



La différence entre régression et corrélation

Tom Poelman, Vakgroep Huisartsgeneeskunde en Eerstelijnsgezondheidszorg, UGent

Pour exprimer la relation ou l'association entre deux ou plusieurs variables, on a recours à la régression et à la corrélation, deux notions souvent confondues (1).

La régression est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres. Lorsqu'un **diagramme de dispersion** suggère que les variations d'une variable sont proportionnelles à celles de l'autre variable, l'analyse de régression linéaire permet de déterminer la droite approchant au mieux cette variation proportionnelle. Sur un graphique, cela revient à chercher la droite de laquelle tous les points sont aussi peu éloignés que possible de cette droite (voir figure 1). La formule mathématique de cette droite de régression est la suivante (2) :

valeur de la variable dépendante = intercept + coefficient de régression β x valeur de la variable indépendante

Le **coefficient de régression** indique dans quelle mesure la valeur d'une variable dépendante varie avec la variation de la valeur de la variable indépendante (aussi appelée variable explicative). L'intercept (= ordonnée à l'origine) est la valeur obtenue lorsque la variable indépendante vaut zéro (voir figure 1).

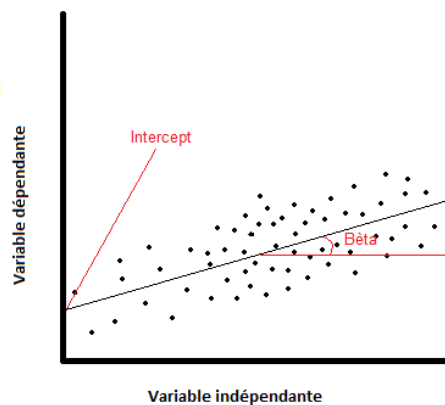


Figure 1 : Diagramme de dispersion avec droite de régression.

Exemple :

L'étude de Little (3,4) cherchait à savoir dans quelle mesure la satisfaction du patient (score MISS) (= variable dépendante) augmentait ou diminuait en fonction du nombre de gestes du médecin (= variable indépendante ou explicative). L'analyse de régression linéaire a montré une association

positive entre les deux variables avec un coefficient de régression (à une variable) de 0,11 et un intervalle de confiance à 95% de 0,02 à 0,19 ($p = 0,018$). Pour chaque geste supplémentaire du médecin, le score MISS augmente donc en moyenne de 0,11 point sur une échelle de 1 à 7. Du fait de l'intervalle de confiance large, cette augmentation peut en réalité n'être que de 0,02 point comme de 0,19 point.

Il arrive souvent que l'on cherche la relation entre une variable dépendante et plusieurs variables indépendantes. L'analyse de régression multiple permet de calculer, pour chaque variable indépendante, un coefficient de régression partielle (5). La formule mathématique de la régression multiple se présente alors comme suit :

$$\text{valeur de la variable dépendante} = \text{intercept} + \text{coefficient de régression } \beta_1 \times \text{valeur de la variable indépendante}_1 + \text{coefficient de régression } \beta_2 \times \text{valeur de la variable indépendante}_2 + \dots + \text{coefficient de régression } \beta_n \times \text{valeur de la variable indépendante}_n$$

Ainsi, dans notre exemple (3,4), le coefficient de régression (à variables multiples) était de 0,08 (avec IC à 95% de 0,01 à 0,15 et $p = 0,046$). Ceci montre que l'on prédira un peu moins bien la satisfaction du patient à partir du nombre de gestes du médecin si d'autres variables indépendantes sont également prises en compte.

Une analyse de régression univariée donne une idée du mode d'association entre deux variables, mais elle ne permet pas de déterminer le degré de cette association. L'intensité de la liaison s'exprime au moyen du **coefficient de corrélation (Pearson, r)** (6), un chiffre se situant toujours entre -1 (relation linéaire parfaite avec pente négative) et +1 (relation linéaire parfaite avec pente positive), et n'établit pas de distinction entre les variables dépendantes et indépendantes. Ce chiffre dépendra donc de la dispersion des différentes données autour de la droite de régression. Dans les figures 2 et 3, on voit deux diagrammes de dispersion différents avec une même droite de régression. A cause d'une faible corrélation entre les variables dans la figure 2, l'angle d'inclinaison (coefficient de régression β) aura un intervalle de confiance large.

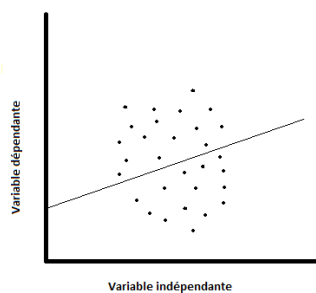


Figure 2 : Faible corrélation

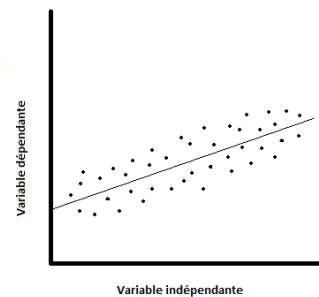


Figure 3 : Forte corrélation

Exemple :

Dans l'étude de Little (3,4), la taille de l'échantillon a été déterminée à partir d'une corrélation de 0,25 entre, d'une part, les aspects de la communication verbale et non verbale de la consultation et, d'autre part, la perception qu'avait le patient de la communication et de la relation médecin-patient. Pour ce faire, les auteurs se sont appuyés sur une étude similaire (7) qui avait trouvé un coefficient de

corrélation de Pearson de $r = 0,28$ avec $p = 0,002$ entre le score MISS et le sentiment d'être au centre de l'acte médical. Pour une meilleure interprétation de ce chiffre, r est mis au carré (r^2) et exprimé sous forme de pourcentage : $(0,28)^2 \times 100\% = 7,8\%$. Ceci permet de dire que seulement 7,8% de la variation du score MISS peut s'expliquer par la focalisation du médecin sur le patient. Il existe sans doute encore beaucoup d'autres facteurs expliquant la variation du score MISS. Malheureusement, les auteurs de l'étude de Little ne donnent pas de coefficient de corrélation, ce qui fait que nous ne pouvons pas estimer correctement les différentes relations (6).

Conclusion

La corrélation mesure l'intensité de la liaison entre des variables, tandis que la régression analyse la relation d'une variable par rapport à une ou plusieurs autres.

Références

1. Sedgwick P. Correlation versus linear regression. *BMJ* 2013; 346:f2686.
2. Sedgwick P. Simple linear regression. *BMJ* 2013; 346:f2340.
3. Little P, White P, Kelly J, et al. Verbal and non-verbal behaviour and patient perception of communication in primary care: an observational study. *Br J Gen Pract* 2015;65:e357-65.
4. Van Nuland M. Importance de la communication verbale et non verbale pendant la consultation. *MinervaF* 2016;15(2):35-8.
5. Sedgwick P. Multiple regression. *BMJ* 2013; 347: f4373.
6. Sedgwick P. Correlation. *BMJ* 2012; 345: e5407.
7. Kinnersley P, Stott N, Peters TJ, Harvey I. The patient-centredness of consultations and outcome in primary care. *Br J Gen Pract* 1999;49:711-6.