Table des matières

[1. À propos 2](#_Toc12012877)

[2. Introduction 3](#_Toc12012878)

[3. Vocabulaire 3](#_Toc12012879)

[4. Lancer Excel 3](#_Toc12012880)

[5. Saisie de données 4](#_Toc12012881)

[6. Bonnes pratiques 4](#_Toc12012882)

[7. Concepts d’Excel 7](#_Toc12012883)

[7.1. Lignes, colonnes, cellules 7](#_Toc12012884)

[7.2. Contenu d'une cellule 8](#_Toc12012885)

[7.2.1. Donnée 8](#_Toc12012886)

[7.2.2. Format d'affichage 9](#_Toc12012887)

[7.2.3. Style d'affichage 9](#_Toc12012888)

[7.2.4. Qu'en retenir 9](#_Toc12012889)

[7.3. Formule 9](#_Toc12012890)

[8. Exemple d’usage des formules 11](#_Toc12012891)

[9. Composition des formules 15](#_Toc12012892)

[9.1. Opérateurs de base 15](#_Toc12012893)

[9.2. Parenthèses 16](#_Toc12012894)

[9.3. Fonctions statistiques 16](#_Toc12012895)

[9.4. Conditions et expressions logiques 16](#_Toc12012973)

[10. Calculs statistiques de base 17](#_Toc12012974)

[10.1. Calculer un pourcentage 17](#_Toc12012975)

[10.1.1. Principe général 17](#_Toc12012976)

[10.1.2. Pourcentages « en ligne » ou « en colonne » 17](#_Toc12012977)

[10.1.3. Cas particulier des études « cas-témoin » ou « exposé / non-exposé » 18](#_Toc12012978)

[10.1.4. Gestion des données manquantes 18](#_Toc12012979)

[10.1.5. Problème du temps 19](#_Toc12012980)

[10.1.6. Problème de l’incertitude 19](#_Toc12012981)

[10.1.7. Problème de dépendance entre les données 20](#_Toc12012982)

[10.1.8. Calcul pratique sous Excel 20](#_Toc12012983)

[10.2. Calculer une moyenne et un écart-type 25](#_Toc12012984)

[11. Graphiques 26](#_Toc12012985)

[11.1. Graphique avec une seule série de données 26](#_Toc12012986)

[11.2. Graphique avec plusieurs séries de données 31](#_Toc12012987)

[12. Tableaux croisés dynamiques 36](#_Toc12012988)

[13. Quelques raccourcis pour se faciliter la vie 39](#_Toc12012989)

[14. Le bug des lignes vides d’Excel 40](#_Toc12012990)

[14.1. Symptômes 40](#_Toc12012991)

[14.2. Historique du bug 40](#_Toc12012992)

[14.3. Comment reproduire le bug 40](#_Toc12012993)

[14.4. Comment identifier le bug 42](#_Toc12012994)

[14.5. Comment corriger le bug 42](#_Toc12012995)

[14.6. Autre solution 45](#_Toc12012996)

[15. En bref 46](#_Toc12012997)

# À propos

Ce cours concernant le tableur Microsoft® Excel® a été élaboré par le Dr André GILLIBERT, assistant spécialiste dans l’unité de biostatistique, au sein du département d’appui à la recherche clinique (DARC) dans la maison de la recherche au CHU de Rouen, Seine-Maritime, Normandie, France.

Il a été conçu dans le cadre de la mission d’enseignant vacataire universitaire, comme support à la séance de travaux pratiques de Microsoft® Excel® à destination des étudiants de la faculté de médecine et pharmacie de l’université de Rouen Normandie, puis a été modifié afin d’être communiqué aux internes en médecine et pharmacie.

Version 2.1

Création le 03/02/2018

Dernière mise à jour 21/06/2019 (ajout des sections de calculs d’un pourcentage et d’une moyenne)

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

# Introduction

Ce petit document résume les concepts de base utiles à l’usage d’Excel dans le contexte de la génération et l’exploitation de base de données en rapport avec de petites études médicales.

L’aspect cosmétique d’Excel est passé vite car il s’apprend assez facilement par l’usage ; alors que certains concepts structurants ne sont pas évidents à découvrir par soi même.

Excel bénéficie d’une documentation très extensive sur le Web. N’hésitez pas à chercher sur le Web (Google Search, Microsoft Bing, Qwant) ainsi que l’aide en ligne (F1 sous Excel).

# Vocabulaire

Colonne, ligne, cellule

Type de données, format d’affichage, style d’affichage

Types numérique, logique et texte

Formule

Référence, plage de cellules, ancre, référence relative, référence absolue

Données agrégées

Effectifs (déjà vu en cours)

Opérateurs

Chaîne de caractères (synonyme texte)

Littéral

Figure/graphique

Série de données

Source de données

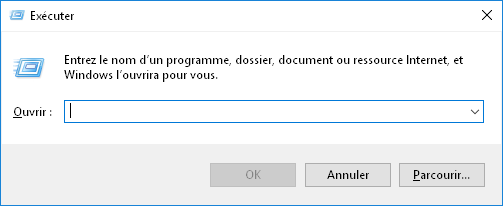
Étiquettes/libellés

# Lancer Excel

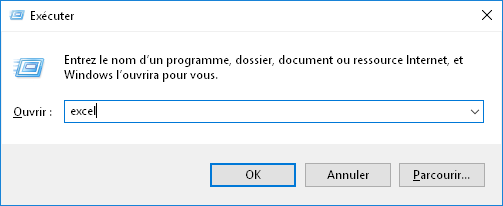
Le menu démarrer change à chaque version de Windows. Les icônes sont tout le temps déplacées. On galère donc un peu à trouver les logiciels.

Pourtant, sur toutes les versions de Windows depuis 1995, on peut toujours lancer Excel, s’il est installé, avec quelques touches magiques.

Tapez [WINDOWS]+[R]. Cela va ouvrir une petite boîte



Tapez maintenant excel



Puis [ENTRÉE]

Il existe un nom pour de nombreux programmes Windows

**excel** pour Microsoft Excel

**winword** pour Microsoft Word

**powerpnt** pour Microsoft PowerPoint

**outlook** pour Microsoft Outlook

**notepad** pour Notepad

**mspaint** pour Microsoft Paint

**control** pour ouvrir le panneau de configuration

**C:** pour ouvrir le lecteur C, généralement le disque dur principal

# Saisie de données

On peut cliquer sur une cellule pour se positionner dessus. On peut aussi utiliser les touches fléchées pour se déplacer.

Lorsqu’on est positionné sur une cellule, il suffit de taper des chiffres pour écrire un nombre ou des lettres pour écrire du texte… Ensuite, on peut valider le contenu de la cellule avec la touche [ENTRÉE] ou la touche [TAB].

[TAB] passe à la colonne suivante

[ENTRÉE] passe à la ligne suivante

# Bonnes pratiques

Un tableur Excel est constitué de cellules dans un très grand espace de lignes et de colonnes.

Les bonnes pratiques de constitution d'une base de données :

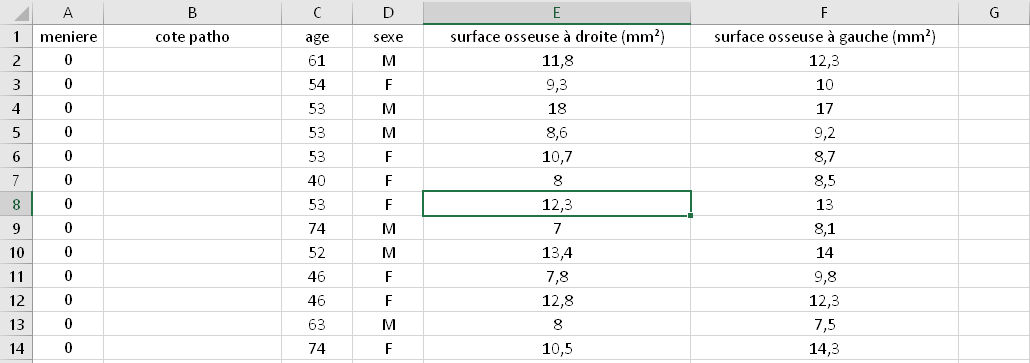
1. Une colonne par variable aléatoire
2. Un codage cohérent de toute une colonne. Par exemple, le sexe, en tant que variable qualitative à 2 modalités, devrait toujours être codé M pour les hommes et F pour les femmes plutôt que H, M et « homme » pour les hommes
3. Une ligne par unité statistique
4. La 1ère ligne réservée aux intitulés de colonnes
5. Ne jamais avoir deux colonnes avec le même intitulé
6. Éviter les colonnes sans intitulé

Par ailleurs, s’y ajoutent les règles déontologiques :

1. Jamais de nom de patients ou données ré-identifiantes (comme le numéro de sécurité sociale) sur clé USB ou par e-mail (réseaux non sécurisés)
2. Conserver un fichier de lien entre les données identifiantes (p.e. numéro CPAGE) et un numéro d’anonymat, sur une machine sécurisée par mot de passe.
3. Seul celui qui a besoin d’accéder au dossier médical a besoin des données identifiantes. La communication avec les autres parties, tel que le biostatisticien, doit se faire avec le numéro d’anonymat.
4. Les fichiers, mêmes anonymisés peuvent parfois être indirectement ré-identifiant (p.e. date de naissance + date d’hospitalisation + comorbidités permettent de retrouver quelqu’un qu’on connaît déjà un peu). C’est pourquoi il faut en plus, crypter les données dès qu’elles transitent sur des réseaux fragiles (clé USB, e-mail). Le système de protection par mot de passe de Microsoft Excel ne vaut rien (il est complètement bidon). Il faut utiliser 7-ZIP avec AES-256 et un mot de passe fort obtenu par tirage au sort de 4 ou 5 mots du dictionnaire.

Voir <https://xkcd.com/936/>

Ci-dessous, une base de données plutôt propre :



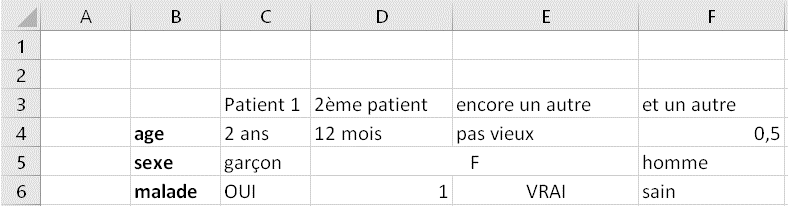
Colonne de titre (conseillé)

Codage cohérent

Une ligne, une unité statistique (ici le patient)

Mais sachez qu’Excel est un bac à sable. Tout est possible. On peut rentrer n’importe quelle donnée dans n’importe quelle cellule et même fusionner des cellules, ce qui va pourtant à l’encontre du principe de données tabulaires.

Ci-dessous une mauvaise base de données (Excel vous laisse faire ça sans sourciller !)



Variables présentées en ligne

plutôt qu’en colonnes

Cellules fusionnées !

Convention de codage différente pour chaque patient

Aucune base n’est parfaite, même la base d’exemple a des défauts. Il faut éviter de renseigner l’âge du patient. Il vaut mieux renseigner sa **date de naissance** et la **date de l’examen**, ce qui permet de recalculer facilement l’âge par différence !

D’une manière générale, il faut éviter de renseigner des données calculées. Il faut renseigner les données brutes. Par exemple, si vous voulez renseigner une échelle constituée de 15 items, plutôt que de calculer le total, il vaut mieux renseigner pour chaque patient, les items individuels, avec 15 colonnes par patient ! Cela évitera certaines erreurs et permettra des analyses beaucoup plus fines, notamment sur la structure des items.

# Concepts d’Excel

## Lignes, colonnes, cellules

Excel est organisé comme un immense tableau avec un nombre presque illimité de lignes et de colonnes.

Les colonnes sont décrites par des lettres A, B, C, D, …, Z, AA, AB, AC, …, AZ, BA, BB, BC, …

Les **lignes** sont numérotées : 1, 2, 3, 4, …

Le nombre de **lignes** maximal est 220 soit 1048576 (Format xlsx)

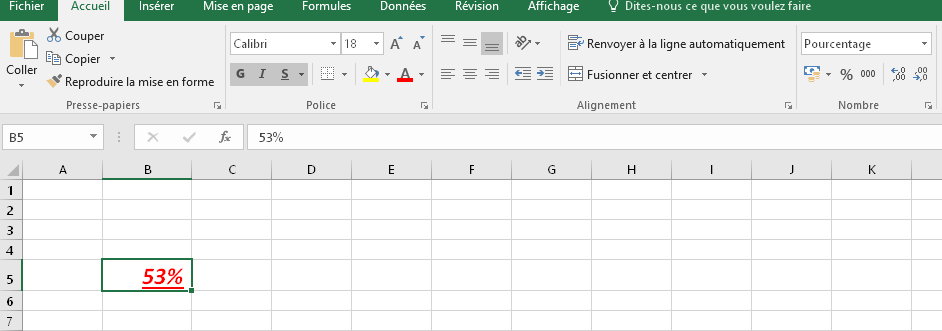
Le nombre de **colonnes** maximal est 214 soit 16384 (Format xlsx)

On atteint très rarement ces limites !

À l'intersection d'une ligne et d'une colonne, on trouve une **cellule**. Toutes les cellules sont identifiées par les lettres de la colonne suivies des chiffres de la ligne. Par exemple, la cellule C5 est à l'intersection de la colonne C et de la ligne 5.

## Contenu d'une cellule

Chaque cellule contient une **donnée ou une formule**, un **format d'affichage** et un **style d'affichage**.



Donnée numérique : 0,53

Format d’affichage : pourcentage avec zéro

chiffre après la virgule

Style d’affichage :

Gras, italique, souligné

Couleur rouge

Police Calibri 18

### Donnée

**Type de données** :

* Numérique (p.e. 3,141592653)
* Logique (VRAI ou FAUX)
* Texte (synonyme : chaîne de caractères)
* Vide (cellule vide)

Il faut rajouter un pseudo-type de données : les dates.

Excel ne distingue pas vraiment les dates des nombres, puisque les dates sont gérées comme le nombre de jours écoulés depuis le 30/12/1899 et ont juste un **format d’affichage** particulier.

On peut alors additionner et multiplier une date ! On peut même prendre la racine carrée d’une date !

Chaque cellule peut a son propre **type de données**. Il est vivement recommandé mais pas obligatoire, de renseigner un **type de donnée** homogène sur une colonne.

#### Type numérique

Le type numérique peut contenir un nombre à virgule (d’une précision d’environ 15 chiffres numériquement significatifs).

#### Type logique

Le type logique peut seulement contenir VRAI ou FAUX

#### Type texte

Le type texte peut contenir une succession de chiffres, lettres, espaces, passages à la ligne, et caractères spéciaux. On parle aussi de **chaîne de caractères** ou, en anglais de **character string** ou, de **string** tout court.

Un caractère, est un symbole élémentaire de l’écriture. Beaucoup de caractères sont directement accessibles par une touche du clavier ou une combinaison de touches.

**ATTENTION** : Excel distingue le nombre 5,43 du texte 5,43 qui lui ressemble énormément, à ceci près que le style d'affichage diffère légèrement par défaut. Les nombres sont centrés alors que les textes sont alignées à gauche. Néanmoins, si on centre un texte, il devient visuellement indifférenciable.

### Format d'affichage

Le format d'affichage n'affecte pas la valeur de la cellule proprement dite mais détermine la représentation. C'est surtout utile pour les valeurs numériques !

Par exemple, on peut demander à Excel de représenter le nombre 0,784621 comme un pourcentage avec un chiffre après la virgule. Dans ce cas il affichera 78,5%

Les formats "date" sont ceux qui permettent de présenter un nombre de jours depuis le 30 décembre 1899 comme une date.

### Style d'affichage

Purement cosmétique, il s'agit de couleur, police, gras, italique, souligné, bordures, etc.

Il ne faut jamais stocker une variable aléatoire dans le style d'affichage ! Si vous souhaitez représenter les malades en rouge et les sujets sains en vert, utilisez plutôt une colonne "malade" contenant l'information VRAI/FAUX. La coloration conditionnelle permet de transformer les données des cellules en couleur, en garantissant la cohérence des résultats.

### Qu'en retenir

Les **données** des **cellules** servent à stocker les réalisations des variables aléatoires. Utilisez des nombres pour les variables quantitatives, des valeurs logiques VRAI/FAUX (ou des nombres 0 et 1) pour les variables binaires, et du texte pour les variables qualitatives.

## Formule

Les formules sont extrêmement utiles dans Excel et il ne faut surtout pas hésiter à les utiliser ! Il faut apprendre à les apprivoiser parce qu’elles peuvent vite être complexes.

Certaines cellules, plutôt que de contenir une donnée brute, contiennent **une formule**. Cette formule récupère les résultats à partir d'une ou plusieurs autres cellules et les synthétise sous forme d'une unique valeur. Excel mémorise la formule mais affiche le résultat du calcul de la formule. Il détecte automatiquement les modifications concernant les dépendances de la formule et mets à jour le résultat immédiatement à la moindre modification. Une formule peut dépendre de cellules contenant des données brutes mais peut aussi dépendre de cellules contenant des formules.

Ci-dessous, l'anatomie d'une formule simple :

=SOMME(A2:B5)

Signe égal

Fonction

Référence à une **plage de cellules**

Dans une formule, on peut faire référence à une seule cellule ou à une **plage** de cellules entière.

Une plage de cellules est un rectangle obtenu en décrivant les coordonnées de la 1ère cellule (en haut à gauche du rectangle) et la dernière cellule (en bas à droite du rectangle) séparées par un symbole deux-points.

Ainsi A2:B5 décrit les cellules A2, A3, A4, A5, B2, B3, B4, B5.

On peut faire référence à une colonne toute entière en omettant le numéro de ligne. Par exemple B:B fait référence à toutes les cellules de la colonne B à la fois.

Lorsqu'on recopie une formule d'une cellule à la suivante, les références sont mises à jour afin de maintenir la position relative des cellules.

Par exemple, si la cellule C3 contient la formule =SOMME(A2:B5) et qu'on recopie cette cellule C3 en C4, la formule est changée en =SOMME(A3:B6). En effet, on a tout décalé d'une ligne vers le bas ! De même, si on recopie la formule de la cellule C3 vers la cellule D3, la formule sera changée en =SOMME(B2:C5)

Parfois il est souhaitable de "bloquer" une dimension afin qu'elle ne soit plus changée par une recopie. On parle d’ancrage.

Dans ce cas, il faut préfixer la dimension en question du symbole dollar.

Par exemple, si la cellule C3 contient la formule =2\*A2 et qu'on recopie cette formule en D4, elle va être transformée en =2\*B3. Par contre, si C3 contient =2\*$A2 et qu'on la recopie en D4, elle va devenir =2\*$A3

On peut bloquer l'une, l'autre ou les deux dimensions à la fois:

A2 : colonne et ligne relatives (déplacées en cas de recopie de formule)

$A2 : colonne ancrée, ligne relative (colonne inchangée en cas de recopie mais ligne déplacée)

A$2 : colonne relative, ligne ancrée

$A$2 : colonne et ligne ancrées

Plutôt que parler d’ancre, on peut aussi parler de référence absolue (synonyme).

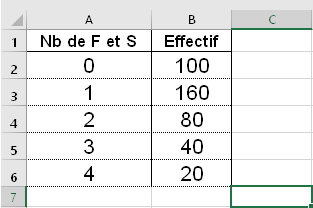
Les formules peuvent contenir les opérateurs habituels : addition (+), soustraction (-), multiplication (\*), division (/), mais aussi la concaténation de chaînes de caractères (&), c’est-à-dire, mettre le texte bout à bout, de telle sorte que 1 & 2 vaut "12"

Les formules peuvent contenir des expressions composées, exploitant de très nombreuses fonctions dont la liste est disponible dans l’onglet « Formules ». L’option « Formules -> Insérer une fonction » peut aider à paramétrer les fonctions.

Il est parfois plus facile de créer des expressions complexes en plusieurs étapes (colonnes intermédiaires).

# Exemple d’usage des formules

basons nous sur un fichier de **données agrégées** de l’échantillon de 400 étudiants. C’est-à-dire, plutôt que d’avoir les données brutes, avec 400 lignes de données, nous n’avons qu’une ligne de titre et 5 lignes de données. Ce sont les effectifs pour chaque valeur du support.



Dans les cellules A2 à A6, se trouvent le nombre de frères et sœurs (valeurs du support de la variable quantitative discrète). Dans les cellules B2 à B6 se trouvent les effectifs correspondant.

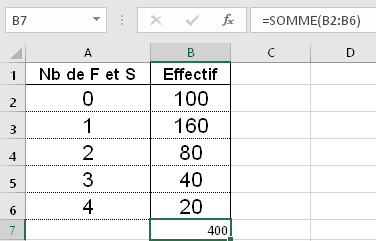
Commencez par recopier ce tableau dans Excel (ça fait un petit exercice de saisie de données). Utilisez [TAB] et [ENTREE] pour vous déplacer.

Ensuite, vous aller calculer l’effectif total par une formule.

Pour ce faire, cliquez sur la cellule B7, puis tapez sur le signe égal pour débuter la formule suivante :

=SOMME(B2:B6)

Une fois la formule tapée, appuyez sur [ENTRÉE]



Résultat du calcul de la formule

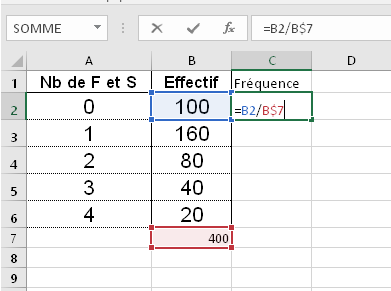
Formule

Vous constaterez que le résultat s’affiche immédiatement, mais dans la barre des formules, la formule =SOMME(B2:B6) apparaît, indiquant qu’Excel n’a pas remplacé la formule par son résultat, mais qu’il a conservé précieusement la formule.

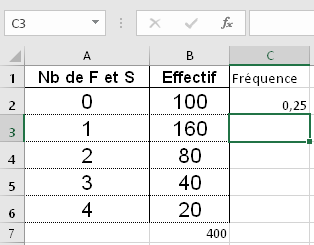
Placez vous ensuite dans la cellule B3 et tapez 180[ENTRÉE]. Vous allez voir le résultat en B7 se mettre à jour immédiatement. Appuyez [CTRL]+[Z] pour annuler cette opération.

Maintenant, placez vous en C1 et tapez Fréquence[ENTRÉE] afin de ne pas laisser une colonne sans intitulé. Excel devrait automatiquement vous positionner en C2 lorsque vous tapez [ENTRÉE]

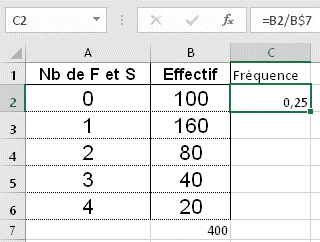
En C2, tapez la formule =B2/B$7



Après avoir appuyé [ENTRÉE], vous allez voir apparaître 0,25 (le résultat du calcul)

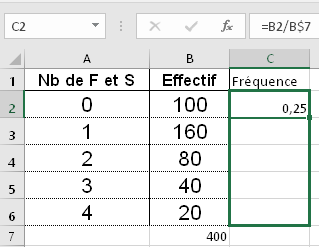


Maintenant, repositionnez vous sur C2 puis cliquez sur le tout petit carré, en bas à droite du rectangle de sélection, puis glissez la sélection jusqu’en C6.

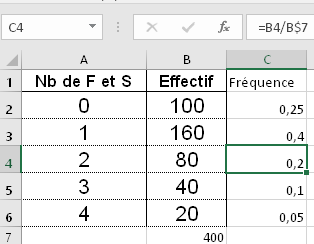


Enfoncez le bouton gauche de la souris sur le tout petit carré vert

(le curseur de la souris doit prendre la forme d’une croix)



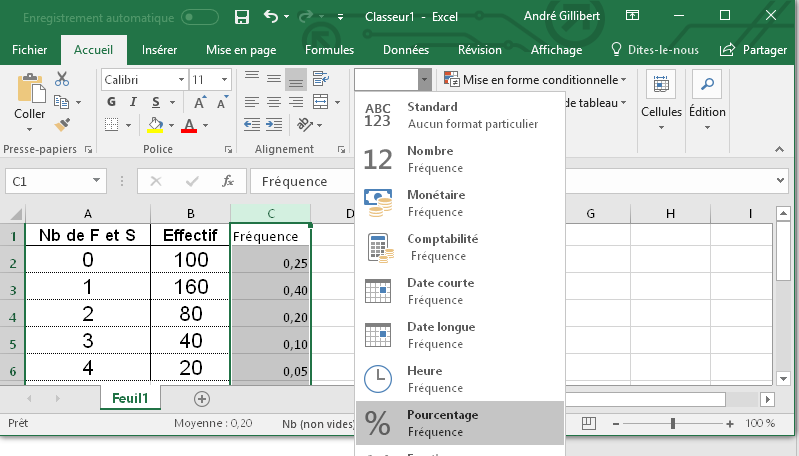
Glissez (en gardant enfoncé le bouton gauche de la souris) jusqu’en C6, puis relâchez le bouton de la souris.



La formule a été recopiée de la cellcule C2 dans la cellule C4 mais la référence a été modifiée B2 est devenu B4 par contre B$7 est resté B$7

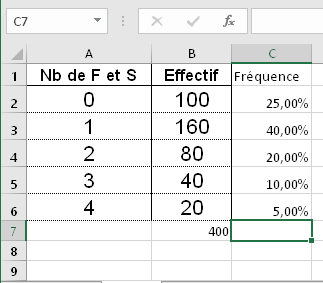
La formule a été recopiée de C2 à C4 avec modification de la 1ère référence (numérateur de la fraction) mais pas de la 2ème référence (le total général au dénominateur) grace à l’ancre indiquée par le symbole dollar.

Sélectionnez toute la colonne C en cliquant sur le C, juste au-dessus de la cellule C1, puis changez le format en « pourcentage »



Cliquez sur la colonne pour la sélectionner

Dans l’onglet Accueil, déroulez le menu des formats et cliquez sur le format « pourcentage »



Vous devez obtenir les pourcentages pour chaque valeur du support (de 0 à 4).

# Composition des formules

## Opérateurs de base

On parle de **valeur littérale** ou plus simplement de **littéral** pour décrire une valeur

Les valeurs littérales sont des nombres comme 5,3, des valeurs logique comme VRAI ou FAUX ou des chaînes de caractères qu’il faut clairement délimiter par des guillements droits comme "bonjour le monde !"

Les opérateurs suivants permettent de faire les opérations des calculatrices de base :

+ pour l’addition

- pour la soustraction

\* pour la multiplication

/ pour la division

^ pour l’exponentiation. Cela se lit « puissance ». Par exemple 10^4 est égal à 10000

= pour l’égalité (renvoie une valeur logique)

> pour la supériorité stricte (renvoie une valeur logique)

< pour l’infériorité strictre

>= pour la supériorité large (supérieur ou égal)

<= pour l’infériorité large

L’opérateur & sert à la concaténation de chaînes de caractères. S’il y a des nombres

Ainsi "bonjour " & "le monde" fait "bonjour le monde"

Les nombres sont automatiquement convertis en chaînes de caractères lorsqu’on utilise cet opérateur de telle sorte que 1 & 23 est égal à "123".

## Parenthèses

Les opérateurs et parenthèses suivent les mêmes règles de priorité qu’en mathématique :

2\*(2+3) est égal à 2\*5 soit 10

Les parenthèses peuvent être imbriquées comme dans l’expression ((1+2)\*(1+2)+1)\*2 qui se réduit à (3\*3+1)\*2 soit 10\*2 soit 20.

## Fonctions statistiques

Les fonctions agrégatrices comme SOMME() fonctionnent avec toute une plage de données.

Les fonctions suivantes sont très utiles en statistique :

MEDIANE, MOYENNE, MOYENNE.SI, SOMME, ECARTYPE (synynoyme ECARTYPE.STANDARD), QUARTILE, CENTILE

**Attention** : ces fonctions sont conçues pour fonctionner sur des données brutes… Sur des données agrégées, elles peuvent ne pas avoir de sens.

La fonction CENTILE ne calcule pas vraiment les centiles, mais plutôt tout **quantile** arbitraire !

Par exemple, sur les données agrégées de nombre de frères et sœurs présentées précédemment, la somme des effectifs a du sens, mais la moyenne, l’écart type ou les centiles n’en ont pas.

Les fonctions QUARTILE et CENTILE prennent deux paramètres, le second devant être séparé du premier par le symbole point-virgule.

=CENTILE(A2:A200; 0,10) pour le 1er décile de la colonne A dans un jeu de données de 200 observations.

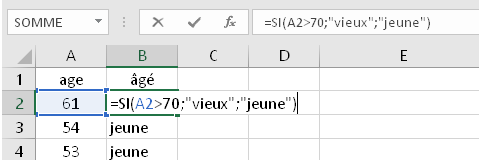
## Conditions et expressions logiques

Les fonctions suivantes permettent de faire des choix :

=SI(expression1 ; expression2, expression3)

La fonction SI va évaluer l’expression1, puis, si elle est vraie, renvoyer le résultat d’expression2, sinon elle va renvoyer le résultat d’expression3.

Exemple : séparer les sujets en deux groupes d’âge



Si on veut séparer en trois catégories, on peut imbriquer les SI

=SI(A2>70 ; "vieux" ; SI(A2>50 ; "intermédiaire" ; "jeune"))

On peut aussi s’aider des opérateurs ET et OU.

=SI(ET(A2>50 ; A2<=70) ; "intermédiaire" ; "jeune ou vieux")

ET(expression1 ; expression2 ; …)

Renvoie VRAI seulement dans la condition ou toutes les expressions sont simultanément vraies ; sinon il renvoie FAUX.

OU(expression1 ; expression2 ; …)

Renvoie VRAI dès qu’au moins une des expressions est vraie et renvoie FAUX lorsque toutes les expressions ton fausses.

On peut noter que la conversion de type logique -> numérique transforme VRAI en 1 et FAUX en 0

Ainsi :

=2\*(A2>50) va renvoyer 2 pour les sujets de plus de 50 ans et 0 pour les sujets de moins de 50 ans.

# Calculs statistiques de base

## Calculer un pourcentage

### Principe général

Le calcul d’un pourcentage nécessite de connaître :

1. Le numérateur (*p.e.* nombre de sujets malades)
2. Le dénominateur (*p.e.* nombre de sujets susceptibles d’être malade)

Le calcul d’un pourcentage nécessite plusieurs étapes de réflexion avant de se lancer dans le calcul. C’est pourquoi, nous vous invitons à lire les sections suivantes dans l’ordre.

### Pourcentages « en ligne » ou « en colonne »

Posez-vous aussi la question du sens dans lequel vous calculez le pourcentage. Par exemple, on peut calculer le pourcentage de diabétiques parmi les obèses ou le pourcentage d’obèses parmi les diabétiques. Si on s’intéresse à l’obésité en tant que facteur de risque du diabète, il serait logique de présenter le pourcentage de diabétiques parmi les obèses alors que si on s’intéresse au traitement de l’obésité chez les patients diabétiques, le second pourcentage pourra être pertinent.

D’une manière générale, le pourcentage ayant un événement de santé négatif parmi les sujets exposés, est plus pertinent que le pourcentage de sujets exposés parmi les sujets ayant l’événement de santé négatif, mais cette règle n’est pas absolue.

### Cas particulier des études « cas-témoin » ou « exposé / non-exposé »

Le calcul de certains pourcentages peut être parfois « interdit » parce que la méthodologie de l’étude a créé une situation dans laquelle la prévalence de la maladie ou de l’exposition est artificielle.

Par exemple, si on construit une étude cas-témoin avec 50% des sujets inclus qui seront diabétiques (cas) et 50% des sujets inclus qui ne seront pas diabétiques, il sera impossible de calculer un pourcentage de sujets diabétiques.

Il sera alors invalide de calculer le pourcentage global de sujets diabétiques, automatiquement égal à 50%, mais il sera tout aussi invalide de calculer le pourcentage de sujets diabétiques parmi les obèses même si le logiciel fournit un résultat tel que 70%, ou le pourcentage de sujets diabétiques parmi les non obèses, même si le logiciel fournit un résultat tel que 40%.

En effet, la véritable prévalence du diabète est de quelques pourcents. La méthodologie cas-témoin augmente artificiellement la prévalence de la maladie.

Ce même principe s’applique à une étude dans laquelle on apparie 2, 3 ou 4 témoins pour chaque cas, ou dans les études dans lesquelles on va artificiellement créer un groupe de sujets exposés (p.e. obèses) et non exposés (p.e. non obèses).

### Gestion des données manquantes

Parfois la donnée est inconnue pour certains patients. Doit-on exclure ces sujets ou considérer qu’ils ne sont pas malades ?

Il n’y a pas de règle générale. L’exclusion de ces sujets, n’apparaissant alors ni au numérateur, ni au dénominateur, fait l’hypothèse forte qu’ils ont le même risque d’avoir la maladie que les autres. Cela est vrai lorsque les données sont manquantes de manière totalement indépendante de la maladie.

S’il y a beaucoup de données manquantes et que les données ne sont pas manquantes au hasard, la variable est inexploitable. Un bon exemple est l’indice de masse corporelle inscrit dans les dossiers médicaux. Si cette valeur est renseignée pour 30% des dossiers, on peut imaginer qu’il est surtout renseigné lorsqu’il est anormalement bas (maigreur) ou élevé (surpoids ou obésité). Cela fausse alors complètement les résultats. Cette variable ne doit donc pas être analysée du tout. Elle est irrécupérable.

On ne peut pas se passer de connaître les conditions de genèse de la variable afin de savoir pourquoi elle est manquante ou ne l’est pas. Certaines méthodes statistiques permettent d’appréhender partiellement ce problème, mais elles sont complexes, nécessitent de grandes bases de données, et ne peuvent pas être réalisées sous Excel.

### Problème du temps

Certains pourcentages n’ont pas de sens si certaines informations contextuelles ne sont pas présentées.

Par exemple, la phrase « 20% des patients suivis après infarctus du myocarde sont décédés » semble anodine, mais ne veut pas dire grand-chose. Cette phrase implique-t-elle que les 80 autres pourcents de patients sont immortels ? 100% des patients finissent par décéder tôt ou tard, mais ce qui importe, c’est le délai avant décès.

Ainsi, en post-infarctus, on peut calculer le pourcentage de patients survivants à un horizon temporel bien défini. Par exemple, on peut calculer le pourcentage de survie à 1 an. Ce calcul de pourcentage par contre, ne peut pas être aisément fait sous Excel si certains patients sont perdus de vue. Il se base sur l’estimateur de Kaplan-Meier et nécessite de bien connaître le statut vital aux dernières nouvelles et de connaître à la fois la date de diagnostic à la date des dernières nouvelles.

De même, on ne peut pas dire 5% des obèses ont développé un diabète sans dire quel était l’âge initial de ces obèses ni combien de temps ils ont été suivis. Idéalement, cela se fait encore avec l’estimateur de Kaplan-Meier, inaccessible sous Excel.

### Problème de l’incertitude

Même si on parle d’analyse descriptive pour un pourcentage, on veut toujours généraliser ce pourcentage à une population complète. Par exemple, si on calcule le taux de complications graves d’une opération chirurgicale en 2018 à Rouen, et que l’on trouve 2 complications graves pour 100 patients, on peut dire qu’on a observé 2% de complications, mais c’est surtout le taux de complications sur les opérations futures à Rouen, voire futures en dehors de Rouen, qui nous intéresse.

Le problème, c’est que ce chiffre de 2% est erroné. En effet, le vrai pourcentage peut très bien être nettement plus grand (p.e. 6%) et avoir été associé à de la « chance » pour ces 100 patients, ou au contraire le vrai pourcentage peut être nettement plus faible (p.e. 3 pour mille) et avoir été associé à de la « malchance » pour ces 100 patients.

On ne peut donc apprécier un pourcentage que si on en connaît l’incertitude. Cette incertitude dépend beaucoup plus du numérateur (nombre d’événements) que du dénominateur (nombre de sujets). Lorsqu’on est proche de 100%, il ne faut plus s’intéresser aux événements mais aux non-événements. Ainsi, c’est la même chose de dire que 2 patients sur 100 ont eu une complication que de dire que 98 patients sur 100 n’ont pas eu de complication, même si la seconde formulation peut faire croire qu’on a une grande quantité de données.

Voici une règle simplifiée :

* En dessous de cinq événements (ou non-événements), les pourcentages sont bidon. Par exemple, si on observe 2 complications pour 10 opérations, on peut dire que le pourcentage de complications est compris quelque part entre 2,5% et 56% (intervalle de confiance à 95%). Bref, on n’a aucune idée du vrai pourcentage ;
* Autour d’une dizaine d’événements, on a un ordre de grandeur à un facteur 3 à 4 près, c'est-à-dire très imprécis, du pourcentage ;
* C’est à partir de 25 ou 30 événements qu’on atteint une précision à un facteur 2 près et qu’on peut commencer à discuter des résultats.

On peut calculer des intervalles de confiance d’un pourcentage avec le site Web suivant :

<http://andre.gillibert.fr/stats/cpmidp.html>

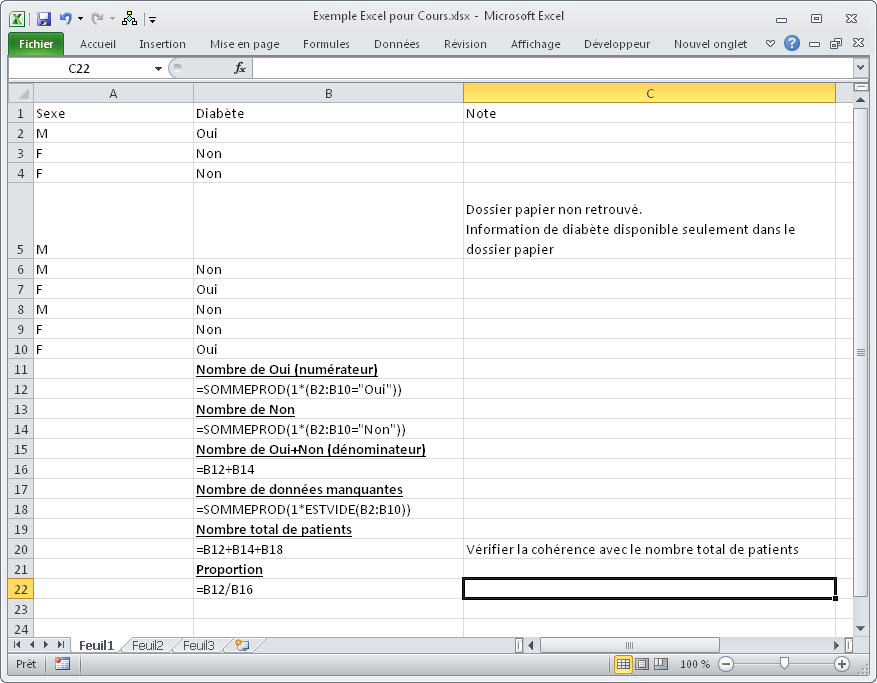
### Problème de dépendance entre les données

Les calculs d’incertitude présentés dans la section précédente sont basés sur l’hypothèse que les observations sont indépendantes. Si vous analysez la cicatrisation des greffes de peau et que certains patients ont plusieurs greffes, vous ne pouvez pas aisément calculer l’intervalle de confiance du pourcentage de greffe qui cicatrisent en moins de 3 mois car on peut imaginer que les deux greffes d’un même patient auront tendance à tous deux cicatriser en même temps (p.e. patient jeune, avec greffes post-traumatiques) ou à ne pas cicatriser en même temps (p.e. patient présentant des ulcères variqueux + une AOMI avec des greffes de peaux aux membres inférieurs). Les méthodes statistiques permettant ces calculs ne sont pas simples.

### Calcul pratique sous Excel

Comme vous aurez compris, le calcul d’un pourcentage nécessite un grand nombre de réflexions et vérifications préalables (voir sections précédentes). Sans ces étapes, la présentation et l’interprétation des résultats risquent fortement d’être erronées.

Ci-dessous une première manière de calculer les pourcentages en comptant les « Oui », les « Non », les données manquantes (cases vides), puis en vérifiant que le total est bien égal au nombre de patients. Enfin, on calcule la proportion comme (Nombre de Oui)/(Nombre de Non + Nombre de Oui)



Afin de compléter ce calcul, il faut rechercher l’intervalle de confiance du pourcentage en s’aidant de <http://andre.gillibert.fr/stats/cpmidp.html>

« Number of successes » représente le numérateur (nombre de Oui)

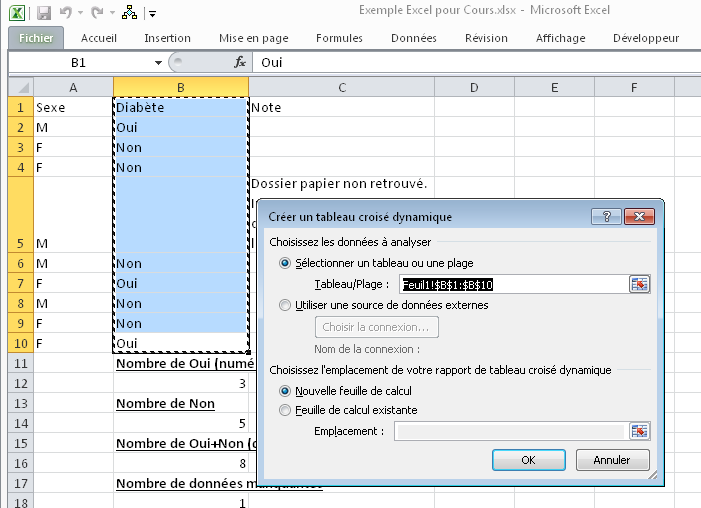
« Number of trials » représente le dénominateur (nombre de Oui + nombre de Non)

Selon les cas, il faut considérer les données manquantes comme des Non ou alors les supprimer à la fois du dénominateur et du numérateur (cf section **Erreur ! Source du renvoi introuvable.**).

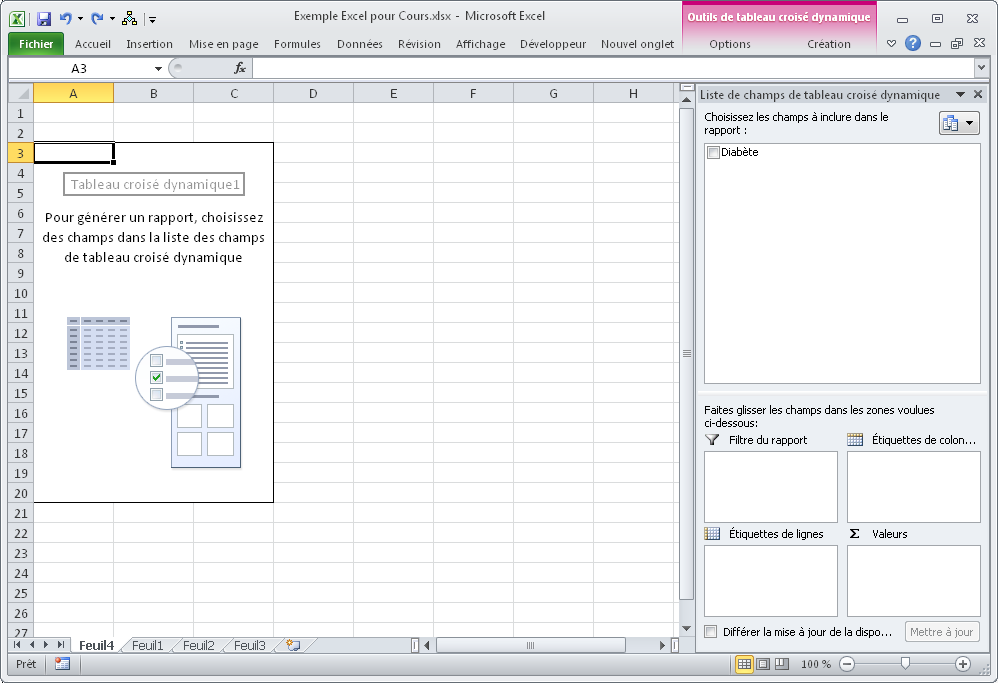
En réalité, le pourcentage calculé dans Excel n’est généralement pas pertinent. Seules les deux bornes de l’intervalle de confiance fourni par le site Web <http://andre.gillibert.fr/stats/cpmidp.html> fournissent le résultat interprétable. Le vrai pourcentage est quelque part entre les deux bornes. Il n’est généralement pas égal au pourcentage observé, tel que calculé sous Excel.

Une autre manière de calculer le pourcentage, sous Excel, c’est d’utiliser un tableau croisé dynamique.

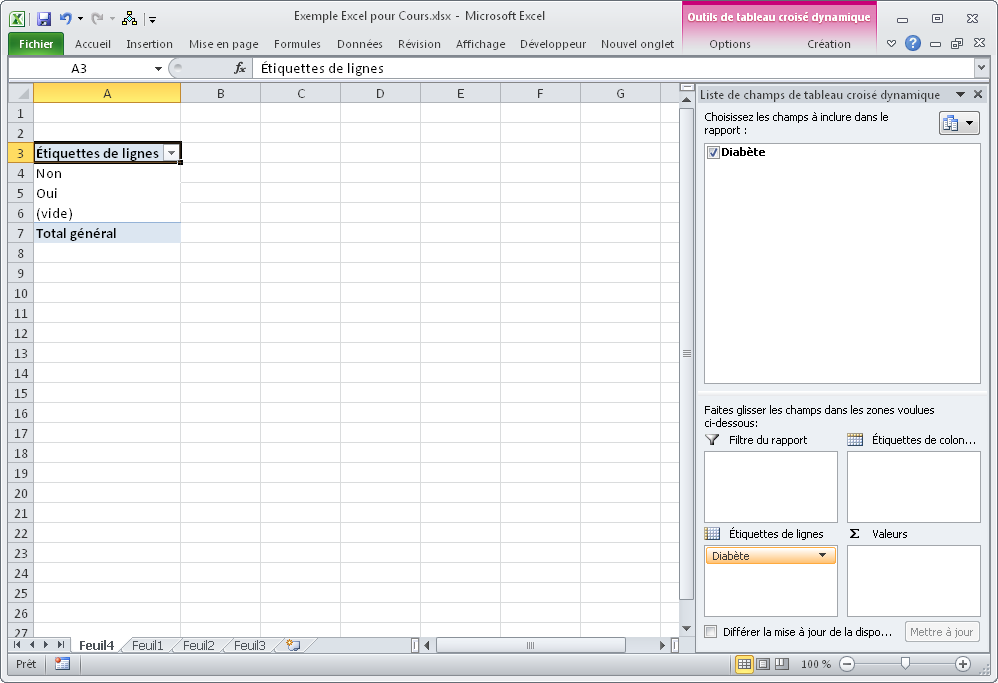
Sélectionnez la zone de données de la colonne vous intéressant, puis cliquez dans l’onglet « Insertion » puis sur « Tableau croisé dynamique », puis cliquez sur [OK].



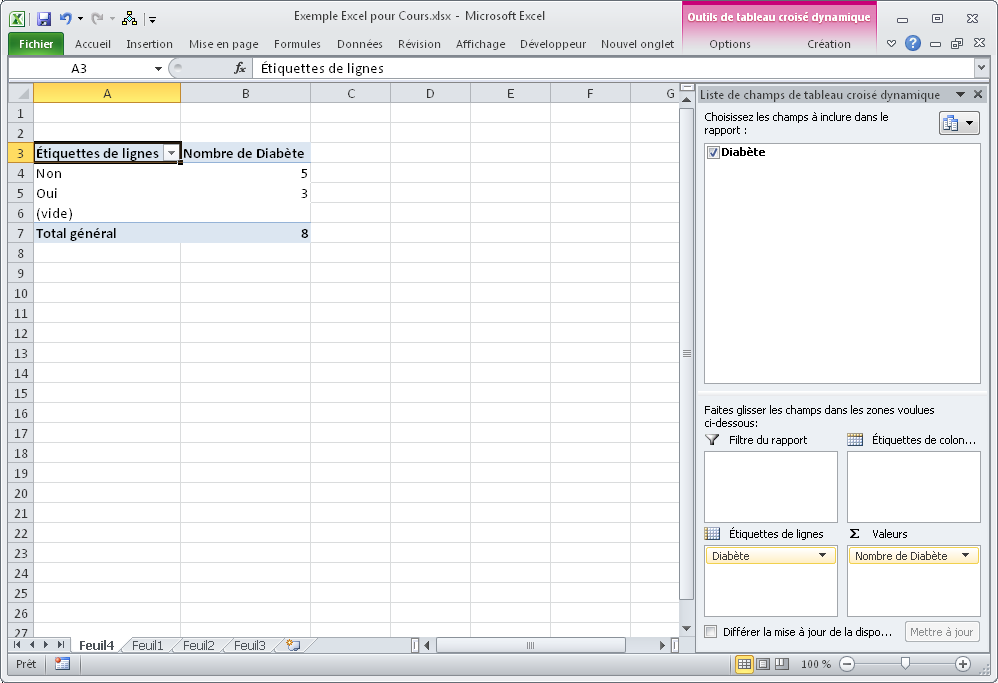
Vous devriez voir le constructeur de Tableau croisé dynamique apparaître :



Glissez-déposez le champ « Diabète » dans la boîte « Étiquettes de lignes ».

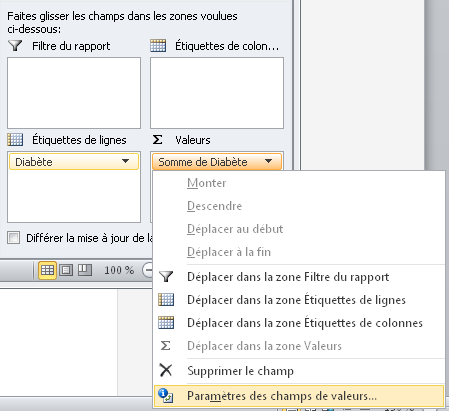


Puis glissez-déposez le même champ « Diabète » dans la boîte « Σ valeurs »

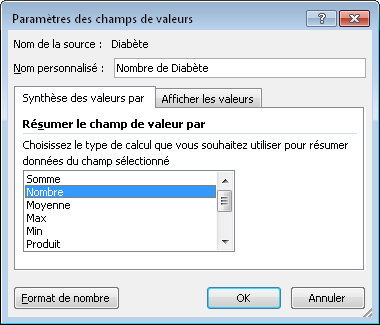


Vérifiez bien que la zone « Σ Valeurs » contient « Nombre de Diabète » et pas « Somme de Diabète ».

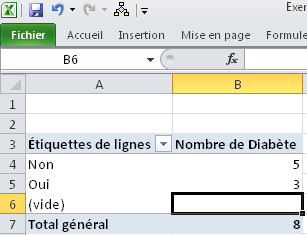
Si c’était « Somme de Diabète », cliquez (clic gauche) sur « Somme de Diabète » puis sur « Paramètres des champs de valeurs… »



Cliquez ensuite sur « Nombre » dans la boîte de dialogue qui apparaît, puis cliquez [OK] :

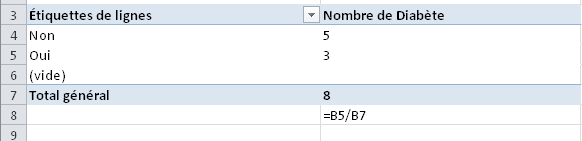


Vous constaterez que vous perdez l’information sur le nombre de cases vides (données manquantes) :



Si vous voulez éviter ce désagrément, faites un tableau croisé dynamique dans lequel vous incluez une variable sur laquelle il n’y a aucune donnée manquante dans la partie « Σ Valeurs », telle que l’identifiant patient.

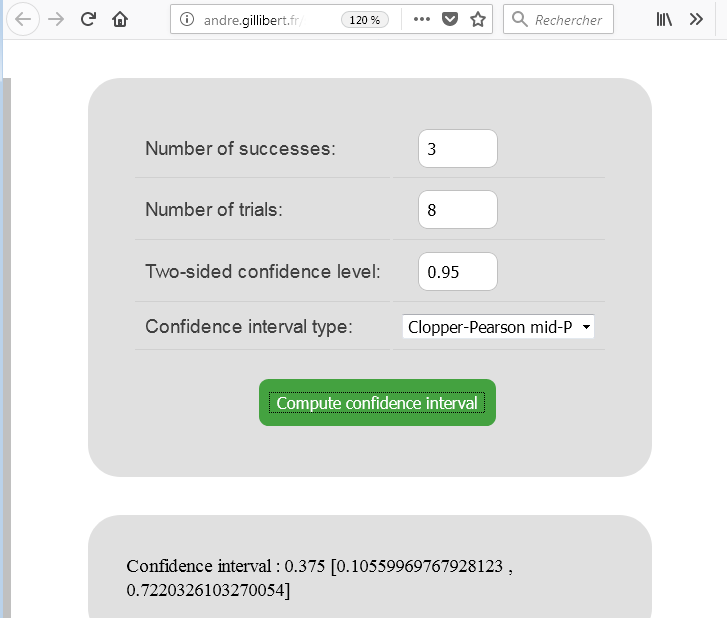
Ensuite, vous pouvez calculer le pourcentage en ajoutant la formule numérateur/dénominateur :



Vous verrez apparaître le pourcentage 37,5%

Ce pourcentage, est pourtant très probablement éloigné de la réalité…

Faites le calcul d’intervalle de confiance :



Et voilà, vous constatez que le vrai pourcentage est quelque part entre 11% et 72%.

## Calculer une moyenne et un écart-type

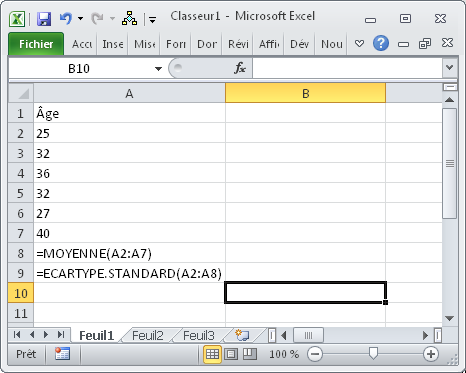
Le problème des données manquantes est globalement le même que pour les calculs de pourcentage (cf section ). Les méthodes de gestion possibles sont l’exclusion et l’imputation par la moyenne ou par la médiane. L’imputation consiste à « inventer » la donnée manquante en lui substituant une valeur.

Le calcul de la moyenne est insuffisant pour décrire une variable quantitative. Il faut au minimum y associer l’écart-type.

Certaines variables qui sont anormales lorsqu’elles sont basses aussi bien que lorsqu’elles sont élevées (p.e. indice de masse corporelle, hémoglobinémie) sont mal décrites par la moyenne et l’écart-type. Mieux vaut définir plusieurs catégories (p.e. IMC < 18.5, 18.5 – 25, 25 – 30, ≥ 30) et présenter les pourcentages de sujets dans chaque catégorie.

La moyenne se calcul sous Excel avec la fonction MOYENNE alors que l’écart-type se calcule avec ECARTTYPE.STANDARD.

Le calcul est très simple (cf ci-dessous) :



# Graphiques

La plupart des **graphiques** (synonyme **figure**) d’Excel sont conçus pour fonctionner sur des données agrégées telles que les effectifs par modalité ou les fréquences par modalité. Le nuage de points est l’exception (chaque point représentant une observation).

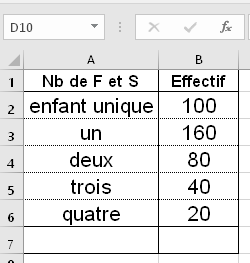
Les figures sont toujours basées sur une **source de données** définie par une **plage** contiguë ou non du classeur Excel. Cette plage permet de définir une ou plusieurs **séries de données** numériques qui sont alors présentées de diverses manières par Excel selon la représentation graphique choisie.

Les **séries de données** peuvent être en colonnes ou en lignes.

Ces concepts abstraits deviendront concrets avec un exemple.

## Graphique avec une seule série de données

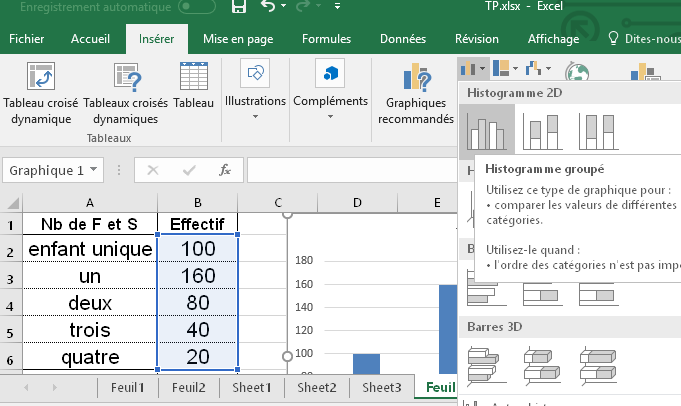
Reprenons l’exemple des décomptes de frères et sœurs avec 400 étudiants en nommant les modalités afin de clairement distinguer les modalités des effectifs. En bref, on présente cette variable quantitative discrète comme une variable qualitative ordinale.



Nous allons faire un diagramme en bâtons à partir de cette figure.

Commençons par une figure de type « Histogramme » (terme impropre utilisé par Excel pour décrire les diagrammes en bâtons).

Sélectionnez la plage B2:B6 en enfonçant le bouton gauche de la souris sur B2 puis en glissant jusqu’à B6 avant de relâcher le bouton gauche. Cliquez ensuite sur l’onglet « Insérer » puis sur histogramme 2D :



Cliquer Insérer

Sélectionner plage B2:B6

Cliquer histogramme 2D

Vous observerez un joli diagramme en barres :

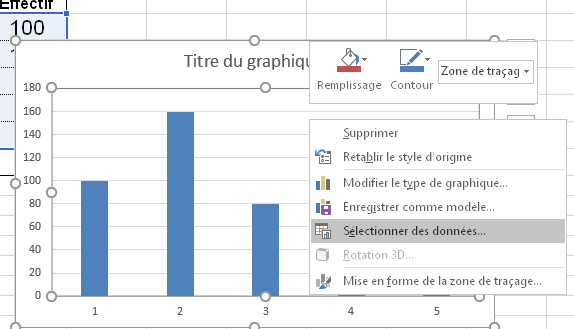


Vous constaterez que les effectifs par modalité apparaissent en ordonnée.

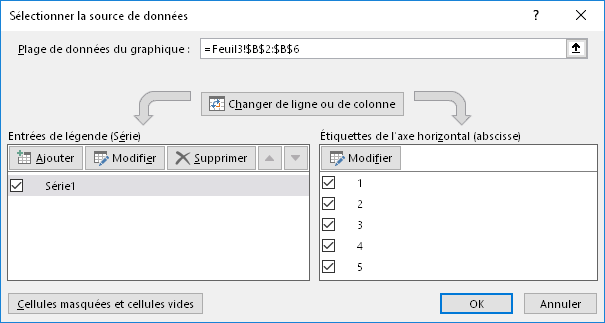
En abscisse, on retrouve bien les 5 modalités (de zéro frères et sœurs à 4 frères et sœurs), mais elles sont improprement nommées : 1 à 5 ! Cela est dû au fait que la **source de données** est constituée d’une plage de 4 cellules ne permettant pas à Excel de deviner les **étiquettes** (synonyme libellés) des modalités !

Pour donner à Excel ces libellés, il faut sélectionner la **plage** A2:B6 comme **source de données**. Ceci peut se faire par modification du graphique existant.

Faites un clic droit n’importe où sur le graphique, puis cliquez sur « Sélectionner des données »



Vous verrez alors apparaître une boîte de dialogue :



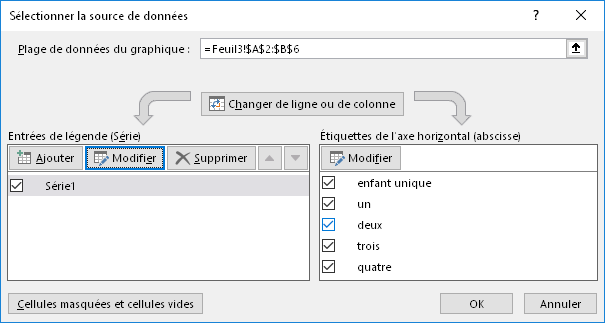
Vous pouvez changer la plage en remplaçant $B$2:$B$6 par $A$2:$B$6

Ou alors, vous pouvez directement utiliser le clic gauche/glisser de la souris pour sélectionner la zone du classeur correspondant.

Cliquez ensuite sur [OK]



Bien que la **source de données** contienne deux colonnes, il n’y a qu’une **série de données** et une **série d’étiquettes**. Vous pouvez de nouveau faire un clic droit sur la figure et aller dans « Sélectionner les données… » pour voir qu’il y a une plage qui a, à gauche, une série de données nommée « Série1 » que vous pouvez éventuellement modifier/renommer en cliquant sur [Modifier] en dessous de « Entrées de légende (Série) » :

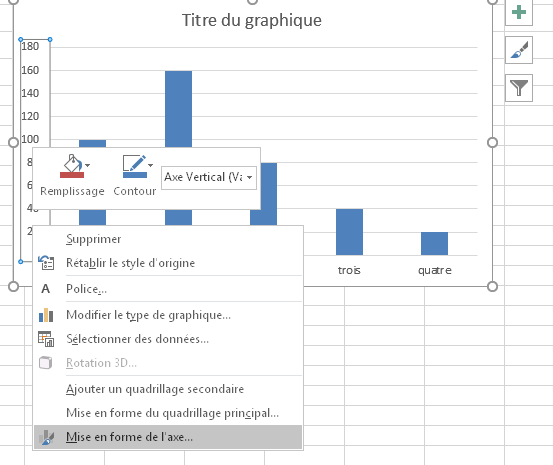


De même, vous pouvez changer la plage correspondant aux étiquettes en cliquant sur le bouton [Modifier] en dessous de « Étiquettes de l’axe horizontal (abscisse) ». Vous pouvez aussi déselectionner des modalités que vous souhaitez omettre de la représentation graphique.

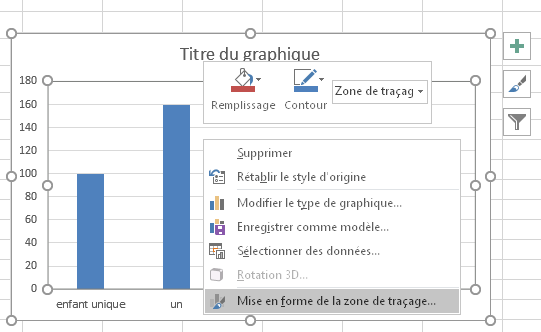
Vous pouvez aussi sélectionner la zone A1:B6 pour inclure les intitulés de colonnes qui serviront alors à définir le nom de la **série de données** de manière automatique.

Le graphique est constitué de plusieurs parties. Un clic droit sur chaque zone permet de régler des options spécifiques à cette zone.

Par exemple, si vous cliquez sur l’axe des ordonnées, vous verrez l’option « Mise en forme de l’axe ».



Si vous faites un clic droit sur le fond du graphique, vous aurez l’option « Mise en forme de la zone de traçage »



Éléments modifiables :

1. Axe des ordonnées (clic droit sur les étiquettes de l’axe telles que 0, 20, 40, 60)
2. Axe des abscisses (clic droit sur les étiquettes de l’axe telles que « enfant unique », « un », « deux », …)
3. Zone de traçage/arrière plan (clic droit sur le blanc du fond de la figure)
4. Quadrillage (clic droit sur les lignes horizontales du fond)
5. Titre du graphique (en haut)
6. Série de données complète (clic droit sur une des barres du diagramme)
7. Valeur de la série de données. Commencez par faire un clic gauche sur une des barres du diagramme (vous verrez toutes les barres surlignées), puis après un délai d’une seconde, faites un second clic gauche sur la barre du diagramme qui vous intéresse (seule cette barre restera surlignée), puis faites un clic droit dessus.

Tous les menus contextuels (accessibles par le clic droit) contiennent des options communes comme « Supprimer », « Rétablir le style d’origine », « Modifier le type de graphique » ou « Sélectionner des données… ». Après ces options générales, il y a des options spécifiques de l’objet sélectionné.

Je ne vais pas détailler toutes les options. Je vous conseille d’explorer systématiquement tout ce qui est possible ici.

Les options les plus importantes sont :

Clic droit sur l’axe des ordonnées -> Mise en forme de l’axe ->Options d’axe -> Limites (minimum et maximum).

Clic droit sur la série de données -> Ajouter des étiquettes de données

## Graphique avec plusieurs séries de données

Prenons l’exemple des mentions au baccalauréat.

Chez les étudiants ayant réussi le baccalauréat, on croise deux variables : type de Bac passé et mention au Bac.

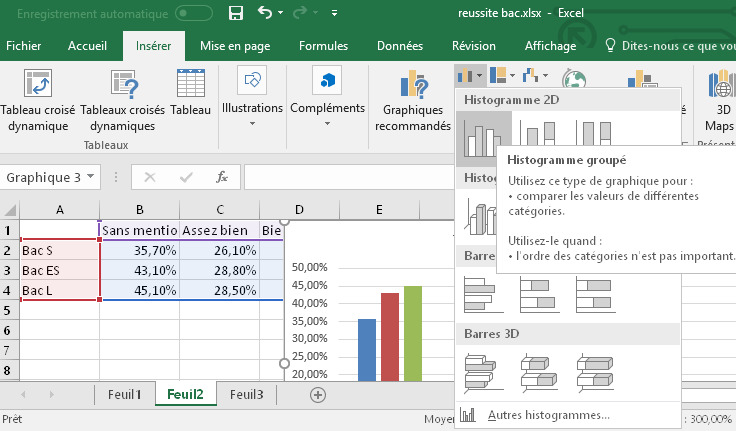


Par exemple 35,7% des étudiants reçus au Bac S n’avaient pas de mention. Les pourcentages font 100% en lignes !

**Note** : le tableau ci-dessous est un classeur Excel insérée dans un document Word. Si vous double-cliquez dessus, vous pourrez modifier ou copier/coller le classeur !

Commencez par changer le **format des données** numériques sous forme de **pourcentages**.

Ensuite, Sélectionnez la plage A1:E4 puis cliquez sur l’onglet Insérer -> histogramme 2D groupé.

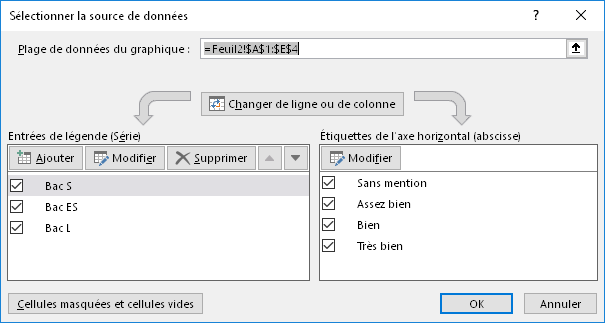


Vous verrez alors apparaître une diagramme en barres avec 3 séries de données : une pour chaque catégorie de Bac (S, ES et L).

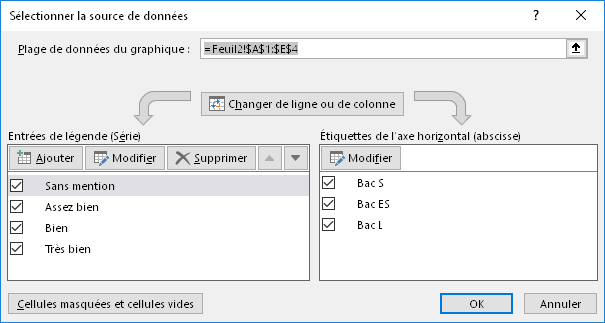
La série « Bac S » est en bleu, la série « Bac ES » est en rouge et « Bac L » est en vert.



Vous pouvez confirmer le fait qu’il y a trois séries de données, avec 4 valeurs par série, avec « clic droit -> Sélection des données ». Cela ouvrira la boîte de dialogue suivante :



Maintenant, supposons que nous préférions présenter les données dans l’autre sens : quatre séries (sans mention, assez bien, bien, très bien) de trois valeurs (Bac S, Bac ES, Bac L), alors, il suffit de cliquer [Changer de ligne ou de colonne] pour inverser les rôles :



Cliquez [OK]

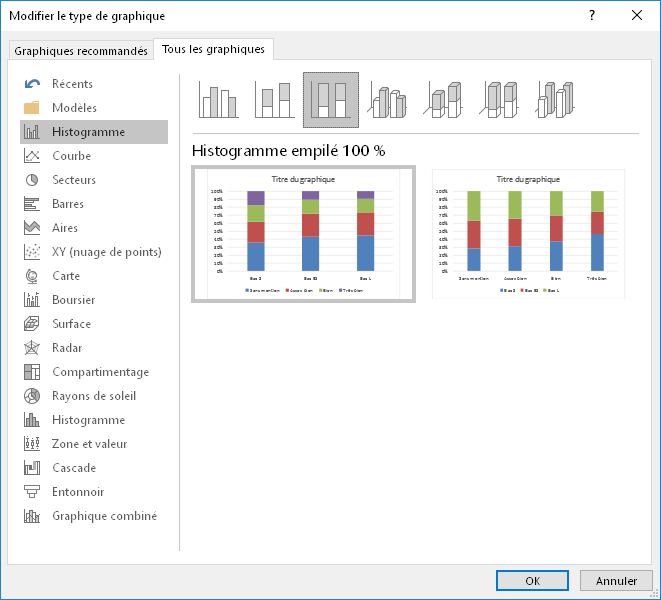


Voilà le travail : 4 séries, 3 valeurs

Plutôt qu’un diagramme en barres juxtaposées/groupées, on peut préférer un diagramme en barres superposées. Pour ce faire, il suffit de changer le type de graphique.

Clic droit n’importe où sur la figure -> modifier le type de graphique

Cliquez sur [Histogramme empilé 100%]



Cliquez [OK]

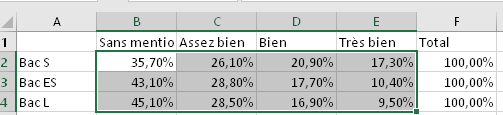


Il y a toujours 4 séries de données, mais elles apparaissent empilées les unes sur les autres, avec un total à 100%. Cela est logique puisque chacun des Bac, on a un pourcentage dont le total fait 100%. Connaissant la base de données agrégées que nous avons, il ne faudrait surtout pas faire ça dans l’autre sens ! Ici, on ne connaît pas la proportion de bacheliers du Bac S, ES et L.

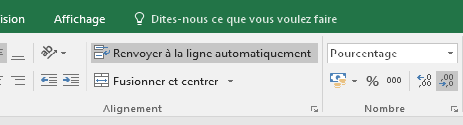
Maintenant, faites un clic droit sur chaque série de données et cliquez « Ajouter des étiquettes de données ».



Sélectionnez de nouveau la plage de données



Puis cliquez sur l’icône pour réduire le nombre de chiffres après la virgule :



Réduit le nombre de chiffres après la virgule

La figure est immédiatement mise à jour :



# Tableaux croisés dynamiques

Comme vu précédemment, les données des graphiques doivent, dans le cas général, avoir été préalablement agrégées.

Les tableaux croisés dynamiques permettent d’agréger les données.

Prenons un exemple assez typique : calculer la moyenne d’âge dans 4 sous-groupes définis par le croisement de deux variables.

Il y a trois variables

Variable maladie oui/non (qualitatif binaire)

Variable sexe M/F (qualitatif binaire)

Variable âge (quantitatif)

On va demander à Excel de calculer les moyennes d’âge dans quatre sous-groupes :

Les femmes indemnes de la maladie

Les femmes malades

Les hommes indemnes de la maladie

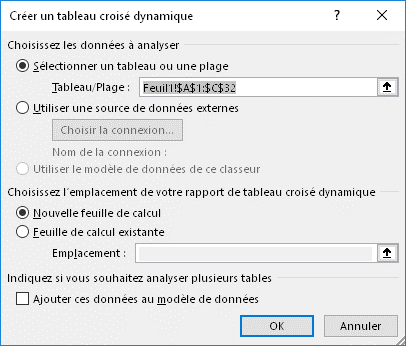
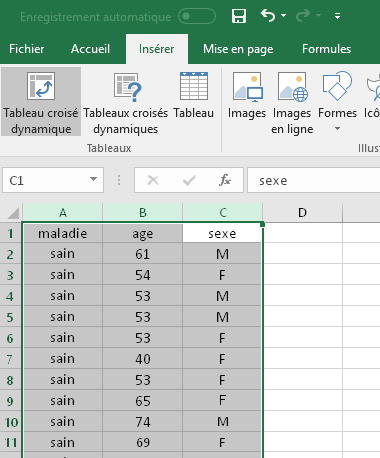
Les hommes malades

Ci-dessous la base de données :



Vous pouvez double-cliquer sur ce tableau et copier/coller dans Excel (c’est un objet Excel inséré dans Word).

Sélectionnez ces trois colonnes.



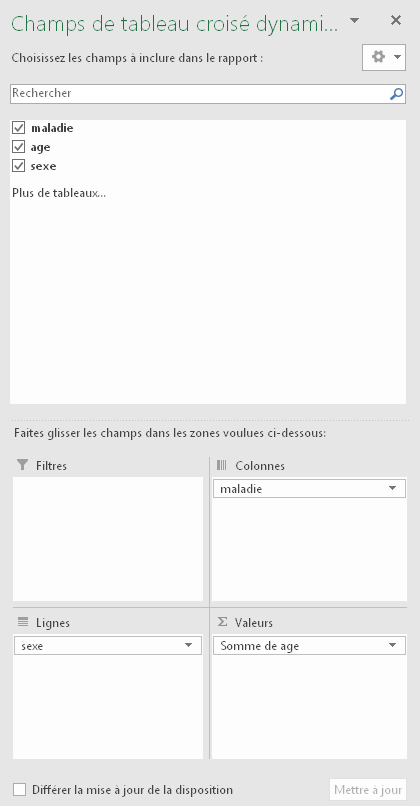
Insérer -> tableau croisé dynamique

Sélectionner colonnes

Vérifier plage

Choisir « nouvelle feuille » ou « feuille existante »

Cliquer [OK]

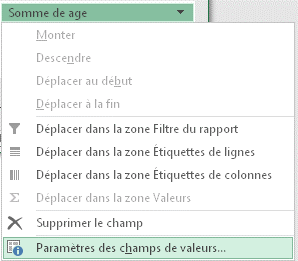


Glisser maladie en « colonnes »

Glisser sexe en « lignes »

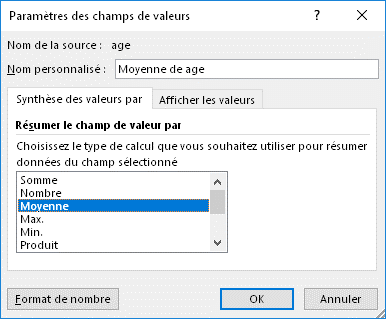
Glisser âge en « ∑ valeurs »

c.à.d. « synthèse des valeurs »



Cliquer sur « Somme de age »

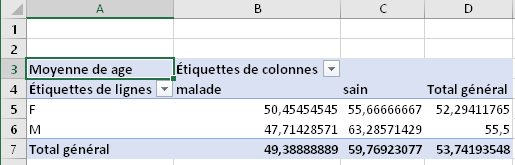
Puis « paramètres des champs de valeurs… »



Sélectionner la statistique « Moyenne »

Puis [OK]

Et voilà !



À l’intersection d’une modalité du sexe en ligne (F ou M) et du statut pathologique en colonne (malade ou sain) se trouve la moyenne d’âge. Par ailleurs, il y a les moyennes totales, tel que 52,29411765 pour la moyenne d’âge des femmes, peu importe leur statut pathologique ou 39,38889 pour la moyenne d’âge des malades, peu importe leur sexe. Enfin, la moyenne générale de l’âge est égal à 53,7419, peu importe le sexe ou le statut pathologique.

La puissance des tableaux croisés dynamiques est à peine effleurée ici. Si on veut juste compter les cas, plutôt que d’utiliser la statistique « Moyenne » on peut utiliser la statistique « Nombre » qui mesure l’effectif. On peut remarquer que le « Nombre de age » ou le « Nombre de sexe » ou le « Nombre de maladie » vont fournir exactement la même information sauf en cas de cellules vides… Plutôt que de compter les unités statistiques, Excel compte le nombre de réalisations d’une variable aléatoire. Dans ce cas, il faut préférer une variable aléatoire pour laquelle il ne peut exister de donnée manquante (cellule vide) telle que l’identifiant patient.

Ce tableau croisé dynamique est bidimensionnel (lignes et colonnes), on peut faire des tableaux en un bien plus grand nombre de dimensions, en glissant plusieurs variables **en lignes**, **en colonnes** ou **en filtres**.

# Quelques raccourcis pour se faciliter la vie

Pour enregistrer votre classeur tapez [CTRL]+[S]

Pour enregistrer sous un nouveau nom ou format de fichier, tapez [F12].

Pour vous déplacer, utilisez [CTRL]+[flèches], [Page Préc] et [Page Suiv] (les arêtes de Poisson / sapins), [CTRL]+[FIN] et [CTRL+ORIGINE].

Utilisez [MAJUSCULE]+[touches de déplacement] pour sélectionner une zone du classeur.

Vous pouvez aussi double cliquer sur les bords du rectangle de sélection (avec ou sans [MAJUSCULE] enfoncé) et double cliquer sur le petit carré en bas du rectangle de sélection pour recopier automatiquement une formule sur toute la colonne jusqu’à la première cellule vide.

Ces astuces fonctionnent sur toutes les versions d’Excel. Vous remarquerez que l’interface graphique change à chaque version (au CHU, on doit travailler avec toutes les versions des quinze dernières années), mais les raccourcis clavier sont beaucoup plus pérennes.

# Le bug des lignes vides d’Excel

**Note** : en aucun cas on ne vous demandera de connaître ce bug (ou son historique) à l’examen. Ce chapitre a pour objectif de vous aider à résoudre le problème si vous y êtes confronté.

## Symptômes

Votre fichier est très gros et lent à ouvrir et à enregistrer.

## Historique du bug

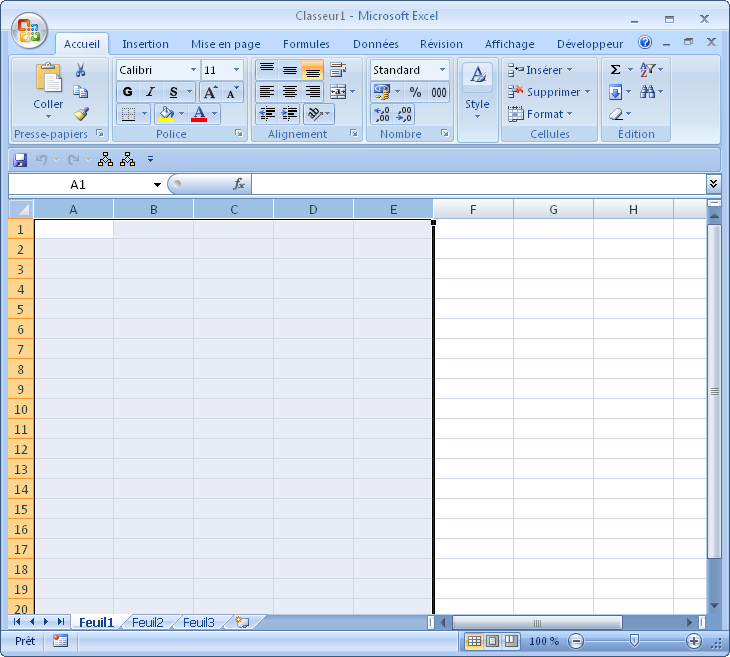
Ce bug a été précieusement conservé depuis les plus anciennes versions d’Excel. Le bug existe depuis la version 1.0 d’Excel en 1985, mais il n’a cessé de s’aggraver au fil des versions :

* Version 1.0 en Janvier 1985 : dégâts aux fichiers limités du fait de messages « Out of memory » qui bloquaient les modifications avant même l’enregistrement
* Version 3.0 en 1990 : il devient facile de créer de très grands fichiers buggés car il est possible de procéder à des opérations dépassant la capacité de la mémoire à condition de rennoncer à la possibilité d’annuler.
* Version 8.0 en 1997 : le nombre de lignes passant de 16384 à 65536, l’impact du bug est multiplié par 4
* Version 12.0 en 2007 : le nombre de lignes passant de 65536 à 1048576, le bug est agravé d’un facteur 16
* Version 16.0 en Septembre 2015 : le bug a dépassé les 30 ans et n’est toujours pas corrigé

## Comment reproduire le bug

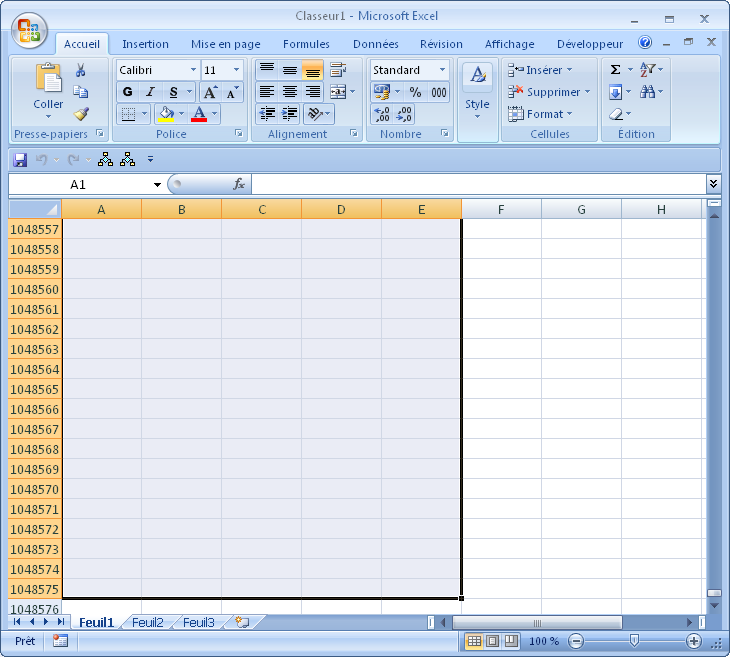
Ouvrez un classeur vide

Sélectionnez une ou plusieurs colonnes



Appuyez [MAJUSCULE]+[FLÈCHE VERS LE HAUT]

Cela sélectionne toutes les lignes de l’espace « infini » d’Excel sauf la dernière !



Changez le style des cellules en cliquant par exemple sur [GRAS] ou en appuyant CTRL+G

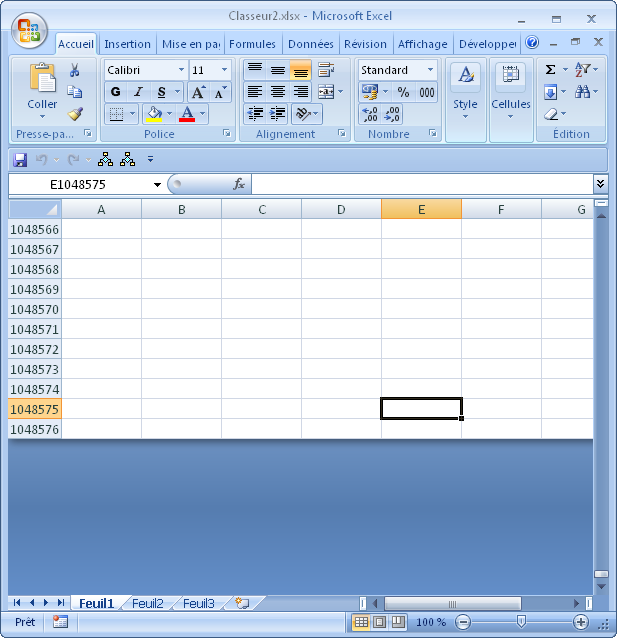
Excel met une petite seconde.

Enregistrez le classeur. Il faut alors quelques secondes et le fichier enregistré est énorme !

Rouvrez le classeur, il lui faut encore plus de temps !

## Comment identifier le bug

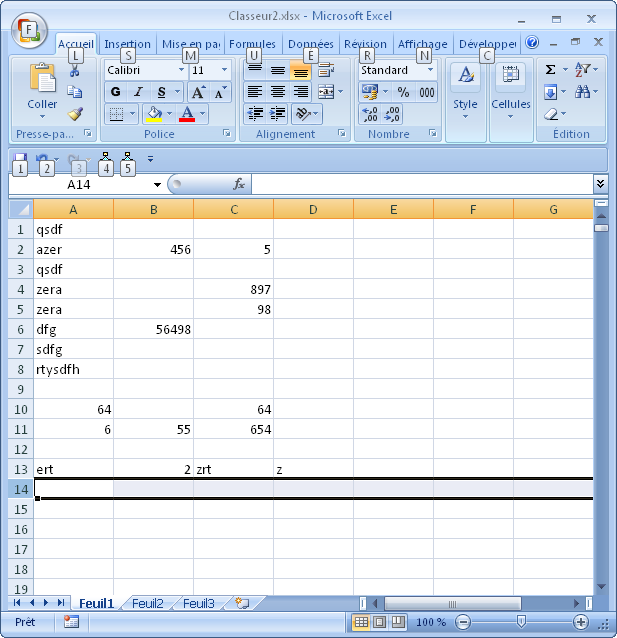
1. Le fichier est énorme
2. Il met beaucoup de temps à s’ouvrir
3. Une fois qu’il est ouvert, rentrez dans chaque feuille, et tapez [CTRL]+[FIN]
4. Si le fichier est bien buggé, vous devriez sauter à une cellule très éloignée



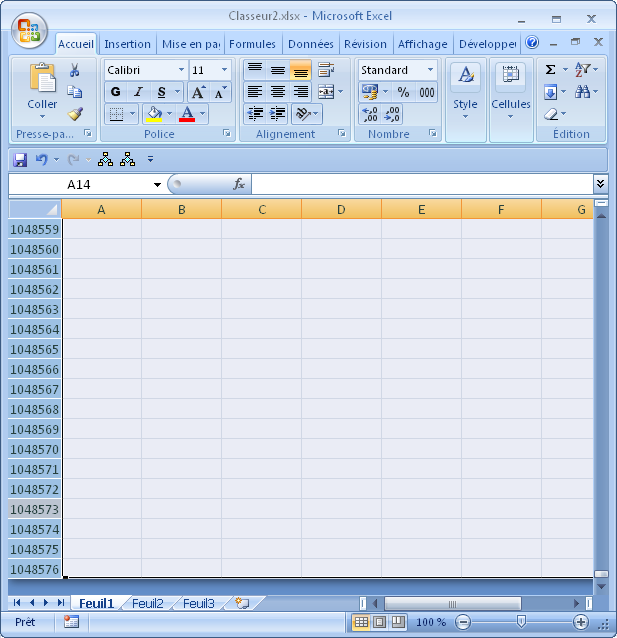
## Comment corriger le bug

Le bug peut affecter une ou plusieurs feuilles. Il va falloir le corriger sur chacune des feuilles buggées.

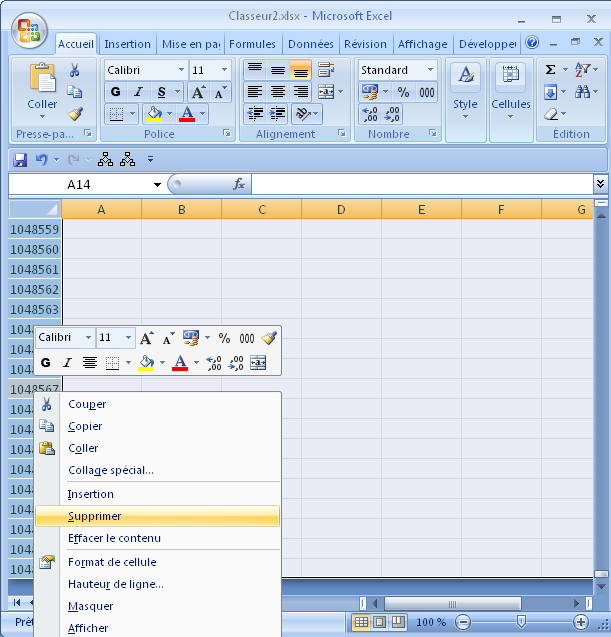
Identifiez la première ligne « vide et inintéressante » du tableur, au-delà des données que vous voulez conserver.



Appuyez[CTRL]+[FLÈCHE VERS LE BAS] pour sélectionnez toutes les lignes jusqu’à la dernière !



Appuyez [CTRL]+[MOINS] du pavé numérique ou alors faites un clic droit sur la colonne de numéros de lignes et cliquez Supprimer



Maintenant, le formatage vide a disparu !

Recommencez la procédure sur toutes les feuilles buggées.

Enregistrez le fichier !

Le problème, dans une moindre mesure, peut arriver avec les colonnes. La procédure est la même, mais on sélectionne alors les colonnes et on utilise [CTRL]+[FLÈCHE DROITE].

## Autre solution

1. Copiez coller la zone d’intérêt de votre feuille dans une nouvelle feuille propre.
2. Supprimez l’ancienne feuille
3. Renommez la nouvelle feuille au même nom que l’ancienne feuille

# En bref

Structurez proprement votre base de données

Ne stockez pas d’information dans le style (couleur, bordure, police)

Utilisez les formules pour construire de nouvelles données à partir d’anciennes

Vérifiez la cohérence de vos données

Utilisez les tableaux croisés dynamiques pour agréger les données

Ensuite, créez des graphiques à partir de ces tableaux croisés dynamiques

Méfiez vous du bug des lignes vides d’Excel

Cherchez sur le Web.

Familiarisez vous avec la structure de l’aide en ligne (F1)

En cas de difficulté technique insoluble, même avec une recherche approfondie, les dernières ressources sont, l’aide d’un expert :

<https://superuser.com/>

[andre.gillibert@chu-rouen.fr](mailto:andre.gillibert@chu-rouen.fr)