ED N° 2 : Statistique bivariée

EXERCICE 1

Soient X et Y deux variables quantitatives mesurées sur les mêmes individus.

- 1) Rappeler les formules de la covariance de X et Y et du coefficient de corrélation linéaire entre X et Y.
- 2) On pose Z = aX+b et T = cY+d. Calculer le coefficient de corrélation linéaire de Z et T

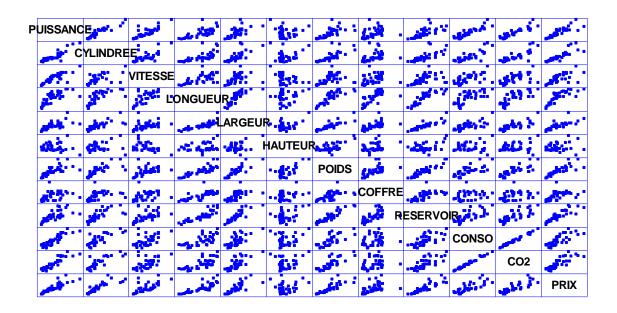
EXERCICE 2

- 1) Un nuage de points présente une symétrie par rapport à l'axe Oy. Montrer que le coefficient de corrélation linéaire entre X et Y calculé sur ces points est nul.
- 2) Peut-on généraliser ce résultat à un nuage de points présentant une symétrie par rapport à :
 - Un axe parallèle à Oy
 - Un axe parallèle à Ox
 - Un axe oblique?

EXERCICE 3

- 1) Commentez les nuages de points ci-dessous. Indiquez pour quels couples de variables la corrélation est proche de 1, de 0.
- 2) Rappelez à partir de quel seuil (pour un risque d'erreur 5%) on peut considérer une corrélation calculée sur un échantillon de taille 40, comme significativement différente de 0.
- 3) Indiquer un couple de variables non corrélées linéairement.
- 4) Ecrire la matrice de corrélation des variables puissance, vitesse hauteur.

NUAGES DE POINTS (Fichier autos2005)



Corrélations

	PUISSANCE	CYLINDREE	VITESSE	LONGUEUR	LARGEUR
PUISSANCE		0,8862	0,8041	0,7106	0,6647
		(40)	(40)	(40)	(40)
		0,0000	0,0000	0,0000	0,0000
CYLINDREE	0,8862		0,5583	0,6594	0,7324
	(40)		(40)	(40)	(40)
	0,0000		0,0002	0,0000	0,0000
VITESSE	0,8041	0,5583		0,6090	0,3615
	(40)	(40)		(40)	(40)
	0,0000	0,0002		0,0000	0,0219
LONGUEUR	0,7106	0,6594	0,6090		0,8154
	(40)	(40)	(40)		(40)
	0,0000	0,0000	0,0000		0,0000
LARGEUR	0,6647	0,7324	0,3615	0,8154	
	(40)	(40)	(40)	(40)	
	0,0000	0,0000	0,0219	0,0000	
HAUTEUR	0,0025	0,2109	-0,4450	0,1820	0,4231
	(40)	(40)	(40)	(40)	(40)
	0,9879	0,1915	0,0040	0,2609	0,0065
POIDS	0,6832	0,7409	0,3261	0,8192	0,8509
	(40)	(40)	(40)	(40)	(40)
	0,0000	0,0000	0,0401	0,0000	0,0000
COFFRE	0,3230	0,3408	0,3035	0,7062	0,5895
	(40)	(40)	(40)	(40)	(40)
	0,0420	0,0314	0,0569	0,0000	0,0001
RESERVOIR	0,6963	0,7283	0,4325	0,8637	0,8684
	(40)	(40)	(40)	(40)	(40)
	0,0000	0,0000	0,0053	0,0000	0,0000

CONSO	0,8905	0,7863	0,5808	0,6644	0,6895	
	(40)	(40)	(40)	(40)	(40)	
	0,0000	0,0000	0,0001	0,0000	0,0000	
CO2	0,9024	0,8100	0,5803	0,7140	0,7259	
	(40)	(40)	(40)	(40)	(40)	
	0,0000	0,0000	0,0001	0,0000	0,0000	
PRIX	0,8647	0,8516	0,6457	0,7873	0,7652	
	(40)	(40)	(40)	(40)	(40)	
	0,0000	0,0000	0,0000	0,0000	0,0000	

Exercice 4:

Rappeler la difference entre le coefficient de corrélation de Pearson et le coefficient de corrélation de Spearman. Les calculer pour les cas suivants :

1.

X	1	2	3	4	5
Y	1	2	4	8	16

2.

X	1	10	100	1000	10000
Y	0	1	2	3	4

3.

X	1	1	2	3	3	50
Y	1	3	2	1	3	50

Exercice 5:

Certaines stations de sports d'hiver ont obtenu le label "France ski de fond". La répartition entre les stations ayant le label et n'ayant pas le label est-elle liée à la région de la station de sports d'hiver?

Répartition des stations labelisées "France ski de fond"

Effectif	Labelisée	Non labelisée
Pyrénées	1	6
Vosges	3	2
Jura	2	1
Auvergne	3	0
Alpes du sud	9	7
Haute Savoie	8	11
Savoie	5	27
Dauphiné Isère	7	6