Corrélation de Spearman

La corrélation de Spearman est égal au coefficient de corrélation de Pearson calculé sur les variables de rang. Prenons deux échantillons $(X_i)_{i=1..n}$, $(Y_i)_{i=1..n}$. La variable de rang $\mathrm{rg}X_i$ est la variable telle que $\mathrm{rg}X_i = j \iff X_i = X_{(j)}$ (X_i est la $j$ème plus petite variable). Le coefficient de corrélation de Spearman est égal à :

$$r_s = \mathrm{corr}(\mathrm{rg}_X, \mathrm{rg}_Y) = \frac{\mathrm{cov}(\mathrm{rg}_X, \mathrm{rg}_Y)}{\sigma_{\mathrm{rg}_X}\sigma_{\mathrm{rg}_Y}}$$

Si toutes les valeurs de X et Y sont différentes, on a :

$$r_s = 1 - \frac{6\sum_i(\mathrm{rg}X_i - \mathrm{rg}Y_i)^2}{n(n^2-1)}$$

Pour tester la significativité de ce coefficient deux méthodes peuvent être utilisées :

1. En utilisant une transformation de Fisher $F(r) = \frac{1}{2}\ln\frac{1+r}{1-r} = \arctan r$ alors $z = \sqrt{\frac{n-3}{1.06}}F(r_s)$ est la cote Z de r_s et suit une loi normale sous l'hypothèse d'une corrélation nulle (H_0).
2. Avec un test de Student (ce qui est fait sous R et Python) : $t = r\sqrt{\frac{n-2}{1-r^2}} \overset{H_0}{\sim} St(n-2)$ sous H_0.

Comparaison de corrélations

Prenons maintenant le cas où l'on a un troisième échantillon $(Z_i)_{i=1..n}$ et on veut effectuer le test suivant :

$$\begin{cases} H_0 & r_s(X,Y) = r_s(X,Z) \\ H_1 & r_s(X,Y) \neq r_s(X,Z) \end{cases}$$

Un version simplifié du test serait de considérer $r_s(X,Z)$ comme étant fixé et non aléatoire : il suffit alors d'utiliser regarder si cette valeur est dans l'intervalle de confiance de fait à partir de la distribution de Student.

Si on considère que $r_s(X,Z)$ est aléatoire, on pourrait comparer les intervalles de confiance des deux corrélations de Spearman. De façon approximative, si les intervalles aux autres 97,5 % se croisent, on pourrait conclure qu'ils ne sont pas significativement différents.

Dans notre cas la situation est un peut différente : nous avons 6 observations de corrélations de spearman. Une idée serait de construire des intervalles de confiance de la façon suivante :

$$\left[\frac{1}{6}\sum_i r_{s,i} - q_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{6}\sum_i \frac{n-2}{1-r_{s,i}^2}}, \frac{1}{6}\sum_i r_{s,i} + q_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{6}\sum_i \frac{n-2}{1-r_{s,i}^2}}\right]$$