

LICENCE DE PSYCHOLOGIE - UV PSY38X2
CORRIGÉ DE L'ÉPREUVE DE TRAITEMENT DES DONNÉES EN
PSYCHOLOGIE DONNÉE EN JUIN 2001

Exercice 1 – Informatique

Une Fondation pour la protection de la nature reçoit des dons pour ses différentes campagnes. Elle utilise un système de gestion de bases de données pour gérer les dons qu'elle reçoit. Plus précisément, les dons sont mémorisés dans une table "Donateurs-dons", de structure :

Donateurs-dons(N°, Nom, Prénom, Date, Montant, Campagne).

On a représenté ci-dessous un exemple d'une telle table :

N°	Nom	Prénom	Date	Montant	Campagne
10	MARTIN	Pierre	01/04/97	500F	Marais Poitevin
12	DUPONT	Jacques	01/04/97	300F	Parcs naturels
15	DUPOND	Bernard	15/04/97	200F	Camargue Sauvage
20	MARTIN	Jean	21/04/97	150F	Parcs Naturels
12	DUPONT	Jacques	15/06/97	300F	Marais Poitevin
20	MARTIN	Jean	15/06/97	150F	Parcs Naturels
10	MARTIN	Pierre	21/04/97	200F	Oiseaux de montagne
20	MARTIN	Jean	15/06/97	150F	Marais Poitevin

- 1) Quelle est la table obtenue par sélection sur la condition "Montant > 250F" ?
- 2) Quelle est la table obtenue par projection sur les attributs N°, Nom et Prénom ?

1) On ne garde que les lignes correspondant à la condition, soit :

<i>N°</i>	<i>Nom</i>	<i>Prénom</i>	<i>Date</i>	<i>Montant</i>	<i>Campagne</i>
<i>10</i>	<i>MARTIN</i>	<i>Pierre</i>	<i>01/04/97</i>	<i>500F</i>	<i>Marais Poitevin</i>
<i>12</i>	<i>DUPONT</i>	<i>Jacques</i>	<i>01/04/97</i>	<i>300F</i>	<i>Parcs naturels</i>
<i>12</i>	<i>DUPONT</i>	<i>Jacques</i>	<i>15/06/97</i>	<i>300F</i>	<i>Marais Poitevin</i>

2) On ne garde que les colonnes citées et on élimine les doublons dans la table ainsi obtenue, soit :

<i>N°</i>	<i>Nom</i>	<i>Prénom</i>
<i>10</i>	<i>MARTIN</i>	<i>Pierre</i>
<i>12</i>	<i>DUPONT</i>	<i>Jacques</i>
<i>15</i>	<i>DUPOND</i>	<i>Bernard</i>
<i>20</i>	<i>MARTIN</i>	<i>Jean</i>

Exercice 2

Une expérience a été menée en utilisant un plan de la forme $\mathcal{S}_8 < \mathcal{A}_3 * \mathcal{B}_2 > * \mathcal{C}_4$.

Combien de groupes indépendants de sujets a-t-on constitué ?

Combien de sujets différents ont été utilisés pour cette expérience ?

Combien d'observations différentes d'un même sujet a-t-on effectué ?

*Les sujets sont emboîtés dans $< \mathcal{A}_3 * \mathcal{B}_2 >$. On a 6 combinaisons de modalités et donc 6 groupes de sujets.*

*L'indice dans $\mathcal{S}_8 < \mathcal{A} * \mathcal{B} >$ montre qu'il y a 8 sujets dans chaque groupe, soit en tout : $6 \times 8 = 48$ sujets.*

Les sujets sont croisés avec \mathcal{C}_4 , on a donc fait 4 observations de chaque sujet.

Exercice 3

N.B. Les parties A, B et C sont indépendantes.

Dans une étude publiée en 1995, A.M. Ferrandez, M. Durup et J. Pailhous ont mené des expériences pour tenter de mieux comprendre comment la marche ralentit avec le vieillissement.

Partie A : Etude d'une reprise de l'expérience

Dans une reprise de l'expérience, on considère 60 sujets, répartis en trois classes d'âge, dont les centres sont respectivement 65 ans, 75 ans et 85 ans. L'effectif de chaque classe est de 20 sujets. Chaque sujet devait traverser une salle, d'abord en marchant normalement; puis il revenait au point de départ et devait refaire une marche, mais le plus vite possible.

Dans chacune des deux conditions, on a mesuré quatre paramètres cinématiques de la marche: la vitesse, la longueur d'enjambée, la durée du cycle et la durée de double appui.

A.1) Etude du plan d'expérience utilisé.

- Quels sont les facteurs de variation qui sont pris en compte?
- On étudie ici plusieurs variables dépendantes. Lesquelles?
- Ecrire les relations d'emboîtement ou de croisement entre les facteurs pris deux à deux, puis le plan d'expérience utilisé.

A.2) Analyse des résultats relatifs à la variable "longueur de l'enjambée".

On s'intéresse ici à la variable "longueur de l'enjambée", plus simplement appelée "longueur" dans les documents ci-dessous. Les données observées pour les 60 sujets sont rassemblées dans le tableau 1.

65 ans		75 ans		85 ans	
Normale	Rapide	Normale	Rapide	Normale	Rapide
0.71	0.93	0.61	0.56	0.41	0.37
0.76	0.78	0.75	0.72	0.44	0.54
0.77	1.08	0.75	0.79	0.52	0.69
0.88	1.10	0.80	0.73	0.54	0.70
0.93	1.11	0.82	0.86	0.54	0.76
0.94	1.12	0.84	0.94	0.60	0.78
0.96	1.16	0.86	0.94	0.64	0.81
0.96	1.16	0.88	0.98	0.66	0.78
0.99	1.15	0.88	1.06	0.72	0.85
1.01	1.23	0.91	1.14	0.74	0.86
1.02	1.19	0.93	1.06	0.78	0.86
1.03	1.24	0.98	1.21	0.78	0.88
1.06	1.27	0.98	1.27	0.80	0.90
1.13	1.25	1.05	1.34	0.85	0.95
1.13	1.31	1.07	1.31	0.86	1.05
1.15	1.29	1.11	1.30	0.87	0.99
1.17	1.45	1.19	1.37	0.90	1.06
1.19	1.44	1.23	1.46	0.93	1.10
1.30	1.41	1.26	1.55	1.02	1.19
1.43	1.53	1.33	1.65	1.06	1.24

Tableau 1: Longueur de l'enjambée

Le tableau d'analyse de variance relatif à ces données se présente ainsi :

Sources de var.	ddl	SC	CM	F
<i>Entre les sujets</i>				
Age	2	2.1728	1.0864	...
$\mathcal{S}(\mathcal{A})$	57	4.8719	0.0855	...
<i>Dans les sujets</i>				
Condition	1	0.7304
$\mathcal{C} \times \mathcal{A}$	2	0.0127
Résidu	57	0.2100	0.00368	
Total	119	7.9979		

a) Compléter ce tableau en calculant les carrés moyens et les statistiques F de Fisher qui sont remplacés par “...” dans le tableau ci-dessus.

b) En utilisant un seuil de 5%, étudier quelles sont les sources de variation dont l’effet est significatif.

c) Illustrer le résultat relatif à l’interaction Age×Condition à l’aide d’un graphique adapté.

A.1) Outre le facteur sujet \mathcal{S} , les facteurs de variation pris en compte sont d’une part l’âge \mathcal{A}_3 (trois niveaux : 65 ans, 75 ans, 85 ans) et d’autre part le type de marche ou condition \mathcal{C}_2 (deux niveaux : normale ou rapide). Les variables dépendantes sont les 4 paramètres cinématiques mesurés : vitesse, longueur d’enjambée, durée du cycle, durée de double appui.

Les sujets sont évidemment emboîtés dans le facteur âge ($\mathcal{S} < \mathcal{A}_3 >$) et croisés avec le type de marche ($\mathcal{S} * \mathcal{C}_2$). Pour chaque niveau du facteur âge, les deux types de marche ont été observés : âge et type de marche sont également croisés ($\mathcal{A}_3 * \mathcal{C}_2$). Finalement, le plan d’expérience est : $\mathcal{S} < \mathcal{A}_3 > * \mathcal{C}_2$.

A.2) Le tableau complété est le suivant :

Sources de var.	ddl	SC	CM	F
<i>Entre les sujets</i>				
Age	2	2.1728	1.0864	12.71
$\mathcal{S}(\mathcal{A})$	57	4.8719	0.0855	
<i>Dans les sujets</i>				
Condition	1	0.7304	0.7304	198.48
$\mathcal{C} \times \mathcal{A}$	2	0.0127	0.00635	1.725
Résidu	57	0.2100	0.00368	
Total	119	7.9979		

Pour $\alpha = 5\%$, on a : $F_{crit}(1, 57) = 4.00$ et $F_{crit}(2, 57) = 3.15$.

Pour le facteur Age, $ddl_1 = 2$, $ddl_2 = 57$ et $F_{cal} = 12.71$. L’effet du facteur Age est donc significatif : il y a un comportement différent selon la classe d’âge.

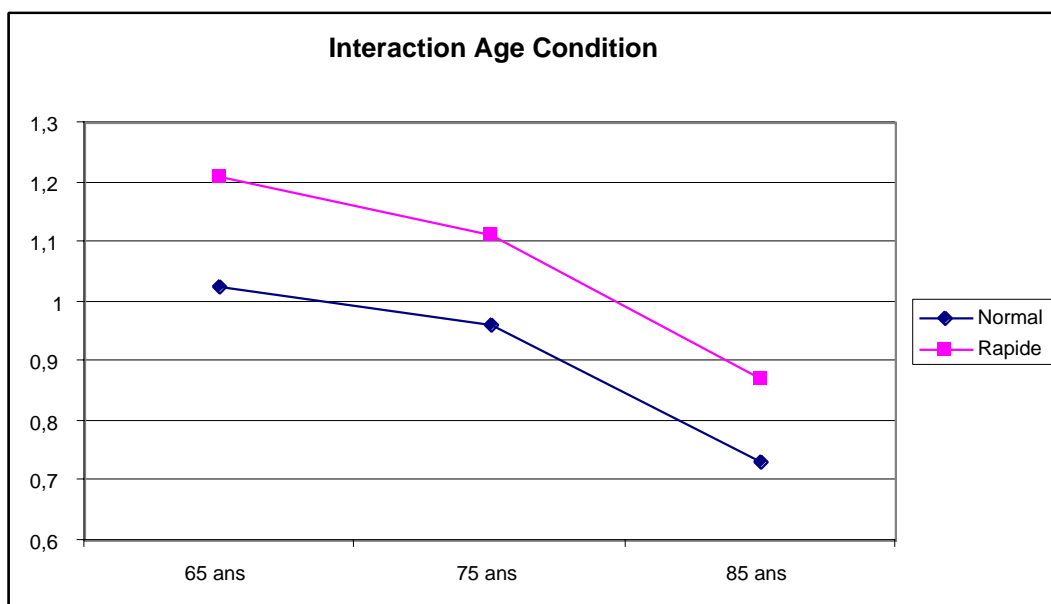
Pour le facteur Condition, $ddl_1 = 1$, $ddl_2 = 57$ et $F_{cal} = 198.47$. L’effet du facteur Condition est donc significatif : il y a un comportement différent selon le type de marche.

Pour l’interaction Age×Condition, $ddl_1 = 2$, $ddl_2 = 57$ et $F_{cal} = 1.725$. L’effet d’interaction n’est donc pas significatif.

Les moyennes pour les 6 groupes de données sont les suivantes :

Marche	65 ans	75 ans	85 ans
Normale	1.026	0.96	0.73
Rapide	1.21	1.11	0.87

D’où le graphe d’interaction suivant, dans lequel le quasi-parallélisme des deux lignes brisées montre l’absence d’interaction :



Partie B : Etude de l'expérience originale

L'expérience originale portait sur 67 sujets. Les effectifs des 3 classes d'âge étaient respectivement de 28, 27 et 12 sujets. Les données relatives aux 4 paramètres cinématiques ont été analysées au moyen d'analyses de variance univariées, une par paramètre cinématique présenté. Les résultats publiés par les auteurs sont donnés dans les tableaux 2 et 3.

Age marche	65(n=28)		75(n=27)		85(n=12)	
	norm	rap	norm	rap	norm	rap
Longueur (m)						
Moyenne	1.09	1.26	0.96	1.11	0.71	0.86
Ecart type	0.20	0.16	0.26	0.24	0.23	0.27
Vitesse (m/s)						
Moyenne	1.00	1.31	0.82	1.08	0.60	0.84
Ecart type	0.27	0.25	0.26	0.23	0.19	0.27
Durée du cycle (s)						
Moyenne	1.13	0.98	1.19	1.04	1.20	1.03
Ecart type	0.16	0.12	0.14	0.11	0.12	0.11
Durée double appui (s)						
Moyenne	0.19	0.14	0.22	0.16	0.27	0.20
Ecart type	0.06	0.03	0.07	0.03	0.07	0.06

Tableau 2: Paramètres cinématiques – expérience originale

B.1) Parmi les affirmations suivantes formulées par les auteurs, lesquelles peuvent être vérifiées à partir des tableaux 2 et 3?

- L'effet de la condition était significatif ($p < .001$) pour toutes les variables considérées.
- Des différences significatives pour l'âge ($p < .001$) ont été observées dans la diminution de vitesse et de longueur d'enjambée et dans l'augmentation de la durée du double appui.
- La durée du cycle tend à croître avec l'âge, mais pas de manière significative.
- Aucune interaction âge×condition n'était significative.

B.2) Utiliser les éléments précédents pour justifier la conclusion partielle des auteurs :

	Condition	Age
	$F(1, 64)$	$F(2, 64)$
Longueur	115.02	21.48
Vitesse	90.89	14.68
Durée	201.61	1.80
Double appui	117.45	10.29

Tableau 3: Résultats des analyses de variance – expérience originale

Ainsi, le processus de ralentissement locomoteur observé avec le vieillissement semble principalement lié à la diminution de longueur d’enjambée, alors que les modulations volontaires de vitesse (modifications de vitesse dues aux conditions) ont consisté à modifier à la fois la longueur d’enjambée et sa durée.

B.1) Pour les 4 variables, la statistique testant l’effet de la Condition suit une loi de Fisher avec $ddl_1 = 1$ et $ddl_2 = 64$. Pour un seuil $\alpha = .001$, on lit dans la table : $F_{crit} = 11.97$. Comme toutes les valeurs calculées de la statistique sont supérieures à cette valeur, la première affirmation est vérifiée.

Pour le facteur Age, $ddl_1 = 2$ et $ddl_2 = 64$. Au seuil $\alpha = .001$, on a alors : $F_{crit} = 7.76$. Les valeurs calculées de la statistique F montrent un effet significatif de l’âge pour les variables Longueur, Vitesse et Double appui. En revanche, une analyse de variance est un test bilatéral, et le test à lui seul ne montre pas que cet effet se traduit par une diminution pour les deux premières et une augmentation pour la dernière, lorsque l’âge augmente. C’est le tableau 2 qui permet d’apporter ces précisions, au niveau descriptif.

De même, le tableau 2 permet d’affirmer que “la durée du cycle tend à croître avec l’âge”; par contre, la valeur calculée de la statistique F , c’est-à-dire 1.80, montre que le résultat observé sur les échantillons, ne met pas en évidence de différence de comportement dans les populations parentes même si l’on choisit un seuil beaucoup moins “exigeant” (5% par exemple).

La dernière affirmation, quant à elle, ne peut pas être vérifiée à l’aide des données fournies.

B.2) On a montré que l’effet de l’âge sur la vitesse est significatif. Compte tenu du sens des variations, on peut donc parler de “ralentissement locomoteur lié à l’âge”. Parmi les effets de l’âge, celui observé sur la longueur d’enjambée est celui qui correspond à la valeur de F la plus élevée, ce qui justifie en partie la première conclusion. De même, les fortes valeurs des statistiques F évaluant l’effet de la condition sur la Longueur et la Durée justifient la seconde. Le fait que seule la Condition a un effet sur la Durée apporte une justification supplémentaire.

Partie C : *Etude, sur une autre reprise de l’expérience, de la corrélation entre durée du double appui et vitesse.*

La durée du double appui augmente avec l’âge. Pour éliminer les différences individuelles, on a divisé la durée du double appui par la durée totale du cycle et on a exprimé cet indice en pourcentage (conformément au texte original, cet indice sera exprimé à l’aide d’un nombre compris entre 0 et 100).

C.1) Calculer l’indice de durée du double appui lorsque la durée de double appui est de 0.22 s et la durée du cycle de 1.19 s.

C.2) Dans une reprise de l’expérience (différente de celle envisagée dans la partie A), on a relevé la vitesse (exprimée en m/s) et l’indice de durée du double appui chez 20 sujets (d’âges

Sujet	Vitesse	Indice	U_i	V_i
s1	0.31	64	-1.171	4.159
s2	0.42	51	-0.868	3.932
s3	0.45	51	-0.799	3.932
s4	0.64	43	-0.446	3.761
s5	0.86	38	-0.151	3.638
s6	1.15	29	0.140	3.367
s7	1.22	30	0.199	3.401
s8	1.23	28	0.207	3.332
s9	1.23	26	0.207	3.258
s10	1.36	27	0.307	3.296
s11	1.48	22	0.392	3.091
s12	1.53	24	0.425	3.178
s13	1.55	26	0.438	3.258
s14	1.62	27	0.482	3.296
s15	1.62	24	0.482	3.178
s16	1.81	23	0.593	3.135
s17	1.83	22	0.604	3.091
s18	2.22	14	0.798	2.639
s19	2.23	15	0.802	2.708
s20	2.60	18	0.956	2.890

Tableau 4: Vitesse et indice - reprise de l'expérience

variés) marchant à plusieurs vitesses. Les résultats observés sont donnés dans les trois premières colonnes du tableau 4.

Représenter les données des colonnes "Vitesse" et "Indice" du tableau 4 à l'aide d'un nuage de points. Commenter.

C.3) Pour étudier la relation existant entre "Vitesse" et "Indice", on construit un protocole dérivé en appliquant une transformation mathématique à chacune des variables¹. On obtient ainsi :

U : protocole dérivé calculé à partir de la variable "vitesse" ;

V : protocole dérivé calculé à partir de la variable "indice".

Les valeurs observées des variables U et V sont données dans les deux dernières colonnes du tableau 4.

On donne par ailleurs : $\sum U_i = 3.597$, $\sum V_i = 66.540$, $\sum U_i^2 = 7.125121$, $\sum V_i^2 = 224.346608$, $\sum U_i V_i = 7.748995$.

a) Représenter graphiquement le nuage de points (U_i, V_i) .

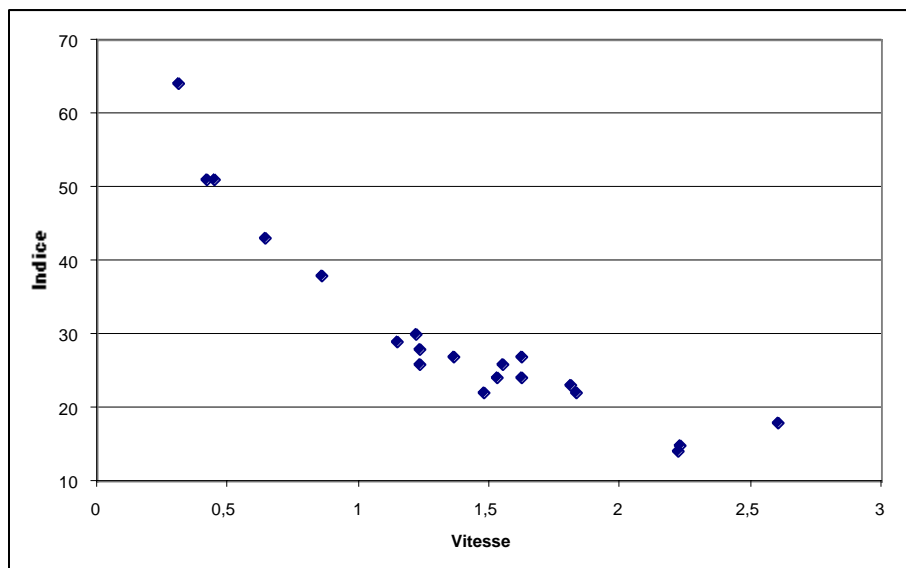
b) Calculer la moyenne et la variance de chacune des deux séries (U_i) et (V_i) , puis la covariance et le coefficient de corrélation des deux séries. La corrélation est-elle significative au seuil de 1%?

c) Déterminer une équation de la droite de régression de V selon les valeurs de U . Tracer cette droite sur le graphique précédent.

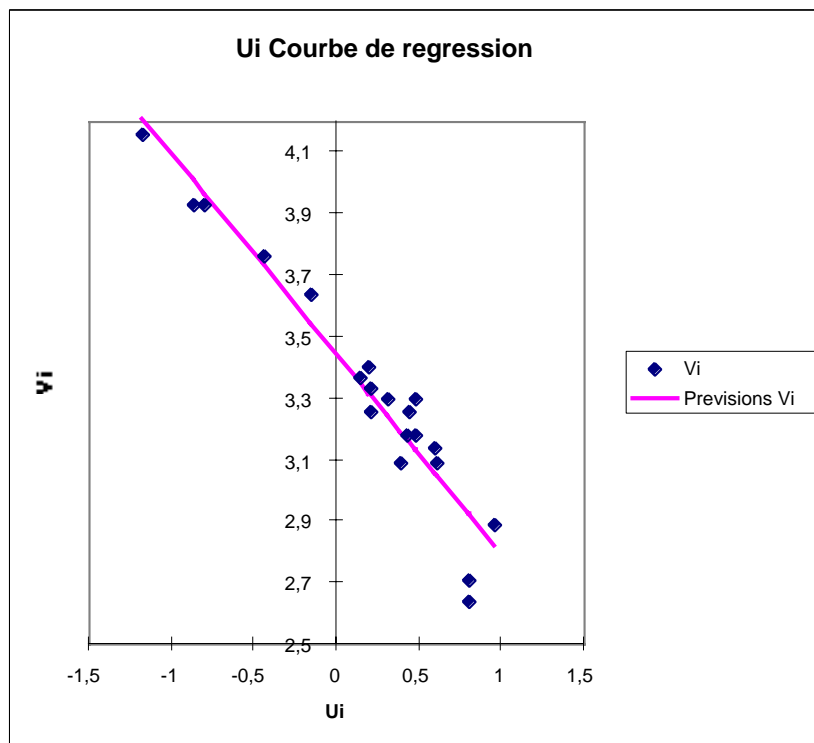
C.1) On trouve : $indice = \frac{0.22}{1.19} \times 100 = 18.5$.

C.2) Le nuage de points est représenté par :

¹Les variables U et V sont en fait définies à l'aide de la fonction logarithme népérien : $U = \ln(\text{Vitesse})$ et $V = \ln(\text{Indice})$. Mais il n'est pas nécessaire de le savoir, ni de savoir ce qu'est la fonction logarithme, pour traiter les questions qui suivent.



C.3) Ce second nuage, et la droite de régression, sont donnés par :



C.4) On trouve : $\bar{U} = 0.18$, $\bar{V} = 3.3$, $\sigma^2(U) = 0.3229$, $\sigma^2(V) = 0.1484$, $Cov(U, V) = -0.2109$ et $r = -0.96$.

Pour un seuil de 1% , 18 ddl et un test bilatéral, on a $r_{crit} = 0.5614$. La valeur trouvée pour r est donc significative d'une corrélation entre U et V .

Les coefficients de l'équation de la droite de régression sont donnés par :

$$a = \frac{-0.2109}{0.3239} = -0.65 ; b = 3.3 + 0.18 \times 0.65 = 3.42.$$