

0. Supports à lire avant de commencer les exercices

Site du cours : http://eric.univ-lyon2.fr/~ricco/cours/cours_econometrie.html

A lire en particulier pour cette séance :

- Analyse de corrélation (slides) : http://eric.univ-lyon2.fr/~ricco/cours/cours/diapo_analyse_de_correlation.pdf
- Analyse de corrélation (fascicule de cours), chapitres 1 et 2, section 4.4, et chapitre 5 : http://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse_de_Correlation.pdf

1. Quelques démonstrations (*correction non fournie*)

Démontrez les propriétés suivantes de la covariance :

1. $COV(X, Y+Z) = COV(X, Y) + COV(X, Z)$
2. $COV(X, a) = 0$ (a est une constante)
3. $COV(X, Y) = 0$ si X et Y sont indépendants (n'oubliez pas que si X et Y sont indépendants, $E[XY] = E[X] * E[Y]$)

2. Tests et intervalle de confiance (1)

On mesure la corrélation entre le salaire (Y) et le nombre d'année d'études (X) sur un échantillon de $n = 40$ individus. Nous obtenons $r^{\wedge} = 0.46$.

1. La corrélation est-elle significative à 5% ? (c.-à-d. significativement différente de 0 $\rightarrow H_0 : r = 0$ vs. $H_1 : r \neq 0$)
2. Est-elle significativement supérieure à 0 à 5% ? ($H_0 : r = 0$ vs. $H_1 : r > 0$)
3. Calculez son intervalle de confiance à 95%.

3. Tests et intervalle de confiance (2)

Des arguments économiques suggèrent que le « taux d'inflation » (Y) est lié négativement au ratio « Masse monétaire/PNB » (X). Des observations sur 14 années consécutives ont permis de recueillir les données suivantes :

$$n = 14 \quad \sum x_i = 59.4 \quad \sum y_i = 2206$$

$$\sum x_i^2 = 280.2 \quad \sum y_i^2 = 630290 \quad \sum x_i y_i = 7350.6$$

1. Calculez les moyennes de X et Y
2. Calculez les variances empiriques de X et Y
3. Calculez le coefficient de corrélation empirique (r^{\wedge}) ?
4. Est-il significativement inférieur à **-0.5** (test à 5%) ? ($H_0 : r = -0.5$ vs. $H_1 : r < -0.5$)

4. Corrélation partielle

On étudie la nocivité des cigarettes via leur teneur en monoxyde de carbone (CO). Avec $n = 24$ observations, on a construit la matrice des corrélations

| CORR (n = 24) | goudron | nicotine | poids | CO |
|---------------|---------|----------|-------|-------|
| goudron | 1 | 0.960 | 0.284 | 0.966 |
| nicotine | 0.960 | 1 | 0.286 | 0.931 |
| poids | 0.284 | 0.286 | 1 | 0.310 |
| CO | 0.966 | 0.931 | 0.310 | 1 |

1. Quelle est la corrélation entre CO et Nicotine ?
2. La liaison entre CO et Nicotine est-elle significativement positive à 5% ? ($H_0 : r = 0$ vs. $H_1 : r > 0$)
3. Cette liaison est-elle significativement positive si on retranche d'influence du goudron ?
(Cette question peut être formulée autrement : à teneur en goudron égale des cigarettes, le CO est-il positivement lié à la nicotine ?)

5. Corrélation de Spearman

On dispose d'un échantillon de $n = 10$ individus d'un couple de variables (X, Y). Voici les valeurs observées :

| X | Y |
|----|-----|
| 1 | 1 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 9 |
| 6 | 9.2 |
| 7 | 9.4 |
| 8 | 9.6 |
| 9 | 9.9 |
| 10 | 10 |

1. Calculez le coefficient de corrélation entre X et Y.
2. Construisez un graphique nuage de points avec X en abscisse et Y en ordonnée. Que constatez-vous ?
3. Transformez les données en rangs.
4. Calculez le coefficient de corrélation de Spearman.

$$\rho = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

Où $D_i = R_i - S_i$; R_i = rang de X_i parmi les X ; S_i = rang de Y_i parmi les Y.

Que peut-on en conclure ?