Corrélation et Régression linéaire

Dr Marc CUGGIA
PCEM 1 – 2007/2008

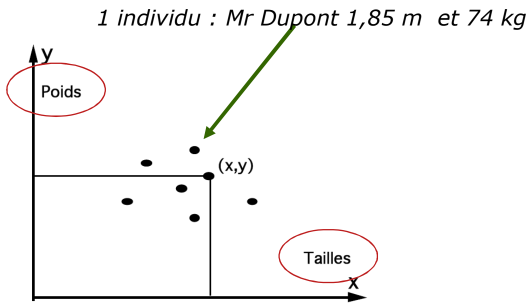
http://www.ea3888.univ-rennes1.fr

Positionnement

- variations respectives de plusieurs grandeurs dans une même population
  - ex: relation entre poids et taille
- courbe associée à la *fonction* y=f(x)
- si la loi est définie, la connaissance de x suffit à déterminer y
  - relation fonctionnelle (sciences exactes)
- si fluctuations statistiques
  - à une valeur d'une des variables correspond une distribution des valeurs de l'autre variable

Représentation graphique

- Représenter les couples de valeurs (x,y)
  - Obtention d'un nuage de point



Correlation et régression

- La régression permet d'étudier l'association entre deux variables quantitatives, en étudiant les variations de l'une en fonction des valeurs de l'autre.

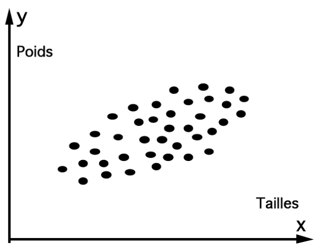
- Le coefficient de corrélation est une mesure d'association entre deux variables quantitatives faisant jouer des rôles symétriques aux valeurs.
  - Les deux variables peuvent être placées indifferement en abscisse ou en ordonnées)
  - On cherche à savoir simplement s'il existe une liaison entre ces deux variables et à quantitfier l'intensité de la liaison

Correlation

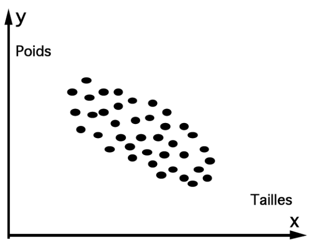
X et Y sont des variables quantitatives

- **Dire que X et Y sont corrélées, c'est affirmer qu'il existe une liaison entre ces deux variables.**
- **Plus X varie dans un sens, plus Y varie.**
- **Si Y varie dans le même sens, ➔ la corrélation est positive**
- **Si Y varie dans le sens opposé ➔ la corrélation est négative**
- **Si X et Y varie indépendamment de l'un de l'autre ➔ les variables ne sont pas corrélées.**

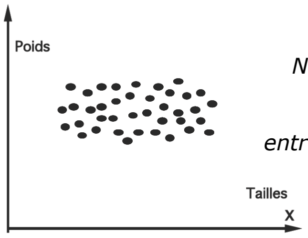
Représentation graphique



Corrélation positive (directe)



Corrélation négative (inverse)



Absence de corélation

Nuage de points diffus
⇔
Absence de liaison
entre les caractères étudiés

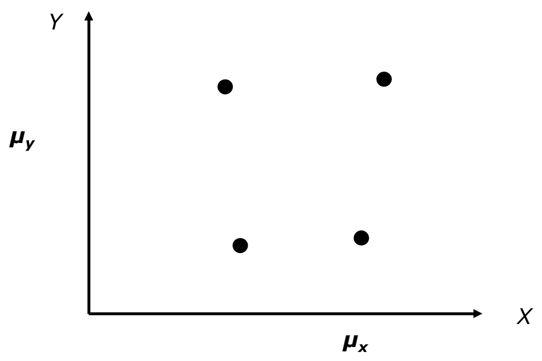
rennes1.fr

Notion de covariance

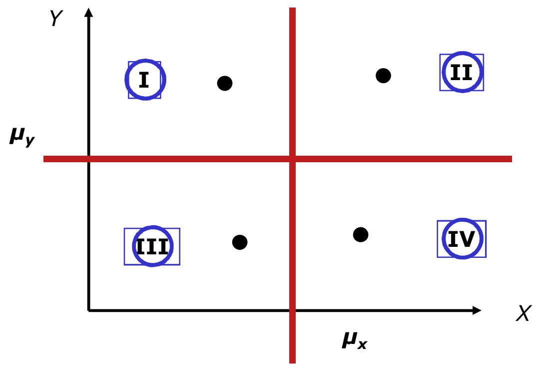
- Indicateur qui mesure la liaison entre deux variables X et Y
- C'est la moyenne des produits des écarts entre X et Y divisé par leurs moyennes respectives μ_x et μ_y

http://www.ea3888.univ-rennes1.fr

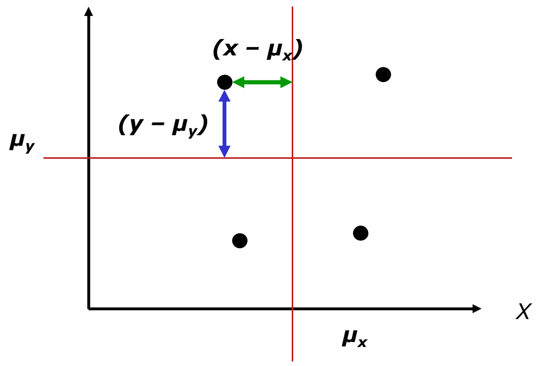
4 couples de valeurs x et y



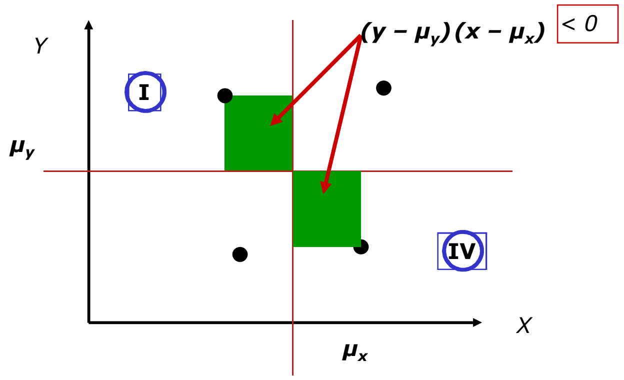
4 couples de valeurs x et y



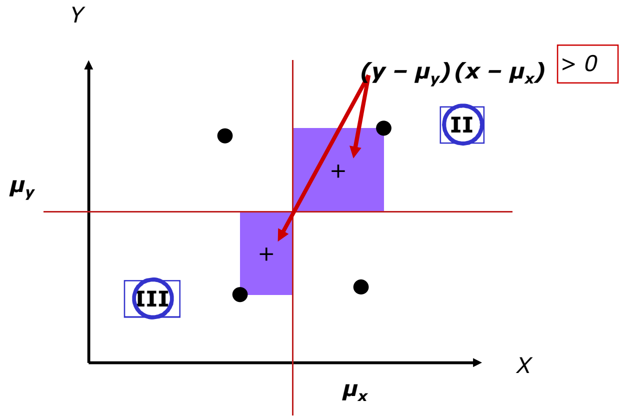
Les distances de chaque point à l'axe des moyennes μ_y et μ_x représentent les écarts aux moyennes



Les aires des rectangles représentante les produits des écarts de chaque couple



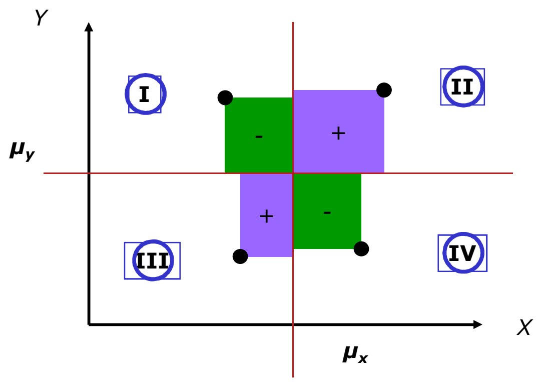
Les rectangles verts dans les zones I et IV représentent les produits négatifs



Les rectangles Mauves dans les zones II et III représentent les produits Positifs

http://www.ea3888.univ-rennes1.fr

La covariance peut être illustrée par la somme des aires des rectangles



http://www.ea3888.univ-rennes1.fr

- **On définit la covariance :**

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{N}$$

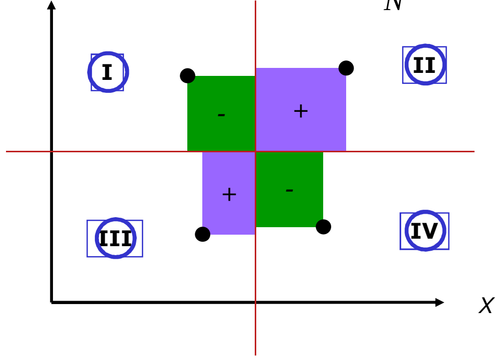
- **3 cas de figure**
  - **1 er cas : pas de liaison en X et Y**
  - **Les points sont répartis uniformement dans les quandrants**
  - **Les aires se compensent :**

$$\sum (x_i - \mu_X)(y_i - \mu_Y) = 0$$

http://www.ea3888.univ-rennes1.fr

- 1 er cas : pas de liaison en X et Y
- Les points sont répartis uniformément dans les quadrants
- Les aires se compensent : **$(y - \mu_y)(x - \mu_x)$** = 0

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{N} = 0$$

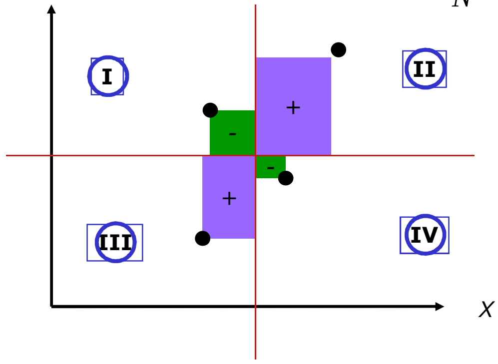


http://www.ea3888.univ-rennes1.fr

Ici, la somme des aires est positive, la covariance est positive.
➔ Il semble exister une liaison positive entre X et Y,
➔ plus X est élevé, plus Y est élevé

$(y - \mu_y)(x - \mu_x) > 0$

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{N} > 0$$

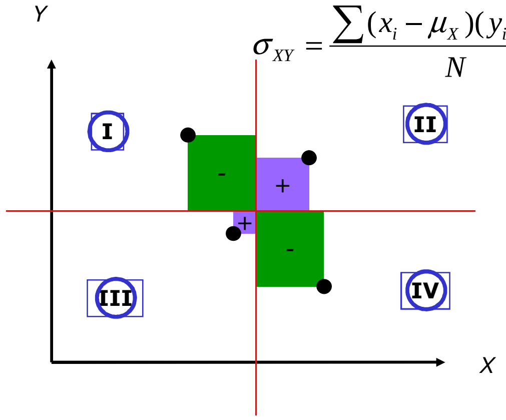


http://www.ea3888.univ-rennes1.fr

Ici, la somme des aires est négative, la covariance est négative.
➔ Il semble exister une liaison négative entre X et Y,
➔ plus X est élevé, plus Y est bas

$$(y - \mu_y)(x - \mu_x) < 0$$

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{N} < 0$$



http://www.ea3888.univ-rennes1.fr

Coefficient de correlation

- **La covariance est le produit de deux termes exprimés en unités qui peuvent être différentes**

$$\sigma_{XY} = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{N}$$

- **Elle ne se prête donc pas à l'analyse statistique**
- **Pour calculer ρ le coefficient de corrélation, on divise σ_{xy} par le produit des écarts types de chaque distribution. On obtient un coefficient sans unité.**

Soit :

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

http://www.ea3888.univ-rennes1.fr

- La covariance entre X et Y pour une population de N sujets dont les valeurs pour les variables X et Y sont (xi,yi) vaut :

$$\sigma_{XY} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$$

- Comme on a

$$\sigma_X{}^2 = \frac{\sum (x_i - \mu_x)^2}{N}$$

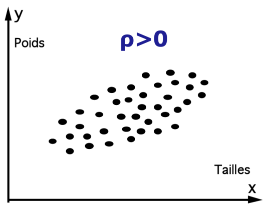
$$\sigma_Y{}^2 = \frac{\sum (y_i - \mu_y)^2}{N}$$

- Après simplification on a :
-

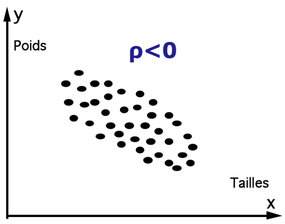
$$\rho = \frac{\sum (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum (x_i - \mu_X)^2 \sum (y_i - \mu_Y)^2}}$$

http://www.ea3888.univ-rennes1.fr

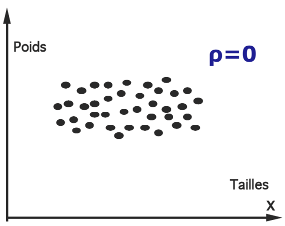
Interprétation de ρ



Corrélation positive (directe)



Corrélation négative (inverse)



Absence de corélation

http://www.ea3888.univ-rennes1.fr

Pour bien comprendre la variation de r en fonction du nuage de point, allez voir la ressource suivante sur internet

http://noppa5.pc.helsinki.fi/koe/flash/corr/ch16.html

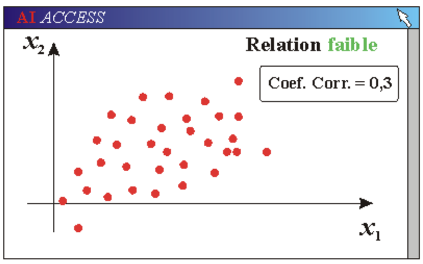
Propriété de ρ

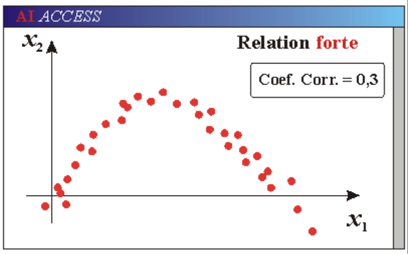
- Le coefficient de corrélation fait jouer un rôle symétrique à X et Y
  - Il ne change pas si on permute X et Y
- ρ a le même signe que β, la pente de la droite de régression de Y en fonction de X
- ρ reste inchangé si on change d'unité ou d'origine pour les X et Y

Propriété de ρ

- ρ est toujours compris entre -1 et 1 et ces bornes ne peuvent être atteintes que si Y = β X + α
  - ρ permet de mesurer la FORCE DE L'ASSOCIATION entre X et Y. Plus ρ est proche de +1 ou de -1, plus l'association est forte
  - Ce n'est pas le cas β qui n'indique rien en elle-même sur la force de l'association, puisqu'elle dépend totalement du choix des unités de ces variables

- Si X et Y sont indépendantes alors ρ=0
  - L'inverse n'est pas vrai :
  - Si ρ~0, les variables peuvent soient être indépendantes mais aussi être liées (mais non linéairement)
  - On peut seulement affirmer que les variables X et Y ne sont pas liées linéairement





~~Pour bien comprendre le problème de la condition~~ de linéarité, vous pouvez aller voir sur internet à l'adresse :

http://noppa5.pc.helsinki.fi/koe/flash/corr/corrx2-en.html

si la condition ne linéarité n'est pas remplie, le coef de corrélation n'est pas un bon moyen d'étudier la ~~force de l'association entre deux variables~~

Estimation ρ ➔ r

- Le coefficient de corrélation d'un échantillon est noté r.
- On l'obtient en remplaçant la covariance et les variances par leurs estimations.
- Estimation de la covariance à partir d'un échantillon de n individus

$$\mathrm{cov}_{XY} = \frac{\sum (x_i - m_X)(y_i - m_Y)}{n-1}$$

- Rappelons l'estimation des variances pour un échantillon de n individus

$$s_x^2 = \frac{\sum (x_i - m_x)^2}{n-1}$$

$$s_y^2 = \frac{\sum (y_i - m_y)^2}{n-1}$$

http://www.ea3888.univ-rennes1.fr

- **Après simplification par (n-1) on obtient pour un échantillon de n sujet où les couples de valeurs de X et Y observés sont (x_i,y_i)**

$$r = \frac{\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^{n}(x_i - m_x)^2 \sum_{i=1}^{n} (y_i - m_y)^2}}$$

Autres formules

$$r = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}(\sum x_i)^2\right]\left[\sum y_i^2 - \frac{1}{n}(\sum y_i)^2\right]}}$$

$$r = \frac{\sum x_i y_i - n.m_x.m_y}{(n-1)\sqrt{s_x^2 s_y^2}}$$

Exemple

http://www.ea3888.univ-rennes1.fr

- Pour 63 nouveau nés, on a relevé le poids maternel avant grossesse (X), et paternel (Y) en kilogramme.

- On cherche a trouver voir s'il existe une corrélation entre les poids maternels et paternels

- On donne :
  - Σx_i=3 644 Σy_i=4 729
  - Σx_i^2=217 502 Σy_i^2=363 527
  - $\Sigma x_i y_i$= 275 480

- Calculer r

http://www.ea3888.univ-rennes1.fr

- On donne :
  - Σx_i=3 644 Σy_i=4 729
  - Σx_i^2=217 502 Σy_i^2=363 527
  - $\Sigma x_i y_i$= 275 480

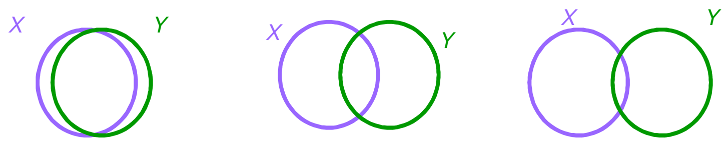
$$r = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}(\sum x_i)^2\right]\left[\sum y_i^2 - \frac{1}{n}(\sum y_i)^2\right]}}$$

$$r = \frac{275480 - \frac{1}{63}4729 \times 3644}{\sqrt{(217502 - \frac{1}{63}3644^2)(363527 - \frac{1}{63}4729^2)}} = \frac{1948,63}{7585,49} = \boxed{0,26}$$

- Interpretation
- r = 0,20 à 0,40 : faible ou quasi absence de correlation
- r = 0,40 à 0,60 : moyenne correlation
- r = 0,60 à 0,80 : bonne corrélation;
  r = 0,80 : corrélation élevée.

- Ici r=0,26 ➔ faible ou quasi absence de corrélation
- Pas de raison, a priori d'avoir des poids de père et de mère corrélées.

- Pour comprendre une valeur de r lorsque il n'est ni proche de 1 ni de 0 ?
- r^2 matérialise la force de la corrélation
- Représente le % de variance que perd une des deux variables quand l'autre est fixée.
- Diagramme de Venn :
  - Cercle = variables
  - Pourcentage de surface commune = r^2
  - Plus le recouvrement est important, plus les variables sont liées :



r=0,9 r=0,6 r=0,3

Test du r

- Rappel : r concerne les variables d'un échantillon
- Le calcul de r peut être sujet à fluctuation.
- Tester r, c'est tenter d'affirmer ou pas que sa valeur est statistiquement significative et ce avec un risque maîtrisé ($p<0,05$)
- Même mécanisme que pour les autres test : hypothèses sur la population

- Ho = Hypothèse nulle : $\rho=0$
- H1 = Hypothèse alternative :
  - $\rho\neq0$ (test bilateral)
  - $\rho<0$ ou $\rho>0$ (test unilateral)

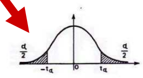
- Test du coefficient de correlation
- 2 types de constructions de tables dans les ouvrages
- ➔ changement dans les hypothèses Ho et H1
- Au concours : 1 seule type de table

VALEURS des CENTILES
pour la
DISTRIBUTION t de STUDENT
en fonction du nombre ν de degrés de liberté
(aire en grisé = p)

ν	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129

Table II

TABLE DU t DE STUDENT



$\nu \backslash \alpha$	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610

- Le test consiste à calculer $$t_o = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$
- Et à le comparer à la valeur seuil lue dans la table du test t de student à (n-2) ddl.
- La règle de décision du test est :

- Test bilatéral :Ho : ρ=0 et H1 : ρ≠0
- on rejette Ho si $$|t_o| \geq t_{n-2;\alpha}$$
- Test unilatéral Ho : ρ=0
  - H1: ρ>0 on rejette Ho si $t_o \geq t_{n-2;2\alpha}$
  - H1: ρ<0 on rejette Ho si $$t_o \geq t_{n-2;-2\alpha} (= -t_{n-2;2\alpha})$$

- Conditions d'applications :
- La régression entre X et Y est linéaire
- Une des deux distributions conditionnelles normales et de variance constante.
  - Les distributions de Y liées à chaque valeur de X doivent être normales et de variance constante.
  - Symétriquement, Les distributions de X liées à chaque valeur de Y doivent être normales et de variance constante.

- Les observation pour chaque variable doivent être indépendantes les unes des autres.
  - Ex : comparaison des données Y en fonction du temps X
  - Les données de la veille ne sont pas indépendantes des données du lendemain.
  - Il ya auto-correlation ➔ nécessite d'autres techniques d'analyse.

Exemple

- Pour 63 nouveau nés, on a relevé le poids maternel avant grossesse (X), et paternel (Y) en kilogramme.
- On cherche a trouver voir s'il existe une corrélation entre les poids maternels et paternels
- r=0,26 pour un échantillon de 63 sujets

$$t_o = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad t_o = \frac{0{,}26\sqrt{61}}{\sqrt{1-0{,}26^2}} = 2{,}10$$

- Rappel en hypothèse bilatérale, on rejette Ho si $|t_o| \geq t_{n-2;\alpha}$
- or to=2,1 et
- $t_{61;5\%}$= ? ➔ on prend le ddl immédiatement au dessous
- $t_{40;5\%}$=2,021
- ➔ On rejette Ho
- Le coefficient de corrélation entre le poids de la mère et le poids paternel est significativement différent de 0.
- Les conditions d'application sont que
  - la régression du poids maternel sur le poids de la mère est linéaire.
  - La distribution conditionnelle de l'une des 2 variables et normale et de variance constante

ν \ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,949
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Spearman

http://www.ea3888.univ-rennes1.fr

Test du coefficient de correlation des rangs de spearman

- **S'il existe un doute sur la normalité des distributions de X et Y, ou sur la linéarité de la relation entre X et Y, on ne peut pas utiliser le coefficient de corrélation de Pearson**
- **On utilise alors un test non paramétrique :**
- **le coefficient de corrélation des rangs de Spearman**
  - **Il étudie l'existence d'une liaison entre 2 variables quantitatives.**

http://www.ea3888.univ-rennes1.fr

- **On ne s'intéresse plus aux valeurs mais à leur rang.**
  - **On appelle rang le numéro d'ordre d'une valeur après classement de la variable par ordre croissant.**
    - Sur la série 1,4,5,8 la valeur 5 a pour rang 3, et la valeur 8 a pour rang 4. En cas d'ex aequo on attribut le rang moyen à chacun d'eux
  - **On définit x'_i et y'_i les rangs des valeurs observées**
- On définit r_s le coefficient de corrélation des rangs de Spearman

$$r_s = 1 - \frac{6\sum(x'_i - y'_i)^2}{n(n^2-1)}$$

- On définit l'écart type du coefficient de Spearman :

$$s_r = \sqrt{\frac{1-r_s^2}{n-2}}$$

- **Le test de Spearman consiste à calculer to**

$$t_o = \frac{|r_s|}{s_r}$$

- **Et à la comparer à une valeur théorique :**
- **Test bilatéral :Ho : ρ=0 et H1 : ρ=/=0**
- **on rejette Ho si** $|t_o| \geq t_{n-2;\alpha}$
- **Test unilatéral Ho : ρ=0**
  - **H1: ρ>0 on rejette Ho si** $t_o \geq t_{n-2;2\alpha}$
  - **H1: ρ<0 on rejette Ho si** $t_o \geq t_{n-2;-2\alpha} (= -t_{n-2;2\alpha})$

Exemple

- On désire vérifier la corrélation entre la taille (en cm) et le poids (en kg) des enfants de 2 ans sur un échantillon de 15 individus.

Taille (x)	82,9	83,4	82,4	82,1	84,8	86,7	84,0	89,0	85,0	85,4	87,7	87,7	86,4	86,4	86,9
Poids (y)	8,7	9,2	9,5	10,1	10,4	10,5	10,8	11,0	11,5	11,6	12,4	13,6	13,8	13,9	14,6

- Il existe un doute sur la linéarité de la relation entre x et y.
- On préconise le calcul du coefficient de Spearman
- Les observations pour chaque variables sont indépendantes les unes des autres

- Conditions d'application vérifiées :
- Le nombre de couples de valeurs >10
- Indépendances
- Pas d'exigence sur la normalité ni sur la linéarité

- On pose
- Ho : il n'existe aucune corrélation entre la taille et le poids
- H1 : il existe une relation entre taille et poids

Taille (x)	82,9	83,4	82,4	82,1	84,8	86,7	84,0	89,0	85,0	85,4	87,7	87,7	86,4	86,4	86,9

valeurs

Ordonner les valeurs de façon croissante

82,1	82,4	82,9	83,4	84	84,8	85	85,4	86,4	86,4	86,7	86,9	87,7	87,7	89

http://www.ea3888.univ-rennes1.fr

Taille (x)	82,9	83,4	82,4	82,1	84,8	86,7	84,0	89,0	85,0	85,4	87,7	87,7	86,4	86,4	86,9

valeurs

Ordonner les valeurs de façon croissante

82,1	82,4	82,9	83,4	84	84,8	85	85,4	86,4	86,4	86,7	86,9	87,7	87,7	89
1	2	3	4	5	6	7	8							

Affecter les rangs

http://www.ea3888.univ-rennes1.fr

Taille (x)	82,9	83,4	82,4	82,1	84,8	86,7	84,0	89,0	85,0	85,4	87,7	87,7	86,4	86,4	86,9

valeurs

Ordonner les valeurs de façon croissante

82,1	82,4	82,9	83,4	84	84,8	85	85,4	86,4	86,4	86,7	86,9	87,7	87,7	89
1	2	3	4	5	6	7	8	9,5	9,5	11	12	13,5	13,5	15

(8+11)/2=9,5 *(12+15)/2=13,5*

Affecter les rangs

http://www.ea3888.univ-rennes1.fr

valeurs

rangs

Taille	x	82,9	83,4	82,4	82,1	84,8	86,7	84,0	89,0	85,0	85,4	87,7	87,7	86,4	86,4	86,9
	x'	3	4	2	1	6	11	5	15	7	8	13,5	13,5	9,5	9,5	12
Poids	y	8,7	9,2	9,5	10,1	10,4	10,5	10,8	11,0	11,5	11,6	12,4	13,6	13,8	13,9	14,6
	y'	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

valeurs

rangs

http://www.ea3888.univ-rennes1.fr

x'	3	4	2	1	6	11	5	15	7	8	13,5	13,5	9,5	9,5	12
y'	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Calcul des carrés des différences des rangs (rang(x)-rang(y))²															
Rang (y)	4	4	1	9	1	25	4	49	4	4	6,25	2,25	12,25	20,25	9

$$\sum (x'-y')^2 = 155$$

$$r_s = 1 - \frac{6\sum (x'_i - y'_i)^2}{n(n^2-1)} \qquad r_s = 1 - \frac{(6 \times 155)}{15 \times (15^2 - 1)} = 0,72$$

$$s_r = \sqrt{\frac{1-r_s^2}{n-2}} \qquad s_r = \sqrt{\frac{(1-0,72^2)}{(15-2)}} = 0,19$$

http://www.ea3888.univ-rennes1.fr

$$t_o = \frac{0,72}{0,19} = 3,79$$

$$ddl = 15 - 2 = 13$$

- Or $t_{13;5\%}$=2,160
- to>$t_{13;5\%}$
- On rejette Ho.
- La valeur t est encore supérieure à $t_{13;1\%}$
- On conclut donc qu'il existe une liaison positive significative entre la taille et le poids des enfants de 2 ans (p<0,01)

ν \ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,949
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

ν \ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,949
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Régression

http://www.ea3888.univ-rennes1.fr