M1 finance: Statistiques non paramétriques
Fiche 2: Test du signe/test de Wilcoxon signe et rang

Emeline Schmisser, `emeline.schmisser@math.univ-lille1.fr`, bureau 314 (bâtiment M3).

Exercice 1 : 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 15 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. 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.
5. On mesure la taille d'élèves de 4ème de 30 élèves au début et à la fin de l'année scolaire. On veut savoir si les élèves ont grandi ou pas.

Exercice 2 : Durées de vies

Un magasin d'électroménager mesure le temps avant la première panne d'un aspirateur. Les résultats sont les suivants : (en années)

$$2,\ 3.6,\ 4,\ 0.3,\ 5,\ 1,\ 2.3,\ 0.6,\ 4.9,\ 1.8,\ >5,\ >5,\ >5,\ >5,\ >5$$

1. Donner un intervalle de confiance à 10% pour le premier quartile.
2. On veut savoir si plus d'un quart des aspirateurs tombent en panne au bout de 3 ans. Tester cette hypothèse avec un niveau $\alpha = 0,1$.

Exercice 3 : Accidents de la route

Aux États-Unis, il est interdit de boire avant un certain âge. Pour lutter contre les accidents de la route, 10 états ont relevé cet âge. Ils veulent savoir si cela a eu un effet positif. Pour cela, pour le groupe d'âge concerné, on mesure le nombre d'accidents de voiture la nuit sur le nombre de conducteurs. On obtient les résultats suivants :

état	âges concernés	Ratio avant	Ratio après	Différence
Floride	18	0.262	0.202	0.060
Géorgie	18	0.295	0.227	0.068
Illinois	19-20	0.191	0.216	-0.025
Iowa	18	0.287	0.209	0.078
Maine	18-19	0.277	0.299	-0.022
Michigan	18-20	0.223	0.151	0.072
Montana	18	0.512	0.471	0.041
Nebraska	19	0.237	0.151	0.086
New Hampshire	18-19	0.348	0.336	0.012
Tennessee	18	0.342	0.307	0.035

1. Utilisez deux techniques différentes pour savoir si la mortalité a diminué avec un niveau $\alpha = 0,1$.

Exercice 4 : examen
24 étudiants ont passé le DS et l'examen. 7 étudiants ont amélioré leurs résultats, 4 ont eu des résultats constants, et 13 ont eu de moins bons résultats.

1. Quelles hypothèses peut-on faire sur les données ? Quel(s) tes(s) peut-on appliquer ?
2. On veut tester si les étudiants ont mieux réussi le partiel que l'examen. Formuler mathématiquement les hypothèses H_0 et H_1.
3. Ici, on ne sait pas trop quoi faire des 4 étudiants qui ont eu des résultats constants. C'est un problème assez courant. Même si F_X est continue, il arrive qu'un des X_i soit égal à la médiane (ou au quantile) supposé. On va faire le test de la façon la plus conservative possible : on ne veut surtout pas écarter H_0 à tort. On considère donc que 11 étudiants ont eu de meilleurs résultats, et 13 de moins bons résultats. Faire le test avec ces données (avec un niveau de 0.1).
4. On décide maintenant que parmis les étudiants qui ont eu des résultats constants, la moitié a eu des résultats légèrement moins bons et l'autre moitié des résultats légèrement meilleurs. Comment va évoluer la p-valeur ? Faire le test avec ces données (pour un niveau de 0.1).

Les données égales au quantile qu'on cherche à tester posent toujours un problème. Suivant la façon dont on les traite, on ne trouve pas le même résultat pour le test. On peut parfois éviter ce problème en prenant plus de chiffres significatifs.

Exercice 5 : Simulation
On veut construire en programme sous R qui soit capable de donner la table de la statistique T_n^+ pour n pas trop grand (on s'autorisera l'utilisation de boucles). On se donne tout d'abord une suite de z_i.

1. La suite de z_i est une suite de 0 et de 1. Elle peut donc être interprétée comme l'écriture en binaire d'un nombre. Quelles sont les valeurs possibles de ces nombres ? Est-ce que toutes les suites sont équiprobables ?
2. Écrire un programme `base2` qui étant donné un nombre entier, renvoie un vecteur donnant son écriture en base 2. Cette fonction va nous permettre de simuler les suites z_i. On pourra utiliser la fonction `%%` qui calcule le reste de la division de deux entiers.
3. Écrire un programme `suite` qui étant donné n, donne une matrice composée de toutes les suites z_i possibles pour n. On utilisera la fonction précédente.
4. Écrire un programme `T` qui calcule la statistique de test T_n^+ étant donné une matrice de configuration de z_i. On commencera par écrire la fonction `T` pour un seul vecteur z, puis on la modifiera pour qu'elle prenne en compte une matrice.
5. Écrire un programme `proba` qui étant donné n, donne la probabilité que $T_n^+ \leq k$ pour toutes les valeurs de k possibles. On utilisera les fonctions `suite` et `T`.
6. Calculer la table de T_n^+ pour n variant de 2 à 5.