

TP n°3 : Fonctions statistiques et Macros

Partie 1 : Fonctions statistiques

Excel permet d'effectuer une analyse statistique d'un fichier de données à l'aide d'un utilitaire qu'il faut au préalable charger comme suit :

1. Cliquer sur l'onglet fichier
2. Choisissez Options
3. Sélectionnez le module Compléments
4. Cliquez sur le bouton Atteindre (en bas de page)
5. Cochez la case « Analysis ToolPak », puis cliquez sur OK.

Vous pouvez à présent manipuler aisément des fonctions statistiques prédéfinies telles que

- La **MEDIANE ()** qui renvoie la valeur qui se trouve au milieu d'une série ;
- L'**ECARTYPE ()** qui mesure la dispersion des valeurs par rapport à la moyenne ;
- Le **coefficient d'asymétrie** qui caractérise le degré d'asymétrie d'une distribution par rapport à sa moyenne. Une asymétrie positive indique une distribution unilatérale décalée vers les valeurs les plus positives. Une asymétrie négative indique une distribution unilatérale décalée vers les valeurs les plus négatives.
- Le **régression linéaire** qui permet d'expliquer la corrélation (comportement) d'une des variables avec les autres données.

Exercice 1 : Football

Télécharger le fichier football.xlsx, disponible sur <https://mycourse.dauphine.fr> ou <https://www.lamsade.dauphine.fr/~mayag/teaching.html>, qui présente quelques chiffres statiques de 20 clubs de football. Dans le ruban du menu Données, à droite, cliquez sur le bouton « Utilitaire d'analyse ».

- a) Parmi les outils d'analyse, cliquez sur Statistiques descriptives. Remplissez la grille qui s'affiche comme suit : la plage d'entrée contient la donnée \$B\$1 : \$F\$21 ; A droite de « Insérer une nouvelle feuille », indiquez RESULTAT. Cochez les cases « Intitulés en première ligne », « Rapport détaillé » et « Niveau de confiance pour la moyenne » (à 95%).

Cliquez enfin sur OK. Vous devriez obtenir une feuille semblable à cette image :

	A	B	C	D	E	F	G	H	I	J
1	Victoires		Nuls		Défaites		Buts marqués		Buts encaissés	
2										
3	Moyenne	14,15	Moyenne	9,7	Moyenne	14,15	Moyenne	45,8	Moyenne	45,8
4	Erreur-type	1,124546692	Erreur-type	0,341051007	Erreur-type	1,16138529	Erreur-type	2,93849227	Erreur-type	2,038575354
5	Médiane	13,5	Médiane	10	Médiane	14	Médiane	44	Médiane	44,5
6	Mode	20	Mode	11	Mode	6	Mode	50	Mode	40
7	Écart-type	5,029125696	Écart-type	1,525226472	Écart-type	5,19387291	Écart-type	13,141337	Écart-type	9,116786137
8	Variance de l'échantillon	25,29210526	Variance de l'éch	2,326315789	Variance de l'échan	26,9763158	Variance de l	172,694737	Variance de l'éc	83,11578947
9	Kurtosis (Coefficient d'aplatissement)	-0,825753236	Kurtosis (Coeff	-0,838065576	Kurtosis (Coeffici	-0,249593	Kurtosis (Co	-0,5273405	Kurtosis (Coeff	-0,624022017
10	Coefficient d'asymétrie	0,095024453	Coefficient d'asy	-0,324359021	Coefficient d'asym	0,24553743	Coefficient d	0,46390305	Coefficient d'as	0,331674943
11	Plage	18	Plage	5	Plage	19	Plage	45	Plage	33
12	Minimum	5	Minimum	7	Minimum	6	Minimum	27	Minimum	29
13	Maximum	23	Maximum	12	Maximum	25	Maximum	72	Maximum	62
14	Somme	283	Somme	194	Somme	283	Somme	916	Somme	916
15	Nombre d'échantillons	20	Nombre d'échan	20	Nombre d'échantill	20	Nombre d'éc	20	Nombre d'écha	20
16	Niveau de confiance(95,0%)	2,353703277	Niveau de confia	0,713827962	Niveau de confian	2,43080735	Niveau de co	6,15033501	Niveau de confi	4,266787253

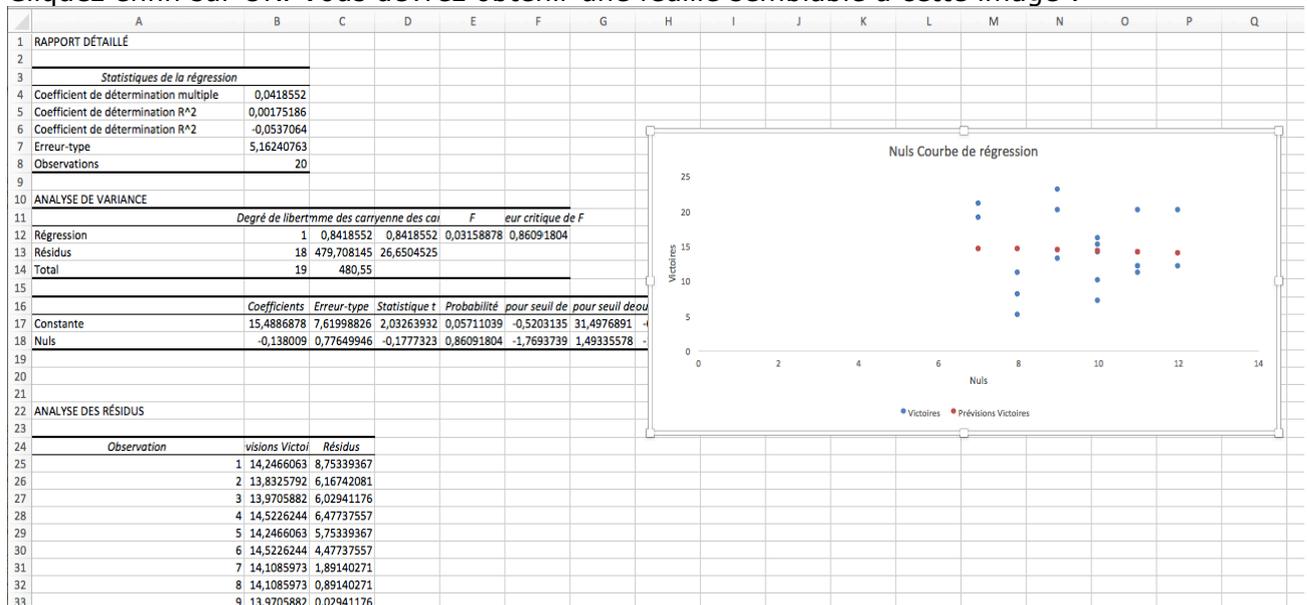
Pour chaque colonne de valeurs quantitatives (Victoires, Nuls, Défaites, Buts marqués, Buts encaissés), on obtient les informations statistiques suivantes :

- **La Moyenne,**
- **L'Erreur-type** (c'est l'écart-type de la moyenne utilisée comme estimateur) qui correspond à l'écart-type divisé par la racine carrée de la taille de l'échantillon. Par exemple, pour la colonne Victoire, l'erreur-type est égal à 5,029/racine (20).
- **La Médiane**

- **Le Mode** : Permet de connaître le nombre présentant le plus d'occurrences dans une série de données
- **L'écart-type** (la racine carrée de la variance)
- **La variance**
- **Le coefficient d'aplatissement de Kurtosis** qui caractérise la forme de pic ou l'aplatissement relatifs d'une distribution comparée à une distribution normale. Un kurtosis positif indique une distribution relativement pointue, tandis qu'un kurtosis négatif signale une distribution relativement aplatie.
- **Le coefficient d'Asymétrie** : Renvoie l'asymétrie d'une distribution. Cette fonction caractérise le degré d'asymétrie d'une distribution par rapport à sa moyenne. Une asymétrie positive indique une distribution unilatérale décalée vers les valeurs les plus positives. Une asymétrie négative indique une distribution unilatérale décalée vers les valeurs les plus négatives.
- **La Plage** (différence entre le maximum et le minimum)
- **Le Minimum**
- **Le Maximum**
- **La Somme**
- **Le nombre d'échantillon** (nombre total d'éléments dans la colonne)
- **Le Niveau de confiance** reflète le degré de certitude de la marge d'erreur (intervalle de confiance)

b) Parmi les outils d'analyse, cliquez sur **Régression linéaire**. Remplissez la plage pour la variable Y par \$B\$1 : \$B\$21 et la plage pour la variable X par \$C\$1 : \$C\$21. Cochez la case « intitulé présent ». A droite de « Insérer une nouvelle feuille », indiquez REGRESSION. Cochez les cases « Courbes de régression » et « Niveau de confiance pour la moyenne » (à 95%).

Cliquez enfin sur OK. Vous devrez obtenir une feuille semblable à cette image :



Sur ce graphique on peut lire les informations suivantes :

- **Coefficient de détermination multiple** : Dans le cas d'une variable explicative unique, il s'agit du coefficient de régression linéaire. Les deux variables ne sont pas linéairement corrélées s'il est nul et elles sont d'autant mieux corrélées s'il est proche de 1 ou -1.
- **Coefficient de détermination R^2** : représente la part de variation expliquée dans le modèle (c'est le carré du coefficient de détermination multiple)
- **Coefficient de détermination R^2 ajusté** : il dépend du nombre de variable explicative
- **Erreur-type** : estimation de l'écart-type supposé constant des résidus

- **Observations**
- **Résidu** : l'écart entre la valeur observée $y(i)$ et la valeur $b x(i) + a$ estimée par la régression
- **Les prévisions pour Y** donnent les valeurs estimées de la variable expliquée.

c) Analysez l'existence d'une corrélation linéaire (régression linéaire) entre les buts marqués et les buts encaissés ? Entre les victoires et les buts marqués ? Entre les défaites et les buts encaissés ?

Exercice 2 : Hôtels de Luxe

Téléchargez le fichier Hôtels Luxe.xlsx disponible sur <https://mycourse.dauphine.fr> ou <https://www.lamsade.dauphine.fr/~mayag/teaching.html> . Il s'agit de données réelles sur des hôtels recueillies par le site internet www.booking.com. Pour chaque hôtel, nous disposons des informations suivantes :

- **Nom Hôtel** : le nom de l'hôtel de luxe ;
 - **Total Négatifs** : le nombre total de mots "négatifs" recensés dans les commentaires laissés par les internautes pour cet hôtel (critère à minimiser) ;
 - **Total Positifs** : le nombre total de mots "positifs" recensés dans les commentaires laissés par les internautes pour cet hôtel (critère à maximiser) ;
 - **Moyenne Reviews** : la moyenne arithmétique des notes (/10) données par les internautes ayant laissé des commentaires pour cet hôtel (critère à maximiser).
 - **Total Reviews** : nombre total de commentaires laissés pour cet hôtel (critère à maximiser) ;
 - **Note de Booking** : Note globale (/10) attribuée par le site internet www.booking.com à cet hôtel.
1. Dans l'onglet France, vous trouvez 457 hôtels de Luxe. Pour cette liste, analysez l'existence d'une corrélation linéaire entre les Total positifs et Total Négatifs, Total négatifs et Moyenne Reviews, Total Positifs et Moyenne Reviews.
 2. Refaire la même analyse avec un autre pays de votre choix, parmi les 5 autres onglets.
 3. Comparez vos deux analyses. Quelles conclusions pouvez-vous en tirer ?

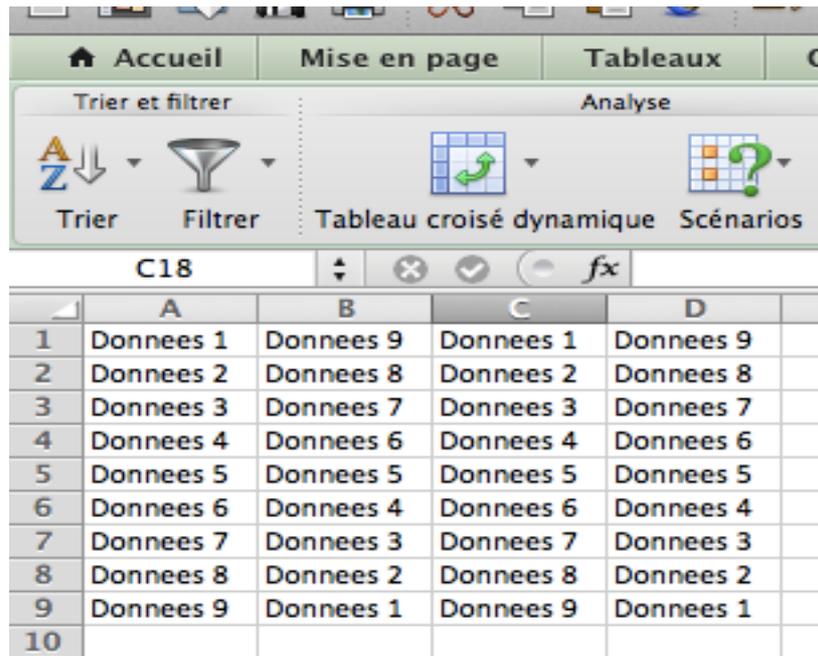
Partie 2 : Macros

Les macros servent essentiellement à automatiser et personnaliser des actions dans le classeur. Vous pouvez ainsi écrire des procédures pour les tâches répétitives, mais aussi adapter l'outil Excel pour qu'il réponde exactement à vos besoins particuliers (interagir avec les manipulations de l'utilisateur, piloter d'autres applications...). En gros, les macros sont pratiques si vous avez régulièrement une longue suite d'actions à effectuer.

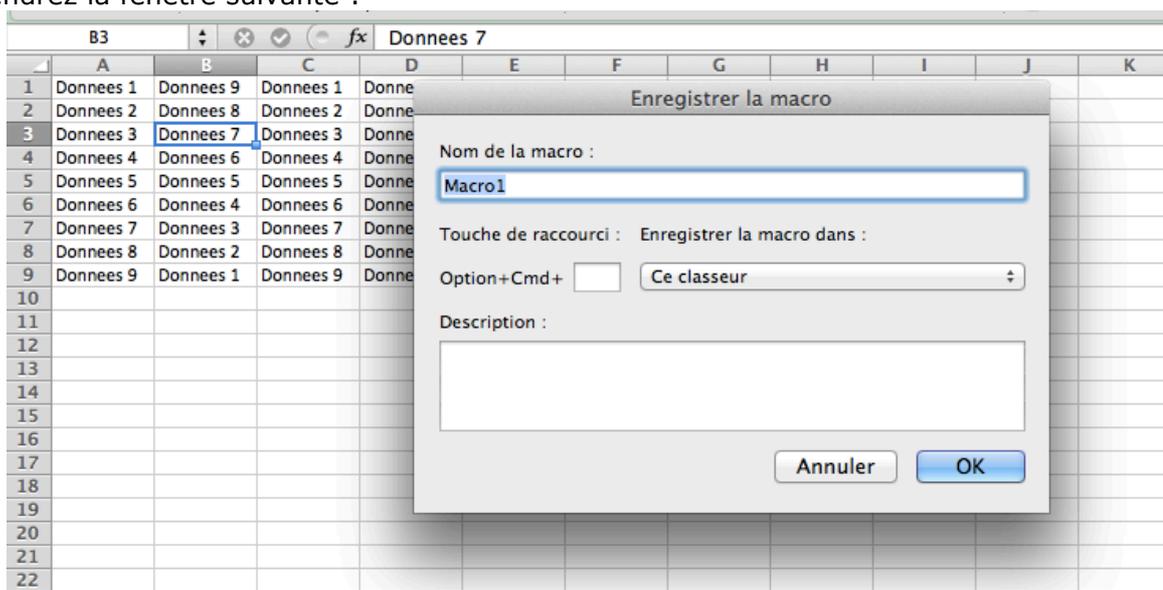
Exercice 3 : Ma première macro.

Dans cet exercice, vous allez pouvoir apprendre à utiliser l'**enregistreur de macro** intégré à Excel et créer ainsi votre première macro qui consiste à mettre en gras et en italique une cellule.

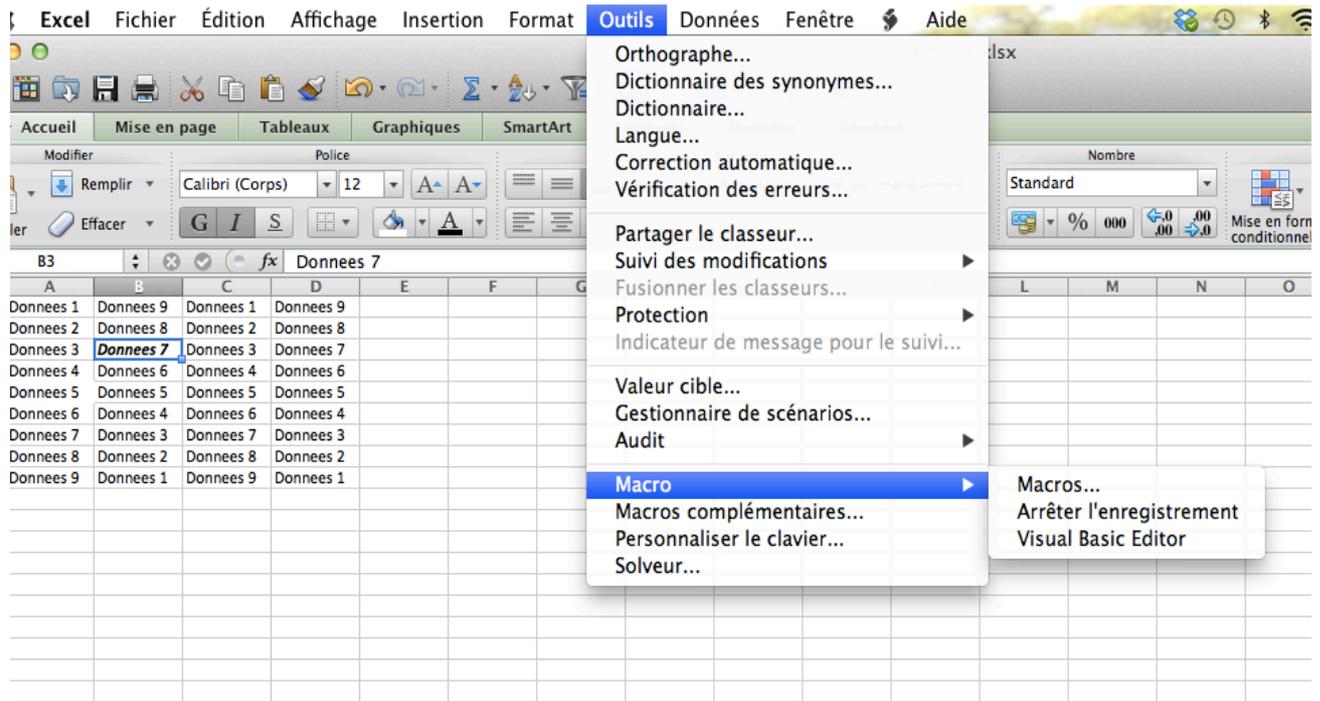
1. Reproduire la feuille Excel suivante :



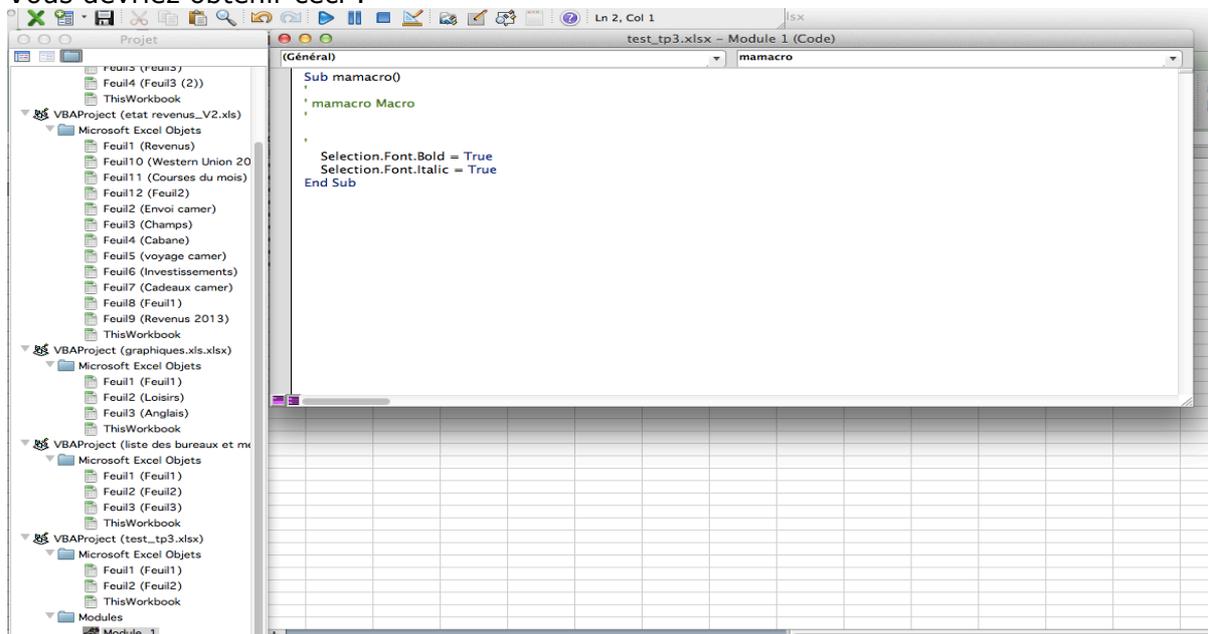
2. Sélectionner une cellule quelconque, puis dans menu faites Outils -> Macro -> Nouvelle macro ou Outils -> Macro -> Enregistrer macro (disponible dans l'onglet « Développeur »). Vous obtiendrez la fenêtre suivante :



3. Donner un nom à cette macro. Ce nom doit toujours commencer par une lettre et les espaces ne sont pas autorisés. Définissez un raccourci clavier (par exemple Ctrl+m) pour rappeler ultérieurement cette macro et vérifiez que l'enregistrement se fait bien dans « Ce classeur ».
4. Cliquez sur OK. L'enregistrement commence à présent. Mettre en gras et en italique la cellule sélectionnée et cliquez sur arrêter l'enregistrement comme ci-dessous (pour certaines versions d'Excel, une barre d'outils flottante permet de stopper l'enregistrement)



5. Sélectionner une autre cellule et exécuter votre macro en faisant dans menu Outils -> Macro -> macro.
 6. Pouvez-vous l'exécuter sur une plage de plusieurs cellules sélectionnées en même temps ?
 7. Nous venons de créer notre première macro grâce à l'enregistreur de macro et nous savons également la rappeler pour reproduire "mécaniquement" une suite d'actions préenregistrées. Ouvrez à nouveau la boîte de dialogue de votre macro et cliquez sur le bouton « Modifier ».
- Vous devriez obtenir ceci :



Vous êtes alors projetés instantanément dans l'Éditeur Visual Basic (VBE) et à l'endroit même où est stocké le code VBA associé à votre macro enregistrée. La macro complémentaire, dans Excel, est un code écrit en VBA.

Exercice 4 :

Vous recevez chaque semaine la liste des nouveaux clients par email. Cette liste est toujours saisie en minuscule. Afin de vous simplifier la tâche, créez une macro de transformation de la colonne A en

majuscules (vous pourrez utiliser la fonction prédéfinie d'Excel **MAJUSCULE**). Le fichier tel que reçu est :

	A	B	C
1	Nouveaux clients	Ventes	
2	angelli		
3	beaud		
4	bernard		
5	blanes		
6	bouisseau		
7	boussy		
8	bremont		
9	chauvet		
10	chneider		
11	claire		
12	clement		
13	cluny		
14	damien		
15			
16			
17			
18			
19			