réponses possibles.

Université de Nice
Département de Mathématiques
NOM :
PRENOM :
Machine n^o (SSSMM) : 21601

L1MASS, année 2015-2016
Statistique

Date : 16.3.2016

Feuille-réponse 07
Covariance et corrélation

1 Calcul à la main

On pose $x = (0, 2, 4, 6)$ et $y = (2, 6, 4, 0)$. Calculer la matrice de variance-covariance Σ des échantillons x et y, ainsi que leur matrice de corrélation R. Ne manquez pas de vérifier le résultat de vos calculs à l'aide de R. Représentez en marge le "nuage" (x, y) et la droite de régression $y = ax + b$.

Moyennes :

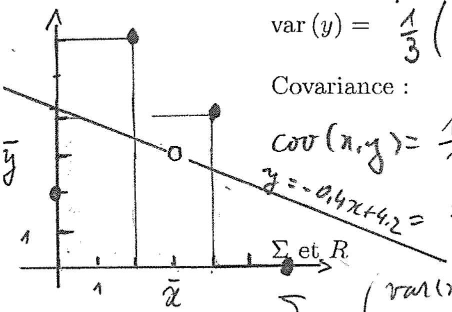
$\bar{x} = \frac{1}{4}(0+2+4+6) = \frac{1}{4}12 = 3$

$\bar{y} = \frac{1}{4}(2+6+4+0) = \frac{1}{4}12 = 3$

Variances :

var $(x) = \frac{1}{4-1}\left((0-3)^2+(2-3)^2+(4-3)^2+(6-3)^2\right) = \frac{1}{3}(9+1+1+9) = \frac{20}{3}$

var $(y) = \frac{1}{3}\left((2-3)^2+(6-3)^2+(4-3)^2+(0-3)^2\right) = \frac{20}{3}.$



Covariance :

$\operatorname{cov}(x,y) = \frac{1}{3}\Big((0-3)(2-3)+(2-3)(6-3)+(4-3)(4-3)+(6-3)(0-3)\Big) =$

$= \frac{1}{3}(+3-3+1-9) = -\frac{8}{3}.$

Σ et R

$\Sigma = \begin{pmatrix} \operatorname{var}(x) & \operatorname{cov}(x,y) \\ \operatorname{cov}(y,x) & \operatorname{var}(y) \end{pmatrix} = \begin{pmatrix} \frac{20}{3} & -\frac{8}{3} \\ -\frac{8}{3} & \frac{20}{3} \end{pmatrix};$ $\quad R = \begin{pmatrix} 1 & \operatorname{cor}(x,y) \\ \operatorname{cor}(x,y) & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{8}{3}\cdot\frac{3}{20} \\ -\frac{8}{3}\cdot\frac{3}{20} & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{2}{5} \\ -\frac{2}{5} & 1 \end{pmatrix}$

Equation de la droite de régression :

$a = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)} = -\frac{8}{3}\cdot\frac{3}{20} = -\frac{2}{5} = -0,4$

$b = \bar{y} - a\bar{x} = 3 + \frac{2}{5}3 = \frac{1}{5}(15+6) = \frac{21}{5} = \frac{42}{10} = 4,2$

l'équation de la droite de régression est donc

$y = -0.4x + 4,2$

2 Reprise du cours ~~06~~ 7

Former trois échantillons gaussiens simulés x, y, et z de taille 200, d'espérance respective 0, 2, et 2 et d'écart-type 1, 1, et 5. Afin de pouvoir reproduire exactement vos échantillons faites précéder vos trois appels de la fonction `rnorm` par l'instruction `set.seed(SSSMM)`, où SSSMM désigne le numéro de votre machine que vous aurez reproduit au début de votre copie, sous votre nom.

1. Former un `data.frame` appelé `MesDonnées` regroupant x, y, et z; quelles sont les trois premières lignes de `MesDonnées`

```
x = rnorm(200, 0, 1)
y = rnorm(200, 2, 1)
z = rnorm(200, 2, 5)
```

```
MesDonnées = data.frame(x, y, z)
head(MesDonnées)
```

	x	y	z
1	0,0561..	0.38..	0,46..
2	-0,6772..	0.76..	3,80..

2. Déterminer la moyenne et l'écart-type de vos trois échantillons x, y, et z. Ces nombre sont-ils égaux/proches de ceux que vous attendiez. Expliquez.

Pour x, y, et z on s'attend à des moyennes proches, respectivement, de 0, 2, et 2.

On trouve mean(x) = 0.006... proche (à $6\cdot10^{-3}$) de 0

mean(y) = 1.865278
mean(z) = 1.8684

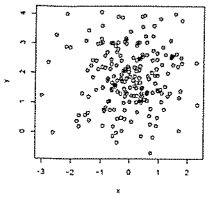
proches de 2, mais seulement à $4\cdot10^{-2}$

Il est surprenant que mean(y) = 1.86 soit beaucoup plus loin de 2 que mean(x) de 0 puisque les échantillons ont le même écart-type ...

3. Quelle est la matrice de variance-covariance Σ de vos échantillons x, y, et z (arrondir à 3 chiffres après le point décimal)

$$\Sigma = var(MesDonnées) = \begin{pmatrix} 0,957.. & 0,030.. & 0,046.. \\ 0,030.. & 1,024.. & -0,225.. \\ 0,046.. & -0,225.. & 21,882.. \end{pmatrix}$$

4. Faites tracer à l'écran puis esquissez (en marge de cette page) le nuage des x et y. Comment se manifeste que les échantillons x et y sont indépendants ?



plot(x,y)

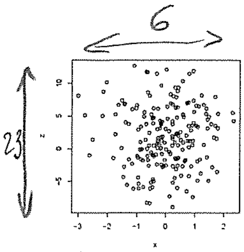
Je ne perçois aucune relation entre x et y

5. Comment se manifeste la différence de moyenne de x et y ?

Les abscisses (x) se distribuent autour de 0 alors que les ordonnées se distribuent autour de 2.

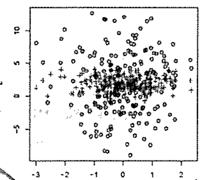
6. Comment se manifeste que ces échantillons suivent une loi normale ?

On perçoit une plus forte concentration autour de (0,2)



7. Faites tracer le nuage des x et z Comment se manifeste la différence d'écart-type de x et z ? Les deux nuages se présentent pourtant de façon similaire : commentez.

Les x sont compris entre -3 et +3 alors que les z vont de -10 à +13 , $(3-(-3)) = 6 << 23 = (13-(-10))$

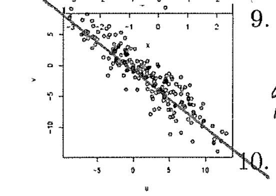


8. Faites *ajouter en rouge* (avec `points`) le nuage des x et y. Esquissez en marge de cette page la figure obtenue à l'écran.

points(x, y, col="red", pch=3)

les points sont représentés par des "+"

On voit que les y sont plus proches de leur moyenne que les z



9. On forme les échantillons u=x+z et v=x-z. Représenter le nuage des u et v. Qu'observez-vous ?

Je compare cette nouvelle figure avec la figure de la question 7. Je note qu'à présent les abscisses u ont également une amplitude de l'ordre de 23. De ce fait, ce nouveau nuage apparaît comme allongé le long d'une droite "descendante"

10. Former le `data.frame` MD2 des u et v. Donner la matrice de variance-covariance $\Sigma(u,v)$ des u et v et leur matrice de corrélation $R(u,v)$. Que vaut $\rho = cor(u,v)$? Commentez.

MD2 = data.frame(u,v).

$$\Sigma(u,v) = var(MD2) = \begin{pmatrix} 22,93.. & -20,92.. \\ -20,92.. & 22,74.. \end{pmatrix}$$

$\rho = cor(u,v) = -0,916...$

ρ est négatif (droite descendante)

$|\rho|$ est proche de 1 (nuage bien allongé)

11. Quelle est l'équation $v = au + b$ de la droite de régression des v sur les u ? (Indication : utiliser `lm` et `coef`). Ajoutez cette droite à votre croquis de la question 9.

Reg v sur u = lm(v ~ u)

coef(Reg v sur u) donne a = -0,912..

b = -0,17..

abline(Reg v sur u).

$v = -0,912..u - 0,17..$

12. Comment peut-on retrouver la pente a de la droite de régression à partir de la matrice de variance-covariance ? Vérifiez que vous retrouvez bien le même résultat que celui donné par R.

On a vu en cours que $a = \frac{cov(u,v)}{var(u)} = \frac{-20,92..}{22,93..} = -0,9124..$

= idem question 11

a <- cov(u,v) / var(u)

13. Comment peut-on retrouver l'ordonnée à l'origine b. Vérifiez que vous retrouvez bien le même résultat que celui donné par R.

On a vu en cours que $b = \bar{v} - a\bar{u}$ = mean(v) - a*mean(u)

= -0,1762.. = idem question 11.