Analyse de corrélation

Caractériser la liaison entre deux variables X et Y

Ricco Rakotomalala
Université Lumière Lyon 2

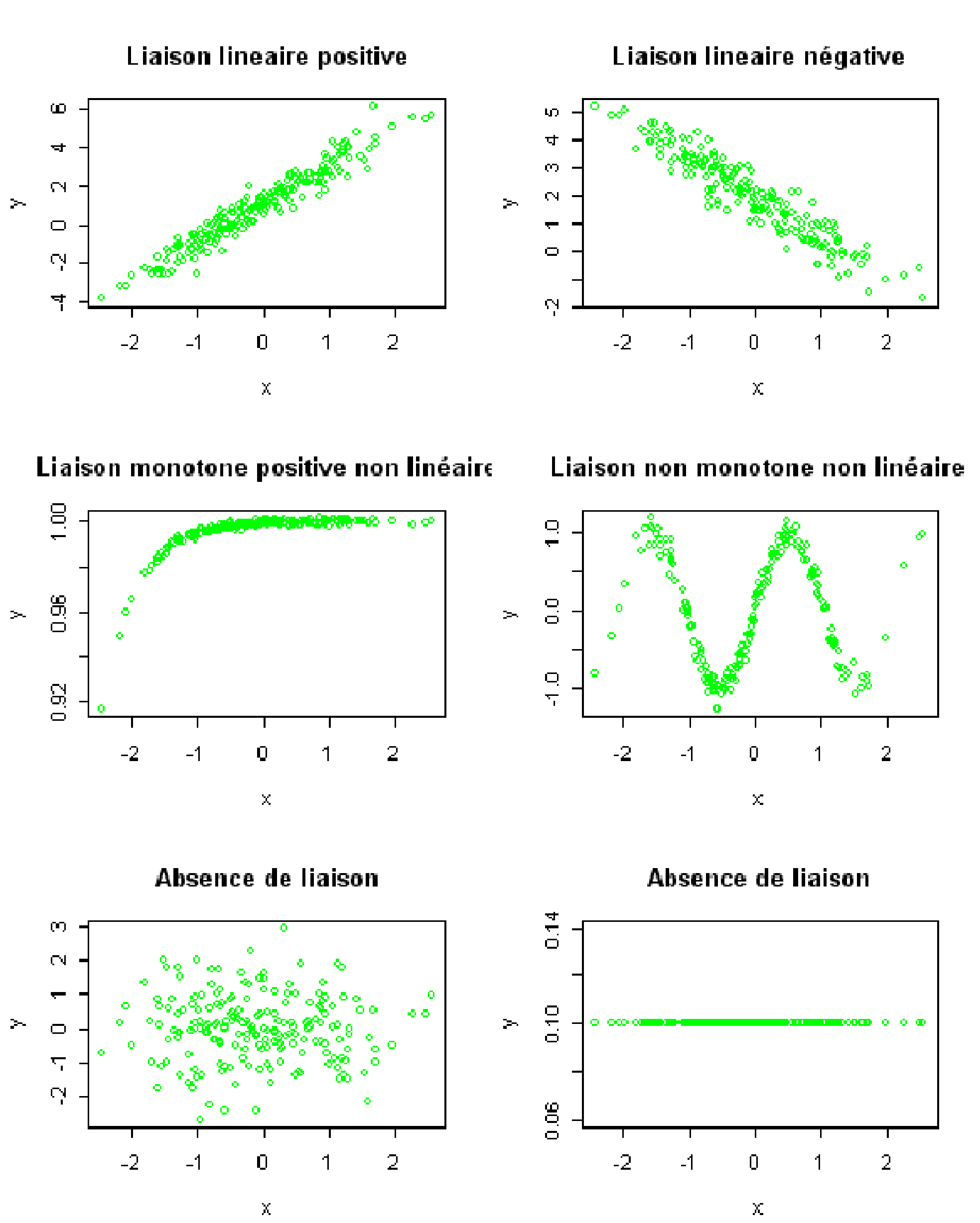
1. Etudier la liaison entre deux variables : Introduction

Position du problème

X et Y sont deux grandeurs statistiques quantitatives, on souhaite :

- déterminer l'existence d'une relation (liaison) entre X et Y ;
- caractériser la forme de le relation ;
- quantifier l'intensité de la liaison si elle existe.

Remarque : la position des variables est symétrique, on ne cherche pas à savoir si l'une détermine l'autre ou pas (ce sera le rôle de la régression)



Analyse graphique

Deux points de vue :

• en termes d'évolution (quand X augmente, Y augmente ou diminue ; quelques mots clés : monotonie, linéarité, sens +/-)

• en termes de positionnement (quand X est élevé, Y est élevé ou faible ; élevé par rapport à quoi ?)

Notations

Variable : Notée en majuscule (X est une variable)

Valeur : Valeur mesurée chez un individu i (x_i) ou à la date t (x_t)

Population : Ω^{pop}

Echantillon observé : Ω (fraction de la population)

Taille de l'échantillon : n = card(Ω)

Pour la corrélation : couples de points $\Omega = \{(x_i,y_i), i=1,\ldots,n\}$

Deux indicateurs usuels →
$$\begin{cases} \bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i \\ s_x^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

2. Covariance et corrélation

Covariance : définition, interprétation

Espérance du produit des variables centrées

Définition :

$$COV(X,Y) = E\{(X - E[X])(Y - E[Y])\}$$
$$= E[XY] - E[X]E[Y]$$

Signification :

- Mesurer la tendance des deux variables à être simultanément au dessus ou en dessous de leurs espérances respectives.
- La référence est l'espérance mathématique des variables (moyenne)
- Caractérise les liaisons monotones et linéaires
- Fournit une indication sur le sens de la liaison :
COV(X,Y) > 0, liaison positive ; COV(X,Y) < 0,liaison négative
- Et sur son intensité : plus |COV| est élevé, plus forte est la liaison
- COV(X,X) = V(X)

Covariance : propriétés

- **Symétrie** : COV(X,Y) = COV(Y,X)
- **Distributivité** : COV(X,Y+Z) = COV(X,Y) + COV(X,Z)
- **Covariance avec une constante** : COV(X,a) = 0
- **Covariance avec une variable transformée** : COV(X,a+b.Y) = b.COV(X,Y)
- **Variance de la somme de 2 variables** : V(X+Y) = V(X) + V(Y) + 2.COV(X,Y)
- **Covariance de deux variables indépendantes** : COV(X,Y) = 0

Démonstrations à faire en exercice par la suite….

Domaine de définition : $-\infty < COV < +\infty$

→ Elle est non normalisée !!!

Covariance : estimation sur un échantillon

Covariance empirique : $$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

C'est un estimateur biaisé : $$E[S_{xy}] = \frac{n-1}{n} COV(X,Y)$$

Ecriture simplifiée : $$S_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}.\bar{y}$$

Estimateur corrigé (sans biais) : $$\hat{COV}(X,Y) = \frac{n}{n-1} S_{xy}$$

Jamais utilisé

Covariance : estimation sous Excel

Numero	Modele	Cylindree	Puissance	XY
1	Daihatsu Cuore	846	32	27072
2	Suzuki Swift 1.0 GLS	993	39	38727
3	Fiat Panda Mambo L	899	29	26071
4	VW Polo 1.4 60	1390	44	61160
5	Opel Corsa 1.2i Eco	1195	33	39435
6	Subaru Vivio 4WD	658	32	21056
7	Toyota Corolla	1331	55	73205
8	Opel Astra 1.6i 16V	1597	74	118178
9	Peugeot 306 XS 108	1761	74	130314
10	Renault Safrane 2.2. V	2165	101	218665
11	Seat Ibiza 2.0 GTI	1983	85	168555
12	VW Golt 2.0 GTI	1984	85	168640
13	Citroen ZX Volcane	1998	89	177822
14	Fiat Tempra 1.6 Liberty	1580	65	102700
15	Fort Escort 1.4i PT	1390	54	75060
16	Honda Civic Joker 1.4	1396	66	92136
17	Volvo 850 2.5	2435	106	258110
18	Ford Fiesta 1.2 Zetec	1242	55	68310
19	Hyundai Sonata 3000	2972	107	318004
20	Lancia K 3.0 LS	2958	150	443700
21	Mazda Hachtback V	2497	122	304634
22	Mitsubishi Galant	1998	66	131868
23	Opel Omega 2.5i V6	2496	125	312000
24	Peugeot 806 2.0	1998	89	177822
25	Nissan Primera 2.0	1997	92	183724
26	Seat Alhambra 2.0	1984	85	168640
27	Toyota Previa salon	2438	97	236486
28	Volvo 960 Kombi aut	2473	125	309125
n		**Moyenne**		**Somme**
28		1809.07	77.71	4451219

Cov.Empirique	**18381.4133**
Cov.Non-Biaisé	**19062.2063**

Cov.Excel	**18381.4133**

Coefficient de corrélation de Pearson (Bravais-Pearson)

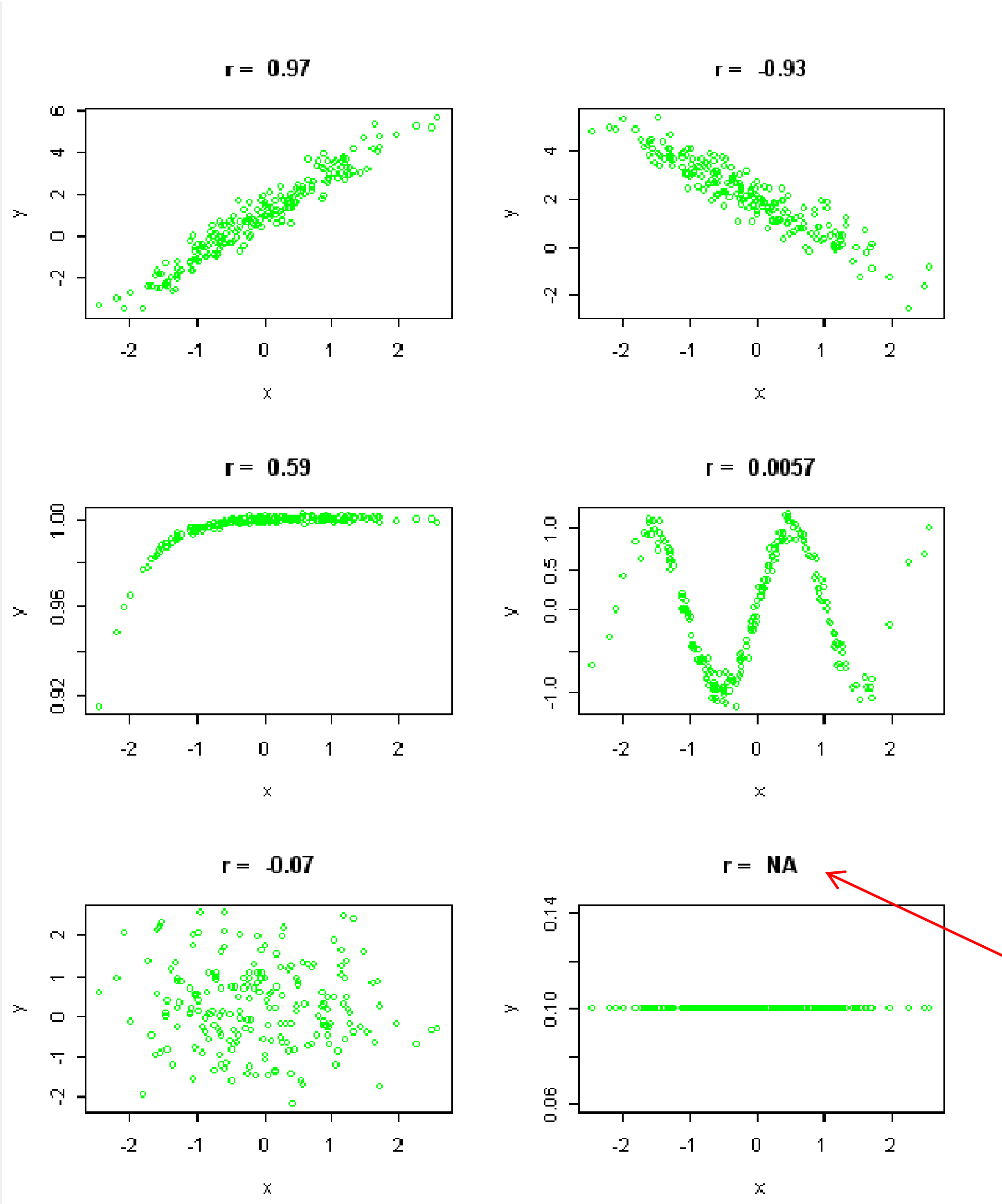
Normalisation de la covariance par le produit des écarts type.

Définition : $$r_{xy} = \frac{COV(X,Y)}{\sqrt{V(X).V(Y)}} = \frac{COV(X,Y)}{\sigma_x.\sigma_y}$$

Première conséquence : domaine de définition $-1 \leq r_{xy} \leq +1$

Commentaires :

- Mesure l'intensité de la liaison monotone linéaire entre 2 variables
- (X,Y) indépendants → r = 0 (la réciproque est en général fausse)
- Corrélation d'une variable avec elle-même : $r_{xx} = 1$
- Corrélation = Covariance pour les variables réduites = Espérance du produit des variables centrées et réduites



Corrélation : quelques exemples

Notions à voir : monotonie, linéarité, absence de liaison

Pourquoi le calcul n'est pas possible ici ?

Corrélation : estimation sur un échantillon

Corrélation empirique :

$$\hat{r} = \frac{S_{xy}}{s_x . s_y}$$

C'est un estimateur asymptotiquement sans biais

$$E[\hat{r}] = r - \frac{r(1-r^2)}{2n}$$

Le biais est très faible dès que n augmente…

Un estimateur non biaisé

$$\hat{r}_{aj} = \sqrt{1 - \frac{n-1}{n-2}\left(1-\hat{r}^2\right)}$$

Très peu utilisé en pratique, la correction est très minime, négligeable dès que n augmente

Corrélation : un exemple sous Excel

$$\hat{r} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \times \sqrt{\sum_i (y_i - \bar{y})^2}}$$

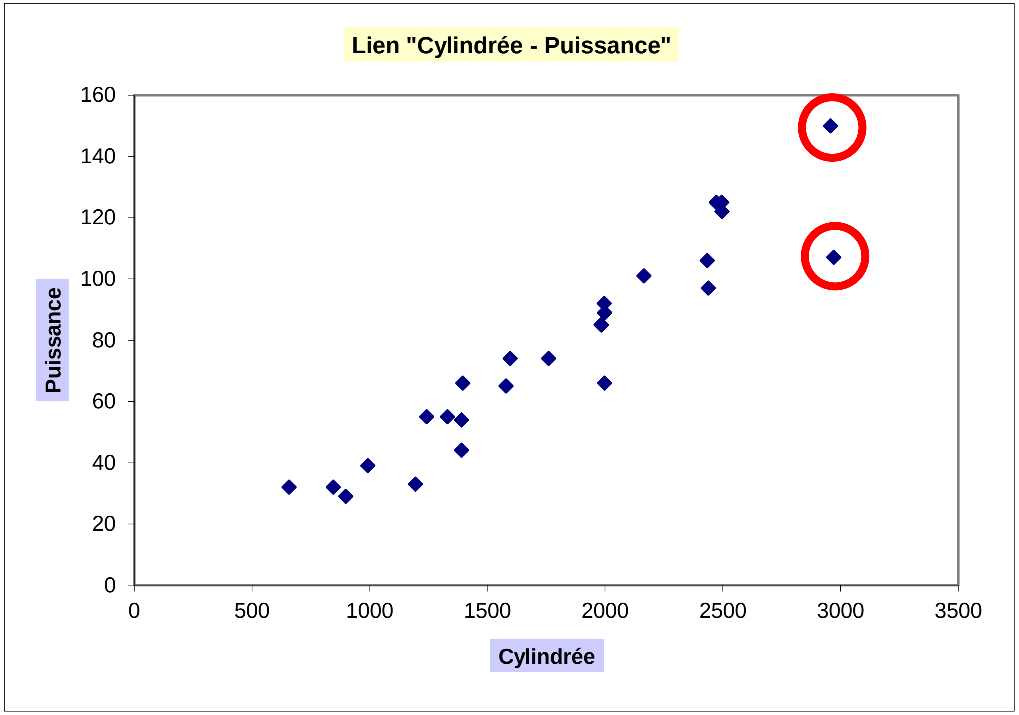
$$= \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_i x_i^2 - n\bar{x}^2} \times \sqrt{\sum_i y_i^2 - n\bar{y}^2}}$$

Numero	Modele	Cylindree	Puissance	XY	X²	Y²
1	Daihatsu Cuore	846	32	27072	715716	1024
2	Suzuki Swift 1.0 GLS	993	39	38727	986049	1521
3	Fiat Panda Mambo L	899	29	26071	808201	841
4	VW Polo 1.4 60	1390	44	61160	1932100	1936
5	Opel Corsa 1.2i Eco	1195	33	39435	1428025	1089
6	Subaru Vivio 4WD	658	32	21056	432964	1024
7	Toyota Corolla	1331	55	73205	1771561	3025
8	Opel Astra 1.6i 16V	1597	74	118178	2550409	5476
9	Peugeot 306 XS 108	1761	74	130314	3101121	5476
10	Renault Safrane 2.2. V	2165	101	218665	4687225	10201
11	Seat Ibiza 2.0 GTI	1983	85	168555	3932289	7225
12	VW Golt 2.0 GTI	1984	85	168640	3936256	7225
13	Citroen ZX Volcane	1998	89	177822	3992004	7921
14	Fiat Tempra 1.6 Liberty	1580	65	102700	2496400	4225
15	Fort Escort 1.4i PT	1390	54	75060	1932100	2916
16	Honda Civic Joker 1.4	1396	66	92136	1948816	4356
17	Volvo 850 2.5	2435	106	258110	5929225	11236
18	Ford Fiesta 1.2 Zetec	1242	55	68310	1542564	3025
19	Hyundai Sonata 3000	2972	107	318004	8832784	11449
20	Lancia K 3.0 LS	2958	150	443700	8749764	22500
21	Mazda Hachtback V	2497	122	304634	6235009	14884
22	Mitsubishi Galant	1998	66	131868	3992004	4356
23	Opel Omega 2.5i V6	2496	125	312000	6230016	15625
24	Peugeot 806 2.0	1998	89	177822	3992004	7921
25	Nissan Primera 2.0	1997	92	183724	3988009	8464
26	Seat Alhambra 2.0	1984	85	168640	3936256	7225
27	Toyota Previa salon	2438	97	236486	5943844	9409
28	Volvo 960 Kombi aut	2473	125	309125	6115729	15625
n		**Moyenne**			**Somme**	
28		1809.07	77.71	4451219	102138444	197200

Numérateur	514679.571
Dénominateur	543169.291
Corrélation	**0.9475**

Coef.Corr.Excel	0.9475

Corrélation : analyse graphique



Un indicateur numérique ne fait pas tout, l’analyse graphique est un complément indispensable (ex. pour repérer les situations atypiques, insolites)

3. Test de significativité

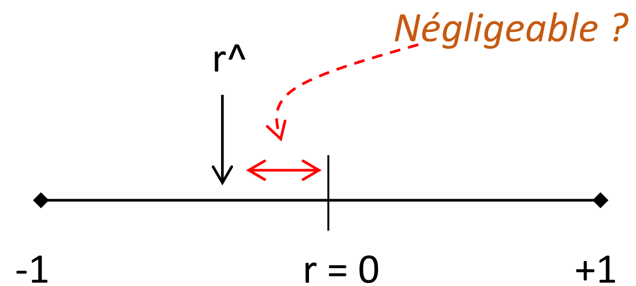
Test de significativité

Tester l'existence d'un lien

linéaire entre X et Y

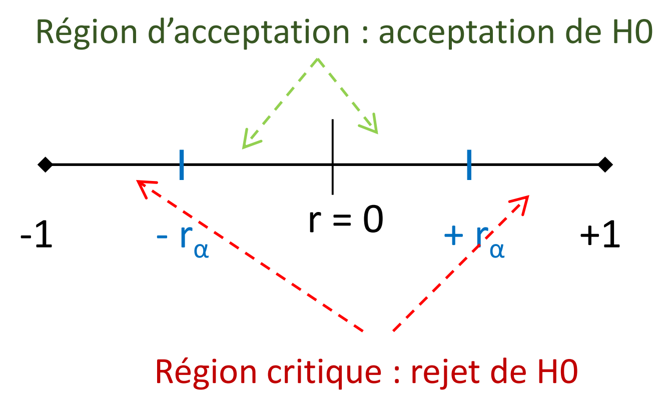
$$\begin{cases} H_0 : r = 0 \\ H_1 : r \neq 0 \end{cases}$$

Attention : (X,Y) indépendants → r = 0 ; mais r = 0 ne veut pas dire que (X,Y) indépendants, ils ne sont pas liés linéairement simplement !!!



Comment procéder ?

- On dispose d'une estimation de r (r^).
- On cherche à savoir si r^ s'éloigne *significativement* de 0.
- Pour définir les « seuils » autour de 0, on fixe (contrôle) la probabilité de conclure à tort α = P(Rejeter H0 au vu de r^ / en réalité H0 est vrai c.-à-d. r = 0) → seuil critique : r_α
- Problème : il faut connaître la loi de distribution de r^ quand H0 est vrai



Test de significativité – Test de Student

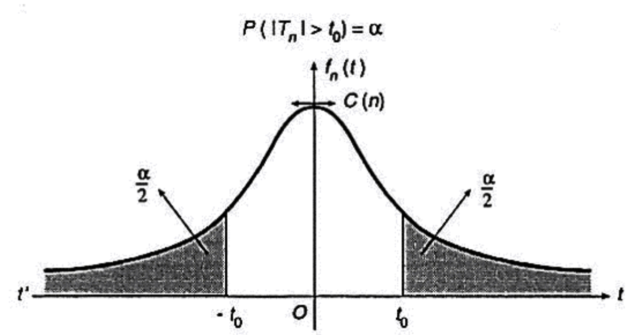
Idée : Sous H0, on ne connaît pas la loi de r^, en revanche on connaît celle d'une transformation de r^

$$t = \frac{\hat{r}}{\sqrt{\frac{1-\hat{r}^2}{n-2}}} \equiv \Im(n-2)$$

La règle de décision devient :

$$\text{Accepter } \mathrm{H}_0 (r = 0) \text{ ssi } |t| < t_{1-\alpha/2}$$

$$\text{Rejeter } \mathrm{H}_0 (r \neq 0) \text{ ssi } |t| \geq t_{1-\alpha/2}$$



Remarque :

• Souvent les logiciels fournissent la p-value (probabilité critique)

• La loi de Student n'est valable que dans le voisinage (r = 0), on ne peut donc pas l'utiliser pour les autres tests de conformité (H0 : r = a) où (a ≠ 0) et pour le calcul des intervalles de confiance.

Test de significativité – Un exemple

Numero	Modele	Cylindree	Puissance
1	Daihatsu Cuore	846	32
2	Suzuki Swift 1.0 GLS	993	39
3	Fiat Panda Mambo L	899	29
4	VW Polo 1.4 60	1390	44
5	Opel Corsa 1.2i Eco	1195	33
6	Subaru Vivio 4WD	658	32
7	Toyota Corolla	1331	55
8	Opel Astra 1.6i 16V	1597	74
9	Peugeot 306 XS 108	1761	74
10	Renault Safrane 2.2. V	2165	101
11	Seat Ibiza 2.0 GTI	1983	85
12	VW Golt 2.0 GTI	1984	85
13	Citroen ZX Volcane	1998	89
14	Fiat Tempra 1.6 Liberty	1580	65
15	Fort Escort 1.4i PT	1390	54
16	Honda Civic Joker 1.4	1396	66
17	Volvo 850 2.5	2435	106
18	Ford Fiesta 1.2 Zetec	1242	55
19	Hyundai Sonata 3000	2972	107
20	Lancia K 3.0 LS	2958	150
21	Mazda Hachtback V	2497	122
22	Mitsubishi Galant	1998	66
23	Opel Omega 2.5i V6	2496	125
24	Peugeot 806 2.0	1998	89
25	Nissan Primera 2.0	1997	92
26	Seat Alhambra 2.0	1984	85
27	Toyota Previa salon	2438	97
28	Volvo 960 Kombi aut	2473	125

r^	0.9475
n	28
ddl (n-2)	26

Test de significativité

t	15.1171
t-théorique (5%)	2.0555
p-value	2.14816E-14

$$t = \frac{\hat{r}}{\sqrt{\frac{1-\hat{r}^2}{n-2}}} = \frac{0.9475}{\sqrt{\frac{1-0.9475^2}{28-2}}} = 15.1171$$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0.975}(26) = 2.0555$$

Conclusion : On rejette l'hypothèse nulle (r = 0). Cette hypothèse n'est pas compatible avec les données au risque α = 5%

4. Intervalle de confiance

Intervalle de confiance

Problème : r^ est un estimateur qui dépend de l'échantillon, on dit qu'il est soumis aux fluctuations d'échantillonnage (avec d'autres données, on aura un résultat – légèrement – différent).

Solution : Déterminer un intervalle qui a une probabilité de (1 – α) de contenir la « vraie » valeur de r. (1 – α) est le **niveau de confiance**, l'intervalle ainsi définit est l'**intervalle de confiance**.

→ Pour y répondre, il faut connaître la loi de distribution de r^ (quelle que soit la vraie valeur de r), et disposer (ou pouvoir estimer) des paramètres de la loi.

→ La loi de Student ne convient plus, elle n'est valable que si « r = 0 »

$$\hat{z} = \frac{1}{2}\ln\frac{1+\hat{r}}{1-\hat{r}}$$

z^ suit une loi normale (quelle que soit la valeur de r)

On ne dispose (toujours) pas de la loi de r^, on passe par une autre transformation

→ La « transformation de Fisher ».

Avec :

$$E[\hat{z}] \approx \frac{1}{2}\ln\frac{1+r}{1-r}$$

$$V[\hat{z}] \approx \frac{1}{n-3}$$

Intervalle de confiance : Calcul pratique

Numero	Modele	Cylindree	Puissance
1	Daihatsu Cuore	846	32
2	Suzuki Swift 1.0 GLS	993	39
3	Fiat Panda Mambo L	899	29
4	VW Polo 1.4 60	1390	44
5	Opel Corsa 1.2i Eco	1195	33
6	Subaru Vivio 4WD	658	32
7	Toyota Corolla	1331	55
8	Opel Astra 1.6i 16V	1597	74
9	Peugeot 306 XS 108	1761	74
10	Renault Safrane 2.2. V	2165	101
11	Seat Ibiza 2.0 GTI	1983	85
12	VW Golt 2.0 GTI	1984	85
13	Citroen ZX Volcane	1998	89
14	Fiat Tempra 1.6 Liberty	1580	65
15	Fort Escort 1.4i PT	1390	54
16	Honda Civic Joker 1.4	1396	66
17	Volvo 850 2.5	2435	106
18	Ford Fiesta 1.2 Zetec	1242	55
19	Hyundai Sonata 3000	2972	107
20	Lancia K 3.0 LS	2958	150
21	Mazda Hachtback V	2497	122
22	Mitsubishi Galant	1998	66
23	Opel Omega 2.5i V6	2496	125
24	Peugeot 806 2.0	1998	89
25	Nissan Primera 2.0	1997	92
26	Seat Alhambra 2.0	1984	85
27	Toyota Previa salon	2438	97
28	Volvo 960 Kombi aut	2473	125

r^	0.9475
n	28

Calcul de z

z	1.8072
Variance(z)	0.0400
Ecart type(z)	0.2000

Quantile 0.975 - Loi normale

u(0.975)	1.9600

Intervalle de conf. pour z^

bb(z)	1.4152
bh(z)	2.1992

Intervalle de conf. pour r^

bb(r)	0.8886
bh(r)	0.9757

Etapes :

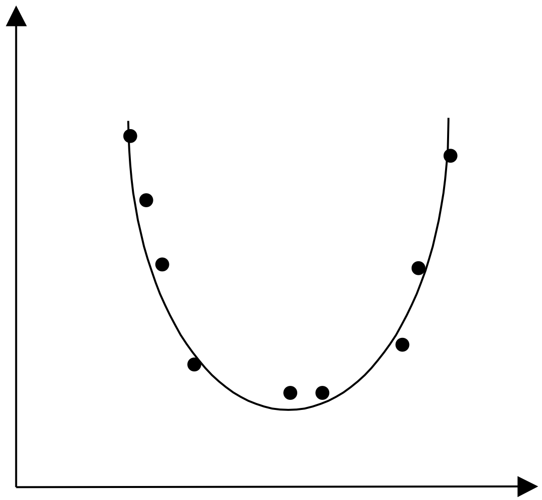
1. Calculer r^
2. Passer à la transformation z^
3. Calculer les bornes de l'intervalle de confiance de z^ au niveau de confiance (1-α)
4. Ramener ces bornes définies pour z^ à des bornes pour r^ (en utilisant la fonction inverse de la transformation de Fisher)

Conclusion : il y a 95% de chances que l'intervalle (0.8886 ; 0.9757) couvre la « vraie » valeur de r.

5. Cas pathologiques

Liaison non linéaire – Transformation de variables

Liaison non linéaire, et non monotone



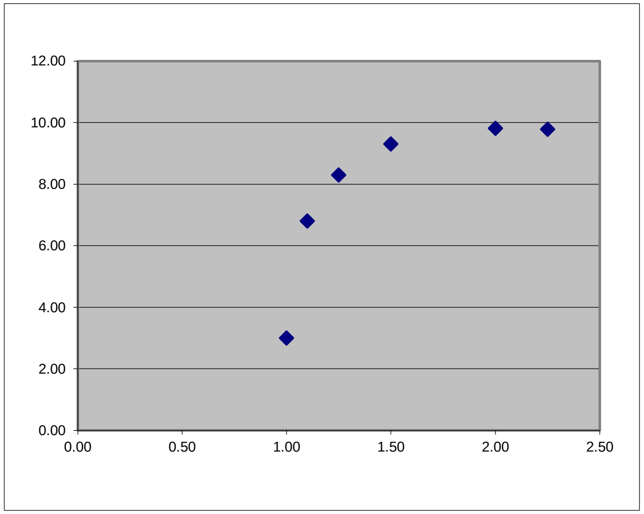
Linéarisation par transformation de variables (ex. $Z = X^2$)

Y	X	X²
9.31	-3	9
4.14	-2	4
1.04	-1	1
0.45	0	0
1.47	1	1
4.82	2	4
9.42	3	9
Corrélation (Y,X)	0.04369908	
Corrélation (Y,X²)	0.99772156	

Problème : comment deviner la bonne transformation ? Elle n'est pas toujours aussi évidente…

Liaison non linéaire – Passage par les rangs

Liaison non linéaire, mais monotone



Transformation des données en rangs

X	Y	RX	RY
1.00	3.00	1	1
1.10	6.80	2	2
1.25	8.30	3	3
1.50	9.30	4	4
2.00	9.81	5	6
2.25	9.78	6	5
Corrélation (XY)	0.77588403		0.94285714

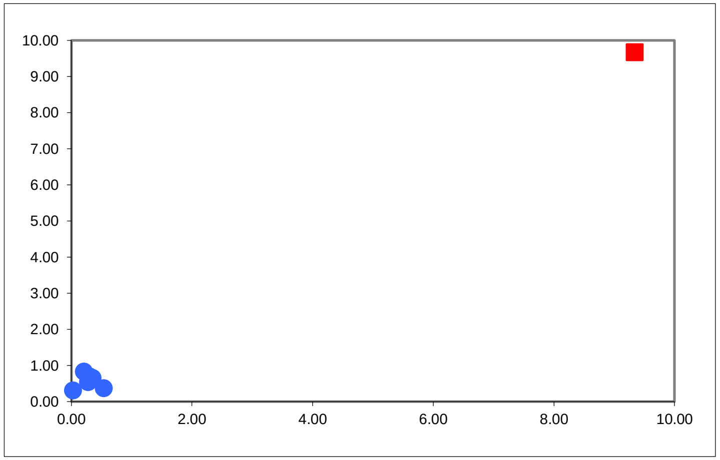
→Coefficient de Pearson calculé sur les rangs = « coefficient de Spearman »
→ Toute la partie inférentielle (test d'hypothèses, intervalle de confiance) reste valable
→ Mais pas bon pour les liaisons non monotones

age	rang moyen	rang aléatoire
15	4	**5**
18	7	7
12	1	1
13	2	2
15	4	**3**
16	6	6
15	4	**4**

En cas d'ex-aequo :

- rangs aléatoires (simple)
- rangs moyens (nécessite plus de calculs, mais plus précis)

Problème des points atypiques



Le coefficient de corrélation de Pearson est très sensible aux points atypiques

	X	Y
1	0.30	0.70
2	0.35	0.65
3	0.54	0.37
4	0.28	0.54
5	0.21	0.83
6	0.03	0.31
7	9.34	9.67

r (6 points)	**0.0185**
r (7 points)	**0.9976**

Passage aux rangs

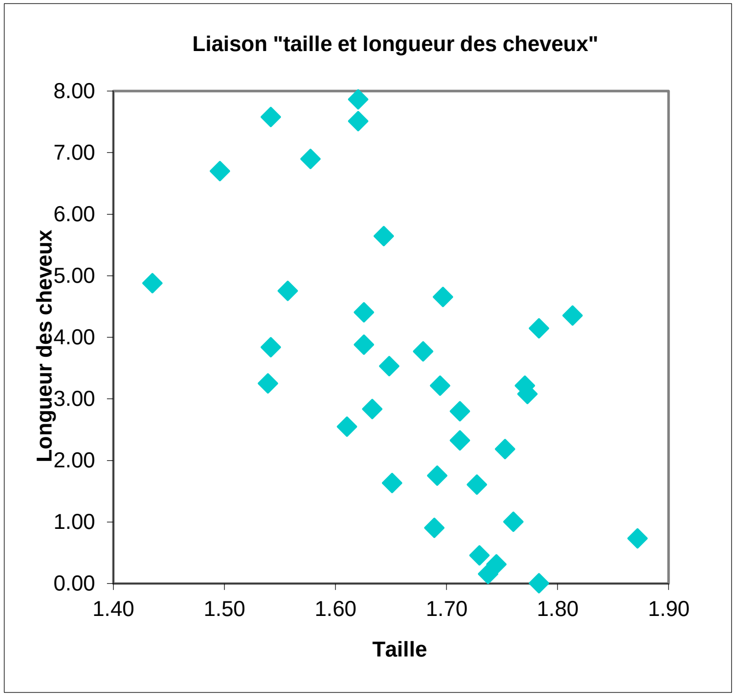
	X	Y	RX	RY
1	0.30	0.70	4	5
2	0.35	0.65	5	4
3	0.54	0.37	6	2
4	0.28	0.54	3	3
5	0.21	0.83	2	6
6	0.03	0.31	1	1
7	9.34	9.67	7	7

Coef. Rang:	0.39285714

Le coefficient calculé sur les rangs (coefficient de Spearman) est moins sensible aux points aberrants – Parce qu'il « lisse » les valeurs.

6. Corrélation partielle

Corrélations suspectes



Qui peut croire qu'il y a un lien entre la taille des personnes (X) et la longueur des cheveux (Y) ?

Il y a sûrement une tierce variable (Z) qui pèse simultanément sur X et Y.

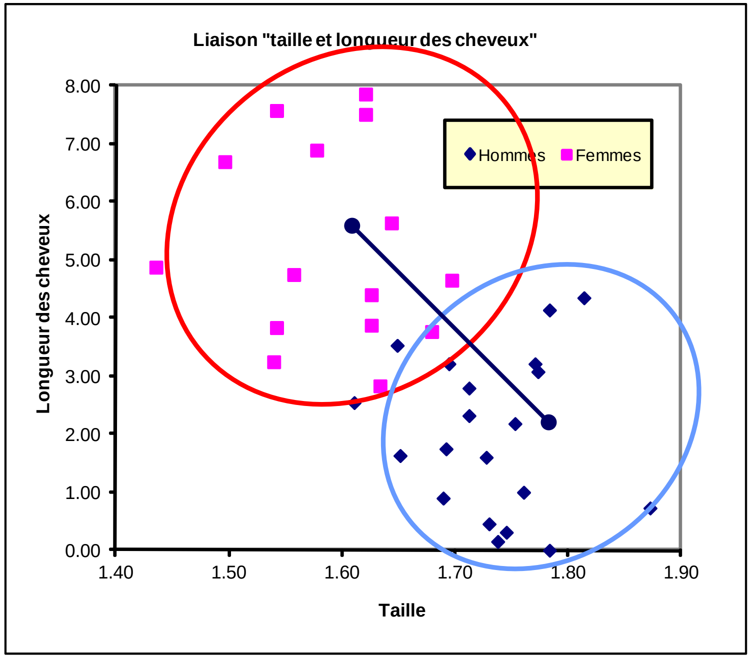
Et, de fait, la relation entre Y et X est essentiellement déterminée par Z.

Cas particulier : Z est binaire

		Cheveux (cm)	Taille (m)
Hommes	1	1.64	1.65
	2	0.32	1.74
	3	1.00	1.76
	4	2.80	1.71
	5	4.35	1.81
	6	2.33	1.71
	7	0.01	1.78
	8	1.75	1.69
	9	3.22	1.77
	10	3.53	1.65
	11	2.55	1.61
	12	3.08	1.77
	13	0.46	1.73
	14	3.22	1.69
	15	2.19	1.75
	16	0.73	1.87
	17	0.16	1.74
	18	0.90	1.69
	19	4.14	1.78
	20	1.61	1.73
Femmes	1	4.66	1.70
	2	3.25	1.54
	3	3.88	1.63
	4	2.84	1.63
	5	4.88	1.44
	6	3.77	1.68
	7	5.64	1.64
	8	4.41	1.63
	9	3.84	1.54
	10	7.58	1.54
	11	7.51	1.62
	12	6.90	1.58
	13	4.76	1.56
	14	6.70	1.50
	15	7.86	1.62

r (hommes)	-0.074
r (femmes)	-0.141

r (global)	-0.602



La corrélation est essentiellement définie par le décalage entre les nuages de points.

Les corrélations intra-nuages sont nulles.

Corrélation partielle (Z quantitative également)

On retranche l'effet de z sur x **et** sur y

Corrélation brute (y, x)

Coefficient de corrélation partielle (corrélation entre X et Y, en contrôlant l'effet de Z)

$$\boxed{r_{xy.z} = \frac{r_{xy} - r_{xz} \times r_{yz}}{\sqrt{(1-r_{xz}^2)}\sqrt{(1-r_{yz}^2)}}}$$

Normalisation pour que $-1 \leq r_{xy.z} \leq +1$

Estimation : il faut utiliser les estimations des corrélations brutes

$$\boxed{\hat{r}_{xy.z} = \frac{\hat{r}_{xy} - \hat{r}_{xz} \times \hat{r}_{yz}}{\sqrt{(1-\hat{r}_{xz}^2)}\sqrt{(1-\hat{r}_{yz}^2)}}}$$

Corrélation partielle d'ordre p (p >1) : formule de récurrence

$$\boxed{r_{xy.zw} = \frac{r_{xy.z} - r_{xw.z} \times r_{yw.z}}{\sqrt{(1-r_{xw.z}^2)}\sqrt{(1-r_{yw.z}^2)}}}$$

p = 2 ici, on utilise un calcul de proche en proche

Corrélation partielle – Inférence statistique

(Noter les degrés de liberté)

$$\hat{r}_{xy.z} = \frac{0.8878 - 0.9475 \times 0.8919}{\sqrt{(1-0.9475^2)}\sqrt{(1-0.8919^2)}} = 0.2955$$

		X	Y	Z
Numero	Modele	Puissance	Conso	Cylindree
1	Daihatsu Cuore	32	5.7	846
2	Suzuki Sw ift 1.0 GLS	39	5.8	993
3	Fiat Panda Mambo L	29	6.1	899
4	VW Polo 1.4 60	44	6.5	1390
5	Opel Corsa 1.2i Eco	33	6.8	1195
6	Subaru Vivio 4WD	32	6.8	658
7	Toyota Corolla	55	7.1	1331
8	Opel Astra 1.6i 16V	74	7.4	1597
9	Peugeot 306 XS 108	74	9.0	1761
10	Renault Safrane 2.2. V	101	11.7	2165
11	Seat Ibiza 2.0 GTI	85	9.5	1983
12	VW Golt 2.0 GTI	85	9.5	1984
13	Citroen ZX Volcane	89	8.8	1998
14	Fiat Tempra 1.6 Liberty	65	9.3	1580
15	Fort Escort 1.4i PT	54	8.6	1390
16	Honda Civic Joker 1.4	66	7.7	1396
17	Volvo 850 2.5	106	10.8	2435
18	Ford Fiesta 1.2 Zetec	55	6.6	1242
19	Hyundai Sonata 3000	107	11.7	2972
20	Lancia K 3.0 LS	150	11.9	2958
21	Mazda Hachtback V	122	10.8	2497
22	Mitsubishi Galant	66	7.6	1998
23	Opel Omega 2.5i V6	125	11.3	2496
24	Peugeot 806 2.0	89	10.8	1998
25	Nissan Primera 2.0	92	9.2	1997
26	Seat Alhambra 2.0	85	11.6	1984
27	Toyota Previa salon	97	12.8	2438
28	Volvo 960 Kombi aut	125	12.7	2473

n	28

Corrélations brutes		
Puissance	Conso	0.88781
Puissance	Cylindrée	0.94755
Conso	Cylindrée	0.89187

Corrélation partielle	
r_xy.z	0.29553

Test de significativité	
t	1.54673
t(0.975 ; 25)	2.38461
p-value	0.13450

Intervalle de confiance à 95%	
f	0.30461
e.t.	0.20412
u(0.975)	1.95996
bb(f)	-0.09546
bh(f)	0.70469
bb (r)	-0.09517
bh (r)	0.60734

Test de significativité

$$t = \frac{\hat{r}_{xy.z}}{\sqrt{\frac{1-\hat{r}^2_{xy.z}}{n-p-2}}} \equiv \Im(n-p-2)$$

Intervalle de Confiance (via la transformation de Fisher)

$$\hat{z} = \frac{1}{2}\ln\frac{1+\hat{r}}{1-\hat{r}}$$

qui suit une loi normale de paramètres…

$$E[\hat{z}] \approx \frac{1}{2}\ln\frac{1+r}{1-r}$$

$$V[\hat{z}] \approx \frac{1}{n-p-3}$$

Bibliographique

- R. Bourbonnais, « Économétrie », Dunod, 1998.

- Y.Dodge, V.Rousson, « Analyse de régression appliquée », Dunod, 2004.

- M. Tenenhaus, « Statistique : Méthodes pour décrire, expliquer et prévoir », Dunod, 2007.