

Fonctions que l'on va utiliser

En plus des fonctions déjà vues la dernière fois, voici de nouvelles fonctions R que nous allons utiliser aujourd'hui :

- charger un package : `library(NomPackage)`
- corrélation : `cor`
- tableau de contingence : `table`, package : `MASS`
- test du Chi2 : `chisq.test`
- test de Student (à partir des données brutes) : `t.test`.
- test de Student (à partir des moyennes) : `tsum.test`, package : `BSDA`.

1 Prise de poids de nouveaux nés

On veut comparer la croissance pondérale de nouveau-nés nourris au biberon selon qu'on leur donne un lait 1 ou un lait 2. Pour rendre les groupes plus comparables, les nouveaux-nés ont été appariés sur leur poids de naissance. On note $M1$ et $M2$ les gains moyens de poids avec les pairs 1 et 2. On supposera ici que les variances des deux groupes sont statistiquement similaires.

1. Charger le fichier `NouveauNes.csv`
2. Quel test devez-vous effectuer?
3. Donnez les hypothèses du test (H_0 , H_1), la statistique du test ainsi que sa loi sous H_0 .
4. Faites le test (attention à bien spécifier les options) et conclure.
5. On souhaite étudier la liaison entre les réponses des sujets d'une même paire. Cela permet notamment de vérifier qu'il était justifié de procéder par appariement. Quel indicateur statistique pouvez-vous utiliser? Calculez-le.

Note: le test de Fisher pour comparer deux variances n'est applicable que si les échantillons sont indépendants.

2 Infarctus

On s'intéresse au rôle des antécédents familiaux dans la survenue d'un infarctus du myocarde. Une enquête a été réalisée où sont enregistrés tous les *cas* d'infarctus dans la région de Strasbourg. Ces cas ont été comparés à des individus n'ayant pas eu d'infarctus (*témoins*) tirés au sort dans la population de la région. Nous nous intéressons ici à l'âge de survenue d'un infarctus chez les parents. Les résultats de l'enquête sont les suivants :

	Cas (n=91)	Témoin (n=177)
mère	âge à l'infarctus : m=64, s =14.9	âge à l'infarctus : m=64.6, s =11.8
père	âge à l'infarctus : m=59.8, s =10.7	âge à l'infarctus : m=73, s =10.7

Les âges moyens à l'infarctus des pères sont-ils différents entre les cas et les témoins?

3 Toujours commencez avec une analyse descriptive!

On souhaite comparer la taille de deux populations d'insectes d'une même espèce vivant dans deux régions différentes. Une étude précédente suggère que la taille de ces insectes suit une distribution gaussienne.

1. Charger les données *Normal.csv*.
2. On souhaite tester l'hypothèse suivante: les deux populations ont une taille similaire. Quel test envisageriez-vous a priori? Justifier votre réponse.
3. Faites une analyse descriptive rapide et représenter graphiquement ces données. Commentez.

```
boxplot(mydata$Taille~ mydata$Groupe, ylab="Taille population")
```

4. Au vu de cette analyse descriptive, reconsidériez vous le test à mettre en oeuvre? Pourquoi?
5. Implémentez un test non paramétrique et conclure.

4 Tableau de contingence (facultatif)

592 étudiants en statistique ont répondu à une enquête sur la couleur de leurs yeux et cheveux. On se propose d'analyser ces résultats.

1. Chargez le fichier *couleur*.
2. Construisez un tableau de contingence.
3. Quelle est la couleur des yeux la plus fréquente parmi les hommes?
4. La couleur des yeux est-elle indépendante de la couleur des cheveux? Pour répondre à cette question on peut utiliser un test du Chi2.
 - (a) Quel test devez vous effectuer? Donnez les hypothèses du test (H_0 , H_1), la statistique de test ainsi que sa loi sous H_0 .
 - (b) Faites le test et enregistrez les résultats dans une variable.
 - (c) Récupérez les effectifs théoriques calculés par la fonction (valeurs attendues sous H_0 = expected values) et vérifiez qu'ils sont tous supérieurs à 5. Si certains effectifs théoriques étaient inférieurs à 5, il aurait fallu appliquer la correction de Yates.
 - (d) Conclure.

References

- Méthodes statistiques-Médecine-Biologie par Jean Bouyez, Editions Inserm (chapitre 13)
- CESAM, Programme de travail, méthodologie statistique, Jean Bouyer