

Pratique des tests Statistiques

A.B. Dufour & J.R. Lobry

La fiche contient quelques exercices portant sur les tests statistiques paramétriques ou non à réaliser sous le logiciel R.

Exercice 1

Les études statistiques portant sur la latéralisation permettent d'estimer à environ 10% de la population la proportion d'individus qui, dans nos sociétés, utilisent préférentiellement leur main gauche dans les tâches motrices et dans l'écriture en particulier. Parmi les meilleurs tennismen et escrimeurs mondiaux, on a dénombré 18 gauchers sur 64. Peut-on admettre que les gauchers sont plus nombreux dans ces sports d'opposition que dans la population totale ?

Exercice 2

Le journal *L'équipe* donne les recettes (en francs) de 14 matches de rugby joués au cours de la journée du 13 janvier 1991 :

22550	22180	25000	38580	41860	39800	178525
31670	75785	17680	62595	23825	40550	144000

- 1) Calculer la moyenne, la variance et l'écart-type des recettes de la journée du 13 janvier.
- 2) Faire une représentation graphique des recettes.
 - (a) Tester la normalité de la distribution.
 - (b) A quelle loi de probabilité pourrait-on associer cette représentation graphique ?
 - (c) Calculer le(s) paramètre(s) de cette distribution et tester son adéquation à l'aide de Kolmogorov-Smirnov.
- 3) En 1990, les mêmes 14 matches ont donné les recettes suivantes :

19910	19570	22340	41320	38700	36740	181730
28500	79080	20390	59740	20830	43390	147120

- (a) Proposer une représentation graphique judicieuse.
 (b) Existe-t-il un gain de recettes entre 1990 et 1991 ?

Exercice 3

On a suivi, sur une période de 20 ans, deux cohortes : 200 sujets fumeurs et 200 sujets non fumeurs. On a noté le nombre d'apparition de cancer dans chacune des cohortes : 40 chez les fumeurs ; 20 chez les non fumeurs. La différence d'apparition de cancer dans les deux cohortes est-elle significative ?

Exercice 4

On a mesuré l'indice de Quételet (ou indice de masse corporelle) chez 13 jeunes filles âgées de 14 ans, atteintes du syndrome de Turner et chez 8 filles atteintes d'un déficit en hormone de croissance.

```
turner <- c(16.67, 18.32, 18.15, 18.65, 19.10, 18.38, 20.21, 17.72, 16.31, 19.05, 16.57, 15.33, 18.47)
deficit <- c(16.79, 19.22, 23, 20.52, 16.06, 20, 16.61, 17.81)
```

Existe-t-il une différence, en moyenne, de l'indice de Quételet entre ces deux pathologies de croissance ?

Exercice 5

En septembre, dans une classe de CP, on a réparti les élèves dans deux groupes. Dans le premier groupe, l'enseignement de la lecture était réalisé selon une méthode dite "uniforme" : tous les élèves franchissent les étapes pas à pas selon le même rythme ; dans le second groupe, l'enseignement est individualisé : chaque élève suit son propre rythme d'apprentissage. A la fin de l'année, chaque élève passe un test de capacité de lecture (adaptée de Conover 1971).

- Premier groupe :

227	55	184	174	176	234	147	194	252	194
88	248	149	247	161	206	16	99	171	89

- Second groupe :

209	271	63	19	14	151	184	127	165	235
53	151	171	147	228	101	292	99	271	179

1. Faire une représentation graphique.
2. Existe-t-il une différence d'apprentissage de la lecture : en variance ? en moyenne ?
3. L'information sur la variance est-elle importante ? Pourquoi ?

Exercice 6

Sept juges ont classé les cinq finalistes d'un concours de beauté local (adaptée de Conover 1971). Existe-t-il une concordance de classement de la part des juges ?

A	B	C	D	E
5	2	1	3	4
5	1	2	3	4
3	1	2	4	5
2	3	4	1	5
3	1	2	5	4
4	1	2	3	5
4	2	3	1	5

Exercice 7

On a défini quatre méthodes différentes pour étudier la croissance du blé dans des terrains de même nature et on a étudié le rendement par demi-hectare issues de ces méthodes notées de M1 à M4 (adaptée de Conover 1971).

M1	M2	M3	M4
83	91	101	78
91	90	100	82
94	81	91	81
89	83	93	77
89	83	93	77
96	83	95	81
91	88	94	80
92	91		81
90	89		
	84		

1. Existe-t-il une différence de rendement entre les quatre méthodes ?
2. Peut-on réaliser les comparaisons des différents rendements deux à deux ?
Si oui, sous quelle condition ?
3. Etudier la fonction `pairwise.wilcox.test(x, fac, p.adjust = "bonferroni")` de R et l'appliquer à la situation.
4. Apparaît parfois le message `impossible de calculer la p-value exacte avec des ex-aequos`. Quand le test est non significatif, cela a peu d'importance. Quand le test est significatif, cela peut poser des problèmes quand on est proche du risque α . Que peut-on conclure de l'étude ?