

TP 5

Tests d'hypothèses

Dans ce TP, on utilisera les jeux de données issus de la bibliothèque *MASS*. Pour les charger en mémoire, on exécute l'instruction suivante :

```
| library(MASS)
```

1 Tests de conformité

Le test de conformité de Student est un test portant sur l'espérance d'une loi gaussienne et s'effectue à l'aide de la fonction *t.test*. Une exécution typique est la suivante :

```
| t.test(x, mu = mu0, alternative = "less")
```

où *x* est l'échantillon que l'on veut tester, *mu0* l'espérance de l'hypothèse simple H_0 et *alternative* la nature du test : unilatéral avec le mot clé "*two.sided*" (comportement par défaut), bilatéral inférieur avec le mot clé "*less*" et bilatéral supérieur avec le mot clé "*greater*". Le niveau de signification peut être changé avec l'argument nommé *conf.level*.

Q1 Le jeu de donnée stocké dans le fichier **bottles.data** contient des quantités effectives de liquide relevées dans 20 bouteilles de 500 ml.

En supposant l'échantillon gaussien, peut-on dire que la quantité de liquide est inférieure à 500 ml ? Tester pour différents niveaux de signification ($\alpha^* = 0.1$, $\alpha^* = 0.05$).

2 Test sur une proportion

Le test sur une proportion s'effectue avec la fonction *prop.test*. Elle s'utilise comme suit :

```
| prop.test(x, n, p)
```

où *x* est le nombre d'expérience positive, *n* le nombre d'expérience total et *p* la proportion que l'on veut tester.

Q2 Le jeu de donnée présent dans le fichier **MM.data** contient les effectifs de M&Ms de différentes couleurs issus de 30 sachets pour un total de 1713. Est-ce qu'une couleur est sur ou sous-représentée ?

3 Tests d'homogénéité (échantillons appariés)

La fonction *t.test* permet également de tester deux échantillons appariés en spécifiant l'argument *paired* = *TRUE*.

Le jeu de donnée *immer* présent ans la bibliothèque *MASS* contient les rendements de plantations d'orges en différents lieux lors de deux années successives. On souhaite tester si le rendement a été différent d'une année à une autre.

Q3 Faire un test de Student apparié sur les deux rendements. Que peut-on conclure au niveau de signification $\alpha^* = 0.05$?

Le test de Student apparié suppose que la différence des deux échantillons suit une loi gaussienne. Lorsque ce n'est pas le cas, on peut faire un test de signe.

Q4 Faire un test de signe sur les deux échantillons précédent. Pour cela :

1. Il faut créer un vecteur booléen qui indique si la différence entre les deux échantillons est négative et compter le nombre de ces différences négatives.
2. Ensuite, on utilise le test de proportion pour tester si la proportion vaut $p = 0.5$.

4 Tests d'homogénéité (échantillons indépendants)

Comparaison de deux variances

La liste *shoes* contient deux vecteurs mesurant l'usure de chaussures de marque A et B.

Q5 A l'aide de la fonction *var.test*, tester si la variance de l'usure est la même pour les deux types de chaussures.

Comparaison de deux espérances

On souhaite à présent tester si l'usure moyenne des deux marques est la même. On sait déjà d'après la question précédente que les variances sont les mêmes.

Pour comparer les espérances, on utilise encore la fonction *t.test* avec les deux échantillons en spécifiant en plus que les variances des deux échantillons sont supposées les mêmes avec le paramètre *var.equal* = *TRUE*.

Q6 Faire un test d'égalité de l'usure sur les deux marques. Que peut-on en conclure ?

5 Test d'adéquation

Le jeu de données *galaxies* regroupe les vitesses calculées de 82 galaxies. On souhaite tester la normalité de ces données.

Q7 Faire un test de normalité à l'aide de la fonction *shapiro.test*. La distribution peut-elle être considérée comme issue d'une loi normale ?

Le fichier de donnée **delai-data.data** contient des délais d'attente en jours pour un rendez-vous. On veut vérifier que cet échantillon est issu d'une loi exponentielle.

Q8 Faire un test de Kolmogorov-Smirnov avec la fonction *ks.test*. Pour cela :

1. Il faut estimer le paramètre λ de la loi parente de l'échantillon.
2. Ensuite, on utilise le test de Kolmogorov-Smirnov pour voir si l'échantillon est issu d'une loi exponentielle de paramètre λ estimé précédemment.

6 Test d'indépendance

On souhaite tester l'indépendance du choix d'un parfum de glace par rapport au caractère homme-femme. Pour cela, on dispose du tableau de contingence suivant :

	chocolat	vanille	fraise
Homme	100	120	60
Femme	350	200	90

Q9 définir la *data.frame* regroupant les données du tableau précédent. Puis, faire un test du χ^2 avec la fonction *chisq.test*. Que peut-on conclure ?