

Chapitre 1 : Statistique univariée : encodage de tableaux de données brutes, calculs de statistiques descriptives

1. Encodage de données brutes

Dans Excel, pas de structure préétablie pour encoder les données. Mais avec quelques bons réflexes, on peut éviter de se perdre dans des fichiers mal structurés et difficiles à lire !

Règles de base :

Une ligne par individu (ou « enregistrement »)

Une colonne par variable (ou « champ »)

Avantage :

- Facilite l'export vers d'autres logiciels (SPSS, par exemple)
- Facilite la collaboration (car convention fréquemment utilisée)

À titre d'exemple, ouvrez le classeur *R1-9.xls/x*, feuille *Données brutes*. Vous allez encoder correctement les données suivantes, relatives à 18 sujets.

I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Score (X_i)	6	9	4	13	14	9	8	12	11	9	12	18	14	10	19	8	15	11
Groupe (Y_i)	1	2	2	2	1	2	2	2	1	2	1	2	2	1	1	2	1	2
Age (Z_i)	17	18	22	16	14	15	20	23	17	18	19	19	19	24	22	16	21	26
Sexe (W_i)	1	1	1	1	1	2	2	1	1	2	1	2	2	1	1	2	2	1

Avant de commencer, quelques règles, de sorte à augmenter la lisibilité :

Règle 1 : créez une première ligne (= ligne 1 du tableur) = nom des variables.

Intérêt : permet de discriminer les colonnes.



Règle 2 : créez une première variable i (=colonne A du tableur) = numéro des individus statistiques.

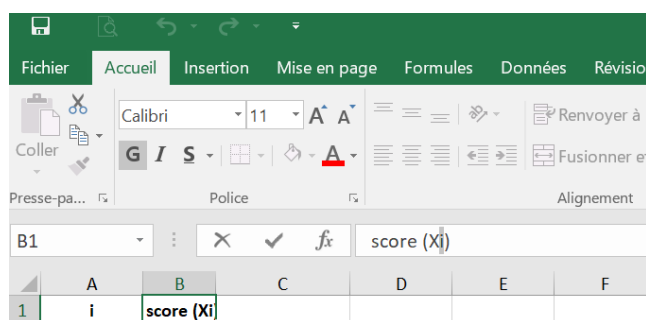
Intérêt : si au cours de vos analyses, vous êtes amenés à modifier l'ordre de présentations des données, cette première colonne pourra toujours servir à retrouver l'ordre initial.

Encoder correctement les noms de variable :

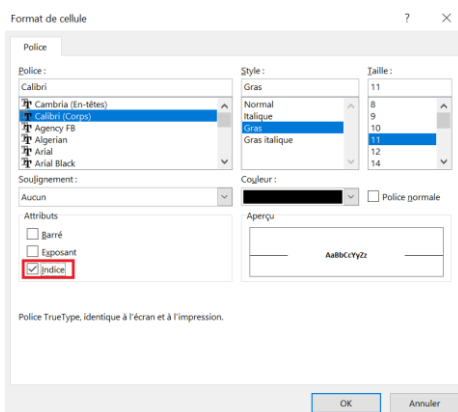
Pour rappel, une variable se symbolise par une lettre majuscule, associée à un indice « i » qui représente l'individu étudié.

Commencez par écrire le nom complet d'une variable dans la cellule adéquate, soit score (Xi) dans la cellule B1.

Ensuite, dans la barre de formule, mettez en surbrillance la partie que vous voulez transformer en indice (soit la lettre « i »). Pour ce faire, vous pouvez maintenir le bouton gauche de la souris enfoncé le temps de mettre en surbrillance ce que vous souhaitez, ou alternativement, placer le curseur entre la lettre majuscule et le « i », dans la barre de formule, enfoncer la touche « Shift »  de votre clavier, et simultanément, la flèche vers la droite .



Dans l'onglet d'affichage « accueil », cliquez sur la flèche en bas à droite de la zone « Police » du ruban pour ouvrir l'interface des paramètres de police, et cochez « indice » dans les attributs. Pour finir, cliquez sur OK.



Reproduisez la même démarche pour encoder correctement le nom des 3 autres variables.

Résultat final attendu :

l11					
	A	B	C	D	E
1	i	score (X_i)	groupe (Y_i)	âge (Z_i)	sexe (W_i)
2	1	6	1	17	1
3	2	9	2	18	1
4	3	4	2	22	1
5	4	13	2	16	1
6	5	14	1	14	1
7	6	9	2	15	2
8	7	8	2	20	2
9	8	12	2	23	1
10	9	11	1	17	1
11	10	9	2	18	2
12	11	12	1	19	1
13	12	18	2	19	2
14	13	14	2	19	2
15	14	10	1	24	1
16	15	19	1	22	1
17	16	8	2	16	2
18	17	15	1	21	2
19	18	11	2	26	1

l11					
	A	B	C	D	E
1	i	score (X_i)	groupe (Y_i)	âge (Z_i)	sexe (W_i)
2	1	6	1	17	1
3	2	9	2	18	1
4	3	4	2	22	1
5	4	13	2	16	1
6	5	14	1	14	1
7	6	9	2	15	2
8	7	8	2	20	2
9	8	12	2	23	1
10	9	11	1	17	1
11	10	9	2	18	2
12	11	12	1	19	1
13	12	18	2	19	2
14	13	14	2	19	2
15	14	10	1	24	1
16	15	19	1	22	1
17	16	8	2	16	2
18	17	15	1	21	2
19	18	11	2	26	1

→ 4 variables

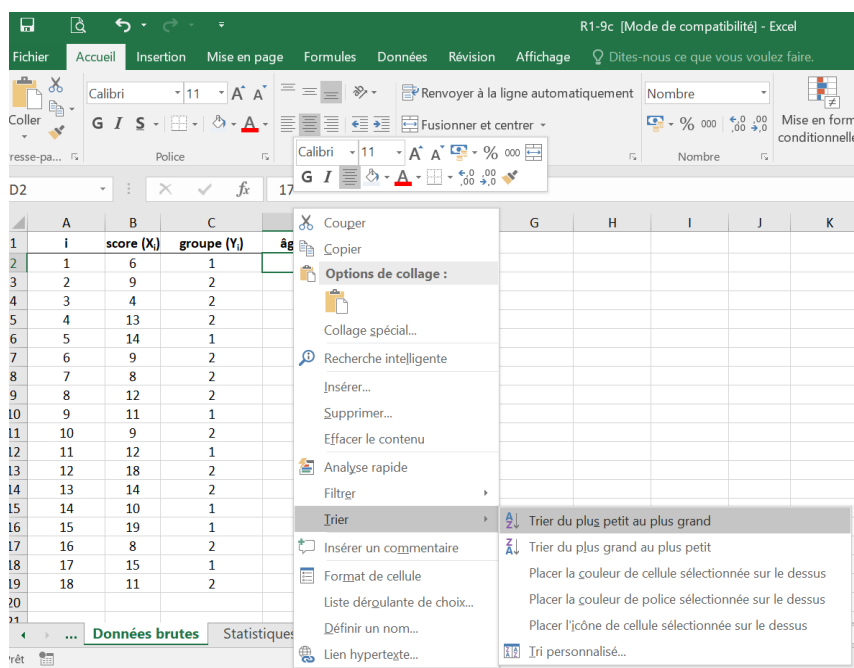
18 sujets

Intérêt de la colonne « i » ?

Nous allons trier les observations par ordre croissant d'âge. Pour ce faire, sélectionnez l'une des cellules de la colonne « âge » (par exemple D2).

	A	B	C	D	E
1	i	score (X_i)	groupe (Y_i)	âge (Z_i)	sexe (W_i)
2	5	14	1	14	1
3	6	9	2	15	2
4	4	13	2	16	1
5	16	8	2	16	2
6	1	6	1	17	1
7	9	11	1	17	1
8	2	9	2	18	1
9	10	9	2	18	2
10	11	12	1	19	1
11	12	18	2	19	2
12	13	14	2	19	2
13	7	8	2	20	2
14	17	15	1	21	2
15	3	4	2	22	1
16	15	19	1	22	1
17	8	12	2	23	1
18	14	10	1	24	1
19	18	11	2	26	1

Ensuite, faites un clic droit sur votre souris et choisissez l'option « Trier du plus petit au plus grand ».



Le tri s'est étendu à l'ensemble de la base de données, ce qui est impératif, dans la mesure où toutes les informations qui figurent sur une ligne correspondent au même individu statistique.

Résultat :

112					
	A	B	C	D	E
1	i	score (X_i)	groupe (Y_i)	âge (Z_i)	sexe (W_i)
2	5	14	1	14	1
3	6	9	2	15	2
4	4	13	2	16	1
5	16	8	2	16	2
6	1	6	1	17	1
7	9	11	1	17	1
8	2	9	2	18	1
9	10	9	2	18	2
10	11	12	1	19	1
11	12	18	2	19	2
12	13	14	2	19	2
13	7	8	2	20	2
14	17	15	1	21	2
15	3	4	2	22	1
16	15	19	1	22	1
17	8	12	2	23	1
18	14	10	1	24	1
19	18	11	2	26	1

Le numéro de chaque ligne (zone grisée qui apparaît juste avant la colonne A) n'a pas changé. Par contre, puisque le tri a été étendu à l'ensemble de la base de données, l'ordre d'apparition des chiffres dans la colonne « i », lui, s'est modifié !

Grâce à cela, on est capable de déduire que le sujet étant le plus jeune était le 5^{ème} à être encodé dans la base de données. $Z_5 =$ le minimum de la variable âge = 14.

Remarque : les variables $\text{groupe}(Y_i)$ et $\text{sexe}(W_i)$ sont des variables nominales. Pourtant, les modalités de la variable semblent être chiffrées. Rappelez-vous le cours théorique : il arrive souvent que l'on attribue des chiffres comme nom d'étiquette aux modalités. Pour voir à quelle catégorie/modalité se rapporte chacune des valeurs numériques pour ces deux variables, regardez la feuille *Légendes* du fichier.

2. Calcul de statistiques descriptives, au départ de données brutes

Si peu de données/variables, les paramètres peuvent être calculés dans le feuillet dédié aux données brutes.

Si beaucoup de données/variables, il est préférable de calculer et encoder les statistiques descriptives sur une feuille à part :

- Créer une première variable dans la colonne A, que l'on nomme *Indicateur statistique*.
- Indiquer le nom de la variable représentée dans la première ligne de chaque colonne
- ATTENTION : il est très important de respecter scrupuleusement l'ordre d'apparition des variables. Cela permet d'étendre facilement les formules.

Exemple : ouvrez le classeur *R1-9.xlsx*, feuille *Statistiques descriptives*.

Rappel : pour tous les calculs ci-dessous, il vous sera demandé d'introduire dans la feuille *Statistiques descriptives* une formule faisant appel à des séries de données figurant dans la feuille *Données brutes*. Pensez à votre cours d'introduction à Excel pour savoir comment vous y prendre.

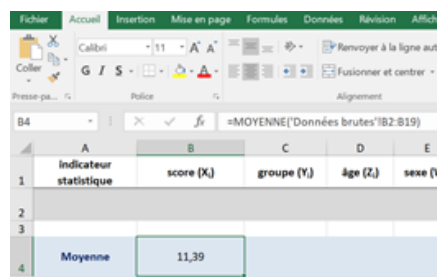
Indicateurs de tendance centrale

La moyenne

La **Moyenne** est un indicateur de tendance centrale qui consiste à additionner toutes les valeurs d'une variable, et à diviser le tout par le nombre de termes additionnés.

Fonction : **=MOYENNE()**. L'argument à introduire entre parenthèses sera la plage de la série de données, soit l'ensemble des valeurs à considérer pour calculer la moyenne (par exemple, la série de données *D2:D19* de la feuille *Données brutes* contient les données des 18 sujets relatives à la variable *Âge(Z_i)*).

Pour éviter des erreurs, un bon réflexe est d'étendre les formules lorsque c'est possible. En *B4*, introduisez la formule adéquate pour calculer la moyenne de la variable *Score(X_i)*. Affichez deux décimales.



Étirez ensuite la formule dans les cellules *C4:E4*.

Avez-vous obtenu le résultat ci-dessous ?

	A	B	C	D	E	F
1	Indicateur statistique	score (X _i)	groupe (Y _i)	âge (Z _i)	sexe (W _i)	
2	Type de variable					
4	Moyenne	11,22	1,61	19,22	1,39	Indice de tendance

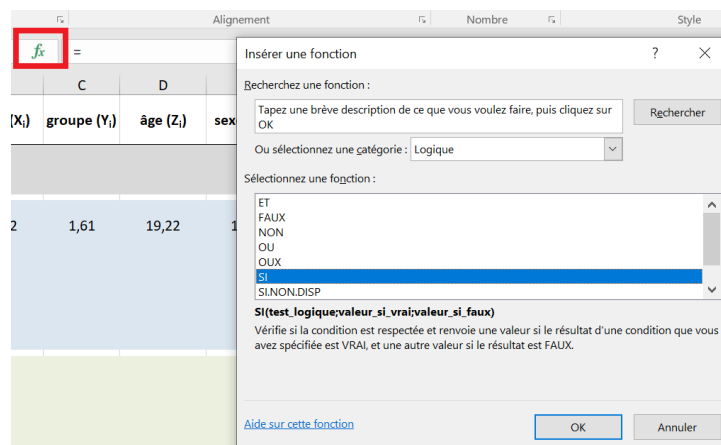
On y est presque, mais un détail devrait vous choquer...À quoi correspond le score 1,39 pour la variable *Sexe(W_i)* ? et le 1,61 pour la variable *Groupe(Y_i)* ?

Souvenez-vous : le calcul de la moyenne n'est pas du tout pertinent pour des variables de type nominal (ni pour n'importe quelle variable de type qualitatif)!

Nous allons corriger cela. Dans un premier temps, indiquez pour chaque variable son type. A niveau de la ligne 2, écrivez pour chaque variable, soit « qualitatif », soit « quantitatif ». Ensuite, modifiez la formule que vous avez introduite en B4 et faites-en sorte que :

- Si la variable est de type quantitatif, la moyenne est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Prérequis : les fonctions conditionnelles (syllabus 7, niveau 2 de votre cours d'introduction à Excel). Un trou de mémoire ? N'hésitez pas à utiliser l'aide disponible dans Excel. Cliquez sur le symbole « insérer une fonction », et faites une recherche dans les fonctions logiques. Vous y trouverez le nécessaire.



La médiane

La **Médiane** est un indicateur plus robuste que la moyenne (=moins influencé par les valeurs extrêmes). Elle consiste à calculer une valeur telle qu'il y ait autant d'observations qui lui soient inférieures ou égales qu'il n'y a d'observations qui lui soient supérieures ou égales (on découpe la distribution en 2 parts contenant le même nombre de sujets).

Fonction : **=MEDIANE()** », où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B5 une formule qui permet de calculer la médiane de la variable *Score(X_i)*, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, la médiane est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez une décimale et étirez la formule dans les cellules C5:E5.

Le mode

Contrairement à la moyenne et à la médiane, le **Mode** est un indicateur de tendance centrale qui convient tant pour décrire les variables qualitatives que les variables quantitatives.

Remarque 1 : Bien que le mode soit assez simple à calculer dans Excel lorsque les modalités d'une variable sont chiffrées, c'est plus compliqué lorsque les modalités sont non chiffrées (si vous vous demandiez pourquoi nous avons choisi des valeurs chiffrées pour décrire le sexe et le groupe, voici une partie de la réponse).

Remarque 2 : il existe la fonction suivante dans Excel : **=MODE()**, où l'argument à introduire entre parenthèses est la plage de la série de données. Bien que cette fonction soit toujours disponible dans les versions actuelles d'Excel, elle est amenée à disparaître. Il est préférable de lui préférer deux nouvelles fonctions plus précises ayant été ajoutée ultérieurement :

- **=MODE.SIMPLE()**
- **=MODE.MULTIPLE()**

L'option **=MODE.SIMPLE()** retourne systématiquement une seule valeur, soit celle associée à la plus grande fréquence. Elle est adéquate uniquement lorsque la distribution est unimodale (ce qui est le cas pour toutes les variables du fichier R1-9.xls/x).

Fonction : **=MODE.SIMPLE()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B6 une formule qui permet de calculer le mode de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, le mode est calculé
- Si la variable est de type qualitatif, une barre « / » est imprimée.

N'affichez aucune décimale et étirez la formule dans les cellules C6:E6.

E21						
	A	B	C	D	E	F
	indicateur statistique	score (X_i)	groupe (Y_i)	âge (Z_i)	sexe (W_i)	
1						
2	Type de variable	quantitatif	qualitatif	quantitatif	qualitatif	
3						
4	Moyenne	11,22	/	19,22	/	Indicateurs de tendance centrale
5	Médiane	11,0	/	19,0	/	
6	Mode	9	2	19	1	

À présent, répondez aux questions suivantes :

- Quel type de famille est le plus représenté dans l'échantillon : les familles monoparentales, ou les familles biparentales ?
- Quel sexe est le plus représenté dans l'échantillon : les hommes ou les femmes ?

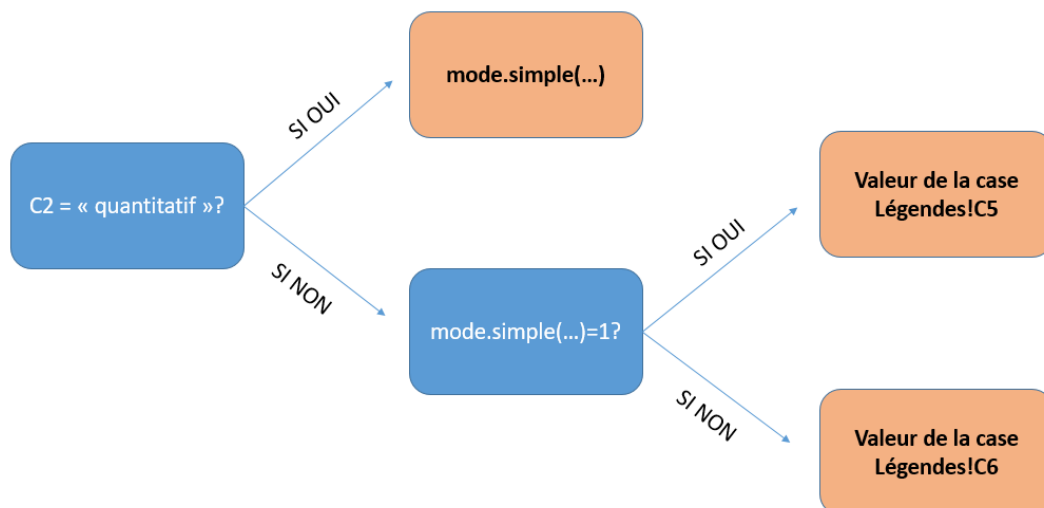
La réponse figure dans la feuille *Légende* (vous y trouverez ce que représente le « 2 » dans la variable *Groupe*(Y_i), et le « 1 » dans la variable *Sexe*(W_i)).

Imaginez que vous ayez quelques minutes pour analyser un grand tableau, constitué de plein de variables, et que vous deviez chaque fois switcher de la feuille *Légendes* vers la feuille *Statistique descriptives*. Peu pratique n'est-ce pas ?

Nous allons modifier la formule, de sorte que lorsque la variable est de type qualitatif, ce ne soit plus un code chiffré qui apparaisse, mais bien le nom de la catégorie représentée par ce chiffre.

Astuce : à nouveau, utiliser les fonctions conditionnelles, mais cette fois, en imbriquant deux conditions. Pensez à placer les « \$ » correctement, pour pouvoir étirer les formules.

Besoin d'un coup de pouce ? Commencez par modifier la formule de la cellule C6 (vous la copiez-collerez ensuite dans les cellules B6, D6 et E6). Cet arbre de décision devrait vous aider :



Résultat attendu :

A1						
	A	B	C	D	E	F
1	indicateur statistique	score (X _i)	groupe (Y _i)	âge (Z _i)	sexe (W _i)	
2	Type de variable	quantitatif	qualitatif	quantitatif	qualitatif	
3						
4	Moyenne	11,22	/	19,22	/	Indicateurs de tendance centrale
5	Médiane	11,0	/	19,0	/	
6	Mode	9	mono-parental	19	Homme	

L'option **=MODE.MULTIPLE()** retourne une seule valeur si la distribution est unimodale. Elle retourne par contre plusieurs valeurs si la distribution est bimodale. Vous l'aurez compris, l'option **=MODE.MULTIPLE()** est plus pertinente que l'option **=MODE.SIMPLE()**, puisqu'elle peut être utilisée tout le temps. Cependant, elle ne fera pas l'objet de ce cours, puisqu'il s'agit d'une fonction matricielle, dont la difficulté dépasse les objectifs de ce cours.

Indicateurs de dispersion

L'**Étendue** des données consiste à calculer l'écart entre les valeurs maximales et minimales. La calculer implique donc d'utiliser deux fonctions : celles du minimum et celle du maximum.

Fonctions :

- **=MIN()**, où l'argument à introduire entre parenthèses est la plage de la série de données.
- **=MAX()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B21 une formule qui permet de calculer l'étendue de la variable *Score(X_i)*, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, l'étendue est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

N'affichez aucune décimale et étirez la formule dans les cellules C21:E21.

L'**Écart interquartile** est l'écart entre le troisième quartile et le premier quartile (il correspond à la boîte centrale de la boîte à moustache).

Rappel :

- Le troisième quartile est égal au 75^{ème} percentile
- Le premier quartile est égal au 25^{ème} percentile

Dans Excel, vous pourrez utiliser soit la fonction quartile, soit la fonction percentile.

Fonctions :

- **=CENTILE()**, où les arguments à introduire entre parenthèses sont, d'une part, la plage de la série de données et d'autre part, une proportion (soit k, la proportion des valeurs de la série de données qui sont inférieures ou égales à la valeur recherchée).
 - k = 0,25 → premier quartile
 - k = 0,5 → médiane
 - k = 0,75 → troisième quartile
- **=QUARTILE()**, où les arguments à introduire entre parenthèses sont, d'une part, la plage de la série de données et d'autre part, le numéro du quartile :
 - 1 → premier quartile ;
 - 2 → second quartile ;
 - 3 → troisième quartile ;

Créez-en B22 une formule qui permet de calculer l'écart interquartile de la variable *Score*(X_i), en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, l'écart interquartile est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C22:E22.

L'**Écart moyen absolu** (ou **EMA**) est une mesure de dispersion très utile, bien que peu utilisée en pratique. Elle consiste à calculer la moyenne des écarts absolus des observations d'une série de données par rapport à la moyenne arithmétique de cette série.

Fonction : **=ECART.MOYEN()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B23 une formule qui permet de calculer l'EMA de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, l'EMA est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C23:E23.

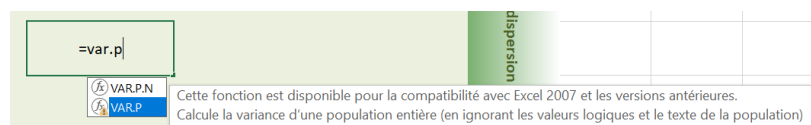
La **Variance** est une alternative à l'EMA beaucoup plus fréquemment utilisée. Elle consiste à calculer la moyenne des écarts au carré des observations d'une série de données par rapport à la moyenne arithmétique de cette série.

Il faut distinguer deux cas, lorsqu'on veut calculer la variance dans Excel :

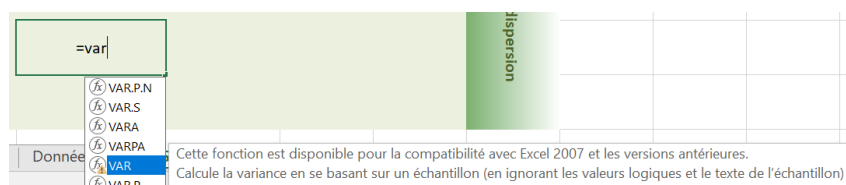
- 1) Celui où l'on veut simplement décrire notre échantillon (fonction **=VAR.P()** = statistiques descriptives).
- 2) Celui où l'on veut estimer ce que serait la variance d'une population plus large au départ de laquelle notre échantillon est récolté (fonction **=VAR()** = statistiques inférentielles).

Faites **très** attention ici. L'aide Excel pourrait vous sembler contre-intuitive et pourrait vous induire en erreur !

Voici l'explication associée à la formule **=VAR.P()** :



Et voici celle associée à la formule **=VAR()** :



Cela vous semble-t-il contradictoire avec mon explication plus haut ?

Si **non**, l'explication ci-dessous ne vous intéressera pas. Si **oui**, voici ce qu'il en est :

Comme expliqué en introduction de ce cours, il arrive souvent que l'on veuille décrire une population sur base d'un échantillon (par exemple, parce qu'il n'est pas possible de récolter les données de la population entière). Cependant, la formule de la variance que l'on vous a apprise dans ce cours est telle que bien souvent, elle aura pour effet de légèrement sous-estimer la variance de la population, c'est pourquoi il convient d'effectuer une correction.

Lorsque dans Excel, vous précisez vouloir utiliser la formule « en se basant sur un échantillon », il sera automatiquement compris que vous voulez estimer le paramètre d'une population plus large.

Lorsqu'au contraire vous précisez vouloir utiliser la formule de la variance d'une population entière, il est compris qu'il n'est pas nécessaire d'effectuer de correction, puisque le but n'est pas d'estimer la variance d'un ensemble contenant plus de sujets que récoltés.

Autrement dit, « leurrer » Excel, en affirmant que notre échantillon est en fait la population est le moyen de calculer la formule de variance sans correction, et c'est ce que j'attends de vous.

Fonction : **=VAR.P()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B24 une formule qui permet de calculer la variance de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, la variance est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C24:E24.

Enfin, l'**écart-type** est la racine-carré de la variance. Il existe deux manières de la calculer :

- Soit vous utilisez la fonction proposée par Excel (à nouveau, il convient de choisir l'option qui permet de calculer l'écart-type d'une population entière ; cf. explication ci-dessus).
- Soit vous calculez la racine-carré de la valeur introduire à la ligne 24.

Fonctions :

- **=ECARTYPE.P()**, où l'argument à introduire entre parenthèses est la plage de la série de données.
- **OU =RACINE()**, où l'argument à introduire entre parenthèses est la cellule contenant la valeur dont on veut prendre la racine carrée.

Créez-en B25 une formule qui permet de calculer l'écart-type de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, l'écart-type est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C25:E25.

Mesures d'asymétrie et d'aplatissement

La mesure d'**asymétrie** sert entre autres à décider si la moyenne sera ou non un résumé pertinent de notre distribution.

Fonction : **=COEFFICIENT.ASYMETRIE()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Créez-en B27 une formule qui permet de calculer le coefficient G1 de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

- Si la variable est de type quantitatif, la mesure d'asymétrie est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C27:E27.

Question : parmi toutes les variables quantitatives, laquelle présente la plus forte asymétrie ? S'agit-il d'une asymétrie positive, ou négative ?

La mesure d'**Aplatissement** est un indicateur de forme de la distribution, essentiellement lié à la densité des extrémités des distributions (plus le kurtosis est élevé, plus forte est la densité des extrémités).

Fonction : **=KURTOSIS()**, où l'argument à introduire entre parenthèses est la plage de la série de données.

Cette équation n'est pas identique à celle introduite dans le cours théorique. Le principe reste cependant identique : si le kurtosis vaut 0, alors la distribution est similaire à la distribution normale (en termes de densité). Si le kurtosis est supérieur à 0, cela signifie que la distribution est plus pointue, avec des extrémités plus lourdes que la normale et enfin, si le kurtosis est inférieur à 0, cela signifie que la distribution est plus aplatie, avec des extrémités moins denses que la normale.

Créez-en B28 une formule qui permet de calculer le coefficient G2 de la variable $Score(X_i)$, en utilisant les fonctions conditionnelles :

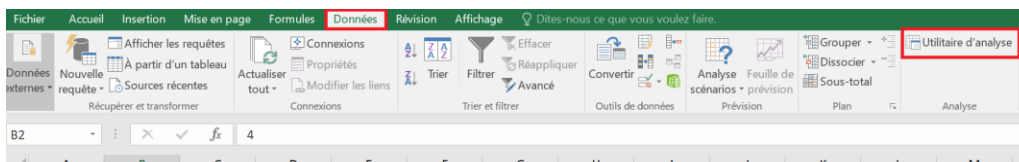
- Si la variable est de type quantitatif, la mesure de kurtosis est calculée
- Si la variable est de type qualitatif, une barre « / » est imprimée.

Affichez deux décimales et étirez la formule dans les cellules C28:E28.

Calculer toutes les statistiques descriptives usuelles d'une variable, en une seule fois

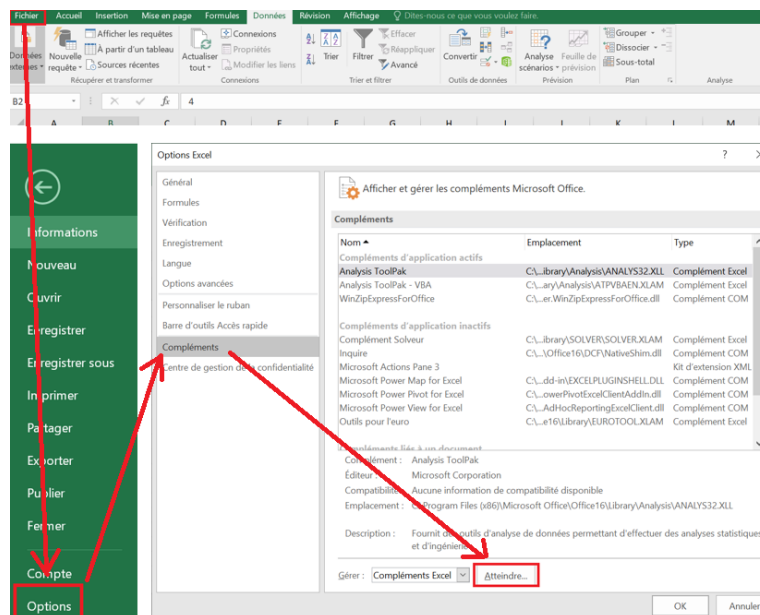
Excel propose un utilitaire d'analyse de données, qui permet entre autres de calculer automatiquement une série de paramètres descriptifs pour une ou plusieurs variables. Cet utilitaire présente cependant quelques limites, que nous exposerons après l'avoir utilisé.

Rendez-vous dans l'onglet « Données », groupe « Analyse », « Utilitaire d'analyse ».

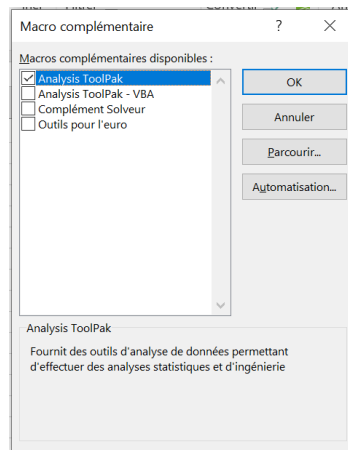


Il est possible que vous ne trouviez pas cette option. Si c'est le cas, cela signifie qu'elle n'est pas activée. Voici les étapes à suivre pour l'activer :

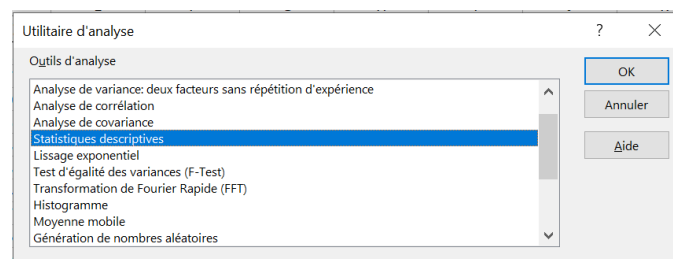
Rendez-vous dans l'onglet « Fichier », et cliquez sur « Options ». Dans l'interface qui s'ouvre, cliquez sur « Compléments ». En bas de la nouvelle interface, cliquez sur « Atteindre ».



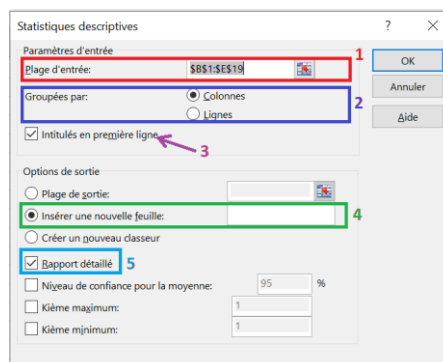
Dans les macros complémentaires disponibles, cochez « Analysis ToolPak », et cliquez sur OK.



À présent, l'utilitaire d'analyse apparaît dans l'onglet « Données. Cliquez dessus. Cochez l'option « Statistiques descriptives », puis cliquez sur OK.



- 1) Dans l'interface ouvrant, commencez par spécifier la plage d'entrée. Vous devez sélectionner toute la base de données, à l'exception de la variable *i*, soit la plage de données **B1:E19** (pensez au chapitre « Comment faire des sélections ? » de votre cours d'introduction à Excel, syllabus 1, niveau 1, p 15).
- 2) Spécifiez que les variables ont été encodées par « Colonnes » (cf. règle de base de l'encodage des données).
- 3) Cochez l'option « Intitulés en première ligne »
- 4) Précisez que vous voulez insérer les résultats dans une nouvelle feuille
- 5) Demandez un rapport détaillé



Cet outil présente de nombreuses limites, du moins dans l'usage que nous voulons en faire :

1. Il n'est pas dynamique : si les données de la feuille *Données brutes* sont modifiées, le compte rendu ne sera pas mis à jour.
2. Il ne fait pas de distinction entre les variables qualitatives et quantitatives. Lorsque les observations d'une variable sont des chiffres, des paramètres tels que la moyenne pourront toujours être calculés, même si ceux-ci ne sont pas intrinsèquement numériques (cf. variables qualitatives).
3. Il calcule par défaut la variance et l'écart-type corrigé (cf. Remarque plus haut, dans la section sur la variance). Idem pour le kurtosis.
4. Il donne une réponse chiffrée pour le mode, et non le nom de l'étiquette qui s'y rapporte.
5. Il ne fournit pas tous les paramètres désirés (il manque l'écart interquartile, et l'EMA).