

Planche de TP 6
Test du χ^2 d'adéquation

Exercice 1 (Naissances et saisons)

On dispose d'un échantillon de 8 760 naissances, groupées selon des saisons. Voici les effectifs observés :

Saison	Effectifs observés
Printemps	2 284
Eté	1 478
Automne	1 424
Hiver	3 574

Peut-on affirmer au vu de ces observations que les naissances se répartissent uniformément suivant les saisons ? On effectuera dans un premier temps les calculs manuellement puis à l'aide de la fonction `chisq.test`. Comparer les résultats obtenus.

Remarque : les saisons sont de longueur variable (printemps : 91 jours, été : 62 jours, automne : 61 jours, et hiver : 151 jours).

Exercice 2 (Générateur de nombre aléatoire de **R**)

On souhaite contrôler dans cette exercice les performances du générateur aléatoire de **R**. Générer 100 nombres de loi de Bernoulli de paramètre 0.5. Tester au niveau 5% si l'échantillon obtenu suit bien une telle loi. Même questions pour une loi Binomiale de paramètre 0.7 et 10 ainsi que pour une loi uniforme sur l'ensemble $\{1, \dots, 10\}$. Commenter les résultats obtenus...

Exercice 3 (Record de naissance à Brisbane)

Cet exercice utilise le jeu de données `babyboom.dat`. La documentation associée est contenue dans le fichier `babyboom.txt`.

1. Créer un `data.frame` contenant l'ensemble des données.
2. Créer un vecteur `poids` contenant le poids des 44 bébés. Calculer le poids moyen de cet échantillon ainsi que la variance. On s'intéresse dans cet exercice à l'hypothèse selon laquelle la distribution du poids suit une loi normale. Cette hypothèse vous semble-t-elle plausible ?
3. En combien de classes peut-on répartir notre échantillon pour pouvoir utiliser le test du χ^2 d'adéquation ? Que produit la commande :

```
> qnorm(c(0,0.12,0.24,0.36,0.48,0.60,0.72,0.84,1),mean(poids),sd(poids))
```


On prendra soins de remplacer les valeurs `Inf` et `-Inf` par, par exemple, 1000 et 5000.
4. Créer l'histogramme de fréquence des poids en fonction des classes précédentes. Tracer sur le même graphique la densité d'une loi gaussienne de moyenne `mean(poids)` et d'écart-type `sd(poids)`. Récupérer le vecteur des fréquences classe par classe.

5. Tester l'hypothèse selon laquelle le poids des bébés suit une loi normale aux niveaux 5% et 1% et commenter. (Rappel : le nombre de degré de liberté de la statistique de test est changé puisque la moyenne et la variance de la loi gaussienne sont estimées à partir des données).
6. Reprendre le même travail mais cette fois-ci en séparant les filles des garçons et commenter les résultats obtenus.
7. Tester l'hypothèse selon laquelle la probabilité d'obtenir une fille est la même que pour un garçon.

Exercice 4 (Générateur de nombre aléatoire, suite)

On s'intéresse cette fois-ci à la façon dont **R** génère des lois continues. Générer un échantillon de tailles 100 correspondant aux réalisations d'une variable aléatoire du χ^2 à 5 degrés de liberté et tester la répartition obtenue. On prendra soins de faire apparaître un histogramme ainsi que la densité de la loi à tester sur le même graphique.

Exercice 5 (Couleurs des cheveux et département)

On dispose d'un échantillon de 77 individus répartis sur 2 départements A et B. Ces individus sont classés en fonction de la couleur de leurs cheveux. Le tableau suivant présente les effectifs classe par classe :

	blonds	bruns	roux
Département A	25	9	3
Département B	13	17	10

On souhaite savoir si la couleur des cheveux des individus se répartissent de la même manière dans ces deux départements. Effectuer un test du χ^2 de comparaison de lois au niveau 5% pour essayer de répondre à cette question.