

Préambule

Evaluation du niveau de compétences en statistique.

I – Notions de base

Révision et précision (théorie & tutorial) :

- La notion de variable dépendante, indépendante et contrôle dans un plan expérimental.
- La notion d'hypothèse opérationnelle et théorique

Exercice I.1 :

En général, le temps mis pour détecter un stimulus à valence négative parmi des stimuli neutres est plus court que le temps mis pour détecter un stimulus positif parmi des stimuli neutres. Pour savoir si ces résultats se manifestent avec des visages, Arne Öhman, professeur de psychologie dans le département de neurosciences cliniques de l'institut Karolinska de Stockholm, a construit un ensemble de visages schématiques. Deux ensembles de tels stimuli sont présentés sur la figure 1 : un visage négatif à détecter au milieu de visages neutres (à droite) et un visage positif à détecter au milieu de visages neutres (à gauche).

De plus, si les visages fournissent des informations nécessaires à l'identification d'une personne, ils nous permettent, grâce à l'analyse des expressions faciales, de percevoir dans quel état se trouve un individu et, par conséquent, de savoir quel type d'interaction nous pouvons avoir avec lui.

Par ailleurs, l'auteur fait l'hypothèse théorique que les personnes mal à l'aise en société (qui ont la phobie de leurs pairs) sont beaucoup plus sensibles à la présence d'individus menaçants que des sujets ne présentant pas ce trouble. Les résultats obtenus avec un groupe de sujets phobiques et un groupe contrôle sont indiqués dans la figure 2.

FIGURE 1

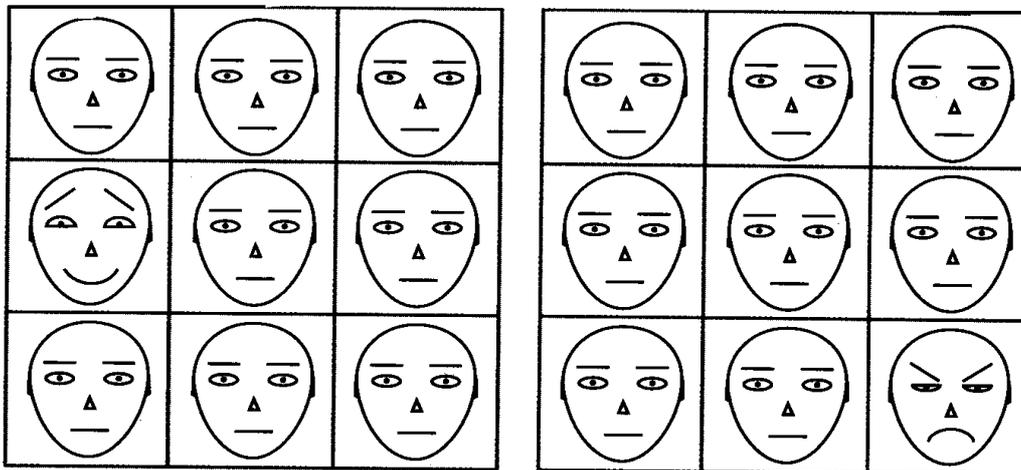


FIGURE 2

	Visage négatif	Visage positif
Phobique	1400	1650
Contrôle	1500	1600

Déterminer:

- 1) la problématique

- 2) l'hypothèse théorique
- 3) les VI, VD, et VC. Précisez si aléatoire, discrète, continue, quantitative, qualitative.
- 4) les hypothèses opérationnelles
- 5) Représentation graphique
- 6) Les résultats sont-ils conformes aux hypothèses ?

Exercice I.2 :

Nathalie Bedoin en 1995 a réalisé une étude portant sur les mécanismes cognitifs permettant l'accès au sens d'un mot écrit et plus particulièrement sur le rôle de la phonologie durant une lecture silencieuse. Selon l'auteur, la représentation phonologique d'un mot interviendrait dans l'accès au sens du mot. Pour cela, 19 étudiants, tous bons lecteurs, de langue maternelle française et dotés d'une vue normale ont été testés individuellement. La procédure était la suivante : un point de fixation apparaissait au centre de l'écran suivi d'un premier mot pendant 500 ms, immédiatement remplacé par un second. Les sujets avaient pour consigne de porter leur attention sur le premier mot et de décider le plus rapidement et le plus précisément possible si le second mot appartenait à la langue française ou non (tâche de décision lexicale). Le matériel était constitué de trois listes de mots construites de la manière suivante :

- Une première série de couples de mots reliés par un lien sémantique (ex : mer-océan)
- Une seconde série de couples de mots où le premier mot de chaque couple de la première série est remplacé par un homophone (ex : maire-océan)
- Une dernière série où les premiers et seconds mots ne partageaient aucun lien (lapin-océan).

Les résultats sont représentés dans le tableau ci-dessous :

Type de lien entre le premier et le second mot	Temps de réponse à la tâche de décision lexicale exprimés en ms
Sémantique	618
Homophone	620
Aucun lien	662

Déterminer:

- 1) la problématique
- 2) l'hypothèse théorique
- 3) les VI, VD, et VC. Précisez si aléatoire, discrète, continue, quantitative, qualitative
- 4) les hypothèses opérationnelles
- 5) Représentation graphique
- 6) Les résultats sont-ils conformes aux hypothèses ?

Exercice I.3 :

Il a été proposé que le contexte dans lequel de nouvelles connaissances sont apprises influence le rappel (rappeler l'item appris en l'absence d'indices) et la reconnaissance (reconnaître l'item appris au sein d'une liste). Il serait plus facile de rappeler ou de reconnaître l'item dans un contexte identique à celui qui existait lors de son apprentissage que dans un contexte différent, par exemple dans la même pièce en comparaison d'une pièce différente. Dans leur étude, Godden et Baddeley (1975) ont cherché à mettre cet effet en évidence en utilisant des contextes très différents, soit à l'air libre soit en plongée sous-marine.

Deux groupes de 40 plongeurs professionnels ont participé à l'expérience. Chacun devait apprendre une liste de 20 mots ; le premier devait rappeler les mots (tâche de rappel libre). Le second devait décider si des mots présentés appartenaient à la liste apprise précédemment (tâche de reconnaissance). Dans chaque groupe, 20 sujets apprenaient sur la terre ferme et 20

en plongée, et dans chaque sous-groupe, 10 sujets réalisaient la tâche (rappel ou reconnaissance) sur la terre ferme et 10 en plongée.

Déterminer:

- 1) la problématique
- 2) l'hypothèse théorique
- 3) les VI, VD, et VC. Précisez si aléatoire, discrète, continue, quantitative, qualitative
- 4) les hypothèses opérationnelles

Exercice I.4 :

De nombreuses recherches ont montré que les performances des personnes âgées peuvent différer de celles des jeunes en fonction du type de tâche de mémoire à réaliser. Certains chercheurs ont ainsi étudié le nombre de mots rappelés et reconnus par deux types de populations (jeune et âgée). Les sujets devaient apprendre une liste de 30 mots. Il leur était demandé dans une première tâche de rappeler le maximum de mots mémorisés (rappel libre). Dans une seconde tâche, une nouvelle liste de mots était présentée contenant des mots anciens (appartenant à l'ancienne liste) ou des mots nouveaux, les participants devant déterminer si chacun des mots étaient anciens ou nouveaux (tâche de reconnaissance). Les chercheurs s'attendaient à une différence de performances entre les deux populations pour la tâche de rappel libre et à aucune différence pour la tâche de reconnaissance.

Le tableau ci-dessous présente le nombre de mots rappelés ou correctement reconnus sur les 30 présentés :

	<i>Rappel libre</i>	<i>Reconnaissance</i>
<i>Population jeune</i>	14	28
<i>Population âgée</i>	5	28

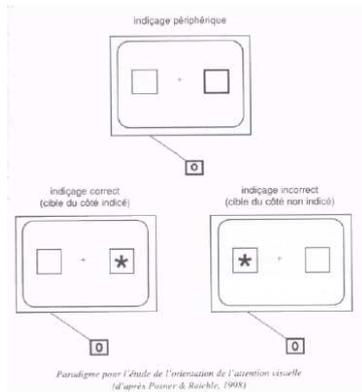
Déterminer:

- 1) la problématique
- 2) l'hypothèse théorique
- 3) les VI, VD, et VC. Précisez si aléatoire, discrète, continue, quantitative, qualitative
- 4) les hypothèses opérationnelles
- 5) Représentation graphique
- 6) Les résultats sont-ils conformes aux hypothèses ?

Exercice I.5 :

Énoncé

Des lésions des aires pariétales de l'hémisphère droit chez l'homme entraînent une hémiparésie gauche et des difficultés d'orientation de l'attention visuelle. Selon Posner, cette orientation dépend d'une séquence d'opérations mentales élémentaires qui comprend le désengagement du foyer attentionnel actuel, le déplacement de l'attention vers la région indiquée et enfin, l'engagement de l'attention sur la cible. Le paradigme de détection avec indicage mis au point par Posner a permis de préciser quelles sont les opérations perturbées dans l'hémiparésie. La tâche du sujet est de répondre le plus rapidement possible à l'apparition d'une cible qui lui a été préalablement spécifiée.



		Cible	
		controlatérale	ipsilatérale
Indiçage	correct	530	470
	incorrect	1010	610

Dans la condition d'indiçage attentionnel, un indice précède l'apparition de la cible. Lorsque l'indice indique correctement la position la plus probable d'apparition de la cible (indiçage correct), les réponses aux cibles ipsi et controlatérales à la lésion ne diffèrent pas. Posner déduit de cette première donnée que les patients atteints de lésions pariétales droites (présentant une héminegligence gauche) conservent la capacité à focaliser leur attention sur une cible lorsque l'attention a été correctement orientée. Par contre, un indiçage incorrect (cible et indice apparaissent du côté opposé) occasionne des difficultés sévères surtout si l'indice apparaît du côté ipsilatéral à la lésion et la cible du côté controlatéral. En effet, dans ce cas, soit la cible n'est pas détectée, soit elle l'est mais avec un temps de réaction très long.

Déterminer:

- 1) la problématique
- 2) l'hypothèse théorique
- 3) les VI, VD, et VC. Précisez si aléatoire, discrète, continue, quantitative, qualitative
- 4) les hypothèses opérationnelles

II – Echelle de mesures

Révision et précision (théorie & tutorial) :

- Les différents types d'échelles. Spécificité et intérêts.

Exercice II.1 :

1. Inventer cinq échelles nominales et décrivez leurs modalités.
2. Inventer cinq échelles ordinales et décrivez leurs modalités.
3. Inventer cinq échelles numériques (intervalle ou de rapport) et décrivez leurs modalités.

Exercice II.2 :

La propriété d'une variable est qu'elle prend plusieurs formes ou éventualités. L'échelle de mesure exprime le type de relation qui existe entre ces éventualités.

Variable	n°1	n°2	n°3	n°4	n°5	n°6
Nom	âge	sexe	poids	profession	pratique le squash	goût pour les maths
Eventualité 1	27	F	62	ingénieur	jamais	beaucoup
Eventualité 2	38	H	78	artisan	souvent	pas du tout

Tableau 1.

Trois types de relation existent entre les éventualités d'une variable, ce qui correspond à trois échelles de mesure. Ces trois types sont présentés dans le tableau 2. Complétez le tableau 2 à partir du tableau 1.

Type de relation	n° de variable (voir le tableau 1)
Les éventualités ne peuvent pas être ordonnées	
Les éventualités peuvent être ordonnées mais l'écart entre chaque éventualité ne peut pas être quantifié	
Les éventualités peuvent être ordonnées et l'écart entre chaque éventualité peut être quantifié	

Tableau 2.

Mise en garde

Les valeurs chiffrées ne représentent pas toujours une valeur numérique. S'il est possible de coter 1 pour « femme » et 2 pour « homme », ces valeurs ne transformeront pas pour autant l'échelle nominale de la variable « sexe » en échelle numérique.

Exercice II.3 :

Il est possible de passer d'une échelle numérique à une échelle ordinale en regroupant les valeurs numériques en classes. Effectuez ce type de regroupement dans le tableau 3 (dans la ligne votre regroupement) et donnez un nom à chacune de ces classes.

Individu	n°1	n°2	n°3	n°4	n°5	n°6	n°7	n°8	n°9	n°10
Age	6	8	10	16	20	24	32	40	54	70
Votre regroupement										
Autre possibilité										
Autre possibilité										

III – Analyses univariées

3.1. Organisation des données

Révision et précision (théorie & tutorial) :

- Comment organiser ces données ? Quelle utilité ?
- Quel graphique choisir en fonction de l'échelle utilisée ?
- Comment faire un bon graphique ? Les paramètres importants
- Quelle est la différence entre un diagramme en bâton et un histogramme ?
- Mise en garde : la mauvaise utilisation de représentations graphiques
- Comment tracer un polygone des fréquences ou des effectifs ?
- Comment tracer un graphique circulaire ou en tronçon ?

Exercice III.1.1 :

Trente personnes âgées résidents dans une maison de retraite ont passé le questionnaire de dépression de Beck (BDI II). Les scores obtenus étaient les suivants :
2,30,20,3,48,57,48,39,20,49,13,12,20,37,10,8,0,3,8,19,23,25,26,28,11,40,38,27,16,58

1. Ordonner les données et calculer les fréquences relatives.

Indice : Les personnes ayant aucune dépression ont un score inférieur à 10, une dépression légère est coté entre 10 et 18, une dépression modérée entre 19 et 29, une dépression sévère est évalué à un score supérieure à 30.

2. Représenter les données sous la forme d'un diagramme des fréquences et d'un diagramme circulaire.

Exercice III.1.2 :

Répartition du nombre de patient d'un psychologue selon le type de phobies.

Type de phobies	Nb de patients	Fréquence relative	%age	Effectif cumulé	fréquence cumulée
Agoraphobie	18				
Claustrophobie	37				
Éreuthophobie (crainte de rougir)	4				
Hypocondrie	16				
Nosophobie	12				
Total	87				

1. Quel est le type d'échelle utilisé pour cette variable ?
2. Calculez les fréquences relatives et les pourcentages.
3. Calculez l'effectif cumulé et les fréquences cumulées.
4. Faire un diagramme en barre, un diagramme en tronçon, un diagramme circulaire.

Exercice III.1.3 :

Dispositif d'évaluation d'un plan de formation : évaluation du degré de satisfaction des stagiaires ayant effectués le stage. Quelle appréciation porteriez-vous à l'égard de la formation que vous venez d'achever?

Pas du tout satisfait	1	2	3	4	5	6	7	très satisfait
-----------------------	---	---	---	---	---	---	---	----------------

Effectifs de chaque catégorie de réponse (sur 85 stagiaires)

Degré de l'échelle	1	2	3	4	5	6	7
Effectif	2	4	7	10	16	34	12

1. Quel est le type d'échelle utilisé pour cette variable ?
2. Faire un graphique de fréquence et un graphique de fréquence cumulée.

Exercice III.1.4 :

Une étude clinique a évalué le poids de personnes souffrant d'anorexie avant et après un traitement thérapeutique. Cette étude a étudié deux groupes : un groupe expérimentale et un groupe contrôle.

Groupe contrôle :

Reprise de poids entre la mesure pré et post-traitement	Eff	Centre	fréquence	cumul Eff	cumul fré
[-1;-0,5[2				
[-0,5;0[3				
[0;0,5[4				
[0,5;1[5				
[1;1,5[8				
[1,5;2[3				

total

Groupe traitement :

Reprise de poids entre la mesure pré et post-traitement	Eff	fréquence	cumul Eff	cumul fré
moins d'un kilo	3			
[0;1]	3			
[1;2]	4			
[2;3]	10			
[3;4]	4			
plus de quatre kilos	1			

1. Quel sont les types d'échelle utilisés pour cette étude ?
2. Calculez les fréquences relatives et les effectifs cumulés.
3. Faire une représentation graphique de ces deux distributions.

3.2. Tendances centrale et dispersion

Révision et précision (théorie & tutorial) :

- Exemples de l'utilité des paramètres de positions et de dispersion
- Quel paramètre de position ou de dispersion choisir en fonction des échelles utilisées ?
- Calcul des paramètres
- Deux manières de calculer l'écart-type
- Comment calculer un quantile de manière précise ?

Exercice III.2.1.

Pour chacun des énoncés suivants calculez les paramètres de position pertinents.

Énoncé 1

On présente à 16 sujets un portrait comportemental d'une personne fictive, composé de 18 comportements. Les sujets sont invités à lire ce portrait puis, après une tâche distractive, on leur demande de se rappeler du plus grand nombre possible de comportements lus.

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Comportements rappelés	8	7	5	9	8	7	10	6	9	7	8	10	11	9	8	8

Énoncé 2

Lors d'une recherche portant sur l'étude des processus de décision chez des jurys de tribunaux, on a demandé à des sujets d'évaluer la sévérité d'une décision de justice dans un contexte donné à l'aide d'une échelle en sept points allant de la valeur -3 (mentionnée sévère) à +3 (mentionnée indulgente).

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13
Évaluation	2	2	1	0	3	2	1	0	-1	3	2	1	2

Énoncé 3

Dans une recherche portant sur le codage visuel et la mémoire, deux chercheurs¹ ont présenté des couples de lettres (AA) en faisant varier l'intervalle de temps séparant la présentation entre la première et la dernière lettre de chacun de ces couples. Selon les groupes, cet intervalle était de 0, ou 2 secondes. Les sujets devaient dire, le plus vite possible, après la présentation de la seconde lettre, si celle-ci était la même que la première lettre présentée. Les tableaux ci-dessous indiquent les temps (ms) de chaque sujet pour chaque intervalle.

Intervalle de 0 seconde

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Tps (ms)	403	385	379	390	368	387	422	435	418	312	378	482	353	510	406	342	419	309	451	312

Sujet	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Tps	408	502	374	328	338	421	483	348	328	422	399	386	354	378	396	371	384	429	467	356

Intervalle de 2 secondes

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Tps	582	421	410	514	471	610	469	536	561	513	493	514	498	501	585	463	425	520	533	487

Sujet	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Tps	496	502	563	476	508	526	587	489	508	511	542	561	541	477	518	535	486	487	528	538

Exercice III.2.2 :

En vue de réaliser un programme de rééducation, une épreuve de neuropsychologie cognitive a été proposée à des enfants dyslexiques. On recueille les données de 150 enfants selon leur nombre de bonnes réponses. Les résultats ainsi récoltés sont les suivants :

Nombre de bonnes réponses	Effectif
[1;5[34
[5;9[38
[9;13[31
[13;17[25
[17;21[22

1. Déterminer la nature de la variable étudiée
2. Déterminer la classe modale de cette série, l'étendue et l'écart-moyen
3. Quel est le score moyen? Et quel est l'écart type?

Interprétez ces indices

4. Calculer la médiane et le troisième quartile
5. Représenter ces données en choisissant la représentation graphique la mieux adaptée.

¹ POSNER & KEELE, « Decay of Visual Information from a Letter », *Science*, 1967, 158, p 137-139

Exercice III.2.3 :

Voici la répartition d'un ensemble d'élèves selon leur taille en mètre :

x_i	Effectif (n_i)
[1.615;1.645[121
[1.645;1.675[287
[1.675;1.705[258
[1.705;1.735[133
[1.735;1.765[78

Déterminer :

1. La nature de la variable
2. La classe modale puis le mode, l'étendue et l'écart-moyen
3. La classe médiane puis la médiane
4. La moyenne
5. Le 1er quartile
6. L'écart type
7. Vous représenterez ces données en choisissant la représentation graphique la mieux adaptée.
8. Donner une approximation de la taille du 800ème individu si on les classait par ordre croissant de taille.
9. On classe ces individus par ordre croissant et on ne considère que les 700 plus petits. Quelle est la taille moyenne de ces individus?

Exercice III.2.4 :

Dans le cadre d'une étude, vous avez demandé à des patients d'auto-évaluer leur anxiété sur une échelle notée de 0 à 70. La répartition est la suivante :

Notes en classes	Effectifs
de 0 à 10	4
de 10 à 20	13
de 20 à 30	19
de 30 à 40	28
de 40 à 50	17
de 50 à 60	11
de 60 à 70	5

1. Présentez sous forme de tableau : Les effectifs simples et cumulés
Les fréquences simples et cumulées
2. Définissez et calculez :
L'étendue et l'écart-moyen de la distribution
Les valeurs centrales des classes
La classe modale
La médiane et l'intervalle inter-quartiles
La moyenne, la variance
3. Représentez graphiquement les fréquences.
4. Commentez les indices et la forme de la distribution.

Exercice III.2.5 :

30 élèves ont passé un test noté sur 10. Voici le résultat de ce test.

1. Compléter le tableau ci-dessous.

Notes	0	1	2	3	4	5	6	7	8	9	10
Effectifs	0	0	3	1	3	6	7	5	3	1	1
Effectifs cumulés											

2. Prévoir la médiane de cette série.

3. Tracer le polygone des effectifs cumulés.

4. Comment peut-on retrouver cette médiane sur le graphique ?

Exercice III.2.6 :

184 étudiants devaient décrire leur personnalité. Plusieurs traits de personnalité leur étaient proposés et, pour chacun d'eux, ils devaient se positionner sur une échelle allant de 0 (je ne suis pas du tout comme ça) à 9 (je suis tout à fait comme ça). L'une de ces échelles concernait le trait « audacieux », l'autre le trait « ambitieux ». Les résultats sont reportés ci-dessous :

X_i	0	1	2	3	4	5	6	7	8	9	
Effectifs pour « audacieux »	0	6	12	22	30	48	26	22	13	5	N = 184
Effectifs pour « ambitieux »	3	3	3	10	18	36	31	34	23	23	N = 184

a) Faites une représentation graphique des distributions pour chacun des traits.

b) pour chaque trait, déterminez le mode, les trois quartiles, les 4^o et 7^o déciles.

c) Calculez la moyenne et l'écart-type.

d) Calculez l'étendue et l'écart-moyen de chaque modalité.

Exercice III.2.7.

a) Représentez graphiquement la distribution des notes de l'échantillon suivant (histogramme des effectifs et polygone des fréquences).

Notes x_i	Effectifs n_i
[0 ; 2[0
[2 ; 4[1
[4 ; 6[2
[6 ; 8[9
[8 ; 10[15
[10 ; 12[8
[12 ; 14[3
[14 ; 16[0
[16 ; 18[1
[18 ; 20[1

- b) Calculez la moyenne, l'écart-type, l'étendue et l'écart-moyen des notes.
 c) déterminez la classe modale, la médiane, ainsi que Q1 et Q3 (par approximation puis précisément).

Exercice III.2.8

150 personnes sont interrogées sur la distance qui sépare leur lieu d'habitation de leur lieu de travail. Voici comment elles se répartissent selon ce critère :

Nombre de km	[0, 2[[2,4[[4,8[[6,8[[8,10[[10,12[[12,14[[14,16[[16,18[[18,20[[20,22[[22,24[
Nombre de sujet	22	20	9	8	4	0	2	15	24	28	14	4

- 1/ Définissez le type d'échelle de mesure de cette étude
- 2/ Dites quel(s) indice(s) de tendance centrale et de dispersion peuvent être employé(s).
- 3/ Calculez ces indices.
- 4/ Représentez graphiquement les données en considérant que toutes les distances supérieures à 20 km sont regroupées dans une seule classe « très longue distance ».

Exercice III.2.9 :

On fait passer une épreuve testant les compétences d'orientation spatiale à tous les enfants des classes de 6^{ème} d'un collège. La note minimale de ce test est 0 et la note maximale est 100. Voici comment se répartissent les résultats.

note	effectif
[0 ;10[4
[10 ;20[11
[20 ;30[52
[30 ;40[38
[40 ;50[68
[50 ;60[140
[60 ;70[0
[70 ;80[18
[80 ;90[15
[90 ;100[4

- 1) Définissez le type de mesure utilisée.
- 2) Calculez la note moyenne obtenue par les enfants.
- 3) Calculez l'écart-type de cette distribution de notes et interprétez-le.

Exercice III.2.10 :

Dans le cadre d'une étude, on a demandé à des patients d'auto-évaluer leur anxiété en s'attribuant une note (de 0 à 70). La répartition de leur réponse est la suivante (données fictives):

Notes	Effectifs
[0;10[4
[10;20[13
[20;30[19
[30;40[28
[40;50[17
[50;60[11
[60;70[5

- 1/ Définissez la nature de la variable
- 2/ Calculez : la classe modale
la médiane et l'intervalle inter-quartiles
la moyenne et l'écart-type
- 3/ Présentez les fréquences sous forme de tableau
- 4/ Représentez graphiquement les fréquences en choisissant le type de graphique le plus approprié aux données.
- 5/ Commentez les indices et la forme de la distribution.

Exercice III.2.11 :

En vue de réaliser un programme de rééducation, une épreuve de neuropsychologie cognitive a été proposée à des enfants dyslexiques. On recueille les données de 150 enfants selon leur nombre de bonnes réponses. Les résultats ainsi récoltés sont les suivants :

Nombre de bonnes réponses	Effectif
[1;5[34
[5;9[38
[9;13[31
[13;17[25
[17;21[22

6. Déterminer la nature de la variable étudiée
7. Déterminer la classe modale de cette série
8. Quel est le score moyen? Et quel est l'écart-type? Interprétez ces indices
9. Calculer la médiane et le troisième quartile (par approximation)
10. Représenter ces données en choisissant la représentation graphique la mieux adaptée.

Exercice III.2.12 :

Soit la série statistique suivante :

Intervalles x_i	Effectifs n_i
[0 ; 5 [6
[5 ; 10 [8
[10 ; 15 [14
[15 ; 20 [8
[20 ; 25 [9
[25 ; 30 [13
[30 ; 35 [0
[35 ; 40 [15
[40 ; 45 [24
[45 ; 50 [6

1. Déterminer le mode de cette distribution.
2. Calculez la moyenne et l'écart type de cette série statistique.
3. Déterminez par approximation puis par calcul précis la valeur de la médiane de cette distribution. Donnez la signification de cette valeur.
4. Représentez ces données sous la forme d'un histogramme.

Exercice III.2.13 :

Un test permettant d'évaluer le niveau d'anxiété a été étalonné sur une population de 500 personnes. Les résultats sont les suivants :

Score au test d'anxiété	Effectif
[0 ; 4[13
[4 ; 8[22
[8 ; 12[39
[12 ; 16[75
[16 ; 20[94
[20 ; 24[100
[24 ; 28[80
[28 ; 32[41
[32 ; 36[26
[36 ; 40[10

1. Déterminez la nature de la variable étudiée (justifiez votre réponse).
2. Définissez puis calculez le mode.
3. Calculez le score d'anxiété moyen puis l'écart-type de cette distribution.
4. Déterminez par approximation puis par calcul précis la valeur du troisième quartile. Quelle

Exercice III.2.14 :

Le questionnaire des peurs de Marks et Mathews a pour objectif d'évaluer brièvement les problèmes phobiques les plus fréquemment rencontrés en clinique. Nous pouvons déduire de ce questionnaire 3 scores, un score d'agoraphobie et un score de phobie sociale. Ce questionnaire a été testé en français (Cottraux et coll, 1987, Bouvard, 1986) :

		moyenne	écart-type	1er quartile	Médiane	3eme quartile
score d'agoraphobie	sujet AGO	27,47	9,35	22	28	34
	sujet PH SOC	10,5	7,52	5	9	15
	Sujet contrôle	4,58	6,42	0	2	7
score de phobie sociale	sujet AGO	23,32	6,64	19	22	29
	sujet PH SOC	14,91	9,07	7	14	18
	Sujet contrôle	7,27	6,11	2	5	10

Tiré de Bouvard & Cottraux (2005)

Nous avons fait passer ce questionnaire à trois patients HM, PJ, JK (résultats fictifs). Au regard, des moyennes et écart-types de la population, pouvez-vous dire pour chaque patient si son profil est plus celui d'un patient agoraphobique, phobique social ou normal ?

	Score Agora	Score PhSoc
Cas HM	12	15
Cas PJ	19	29
Cas JK	10	3

Indice : aidez-vous des valeurs centrées réduites !

Exercice III.2.15 :

Deux patients ont passé 3 tests, le questionnaire de Liebowitch pour la phobie sociale (moyenne=67,2, écart-type=27,5), l'inventaire de Padoue pour les TOC (moyenne= 83,6; écart-type=34,8) et le PCL-S pour le stress post-traumatique (moyenne= 60,18; écart-type=9,4). Ces questionnaires n'ont pas la même échelle : Liebowitch est scoré sur 72; le Padoue est scoré sur 240 et le PCL-S est scoré sur 85. Voici les résultats des deux patients :

		Liebowitch	Padoue	PCL-S
		sur 72	sur 240	sur 85
Scores obtenu	patient A	10	120	70
	patient B	50	50	14

Comparer chacun des scores aux trois tests de chacun de ces patients.

IV – Analyses bivariées**4.1. Organisation des données****Exercice IV.1.1 :**

Une entreprise s'est informatisée. Elle comprend deux services, un seul ayant suivi une formation avant l'installation du matériel. Le responsable des ressources humaines souhaite faire le bilan après 3 semaines d'utilisation. Il interroge donc le personnel de l'entreprise : 44 personnes ont suivi la formation et 52 personnes n'ont pas suivi la formation. Une des questions proposait un choix (et un seul) entre 4 possibilités :

P1 : l'ordinateur bouleverse ma façon de travailler

P2 : l'ordinateur améliore la qualité de mon travail

P3 : l'ordinateur augmente la rapidité de mon travail

P4 : l'ordinateur rend mon travail plus agréable

	Bouleverse ma façon de travailler	Améliore la qualité de mon travail	Augmente la rapidité de mon travail	Rend mon travail plus agréable
Formation	5	10	9	20
Sans formation	10	18	15	9

- 1) Parmi les personnes interrogées, quel est le pourcentage de personnes qui n'ont pas suivi de formation et qui estime que l'ordinateur a bouleversé leur façon de travailler ?
- 2) Parmi les personnes estimant que l'ordinateur rend leur travail plus agréable, quel est le pourcentage n'ayant pas suivi de formation ?
- 3) Parmi les personnes sans formation, quel est le pourcentage de celles qui estiment que l'ordinateur a amélioré la qualité ou la rapidité de leur travail ?
- 4) Comparer ce dernier pourcentage au pourcentage des personnes estimant que l'ordinateur a amélioré la qualité ou la rapidité du travail parmi celles qui ont suivi une formation.

Exercice IV.1.2 :

Des femmes et des hommes ont été questionnés sur ce qui comptait le plus dans leur vie. Voici la répartition de l'échantillon interviewé (résultats fictifs):

	Travail		Famille		Loisirs	
	Classique	Associatif	Conjoint	Enfant	Individuels	Collectifs
Femme	30	15	20	12	20	38
Homme	90	3	18	3	27	21

- 1) Parmi les personnes interviewés, quelle est la proportion de ceux pour lesquels les enfants sont le plus important ? Le travail classique ?
- 2) Quelle est la proportion de femmes interviewées dont le conjoint est important ? Quelle est la proportion d'hommes ?
- 3) Quelle est la proportion de femmes parmi les personnes interviewées qui ont répondu que les loisirs individuels étaient le plus important ?

Exercice IV.1.3 :

La diversité des échelles de mesure conduit à des analyses statistiques différentes et à des fins différentes. Quatre variables sont présentées dans le tableau ci-dessous.

Individu	n°1	n°2	n°3	n°4	n°5	n°6	n°7	n°8	n°9	n°10	n°11	n°12	n°13	n°14
Note en math	6	8	8	12	15	18	5	4	3	20	7	2	17	19
Poids	65	65	66	69	73	75	64	62	63	80	65	60	75	77
Sexe	F	F	F	H	F	H	H	F	F	H	F	H	H	H
Pratique le foot	souv.	jam.	jam.	jam.	jam.	souv.	souv.	jam.	souv.	souv.	jam.	souv.	souv.	souv.

Tableau (souv. = souvent ; jam. = jamais).

- 1). Faites le nuage de point du croisement de deux variables d'intervalles. A votre avis, les deux variables sont-elles corrélées ?
- 2) Le croisement de deux variables non numérique permet de comparer des fréquences. Inscrive les effectifs correspondant au croisement des deux variables non numériques dans le tableau d'effectifs suivant. Que peut-on en conclure ?

Tableau des effectifs.

- 3) Représenté graphiquement le rapport entre les variable « poids » avec la variable « sexe ». A votre avis, quel type de représentation est ici nécessaire ?

Exercice IV.1.4 :

A partir du tableau ci-dessous, tracez le tableau des fréquences conjointes et les deux tableaux des fréquences conditionnelles. Faire différents graphiques.

Effectif	introvertis	extravertis	neutre
moins de 20 ans	54	34	19
Entre 20 et 40 ans	34	12	39
plus de 40 ans	45	15	98

4.2. Relations entre deux séries

Révision et précision (théorie & tutorial) :

Relation entre 2 V nominales: chi deux (exo2.1 à 2.4)

Relation entre 2 V ordinales: Corrélation Rang de Spearman (exo 2.5 à 2.9)

Relation entre 2 V numérique: Covariance / Corrélation linéaire : R de Bravais Pearson (exo 2.10 à 2.15)

Nuage de points

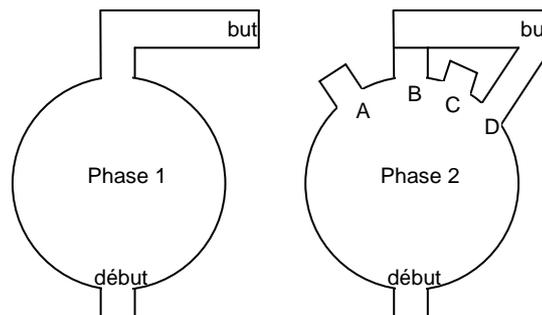
Courbe d'ajustement / droite de régression

Comment classer des données (la notion de rang et ces avantages)

Variable Xi	Position in the descending order	Rank xi
0.8	5	5
1.2	4	(3+4)/2=3.5
1.2	3	(3+4)/2=3.5
2.3	2	2
18	1	1

Exercice IV.2.1 : (tiré de Howell)

Tolman, Ritchie et Kalish (1946) ont mené une expérience pour déterminer si les rats étaient capables d'avoir une représentation mentale de l'espace. A l'époque, Hull pensait que les rats placés dans un labyrinthe apprenaient un ensemble discret de réponse motrice ; selon Tolman, les rats créaient une carte cognitive. L'enjeu était la question fondamentale de savoir si l'apprentissage des animaux repose sur une conception stimulus-réponse ou par une interprétation cognitive. Dans leur expérience, Tolman et al. ont appris à des rats à descendre l'allée de départ d'un labyrinthe pour aboutir dans une zone circulaire, une autre allée sortait tout droit par rapport à l'aller de départ puis tournait à droite pour aboutir vers un but à atteindre (nourriture). Dans une deuxième phase, d'autres allées étaient mises à disposition, deux sans issues, une autre qui pointait directement vers le but (D sur la figure), et l'allée d'origine était condamnée. Les rats étaient donc confrontés à différent choix.



Si Hull avait raison alors les rats continueraient de prendre la même allée dans les phases 1 et 2. Si Tolman avait raison alors les rats prendraient directement le raccourci. 32 rats ont réalisés cette expérience. Voici leurs réponses :

allée choisie	A	B	C	D
Effectif	4	5	8	15

Comment savoir si ces résultats sont le fruit du hasard ou si Tolman avait raison ?

Exercice IV.2.2 : (tiré de Howell)

Pugh (1983) a mené une étude sur la manière dont les jurys prennent une décision dans des cas de viols. Il a étudié parmi plusieurs variable : (1) Le prévenu a-t-il été reconnu innocent

ou coupable ? (2) La défense a-t-elle soutenu que la victime était partiellement responsable du crime ? Les données réelles de Pugh sont les suivantes :

décisions prises ds des cas de viol		verdict		<i>total</i>
		coupable	non coupable	
faute	légère	153	24	
	grave	105	76	
	<i>total</i>			

Ce tableau tend à prouver que les juges basent partiellement leur décision quant à la culpabilité de l'accusé sur les fautes perçues dans le chef de la victime. Notons que quand on considère que la victime a commis une faute légère, le prévenu est reconnu coupable dans 86% des cas environ (153/177). Par contre, quand on considère que la victime a commis une faute grave, l'accusé n'est reconnu coupable que dans 58% des cas (105/181). Cette observation n'est-elle pas le fruit du hasard ?

Exercice IV.2.3 : (tiré de Howell)

Geller, Witmer et Orebaugh (1976) étudiaient la tendance à jeter des dépliant sur la voie publique et voulaient savoir s'il serait efficace d'inclure le message « Ne pas jeter sur la voie public » dans un dépliant de promotion distribué à l'entrée d'un supermarché. Nous résumerons leurs études comme ceci : le soir, Geller et al. ont compté le nombre de dépliant retrouvé par terre (jeté), dans les poubelles (poubelles) ou pris par les clients (enlevé). Voici leurs résultats :

	poubelle	jeté	enlevé
sans message	41	385	477
avec message	80	290	499

Quelle conclusion tirer de cette étude ?

Exercice IV.2.4 : (tiré de Howell)

Fumer porte-t-il atteinte à la fertilité ? Weinberg et Gladen (1986) ont étudié la relation entre le fait de fumer et la facilité avec laquelle les femmes deviennent enceintes. Ces chercheurs ont questionné 586 femmes qui envisageaient une grossesse pour connaître le nombre de cycles menstruels qu'il leur a fallu pour être enceinte après avoir interrompu la contraception.

Smoke	Cycle	Freq
1	1	29
1	2	16
1	3	55
2	1	198
2	2	107
2	3	181

1: fumeuse
2: non fumeuse
1: 1 cycle
2: 2 cycles
3: 3+ cycles

Fumer porte-t-il atteinte à la fertilité ?

Exercice IV.2.5 :

Lors de la passation d'un questionnaire, 10 participants ont répondu aux deux questions suivantes :

Je suis une personne qui s'inquiète pour un rien :

Pas du tout 0 _____ 1 _____ 2 _____ 3 _____ 4 _____ 5 _____ Complètement

Je suis quelqu'un qui est hyperactif :

Pas du tout 0 _____ 1 _____ 2 _____ 3 _____ 4 _____ 5 _____ Complètement

Leurs résultats (fictifs) ont été compilés dans le tableau suivant :

Participant	1	2	3	4	5	6	7	8	9	10
Stress	0	5	2	5	0	0	4	3	2	4
Hyperactivité	1	4	2	3	0	2	5	4	1	4

D'après cet échantillon existe-t-il un lien entre le degré d'inquiétude et l'hyperactivité ? Représenter graphiquement les données.

Exercice IV.2.6 : (tiré de wikipédia.org)

Existe un lien entre le nombre d'heures passées à regarder la télévision et le quotient intellectuel ? Dix personnes ont passé un questionnaire de QI et ont donné le nombre d'heures moyen hebdomadaire à regarder la télévision. Leur résultat sont les suivants (fictifs) :

QI	Heures passées devant TV
106	7
86	0
100	27
101	50
99	28
103	29
97	20
113	12
112	6
110	17

1. Calculer le rapport entre ces deux variables par un coefficient de rang de Spearman.
2. Représenter graphiquement les données.

Exercice IV.2.7 :

L'exposition à un événement traumatisant peut engendrer un stress post-traumatique (STP°) chez un quart de la population exposée. Après le crash d'un avion, la fréquence d'apparition d'un stress post-traumatique a été évaluée en fonction du degré d'atteinte corporelle de l'accident. Voici les résultats obtenus (fictifs).

Blessure corporelle	fréquence du SPT
indemne	70%
accrochage	30%
blessure légère	60%
blessure grave	30%
amputation	70%
risque mortel	90%

Existe-t-il un lien entre le degré de blessure et l'apparition d'un STP ? Représenter graphiquement les données.

Exercice IV.2.8 :

Y a-t-il un marqueur physiologique du stress ? Les réponses de stress activeraient le système nerveux végétatif. Une étude à faire remplir à un échantillon de 50 personnes l'échelle de stress pour l'Adulte de Holmes Rahe et on mesuré leur tension artérielle systolique (TAS). Cette étude n'a pas trouvé de corrélation entre le score de l'échelle de Stress et la TAS. Les chercheurs ont regardé s'il y avait une corrélation entre les douze premiers items de l'échelle (correspondant aux événements les plus stressants) et la moyenne relative de la TAS enregistrée par les personnes ayant choisie l'item cible. Les résultats (fictifs) sont les suivants :

Items de l'échelle de stress pour l'Adulte de Holmes Rahe - Événement vécu dans le mois courant	Score	TAS moyenne
Décès du conjoint	100	160
Divorce	73	180
Ménopause	60	163
Séparation amoureuse	65	161
Probation ou emprisonnement	63	145
La mort d'un proche parent autre que le conjoint	63	130
Blessures ou Maladies sérieuses	53	139
Mariage ,ou début d'une relation	50	139
Perte d'emploi	47	160
Réconciliation avec le conjoint	45	140
Retraite	45	110
Problème de santé d'un membre de la famille immédiat	44	140

Les scores de chaque item sont indiqués ici uniquement pour donner l'ordre de grandeur des stresseurs. La tension artérielle est-elle un marqueur du stress ? Utiliser le coefficient de rang de Spearman et représenter graphiquement les données.

Exercice IV.2.9 :

Une étude (fictive) a étudié le lien entre l'indice de masse corporelle (IMC) et la fréquence des crises de boulimies, ainsi que la représentation du corps et la fréquence des comportements compensatoires (i.e. vomissement) chez un échantillon de personne souffrant de troubles du comportement alimentaire (TCA).

Interprétation de l'IMC	fréquences crises de boulimies
denutrition	64%
maigre	70%
corpulence normale	30%
surpoids	40%
obésité modérée	43%
obésité sévère	20%
obésité morbide ou massive	30%

Echelle de représentation du corps	moyenne des comportements compensatoires
Je déteste mon corps	80%
Je me sens mal dans ma peau	80%
Je me trouve bof et cela me gêne un peu	50%
Je me trouve bof, mais cela ne me préoccupe pas	60%
Je me trouve normal	20%
Je me sens bien dans ma peau	1%

1. Existe-t-il une relation entre l'IMC et la fréquence des crises de boumilies ?
2. Existe-t-il une relation entre la représentation de son corps et la fréquence des comportements compensatoires ?
3. Représenter graphiquement ces deux relations.

Exercice IV.2.10 :

Après un travail de recensement de tous les mots contenus dans les manuels scolaires de 5^{ème}, moins les mots définis du vocabulaire courant, des chercheurs ont conçu un questionnaire à choix multiples comprenant 800 questions, lesquelles étaient assorties de quatre propositions de réponse. Ce questionnaire était destiné à mesurer le taux d'acquisition de vocabulaire d'élèves de 5^{ème} au terme de l'année scolaire, et ce dans toutes les disciplines.

Afin d'étudier le lien existant entre la performance sémantique et la performance scolaire, les scores de 20 élèves au QCM sont comparés à leur moyenne générale sur l'année.

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Score au QCM (sur 80)	60	26	53	34	46	56	52	31	39	50	33	35	56	45	47	43	38	51	48	43
Moyenne générale	18	4	13	9	12	16	14	7	12	10	9	11	12	10	8	9	6	10	16	14

1. Représentez graphiquement ces données. Quel lien semble-t-il exister entre les deux variables ?
2. Calculer ce lien par le calcul approprié et trouver la fonction de la droite de régression linéaire.

Exercice IV.2.11 :

Dans ses recherches sur les phénomènes d'influence sociale en matière de production de comportements extrêmes, D.P. Phillips (1974) a étudié les répercussions de la publicité faite par les médias à propos du suicide d'une personnalité connue sur la répartition du nombre de suicide et du nombre d'accident de la route observés dans les jours qui suivaient une telle médiatisation. Les résultats sont présentés ci-dessous. Ils correspondent au nombre d'accident et de suicide observés dans cette période moins le nombre d'accident ou de suicide observés en moyenne dans les périodes sans événement exceptionnel.

Jour	+ 1	+ 2	+ 3	+ 4	+ 5	+ 6	+ 7	+ 8	+ 9	+ 10	+ 11
suicide	50	72	76	63	59	52	50	69	65	62	51
accident	270	313	338	266	246	235	228	308	331	290	245

- 1/ Calculez la moyenne des suicides et la moyenne des accidents observés ces onze jours post-médiatisation
 2/ Le nombre d'accident et de suicide sont-ils corrélés? Interprétez ce résultat.

Exercice IV.2.12 :

À la sortie d'un supermarché, 17 personnes sont interrogées sur le montant des dépenses qu'elles viennent d'effectuer et sur la faim qu'elles ressentent. L'évaluation de la faim se fait sur une échelle allant de 1 à 20 (plus le chiffre donné est élevé, plus la faim est forte).

Existe-t-il un lien entre l'estimation subjective de la faim et les dépenses des consommateurs ? Aidez-vous des résultats présentés dans les tableaux ci dessous.

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Évaluation de la faim	14	8	9	6	18	12	15	9	10	15	16	13	8	14	17	19	5
Dépenses (en francs)	318	106	123	48	387	158	213	87	80	214	346	199	62	145	198	231	58

Exercice IV.2.12 :

On souhaite tester l'existence d'un lien d'une part entre la santé mentale et le stress et d'autre part entre la santé mentale et les antécédents familiaux.

On utilise pour cela une échelle qui donne un indice de la santé mentale (en fonction de la présence ou de l'absence de 57 symptômes psychologiques) chez 10 sujets. Une seconde échelle permet d'indiquer le niveau de stress de ces mêmes sujets en fonction de différents paramètres (événements négatifs récents, auto-évaluation...). Par ailleurs, on relève le nombre de personnes présentant des troubles psychologiques dans l'entourage familial. Les résultats sont présentés dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10
Santé mentale	99	94	80	70	100	109	62	81	74	121
Stress	29	27	8	19	3	14	5	10	22	3
Antécédents familiaux	5	6	3	4	6	6	1	4	2	8

Déterminer si la santé mentale est liée au stress et/ou aux antécédents familiaux. Représenter graphiquement ces relations et déterminer les droites de régressions des moindres carrés.

Exercice IV.2.13 :

Le test de RISKO-MARGOLIN évalue le risque de contracter une maladie cardiovasculaire. A la suite de sa passation, on obtient un score pouvant aller de 1 à 10. Plus le score est proche de 10, plus le risque est élevé.

Les médecins pensent que le rythme cardiaque mesuré après une minute d'effort peut également être un critère indicateur de ce risque. Qu'en pensez-vous? (justifiez votre réponse). Représenter graphiquement cette relation et déterminer l'équation de la droite de régression linéaire.

Patient	1	2	3	4	5	6	7	8	9	10
Score au RISKO MARGOLIN	3	5	2	4	6	6	2	4	3	8
Rythme cardiaque après 1 mn d'effort	80	120	75	138	83	140	90	92	102	153

Exercice IV.2.14 :

La technique du « pied-dans-la-porte » est un paradigme classique en psychologie sociale dont l'objectif est d'augmenter la probabilité qu'un sujet accepte de satisfaire une requête donnée en formulant au préalable une autre requête moins coûteuse. Par exemple, si une personne demande 2 euros à un passant (requête principale), la probabilité qu'il obtienne les 2

euros est plus élevée s'il commence par demander l'heure à ce même passant (requête peu coûteuse) plutôt que de lui demander directement les 2 euros. Un chercheur veut tester l'hypothèse selon laquelle plus le nombre de requêtes peu coûteuses préalables est important plus le sujet accède à la requête principale. Que pouvez-vous lui répondre ?

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nombre de requêtes peu coûteuses	1	3	4	2	5	5	4	3	1	2	3	4	2	1	5
Pourcentages d'acceptation de la requête principale	56,3	62,6	61,2	58,4	72,5	77,4	68,5	64,1	54,3	55,2	58,9	65,6	55,2	45,7	78,9

Exercice IV.2.15 :

Le vieillissement de la population est devenu un thème de recherche très prisé depuis quelques années puisque le nombre de personnes âgées ne cesse de croître et avec lui, les troubles perceptifs (déficits auditifs et visuels), psychologiques (anxiété, dépression...) et cognitifs (troubles mnésiques, attentionnels...). Plusieurs psychologues ont étudié l'existence d'un lien éventuel entre l'état dépressif des personnes âgées de plus de 70 ans et leur plainte mnésique. Selon eux, plus les personnes âgées sont dépressives plus leurs performances à des tâches de mémoire sont faibles. Afin de tester cette hypothèse, 2 tests ont été élaborés ; le premier est un questionnaire visant à évaluer l'état dépressif des individus (score maximum : 20 ; plus le score est élevé, plus les personnes sont dépressives), et le second, un test de rappel libre (consistant à rappeler le maximum de mots parmi ceux présentés auparavant) visant à étudier un aspect de la mémoire de ces mêmes individus (score maximum : 15 ; plus le nombre de mots rappelés est important, plus la mémoire est préservée).

Confirmez-vous l'hypothèse des psychologues et pourquoi ? Représenter graphiquement cette relation et déterminer l'équation de la droite de régression linéaire.

Sujet	1	2	3	4	5	6	7	8	9	10
Score au test de dépression	10	7	17	14	3	5	19	9	11	15
Nombre de mots rappelés	7	10	3	11	12	5	2	9	7	4

Références

Liens :

<http://www.bibmath.net/dico/index.php3?action=rub&quoi=50301>

Livre de référence :

« Méthodes statistiques en Sciences Humaines » de D.C. Howell traduit chez Deboeck.

Livre révision et exercices :

« Les stats pour psycho » de Nicolas Gauvrit (Deboeck)

« Les stats pour psycho – 500 exercices corrigés » de Nicolas Gauvrit (Deboeck)