

Section : Psychologie - Master 1ère année

STATISTIQUES PARAMÉTRIQUES ET NON PARAMÉTRIQUES
SUJET DONNÉ EN JANVIER 2006
ÉNONCÉ ET INDICATIONS DE RÉPONSES

N.B. Calculatrices, tables des lois statistiques et résumé de cours autorisés.

Exercice 1

Les chercheurs en psychologie du sport ont utilisé le terme d'*élan psychologique*¹ pour décrire les variations de performance fondées sur des succès ou échecs récents qui modifient les croyances ou la psychologie des athlètes.

Pour étudier la réalité éventuelle de cet effet au niveau des sports d'équipe, un chercheur² a relevé les séries de défaites et de victoires des équipes de la National Basketball Association en 1996/97 et 1997/98.

Chaque équipe joue 82 matches par an. Pour l'une d'entre elles, on observe 45 victoires et 37 défaites, échelonnées chronologiquement comme suit :

V V D V V V D D D D D V V D D V V D V V
D D V V D D V V D D D V V D D D V D D
V V V V V D D V V V V D V V V D D D D V V V V
D V V D D V V D D V V V D V V D D D V V

a) Quel test statistique peut-on utiliser pour étudier si la succession des victoires et des défaites est aléatoire ?

b) Mettre en œuvre le test et conclure au seuil de 5%.

Cf. exercice fait en TD. On réalise un test de Wald-Wolfowitz. Le nombre de runs est $u = 35$. On obtient $Z_{obs} = -1.37$ et le test conclut donc sur H_0 .

Exercice 2

On interroge un groupe de 30 étudiants parmi l'ensemble des étudiants en psychologie d'une université parisienne.

¹*psychological momentum* dans le texte original

²Réf. Vergin R.C., Winning streaks in sports and the misperception of momentum, Journal of Sport Behavior, Vol. 23 No 2, 2000, pp. 181-197

On note leur sexe (variable SEXE), on leur demande la série du bac qu'ils ont obtenu (variable BAC : L / S / ES / Autre) et le nombre d'années d'études de psychologie qu'ils ont suivies (variable ANPSY : 1 à 5).

On leur demande de bien vouloir passer une expérience visant à étudier la charge mentale induite par la lecture de deux textes différents, un article de journal et un texte sur la psychanalyse. Pour cela on introduit une tâche secondaire consistant à demander aux sujets de donner, pendant la lecture, des chiffres au hasard. On compte le nombre de chiffres énoncés par minute de lecture dans chacune des deux conditions et on calcule la différence dans le sens "Journal - Psychanalyse". On construit ainsi un indice qui mesure la charge mentale induite par la lecture du texte sur la psychanalyse (CHARG). Un nombre positif élevé signifie une charge mentale élevée pour la lecture du texte de psychanalyse.

On trouve ci-dessous le tableau des données recueillies :

	ANPSY	CHARG	SEXE	BAC		ANPSY	CHARG	SEXE	BAC
e01	1	3.5	F	L	e16	3	1	F	L
e02	1	5	G	Autre	e17	3	0.8	F	L
e03	1	2.8	G	S	e18	3	-0.5	F	Autre
e04	1	2.5	F	ES	e19	3	-0.4	F	ES
e05	1	1	F	L	e20	3	0.1	G	ES
e06	2	0.4	G	S	e21	4	0.7	F	S
e07	2	2.4	F	Autre	e22	4	1.1	G	Autre
e08	2	3	F	L	e23	4	1	F	ES
e09	2	2	G	L	e24	4	-0.1	F	L
e10	2	1.7	F	Autre	e25	4	-0.8	F	S
e11	2	0.9	F	ES	e26	5	0	F	L
e12	2	2.2	F	S	e27	5	0.6	F	ES
e13	2	0.5	F	S	e28	5	0.2	G	Autre
e14	3	-1.1	G	L	e29	5	-0.3	F	L
e15	3	2.6	G	S	e30	5	-1.2	G	S

1) On souhaite comparer les 5 groupes définis par les 5 niveaux de la variable ANPSY, du point de vue de la variable CHARG. On a réalisé une analyse de variance à un facteur sur les données observées. Le tableau d'analyse de variance se présente ainsi :

Source de variation	SC	ddl	CM	F
Inter-groupes	33.69
Résiduelle	28.39
Total	62.08	...		

a) Compléter ce tableau en calculant les valeurs qui sont remplacées par "...".

b) En utilisant un seuil de 5%, étudier si les cinq groupes sont significativement différents.

2) On souhaite comparer les 4 groupes définis par la série du bac obtenu du point de vue de la charge mentale, et on choisit, pour cette comparaison, de faire un test non paramétrique.

a) Proposer un test non paramétrique adapté à la situation envisagée.

b) Le protocole des rangs de la variable CHARG pour les 30 sujets est donné par :

Autre	Bac ES	Bac L	Bac S
30	25	29	27
24	16	18	11
21	5	28	23
4	9	22	12
20	18	2	26
10	13	18	14
		15	3
		7	1
		8	
		6	

Réaliser le test proposé et conclure au seuil de 5%.

Réponses. 1) a)

Source de variation	SC	ddl	CM	F
Inter-groupes	33.69	4	8.42	7.41
Résiduelle	28.39	25	1.136	
Total	62.08	29		

1) b) Au seuil de 5%, $F_{crit} = 2.76$. Comme $F_{cal} > F_{crit}$, l'hypothèse H_1 est retenue : les cinq groupes sont significativement différents du point de vue de la variable étudiée.

2) a) Test de Kruskal-Wallis.

2) b) On trouve : $\bar{R}_1 = 18.17, \bar{R}_2 = 14.33, \bar{R}_3 = 15.30, \bar{R}_4 = 14.63$ d'où $K = 0.74$. La loi suivie par K est une loi du χ^2 à 3 ddl. Au seuil de 5%, la valeur critique est $\chi_{crit}^2 = 7.815$. On conclut donc sur H_0 .

Exercice 3

On a conçu un questionnaire d'enquête comportant deux blocs de 5 items sous forme d'échelle de Likert à 5 niveaux, codés de 1 à 5. Pour chacun des deux blocs, on définit un score composite en additionnant les niveaux observés pour chacun des 5 items. L'intervalle de valeurs de ce score est donc [5, 25].

Ce questionnaire est soumis à un échantillon de 20 sujets. Pour le premier bloc, les réponses aux 5 items sont très dispersées, de sorte que les scores composites observés sont donnés par la variable Var1.

Suj.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Var1	5	6	8	8	8	9	11	12	13	13	16	18	19	20	20	20	21	23	24	24

Pour le deuxième bloc, les réponses, beaucoup moins dispersées, conduisent aux scores composites indiqués dans la variable Var2.

Suj.	A'	B'	C'	D'	E'	F'	G'	H'	I'	J'	K'	L'	M'	N'	O'	P'	Q'	R'	S'	T'
Var2	15	16	16	16	16	16	17	17	17	20	20	21	21	22	22	23	23	24	24	24

Dans chacun des deux cas, on souhaite étudier, à l'aide des tests de Kolmogorov-Smirnov et de Lilliefors, si les scores effectivement observés sont compatibles avec l'hypothèse de normalité de la distribution dans la population parente.

Le module "Ajustement de distributions" de Statistica permet d'obtenir les tableaux de résultats détaillés suivants :

Variable : Var1, Distribution : Normale (Feuille de données 1) Kolmogorov-Smimov, d= 0,14162									
Borne Sup.	Observé (effectifs)	Cumul Observé	%age Observé	% Cumulé Observé	Théorique (effectifs)	Cumul Théorique	%age Théorique	% Cumulé Théorique	Observé-Théorique
<= 4,00000	0	0	0,00	0,00	0,8477	0,8477	4,2384	4,2384	-0,8477
6,00000	2	2	10,00	10,00	0,7454	1,5931	3,7271	7,9655	1,2546
8,00000	3	5	15,00	25,00	1,1591	2,7522	5,7953	13,7608	1,8409
10,00000	1	6	5,00	30,00	1,6321	4,3842	8,1603	21,9211	-0,6321
12,00000	2	8	10,00	40,00	2,0811	6,4653	10,4053	32,3264	-0,0811
14,00000	2	10	10,00	50,00	2,4030	8,8683	12,0149	44,3413	-0,4030
16,00000	1	11	5,00	55,00	2,5127	11,3809	12,5634	56,9047	-1,5127
18,00000	1	12	5,00	60,00	2,3793	13,7602	11,8963	68,8010	-1,3793
20,00000	4	16	20,00	80,00	2,0402	15,8004	10,2009	79,0019	1,9598
22,00000	1	17	5,00	85,00	1,5842	17,3846	7,9211	86,9229	-0,5842
24,00000	3	20	15,00	100,00	1,1140	18,4986	5,5699	92,4928	1,8860
< Infini	0	20	0,00	100,00	1,5014	20,0000	7,5072	100,0000	-1,5014

Variable : Var2, Distribution : Normale (Feuille de données 1) Kolmogorov-Smimov, d= 0,22782									
Borne Sup.	Observé (effectifs)	Cumul Observé	%age Observé	% Cumulé Observé	Théorique (effectifs)	Cumul Théorique	%age Théorique	% Cumulé Théorique	Observé-Théorique
<= 15,0000	1	1	5,00	5,00	1,6860	1,6860	8,4302	8,4302	-0,6860
16,00000	5	6	25,00	30,00	1,1567	2,8427	5,7835	14,2136	3,8433
17,00000	3	9	15,00	45,00	1,6010	4,4437	8,0048	22,2184	1,3990
18,00000	0	9	0,00	45,00	2,0194	6,4630	10,0968	32,3152	-2,0194
19,00000	0	9	0,00	45,00	2,3212	8,7842	11,6060	43,9212	-2,3212
20,00000	2	11	10,00	55,00	2,4315	11,2158	12,1576	56,0788	-0,4315
21,00000	2	13	10,00	65,00	2,3212	13,5370	11,6060	67,6848	-0,3212
22,00000	2	15	10,00	75,00	2,0194	15,5563	10,0968	77,7816	-0,0194
23,00000	2	17	10,00	85,00	1,6010	17,1573	8,0048	85,7864	0,3990
24,00000	3	20	15,00	100,00	1,1567	18,3140	5,7835	91,5698	1,8433
< Infini	0	20	0,00	100,00	1,6860	20,0000	8,4302	100,0000	-1,6860

1) Indiquer comment a été calculé chacun des nombres figurant dans la deuxième ligne du premier tableau. N.B. La moyenne et l'écart type corrigé de la variable Var1 sont donnés par : $\bar{x} = 14.9$, $s_c = 6.3237$.

2) Dans l'en-tête du deuxième tableau figure l'indication $D = 0.22782$. Indiquer comment a été obtenue cette valeur à partir des données figurant dans le tableau.

3) Dans chacun des deux cas, l'hypothèse de normalité de la distribution parente peut-elle être acceptée (utiliser le test de Lilliefors et un seuil de 5%) ?

Réponses. 1) Col. 1 : Deux sujets (A et B) ont un score X vérifiant $4 < X \leq 6$ d'où l'effectif indiqué dans cette colonne.

Col. 2 : Cumul, obtenu en additionnant le cumul précédent (case au-dessus) et l'effectif de la

classe (case à droite).

Col. 3 : Pourcentage observé : Col. 1 divisée par 20.

Col. 4 : Pourcentage cumulé observé : Col. 2 divisée par 20.

Col. 5 : Effectif théorique : $20 \times P(4 < X \leq 6)$ où X désigne une variable normale de moyenne 14.9 et d'écart type 6.3237.

Col. 6 : Effectif cumulé théorique : $20 \times P(X \leq 6)$ où X désigne une variable normale de moyenne 14.9 et d'écart type 6.3237.

Col. 7 : Pourcentage (ou fréquence) théorique : $P(4 < X \leq 6)$ où X désigne une variable normale de moyenne 14.9 et d'écart type 6.3237.

Col. 8 : Pourcentage (ou fréquence) cumulé(e) théorique : $P(X \leq 6)$ où X désigne une variable normale de moyenne 14.9 et d'écart type 6.3237.

Col. 9 : Observés - Théoriques : Col. 1 moins Col. 5.

2) La valeur 0.22782 est obtenue comme maximum des valeurs absolues des différences entre les fréquences cumulées observées (col. 4) et les fréquences cumulées théoriques (col. 8). Ce maximum est atteint à la ligne 3 ($0.45 - 0.222184 = 0.227816$). N.B. Le tableau est construit en utilisant des classes de la forme $X \leq a$. En fait, sont également intervenues dans le calcul les fréquences obtenues en considérant des classes de la forme $X < a$

3) L'hypothèse de normalité de la distribution parente est H_0 , alors que H_1 correspond à l'absence de normalité. Pour un échantillon de taille 20 et un seuil de 5%, on lit dans la table du test de Lilliefors : $L_{crit} = 0.19$. L'hypothèse de normalité peut donc être acceptée dans le premier cas, et doit être refusée dans le second.