ECONOMETRIE

MODULE 1

J. Paul Tsasa Vangu

Rédigée sous contrôle des –
Prof. J. Pierre Bosonga et Prof. Daniel Mukoko

UNIVERSITE PROTESTANTE AU CONGO
Centre Congolais-Allemand de Microfinance

Frankfurt School of Finance & Management Bankakademie | HfB

DAAD Deutscher Akademischer Austausch Dienst German Academic Exchange Service

" *Pourquoi s'obstiner à expliquer très péniblement et très incorrectement, comme l'a fait souvent* David Ricardo *et comme le fait chaque instant* John Stuart Mill *dans son Principe d'Economie Politique, en se servant de la langue usuelle, de choses qui, dans une langue mathématique, peuvent s'énoncer en bien moins de mots, d'une façon bien plus exacte et plus claire ?* "

– Léon Walras

PLAN SOMMAIRE

MODULE :

1. Théorie de corrélation
2. Hypothèses classiques des moindres carrés ordinaires : Implications
3. Hypothèses classiques des MCO : Relâchement, Correction et Dépassements
4. Modèle de régression linéaire multiple
5. Modèles de régression non linéaires
6. Modèle multiéquationnel
7. Modèles avec variables qualitatives : LOGIT – TOBIT – GOMBIT
8. Modèles dynamiques
9. Modèles à correction d'erreur et autorégressifs vectoriels
10. Introduction à l'économétrie bayésienne

AVANT-PROPOS

Ce recueil d'applications est rédigé en un moment où l'enseignement de l'économétrie est entaché de nombreuses erreurs techniques. Il suffit de lire la plupart de recueils d'applications, il s'attarde soit à bombarder les démonstrations mathématiques souillées c'est-à-dire sans respect de fondamentaux de l'inférence statistique, soit à confondre les tests paramétriques de tests non paramétriques, soit à proposer une recette purement informatique embellie par de raccourcis et de commandes de logiciels informatiques.

Partant de cette constatation, ce recueil se propose :

1. Dune part, d'aider l'étudiant de comprendre le cours, tel qu'enseigné dans la partie théorique ;
2. Et d'autre part, de recadrer les applications afin de retrouver l'orthodoxie, tel que voulu par le Maître.

En vue de ne pas condamner les apprenants à une illusion scientifique, nous avons pensé rédiger, sous le contrôle des Professeurs Bosonga et Mukoko (notamment, la partie concernant modélisation ARCH), un support qui soit non seulement très pratique, mais aussi techniquement pur ! Cette démarche garantie l'objectivité dans l'analyse, met en avant l'élégance mathématique et évite la « violence symbolique », source d'exclusion de ceux qui ne maîtrisent pas le langage mathématique complexe.

Les exercices recensés dans plus d'une vingtaine d'ouvrages d'économétrie et de statistique appliquée en économie [voir bibliographie] ont été soigneusement sélectionnés, en vue de faciliter la pratique et l'illustration de l'analyse théorique. Par ailleurs, en vue de renforcer notre vision de la *rigueur dans la simplicité*, nous avons conçu quelques exercices dans chaque chapitre.

Les lecteurs trouveront également des exercices se rapportant aux examens et interrogations de différentes universités, notamment l'Université Protestante au Congo [UPC], l'Université de Kinshasa [UNIKIN], l'Université William-Booth [UWB], l'Université Libre de Kinshasa [ULK], l'Université Cheik Anta Diop, l'Université de Princeton, l'Université Impériale de Kyoto, l'Université de Sorbonne, l'Université de Lyon, l'Université de Jérusalem et l'Université Laval.

In fine, nous espérons que l'autorité et la rigueur avec lesquelles nous présentons ces exercices, aideront les étudiants à maîtriser la plupart de questions fondamentales que soulève la discipline en cause, tant sur les plans technique et théorique, que sur les plans pratique et réel.

Remerciements

Le défi n'était pas moins grand pour nous de réaliser ce recueil d'économétrie qui certainement va relancer le débat sur la manière de présenter les travaux pratiques dans nos universités.

A présent, rendons à César, ce qui est à César ! Je tiens à remercier les Professeurs Jean-Pierre Bosonga et Daniel Mukoko que j'assiste respectivement au cours d'Econométrie et de Statistique appliquée à la finance. Leurs concours, par effet de diffusion et à travers les multiples discussions engagés avec eux, m'ont permis de voir au-delà de l'horizon.

De même, je ne saurai passer sous silence les multiples discussions engagées avec mes ainés scientifiques, particulièrement aux chefs de travaux : Alexandre Nshue Mbo Mokime et Blaise Nlemfu. Et aussi vifs remerciements aux assistants et chefs de travaux : Albert Lomboto et Karim Omonga.

Pour tous les sacrifices imposés aux personnes que nous aimons, nous tenons à leur dédier cet ouvrage.

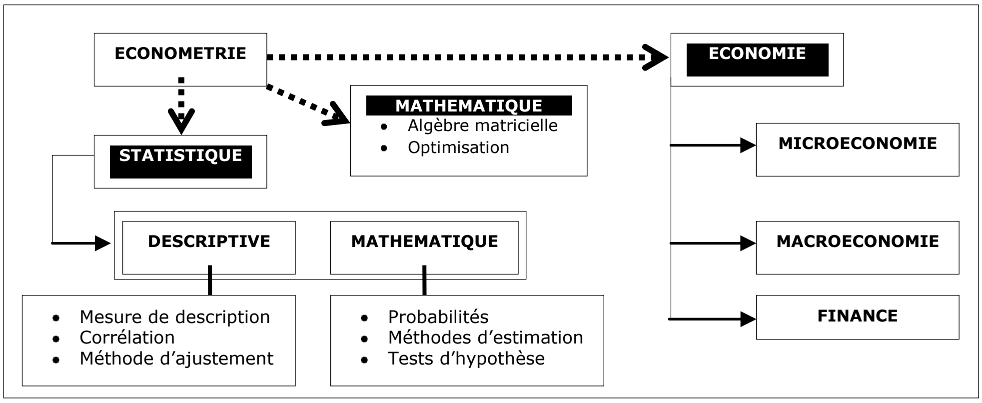
Bien entendu, selon la formule consacrée, ce document n'engage que son auteur. Toutes suggestions ou commentaires qui peuvent en améliorer le contenu sont le bienvenu.

Jean-Paul TSASA
Kinshasa, 26 Janvier 2010

INTRODUCTION

L'économétrie est une discipline technique qui est très intéressante puisqu'elle peut servir à la fois d'outil d'évaluation des théories et d'outil d'analyse de phénomènes et faits économiques. Ainsi, avant de chercher à comprendre « pourquoi étudier l'économétrie », il faut, tout d'abord, être à même de comprendre « qu'est-ce que l'économétrie ? » En voici une de réponses :

Schéma 1. Définition de l'économétrie



Simplement, l'économétrie peut être définie comme une branche de la statistique appliquée à l'économie. Par conséquent, pour une bonne malléabilité, elle nécessite une connaissance préalable de la statistique, analyse mathématique et théorie économique.

En vue de bien relayer le cours théorique à la pratique, nous construirons une passerelle nommée « RAPPELS & RESUME » pour aider les étudiants à joindre les connaissances acquises lors des enseignements [partie théorique] à l'esprit des applications retenues.

S'inscrivant dans cette logique, les applications retenues dans ce recueil seront, à chaque fois, précédées par une série de rappels et un résumé ; car comme l'affirme d'ailleurs Henri Poincaré, *rien n'est plus pratique qu'une bonne théorie* !

In fine, précisons que l'effort fourni dans cet ouvrage est de proposer des exercices qui correspondent au niveau d'exigence internationale. Ce n'est qu'à ce prix qu'il est possible de garantir la convertibilité nationale et internationale des savoirs acquis tout au long de la partie théorique.

MODULE 1-
THEORIE DE CORRELATION

1.1- CORRELATIONPARAMETRIQUE : coefficient de Bravais-Galton-Pearson

1.2- CORRELATION NON PARAMETRIQUE : coefficient de rang de Spearman et coefficient de rang de Kendall

1.3- CORRELATION TETRACHRORIQUE ET COEFFICIENT DE CONTINGENCE

OBJECTIF :

- Mesurer le degré et le sens de liaison entre deux ou plusieurs variables quantitatives ou qualitatives.
- Tester la significativité d'une relation entre variables supposées liées.
- Evaluer la normalité des distributions des échantillons.

Ce manuel d'exercices commence par la théorie de la corrélation. Après avoir étudié à fonds l'histoire et le développement de l'économétrie, nous nous sommes rendu compte que la théorie de la corrélation constitue l'ingrédient de base de tous les modèles de régression. L'analyse du coefficient de corrélation s'articule autour de huit points essentiels, pouvant faire l'objet d'un ouvrage sérieux ! Dans ce recueil d'applications, nous nous intéressons qu'à trois de ces point, il s'agit de la corrélation simple ; corrélation de rang et corrélation de données qualitatives. Cette analyse s'appliquera aussi bien aux chroniques qu'aux données en cross section.

1.1- CORRELATION PARAMETRIQUE : coefficient de Bravais-Galton-Pearson

1.1.1- Définition, calcul et test de significativité

Personne n'ignorerait que la *covariance* ou *variance généralisée* est une mesure qui permet de qualifier l'indépendance entre deux variables quantitatives. Mais puisque cet indicateur ne permet que de préciser la nature de dispersion des points, ne renseigne pas sur l'intensité de la relation et que par ailleurs, il est affecté par l'unité de mesure de variables considérées[1], Bravais Francis Galton et Karl Pearson se sont proposés de trouver un autre indicateur qui ne pourrait être affecté par l'échelle des mesures, mais aussi qui précise à la fois le sens et l'intensité ou le degré de dépendance entre séries statistiques[2]. Ainsi, en standardisant la covariance, ils obtiennent : le *coefficient de BGP*, généralement appelé « coefficient de corrélation linéaire simple ».

---

[1] En effet, la covariance renseigne sur l'inclinaison du nuage de points, mais elle ne donne aucune idée de l'intensité de la liaison existant entre les variables x et y. En conséquence, la covariance peut augmenter alors que la liaison entre x et y reste constante.

[2] Deux séries sont statistiques lorsqu'au moins une des variables est aléatoire.

A RETENIR !

Covariance: mesure de la dispersion conjointe de deux variables quantitatives (x et y) autour de leur moyenne.
Corrélation: mesure de la liaison entre deux variables x et y.
Corrélation linéaire de BGP : mesure de la liaison linéaire entre deux variables quantitatives x et y.

Le coefficient de corrélation linéaire simple[3] est *symétrique* et *sans dimension* ; il est noté par :

- $\rho(X_t, Y_t)$, lorsque les distributions des variables X et Y sont (ou supposées) connues ;

- $r(X_t, Y_t)$, lorsque les distributions des variables X et Y ne sont connues que pour un échantillon.

Ce coefficient ne représente que la partie linéaire de la liaison entre les variables X et Y. Et par ailleurs, il assume que les deux distributions sont à peu près normalement distribuées.
Le coefficient de corrélation est calculé par la formule :

$$\rho(X_t, Y_t) = \frac{Cov(X_t, Y_t)}{\sigma_X \sigma_Y}$$

Où $Cov(X, Y)$: désigne la covariance entre X et Y ; σ_X et σ_Y, respectivement les écarts-type de variables X_t et Y_t.

Et Puisque, généralement, les paramètres de la population ne sont pas connus, nous allons calculer $r(X, Y)$ qui est un estimateur du coefficient de corrélation. Ainsi, on a :

$$r(X_t, Y_t) = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum (X_t - \bar{X})^2 \sum (Y_t - \bar{Y})^2}}$$

En recourant aux variables centrées [écarts par rapport à la moyenne], la formule du coefficient de corrélation devient :

$$r(X_t, Y_t) = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$$

Où $x_t = X_t - \bar{X}$ et $y_t = Y_t - \bar{Y}$

Le coefficient de Bravais-Pearson peut également s'écrire sous forme d'une formule de « moyenne-écart type » :

$$r(X_t, Y_t) = \frac{\frac{\sum X_t Y_t}{n} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

[3] Une analyse comparative permet d'associer la covariance à la productivité marginale et le coefficient de corrélation linéaire à l'élasticité !

NOTE : Il est généralement souhaitable de présenter tout d'abord le graphique en nuage des points de variables sous analyse, avant de calculer le coefficient de Bravais-Pearson. Ce graphique permet de se rendre compte de la linéarité ou non de la relation des variables en cause.

1.1.2- Test de significativité d'un coefficient de Bravais-Pearson

Le résultat obtenu, après application de la formule du coefficient de Bravais-Pearson, doit être soumis à un test statistique avant sa validation. Ainsi, on applique le test de Student [test mis à jour par Gosset]. Ce test passe par 5 étapes.

(1) La spécification des hypothèses :

H0 : $\rho(X_t, Y_t) = 0 \; (absence\ de\ corrélation)$ Contre H1 : $\rho(X_t, Y_t) \neq 0 \; (absence\ de\ decorrélation)$

Il s'agit donc d'un test bilatéral ;

(2) La détermination du seuil de signification avec :

$$\alpha = Probabilité\ de\ rejeter\ H0, alors\ qu'elleest\ vraie$$

(3) Le calcul de la statistique du test qui est donnée par la formule :

$$t_{calculé} = \frac{|r(X_t, Y_t) - \rho(X_t, Y_t)|}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

Et sous l'hypothèse nulle, on a :

$$t_{calculé} = \frac{|r(X_t, Y_t)|}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

(4) La détermination de la statistique t sur la table de Student, avec un seuil de signification α et un degré de liberté $n - 1$:

$$t_{table} = t_{n-2}^{\alpha}$$

(5) la prise de décision : le coefficient $r(X_t, Y_t)$ est significatif si et seulement si la valeur de la statistique t calculé est supérieure à celle de la table.

Il existe également une table qui permet d'évaluer, pour un niveau de risque donné, la significativité du coefficient de corrélation simple, c'est la *table de Pearson-Hartley* [voir en annexe]. Ainsi, le coefficient de Bravais-Pearson serait significatif lorsque sa valeur est supérieure à celle de la table, pour un seuil de signification α quelconque.

A RETENIR !

L'interprétation du coefficient de Bravais-Pearson comprend toujours trois volets :

- Une interprétation par rapport au signe ou au sens de la liaison entre variables ;
- Une interprétation par rapport au degré ou à l'intensité de la liaison entre variables ;
- Et enfin, une interprétation par rapport à la significativité de la liaison supposée entre variables.

PAR RAPPORT AU SIGNE : le coefficient de Bravais-Pearson définit le sens de variation entre deux variables.

- Lorsque $r(X_t, Y_t) < 0$; la corrélation est dite négative c'est-à-dire : plus X_t augmente, plus Y_t diminue ou plus X_t diminue, plus Y_t augmente ;
- Lorsque $r(X_t, Y_t) > 0$; la corrélation est dite négative c'est-à-dire : plus X_t augmente, plus Y_t augmente ou plus X_t diminue, plus Y_t diminue.

PAR RAPPORT A LA VALEUR DU COEFFICIENT (signification clinique), le coefficient de Bravais-Pearson définit la force ou le degré de liaison entre deux variables. Lorsque la valeur du coefficient est [en module] :

- Proche de 0, ce qu'il y a absence de relation ;
- Entre 0,025 et 0,25, ce qu'il y existe une relation très faible entre les deux variables considérées ;
- Entre 0,25 et 0,50, ce que la relation entre $\mathrm{X_t}$ et $\mathrm{Y_t}$ est faible ;
- Entre 0,50 et 0,65, indique une relation modérée ;
- Entre 0,65 et 0,80 indique une relation forte [élevée] ;
- Entre 0,80 et 1 indique une relation très forte ; et de fois, cela traduit la présence de tautologie ou redondance.

PAR RAPPORT AU TEST DE SIGNIFICATIVITE : une relation entre variables n'est pas fictive si l'hypothèse nulle de présence de décorrélation entre variables est rejetée.

1.1.3- Comparaison de deux coefficients de corrélation

Pour savoir si deux coefficients de corrélation obtenus à partir d'échantillons de tailles respectives T1 et T2, diffèrent l'un de l'autre de façon significative, on calcule W1 et W2 correspondant au coefficient r1 et r2, en utilisant la *transformation de Fischer* :

$$Wi = \frac{1}{2} Ln\left(\frac{1+ri}{1-ri}\right) \qquad , avec\ i = 1{,}2$$

La statistique Wi suit approximativement une loi normale de moyenne μ_{Wi} et d'écart-type σ_{Wi}, tels que :

$$\mu_{Wi} = \frac{1}{2} Ln\left(\frac{1+\rho i}{1-\rho i}\right) \quad et \quad \sigma_{Wi} = \frac{1}{\sqrt{Ti-3}}$$

$$Où\ \rho i\ est\ le\ coefficient\ de\ corrélation\ théorique$$

Nous allons évaluer :

H0 : $\mu_{W1} = \mu_{W2}$ contre H1 : $\mu_{W1} \neq \mu_{W2}$; Il s'agit donc d'un test bilatéral.

Dès lors, on utilise la statistique z qui suit également la loi de Laplace-Gauss :

$$Z_C = \frac{W1 - W2 - \mu_{W1-W2}}{\sigma_{W1-W2}}$$

Avec :

$$\mu_{W1-W2} = \mu_{W1} - \mu_{W2}\ et\ \sigma_{W1-W2} = \sqrt{\sigma_{W1}^2 + \sigma_{W2}^2}$$

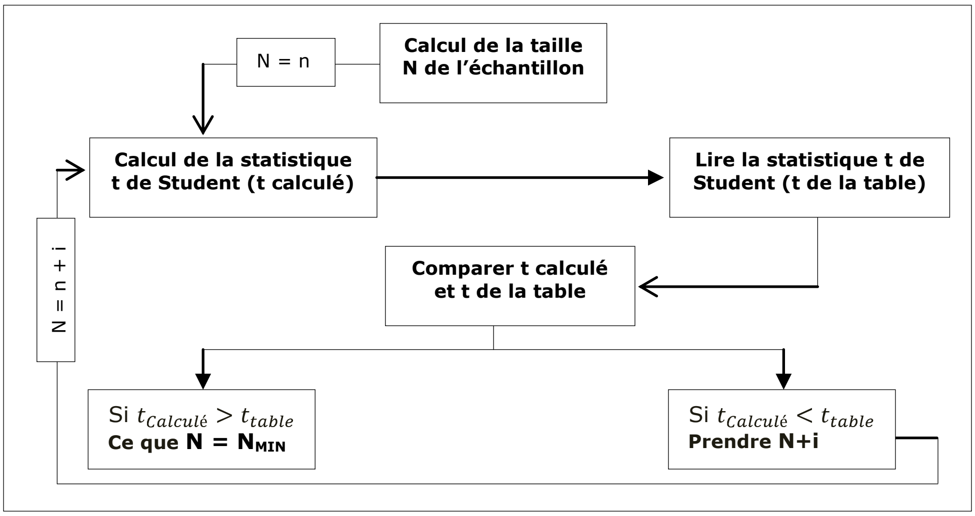
On rejette l'hypothèse nulle si, et seulement si : $Z_C > Z_t$ ou si $Z_C < -Z_t$; où Z_t est lue sur la table de la loi normale centrée-réduite, au seuil de signification quelconque.

1.1.4- Significativité du coefficient de corrélation et taille d'échantillon minimum

Les économistes s'intéressent de fois à la détermination de la taille minimale N_{MIN} nécessaire d'un échantillon correspondant à un coefficient de corrélation significativement différent de zéro. Ainsi, connaissant le degré de corrélation et le seuil de signification, il suffit donc de : (1) calculer N avec un degré de liberté infini par la formule $n = \frac{1}{[r(X_t,Y_t)]^2}\left\{2[r(X_t,Y_t)]^2 + [1 - r(X_t,Y_t)]\left(t_{\alpha,n-2}\right)^2\right\}$; (2) connaissant dès lors, en plus de degré de corrélation et de seuil de signification, la taille de l'échantillon ; calculer les statistiques t_c ; (3) comparer $t_{Calculé}$ et t_{table}, au seuil α et avec un degré de liberté n – 2. Si $t_{Calculé} > t_{table}$, ce que la taille « n » correspond à N_{MIN}. Dans le cas contraire, refaire l'étape (2) et (3), en considérant le même seuil de signification, mais avec n + i cette fois-ci ; et ainsi de suite jusqu'à ce que $t_{Calculé} > t_{table}$, où « i » est un entier naturel correspondant à la $i^{ième}$ itération {i = 1, …, k, …, n}.

La démarche ci-dessus peut se résumer comme suit (voir schéma ci – après).

Schéma 2. Significativité du coefficient de corrélation et taille d'échantillon minimum



1.1.5- Limites de confiance du coefficient de corrélation théorique

Jusque là, nous nous sommes intéressés à l'estimateur ponctuel du coefficient de corrélation théorique. A présent, nous cherchons à construire un intervalle de confiance qui doit contenir, à un seuil de signification quelconque, la vraie valeur du coefficient de la corrélation des variables de la population considérée.

La détermination de limites de confiance du coefficient de corrélation théorique passe par 2 étapes ; la détermination de : (1) limites inférieure et supérieure de la moyenne de la statistique W et puis, (2) celles du coefficient de corrélation théorique.

Soit la statistique $W = \frac{1}{2} Ln\left(\frac{1+r}{1-r}\right)$ qui suit la loi normale de :

- Moyenne :

$$\mu_W = \frac{1}{2} Ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$$

- Et d'écart-type :

$$\sigma_W = \frac{1}{\sqrt{n-3}};$$

Connaissant l'estimation du coefficient de corrélation de la population, le nombre d'observations et le seuil de signification, les intervalles de la moyenne de la statistique W sont obtenus par la formule ci-après :

$$\mu_W = \begin{cases} W - Z_\alpha \sigma_W & (limite\ inférieure) \\ W + Z_\alpha \sigma_W & (limite\ supérieure) \end{cases}$$

Où la statistique Z_α est lu dans la table de l'écart-réduit au seuil de signification de α.

Dès lors, il devient possible de déterminer les limites du coefficient de corrélation de la population en appliquant la formule que voici :

$$\rho_0 = \frac{e^{2\mu_W} - 1}{1 + e^{2\mu_W}}$$

A RETENIR !

Notez que :

- Le coefficient de Bravais-Pearson est égal à la covariance si les variables considérées ont de variances unités.
- Puisque la covariance est toujours inférieure ou égale au produit des écarts-type, le domaine de définition du coefficient de corrélation est donc : [-1, +1].
- Une valeur de $r(X_t, Y_t) = 1$ [en module] implique un lien fonctionnel linéaire tel que $\alpha_0 + \alpha_1 X_t + \alpha_2 Y_t = 0$; dans ce cas, on parle de relation linéaire parfaite ou tautologie.
- Pour $r(X_t, Y_t) = 0$ ou proche de zéro, on affirme que le lien entre les variables X et Y n'est faible que si ce dernier est linéaire ; puisqu'en réalité, un lien non linéaire fort peut conduire à une valeur faible du coefficient de corrélation.
- Les variables $\mathrm{X_t}$ et $\mathrm{Y_t}$ sont décorrélées lorsque le coefficient de corrélation est nul.
- La décorrélation n'implique pas nécessairement l'indépendance ! Deux variables indépendantes sont toujours décorrélées, alors que l'inverse n'est pas toujours vrai, sauf dans le cas où les variables sont toutes deux normales et de distribution conjointe binormale.
- Deux variables normales décorrélées et de distribution conjointe binormale sont indépendantes.

Précisons par ailleurs que le coefficient de Bravais-Pearson peut être trompeur pour les cas suivants :

(i) Lorsque les variables $\mathrm{X_t}$ et $\mathrm{Y_t}$ ont :

- Une cause commune ;
- Une relation non linéaire ;
- Une relation fortuite ;

(ii) Lorsque les variables $\mathrm{X_t}$ et $\mathrm{Y_t}$ sont simultanément considérées avec d'autres variables.

Pour faire face, respectivement, à ces faiblesses, les économètres déterminent : le sens de causalité, le coefficient de corrélation non linéaire, la cointégration et le coefficient de corrélation partielle.

Calcul du coefficient de Bravais-Pearson sur SPSS

Voici les étapes à suivre pour calculer le coefficient de Bravais-Pearson :

(1) Lancer le logiciel SPSS et cliquer sur « saisir des données », puis sur « OK »

(2) Aller dans la feuille « Affichage des variables » :
- Pour créer les variables, aller sur la colonne « Nom » et saisir par exemple sur la première ligne de ladite colonne « INFL » et sur la deuxième ligne « PIB » ;
- Aller sur la colonne « étiquette », spécifier : TAUX D'INFLATION et PRODUIT INTERIEUR BRUT

(3) Ensuite, aller dans la feuille « Affichage des données » :
- Saisir les valeurs de chaque variable pour chaque observation.

(4) Une fois la saisie des données de chaque variable finie, cliquer sur l'option « Analyse », puis sur « Corrélation » → « **Bivariée** » → « Pearson »

Ainsi, on obtient le coefficient de Bravais-Pearson.

NOTA : Contrairement à EVIEWS ou STATA, SPSS est parmi les logiciels qui, après calcul du coefficient de Bravais-Pearson, présentent le test de significativité.

1.2- CORRELATION NON PARAMETRIQUE : coefficient de rang de Spearman et coefficient de rang de Kendall

Comme vu précédemment, le coefficient de corrélation de Bravais-Pearson mesure la liaison pouvant exister entre deux variables. Cependant, comme nous le verrons par la suite, cet outil devient inopérant lorsque les distributions ne sont pas gaussiennes ou lorsqu'on est en face de variables ordinales ou discrètes.

Dans ces cas, on recourt à la corrélation des rangs, celle-ci ne prend plus en compte les observations mais leur classement par rapport à l'ordre observé sur l'autre variable. Dans ce recueil, nous nous intéressons au coefficient développé Spearman et à celui proposé par Kendall.

1.2.1- Coefficient de rang de Spearman

Le coefficient de rang de Galton-Spearman sert à vérifier l'existence de liens entre deux variables quantitatives mesurées sur une même chronique, à partir d'effectifs faibles ne permettant pas d'utiliser le coefficient de Bravais-Pearson. On remplace les valeurs de variables par leur rang. De même, lorsqu'on est en face de chroniques Xt et Yt qui ne reflètent pas les valeurs précises des variables ou encore quand la précision des mesures n'est pas possible, on peut ranger les données par ordre croissant, par exemple, de taille ou d'importance, en utilisant les nombres 1, 2, …, k, …, N.

On note Di, la différence entre les rangs des valeurs correspondantes de X et Y ; N, le nombre de couples de valeurs (X, Y). Ainsi, il devient possible de mesurer la corrélation entre Y_t et Y_{t-k} par le *coefficient de rang de Spearman* :

$$r_{RANG} = 1 - \frac{6\sum D_i^2}{N(N^2-1)}$$

On applique également le calcul de rang de Spearman lorsque l'hypothèse de normalité ou de symétrie est violée.

Noter que si les rangs sont identiques de variables, la corrélation de rang est absolue.

1.5.2- Test de significativité d'un coefficient de Spearman

Pour tester la significativité du coefficient de Spearman, on recourt au test de Student. Ce test passe par 5 étapes.

(1) La spécification des hypothèses :

$$\text{H0 : } \rho(X_t, Y_t) = 0 \ (il\ n'ya\ pas\ de\ corrélation\ des\ rangs)$$
$$\text{Contre H1 : } \rho(X_t, Y_t) \neq 0 \ (il\ existe\ une\ corrélation\ monotone)$$

Il s'agit donc d'un test bilatéral ;

(2) La détermination du seuil de signification avec :

$$\alpha = Probabilité\ de\ rejeter\ H0, alors\ qu'elleest\ vraie$$

(3) Le calcul de la statistique du test qui est donnée par la formule :

$$t_{calculé} = |r_{RANG} - \rho(X_t, Y_t)| \sqrt{\frac{n-2}{1-(r_{RANG})^2}}$$

Et sous l'hypothèse nulle, on a :

$$t_{calculé} = \frac{|r_{RANG}|}{\sqrt{\frac{1-(r_{RANG})^2}{n-2}}}$$

(4) La détermination de la statistique t sur la table de Student, avec un seuil de signification α et un degré de liberté $n-1$:

$$t_{table} = t_{n-2}^{\alpha}$$

(5) la prise de décision : le coefficient r_{RANG}est significatif si et seulement si la valeur de la statistique t calculé est supérieure à celle de la table.

Il existe également une table qui permet d'évaluer, pour un niveau de risque donné, la significativité du coefficient de corrélation de rang, c'est la *table de Spearman-Hartley* [voir en annexe]. Ainsi, le coefficient de corrélation de rang serait significatif lorsque sa valeur est supérieure à celle de la table, pour un seuil de signification α quelconque.

A RETENIR !

L'interprétation du coefficient de Spearman comprend toujours <u>trois volets</u> :

- Une interprétation par rapport au signe ou au sens de la liaison entre variables ;

- Une interprétation par rapport au degré ou à l'intensité de la liaison entre variables ;

- Et enfin, une interprétation par rapport à la significativité de la liaison supposée entre variables.

<u>PAR RAPPORT AU SIGNE</u> : le coefficient de Spearman définit le sens de variation entre deux variables.

- Lorsque $r_{RANG} < 0$; la corrélation est dite négative c'est-à-dire : plus X_t augmente, plus Y_t diminue ou plus X_t diminue, plus Y_t augmente ;

- Lorsque $r_{RANG} > 0$; la corrélation est dite négative c'est-à-dire : plus X_t augmente, plus Y_t augmente ou plus X_t diminue, plus Y_t diminue.

<u>PAR RAPPORT A LA VALEUR DU COEFFICIENT</u> (signification clinique), le coefficient de Spearman définit l'intensité de liaison entre deux variables. Lorsque la valeur du coefficient de corrélation de rang est [en module] :

- Proche de 0, ce qu'il y a absence de relation ;
- Entre 0,025 et 0,50, ce qu'il y existe une relation très faible entre les deux variables considérées ;
- Entre 0,50 et 0,70, ce que la relation entre $\mathrm{X_t}$ et $\mathrm{Y_t}$ est faible ;
- Entre 0,70 et 0,80, indique une relation modérée ;
- Entre 0,80 et 0,90 indique une relation forte [élevée] ;
- Entre 0,90 et 1 indique une relation très forte.

<u>PAR RAPPORT AU TEST DE SIGNIFICATIVITE</u> : une relation entre variables n'est pas fictive si l'hypothèse nulle de présence de décorrélation entre variables est rejetée.

Calcul du coefficient de Spearman sur SPSS
Voici les étapes à suivre pour calculer le coefficient de Spearman : (1) Lancer le logiciel SPSS et cliquer sur « saisir des données », puis sur « OK » (2) Aller dans la feuille « Affichage des variables » : - Pour créer les variables, aller sur la colonne « Nom » et saisir par exemple sur la première ligne de ladite colonne « INFL » et sur la deuxième ligne « PIB » ; - Aller sur la colonne « étiquette », spécifier : TAUX D'INFLATION et PRODUIT INTERIEUR BRUT (3) Ensuite, aller dans la feuille « Affichage des données » : - Saisir les valeurs de chaque variable pour chaque observation. (4) Une fois la saisie des données de chaque variable finie, cliquer sur l'option « Analyse », puis sur « Corrélation » → « **Bivariée** » → « Spearman » Ainsi, on obtient le coefficient de Spearman.

1.3- CORRELATION TETRACHRORIQUE ET COEFFICIENT DE CONTINGENCE

Les méthodes de calcul du coefficient de corrélation décrites précédemment ne permettent pas de mesurer le degré de dépendance de variables qui ne sont pas numériques par nature, notamment les *caractéristiques d'individus* ou *caractéristiques d'objets* [profession (économiste, médecin, ingénieur, …) ; grade (cadre, ouvrier, … ; gradué, licencié, docteur) ; couleurs de cheveux ou des yeux].

Pour mesurer la corrélation entre les caractéristiques d'individus ou d'objet [corrélation des attributs ou corrélation des caractéristiques], on recourt au tableau de contingence. Pour de tableaux de formats n x n, on définit la statistique de Khi-deux ou Khi-carré [inventée par Pearson], notée χ^2. Comme on s'intéresse à la *corrélation tetrachrorique*, n = 2 et ainsi obtient-on le tableau 1.1.

A chaque fréquence observée d'un tableau de contingence k x k, correspond une *fréquence théorique* qui, soumise à une hypothèse, s'obtient par le calcul des probabilités ; ces fréquences sont appelées *fréquences des cases*. Et, la fréquence totale sur chaque ligne ou chaque colonne est appelée *fréquence marginale*.

Tableau 1.1- Tableau de contingence k x k

	$A2$	…	$\overline{An}$	**Totaux**
$A1$	c_{11}	…	c_{1n}	M1
. . .	. . .	. . .	. . .	. . .
$\overline{An}$	c_{n1}	…	c_{nn}	Mn
Totaux	N1	…	Nn	T

A titre illustratif, Ai [en ligne] peut être la nationalité des individus, et Aj [en colonne], le niveau d'étude des individus [gradué, licencié, docteur].

Pour calculer le coefficient de corrélation des attributs, noté r^a, on utilise la formule :

$$r^a = \sqrt{\frac{\chi^2}{T(k-1)}}$$

Ainsi, le coefficient de corrélation des attributs varie entre 0 et 1. Plus ce coefficient est proche de l'unité, plus le degré de dépendance entre les variables considérées est élevé.

En considérant un tableau de contingence 2 x 2, on obtient :

Tableau 1.2- Tableau de contingence 2 x 2

	$\boldsymbol{A2}$	$\overline{\boldsymbol{A2}}$	**Totaux**
$\boldsymbol{A1}$	c_{11}	c_{12}	M1
$\overline{\boldsymbol{A1}}$	c_{21}	c_{22}	M2
Totaux	N1	N2	T

Et par conséquent, le coefficient de corrélation tetrachrorique, noté r^t, est donné par :

$$r^t = \sqrt{\frac{\chi^2}{T}}$$

Pour calculer la statistique χ^2, on utilise la formule suivante[4] :

$$\chi^2 = \frac{T(c_{11}c_{22} - c_{12}c_{21})}{(c_{11} + c_{21})(c_{12} + c_{22})(c_{11} + c_{12})(c_{21} + c_{22})}$$

Comme le suggère la statistique inférentielle, lorsqu'on applique les résultats d'une distribution, telle que celle de Khi-deux, à des données discrètes, on doit procéder à certaines corrections de « continuité », en remplaçant la formule de χ^2 par celle de $\chi^2(corrigé)$. Pour ce faire, on applique la *correction de continuité de Yates*, du nom du statisticien anglais Frank Yates :

$$\chi^2(corrigé) = \frac{T\left(|c_{11}c_{22} - c_{12}c_{21}| - \frac{1}{2}T\right)^2}{(c_{11} + c_{21})(c_{12} + c_{22})(c_{11} + c_{12})(c_{21} + c_{22})}$$

In fine, le coefficient de corrélation tetrachrorique devient :

$$r^t(corrigé) = \sqrt{\frac{\chi^2(corrigé)}{T}}$$

Parallèlement, le *coefficient de contingence* permet également de mesurer le degré de dépendance et de corrélation des caractères ou attributs étudiés dans un tableau de contingence. Son domaine de variation va de zéro à l'unité.

$$r^c = \sqrt{\frac{1}{1 + \frac{T}{\chi^2}}}$$

---

[4] Pour un tableau de contingence m x n, on a : $\chi^2 = \sum_{i=1}^{m} \frac{T}{Mi}\left[\sum_{j=1}^{n} \frac{(c_{IJ})^2}{Nj}\right] - T$

La valeur maximale[5] que prend le coefficient de corrélation des tables de contingence est :

$$\sqrt{\frac{(k-1)}{k}}$$

Partant, il est possible de montrer que le coefficient de corrélation des attributs varie de 0 à 1 :

$$\sqrt{\frac{1}{1+\frac{T}{\chi^2}}} \leq \sqrt{\frac{(k-1)}{k}}$$

Une manipulation algébrique simple nous permet d'écrire :

$$\frac{\chi^2}{\chi^2+T} \leq \frac{k-1}{k}$$

Ainsi, on a :

$$\chi^2 \leq (k-1)T$$

Et donc :

$$r^a = \sqrt{\frac{\chi^2}{T(k-1)}} \leq 1$$

Calcul du coefficient de caractère sur Eviews

Pour déterminer si un attribut **X** [par exemple, le sexe de la personne] intervient significativement dans la réalisation d'une autre variable **Y** [par exemple, le fait de rembourser ou non le crédit à l'échéance], nous allons effectuer le test suivant :

- **Spécification des hypothèses :**
  H0 : indépendance des caractères
  H1 : dépendance des caractères

- **Règle de décision :**
  Rejeter H0 si la p-value est inférieure à 0,05

DETERMINATION DU TABLEAU EFFECTIF

Commande :
- QUICK → SHOW → X Y puis OK
- VIEW → N-WAY TABULATION → COUNT puis OK

CALCUL DU COEFFICIENT DE CARACTERE

Commande :
- QUICK → SHOW → X Y puis OK
- VIEW → N-WAY TABULATION → COUNT & CHI-SQUARE TESTS puis OK

[5] Lorsque le coefficient de contingence atteint sa valeur maximale, les deux attributs sont totalement dépendants et les fréquences théoriques sont tous égaux.

1.4- TESTs DE NORMALITE : Agostino – Pearson ; de Jarque – Bera & Shapiro - Wilk

En pratique, le choix entre l'utilisation du coefficient de Bravais-Pearson et celui de Spearman est dicté par la nature de la distribution de l'échantillon. Lorsque la distribution est gaussienne, on recourt au calcul du coefficient de corrélation paramétrique, notamment le coefficient de corrélation du produit des moments de Pearson. Dans le cas contraire, c'est-à-dire lorsque la distribution de l'échantillon ne suit pas une loi normale, on calcule le coefficient de corrélation non paramétrique, notamment le coefficient de rang de Spearman.

Cependant, jusque là, nous ne nous sommes pas encore intéressés aux méthodes ou techniques permettant de préciser la distribution que suit une variable donnée. Ainsi, dans cette section, il sera question d'aborder le test permettant de préciser la nature de la distribution d'un échantillon : il s'agit du test de Jarque-Bera.

Le test d'hypothèse est le suivant :

H0 : Xt suit une loi normale *contre* **H1** : Xt ne suit pas une loi normale

La statistique de Jarque-Bera, en se basant sur le calcul des coefficients de dissymétrie [skewness] et d'aplatissement [kurtosis], est définie par :

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

Avec :

- n : la taille de l'échantillon

- Coefficient de dissymétrie (S) et Coefficient d'aplatissement (K) :

$$S = \frac{\mu_3}{\mu_2\sqrt{\mu_2}} \quad et \quad K = \frac{\mu_4}{\mu_2^2} - 3$$

Où μ_i désigne le moment d'ordre i

Test de Jarque-Bera sur Eviews

Après avoir saisi les données des variables X et Y, effectuer les commandes suivantes :

- *Aller dans* : QUICK → GROUP STATISTICS → DESCRIPTIVE STATISTICS → COMMON SAMPLE → X Y
- *Puis cliquer sur* : OK

La statistique JB suit, sous l'hypothèse de normalité, une loi du Khi-Deux à deux degré de liberté. Au seuil de 5 %, on lit dans la table du Khi-Deux : 5,99. Ainsi, on ne rejette pas l'hypothèse nulle de normalité si la statistique de Jarque-Bera est inférieure à 5,99 [pour les grands échantillons] ; autrement, lorsque la probabilité critique est supérieure à 0,05, on admet l'hypothèse de normalité.

APPLICATIONS

APPLICATION 1

La SONAS a détecté une forte corrélation positive entre le nombre de pompiers présents sur les sites d'incendies et le montant des remboursements qui lui sont réclamés.

TRAVAIL A FAIRE :

(a) Doit-elle en conclure que les pompiers sont de gens néfastes car « plus il y a de pompiers, et plus ça nous coûte cher » ? Si oui, pourquoi ? Si non, quelle variable peut-elle prendre en compte pour expliquer cette corrélation ?
(b) Peut-on conclure qu'il existe une relation causale entre le nombre de pompiers et l'importance des remboursements ?

APPLICATION 2

Les prix moyens des actions et des obligations en Bourse pour la période 1950 à 1959 sont donnés au tableau ci-après :

On dispose des données suivantes :

Temps	**Prix moyens des actions [en Euro]**	**Prix moyens des obligations [en Euro]**	Temps	**Prix moyens des actions [en Euro]**	**Prix moyens des obligations [en Euro]**
2000	35,22	102,43	2005	53,29	100,07
2001	39,87	100,93	2006	54,14	97,08
2002	41,85	97,43	2007	49,12	91,59
2003	43,23	97,81	2008	40,71	94,85
2004	40,06	98,32	2009	55,15	94,65

TRAVAIL A FAIRE :

(a) Calculez le coefficient de Bravais-Pearson et interprétez les résultats.
(b) Peut-on conclure au seuil de signification de 0,05 que le coefficient de corrélation de la population n'est pas nul ?
(c) Calculez le coefficient de corrélation de rang des données et comparez les résultats avec la réponse trouvée en (a).

RESOLUTION

Notons par Xt, les prix moyens des actions et par Yt, les prix moyens des obligations.

Connaissant la formule du coefficient de corrélation de Bravais-Pearson :

$$r(X_t, Y_t) = \frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum(X_t - \bar{X})^2 \sum(Y_t - \bar{Y})^2}}$$

Voici la démarche qui nous conduira à la détermination du coefficient $r(X_t, Y_t)$:

- Calcul des moyennes des variables Xt et Yt : $\bar{X}$ et $\bar{Y}$
- Calcul de la covariance
- Calcul des écarts-types de Xt et Yt

Nous allons nous servir d'un tableau ad hoc à cet effet.

Temps	X_t	Y_t	$X_t - \bar{X}$	$(X_t - \bar{X})^2$	$Y_t - \bar{Y}$	$(Y_t - \bar{Y})^2$	$(X_t - \bar{X}) * (Y_t - \bar{Y})$
2000	35,22	102,43	-10,044	100,881936	4,914	24,147396	-49,356216
2001	39,87	100,93	-5,394	29,095236	3,414	11,655396	-18,415116
2002	41,85	97,43	-3,414	11,655396	-0,086	0,007396	0,293604
2003	43,23	97,81	-2,034	4,137156	0,294	0,086436	-0,597996
2004	40,06	98,32	-5,204	27,081616	0,804	0,646416	-4,184016
2005	53,29	100,07	8,026	64,416676	2,554	6,522916	20,498404
2006	54,14	97,08	8,876	78,783376	-0,436	0,190096	-3,869936
2007	49,12	91,59	3,856	14,868736	-5,926	35,117476	-22,850656
2008	40,71	94,85	-4,554	20,738916	-2,666	7,107556	12,140964
2009	55,15	94,65	9,886	97,732996	-2,866	8,213956	-28,333276
Total :	**452,64**	**975,16**		**449,39204**		**93,69504**	**-94,67424**
Moyenne :	45,264	97,516					
Ecart-type :	**6,70367**	**3,06096**			**Covariance =**		**-9,467424**

Par conséquent, le coefficient de Bravais-Pearson est :

$$r(X_t, Y_t) = \frac{-9{,}46742400000001}{(6{,}70367093464469)(3{,}06096455386207)}$$

Et on obtient : $r(X_t, Y_t) = -0{,}46138220392328$

Il existe une corrélation faible et négative entre les prix moyens des actions et les prix moyens des obligations.

Pour s'assurer de la significativité de cette faible liaison négative entre les prix moyens des actions et les prix moyens des obligations, nous devons effectuer le test de Student, en évaluant : **H0 :** *les variables Xt et Yt sont décorrélées* contre **H1 :** *les variables Xt et Yt sont corrélées*.

Seuil de signification : 0,05

Statistique de Student :

$$t_{calculé} = \frac{|r(X_t, Y_t) - \rho(X_t, Y_t)|}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

Ainsi, sous l'hypothèse nulle, on obtient : $t_{calculé} = \frac{|-0{,}46138220392328|}{\sqrt{\frac{1-(-0{,}46138220392328)^2}{10-2}}}$

Après calcul, on a : $t_{calculé} = 1{,}47090145064091$

En confrontant le t de Student calculé au t de Student de la table au seuil de signification de 5 % [$t_{table} = 2.228$], il ressort que l'on ne peut pas rejeter l'hypothèse nulle. Ainsi, on admet qu'il n'existe pas de relation significative entre les prix moyens des actions et les prix moyens des obligations. Cette liaison linéaire semble donc fictive !

NOTE : On peut également recourir à la table de Pearson-Hartley pour décider de la significativité du coefficient de Bravais-Pearson. Connaissant : $r(X_t, Y_t) = -0{,}46138220392328$

Nombre d'observations : 10 ; Degré de liberté : 2. On obtient la valeur de la table de Pearson au seuil de signification de 5% : 0,6319.

Puisque la valeur du coefficient de Bravais-Pearson est inférieure à celle de la table de Pearson-Hartley, on conclut donc que la relation entre les prix moyens des actions et les prix moyens des obligations n'est pas significative.

Il est donc clair qu'au seuil de signification de 0,05, le coefficient de corrélation de la population est nul.

Le coefficient de corrélation de rang est donné par : $r_{RANG} = 1 - \frac{6\sum D_i^2}{N(N^2-1)}$

Déterminons le rang de Xt et celui de Yt, afin de déterminer la différence de rang Di, ainsi que son carré :

Temps	X	Y	Rang X	Rang Y	Di	Di²
2000	35,22	102,43	1	10	-9	81
2001	39,87	100,93	2	9	-7	49
2002	41,85	97,43	5	5	0	0
2003	43,23	97,81	6	6	0	0
2004	40,06	98,32	3	7	-4	16
2005	53,29	100,07	8	8	0	0
2006	54,14	97,08	9	4	5	25
2007	49,12	91,59	7	1	6	36
2008	40,71	94,85	4	3	1	1
2009	55,15	94,65	10	2	8	64
TOTAL	452,64	975,16			0	272

Partant de données du tableau ci-dessus, il y a lieu de calculer le coefficient de rang de Spearman : $r_{RANG} = 1 - \frac{6(272)}{10[(10)^2-1]}$

On obtient :

$$r_{RANG} = -0{,}648484848484848$$

Au regard du coefficient de corrélation de rang (corrélation non paramétrique), il existe une faible liaison négative entre les prix moyens des actions et les prix moyens des obligations.

Précisons à présent si le coefficient de rang est significatif ou non :
H0 : il y a corrélation de rang *contre* H1 : il n'y a pas corrélation de rang

Seuil de signification : 0,05

Statistique de Student :

$$t_{calculé} = \frac{|r_{RANG}|}{\sqrt{\frac{1-(r_{RANG})^2}{n-2}}}$$

Ainsi, on a :

$$t_{calculé} = \frac{|-0{,}648484848484848|}{\sqrt{\frac{1-(-0{,}648484848484848)^2}{10-2}}}$$

Après calcul, on a :

$$t_{calculé} = 2{,}40951881774409$$

En comparant le t de Student calculé au t de Student de la table au seuil de signification de 5 % [t_{table} = 2.228], il ressort que :

$$t_{calculé} > t_{table}$$

En conséquence, on rejette l'hypothèse nulle. Ainsi, on admet qu'il existe une corrélation de rang significative entre les prix moyens des actions et les prix moyens des obligations. Cette liaison linéaire semble donc fictive !

NOTE : On peut également recourir à la table de Spearman pour décider de la significativité du coefficient de corrélation de rang. Connaissant :

$$r_{RANG} = -0{,}648484848484848$$

<u>Nombre d'observations</u> : 10 ; <u>Degré de liberté</u> : 2. On obtient la valeur de la table de Pearson au seuil de signification de 5% : 0,6483. Puisque la valeur du coefficient de rang est supérieure à celle obtenue dans la table de Spearman, on conclut donc que la corrélation de rang entre les prix moyens des actions et les prix moyens des obligations est significative.

Tableau récapitulatif

	Coefficient de Bravais-Pearson	**Coefficient de rang de Spearman**
Formule	$\frac{\sum(X_t-\bar{X})(Y_t-\bar{Y})}{\sqrt{\sum(X_t-\bar{X})^2\sum(Y_t-\bar{Y})^2}}$ *Corrélation paramétrique*	$1-\frac{6\sum D_i^2}{N(N^2-1)}$ *Corrélation non paramétrique*
Valeur	$-\,0{,}46138220392328$	$-0{,}648484848484848$
Intensité de liaison	Faible	Faible
Sens de liaison	Relation inversement proportionnelle	Relation inversement proportionnelle
Significativité de liaison	Non significative	Significative

APPLICATION 3

A partir d'un échantillon de 27 objets, on a trouvé que la valeur d'un coefficient de corrélation état 0,4. Peut-on en conclure, à un seuil de signification de 0,05 que le coefficient de corrélation diffère significativement de la valeur zéro ? Qu'adviendrait la réponse obtenue précédemment si l'on considère un seuil de signification de 0,01.

APPLICATION 4

Le coefficient de corrélation entre les notes de macroéconomie et de microéconomie d'un groupe de 21 étudiants est de 0,80. Calculez les limites de confiance à 0,95 de ce coefficient.

RESOLUTION

On sait :

- Taille de l'échantillon : n = 21
- $r(X_t, Y_t) = 0{,}80$
- Seuil de signification : 0,05

Déterminons les limites inférieure et supérieure de la moyenne de la statistique W :

$$\mu_W = \begin{cases} W - Z_\alpha \sigma_W & (limite\ inférieure) \\ W + Z_\alpha \sigma_W & (limite\ supérieure) \end{cases}$$

Connaissant : $W = \frac{1}{2} Ln\left(\frac{1+r}{1-r}\right) = 0{,}5 * Ln(9)$ et $\sigma_W = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{18}}$

On obtient :

$$\mu_W = \begin{cases} W - Z_\alpha \sigma_W \rightarrow \mu_W = 1{,}09861228866811 - 1{,}96 * 0{,}235702260395516 \\ W + Z_\alpha \sigma_W \rightarrow \mu_W = 1{,}09861228866811 - 1{,}96 * 0{,}235702260395516 \end{cases}$$

Les limites du coefficient de corrélation de la population sont :

$$\rho_0 = \begin{cases} 0{,}562604594776099\ (limite\ inférieure) \\ 0{,}915515780589372\ (limite\ supérieure) \end{cases}$$

APPLICATION 5

La valeur d'un coefficient de corrélation, calculée à partir d'un échantillon de taille 18, est égale à 0,32. Peut-on en conclure au seuil de signification de 0.01 que le coefficient de corrélation de la population n'est pas nul ?

APPLICATION 6

Quelle serait la taille minimale nécessaire d'un échantillon pour pouvoir conclure que le coefficient de corrélation 0,32 diffère significativement de zéro au seuil de 5 % ?

RESOLUTION

On sait :

- $r(X_t, Y_t) = 0{,}32$
- Seuil de signification : 0,05

Déterminons la taille de l'échantillon en appliquant la formule suivante :

$$n = 2 + [1 - r(X_t, Y_t)]\left[\frac{t_{\alpha,n-2}}{r(X_t, Y_t)}\right]^2$$

Ainsi, on a :

$$n = 28$$

Connaissant, à présent, la taille de l'échantillon, calculons la statistique t de Student :

$$t_{calculé} = \frac{|0{,}32|}{\sqrt{\frac{1-(0{,}32)^2}{28-2}}} = 1{,}72224618080818$$

En comparant le t calculé au t de la table de Student au seuil de 0,05 ; il ressort que la taille de l'échantillon *n = 28* n'est pas minimum. Après plusieurs itérations, il ressort que la taille minimale est : *n = 36*.

APPLICATION 7

Trouvez les limites de confiance à 0.95 pour un coefficient de corrélation dont la valeur calculée à partir d'un échantillon de 28 objets est égale à 0.6.

APPLICATION 8

On a calculé que deux coefficients de corrélation obtenus à partir d'échantillons de taille 23 et 28 ont pour valeurs respectives 0.8 et 0.95. Peut-on en conclure, au seuil de signification de 0.01 qu'il y a différence significative entre les deux coefficients ? Qu'adviendrait le résultat obtenu précédemment si l'on considère un niveau de confiance de 0.95 ?

APPLICATION 9

On a indiqué sur le tableau suivant, pour 10 étudiants, leur classement respectif en travaux pratiques et en questions de cours de Modèles macroéconomiques.

Etudiants	TOMBOLA Muke	BILETIKA Arnaud	ELONGA Mboyo	NTUMBA Ngandu	NZINGA Senda
Travaux pratiques	8	3	9	2	7
Cours	9	5	10	1	8

Etudiants	HANGI Banyene	KABATA Jibikilay	KABONGO Ntamb.	KULUNGU Malon.	MALENGERA Sya.
Travaux pratiques	10	4	6	1	5
Cours	7	3	4	2	6

Calculez le coefficient de corrélation de rang de ces notes. Et interprétez.

APPLICATION 10

On a calculé les coefficients de corrélation pour deux échantillons de tailles N1 = 28 et N2 = 35 et l'on a trouvé respectivement r1 = 0.50 et r2 = 0.30. Y a-t-il une différence significative entre les deux coefficients au seuil de 5 % ?

APPLICATION 11

On prélève un échantillon de malade ayant suivi ou non un traitement contre une épidémie quelconque. Au bout de la période de traitement, les enquêteurs constatent que dans ce groupe : sur les 200 individus malades, il y a 100 qui ont suivi un traitement et parmi les 60 individus qui ne sont pas guéris, 25 ont suivi le traitement.

TRAVAIL A FAIRE :

(a) Construire le tableau de contingence.
(b) Quel est le coefficient de corrélation sans la correction de Yates ?
(c) Qu'adviendrait à la réponse obtenue en (b), si l'on intègre dans le calcul la correction de Yates ?
(d) Trouver la valeur maximale du coefficient de contingence.

APPLICATION 12

On a demandé à deux membres du jury de Miss VODACOM de ranger 8 candidates A, B, C, D, E, F, G et H par ordre de préférence. Les choix sont donnés par le tableau suivant.

Candidates miss VODACOM	A	B	C	D	E	F	G	H
1^{ier} membre	5	2	8	1	4	6	3	7
2^{ième} membre	4	5	7	3	2	8	1	6

Calculez le coefficient de corrélation de rang et dire si les juges sont d'accord sur leurs choix.

APPLICATION 13

Le tableau ci-dessous indique le nombre des étudiants de deux classes A et B qui ont réussi et échoué à un examen subi par les deux classes.

	ont réussi	*ont échoué*
Classe A	72	17
Classe B	64	23

TRAVAIL A FAIRE :

(a) Dérivez le coefficient de corrélation tetrachrorique sans la correction de Yates.
(b) Même question que (a), mais avec la correction de Yates.
(c) Qu'en est-il du coefficient du coefficient de contingence avec, et sans la correction de Yates ?

APPLICATION 14

Le tableau ci-dessous montre la relation qu'il y a entre les résultats obtenus par des étudiants en économétrie et en microéconomie.

		ECONOMETRIE		
		Bonnes notes	**Notes moyennes**	**Mauvaises notes**
MICROECONOMIE	**Bonnes notes**	56	71	12
	Notes moyennes	47	163	38
	Mauvaises notes	14	42	85

Calculez le coefficient de corrélation des notes d'économétrie et de microéconomie.

APPLICATION 15

Soit le tableau ci-dessous représentant respectivement les coûts de la promotion d'un produit et les recettes réalisée par la vente de ce produit [les montants sont exprimés en millions de CDF].

On dispose des données suivantes :

Temps	**Coûts de la promotion**	**Recettes réalisées**
1	32.5	136
2	32.5	132
3	33.5	136
4	32	130
5	34	138
6	31	132

Temps	**Coûts de la promotion**	**Recettes réalisées**
7	35	136
8	33	130
9	34	142
10	33.5	134
11	34.5	136
12	35.5	140

TRAVAIL A FAIRE :

(a) Calculez les variations totale, résiduelle et expliquée.
(b) Calculez le coefficient de signification.
(c) Dérivez le coefficient de corrélations.
(d) Calculez le coefficient de corrélation de rang des données et comparez le résultat avec la réponse trouvée en (c).

APPLICATION 16

Si le coefficient de corrélation entre X et Y vaut 0.5, quel est le pourcentage de la variance totale qui reste inexpliquée par l'équation de régression ?

APPLICATION 17

Le tableau suivant donne l'âge X et la tension artérielle Y de 12 femmes :

Individu	1	2	3	4	5	6	7	8	9	10	11	12
Age (X)	56	42	72	36	63	47	55	49	38	42	68	60
Tension artérielle (Y)	136	132	136	130	138	132	136	130	142	134	136	140

TRAVAIL A FAIRE :

(a) Calculez le coefficient de corrélation des variables X et Y.
(b) Déterminez l'équation de la droite de régression des moindres carrés de Y et X.
(c) Estimez la tension artérielle d'une femme âgée de 50 ans.

APPLICATION 18

On dispose des données suivantes :

Temps	$X_t - \overline{X}$	$Y_t - \overline{Y}$	**Temps**	$X_t - \overline{X}$	$Y_t - \overline{Y}$
1	-6	-4	5	1	0
2	-4	-3	6	2	2
3	-3	-1	7	4	3
4	-1	-1	8	7	4

Calculez le coefficient de corrélation linéaire des variables X et Y.

APPLICATION 19

Le tableau suivant montre les taux de naissance et de mortalité pour 1000 personnes aux Etats-Unis de 1915 à 1955.

Période	**Taux de naissance** (pour 1000 personnes)	**Taux de mortalité** (pour 1000 personnes)
1915	25.0	13.2
1920	23.7	13.0
1925	21.3	11.7
1930	18.9	11.3
1935	16.9	10.9
1940	17.9	10.8
1945	19.5	10.6
1950	23.6	9.6
1955	24.6	9.3

Calculez le coefficient de corrélation et interprétez le résultat.

APPLICATION 20

Le coefficient de corrélation de deux variables X et Y est r = 0.60. Si les écarts-type de X et Y sont respectivement 1.50 et 2 ; et leurs moyennes, respectivement, 10 et 20. Trouvez les équations de régression de Y en X et de X en Y.

APPLICATION 21

Un enquêteur prélève les données suivantes auprès d'une institution de micro-finance basée à Kinshasa.

observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0
Y	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1

Où X = 1, si la personne a remboursé son crédit à l'échéance ; X = 0, si c'est le contraire ; Y = 1 si la personne est du genre masculin et Y = 0, c'est féminin.

TRAVAIL A FAIRE :

Tester si le sexe de la personne intervient significativement dans le fait de rembourser ou non le crédit à l'échéance.

BIBLIOGRAPHIE SOMMAIRE

- Akaike, H., 1974, A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, AC-19: 716-723.
- Anderson, D.R., K.P. Burnham and G.C. White, 1998, Comparison of Akaike Information Criterion and Consistent Akaike Information Criterion for Model Selection and Statistical Inference from Capture-recapture Studies, Journal of Applied Statistics, 25: 263-282.
- Borgard, F. et D. Guégan, 1996, *Etudes de séries chroniques linéaires à temps discret. Comparaison de logiciels,* Revue de Statistique Appliquée, tome 44, n°4 (1996), pp. 59-80.
- Brockwell, P.J. and R.A. Davis, 1991, *Time Series: Theory and Methods*, 2nd Edition, Springer.
- Buckland, S.T., K.P. Burnham and N.H. Augustin, 1997, Model Selection: An Integral Part of Inference, Biometrics, 53: 603-618.
- Ducharme, G.R. 1997, Consistent Selection of the Actual Model in Regression Analysis, Journal of Applied Statistics, 24: 549-558.
- Galbraith, J.W. et V. Zinde-Walsh, 2002, Measurement of the Quality of Autoregressive Approximation, with Econometric Applications, in A. Ullah, A. Wan et A. Chaturvedi (éds), Handbook of Applied Econometrics and Statistical Inference, Marcel Dekker, New York, p. 401-421.
- Geweke, J. et R. Meese, 1981, Estimating Regression Models of Finite but Unknown Order, International Economic Review, 22 : 55-70.
- Gouriéroux, C. et A. Montfort, 1995, *Séries temporelles et modèles dynamiques*, 2ième édition, Economica.
- Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.
- Hannan, E., and B. Quinn, 1979, *The determination of the order of an autoregression*, Journal of the Royal Statistical Society, B41, 190–195.
- Hannan, E.J. et B.G. Quinn, 1979, The Determination of the Order of an Autoregression, Journal of the Royal Statistical Society Ser. B, 41 : 190-195.
- Hurvich, C.M. et C. Tsai, 1989, Regression and Time Series Model Selection in Small Samples, Biometrika, 76: 297-307.
- King, M. et G. Bose, 2002, Finding Optimal Penalties for Model Selection in the Linear Regression Model, in Giles, D.E.A., (éd.), Computer-Aided Econometrics, Marcel Dekker, New York.
- Koehler, A.B. and E.S. Murphree, 1988, A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order, Applied Statistics, 37.
- Kullback, L. et R.A. Leibler, 1951, On Information and Sufficiency, Annals of Mathematical Statistics, 22 : 79-86.
- Lütkepohl, H. and M. Krätzig, 2004, *Applied Time Series Econometrics*, Themes in Modern Econometrics, Cambridge University Press, Cambridge.

ANNEXE
Les tests d'hypothèse les plus usuels en Statistique inférentielle

1. Tests portant sur un seul échantillon

Lorsque le test porte sur un échantillon, il y a lieu de distingues, selon les ca, trois types de tests statistiques : les tests de conformité à un standard ; les tests d'adéquation à une loi et les tests de symétrie des répartitions. Certains sont paramétriques [c'est-à-dire tests portant sur de données issues d'une distribution paramétrique] et d'autres non paramétriques [c'est-à-dire lorsque les tests ne font aucune hypothèse sur la distribution sous-jacente des données].

Type de test	Tests paramétriques	Tests non paramétriques
Tests de conformité à un standard	Tests de conformité [test de Student] - d'une moyenne - d'un écart type - d'une proportion	
Tests d'adéquation à une loi		- Test de Kolmogorov-Smirnov - Test d'adéquation du $\chi 2$ - Test de Shapiro-Wilks - Test de Lilliefors - Test d'Anderson-Darling - Test de D'Agostino - Test de Jarque-Bera
Tests de symétrie des répartitions		- Test de Wilcoxon - Test de Van der Waerden

2. Tests de comparaison de populations

Lorsque le test porte sur plus de deux échantillons, on distingue généralement : les tests de comparaison [ou test d'homogénéité] ; les tests d'appariement et les tests multivariés [c'est-à-dire lorsqu'ils mettent en jeu simultanément plusieurs variables]. Les tests de comparaison portant sur les échantillons indépendants peuvent classifiés selon que l'on s'intéresse : soit aux caractéristiques de tendance centrale ou modèles de localisation, soit aux caractéristiques de dispersion ou modèle d'échelle.

Type de test	Tests paramétriques	Tests non paramétriques
Tests de comparaison de populations, les fonctions de répartition sont les mêmes dans les groupes		- Test de Kolmogorov - Smirnov - Test de Kuiper - Test de Cramer - von Mises
Tests de comparaison de K échantillons indépendants [différenciation selon les caractéristiques de tendance centrale, modèle de localisation]	- Test de comparaison de moyennes (K = 2) - ANOVA [Analyse de la Variance] à 1 facteur	- Test de la somme des rangs de Wilcoxon (K=2) - Test de Mann - Whitney (K=2) - Test de Kruskal - Wallis - Test des médianes - Test de Van der Waerden - Test de Jonckheere - Terpstra [alternatives ordonnées]

Tests de comparaison de K échantillons indépendants [différenciation selon les caractéristiques de dispersion, modèle d'échelle]	- Test de Fisher (K=2) - Test de Bartlett - Test de Cochran - Test F-max de Hartley - Test de Levene - Test de Brown-Forsythe	- Test de Ansari - Bradley - Test de Klotz - Test de Mood - Test de Siegel-Tukey - Test des différences extrêmes de Moses
Tests d'appariement pour K échantillons [mesures répétées ou blocs aléatoires complets]	- Test de Student de comparaison de moyennes pour échantillons appariés (K=2) - Test de comparaison de variances pour échantillons appariés (K=2) - ANOVA pour blocs aléatoires complets	- Test des signes (K=2) - Test des rangs signés de Wilcoxon (K=2) - Test de Friedman - Test de Page [alternatives ordonnées] - Test de McNemar [K=2, variables binaires] - Test Q de Cochran [variables binaires]
Tests multivariés pour K échantillons indépendants	- T² de Hotelling, comparaison de K=2 barycentres [vecteur des moyennes] - MANOVA [analyse de variance mutlivariée], comparaison de K barycentres : Lambda de Wilks, Trace de Pillai, Trace de Hotelling-Lawley, La plus grande valeur propre de Roy - Test M de Box de comparaison de matrices de variance covariance	

- **K=nombre d'échantillons.**

3. Tests d’indépendance entre variables

Les tests d’indépendance [ou tests d’association] consistent à éprouver l’existence d’une liaison entre 2 variables. On distingue généralement trois catégories de tests d’indépendance selon que les variables sont qualitatives nominales, ordinales ou quantitatives.

Type de test	Tests paramétriques	Tests non paramétriques
Liaison entre **deux variables quantitatives**	Coefficient de corrélation de Pearson	- Rho de Spearman - Tau-a de Kendall
Liaison entre **deux variables ordinales**		- Gamma de Goodman - Kruskal - Tau-b et Tau-c de Kendall - d de Sommers - Test de Mantel - Haenszel [variables binaires]
Liaison entre **deux variables nominales**		- Test d'indépendance du χ^2 - t de Tschuprow et v de Cramer - Coefficient phi [variables binaires] - Coefficient Q de Yule [variables binaires] - Lambda de Goodman - Kruskal - Tau de Goodman - Kruskal - U de Theil
Liaison entre **deux** ou **plus de deux variables**		- Coefficient de concordance de Kendall [variables quantitatives ou ordinales] - Coefficient Kappa de Fleiss, concordance de p jugements [variables ordinales ; Kappa de Cohen pour p = 2]

Table de distribution du coefficient de corrélation de Bravais-Pearson

[Table de Pearson-Hartley]

La table donne la probabilité α pour que le coefficient de corrélation r(X, Y) égale ou dépasse, en valeur absolue, une valeur donnée r, c'est-à-dire soit observé en dehors de l'intervalle]-r, +r[, en fonction du nombre de degré de liberté ddl.

ddl \ α	0,10	0,05	0,02	0,01	0,001
1	0,9877	0,9969	0,9995	0,9999	0,9999
2	0,9000	0,9500	0,9800	0,9900	0,9990
3	0,8054	0,8783	0,9343	0,9587	0,9912
4	0,7293	0,8114	0,8822	0,9172	0,9741
5	0,6694	0,7545	0,8329	0,8745	0,9507
6	0,6215	0,7067	0,7887	0,8343	0,9249
7	0,5822	0,6664	0,7498	0,7977	0,8982
8	0,5494	0,6319	0,7155	0,7646	0,8721
9	0,5214	0,6021	0,6851	0,7348	0,8471
10	0,4973	0,5760	0,6581	0,7079	0,8233
11	0,4762	0,5529	0,6339	0,6835	0,8010
12	0,4575	0,5324	0,6120	0,6614	0,7800
13	0,4409	0,5139	0,5923	0,6411	0,7603
14	0,4259	9,4973	0,5742	0,6226	0,7420
15	0,4124	0,4821	0,5577	0,6055	0,7246
16	0,4000	0,4683	0,5425	0,5897	0,7084
17	0,3887	0,4555	0,5285	0,5751	0,6932
18	0,3783	0,4438	0,5155	0,5614	0,6787
19	0,3687	0,4329	0,5034	0,5487	0,6652
20	0,3598	0,4227	0,4921	0,5368	0,6524
25	0,3233	0,3809	0,4451	0,4869	0,5974
30	0,2960	0,3494	0,4093	0,4487	0,5541
35	0,2746	0,3246	0,3810	0,4182	0,5189
40	0,2573	0,3044	0,3578	0,3932	0,4896
45	0,2428	0,2875	0,3384	0,3721	0,4648
50	0,2306	0,2732	0,3218	0,3541	0,4433
60	0,2108	0,2500	0,2948	0,3248	0,4078
70	0,1954	0,2319	0,2737	0,3017	0,3799
80	0,1829	0,2172	0,2565	0,2830	0,3568
90	0,1726	0,2050	0,2422	0,2673	0,3375
100	0,1638	0,1946	0,2301	0,2540	0,3211

Table de distribution du coefficient de corrélation de Spearman

[Table de Pearson-Hartley]

La table donne les valeurs critiques du coefficient de corrélation de Spearman à DDL = n-2

n \ α	0,20	0,10	0,05	0,02	0,01	0,001	0,0001	0,00001
4	1,0000	1,0000	-	-	-	-	-	-
5	0,8005	0,8898	1,0000	1,0000	-	-	-	-
6	0,6573	0,8294	0,8864	0,9432	1,0000	-	-	-
7	0,5709	0,7145	0,7858	0,8934	0,9294	1,0000	-	-
8	0,5243	0,6432	0,7382	0,8326	0,8807	0,9762	-	-
9	0,4828	0,5997	0,7001	0,7831	0,8332	0,9328	-	-
10	0,4554	0,5643	0,6483	0,7455	0,7836	0,9035	-	-
11	0,4288	0,5381	0,6179	0,7092	0,7548	0,8729	-	-
12	0,4065	0,5028	0,5874	0,6709	0,7271	0,8599	-	-
13	0,3849	0,4845	0,5602	0,6485	0,7035	0,8352	-	-
14	0,3668	0,4637	0,5377	0,6224	0,6747	0,8112	-	-
15	0,3542	0,4432	0,5214	0,6038	0,6539	0,7856	1,0000	-
16	0,3415	0,4292	0,5035	0,5823	0,6346	0,7655	0,9973	-
17	0,3277	0,4136	0,4848	0,5665	0,6154	0,7477	0,9727	-
18	0,3175	0,4008	0,4722	0,5501	0,6002	0,7281	0,9436	-
19	0,3089	0,3913	0,4605	0,5354	0,5839	0,7120	0,9170	1,0000
20	0,2994	0,3802	0,4474	0,5204	0,5696	0,6958	0,8926	0,9992
21	0,2918	0,3667	0,4346	0,5076	0,5558	0,6807	0,8700	0,9878
22	0,2844	0,3615	0,4249	0,4958	0,5443	0,6668	0,8490	0,9639
23	0,2776	0,3531	0,4153	0,4863	0,5325	0,6543	0,8295	0,9418
24	0,2711	0,3438	0,4062	0,4757	0,5206	0,6425	0,8113	0,9211
25	0,2653	0,3368	0,3980	0,4659	0,5114	0,6304	0,7942	0,9017
26	0,2587	0,3307	0,3903	0,4573	0,5012	0,6187	0,7781	0,8835
27	0,2552	0,3238	0,3816	0,4481	0,4909	0,6084	0,7630	0,8863
28	0,2548	0,3174	0,3748	0,4402	0,4833	0,5976	0,7488	0,8501
29	0,254	0,3117	0,3677	0,4335	0,4748	0,5892	0,7353	0,8348
30	0,2398	0,3062	0,3624	0,4254	0,4669	0,5802	0,7225	0,8203
31	0,2366	0,3009	0,3562	0,4183	0,4587	0,5706	0,7103	0,8065
32	0,2318	0,2961	0,3501	0,4125	0,4518	0,5631	0,6988	0,7934
33	0,2296	0,2910	0,3447	0,4048	0,4461	0,5539	0,6878	0,7809
34	0,2248	0,2873	0,3403	0,3991	0,4388	0,5475	0,6773	0,7690
35	0,2224	0,2828	0,3348	0,3937	0,4335	0,5388	0,6672	0,7576
36	0,2187	0,2788	0,3305	0,3882	0,4267	0,5332	0,6576	0,7467
37	0,2165	0,2753	0,3249	0,3829	0,4214	0,5258	0,6484	0,7362
38	0,2124	0,2714	0,3214	0,3779	0,4153	0,5189	0,6396	0,7262
39	0,2089	0,2673	0,3172	0,3733	0,4097	0,5134	0,6312	0,7166
40	0,2068	0,2638	0,3135	0,3679	0,4052	0,5074	0,6230	0,7073
50	0,1845	0,2355	0,2787	0,3291	0,3628	0,4563	0,5558	0,6311
60	0,1677	0,2136	0,2554	0,2998	0,3315	0,4179	0,5065	0,5751
70	0,1553	0,1983	0,2352	0,2776	0,3074	0,3884	0,4684	0,5318
80	0,1442	0,1851	0,2205	0,2617	0,2898	0,3702	0,4377	0,4970
90	0,1358	0,1744	0,2078	0,2466	0,2730	0,3488	0,4124	0,4682
100	0,1288	0,1653	0,1970	0,2338	0,2589	0,3307	0,3910	0,4440
200	0,0908	0,1166	0,1389	0,1649	0,1826	0,2333	0,2758	0,3131
300	0,0741	0,0951	0,1133	0,1345	0,1490	0,1903	0,2250	0,2555
400	0,0642	0,0823	0,0981	0,1165	0,1290	0,1647	0,1948	0,2211
500	0,0574	0,0736	0,0877	0,1041	0,1153	0,1473	0,1742	0,1977

Table de l'écart-réduit [Table de la loi normale centrée-réduite]

La table donne la probabilité α pour que l'écart-réduit T égale ou dépasse, en valeur absolue, une valeur donnée t, c'est-à-dire la probabilité extérieure à l'intervalle [-t, +t].

$$1 - \alpha = P\{-t \leq T \leq +t\}$$

α	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	∞	2,576	2,326	2,17	2,054	1,96	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,44	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,2	1,175	1,15	1,126	1,103	1,08	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,86
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,69
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,51	0,496	0,482	0,468	0,454	0,44	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,24	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,1	0,088	0,075	0,063	0,05	0,038	0,025	0,013

- Pour chaque valeur de α, on lit t(a) tel qu'une variable aléatoire de loi N(0,1) a la probabilité (1 – α) de se trouver dans l'intervalle [-t(α), t(α)].
- La probabilité α s'obtient par addition des nombres inscrits en marge. Ainsi, pour t = 1,96 ; la probabilité est α = 0,00 + 0,05 = 0,05

Table de distribution de la loi T de Student [Test bilatéral]

(Valeurs de T ayant la probabilité P d'être dépassée en valeur absolue)

La table donne, en fonction du nombre de degrés de liberté ddl, la probabilité α pour que Tn égale ou dépasse, en valeur absolue, une valeur donnée tα.

$$1 - \alpha = \Pr\{-t\alpha \leq Tn \leq +t\alpha\}$$

Pr / ddl	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	5,959
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	2,263	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	0,961	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	0,137	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	3,649
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	3,656
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

Table de distribution de la loi F de Fisher-Snedecor

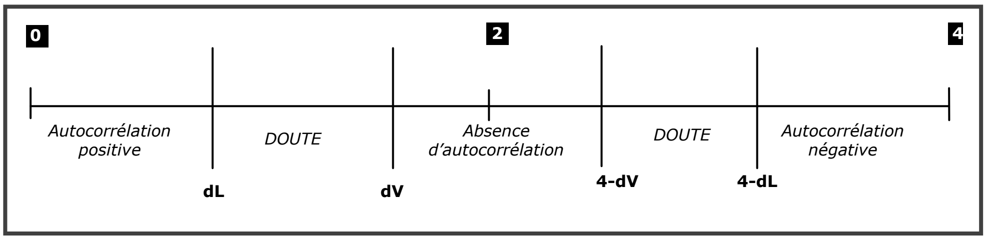
(Valeurs de F ayant la probabilité α d'être dépassées : $F = S_1^2/S_2^2$)

Nu2	Nu1 = 1		Nu1 = 2		Nu1 = 3		Nu1 = 4		Nu1 = 5	
	α = 0,05	α = 0,01	α = 0,05	α = 0,01	α = 0,05	α = 0,01	α = 0,05	α = 0,01	α = 0,05	α = 0,01
1	161,4	4052,00	199,5	4999,00	213,7	3403,00	224,6	5625,00	230,2	5764,00
2	18,51	98,49	19,00	99,00	19,16	99,17	19,25	99,25	19,30	99,30
3	10,13	34,12	9,55	30,81	9,28	29,46	9,12	28,71	9,01	28,24
4	7,71	21,20	6,94	18,00	6,59	16,69	6,39	13,98	6,26	13,32
3	6,61	16,26	5,79	13,27	5,41	12,06	5,19	11,39	5,03	10,97
6	3,99	13,74	3,14	10,91	4,76	9,78	4,53	9,13	4,39	8,75
7	3,39	12,23	4,74	9,35	4,33	8,43	4,12	7,85	3,97	7,45
8	3,32	11,26	4,46	8,63	4,07	7,39	3,84	7,01	3,69	6,63
9	5,12	10,56	4,26	8,02	3,86	6,99	3,63	6,42	3,48	6,06
10	4,96	10,04	4,10	7,56	3,71	6,33	3,48	5,99	3,33	5,64
11	4,84	9,65	3,98	7,20	3,59	6,22	3,36	5,67	3,20	5,32
12	4,75	9,33	3,88	6,93	3,49	5,93	3,26	5,41	3,11	5,06
13	4,67	9,07	3,80	6,70	3,41	5,74	3,18	5,20	3,02	4,86
14	4,60	8,86	3,74	6,31	3,34	5,56	3,11	5,03	2,96	4,69
15	4,34	8,68	3,68	6,36	3,29	5,42	3,06	4,89	2,90	4,56
16	4,49	8,53	3,63	6,23	3,24	5,29	3,01	4,77	2,85	4,44
17	4,45	8,40	3,59	6,11	3,20	5,18	2,96	4,67	2,81	4,34
18	4,41	8,28	3,53	6,01	3,16	5,09	2,93	4,58	2,77	4,25
19	4,38	8,18	3,52	5,93	3,13	5,01	2,90	4,50	2,74	4,17
20	4,35	8,10	3,49	5,85	3,10	4,94	2,87	4,43	2,71	4,10
21	4,32	8,02	3,47	5,78	3,07	4,87	2,84	4,37	2,68	4,04
22	4,30	7,94	3,44	5,72	3,05	4,82	2,82	4,31	2,66	3,99
23	4,28	7,88	3,42	5,66	3,03	4,76	2,80	4,26	2,64	3,94
24	4,26	7,82	3,40	5,61	3,01	4,72	2,78	4,22	2,62	3,90
25	4,24	7,77	3,38	5,37	2,99	4,68	2,76	4,18	2,60	3,86
26	4,22	7,72	3,37	5,33	2,98	4,64	2,74	4,14	2,39	3,82
27	4,21	7,68	3,33	5,49	2,96	4,60	2,73	4,11	2,37	3,78
28	4,20	7,64	3,34	5,43	2,95	4,57	2,71	4,07	2,56	3,75
29	4,18	7,60	3,33	5,42	2,93	4,34	2,70	4,04	2,34	3,73
30	4,17	7,56	3,32	5,39	2,92	4,31	2,69	4,02	2,53	3,70
40	4,08	7,31	3,23	5,18	2,84	4,31	2,61	3,83	2,43	3,31
60	4,00	7,08	3,15	4,98	2,76	4,13	2,32	3,65	2,37	3,34
120	3,92	6,85	3,07	4,79	2,68	3,93	2,43	3,48	2,29	3,17
∞	3,84	6,64	2,99	4,60	2,60	3,78	2,37	3,32	2,21	3,02

- S_1^2 est la plus grande des deux variances estimées, avec n degrés de liberté au numérateur.

Table de Durbin-Watson

La table donne les limites inférieures et supérieures des seuils de signification du test de Durbin et Watson pour α = 5 %.

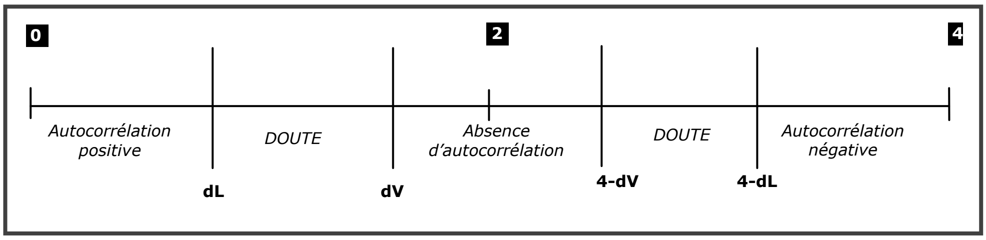


	k = 1		k = 2		k = 3		k = 4		k = 5	
n	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,91	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

- k : nombre de variables explicatives, constante exclue ; n : nombre d'observations et α = 0.05.

Table de Durbin-Watson

La table donne les limites inférieures et supérieures des seuils de signification du test de Durbin et Watson pour α = 1 %.



	k = 1		k = 2		k = 3		k = 4		k = 5	
n	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**	**dL**	**dV**
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,42	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

- k : nombre de variables explicatives, constante exclue ; n : nombre d'observations et α = 0.01.