

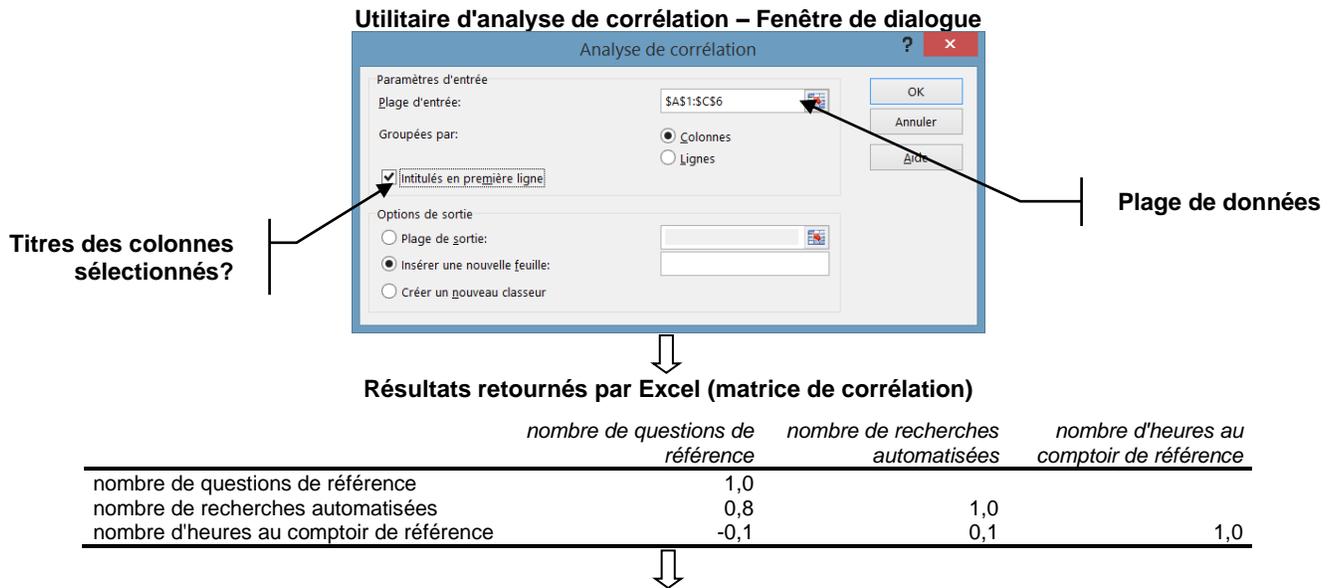
# Analyse de corrélation

L'analyse de corrélation permet de **quantifier la force du lien** entre des variables. Elle ne permet pas cependant de parler de cause-effet mais seulement d'association. Par exemple, soit le **nombre de questions de référence** reçues par un bibliothécaire dans une semaine, le **nombre de recherches automatisées** que ce même bibliothécaire a faites, et finalement le **nombre d'heures passées au comptoir de référence** par jour (cf tableau à droite). Y a-t-il un lien entre le nombre de recherches automatisées et le nombre de questions de référence? Entre le nombre de recherches et le temps passé au comptoir? Si oui, est-ce un lien fort ou faible?

	Données		
nombre de questions de référence	nombre de recherches automatisées	nombre d'heures au comptoir de référence	
10	7		3
8	7		2
6	8		5
4	2		5
11	10		6

Pour le savoir, il faut faire une **analyse de corrélation** en utilisant les données des variables qui nous intéressent.

## Analyse de corrélation dans Excel [Utilitaire d'analyse – Analyse de corrélation]



### Interprétation de la matrice de corrélation retournée par Excel

#### Explications

Chaque cellule présente une valeur correspondant à la force et le sens du lien entre la variable de la ligne où la cellule se trouve et la variable de la colonne où la cellule se trouve

- Par exemple, la valeur 0,8 est la force du lien entre le **nombre de recherches automatisées** et le **nombre de questions de référence**

#### Interprétation du lien

- Signe de la valeur (positif/négatif) : sens du lien**
  - Valeur positive : lien direct (i.e. une variable augmente, l'autre augmente)
  - Valeur négative : lien inverse (i.e. une variable augmente, l'autre diminue)
- Valeur (en valeur absolue i.e. sans tenir compte du + ou du -) : force du lien**
  - Repères (valeur arrondie; entre 0 et 1)
    - De 0 à 0,1 : Aucun lien ou très faible
    - De 0,2 à 0,3 : Lien faible
    - De 0,4 à 0,5 : Lien modéré
    - De 0,6 à 0,7 : Lien fort
    - 0,8 et plus : Lien très fort (multicolinéarité i.e. le lien est tellement fort que les variables sont presque identiques et l'une peut être pratiquement remplacée par l'autre)

#### Exemple tiré des résultats ci-dessus

- Lien entre le **nombre de questions de référence** et le **nombre de recherches automatisées**
  - Valeur (arrondie) : 0,8
  - Sens du lien : direct car valeur positive
  - Force du lien : Lien très fort
- Lien entre le **nombre de questions de référence** et le **nombre d'heures au comptoir de référence**
  - Valeur (arrondie) : -0,1
  - Sens du lien : inverse car valeur négative
  - Force du lien : Aucun lien ou très faible
- Lien entre le **nombre de recherches automatisées** et le **nombre d'heures au comptoir de référence**
  - Valeur (arrondie) : 0,1
  - Sens du lien : direct car valeur positive
  - Force du lien : Aucun lien ou très faible

**Conclusion?** On observe, dans nos données, un lien direct très fort entre le nombre de questions de référence et le nombre de recherches automatisées. Cependant, il n'y a aucun lien (ou un lien très faible) entre le nombre de questions de référence et le nombre d'heures au comptoir de référence, de même qu'entre le nombre de recherches automatisées et le nombre d'heures au comptoir de référence.