

Fiche TD $N = 01-01$

Exercice 01: Qu'est-ce qu'une méthode non paramétrique ? Qu'est-ce qu'un test statistique ?
 Qu'est-ce qu'un test non paramétrique ? Quelles sont les limites des tests non paramétriques ?

Exercice 02: Formulation d'hypothèses, choix d'un test: Pour chaque situation, formuler H_0 et H_1 , d'abord avec une phrase, puis mathématiquement. Dire quel(s) test(s) on pourrait appliquer.

- 1-) Une enseigne d'électroménager vend avec des aspirateurs une assurance de 3 ans. Elle veut tester si au moins $3/4$ des aspirateurs tombent en panne pour la première fois après 3 ans. La durée de vie de l'appareil avant la première panne suit une loi exponentielle. On observe au bout de combien de temps intervient la première panne sur 200 appareils.
- 2-) Une enseigne d'électroménager vend avec des aspirateurs une assurance de 3 ans. Elle veut savoir si au moins $3/4$ des aspirateurs tombent en panne pour la première fois après 3 ans. Pour cela, elle observe au bout de combien de temps intervient la première panne sur 200 appareils.
- 3-) Un même groupe de 100 étudiants passe un partiel et un examen. Le professeur veut savoir si les étudiants ont réussi de la même façon le partiel et l'examen.
- 4-) Supposons qu'on ait demandé à chacune des 13 électrices inscrites choisies au hasard d'indiquer si elle allait voter pour le candidat A ou le candidat B lors d'une élection à venir. Les résultats montrent que 9 des sujets ont préféré A. Est-ce une preuve suffisante pour conclure que le candidat A est préféré à B par les électrices ?
- 5-) On mesure la taille d'élèves de 4^{ème} de 30 élèves au début et à la fin de l'année scolaire. On veut savoir si les élèves ont grandi ou pas.

Exercice 03: Test des signes

- a) Neuf laboratoires pharmaceutiques ont coopéré à une étude pour déterminer le niveau de dose efficace médian d'un certain médicament. Chaque laboratoire réalisait des expériences et rapportait sa dose efficace. Pour les résultats 0.41, 0.52, 0.91, 0.45, 1.06, 0.82, 0.78, 0.68, 0.75, estimer l'intervalle de dose efficace médiane avec un niveau de confiance de $0.95 (\alpha = 0.05)$
- b) On considère une route avec une limitation de vitesse fixée à 75 km/h. On suspecte que les conducteurs empruntant cette route dépassent en moyenne la vitesse limite. On mesure la vitesse pour un échantillon de 10 voitures et on obtient les mesures 70, 74, 79, 80, 73, 77, 78, 67, 79, 81, 80, 77. On suppose que la vitesse ne fait aucune hypothèse sur la distribution.
 - Donner une description des données : calculer moyenne, médiane, écart type, tracer la boîte à moustache.
 - Proposer un test d'hypothèse pour vérifier si la vitesse moyenne est supérieure à 75 km/h.

Exercice 04: Wilcoxon signed-rank test

- a-) On veut évaluer l'efficacité d'un nouveau régime alimentaire pour réduire le niveau de glucose dans le sang chez les patients diabétiques. Pour cela, les niveaux de glucose de 10 patients diabétiques ont été mesurés avant et après un mois de régime

Sujet	1	2	3	4	5	6	7	8	9	10
Avant	268	225	252	192	307	228	246	298	231	185
Après	106	186	223	110	203	101	211	176	194	203

1. On suppose que le taux de glucose dans le sang est normalement distribué. Afficher les boxplots de deux échantillons : qu'aurait-on envie de conclure concernant l'efficacité du régime ? Les lois des échantillons Avant et Après ont-elles le même écart type ? Faire un test de comparaison de la variance pour le vérifier.

2. Quel test appliquer si l'hypothèse de normalité n'est pas raisonnable ?
3. Illustration avec R :

```
avant <- c(268,225,252,192,307,228,246,298,231,185)
apres <- c(106,186,223,110,203,101,211,176,194,203)
ech <- c(avant,apres)
group <- c(rep("avant",10),rep("apres",10))
boxplot(ech~group)
var.test(avant,apres)
t.test(avant,apres,alternative="greater",paired=TRUE)
wilcox.test(avant-apres,mu=0,alternative='greater')
```

- c-) On mesure le stress de 14 étudiants à utiliser un ordinateur avant et après un cours de 5 semaines sur l'utilisation des logiciels courants. On veut savoir si ce cours a permis de faire baisser le niveau de stress. On note X le niveau de stress avant le cours et Y le niveau de stress après le cours. Voici les résultats :

étudiant	A	B	C	D	E	F	G	H	I	J	K	L	M	N
X	19	21	23	26	32	27	38	34	28	20	29	22	30	25
Y	20	18	10	16	11	20	20	19	13	21	12	15	14	17

- 1- Est-ce que les vecteurs X et Y sont indépendants ?
- 2- Quel(s) test(s) peut-on utiliser ici ?
- 3- On veut utiliser le test de Wilcoxon signe et rang. Quelle hypothèse doit-on faire sur les données ? Formuler H_0 et H_1 et faire le test?

Exercice 05: Wilcoxon sum signed-rank test: Test de mann-whitney

On ne dispose d'aucune information sur la loi suivie par les observations dans les deux échantillons.

Sample1	38	49	45	29	31	35			
Sample2	31	42	22	26	43	37	25	30	47

Quel(s) test(s) peut-on utiliser ici ? Formuler H_0 et H_1 et faire le test?
illustration avec R :

Exercice 06:

On a recensé ci-dessous le temps, exprimé en semaines, qui leur a été nécessaire pour avoir un emploi après l'obtention de leur diplôme et éventuellement déduction faite de la période de temps correspondant à la durée du service militaire. On se demande si l'on peut considérer que le temps Y mis par le titulaire d'un bac professionnel(P) est significativement supérieur à celui X mis par un diplômé de la filière universitaire(U).

Diplômé U	1	2	3	4	5	6	7	8	9	10	11	12
Durée X_i	38	35	40	45	51	41	42	39	34	33	30	25
Rang r_i	6	5	8	?	17	9	?	7	4	3	2	1
Diplômé P	1	2	3	4	5	6	7	8	9	10	11	12
Durée Y_j	49	50	57	61	44	55	60	43	46	56	62	65
Rang s_j	15	16	20	22	?	18	21	?	14	19	23	24

1. Quel est le test, qui a été employé pour étudier ces données ? Rappeler brièvement les hypothèses et la procédure de ce test.
2. Expliquer pourquoi il est intéressant dans cette situation de réaliser soit même ce test.

Exercice 07

X and Y are two continuous random variables with cumulative distribution functions $F(t)$ and $F(t - \theta)$ respectively. Let X_1, X_2, \dots, X_m and Y_1, Y_2, \dots, Y_n denote random samples from the two populations and suppose $P(Y < X) = \omega$:

a) Prove that $\omega = \int_{-\infty}^{+\infty} F(x - \theta)f(x)dx$: Hence verify that $\omega = 0,5$ when $\theta = 0$.

b) Consider the Mann-Whitney statistic $U = \sum_{i=1}^m \sum_{j=1}^n h(X_i - Y_j)$, $h(u) = \mathbb{I}_{(0,\infty)}(u)$, Find $E(U)$ and an unbiased estimator of ω ,
rep:b) $mn\omega$, $U/(mn)$