Travail SCILAB n° 4 - Calculs de covariance et modèle de régression linéaire

Il existe plusieurs commande permettant de calculer la covariance d'un couple de variables aléatoires finies (contexte probabiliste) ou de deux échantillons de données (contexte statistiques) :
• Si `x` et `y` sont les vecteurs de même dimension représentant deux échantillons stastitiques `x`$= (x_1, \dots, x_n)$ et `y`$= (y_1, \dots, y_n)$ de même taille alors :

`corr(x,y,1)`

permet le calcul de la covariance empirique

$$cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i y_i \;-\; \overline{xy}$$

• Si `x` et `y` sont des vecteurs réprésentant les valeurs possibles correspondant à deux variables aléatoires finies X et Y et si `matrix` est la matrice (de taille (taille de `x`) lignes et (taille de `y`) colonnes) donnant la loi jointe du couple (X, Y) alors

`covar(x,y,matrix)`
`correl(x,y,matrix)`

permettent les calculs de la covariance empirique et du coefficient de corrélation :

$$cov(X, Y) = E(XY) - E(X)E(Y) \qquad \rho(X, Y) = \frac{cov(X, Y)}{V(X)V(Y)}$$

☛ Attention la commande `variance(x)` ne calcule pas la variance de la série x mais une quantité voisine

$$\frac{n}{n-1}\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

Afin d'obtenir la variance d'une série statistique x, l'on peut utiliser la commande `corr(x,x,1)`.
De la même manière la commande `st deviation`(x) retourne la racine de la valeur obtenue par `x` soit :

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

valeur voisine de l'écart-type de x.

Exercice 1 - Covariance d'un couple de variables aléatoires discrètes

1. Calculer, grâce à la commande `covar` de SCILAB, la covariance du couple (X, Y) dont voici la loi jointe

$X\backslash Y$	-3	3	4	9
2	1/8	1/9	1/6	1/12
4	1/8	1/9	1/12	0
6	0	1/9	0	1/12

2. Compléter la fonction programmée dans le fichier cov.sci (à télécharger sur la page algorthmique du site math de la classe) qui demande en entrée une valeur entière $n \geqslant 2$, et qui affiche en sortie la covariance du couple de variables $(X,\, Y)$ sont la loi est définie dans l'exercice 5 de la feuille de TD - **Loi de couples de variables finies - Calculs de covariance**

Exercice 2 - Nuage de points et équations des droites de régression.

1. Repcopier et commenter (en complétant les pointillés) les actions liées à chacune des lignes de commande suivantes dans la console SCILAB et interpréter le rendu observé :

```
x=rand(1,50);
y=-2*x+3+0.3*rand(x,'n');
plot2d(x,y,-9) ;
mx=mean(x);my=mean(y);
plot(mx,my,'or');         //.........................................
cov=corr(x,y,1);          //..........................................
vx=corr(x,x,1);           //...........................................
t=[min(x),max(x)];
z=my+cov*(t-mx)/vx;
plot2d(t,z,style=5);       //...........................................
```

2. Compléter les lignes suivantes afin que celles-ci permettent le tracé de la droite de regression de x en y

```
vy=..................... ;           // variance de la variable y
u=[min(y),max(y)];           // vecteur des ordonnées de 2 points à relier
                                           pour former la dte de regression de x en y
v=............................ ;    // vecteur des abscisses de 2 points à relier
                                           pour former la dte de regression de x en y
plot2d(v,u,style=3);
```

Exercice 3 - Droite de régression et commandes `regress` et `reglin`

Scilab possède plusieurs commandes permettant de calculer les coefficients de régression et même de tracer la droite de régression sans avoir à recourir à ces formules. Les fonctions `regress` et `reglin`.

☛ La fonction `reglin` permet d'obtenir de la variance des résidus $\left(e_i = y_i - (\widehat{(a)}x_i + \widehat{b})\right)_{1\leqslant i\leqslant n}$

```
coeffs=regress(x,y)
[a,b]=reglin(x,y)
```

1. Prise en main des commandes `regress` et `reglin`. Tapez puis observer les résultats obtenus après validation des commandes suivantes :

```
x=[ 1 : 5] ;
 y=[-2 , -1 , 0 , 1 , 2] ;
R = regress(x,y)
[a,b,res] = reglin(y,x)
```

```
x=[ 1 : 5] ;
 y=[-1.9 , -1.1 , 0.2 , 1.3 , 1.98] ;
R = regress(x,y)
[a,b,res] = reglin(y,x)
u=[max(x),min(x)]
v=R(1)+R(2)*u
plot2d(u,v,style=3)
```

2. **Représentation et calculs sur données simulées**
   (a) Simuler un échantillon x=$[x_1, \dots, x_{20}]$ pour lequel les $x_i = i + u_i$ lorsque u_i est la réalisation d'une variable uniforme sur l'intervalle $\left[-\frac{1}{2}; \frac{1}{2}\right]$.
   (b) Simuler un échantillon z=$[z_1, \dots, z_{20}]$ pour lequel les z_i est la réalisation d'une variable normale centrée et de variance $\frac{1}{4}$.
   (c) Créer le vecteur $y = x + z$ puis dessiner le nuage de points associé au couple (x, y).
   (d) A l'aide des indications ci-dessus déterminer l'équation puis tracer la droite de régression de y en x.
   (e) A l'aide des indications ci-dessus déterminer l'équation puis tracer la droite de régression de x en y.