

ISFA 2A Analyse de données

D. Chessel

Contrôle terminal - Mercredi 7 septembre 2005 - 8h 10h

Tous documents autorisés - Machine personnelle avec R disponible imposée.
Indiquer clairement sur la copie le numéro de la question à laquelle vous répondez et faites une réponse courte et motivée.

1 Éléments théoriques

1.1

Soient $\mathbf{x} = (1, 1, 0)$ et $\mathbf{y} = (2, 1, 4)$ deux vecteurs de \mathbb{R}^3 . Donner les coordonnées de \mathbf{z} , projeté orthogonal de \mathbf{y} sur \mathbf{x} pour la métrique canonique.

1.2

Vrai ou faux ? Choisir et justifier. Toute matrice carrée de nombre réels a au moins un vecteur propre.

1.3

Vrai ou faux ? Choisir et justifier. Toute matrice carrée de nombre réels a au moins un vecteur non nul qui n'est pas propre.

1.4

Vrai ou faux ? Choisir et justifier. Toute matrice carrée de nombre réels qui a des vecteurs propres est diagonalisable.

1.5

Vrai ou faux ? Choisir et justifier. Les valeurs propres d'une matrice de corrélation sont toujours positives ou nulles.

1.6

Vrai ou faux ? Choisir et justifier. Pour deux variables statistiques centrées et de variances non nulles, le nuage des points est sur une droite si et seulement si le coefficient de corrélation égale à 1.

1.7

Vrai ou faux ? Choisir et justifier. Le pourcentage de variance expliquée par la prédiction linéaire de \mathbf{x} par \mathbf{y} est toujours égal au pourcentage de variance expliquée par la prédiction linéaire de \mathbf{y} par \mathbf{x} .

1.8

Vrai ou faux ? Choisir et justifier. Un tableau \mathbf{X} a n lignes-individus et p colonnes-variables et \mathbf{C} est sa matrice de covariances. Si la somme par lignes dans \mathbf{X} vaut 1 pour toutes les lignes, 0 est toujours valeur propre de \mathbf{C} .

1.9

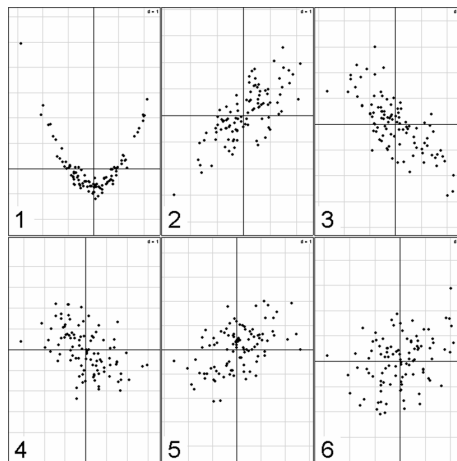
Vrai ou faux ? Choisir et justifier. Le rang de la matrice de corrélation d'une ACP normée sur p variables vaut toujours p .

1.10

Vrai ou faux ? Choisir et justifier. L'inertie projetée sur le premier axe principal d'une ACP normée sur p variables vaut au moins 1.

2 Éléments statistiques

2.1



Attribuer à chacune des six figures qui précèdent son coefficient de corrélation sachant que ces valeurs figurent parmi l'ensemble :

$$\{-1, -0.73, -0.49, -0.04, 0.33, 0.50, 0.74, 1\}$$

2.2

Soient $\mathbf{x} = (1, \sqrt{2}/2, 0, -1, -\sqrt{2}/2, 0)$ et $\mathbf{y} = (0, \sqrt{2}/2, 1, 0, -\sqrt{2}/2, -1)$ deux vecteurs de \mathbb{R}^6 . Considérer qu'on a affaire à 6 points de \mathbb{R}^2 et donner

l'inertie autour de l'origine du nuage de ces 6 points avec la pondération uniforme et la métrique canonique.

2.3 suite

Considérer que \mathbf{x} et \mathbf{y} sont deux variables sur 6 individus. Donner l'équation de la droite de régression de \mathbf{y} sur \mathbf{x} .

2.4 suite

Donner l'équation de la droite de régression de \mathbf{x} sur \mathbf{y} .

2.5 suite

Considérer que cette information donne un tableau à 6 lignes-individus et 2 colonnes-variables. Donner les axes principaux de l'ACP centrée de ce tableau.

2.6 suite

Donner les axes principaux de l'ACP normée de ce tableau. Indiquer pourquoi on n'a pas besoin de faire de calculs pour répondre.

2.7 suite

Donner les valeurs singulières de ce tableau.

2.8 suite

Sur une figure, placer les 6 points, leur centre de gravité, les deux droites de régression et les deux axes principaux.

2.9

Pourquoi un coefficient de corrélation est toujours compris entre -1 et $+1$?

2.10

A quoi sert un *boxplot* ?

3 Éléments pratiques

3.1

Comment fait-on afficher le dossier de travail dans une session du logiciel R ?

3.2

Quelle est la probabilité pour qu'une variable aléatoire dépasse la valeur 3.26 si elle suit une loi normale ?

3.3

Donner la valeur de l'expression en langage S :

```
!(1>2) & is.complex(sqrt(1i))
```

3.4

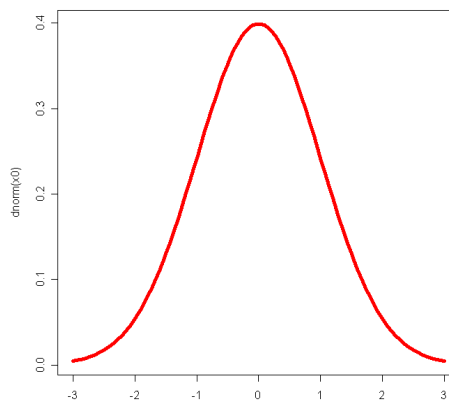
Que vaut le logarithme à base 22 de 3.5 ?

3.5

Quelle est la différence entre les deux graphes suivants ?

```
> plot(rep(1:5,rep(4,5)),rep(1:4,rep(5,4)))  
> sunflowerplot(rep(1:5,rep(4,5)),rep(1:4,rep(5,4)))
```

3.6



Peut-on reproduire ce dessin en une ligne de S ?

3.7

Combien de paramètres sont-ils utilisés par la fonction `dudi.pca` de la librairie `ade4` ?

3.8 suite

Parmi ces paramètres combien ont-ils une valeur par défaut ?

3.9

A quoi sert le paramètre `as.is` dans la fonction `read.table` ?

3.10

Expliquer le mécanisme qui donne ce genre de résultat :

```
> c(T,T,0,0,1)
[1] 1 1 0 0 1
> c(T,T,0,0,"a")
[1] "TRUE" "TRUE" "0"    "0"    "a"
```