

Mesures et analyses statistiques de données - Probabilités - TPs

Décembre 2013 - Contrôle Terminal, Semestre 1, Session 1

Durée de l'épreuve : 2h00

Documents autorisés. Calculatrice autorisée.

(Tous les exercices sont indépendants. Un soin tout particulier sera apporté à la rédaction des réponses)

**Exercice 1**

1. À l'aide des fonctions `rep`, `seq` et `c` exclusivement, générer les séquences suivantes :
  - (a) 0 6 0 6 0 6
  - (b) 1 4 7 10
  - (c) 1 2 3 1 2 3 1 2 3 1 2 3
  - (d) 1 2 2 3 3 3
  - (e) 1 1 1 2 2 3
  - (f) 1 5.5 10
  - (g) 1 1 1 1 2 2 2 2 3 3 3 3
2. Générer les suites de nombres suivantes à l'aide des fonctions `c` et `rep` exclusivement, donc sans utiliser la fonction `seq` :
  - (a) 1,1 1,2 1,3 1,4 1,5 1,6 1,7 1,8 1,9 2
  - (b) 1 3 5 7 9 11 13 15 17 19
  - (c) -2 -1 0 1 2 -2 -1 0 1 2
  - (d) -2 -2 -1 -1 0 0 1 1 2 2
  - (e) 10 20 30 40 50 60 70 80 90 100
3. Reprendre la séquence (d) précédente.
  - (a) Remplacer toutes les valeurs négatives par NA
  - (b) Compter le nombre de valeurs NA.
  - (c) Remplacer les valeurs négatives par -10.

**Exercice 2** Soit la matrice : `x<-matrix(1 :120,ncol=12)`

1. Afficher toutes les lignes de la matrice commençant par un nombre pair.
2. Afficher toutes les lignes de la matrice dont la moyenne est inférieure à 60
3. Afficher toutes les colonnes de la matrice dont la somme est inférieure à 500
4. Sélectionner la sous-matrice formée des lignes de la matrice dont la moyenne est inférieure à 60 et des colonnes dont la somme est inférieure à 500.
5. Sélectionner la même sous-matrice, sauf la troisième ligne.

Bien-sûr, une seule commande est nécessaire à chaque fois pour effectuer ces opérations...

**Exercice 3** On utilise la base `airquality`, disponible dans R. Si l'on veut extraire la sous-base pour laquelle la température est supérieure à 92°F, on utilise les commandes :

```
> air1 <- subset(airquality, Temp > 92)
> air1
```

	Ozone	Solar.R	Wind	Temp	Month	Day
42	NA	259	10.9	93	6	11
120	76	203	9.7	97	8	28
..						
127	91	189	4.6	93	9	4

On peut également utiliser :

```
> air1<-airquality[airquality$Temp>92,]
```

- Créer à partir de `air1` le dataframe `air2`, où :
  - la variable `Ozone` n'est pas manquante,
  - la température est inférieure à 94°F.
- Créer à partir de `airquality` le dataframe `air3`, où la variable `Ozone` n'est pas manquante.
- Ajouter une colonne à `air3` représentant une variable valant 1 si :
  - on est dans les 6 premiers mois de l'année,
  - la température est supérieure à 80°F,
 et 0 sinon.

**Exercice 4** On vous donne un relevé des profondeurs de glace relevées dans une station météo avec les dates correspondantes :

```
> dates<-c("1971-01-20","1971-01-28","1971-02-03","1971-02-11","1971-02-18",
+ "1973-01-17","1973-01-25","1973-01-31","1973-02-17","1974-01-07","1974-01-10",
+ "1974-01-15","1974-01-22","1974-01-29","1974-02-05","1974-02-12","1974-02-19")
> mesure<-c(64,69,71,71,71,32,42,28,32,18,25,29,34,36,42,50,61)
```

Extraire de ce vecteur la première incidence de chaque profondeur mesurée, en utilisant la fonction `match`. Trier les mesures de glace dans l'ordre croissant et créer une matrice avec comme première colonne les mesures triées et comme deuxième colonne les dates correspondantes (nécessite un tri simultané!)

**Exercice 5**

- Créer une matrice `x` portant les chiffres de 1 à 100, en 4 colonnes.
  - Calculer les quantiles 0.1 et 0.9 de chaque colonne.
  - Calculer les quantiles 0.1 et 0.9 de chaque ligne.
  - Calculer la moyenne et la variance de `x`, des colonnes de `x`, des 3 premières lignes de `x`.
- Prendre le jeu de données interne à R appelé `ToothGrowth`.
  - Le visualiser pour se faire une idée de son contenu.
  - La première colonne `len` représente la longueur des dents de cochons d'Inde soumis à différents régimes. Ranger la première colonne, par colonnes, dans une matrice à 6 colonnes `mat`.
  - Calculer la matrice de corrélations entre les colonnes de `mat`.

**Exercice 6**

- Créer un vecteur de 10 valeurs en tirant les 6 premières dans  $\mathcal{N}(2; 5)$  et les 4 suivantes dans  $\mathcal{N}(1; 4)$ .
- Tirer un vecteur  $v_1$  de 10 valeurs dans une mixture de 2 lois normales,  $\mathcal{N}(2; 5)$  et  $\mathcal{N}(1; 4)$ , avec une probabilité 50-50.
  - Calculer la moyenne de  $v_1$ . Répéter l'opération 10 fois en conservant les moyennes des 10 échantillons  $v_1$  successifs dans un vecteur.
- Recommencer la même opération en tirant un échantillon  $v_2$  dans une mixture avec une probabilité 10-90. Comparer les moyennes obtenues pour  $v_1$  et  $v_2$ .