



INF112 - contrôle continu 2

L'objectif du contrôle continu 2 est de répondre à une question scientifique et de présenter les résultats de votre travail dans un mini-site web. Pour répondre à la question scientifique, vous devrez programmer quelques macros.

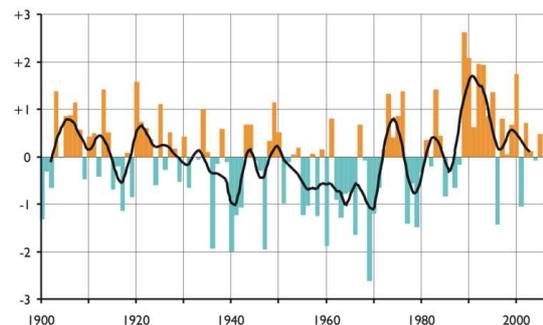
Plusieurs sujets vous sont proposés. Vous devez n'en traiter qu'un seul, au choix. La complexité des sujets est similaire.

1. Présentation des sujets.

Sujet 1 : ONA et pluviométrie.

L'**Oscillation Nord-Atlantique** (ONA) désigne un phénomène météorologique basé sur l'Atlantique Nord (en anglais *North Atlantic Oscillation* ou NAO est souvent utilisé dans la littérature). L'indice ONA mesure la différence de pression atmosphérique entre l'Anticyclone des Açores et la dépression d'Islande¹. La figure ci-contre montre la valeur de l'indice ONA en fonction des années².

L'ONA a été découverte durant les années 1920 par Sir Gilbert Walker. C'est un facteur déterminant dans le climat car elle est reliée à la position et la trajectoire des systèmes météorologiques du bassin atlantique-nord. Ainsi, s'attend à ce que certains phénomènes climatiques, tels que les précipitations (en France, et plus généralement autour de l'océan atlantique) soient influencés par les ONA.



Le but du sujet 1 est de rechercher s'il existe un lien entre les ONA et les précipitations.

Sujet 2 : Valeur-C et nombre de chromosomes.

Le **génome** est l'ensemble du matériel génétique d'un individu ou d'une espèce contenu dans son ADN (à l'exception de certains virus dont le génome est porté par des molécules d'ARN). Il contient en particulier toutes les séquences codantes (transcrites en ARN messagers, et traduites en protéines) et non-codantes (non transcrites, ou transcrites en ARN, mais non traduites) [wikipédia].

La **valeur-C** représente la taille d'un génome présent dans une cellule. Par convention, la valeur-C est mesurée sur les cellules haploïdes³. Elle est exprimée en paire de bases, ou en pico-gramme (10^{-12} gramme). Cette valeur-C est différente pour chaque espèce mais globalement constante pour chaque individu d'une même espèce.



Intuitivement, on pourrait s'attendre à ce que plus il y a de chromosome dans une cellule, plus le matériel génétique est important.

Le but du sujet 2 est de déterminer si la valeur-C est corrélée au nombre de chromosomes.

¹ http://fr.wikipedia.org/wiki/Oscillation_nord-atlantique

² <http://fr.wikipedia.org/wiki/Fichier:Winter-NAO-Index.png>

³ Une cellule biologique est haploïde lorsque les chromosomes qu'elle contient sont chacun en un seul exemplaire.

2. Analyse préliminaire

Avant de commencer un sujet, prenez le temps de réfléchir aux questions suivantes :

- Que va-t-on comparer ?
- Comment les données doivent-elle être disposées ?
- Quel type de comparaison va-t-on faire ?

3. Sujet 1 : l'étude de l'oscillation Nord-Atlantique

3.1 Les données

Sur Internet, il est possible de récupérer des valeurs des indices NOA sur plusieurs années. Selon ; les sites, on obtient des valeurs annuelles ou mensuelles. Le formatage des données n'est pas forcément identique pour tous les fichiers. Par ailleurs, selon le mode de calcul, les valeurs obtenues peuvent être un peu différentes.

Lorsque l'on récupère des données sur Internet, on est souvent amené à faire des traitements pour mettre les données dans un format adéquat pour la suite. Ici, nous nous sommes chargés de ce traitement. Vous avez à votre disposition un fichier Excel lequell on a 3 colonnes. La première indique l'année, la seconde indique les valeurs NOA récoltée sur un premier site et la troisième indique les valeurs récoltées sur un second site.

	A	B	C
1	année	indice 1	indice 2
2	1864	-1,02	-0,44
3	1865	-1,24	0,33
4	1866	0,54	0,31
5	1867	-1,38	-0,07
6	1868	2,81	1,58

En ce qui concerne les données pluviométriques, Météo France nous a fourni 3 fichiers, comportant indiquant la quantité de pluie récoltée sur 3 sites différents. Les données sont mensuelles et sont récoltées entre 1950 et 2008.

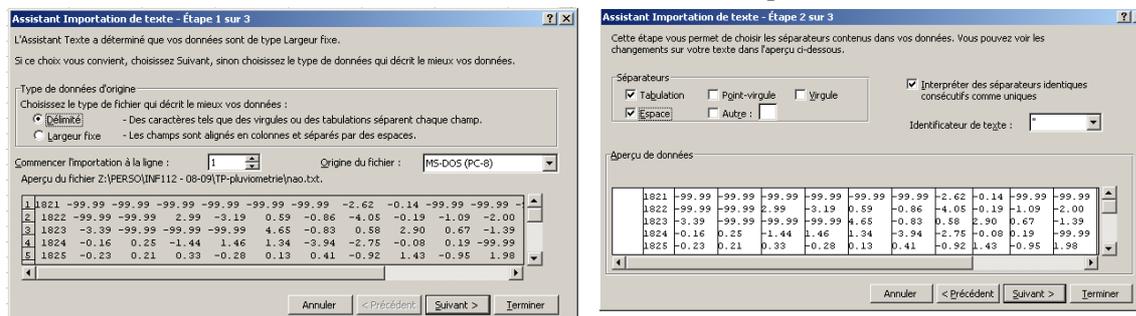
Les indices ONA étant annuels, il faut transformer les données pluviométriques mensuelles en données annuelles. Les exercices ci-après qui vous guideront vers une réalisation pas à pas.

Les fichiers fournis par Météo-France sont textuels. Pour les ouvrir sous Excel, il faut procéder avec la procédure donnée ci-dessous.

ATTENTION : un des fichiers a des données manquantes. **Commencer** par traiter le fichier de **Chamonix**, qui est complet.

Les fichiers textuels pour s'ouvrir sous Excel de la façon suivante :

- Ouvrir le fichier texte avec Excel
- Dans la fenêtre qui s'ouvre, choisir l'option « délimité », puis cliquer sur « suivant ».
- Dans la fenêtre suivante, cocher « espace », puis cliquer sur « suivant ».
- Dans la troisième fenêtre, choisir le format des données (par défaut, standard est OK).



Exercice 0. Afficher les années

Produire une itération qui inscrit les années de 1950 à 2008 dans la colonne 5 (une année par ligne).

Exercice 1.

Proposer un algorithme qui fait la somme de 12 cellules de la colonne 3 comprises entre les lignes 2 et 13. On utilisera une itération et une variable intermédiaire Som. Pensez à initialiser Som **avant** l'itération. Copier la valeur de Som dans la première cellule de la colonne 6 **après** l'itération.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 2.

Il faut répéter ce traitement pour chaque paquet de 12 lignes.

Introduire une action paramétrée qui fait le même traitement que pour l'exercice 1. On choisira comme paramètre le numéro de la ligne de départ pour la somme et le numéro de la ligne où sera rangé le résultat.

Traduire votre algorithme en une macro, vérifier vos résultats.

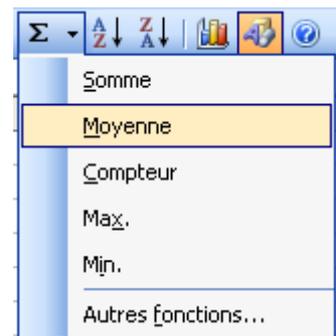
Exercice 3.

Ecrire l'algorithme qui appellera l'action précédente pour chaque paquet de 12 lignes entre les lignes 2 et 709. Traduire votre algorithme en une macro, vérifier vos résultats.

Un autre traitement des données peut être utile. Les valeurs ONA sont calculées par rapport à une moyenne. Elles sont positives ou négatives par rapport à cette moyenne. On se propose de calculer les différences de relevés pluviométriques par rapport à la moyenne de données (calculées pour chaque ville). Il faut donc (1) calculer la moyenne des données pluviométriques sur les années et (2) calculer la différence à la moyenne.

Pour calculer la moyenne des données pluviométriques, on utilisera la fonction Excel.

- Sélectionner les données dont vous voulez calculer la moyenne, plus une cellule.
- Dans le menu adéquat, choisissez d'effectuer une moyenne

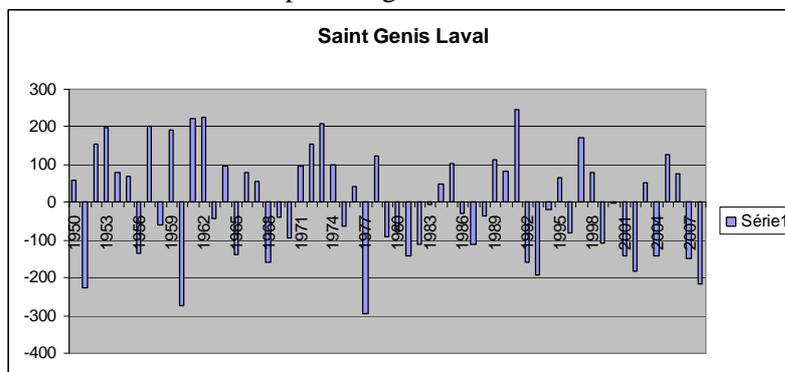
**Exercice 4. Différence à la moyenne**

Ecrire l'algorithme pour calculer la différence entre la moyenne et la cellule colonne 6 de la même ligne. Ecrire le résultat dans la colonne 7.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 5. Graphe de variation

Sélectionner les données « différences à la moyenne », dessiner un graphe représentation la variation des données. N'oubliez pas la légende.

**Exercice 6**

Exécuter les macros nécessaires pour mettre en forme les données sur chaque page contenant des données pluviométriques.

3.2 Vers une comparaison

Pour chacune des sources de valeurs ONA dans le fichier Excel fourni, produire à graphe similaire à ceux faits pour la pluviométrie.

Exercice 8. Comparaison visuelle.

Observer les différents graphes. Observez-vous une évolution similaire ? Qu'en concluez vous ?

Il n'est pas toujours très facile de décider s'il existe ou non une corrélation entre des données sur la seule base de l'observation visuelle. C'est pourquoi, il est intéressant d'utiliser des outils mathématiques pour prendre une telle décision.

Exercice 9. Comparaison à l'aide d'outil mathématique.

Reportez-vous à la section 7 pour calculer les différents coefficients de corrélation entre d'une part, des valeurs ONA et d'autres part des données pluviométriques. Attention à bien indiquer ce que vous comparez.

Exercice 10. Analyse

Que pensez-vous des résultats ?

3.3 Autre comparaison

On propose de mener une autre analyse permettant de comparer le bilan de masse glaciaire (gain ou perte de masse des glaciers d'octobre à septembre) sur les massifs alpins et l'indice ONA. Pour cela, on a mesuré les bilans sur 5 glaciers différents. On a calculé les indices ONA sur ces mêmes périodes.

Exercice 11. Analyse

Comparer les bilans de masse glaciaire et les indices ONA des périodes correspondantes selon les méthodes vues précédemment. Que pensez-vous des résultats ?

4. Sujet 2 : étude de la valeur-C

3.1 Les données

Nous avons téléchargé des données depuis le site « Animal Genome Size Database⁴ ».

Les données ne permettent pas de répondre directement à la question car

- pour chaque espèce, plusieurs valeurs-C ont parfois été mesurées,
- le nombre de chromosomes n'est pas toujours indiqué.

Pour répondre à la question de recherche, il faut disposer des les données de telle sorte que l'on puisse comparer pour chaque espèce, la valeur-C et le nombre de chromosomes.

- Ouvrez le fichier Excel : *genome_size_data_141112_07_01_56.xls*
- Observez les données.

	A	B	C	D	E	F	G	H	I
1	Animal Genome Size Data:: www.genomesize.com								
2									
3	Data downloaded under intellectual property agreement by lydie du Bousquet, lydie.du-bousquet@imag.fr								
4									
5									
6	Phylum	Sub Phylum	Class	Order	Family	Species	Common Name	C-value	Chrom num
7	Chordata	Vertebrata	Amphibia	Anura	Arthroleptidae	Arthroleptis stenodactylus	Common squeaker	3,37	
8	Chordata	Vertebrata	Amphibia	Anura	Ascaphidae	Ascaphus truei	Tailed frog	3,95	46,00
9	Chordata	Vertebrata	Amphibia	Anura	Ascaphidae	Ascaphus truei	Tailed frog	4,10	
10	Chordata	Vertebrata	Amphibia	Anura	Ascaphidae	Ascaphus truei	Tailed frog	4,22	
11	Chordata	Vertebrata	Amphibia	Anura	Ascaphidae	Ascaphus truei	Tailed frog	4,26	
12	Chordata	Vertebrata	Amphibia	Anura	Bombinatoridae	Bombina bombina	European fire-bellied toad	10,30	24,00
13	Chordata	Vertebrata	Amphibia	Anura	Bombinatoridae	Bombina bombina	European fire-bellied toad	10,57	

Sur chaque ligne figure, le nom de l'espèce (entre les colonnes A et G), la valeur-C (colonne H) et le nombre de Chromosome). Ligne 7, décrit le « Common squeaker ». La valeur-C est connue, mais pas le nombre de chromosome. Lignes 8 à 11, sont indiquées quatre Valeur-C pour le « Tailed frog » et le nombre de chromosome (seulement sur la ligne 8).

La première partie du travail vise à calculer la moyenne de la valeur-C pour chaque espèce. Les résultats seront écrits sur la seconde feuille 2 du fichier Excel.

Il faut utiliser un algorithme car il y a de nombreuses espèces à considérer et le nombre de valeur-C mesurées est différent pour chaque espèce. Les exercices 0 à 2 sont des exercices préparatoires pour calculer la moyenne de chaque espèce.

Exercice 0. Somme

Proposer un algorithme qui calcule la somme des valeurs qui se trouvent colonne 8 entre les lignes 8 et 11 (inclus). Le résultat sera écrit ligne 10, colonne 2, sur la feuille « Feuil2 ».

Pour désigner la cellule ligne 10 colonne 2 de « Feuil2 », on écrit Feuil2.cellule(10,2).

On prendra soin d'introduire une itération (même si cela concerne que 4 valeurs) car c'est un exercice préparatoire qui va être utilisé dans la suite. On introduira aussi une variable intermédiaire pour calculer la somme. On réfléchira soigneusement au type de cette variable.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 1. Moyenne.

Modifier l'algorithme précédent pour calculer la moyenne des valeur-C pour les valeurs-C entre les lignes 8 et 11. On écrira le résultat ligne 10, colonne 2, sur la feuille « Feuil2 ».

Sur la colonne 1, ligne 10, on écrira le nom de l'espèce donné colonne 7 (feuille de données)

⁴ Gregory, T.R. (2011). Animal Genome Size Database. <http://www.genomesize.com>.

Pour calculer la moyenne, il faut d'une part la somme des valeurs et d'autre part le nombre de valeur. On utilisera une variable intermédiaire NbVal pour calculer le nombre de valeurs. Pensez à initialiser les variables **avant** l'itération.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 2. Moyenne pour un nombre indéfini de ligne

Comme nous l'avons dit plus haut, pour chaque espèce, nous avons plusieurs valeur-C, mais on ne sait pas dire combien. On va donc faire la somme des valeur-C à partir d'un point donné, « tant que l'espèce ne change pas » (l'espèce est indiquée colonne 6).

Proposer un algorithme avec une itération « **tant que** » pour calculer la moyenne des valeur-C à partir la ligne 12. On écrira le résultat ligne 10, colonne 2, sur la feuille « Feuil2 ».

Sur la colonne 1, ligne 10 de la feuille « Feuil2 » on écrira l'espèce.

Traduire votre algorithme en une macro, vérifier vos résultats. Sauvegarder votre fichier avant d'exécuter la macro (les macros contenant un « tant que » contiennent souvent des erreurs !).

Il faut maintenant automatiser le traitement pour toutes les espèces. On procède en deux temps. On produit une version paramétrée du calcul de la moyenne puis on l'utilise pour faire le traitement sur toute la page. Le résultat attendu est donné ci-contre.

	A	B	C	D
1	Arthroleptis stenodactylus	3,37		8
2	Ascapus truei	4,13	46	12
3	Bombina bombina	11,38	24	18
4	Bombina orientalis	8,18	24	21
5	Bombina pachypus	10,55	24	22
6	Bombina variegata	8,90	24	28
7	Bombina variegata scabra	10,72		29
8	Bombina variegata variegata	10,36		30
9	Atelopus flavescens	6,86		31
10	Atelopus sp.	3,56		32
11	Bufo alvarius	5,65	22	33
12	Bufo americanus	5,17		38
13	Bufo andrewsi	5,88		39
14	Bufo arabicus	5,99		40
15	Bufo brongersmai	5,70		41
16	Bufo bufo	6,69	22	52

Exercice 3. Calcul de moyenne paramétrée

Proposer un algorithme paramétré Calcul_moyenne, ayant 2 paramètres : le numéro de la ligne de cellule du départ et le numéro de la ligne où doit être rangé le résultat.

En plus du nom de l'espèce et de la moyenne, l'algorithme Calcul_moyenne écrit le numéro de la ligne de la prochaine espèce à étudier en colonne 4 de la Feuil2.

Traduire votre algorithme en une macro.

Exercice 4

Ecrire l'algorithme qui appellera l'action précédente pour traiter quelques lignes.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 5

Ecrire l'algorithme qui appellera l'action précédente pour toute la feuille de calcul.

Traduire votre algorithme en une macro, vérifier vos résultats.

Exercice 6. Reporter le nombre de chromosomes sur la feuille 2.

Pour reporter le nombre de chromosomes sur la feuille 2, il faut savoir à quelle ligne le nombre de chromosome est indiqué pour chaque espèce. Comme on ne le sait pas, on va traiter les données pour que l'information soit sur la première ligne de chaque espèce. Avez-vous une idée de comment faire (simplement) ?

Une façon de faire consiste à trier les données avec la fonction de tri proposé par Excel.

Prendre soin de sélectionner les données de la ligne 7 à la dernière et toutes les colonnes qui comportent des données. Trier d'abord par nom (colonne 7) puis par nombre de chromosome (croissant). Lorsque le nombre de chromosome est connu, il sera alors affiché sur la première ligne de l'espèce.

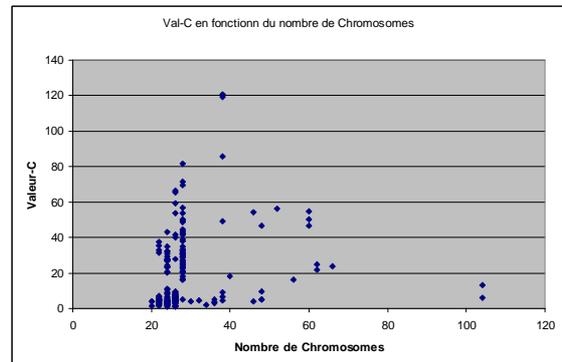


Seules les espèces dont on connaît le nombre de chromosome sont intéressantes.
 Une solution plus pragmatique consiste à trier les résultats affichés sur la Feuil2 et ne sélectionner que ceux dont on connaît le nombre de Chromosomes.

4.2 Vers une comparaison

Exercice 7. Comparaison visuelle.

Si la valeur-C augmente avec le nombre de chromosomes, on devrait observer visuellement l'augmentation sur un graphe. Tracer un graphe de type « nuage de points » permettant d'afficher la valeur-C en fonction du nombre de chromosomes. N'oubliez pas la légende. Vous devriez obtenir un résultat similaire au graphe ci-contre.



Il n'est pas toujours très facile de décider s'il existe ou non une corrélation entre des données sur la seule base de l'observation visuelle. C'est pourquoi, il est intéressant d'utiliser des outils mathématiques pour prendre une telle décision.

Exercice 8. Comparaison à l'aide d'outil mathématique.

Reportez-vous à la section 7 pour calculer le coefficient de corrélation entre les valeurs-C d'une part et le nombre de chromosomes d'autre part. Attention à bien indiquer ce que vous comparez.

Exercice 9. Analyse

Que pensez-vous des résultats ?

4.3 Autre comparaison

Il existe plusieurs sites web permettant d'obtenir la valeur-C. Par exemple, sur le site <http://data.kew.org/cvalues/>, nous avons téléchargé des données sur les végétaux et les avons sauvegardées dans le fichier *RBG_Kew_DNA_C-values.xls*.

Exercice 10. Analyse

Adapter et appliquer les traitements précédents sur ces nouvelles données.

Que pensez-vous des résultats ?

Ce sujet a été inspiré du document http://pbil.univ-lyon1.fr/R_svn/pdf/bem4.pdf

5. Mini-site web (Pour tous les sujets)

On s'attend à ce que votre mini-site web comporte 4 pages html et une feuille de style.

- *index.html* : présente le contexte et votre hypothèse. A titre **exceptionnel**, vous pouvez récupérer des informations sur d'autres sites web et les copier-coller⁵. Mais vous devez clairement identifier vos sources et mettre des liens vers les pages que vous avez copiées.
- *exp.html* : présente votre expérience, c'est-à-dire
 - les données que vous avez utilisées (nature, format, origine),
 - une description des les traitements (principe des algorithmes) que vous avez faits.
- *res.html* : décrit les résultats.
- *quizz.html* : 2 à 5 questions de votre choix et un bouton qui permet de calculer le score.

Les exigences sur l'organisation du mini-site :

- Tous les fichiers doivent être regroupés dans un répertoire **ayant votre nom**.
- Ce répertoire principal doit comporter deux sous-répertoires : **images** et **data**.
- Les pages HTML doivent être dans le répertoire principal.
- Les pages *exp.html*, *res.html* et *quizz.html* doivent être accessibles depuis *index.html*.
- Chacune des ces pages doivent pouvoir retourner à *index.html*.
- Il doit y avoir des liens qui permettent de passer directement de *exp.html* à *res.html* et vice-versa.
- Une feuille de style doit être utilisée pour toutes les pages HTML. Choisissez une feuille de style simple, qui permet une lecture facile de vos pages et qui change de la présentation par défaut (sans feuille de style).
-
- Le mini-site doit contenir des images.
- Toutes les images utilisées doivent être rassemblées dans le répertoire **images**.
- L'ensemble des fichiers de données brutes (fichiers textes fournis et ceux éventuellement trouvés) doivent être rassemblés dans le répertoire **data**.
-
- Des liens doivent permettre d'ouvrir ces fichiers depuis *exp.html*.
- Le fichier Excel comportant vos macros, vos données traitées et vos résultats doit aussi être sauvegardé dans le répertoire data.
- Il doit être accessible depuis *exp.html* ou *res.html*.

Les exigences sur le contenu du mini-site :

- La page *index.html* doit faire apparaître vos noms, prénoms et groupes.
- Votre mini-site doit comporter au moins une image, un tableau, une liste et une image map.
- Il doit y avoir des liens externes vers d'autres sites web que le vôtre.
- Il doit y avoir un lien interne.
- Il doit y avoir un bouton dynamique.

⁵ Recopier des informations prévenant d'Internet ou de livres sans citer ses sources s'appelle du plagiat. L'UJF est particulièrement attentive à ce type de faute, qui peut être sanctionnée gravement.

6. Rendu du travail

Suivez les instructions de votre enseignant, qui pourra choisir entre récupération d'un fichier compressé par mail ou copie de votre répertoire sur une clef USB.

Créer un fichier compressé

Un fichier compressé est un moyen de rassembler différents répertoires et/ou fichiers dans un seul fichier, tout en compressant les données.

- Sur Sarado, sélectionner le répertoire qui contient les fichiers de votre mini-site.
- Cliquer sur ce répertoire avec le bouton droit de la souris,
- Sur le menu déroulant, choisir de compresser le répertoire.

Ce fichier compressé pourra vous être demandé par votre enseignant. **Vérifiez que les noms des membres du binôme apparaissent dans le nom du fichier.**

Le travail attendu pour le CC2 est donc de traiter les données qui vous sont fournies et de rédiger votre compte-rendu sur la forme d'un mini-site web. Cela correspond travail décrit dans les pages précédentes.

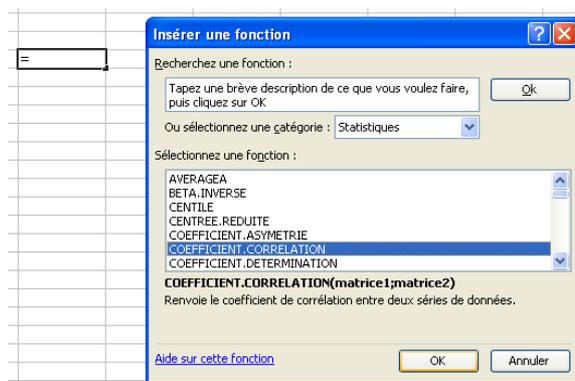
7. Outils pour l'analyse de corrélation.

Il n'est pas toujours très facile de décider s'il existe ou non une corrélation entre des données sur la seule base de l'observation visuelle. C'est pourquoi, il est intéressant d'utiliser des outils mathématiques pour prendre une telle décision.

Il existe plusieurs « outils » mathématiques qui permettent de décider si des ensembles de données sont corrélés (corrélation de Pearson par exemple). Toutefois ces outils ne sont pas tous équivalents et doivent être utilisés dans des contextes appropriés.

Vous verrez ces outils dans vos UEs de statistiques ou de mathématiques.

Excel dispose d'une fonction pré-programmée qui lorsqu'on lui fournit deux ensembles de données, calcule un entier entre -1 et 1 (coefficient de corrélation) qui représente si les valeurs des 2 ensembles sont corrélées.



Pour utiliser cette fonction, choisissez une cellule vide dans laquelle vous souhaitez récupérer le coefficient de corrélation, puis dans le menu « insertion », choisissez « insérer une fonction » puis « coefficient de corrélation ».

Interprétation. Plus le coefficient de corrélation est proche de 1, plus les données sont fortement corrélées (une augmentation de l'un s'observe chez l'autre. Si le coefficient est proche de -1, les données sont négativement corrélées : lorsqu'une augmentation de l'un s'observe, une diminution s'observe chez l'autre). Si la valeur est proche de 0 (négative ou positive), on dit qu'il n'y a pas de corrélation.