DEUG de Psychologie
Corrigé de l'épreuve de Statistiques appliquées à la Psychologie donnée en janvier 2001

Exercice 1

1) Soit X une variable distribuée selon une loi binomiale de paramètres $N = 5$ et $p = 0.5$.
Quelles sont les modalités de la variable X ? Quelles sont leurs fréquences d'apparition ? Calculer la fréquence de $X \geq 3$.
2) On considère un groupe de 5 enfants auxquels on demande choisir entre un jouet rouge et un jouet vert. Trois d'entre eux choisissent le jouet vert.
On fait l'hypothèse que les enfants choisissent au hasard entre les deux couleurs de jouets. Quelle chance a-t-on alors d'observer au moins 3 choix en faveur du jouet vert ? Selon vous, le hasard suffit-il à expliquer ce qui a été observé ?
3) Dans une étude devenue classique (1939), deux chercheurs ont montré à des enfants noirs une poupée noire et une poupée blanche en leur demandant de choisir celle avec laquelle ils voudraient jouer. Sur 252 enfants, 169 ont choisi la poupée blanche tandis que 83 préféraient la poupée noire.
a) Comme précédemment, on fait l'hypothèse que les enfants choisissent au hasard. Quelle est alors la loi de la variable Y "nombre d'enfants ayant choisi la poupée blanche parmi les 252 enfants" ?
b) A l'aide d'une approximation de la loi binomiale par la loi normale, évaluer la fréquence de $Y \geq 169$.
c) Selon vous, le hasard constitue-t-il une explication satisfaisante du comportement observé ?

1) Par définition d'une loi binomiale, une telle variable a comme modalités: 0, 1, 2, 3, 4, 5. Leurs fréquences sont données par :

$P(X = 0) = C_5^0\ 0.5^0\ 0.5^5 = 0.0313$
$P(X = 1) = C_5^1\ 0.5^1\ 0.5^4 = 0.1563$
$P(X = 2) = C_5^2\ 0.5^2\ 0.5^3 = 0.3125$
$P(X = 3) = C_5^0\ 0.5^3\ 0.5^2 = 0.3125$
$P(X = 4) = C_5^0\ 0.5^4\ 0.5^1 = 0.1563$
$P(X = 5) = C_5^0\ 0.5^5\ 0.5^0 = 0.0313$
$P(X \geq 3) = 0.3125 + 0.1563 + 0.0313 = 0.50$

2) Le nombre de choix d'un jouet rouge parmi 5 enfants suit une loi binomiale de paramètres $N = 5$ et $p = 0.5$. D'après la question précédente, on a $P(X \geq 3) = 0.5$. On a donc a 50% de chances d'observer au moins trois choix en faveur du jouet vert. Le hasard constitue donc une explication tout à fait satisfaisante des observations faites.
3) a) La loi de Y est une loi binomiale de paramètres $N = 252$ et $p = 0.5$.
3) b) Soit $\tilde{Y}$ la variable normale approchant Y et   la variable normale centrée réduite associée.
$\tilde{Y}$ a pour moyenne $\mu(\tilde{Y}) = 252 \times 0.5 = 126$ et pour écart type $\sigma(\tilde{Y}) = \sqrt{252 \times 0.5 \times 0.5} = 7.94$.
$Y \geq 169$ correspond à $\tilde{Y} \geq 168.5$ ou encore ≥ 5.35.
Cette dernière valeur est en dehors des limites de nos tables. On en déduit que $P(Y \geq 169) < 10^{-4}$.
3) c) Sous l'effet du hasard, on avait moins d'une chance sur 10000 d'observer un protocole comportant au moins 169 choix pour la poupée blanche. Il faut donc rechercher une autre explication au comportement des enfants.

Exercice 2

Dans un examen comportant un très grand nombre de candidats, on assimile la série des notes à une variable continue distribuée selon une loi normale.
1) On sait que 50% des candidats ont obtenu une note supérieure ou égale à 10.9 et que 34.13% ont obtenu une note comprise entre 10.9 et 16.3.
En déduire la moyenne et l'écart type de la distribution des notes.
2) a) Déterminer la proportion de candidats qui ont obtenu moins de 7.
b) Déterminer la proportion de candidats qui ont obtenu une note comprise entre 8 et 11.
3) On tire au hasard des échantillons de taille 100 parmi les copies. Quelle est la loi suivie par la variable $\overline{X}$ "moyenne observée sur un échantillon de 100 copies" ? Quels sont ses paramètres ?

1) La distribution normale étant symétrique, la médiane 10.9 est aussi la moyenne. D'où $\mu = 10.9$*. D'autre part, pour une variable normale centrée réduite,* $P(0 \leq \quad \leq a) = 38.13\%$ *pour* $\quad = 1$*. L'écart type de la distribution des notes est donc :* $\sigma = 16.3 - 10.9 = 5.4$*.*
2) On utilise la variable normale centrée réduite $\quad = \dfrac{X - 10.9}{5.4}$.
2) a) $P(X \geq 7) = P(\quad \geq -0.72) = 0.2358$
2) b) $P(8 \geq X \geq 11) = P(-0.54 \geq \quad \geq 0.02) = 0.2134$
3) D'après le théorème de l'échantillonnage, la loi suivie par la variable $\overline{X}$ *est la loi normale de paramètres* $Moy(\overline{X}) = 10.9$ *et* $\sigma(\overline{X}) = \frac{5.4}{\sqrt{100}} = 0.54$

Exercice 3

N.B. Les quatre questions sont indépendantes

Nurcombe et it al. ont mené en 1984 une étude sur les enfants présentant un poids réduit à la naissance (PRN). Ces enfants posent des problèmes particuliers à leurs parents parce qu'ils sont, en apparence, apathiques et imprévisibles; en outre, ils risquent de connaître des problèmes physiques et comportementaux. L'étude a porté sur trois groupes d'enfants ;

- Un groupe expérimental de 25 enfants PRN dont les mères bénéficiaient d'un apprentissage particulier : elles étaient sensibilisées aux signaux émis par ces enfants, afin de leur permettre de mieux répondre à leurs besoins ;
- Un groupe témoin de 31 enfants PRN dont les mères ne bénéficiaient d'aucun programme particulier ;
- Un groupe d'enfants dont le poids à la naissance était normal.

Le tableau 1 ci-dessous représente une partie des données observées. Il indique d'une part, l'indice de développement mental (IDM) à 6 mois et à 24 mois pour le groupe témoin PRN, ainsi que la différence entre les deux indices et d'autre part l'IDM à 24 mois pour le groupe expérimental PRN.

1) Préciser les populations et les variables statistiques étudiées. Quel est le type (protocole, effectifs, etc.) du tableau 1 ?

2) Soit X la variable IDM à 6 mois pour le groupe PRN témoin.
a) Déterminer la médiane et la moyenne arithmétique μ de cette variable.
b) Dresser le tableau des écarts à la moyenne $(x_i - \mu)$ et utiliser ce tableau pour calculer la variance et l'écart type de cette variable.
c) Procéder à un regroupement de la variable X en 4 classes en utilisant comme bornes des classes les valeurs :

80 100 110 120 140.

d) Représenter la variable X à l'aide d'un histogramme en utilisant le regroupement en classes précédent.

3) Soit Y la variable IDM à 24 mois pour le groupe PRN témoin. On cherche à déterminer s'il existe un lien de corrélation linéaire entre les variables X et Y.
a) Construire le nuage des points (x_i, y_i).
b) On donne : $\text{Moy}(Y) = 106.71$; $\text{Var}(Y) = 162.40$; $\sum x_i y_i = 368725$. Calculer la covariance, puis le coefficient de corrélation linéaire des variables X et Y. Commenter les résultats obtenus.
4) On veut comparer les IDM à 24 mois dans les deux groupes. A cette fin, on procède à un regroupement de la variable IDM en deux classes : $[80, 112[$ et $[112, 143]$ et on procède à un test du χ^2 sur le tableau de contingence ainsi obtenu.
a) Compléter le tableau de contingence suivant :

IDM-24	Témoin	Expérimental
$[80, 112[$	...	...
$[112, 143]$	...	...

b) Réaliser un test du χ^2 au seuil de 5% sur le tableau de contingence précédent et conclure.

1) La population étudiée est constituée de 56 enfants PRN. Les variables sont le groupe d'appartenance (groupe expérimental ou groupe témoin), l'IDM à 6 mois et l'IDM à 24 mois (ou de manière équivalente, l'âge (6 ou 24 mois) et l'IDM). Le groupe est une variable qualitative, l'IDM est une variable numérique continue.
2) a) La médiane est le score du 16è enfant après classement par IDM croissants. D'où $Md = 115$. *La moyenne est* $\mu = \frac{114+\cdots+132}{31} = 111$.
2) b) Un extrait du tableau des écarts à la moyenne est donné ci-dessous.

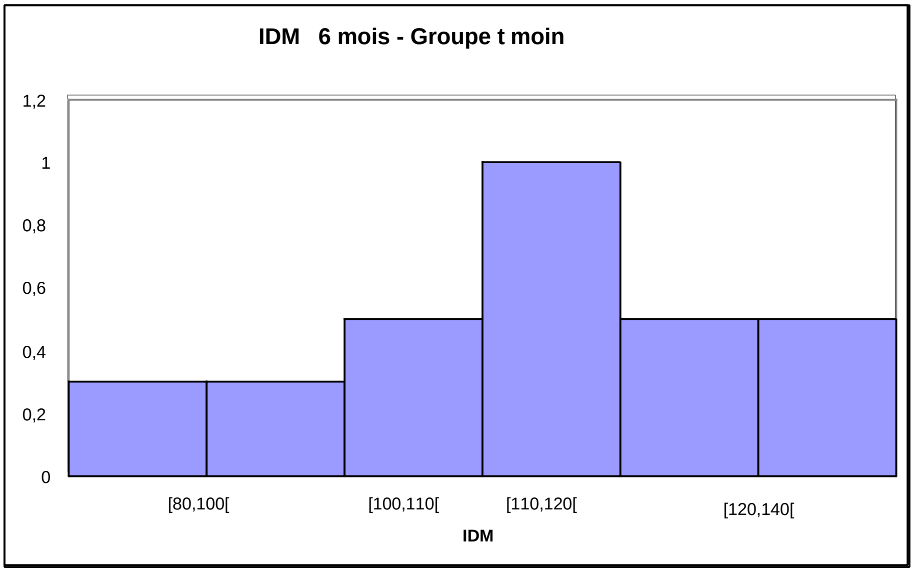
x_i	$x_i - \mu$	$(x_i - \mu)^2$
84	*-27*	*729*
...	...	...
139	*28*	*784*

On obtient : $\sigma^2 = \frac{5756}{31} = 185.68$ *et* $\sigma = 13.63$.
2) c) Le regroupement en classes conduit au tableau suivant :

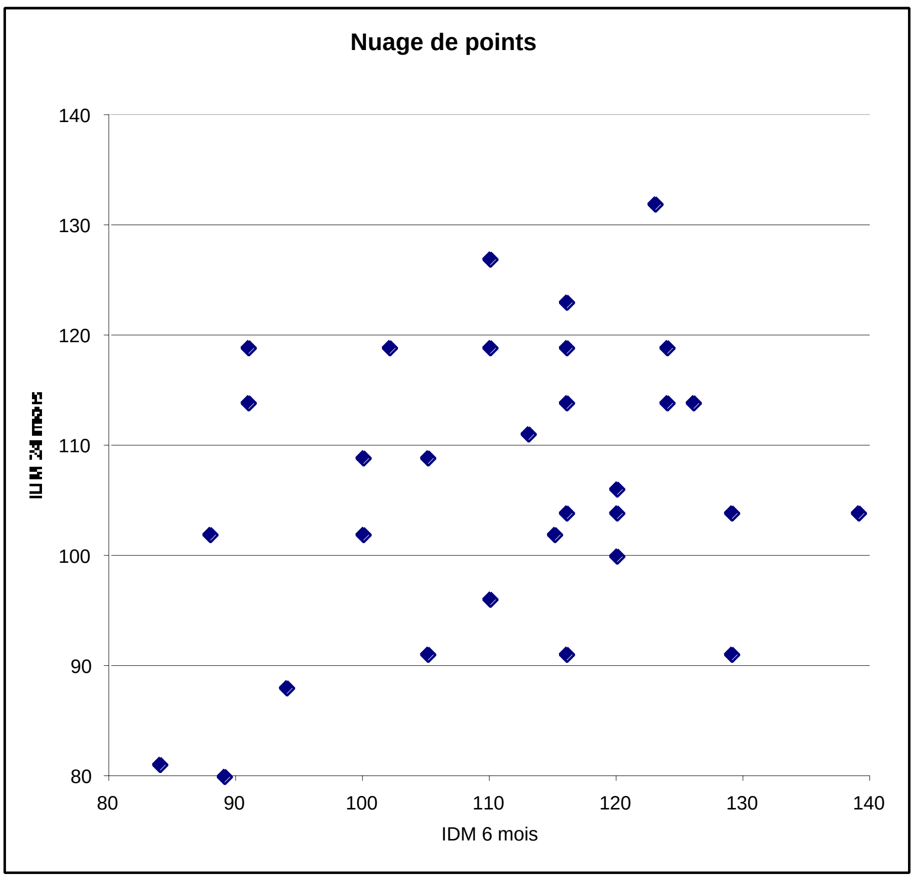
Classe	$[80, 100[$	$[100, 110[$	$[110, 120[$	$[120, 140[$
Effectif	*6*	*5*	*10*	*10*
Amplitude	*20*	*10*	*10*	*10*
Densité	*0.3*	*0.5*	*1*	*0.5*

2) d) On obtient ainsi l'histogramme suivant :

3) a)



3) a) Le nuage de points a l'allure suivante :



3) b) $Cov(X,Y) = \frac{368725}{31} - 111 \times 106.71 = 49.55$. $\rho = \frac{49.55}{\sqrt{162.40 \times 13.63}} = 0.285$.
La corrélation entre IDM-6 mois et IDM-24 mois est donc faible (ce que laissait prévoir le nuage de points).
4) a)

IDM-24	*Témoin*	*Expérimental*	*Total*
< 112	*19*	*9*	*28*
≥ 112	*12*	*16*	*28*
	31	*25*	*56*

4) b) Les effectifs théoriques et les contributions au χ^2 *sont donnés par :*

IDM-24	*Témoin*	*Expérimental*
< 112	*15.5*	*12.5*
≥ 112	*15.5*	*12.5*

IDM-24	*Témoin*	*Expérimental*
< 112	*0.79*	*0.98*
≥ 112	*0.79*	*0.98*

Soient les deux hypothèses statistiques suivantes :
H_0 *: La distribution des IDM est indépendante d'un éventuel apprentissage de la mère.*
H_1 *: Le programme d'apprentissage a un effet sur l'IDM de l'enfant.*
On a ici : $\chi^2_{obs} = 3.54$*. Or, pour* $ddl = 1$ *et* $\alpha = 5\%$*, on lit dans la table :* $\chi^2_{crit} = 3.84$*. C'est donc l'hypothèse* H_0 *qui est ici retenue : on n'a pas mis en évidence d'effet significatif du programme d'apprentissage.*

PRN Témoin				PRN expérimental	
	IDM-6	IDM-24	Diff		IDM-24
s1	124	114	-10	s'1	96
s2	94	88	-6	s'2	127
s3	115	102	-13	s'3	127
s4	110	127	17	s'4	137
s5	116	104	-12	s'5	114
s6	139	104	-35	s'6	119
s7	116	91	-25	s'7	109
s8	110	96	-14	s'8	109
s9	129	104	-25	s'9	143
s10	120	106	-14	s'10	109
s11	105	91	-14	s'11	116
s12	88	102	14	s'12	114
s13	120	104	-16	s'13	143
s14	120	100	-20	s'14	109
s15	116	114	-2	s'15	117
s16	105	109	4	s'16	127
s17	100	109	9	s'17	112
s18	91	119	28	s'18	112
s19	129	91	-38	s'19	98
s20	84	81	-3	s'20	137
s21	91	114	23	s'21	112
s22	116	119	3	s'22	109
s23	100	102	2	s'23	119
s24	113	111	-2	s'24	106
s25	89	80	-9	s'25	109
s26	102	119	17		
s27	110	119	9		
s28	116	123	7		
s29	124	119	-5		
s30	126	114	-12		
s31	123	132	9		

Tableau 1: Données PRN