du TP2).

Finesse	[0,1 9,8]	]9,8 19,5]	]19,5 29,2]	]29,2 38,9]	]38,9 48,6]	]48,6 58,3]	]58,3 68]	
n_i	75	24	3	1	0	1	1	105
N_i	75	99	102	103	103	104	105	

a) Calculer à partir de ce tableau la variance et l'écart-type de X. Comparer et commenter.

EXERCICE I :

- 1) De 1997 à 1998 la population d'un village a augmenté de 10%, l'année d'après elle a baissé de 10%. Retrouve-t-on le même nombre d'habitants de ce village en 1999 ?
- 2) Pendant la période des soldes un magasin offre une réduction de 30% sur tous les produits en vente. Pendant cette période de soldes ce magasin organise de façon aléatoire des « flashes » en un temps limité de 5 minutes en offrant en plus 50% de réduction exceptionnelle sur tous les produits soldés. Ces « flashes » sont évidemment accompagnés de messages publicitaires tels que « Chers clients vous avez 5 minutes pour faire des affaires en or, Nous vous offrons 50% de réduction en plus sur tous nos produits déjà soldés à 30% ! Profitez vite de ces 80% de réduction vous n'avez que ». Commenter.

EXERCICE II : D'après examen juin 00

Le Stylo de l'Homo Sapiens : Journal de l'ASSO SHS N°11 Mai 2000 page 5

Féodalité : Au Zimbabwe, 1% de la population blanche possède 70% de la terre agricole ...

Donner toutes les interprétations possibles de cette affirmation (en termes de répartition de la population et des terres). Quelle est l'interprétation voulue par l'auteur (en principe celle qui correspond à la réalité) ?

Rappelons, à tout hasard, que le Zimbabwe est un pays africain !

Expliquer à cette journaliste son erreur et dévoiler le « mystère ».

EXERCICE III :

En 1995 le taux de suicide chez les policiers a été de 50 pour 100.000. Ce taux est plus du double de taux national moyen (21,5 pour 100.000). Le risque de suicide serait donc beaucoup plus grand chez les policiers ? La presse ne manqua point de l'interpréter ainsi et de dénicher un suicide quasi quotidien chez les policiers. Le psychiatre Xavier Pommereau (Libération du 23 Avril 1996) écrit à juste titre, « *la lecture sommaire de ces chiffres suffit à démanteler l'hypothèse d'une fragilité policière spécifique en regard du risque suicidaire* ».

Sur quels arguments statistiques se base l'analyse (correcte) de ce psychiatre ?

EXERCICE IV :

Une société cherche à embaucher un nouvel ouvrier. Le directeur annonce à un candidat : ici les salaires sont très intéressants. Le salaire moyen est de 12000 F et le salaire médian est de 8000 F. Pendant la période d'essai vous toucherez 3000 F par mois mais votre salaire augmentera vite. Quelle n'a pas été la déception de ce jeune ouvrier en constatant au bout de quelques jours que les dix ouvriers touchent à peine 4000 F et ceci depuis longtemps.

La société est composée d'un directeur de son frère et de 6 membres de sa famille qui occupent des postes administratifs. Le personnel comporte 5 contremaîtres et 10 ouvriers.

- 1) Donner une répartition des salaires qui correspond bien aux dires du directeur.
- 2) Comment pouvez-vous aider ce jeune ouvrier à y voir clair ?

EXERCICE V : D'après examen septembre 02

Lors de l'avant-dernière coupe du monde de football, les femmes constituaient 48% des téléspectateurs français en demi-finale et 49,6% en finale. On sait aussi qu'il y'avait 20,6 Millions de téléspectateurs français (et françaises) qui ont regardé le match de la demi-finale (France-Croatie) et 23,5 Millions pour la finale (France-Brésil).

- 1) Calculer le nombre de téléspectatrices françaises à chacun de ces deux matchs.
- 2) Calculer la moyenne (arithmétique) de ces deux nombres
- 3) Est-il exact de dire que les femmes françaises constituaient $\left(\frac{48 + 49,6}{2}\right)\%$ des téléspectateurs français en moyenne lors de ces deux matchs ? Comment expliquez-vous cela ?

EXERCICE VI :

Cinq collèges A, B, C, D et E participent à une rencontre sportive scolaire. Chaque collège présente un groupe (que l'on désigne par la même lettre que son collège) d'élèves sportifs. Ces cinq groupes sont formés par un même nombre de participants. Une note finale sur 100 est attribuée à chaque participant. Cette note est synthétisée à partir des différentes épreuves de la rencontre. Elle mesure la performance globale du participant.

Groupes	A	B	C	D	E
Moyenne	87	75	86	87	75
Ecart-type	12	4	11	22	20

- 1) Si on vous demande de désigner les groupes qui se ressemblent le plus du point de vue de la performance lesquels choisiriez-vous ? Et pour quelle(s) raison(s) ?
- 2) Si on vous demande de désigner le groupe contenant le plus de participants performants lesquels choisiriez-vous ? Et pour quelle(s) raison(s) ?
- 3) Si on vous demande de désigner le groupe contenant les participants les plus performants lequel choisiriez-vous ? Et pour quelle(s) raison(s) ?
- 4) (facultative +1 point). Expliquer pourquoi on peut calculer la moyenne des performances des cinq groupes réunis et calculer cette moyenne globale.
- 5) (facultative +1 point). Peut-on calculer la variance des performances des cinq groupes réunis ? Si oui calculer cette variance globale sinon expliquer.

EXERCICE VII : Effet de structure

Le tableau ci-contre donne le salaire horaire moyen (moyenne arithmétique des salaires horaires) dans deux entreprises en fonction du sexe des salariés.

		Proportion	Salaire horaire moyen
Entreprise A	Hommes	75%	160
	Femmes	25%	120
Entreprise B	Hommes	40%	170
	Femmes	60%	130

- 1) Quelle est l'entreprise qui paie (en moyenne) le mieux les hommes?
- 2) Quelle est l'entreprise qui paie (en moyenne) le mieux les femmes?
- 3) Quelle conclusion (hâtive !) a-t-on envie de tirer ?
On se propose d'analyser plus finement cette situation :
- 4) Calculer le salaire moyen horaire dans chaque entreprise (indépendamment du sexe) m_A et m_B .
- 5) Quelle est l'entreprise qui paie (en moyenne) le mieux ses salariés?
- 6) Que constatez-vous ? Comment expliquer ce résultat ?
 - a) Calculer le salaire moyen horaire m'_B de l'entreprise B avec une répartition des sexes identique à celle de l'entreprise A.
 - b) Ecrire le rapport m_B / m_A en faisant intervenir m'_B .
 - c) Conclure

EXERCICE I :

On fait passer à 100 sujets une épreuve notée sur une échelle continue de 0 à 20. On obtient la distribution suivante :

0	0,5	1	1	1,5	2	2	2	3	3	4	4	4	4	4	4	4
4	4	4	4	4	4	4,5	4,5	4,5	4,5	4,5	4,5	4,5	5	5	5	5
5	5	5	5	5	5	5	5	5	5,5	5,5	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6,5	6,5	6,5
6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	6,5	7	7	7	7	7
7	7	7	7	7	8	8	8	8	8	8,5	8,5	8,5	9	10		

1) Donner le(s) mode(s), les valeurs extrêmes et l'étendue de cette série statistique.

2) Calculer la médiane ainsi que les quartiles

Box-plots (box-and-whiskers-plots)

Les *box-plots* sont des outils graphiques de représentation des données appartenant à un ensemble de méthodes regroupées sous la rubrique : **Analyse Exploratoire des Données**

Les *box-plots* sont définis pour des variables quantitatives (numériques). Ils nécessitent la détermination de la médiane Me , des quartiles Q_1 , Q_3 et de l'écart interquartiles $I_Q = Q_3 - Q_1$.

On distingue deux, voire trois, types d'observations :

- les observations situées dans l'intervalle interquartiles $[Q_3 - Q_1]$
- les données situées à l'extérieur de l'intervalle interquartile.

Parmi ces dernières on distinguera les données qui sont considérées comme « des observations extérieures » et les données qui sont considérées comme « des observations très extérieures ».

Pour préciser comment sont situées ces données extérieures et/ou très extérieures, on définit les valeurs suivantes :

Les deux limites intérieures : $\ell_G = Q_1 - (1,5 I_Q)$ et $\ell_D = Q_3 + (1,5 I_Q)$.

Les deux limites extérieures : $L_G = Q_1 - (3 I_Q)$ et $L_D = Q_3 + (3 I_Q)$.

3) Calculer les limites intérieures et extérieures des données précédentes.

On considère comme « valeurs extérieures », les valeurs situées entre les limites intérieures et les limites extérieures.

On considère comme « valeurs très extérieures », les valeurs situées au delà des limites extérieures.

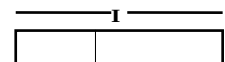
On détermine aussi les deux valeurs adjacentes A_1 et A_3 :

- A_1 est la valeur non extérieure la plus petite et la plus éloignée de Q_1 : $A_1 = \text{Min} (x_i \geq \ell_G)$

- A_3 est la valeur non extérieure la plus grande et la plus éloignée de Q_3 : $A_3 = \text{Max} (x_i \leq \ell_D)$

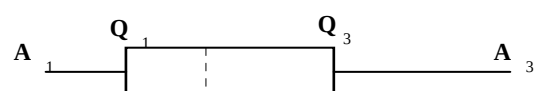
4) Calculer les valeurs adjacentes A_1 et A_3 des données précédentes.

L'ensemble des observations situées dans l'intervalle interquartile est figuré sous la forme d'une boîte dont la longueur est égale à l'écart interquartile : $I_Q = Q_3 - Q_1$.



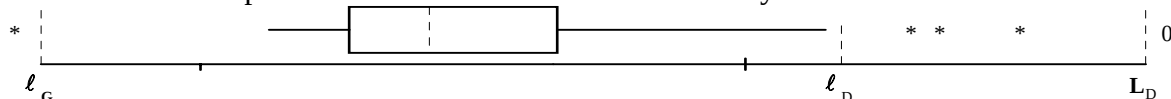
De chaque côté de la boîte, on dessine des « moustaches » (whiskers) de longueurs respectives :

$Q_1 - A_1$ et $A_3 - Q_3$.



Les valeurs extérieures et les valeurs très

extérieures sont représentées individuellement à l'aide de symboles distincts :



5) Construire le box-plot correspondant aux données précédentes. Commenter ce graphique.

Table des matières

<i>Travaux Dirigés N°1</i>	<i>1</i>
<i>Travaux Dirigés N°2</i>	<i>5</i>
<i>Travaux Dirigés N°3</i>	<i>8</i>
<i>Travaux Dirigés N°4</i>	<i>10</i>
<i>Travaux Dirigés N°5</i>	<i>13</i>
<i>Hors-série N°1</i>	<i>16</i>
<i>Hors-série N°2</i>	<i>18</i>