
INSTITUT DU CONSTAT FONDAMENTAL

icf.news · ISSN 2025-0001 (fictif)

Fondements Statistiques des Constats Fondamentaux

Document pédagogique à destination du grand public

— Document de vulgarisation —

Dr. Kwame Osei-Mensah

Statisticien — Institut du Constat Fondamental

Actuellement à Accra · Correspondance : en cours d'acheminement

Produit à la demande du Pr. Parmentier, transmise par message vocal un mardi à 7h12 sans cahier des charges, sans délai et sans budget supplémentaire. Le niveau de complexité a été calibré selon le jugement professionnel de l'auteur, deux demandes de clarification étant restées sans réponse.

Juillet 2026 · Document interne ICF · Non soumis à peer review (par nos soins)

Note liminaire de l'auteur

Le Professeur Parmentier m'a demandé de produire un document permettant au grand public de comprendre les outils statistiques mobilisés dans les études de l'ICF.

J'ai produit ce document.

Le lecteur qui estimerait que le contenu dépasse le niveau habituel d'un document de vulgarisation destiné au grand public est invité à considérer deux hypothèses concurrentes : soit le document est trop complexe, soit le grand public est sous-estimé. L'auteur penche pour la seconde.

Les démonstrations qui suivent sont mathématiquement exactes. Leur accessibilité est laissée à l'appréciation du lecteur.

— *K. Osei-Mensah, Accra, juillet 2026*

Table des matières

1. Le coefficient de corrélation de Pearson (r) — Définition, propriétés et interprétation dans l'espace vectoriel $\mathbb{R}^n$
2. Les seuils de significativité statistique — p-value, test t de Student et intervalles de confiance bootstrappés
3. La droite de régression linéaire — Méthode des moindres carrés ordinaires et ses implications pour les constats fondamentaux
4. Le d de Cohen — Mesure de la taille d'effet et puissance statistique en contexte d'échantillonnage autoroutier
5. L'Indice de Sélectivité Bibliographique (ISB) — Contribution méthodologique originale de l'ICF
6. Tableau récapitulatif des indices ICF et leur interprétation formelle
7. Conclusion et recommandations pour la lecture des études ICF

1. Le coefficient de corrélation de Pearson (r)

1.1 Définition formelle

Soit deux variables aléatoires X et Y définies sur un espace de probabilité $(\Omega, \mathcal{F}, P)$. Le coefficient de corrélation linéaire de Pearson est défini comme le rapport de leur covariance au produit de leurs écarts-types respectifs :

$$r(X, Y) = \text{Cov}(X, Y) / (\sigma_X \cdot \sigma_Y) = E[(X - \mu_X)(Y - \mu_Y)] / \sqrt{E[(X - \mu_X)^2] \cdot E[(Y - \mu_Y)^2]}$$

Équation 1 — Définition du coefficient r de Pearson dans l'espace $L^2(\Omega, P)$

Sur un échantillon fini de n observations $\{(x_i, y_i)\}_{i=1}^n$, l'estimateur empirique s'écrit :

$$r = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{[\sum_i (x_i - \bar{x})^2] \cdot [\sum_i (y_i - \bar{y})^2]}$$

Équation 2 — Estimateur empirique de r (forme de Bravais-Pearson, 1896)

1.2 Propriétés algébriques fondamentales

Le coefficient r appartient à l'intervalle fermé $[-1, +1]$, ce qui découle directement de l'inégalité de Cauchy-Schwarz dans l'espace de Hilbert $L^2(\Omega, P)$:

$$|\langle X - \mu_X, Y - \mu_Y \rangle|^2 \leq \|X - \mu_X\|^2 \cdot \|Y - \mu_Y\|^2$$

Équation 3 — Inégalité de Cauchy-Schwarz, fondement de la borne de r

$r = 1$ indique une relation linéaire positive parfaite : $Y = aX + b$ avec $a > 0$. $r = -1$ indique une relation linéaire négative parfaite : $a < 0$. $r = 0$ indique l'absence de relation linéaire, sans préjuger de l'existence de relations non linéaires (quadratiques, exponentielles, logarithmiques).

Note de Kwame : $r = 0$ ne signifie pas que les variables sont indépendantes. Deux variables peuvent être fortement dépendantes et présenter $r = 0$. Si vous rencontrez quelqu'un affirmant le contraire lors d'un cocktail académique, l'ICF recommande de changer de conversation.

1.3 Interprétation conventionnelle des niveaux de r

$ r $	Interprétation conventionnelle	Occurrence dans les études ICF
0,00 – 0,10	Corrélation négligeable	Études sur la climatisation ($r = 0,04$)
0,10 – 0,30	Corrélation faible	Non observé à ce jour
0,30 – 0,50	Corrélation modérée	Non observé à ce jour
0,50 – 0,70	Corrélation forte	Non observé à ce jour
0,70 – 0,90	Corrélation très forte	Tapis rouge ($r = 0,79$), SDAF ($r = 0,88$)
0,90 – 1,00	Corrélation quasi-parfaite	Majorité des études ICF ($r = 0,91$ à $0,97$)

Tableau 1 — Niveaux conventionnels d'interprétation du r de Pearson (Cohen, 1988) et leur occurrence dans le corpus ICF. La colonne de droite suggère que l'ICF opère principalement dans des zones que les statisticiens qualifient de « surprenantes pour des sciences sociales ».

Avertissement méthodologique : Les études ICF produisant systématiquement des $r > 0,85$ sur des évidences préexistantes confirment le postulat fondateur de l'Institut — à savoir que les évidences sont, en effet, évidentes. Ce résultat ne constitue pas un biais de confirmation. Il constitue une validation empirique de l'intuition commune.

2. Les seuils de significativité statistique

2.1 La p-value — définition rigoureuse

Soit H_0 l'hypothèse nulle (absence d'effet, $r = 0$) et H_1 l'hypothèse alternative ($r \neq 0$). La p-value est la probabilité d'observer une statistique de test T au moins aussi extrême que la valeur observée t_{obs} , sous l'hypothèse que H_0 est vraie :

$$p = P(|T| \geq |t_{\text{obs}}| \mid H_0) = 2 \cdot P(T \geq |t_{\text{obs}}| \mid H_0)$$

Équation 4 — Définition de la p-value bilatérale sous H_0

Pour tester la significativité du coefficient r sur n observations, la statistique de test suit une distribution de Student à $(n-2)$ degrés de liberté :

$$t = r \cdot \sqrt{(n-2)} / \sqrt{(1-r^2)} \sim t(n-2) \text{ sous } H_0$$

Équation 5 — Statistique de test pour $H_0 : \rho = 0$

2.2 Les seuils conventionnels et leur interprétation

La communauté scientifique a adopté par convention les seuils suivants :

Seuil α	Niveau de confiance	Interprétation	Notation
$p < 0,05$	95 %	Significatif	*
$p < 0,01$	99 %	Très significatif	**
$p < 0,001$	99,9 %	Hautement significatif	***
$p < 0,0001$	99,99 %	Ce que produit systématiquement l'ICF	****

Tableau 2 — Seuils de significativité conventionnels. La dernière ligne a été ajoutée par l'auteur pour refléter la réalité du corpus ICF.

Note de Kwame : La p-value ne mesure pas la probabilité que H_0 soit vraie. Elle mesure la probabilité d'obtenir des données aussi extrêmes si H_0 était vraie. Cette distinction, ignorée par environ 78 % des lecteurs de résultats statistiques (estimation personnelle non publiée), constitue l'une des sources de confusion les plus productives de la littérature scientifique contemporaine.

2.3 Les intervalles de confiance

L'intervalle de confiance à 95 % pour r est obtenu via la transformation z de Fisher, qui stabilise la variance de l'estimateur :

$$z = (1/2) \cdot \ln[(1+r)/(1-r)] = \text{arctanh}(r)$$

Équation 6 — Transformation z de Fisher (arctanh)

L'intervalle de confiance sur z est $[z - 1,96/\sqrt{(n-3)} ; z + 1,96/\sqrt{(n-3)}]$, puis retransformé en r via la fonction inverse $\tanh$.

Application aux études ICF : avec $n = 1\,102$ (étude SDAF), $r = 0,88$, $z = \text{arctanh}(0,88) = 1,376$, $\text{IC}_{95\%} = [1,376 \pm 1,96/\sqrt{1099}] = [1,376 \pm 0,059] = [1,317 ; 1,435]$, soit en r : $[0,866 ; 0,892]$. La corrélation est robuste.

3. La droite de régression linéaire

3.1 Le modèle linéaire simple

Le modèle de régression linéaire simple postule que Y est une fonction affine de X à laquelle s'ajoute un terme d'erreur ε de distribution normale centré :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n$$

Équation 7 — Modèle de régression linéaire simple (Gauss-Markov)

3.2 Estimation par les moindres carrés ordinaires (MCO)

Les estimateurs MCO minimisent la somme des carrés des résidus :

$$\min_{\{\beta_0, \beta_1\}} \sum_i (Y_i - \beta_0 - \beta_1 X_i)^2$$

Équation 8 — Critère de minimisation MCO

La solution analytique donne :

$$\beta_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} = \text{Cov}(X, Y) / \text{Var}(X)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

Équations 9-10 — Estimateurs MCO de β_0 et β_1

3.3 Le coefficient de détermination R^2

Le R^2 mesure la proportion de variance de Y expliquée par le modèle linéaire en X :

$$R^2 = 1 - \text{SS}_{\text{res}} / \text{SS}_{\text{tot}} = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2}$$

Équation 11 — Coefficient de détermination R^2

Dans le cas de la régression simple, $R^2 = r^2$. Ainsi, pour l'étude sur la bande centrale ($r = 0,93$), $R^2 = 0,865$: 86,5 % de la variance du narcissisme positionnel est expliquée par le comportement SVC. Les 13,5 % restants correspondent probablement à des facteurs que le drone n'a pas pu capturer.

Note de Kwame : Les hypothèses du modèle MCO (normalité des résidus, homoscedasticité, absence d'autocorrélation, exogénéité des régresseurs) n'ont pas toutes pu être vérifiées dans les conditions d'interception autoroutière. L'auteur estime que cela n'invalide pas les résultats mais reconnaît que le Comité de lecture aurait pu soulever ce point s'il n'avait pas été bloqué dans les embouteillages.

4. Le d de Cohen — Mesure de la taille d'effet

4.1 Définition et motivation

La p-value indique si un effet est statistiquement détectable. Elle n'indique pas si cet effet est substantiellement important. Avec n suffisamment grand, n'importe quel effet infinitésimal devient statistiquement significatif. Le d de Cohen mesure la magnitude de l'effet, indépendamment de la taille de l'échantillon :

$$d = (\mu_1 - \mu_2) / \sigma_{\text{pooled}}$$
$$\sigma_{\text{pooled}} = \sqrt{(\sigma_1^2 + \sigma_2^2) / 2} \quad (\text{estimateur poolé sous hypothèse d'homoscédasticité})$$

Équations 12-13 — d de Cohen et écart-type poolé

4.2 Seuils d'interprétation conventionnels

d de Cohen	Taille d'effet	Équivalent en percentile	Exemple ICF
0,20	Petit	58e percentile	Non observé
0,50	Moyen	69e percentile	Non observé
0,80	Grand	79e percentile	Non observé
2,00	Très grand	98e percentile	Habituel dans le corpus
4,87	Exceptionnel	>99,9e percentile	Étude ICF n°6 : l'eau mouille

Tableau 3 — Seuils d'interprétation du d de Cohen (1988) et leur occurrence dans le corpus ICF. L'étude n°6 (l'eau mouille, $d = 4,87$) constitue selon l'auteur « un résultat dont la robustesse n'est plus à démontrer depuis l'Antiquité ».

Note de Kwame : Jacob Cohen lui-même, dans sa Statistical Power Analysis (1988), précise que ces seuils sont arbitraires et contextuels. Il est probable qu'il n'avait pas envisagé le contexte de l'ICF lors de la rédaction de cet ouvrage.

5. L'Indice de Sélectivité Bibliographique (ISB)

L'ISB est une contribution méthodologique originale de l'ICF, introduite dans l'étude n° 17 (Indétermination catégorielle dans les épreuves sportives). Il mesure le degré de sélectivité avec lequel un chercheur ou un commentateur choisit les sources qu'il cite à l'appui de sa position préexistante.

5.1 Définition formelle

Soit C le corpus total des études disponibles sur un sujet donné, et C_{cit} le sous-ensemble des études effectivement citées par l'auteur. L'ISB est défini comme :

$$\text{ISB} = 1 - |C_{\text{cit}} \cap C_{\text{contra}}| / |C_{\text{cit}}|$$

où C_{contra} désigne le sous-ensemble des études contredisant la position de l'auteur.

Un ISB de 0 indique un chercheur citant autant d'études contradictoires que d'études confirmatoires. Un ISB de 1 indique un chercheur n'ayant cité aucune étude contredisant sa position préexistante.

Valeurs observées : FCO (Fixation Chromosomique Obsessionnelle) : ISB = 0,94 ; DGT (Déconstruction Gazeuse Totale) : ISB = 0,91. Différence non significative ($p = 0,43$), ce que les auteurs jugent « remarquablement équitable ».

6. Tableau récapitulatif des indices ICF

Indice	Signification	Échelle	Valeur typique ICF
r de Pearson	Corrélation linéaire	$[-1 ; +1]$	0,79 à 0,97
p-value	Significativité statistique	$[0 ; 1]$	$< 0,0001$
R ²	Variance expliquée	$[0 ; 1]$	0,62 à 0,94
d de Cohen	Taille d'effet	$[0 ; +\infty[$	2,00 à 4,87
ISB	Sélectivité bibliographique	$[0 ; 1]$	0,91 à 0,94
IAA	Anosognosie arbitrale	$[0 ; 1]$	0,94 (sport)
LDO	Labilité doctrinale	$[0 ; 1]$	0,91 (politique)
IIC	Invisibilité civile	$[0 ; 1]$	0,91 (admin.)
CPE	Préservation épidermique	$[0 ; 1]$	$r = 0,91$ vs revenu
ISA	Solipsisme autoroutier	$[0 ; 1]$	0,88 (E42/E19)
CTR	Tapis rouge diplomatique	$[0 ; 1]$	0,44 à 0,73

Tableau 4 — Récapitulatif des principaux indices statistiques et composites mobilisés dans le corpus ICF (20 études, juillet 2026). La colonne « Valeur typique ICF » reflète la distribution observée et non une norme établie. L'auteur note que ces valeurs seraient inhabituellement élevées pour la plupart des revues académiques. Pour l'ICF, elles constituent la norme, ce qui est cohérent avec le caractère fondamental des constats.

7. Conclusion et recommandations

Le présent document a fourni au lecteur les outils nécessaires à une lecture éclairée des études de l'ICF. Ces outils sont mathématiquement rigoureux, universellement reconnus dans la littérature scientifique, et, selon l'auteur, parfaitement accessibles au grand public tel qu'il le conçoit.

Le lecteur qui, à l'issue de ce document, se sent moins confiant qu'avant dans sa compréhension des statistiques est invité à considérer que c'est peut-être là le signe d'une compréhension plus profonde. La confusion informée est supérieure à la certitude naïve.

Le lecteur qui, au contraire, a trouvé ce document trop simple est prié de contacter l'auteur directement. L'auteur est actuellement à Accra. Sa réponse est en cours d'acheminement.

Recommandations pratiques pour la lecture des études ICF :

1. Un $r > 0,85$ dans une étude ICF confirme une évidence. C'est normal. C'est l'objet.
2. Une $p\text{-value} < 0,0001$ signifie que le résultat ne doit rien au hasard. Les évidences ne doivent effectivement rien au hasard.
3. Un d de Cohen élevé signifie que l'effet est substantiel. Les évidences fondamentales sont, par définition, substantielles.
4. L'ISB élevé observé dans les positions rhétoriques adverses confirme que la sélectivité bibliographique est une constante humaine, indépendamment du camp.
5. La droite de régression ne dit pas pourquoi deux variables sont liées. Elle dit qu'elles le sont. Le « pourquoi » est laissé à la sagacité du lecteur.

« Les statistiques ne mentent pas. Les chercheurs qui les sélectionnent, parfois. L'ICF, pour sa part, sélectionne des évidences. Les évidences n'ont pas besoin de mentir. »

— K. Osei-Mensah, Accra, juillet 2026
