**Zeitschrift:** Acta Tropica
**Herausgeber:** Schweizerisches Tropeninstitut (Basel)
**Band:** 37 (1980)
**Heft:** (11): Santé de l'enfant d'âge scolaire en Côte d'Ivoire

**Artikel:** Méthodologie statistique
**Autor:** Vuataz, L.
**DOI:** https://doi.org/10.5169/seals-312687

**Nutzungsbedingungen**
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

**Conditions d'utilisation**
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

**Terms of use**
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

**Download PDF:** 12.07.2025

**ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch**

# Méthodologie statistique

L. VUATAZ

Nous ne donnons ici qu'un aperçu de la méthodologie statistique, dépourvu de formules mathématiques. Les lecteurs intéressés pourront se procurer un texte détaillé auprès de l'auteur.

Le plan expérimental de cette étude est un tableau 3 × 4 provenant des trois passages dans les quatre villages, donnant lieu à 12 situations. Chaque situation comporte un grand nombre d'observations multivariées: les enfants sur lesquels de nombreuses variables ont été mesurées. Pour l'étude de certaines variables, un groupe d'enfants européens d'Abidjan représente une 13e situation.

Les analyses de variance et comparaisons entre villages lors d'un même passage sont des analyses traditionnelles, univariées ou multivariées selon que l'on considère une seule variable, ou plusieurs variables simultanément. La comparaison univariée transversale de deux situations se réduit au test-t de Student. On applique la version de Welch (1947) lorsque les variances sont hétérogènes.

Pour les analyses longitudinales univariées, on applique la technique des mesures répétées (Winer, 1971). Dans ces conditions, la comparaison de deux situations seulement se réduit au test-t pour mesures appariées.

Dans cette expérimentation, les corrélations entre variables présentent un intérêt particulier. L'étude de ces corrélations pose un problème, en ce sens qu'il faut distinguer:

1. les corrélations intra-situations, calculées à partir des observations d'une même situation et cumulées sur les situations,
2. les corrélations inter-situations, calculées à partir des moyennes des situations, pondérées par les nombres d'observations,
3. les corrélations globales, calculées sur l'ensemble des observations.

On ne tient pas compte de ce que les observations sur les mêmes enfants lors de passages différents ne sont pas forcément indépendantes, et on néglige le fait que les douze situations forment une classification croisée.

Les corrélations intra-situations font l'objet de deux tests:

- un test d'homogénéité (Biometrical Tables, 1970),
- un test de signification si le premier test ne rejette pas l'hypothèse d'homogénéité.

Une corrélation intra-situations est une corrélation naturelle indépendante des situations, due à des causes souvent endogènes (par rapport aux enfants), et dans tous les cas à des causes intrinsèques aux situations.

Les corrélations inter-situations font l'objet d'un test de signification: c'est le test-t classique du coefficient de corrélation avec g-2 degrés de liberté, où g est ici le nombre de situations. *En principe*, une corrélation inter-situations significative peut être indicatrice d'effets corrélés propres aux situations, dus à l'environnement au sens large du mot (donc incluant l'alimentation), et souvent d'origine exogène (par rapport aux enfants). «En principe» disons-nous, car en l'absence d'effets corrélés propres aux situations, les corrélations inter-situations ont une espérance mathématique égale à celle des corrélations intra-situations (Kendall, 1951). Il s'agit donc de tester l'hypothèse de l'égalité possible d'une corrélation inter-situations et de la corrélation intra-situations correspondante. Cela pourra se faire par la transformation normalisante de Fisher (Rao, 1952), ou par l'examen des intervalles de confiance attachés à ces corrélations (Biometrical Tables, 1970), lorsque les conditions ne sont pas requises pour le premier test. On prendra, pour déterminer les intervalles de confiance à partir des courbes de David (Biometrical Tables, 1970), des nombres d'observations

égaux à g pour une corrélation inter-situations, et à N-g + 1 pour une corrélation intra-situations, où N est le nombre total d'observations. S'il existe une relation de dépendance à sens unique entre deux variables, on pourra recourir, pour l'ensemble des tests indiqués ci-dessus, à l'analyse généralisée de covariance (Bennett et Franklin, 1954) qui teste des coefficients de régression.

Le coefficient de corrélation globale, quant à lui, fait l'objet d'un test de signification classique avec N-2 degrés de liberté. Ce test n'a de sens qu'en l'absence de variation significative entre situations. D'une façon générale, ce coefficient global peut conduire à des conclusions erronées (voir un exemple ci-dessous), car il est une résultante des deux autres corrélations.

Dans les tables de corrélations, l'information est présentée de la façon suivante, pour chaque couple de variable:

Variable X

| | | | |
|---|---|---|---|
| | $r_a$ | test de $H_1$ | test de $H_2$ |
| Variable y | $r_m$ | test de $H_4$ | g |
| | $r_o$ | | N |

$r_a$ = coefficient de corrélation intra-situations
$r_m$ = coefficient de corrélation inter-situations
$r_o$ = coefficient de corrélation globale

g = nombre de situations
N = nombre total d'observations
$H_1$ = hypothèse d'homogénéité des corrélations intra-situations. NS (non significatif) veut dire que l'hypothèse $H_1$ n'est pas rejetée au niveau de 10%
$H_2$ = hypothèse stipulant que le coefficient de corrélation intra-situations est nul dans la population échantillonnée. NS veut dire que cette hypothèse n'est pas rejetée au niveau de 10%
$H_4$ = hypothèse stipulant que le coefficient de corrélation inter-situations est nul dans la population échantillonnée. NS veut dire que cette hypothèse n'est pas rejetée au niveau de 10%

+ = significatif à 10%
* = significatif à 5%
** = significatif à 1%
*** = significatif à 0,1%

L'hypothèse d'égalité entre les coefficients réels mais inconnus dont $r_a$ et $r_m$ sont les estimateurs peut être éprouvée approximativement selon les indications données dans le texte.

*Exemple 1*

Fer sérique

| | | | |
|---|---|---|---|
| | 0,15 | NS | * |
| Vitamine C | −0,62 | * | 12 |
| | −0,04 | | 219 |

N = 219, g = 12
l'hypothèse $H_1$ n'est pas rejetée
$r_a$ = 0,15 et l'hypothèse $H_2$ est rejetée au niveau de 5%
$r_m$ = −0,62 et l'hypothèse $H_4$ est rejetée au niveau de 5%

L'intervalle de confiance à 95% associé à $r_a$ va de 0,02 à 0,27. L'intervalle de confiance à 95% associé à $r_m$ va de −0,87 à −0,07. Il n'y a pas de chevauchement entre ces intervalles de confiance, on en déduit qu'il y a des effets négativement corrélés significatifs propres aux situations.

Notons que la corrélation globale seule laisserait supposer l'absence de corrélation entre Fer sérique et vitamine C. La conclusion correcte est la suivante: dans une même situation, il y a une tendance faible mais significative d'association positive entre ces deux variables: un enfant ayant un taux sérique en fer supérieur à la moyenne (de la situation) aura tendance à présenter un taux en vitamine C également supérieur à la moyenne. Par contre, on observe une corrélation négative propre aux situations. On pourrait l'expliquer en supposant que, lors de la cuisson des aliments, la vitamine C disparaît d'autant plus que ces aliments contiennent davantage de fer, ce dernier provenant du sol ou des ustensiles.

*Exemple 2*

Hématocrite

| | | | |
|---|---|---|---|
| | 0,25 | NS | *** |
| Vitamine A | 0,62 | * | 12 |
| | 0,32 | | 220 |

L'intervalle de confiance à 95% associé à $r_a = 0{,}25$ va de 0,13 à 0,36.
L'intervalle de confiance à 95% associé à $r_m = 0{,}62$ va de 0,07 à 0,87

Le dernier intervalle contient entièrement le premier. La corrélation positive significative inter-situations n'est que le reflet au niveau des moyennes des situations, de la corrélation intra-situations.


Bennett C.A., Franklin N.L.: Statistical analysis in chemistry and the chemical industry, p.441. J. Wiley & Sons, New York 1954.

Biometrical Tables for Statisticians, Vol. I (ed. by E.S. Pearson and H.O. Hartley), 3rd ed. University Press, Cambridge 1970.

Kendall M.G.: The advanced theory of statistics, Vol. I, p.340. Griffin & Co. Ltd, High Wycombe 1951.

Rao C.R.: Advanced statistical methods in biometric research, p.230. J. Wiley & Sons, New York 1952.

Welch B.L.: The generalization of Student's problem when several different population variances are involved. Biometrika *34*, 28 (1947).

Winer B.J.: Statistical principles in experimental design. McGraw-Hill Book Co., New York 1971.