



Cours Statistiques descriptives



Auteur : C. Terrier ; <mailto:webmaster@cterrier.com> ; <http://www.cterrier.com>

Utilisation : Reproduction libre pour des formateurs dans un cadre pédagogique et non commercial

2 – Série statistique à deux variables et corrélation

Une série statistique à deux variables décrit deux caractères qui évoluent parallèlement. Il est parfois intéressant d'étudier la relation ou la non relation susceptible d'exister entre ces deux variables.

Exemple 1 : Question : Existe-t-il une relation entre le poids d'une population et sa taille ?

Réponse : Il paraît logique de penser qu'il existe une relation forte entre ces deux caractéristiques.

Exemple 2 : Question : Existe-t-il une relation entre les investissements publicitaires d'une entreprise et l'évolution de son chiffre d'affaires ?

Réponse : On peut penser qu'il existe une relation plus ou moins forte selon l'efficacité de la publicité. (Une bonne pub => une augmentation du CA ; une mauvaise pub n'a pas d'effet sur le CA).

Exemple 3 : Question : Existe-t-il une relation entre les dépenses d'un ménage et la couleur des yeux du couple ?

Réponse : Il est peu vraisemblable qu'il existe une relation entre ces deux caractéristiques.

L'étude de la corrélation entre deux séries permet d'identifier la dépendance ou l'indépendance qui existe entre les deux séries. Ce degré de dépendance peut être vérifié en calculant le **coefficient de corrélation** ou vérifié par les **droites d'ajustement**.

□ Coefficient de corrélation

Le coefficient de corrélation est un nombre qui indique la plus ou moins grande dépendance entre deux séries statistiques

Formule de calcul:
$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \times \sum Y_i^2}} \quad X_i = x_i - \bar{x} \quad Y_i = y_i - \bar{y}$$

Plus le coefficient se rapproche de 1 ou de -1 plus la corrélation est forte

Exercice illustré :

On vous demande s'il existe une corrélation entre la taille et le poids des athlètes d'un club d'escalade. Les données individuelles sont indiquées dans le tableau suivant :

Athlète	A	B	C	D	E	F	G	H	I	J
Taille	1,78	1,65	1,82	1,68	1,75	1,78	1,90	1,60	1,72	1,80
Poids	82	63	90	62	69	72	95	56	68	84

Calculer le coefficient de corrélation entre ces deux séries

Formule de calcul :
$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \times \sum Y_i^2}} \quad X_i = x_i - \bar{x} \quad Y_i = y_i - \bar{y}$$

	Taille xi	Poids yi	Xi	Yi	Xi Yi	Xi ²	Yi ²
	1,78	82	0,032	7,90	0,2528	0,001024	62,41
	1,65	63	- 0,098	- 11,10	1,0878	0,009604	123,21
	1,82	90	0,072	15,90	1,1448	0,005184	252,81
	1,68	62	- 0,068	- 12,10	0,8228	0,004624	146,41
	1,75	69	0,002	- 5,10	- 0,0102	0,000004	26,01
	1,78	72	0,032	- 2,10	- 0,0672	0,001024	4,41
	1,90	95	0,152	20,90	3,1768	0,023104	436,81
	1,60	56	- 0,148	- 18,10	2,6788	0,021904	327,61
	1,72	68	- 0,028	- 6,10	0,1708	0,000784	37,21
	1,80	84	0,052	9,90	0,5148	0,002704	98,01
Moy	1,75	74,10		Total	9,7720	0,06996	1 514,90

coef corrélation = $\frac{9,7720}{\sqrt{0,06996 \cdot 1514,90}} = \frac{9,7720}{\sqrt{10,294776}} = 0,949219 \Rightarrow$ La corrélation est forte

□ Droite d'ajustement (ou de régression des moindres carrés)

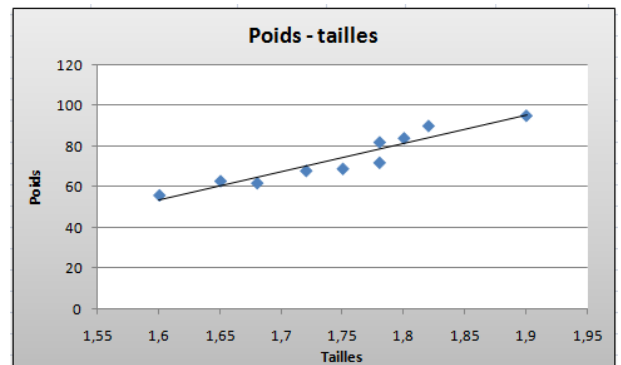
Nous avons vu dans un chapitre précédent qu'il était possible de calculer la droite d'ajustement d'une série de données. De la même façon il est possible de calculer la droite d'ajustement d'une série à 2 variables. **Attention cette dernière n'a de sens que s'il existe une corrélation entre les deux séries de données.**

Rappel des formules :

Droite de la forme $y = ax + b$ de telle sorte que la somme des écarts positifs à la droite soit égale à la somme des écarts négatifs

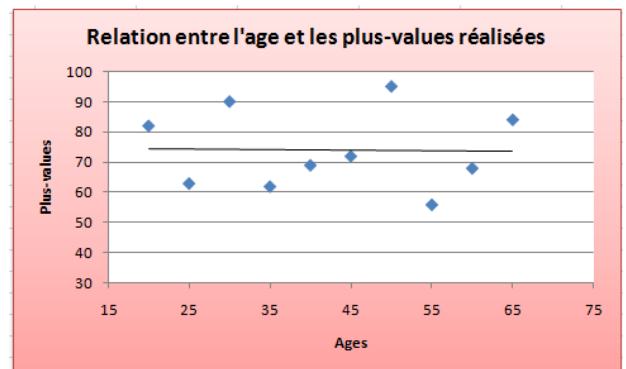
$$a = \frac{\sum Xi Yi}{\sum Xi^2} \quad b = \bar{y} - a \bar{x} \quad Xi = xi - \bar{x} \quad Yi = yi - \bar{y}$$

La droite tracée se présente ainsi, la corrélation de 0,949 est forte et les points sont resserrés autour de la courbe de tendance.



Dans l'exemple suivant nous recherchons la corrélation entre l'âge d'investisseurs et les plus-values réalisées. La droite tracée correspond à une corrélation faible, et les points sont dispersés autour de la droite :

Ages	Plus-values
20	82
25	63
30	90
35	62
40	69
45	72
50	95
55	56
60	68
65	84

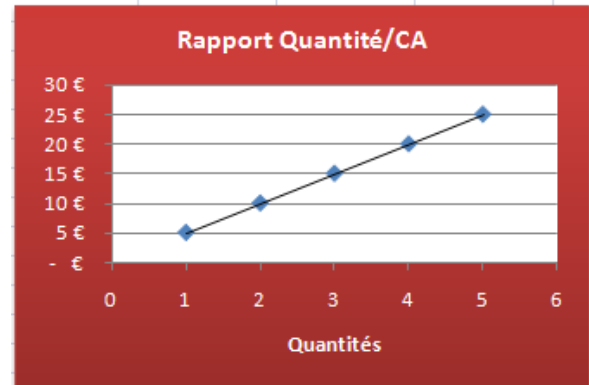


Un coefficient de corrélation de +1 correspond à une courbe de tendance ascendante et une corrélation négative à une courbe descendante.

Exemple 1 :

Nous étudions la corrélation entre le prix de vente d'un article et le chiffre d'affaires réalisé. Un article vendu génère 5 € de CA et 5 articles génèrent 25 € de CA. La corrélation est total le coefficient de +1 => La courbe est ascendante :

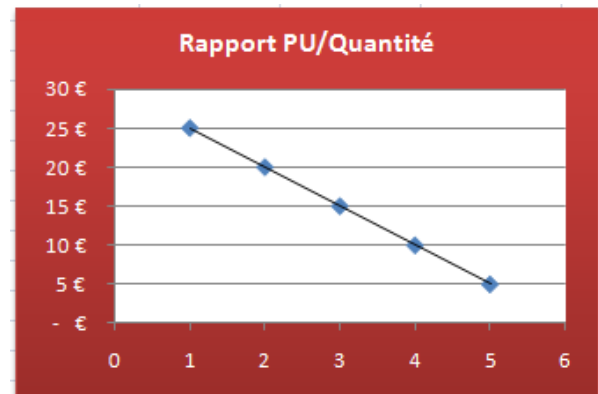
Quantité	CA
1	5 €
2	10 €
3	15 €
4	20 €
5	25 €



Exemple 2 :

Nous étudions la corrélation entre le nombre d'articles acheté et son prix d'achat. Plus la quantité achetée est importante et plus le prix d'achat unitaire est faible. La corrélation est total le coefficient de -1 => La courbe est descendante :

Quantité	PU
1	25 €
2	20 €
3	15 €
4	10 €
5	5 €



Exercice 8

On vous demande s'il existe une corrélation entre la population de chamois d'une commune des Bauges et le nombre de permis de chasse enregistré par l'association de chasse locale.

Années	2005	2006	2007	2008	2009	2010
Chamois	3200	3650	3430	3890	4200	4350
Permis	202	231	240	225	245	263

Travail à faire :

- Calculer le coefficient de corrélation entre ces deux séries
- Tracer la droite d'ajustement

Exercice 9

On vous demande s'il existe une corrélation l'évolution du prix des actions et l'évolution du prix des obligations

Années	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Actions	352	360	358	361	366	382	398	406	450	445
Obligations	1024	998	980	970	982	972	935	902	895	900

Travail à faire :

- Calculer le coefficient de corrélation entre ces deux séries
- Tracer la droite d'ajustement

Exercice 10 de synthèse N° 1

La société **Pol-Arctique** est spécialisée dans la commercialisation de vêtements de sport.

On vous communique ci-dessous le CA des 6 dernières années et les dépenses de publicité

		CA	Publicité
2004	T1	35 000 €	3 500 €
	T2	15 000 €	1 600 €
	T3	5 000 €	400 €
	T4	45 000 €	5 000 €
2005	T1	42 000 €	4 000 €
	T2	15 000 €	2 000 €
	T3	7 000 €	800 €
	T4	58 000 €	5 000 €
2006	T1	57 000 €	5 000 €
	T2	25 500 €	2 000 €
	T3	15 000 €	500 €
	T4	52 500 €	6 000 €
2007	T1	58 800 €	6 000 €
	T2	19 600 €	1 000 €
	T3	16 800 €	1 000 €
	T4	44 800 €	4 000 €
2008	T1	74 100 €	7 000 €
	T2	25 350 €	2 000 €
	T3	23 400 €	2 500 €
	T4	72 150 €	700 €
2009	T1	71 440 €	8 000 €
	T2	28 200 €	2 500 €
	T3	15 040 €	1 000 €
	T4	73 320 €	8 000 €

TRAVAIL A FAIRE

1. Calculer le CA Prévisionnel de 2010
2. Calculer la répartition par trimestre du CA 2010
3. Existe-t-il une corrélation entre les dépenses de publicité et le CA ?