

Transcription de l'épisode 072 – Les résultats (02) – DEFI6M

Analyse de données quantitatives : analyser la normalité de la distribution

Aujourd'hui, nous allons commencer à aborder l'analyse de données quantitatives, qui utilise les statistiques.

Je vois beaucoup trop d'étudiants et d'étudiantes qui se lancent dans l'analyse de leurs données sans vérifier les conditions préalables des tests statistiques, notamment des tests inférentiels qui vont tester leurs hypothèses.

En effet, si les données ne suivent pas la loi normale, qui correspond à une courbe de Gauss, vous ne devriez pas utiliser des tests que l'on qualifie de tests paramétriques. En ce sens, il faut donc utiliser des tests non paramétriques équivalents.

Un test non paramétrique est un test basé sur l'étude des rangs des observations qui ne fait pas d'hypothèses particulières sur la forme de la distribution d'origine.

Toutefois, il est vrai qu'un certain nombre de tests sont dits robustes à la non-normalité des données. Nous aurons probablement l'occasion d'y revenir.

Pour tester préalablement la normalité de la distribution et donc décider du test à effectuer en conséquence, il existe plusieurs méthodes : la plus simple est d'examiner les coefficients d'asymétrie (Skewness) et les coefficients d'aplatissements (Kurtosis) dans un logiciel de statistiques. Des valeurs d'asymétrie et de Kurtosis comprises entre -1,96 et +1,96 sont considérées comme acceptables afin de prouver une distribution normale univariée.

La méthode la plus répandue pour tester la normalité de la distribution réside dans l'utilisation d'un de ces 2 tests, qui sont les plus courants : le 1^{er} le test de Shapiro-Wilk et le test de K-S (Kolmogorov-Smirnov). En règle générale, à partir de 50 cas, vous devez utiliser un test de K-S. S'il est inférieur à cela, il est préférable d'utiliser le test de Shapiro-Wilk.

Attention, car en général, la significativité est inversée dans ce genre de test. Habituellement, pour qu'un test soit significatif, il faut que valeur de p soit inférieure à 0.05. Dans les tests de Shapiro-Wilk et de K-S, la normalité est établie lorsque la significativité (la valeur de p, la p value) est supérieure à 0.05. Dans le cas contraire, cela révèle que les données ne suivent pas une loi normale.

Voilà, c'est terminé pour aujourd'hui et je vous remercie de votre écoute ; je vous invite à visiter mon blog methodorecherche.com et on se retrouve lundi pour un nouvel épisode du podcast de Methodo Recherche. Bon WE et à lundi !

Abonnez-vous au Podcast suivant votre préférence d'écoute. Vous trouverez toutes les possibilités et les explications à l'URL :

<https://methodorecherche.com/subscribe-to-podcast/>

En complément, vous êtes libre de vous abonner à ma newsletter et recevoir gratuitement le bonus "6 clés essentielles pour réussir brillamment votre mémoire de recherche (ou votre thèse)".

<http://bit.ly/2RsYpll>



A très bientôt,

Christophe

