

Conseils pour les travaux pratiques Microsoft Excel

Contents

Organiser les feuilles de calcul (sheets)	3
Renommer les feuilles de calcul.....	3
Changer la couleur de l'onglet pour la feuille de calcul.....	3
Insérer un nouvelle feuille de calcul	3
Supprimer un feuille de calcul.....	3
Pour formater le texte d'une cellule	3
Choisir un format désiré aux cellules calculées (Nombre/texte/date).....	3
Changer le nombre des décimales à montrer	4
Pour fusionner multiple cellules dans une seule	5
Ecrire le texte dans une cellule sur plusieurs lignes.....	5
Classification ascendante des données:	5
Les références relatives et absolues:	5
Les références relatives	5
Les références absolues	6
Formules et fonctions	7
Ecriture d'une formule donnée dans une feuille de calcul:	7
Multiplication d'une formule dans plusieurs cellules	7
Insérer des fonctions prédéfinis (ex. countif, average, min, max, ...)	7
Calcul les paramètres de statistiques descriptives:.....	11
Installation du module Data Analysis pour l'analyse statistique	13
Création des tableaux	13
Création des tableaux de fréquence (en utilisant COUNTIF)	13
Création des tableaux de fréquence (en utilisant PivotTable):.....	13
Graphiques - generalitees	15
La création d'un graphique.....	15
Changer la couleur des éléments graphiques (colonnes, lignes,...).....	16
Effacer des éléments graphiques (titre, légende, ...).....	16
Ajouter un titre a un graphique	16
Ajouter la définition de l'axes du graphique	16
Changer les marqueurs dans les graphiques Nouage des points et Lignes	16
Changer les intervalles des valeurs pour les axes d'un graphique (Scatter)	16
Effacer une série des données dans un graphique (Line (ligne), ...)	17
Ajoutez lez étiquetés sur l'axe pour un graphique de type Line (ligne).....	17
Montrer le nombre des sujets sur les colonnes (autre entités) d'un graphique.	18
Joindre les colonnes (effacer l'espace entre les colonnes).....	18
Montrer les pourcentages sur un graphique	18
Changer la position de la légende	18
Le choix du type du graphique en fonction des types des variables et but !!!	18
Description d'une seule variable	18
Qualitative	18
Quantitative	18
La relation entre deux variables	19
Qualitative	19
Quantitative	19
Une variable quantitative en fonction d'une variable qualitative.....	19

L'évolution dans le temps d'une variable qualitative ou quantitative	19
La relation entre trois variables quantitatives.....	19
Une variable qualitative en fonction des intervalles d'une variable quantitative	19
Création des graphiques (la technique)	19
Graphique d'une seule variable qualitative (en utilisant COUNTIF):	19
Graphique d'une seule variable qualitative (en utilisant PivotTable):	20
Graphique sectoriel (camembert – Pie)	22
Graphique des colonnes	22
Graphique pour la relation entre deux variables qualitatives :.....	22
Graphique pour une seule variable quantitative continue ou discrète.....	27
Création d'un histogramme (variante simple).....	27
Création d'un histogramme (variante difficile – si la variante simple ne marche pas) (sans préciser les classes) :	27
Création d'un histogramme (variante difficile – on ne le demande pas) ayant un certain nombre de classes :.....	28
Graphique des deux variables quantitatifs – nuage des points - scatter	29
Graphique boîte a moustaches (box and whiskers) – pour la relation entre une variable quantitative et une variable qualitative.....	33
Modifications des graphiques (ajouter des titres, légende ...)	35
Ajouter les titres des axes.....	35
Ajouter la légende	36
Pour enlever les boutons depuis un graphique réalisé avec PivotTable.....	37
Interprétations.....	38
Les types des variables:	38
Qualitatives (non métriques):	38
Quantitative (métriques).....	38
Variables de survie	38
L'évaluation de normalité des données en utilisant les indicateurs de tendance centrale, d'aplatissement et d'asymétrie.....	39
Interprétation des mesures de tendance centrale, emplacement, dispersion,	40
Interprétation du graphique nuage des points	42
Interprétation de la corrélation	43
Interprétation du coefficient de détermination	43
Interprétation des coefficients d'une équation de régression linéaire simple	44
Intervalles de confiance.....	44
L'interprétation de l'Intervalle de confiance:.....	44
Le calcul des intervalles de confiance	45
Notations :	45
Le calcul de l'écart type dans Excel	45
Formules de calcul	45
La moyenne d'une variable (caractéristique) quantitative	45
La différence des deux moyennes	45
Une fréquence relative d'une caractéristique qualitative: $np >= 10$ et $nq >= 10$	46
La différence des deux fréquences relatives.....	46
Tests statistiques.....	46
Les étapes d'un test statistique	46
Exemples des hypothèses statistiques	47
Exemple général	47
Exemple pour le test t (Student) pour échantillons indépendants	48
Exemple pour le test t (Student) pour échantillons dépendants (appariées).....	48

Exemple pour le test F pour les variances pour échantillons indépendants	48
Exemple pour le test Khi-deux, ou Fisher exact	48
Exemple pour le test statistique pour le coefficient de corrélation Pearson ou Spearman	50
Exemple pour le test statistique pour le pente de la régression linéaire.....	50
Exemples de créations des hypothèses statistiques avec le choix des hypothèses en fonction du p.....	50
P<0.05	50
P>0.05	51
Comment faire le choix entre des tests statistiques avec échantillons indépendants ou dépendants (appariées)	51
Le choix du test t - Student.....	51
Réaliser les tests statistiques dans Excel avec Data Analysis	52
Le test Student - Pour échantillons indépendants.....	52
Le test Student - Pour échantillons dépendants.....	53
Le test F pour comparer les variances	54
Réaliser le test Khi-deux dans Excel.....	55
Probabilités.....	56
Théorie	56
Probabilités conditionnels - Risque relatif (RR)	57
Interpretation	57
Probabilités conditionnels – Tests diagnostiques.....	57

Organiser les feuilles de calcul (sheets)

Renommer les feuilles de calcul

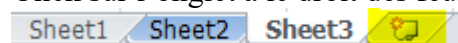
Double click sur l'onglet de la feuille de calcul ou click droit sur l'onglet de la feuille de calcul – **Rename** (renommer).

Changer la couleur de l'onglet pour la feuille de calcul

Click droit sur l'onglet de la feuille de calcul d'intérêt, et dans le menu **Tab color** vous pouvez choisir la couleur désirée

Insérer un nouvelle feuille de calcul

Click sur l'onglet a le droit des feuilles des calculs existantes



Supprimer un feuille de calcul

Click droit sur l'onglet de la feuille de calcul d'intérêt et sélectionnez **Delete** (effacer).

Pour formater le texte d'une cellule

Voir comment se fait dans Word !

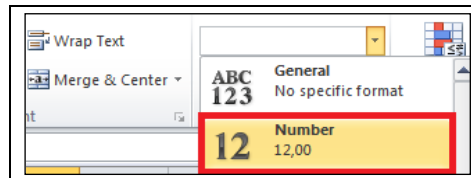
Choisir un format désiré aux cellules calculées (Nombre/texte/date)

S'il vous plaît de ne pas sélectionner la colonne entière en cliquant sur la lettre de la colonne pour changer le type, parce que ça peut nuire a des calculs suivants

S'il vous plaid de sélectionner seulement les cellules avec les valeurs d'intérêt, sans le titre ou des cellules vides.

Sélectionnez la/les cellules à changer le type du contenu, click droit, **Format Cells**, dans l'onglet **Number** (nombre), il y a une liste l'options **Category**, choisissez une de Number (nombre), Date (date), Time (temps), Text (texte) ou Scientific (Scientifique).

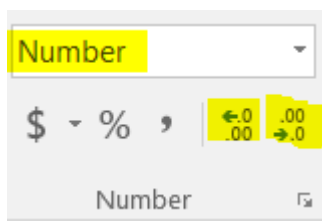
Ou dans l'onglet **Home**, section **Number** utilisez le menu pour choisir le type des cellules selectionees.



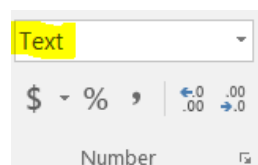
Changer le nombre des décimales à montrer

Sélectionnez la/les cellules à changer le type du contenu, click droit, **Format Cells**, dans l'onglet **Number** (nombre), il y a une liste l'options **Category**, choisissez **Number** (nombre). Puis dans **Decimal places** (nombre des décimales), écrivez la valeur désirée.

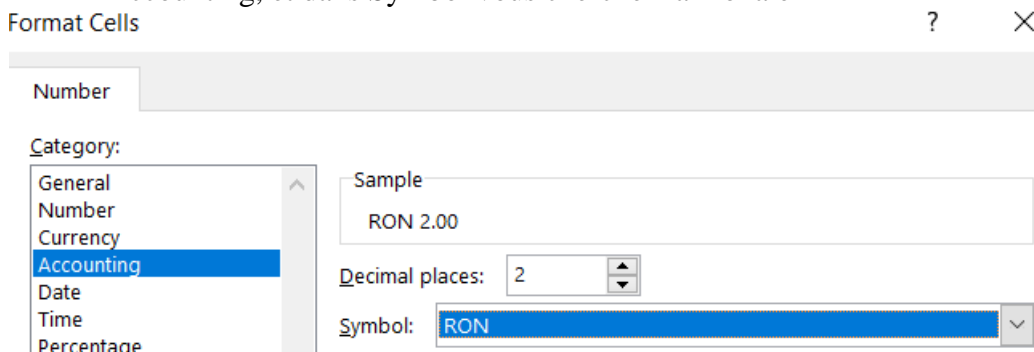
Ou dans l'onglet **Home**, section **Number** (nombre), click sur les images avec flèches pour des décimales.

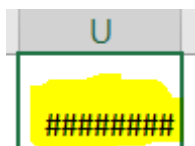


Avec les fleches on peut changer le nombre des decimales



- Pour **Texte**:
- Pour la **Monaie** (RON, EUR, USD, ...) – Click droit - Format Cells – Number – Accounting, et dans Symbol vous cherchez la monaie

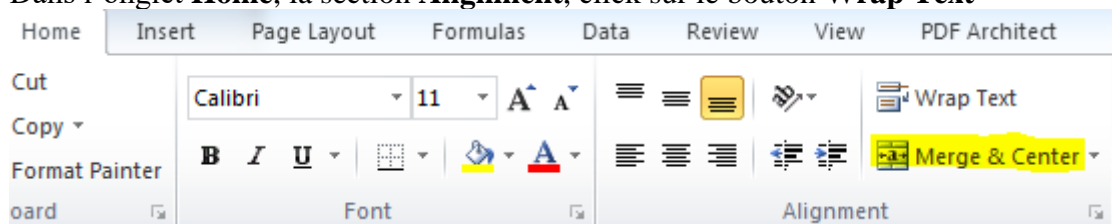




Si vous observez : ##### - ca n'est pas une erreur, mais Excel vous montre que les valeurs numériques ne sont pas bien visibles, et vous devez élargir la colonne pour les voir correctement.

Pour fusionner multiple cellules dans une seule

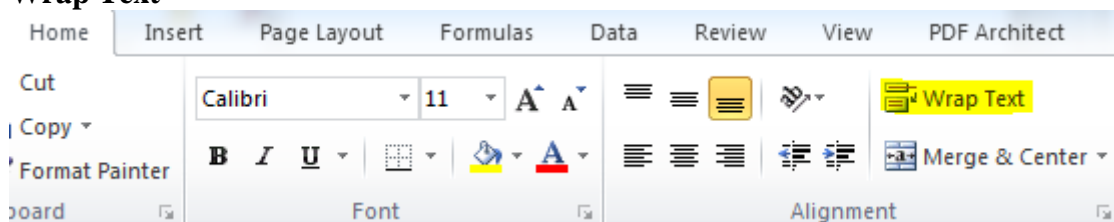
Sélectionnez les cellules désires (qui doivent être fusionné), puis Dans l'onglet **Home**, la section **Alignment**, click sur le bouton **Wrap Text**



Ou, click droit et choisissez l'options :
Format Cells... - Alignment – Text control – Merge cells

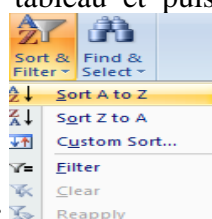
Ecrire le texte dans une cellule sur plusieurs lignes

Sélectionnez la cellule et dans l'onglet **Home**, la section **Alignment**, click sur le bouton **Wrap Text**



Classification ascendante des données:

- pour trier des données, sélectionnez tous le tableau et puis choisissez dans le tab



Editing, vous trouverez l'option **Sort&Filter**

- choisissez les variables désirées comme la première clé et deuxième clé ;

Les références relatives et absolues:

Les références relatives

- Une référence est nommée **relative** si en copiant une formule qui le contienne dans une autre cellule elle va changer en fonction de la nouvelle position (relative au nouveau ligne ou colonne).
- **Format** : Les références relatives contient seulement le nom de la ligne et de la colonne qui identifie la cellule (ex. C3, A7)

- Relativité en fonction des colonnes. Par **exemple**, si vous copiez une référence relative dans la cellule J3 = $(C3+D3)/2$ vers la cellule J4, la formule est automatiquement transformée en $(C4+D4)/2$. Ici on observe que les références des cellules dans les formules sont identiques du point de vue du nom de la colonne (J), mais différente du point de vue du nom de la ligne (3 et 4). Donc la formule dans J4 aura les références avec les mêmes noms des colonnes que les références dans la formule originale (J3), mais le nombre des lignes va changer à 4 (comme le nom de ligne pour le J4). **Essayez !**
- Relativité en fonction des lignes. Par **exemple**, si vous copiez une référence relative dans la cellule J3 = $(C3+D3)/2$ vers la cellule K3, la formule est automatiquement transformée en $(D3+E3)/2$. Ici on observe que les références des cellules dans les formules sont identiques du point de vue du nom de la ligne (3), mais différente du point de vue du nom de la colonne (J et K). Donc la formule dans K3 aura les références avec les mêmes noms des lignes que les références dans la formule originale (J3), mais le nom des colonnes va changer dans la lettre suivante dans l'alphabet (le C devient D, et le D devient E – parce que entre les références des formules J3 et K3 il y a une lettre différence entre les colonnes). **Essayez !** si vous copiez le J3 dans le L3 la formule deviendra $(E3+F3)/2$ (donc deux lettres différence)
- Relativité en fonction des lignes et colonnes. Par **exemple**, si vous copiez une référence relative dans la cellule J3 = $(C3+D3)/2$ vers la cellule K4, la formule est automatiquement transformée en $(D4+E4)/2$. Ici on observe que les références des cellules dans les formules ne sont pas identiques du point de vue du nom de la ligne (3 et 4), et aussi du point de vue du nom de la colonne (J et K). **Essayez !**

Les références absolues

- Une référence est nommée **absolue** si en copiant une formule qui le contienne dans une autre cellule elle ne va pas changer d'une façon relative sélective pour les lignes ou pour les colonnes ou pour les deux,
- **Format** : Les références absolues contiennent seulement le nom de la ligne et de la colonne qui identifie la cellule (ex. absolues pour les lignes : C\$3, A\$7 ou absolues pour les colonnes \$C3, \$A7, ou absolues pour les lignes et colonnes \$C\$3, \$A\$7)
- Absolue en fonction des lignes. Par **exemple**, si vous copiez une référence relative dans la cellule M3 = $(E$3+F$3)/2$ vers la cellule M4, la formule est restée identique $(E$3+F$3)/2$, parce que le nom des lignes est fixé avec le signe \$. **Essayez !** Mais si vous copiez la cellule M3 dans la cellule N4 elle deviendra $(F$3+G$3)/2$ parce que le nom des colonnes n'est pas fixé (absolue) avec le signe \$ - il est donc relative (voir les références relatives). **Essayez !**
- Absolue en fonction des colonnes. Par **exemple**, si vous copiez une référence relative dans la cellule M5 = $(E$5+$F5)/2$ vers la cellule N5, la formule est restée identique $(E$5+$F5)/2$ parce que le nom des colonnes est fixé avec le signe \$. **Essayez !** Mais si vous copiez la cellule M5 dans la cellule M6 elle deviendra $(E$6+$F6)/2$ parce que le nom des lignes n'est pas fixé (absolue) avec le signe \$ - il est donc relative (voir les références relatives). **Essayez !**
- Absolue en fonction des lignes et colonnes. Par **exemple**, si vous copiez une référence relative dans la cellule M7 = $(E$7+F7)/2$ vers n'importe quelle

cellule, la formule est reste identique = $(\$E\$7+\$F\$7)/2$ parce que le nom des colonnes et des lignes est fixée avec le signe \$. **Essayez !**

Formules et fonctions

Ecriture d'une formule donnée dans une feuille de calcul:

- choisissez une nouvelle colonne et renommez-la avec un nom désiré (par exemple, « **Besoins énergétiques** »)
- **appliquez la formule en commençant par le symbole d'égalité (=) ;**
- n'oubliez pas de vérifier si votre variables ont les mêmes unités de mesure que les unités utilisées dans la formule donnée ;
- écrivez la formule en utilisant les opérateurs arithmétiques suivants:
 - addition : **+**
 - soustraction : **-**
 - multiplication : *****
 - division : **/**
 - exponentiation : **^**
 - racine carrée : **^(1/2)**
- pour indiquer l'ordre d'opérations, on utilise toujours les parenthèses **rondes ()** ;


Multiplication d'une formule dans plusieurs cellules

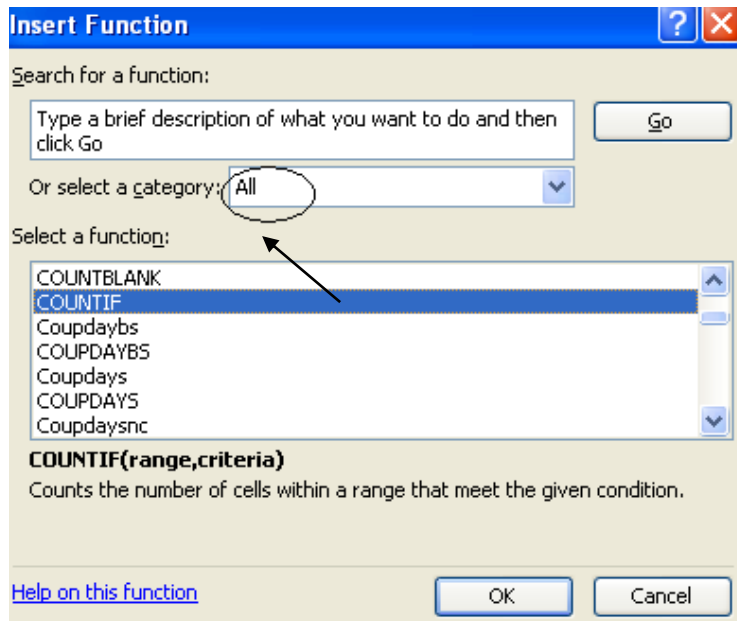
Insérez la formule d'intérêt. Puis sélectionnez cette cellule. Une cellule sélectionnée a un contour foncé, et un petit carré noir dans la partie inférieure droite. Faites click sur ce carré, et tirez le souris en sélectionnant les cellules dans lesquels vous voulez multiplier la formule.

Insérer des fonctions prédéfinies (ex. countif, average, min, max, ...)

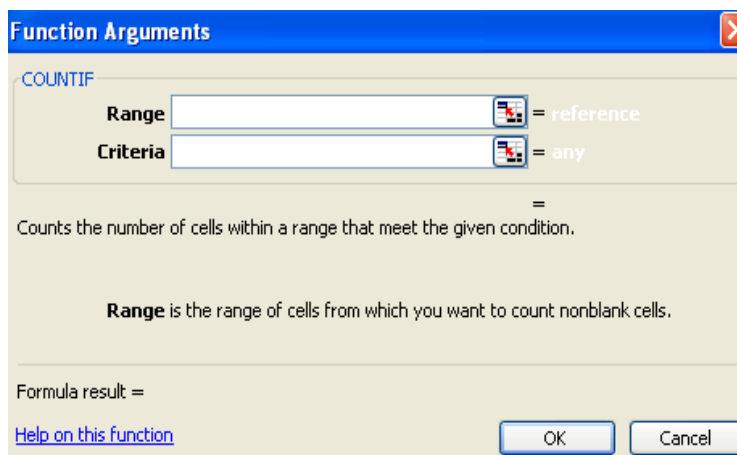
Comment insérer une fonction prédéfinie

Par exemple pour la fonction COUNTIF

- **Formulas → Insert Functions**  la fonction **COUNTIF** :



- la syntaxe de la fonction COUNTIF a deux arguments : le domaine et critère
- Le domaine = domaine des cellules à partir duquel vous souhaitez compter les cellules de la variables désiré ;
- Le critère = ce que vous souhaitez compter;



- donc, comme « range » on va sélectionner les cellules où vous trouvez les catégories de la variable de l'obésité et comme le « critère », on va sélectionner une seule cellule où se trouve un sujet obèse.

Les noms des fonctions les plus importantes (moyenne, écart type, ...)

Mesure	Fonction dans Excel en Anglais	En Français
Moyenne	AVERAGE	MOYENNE
Médiane (= quartile 2)	MEDIAN	MEDIANE
Mode	MODE	MODE
Valeur centrale	(Maximum + minimum)/2 – voir les fonctions en bas	
Amplitude	Maximum - minimum	
Variance	VAR	VAR


d'échantillonnage		
Moyenne de la déviation (valeur absolue)	A VEDEV	ECART.MOYEN
Déviation standard/écart type	STDEV	ECARTYPE
Erreur standard	$ES = \frac{s}{\sqrt{n}}$ ou s = la déviation standard, et n est la taille de l'échantillon	
Coefficient de variation	=déviation standard/moyenne (STDEV/AVERAGE)	
Minimum	MIN	MIN
Quartile 1	= QUARTILE (intervalle,1)	= QUARTILE (intervalle,1)
Quartile 2 (= médiane)	= QUARTILE (intervalle,2)	= QUARTILE (intervalle,2)
Quartile 3	= QUARTILE (intervalle,3)	= QUARTILE (intervalle,3)
Maximum	MAX	MAX
Percentile	= PERCENTILE (intervalle, percentile)	= CENTILE (intervalle, percentile)
Intervalle interquartile	La différence entre le quartile 3 et le quartile 1	
Coefficient d'aplatissement	KURT	KURTOSIS
Coefficient d'asymétrie	SKEW	COEFFICIENT.ASYMETRIE
Si	IF (condition, valeur si condition est vraie, valeur s'il e fausse)	SI
Ou	OR (condition 1, condition 2, ...)	OU
Et	AND (condition 1, condition 2, ...)	ET
Compter	COUNT – compter le nombre des cellules avec des valeurs numériques – pour trouver la taille de l'échantillon	COMPTE
Compter si	COUNTIF (intervalle des cellules, critère pour compter)	NB.SI
La somme	SUM	SOMME
La racine carrée	SQRT	RACINE

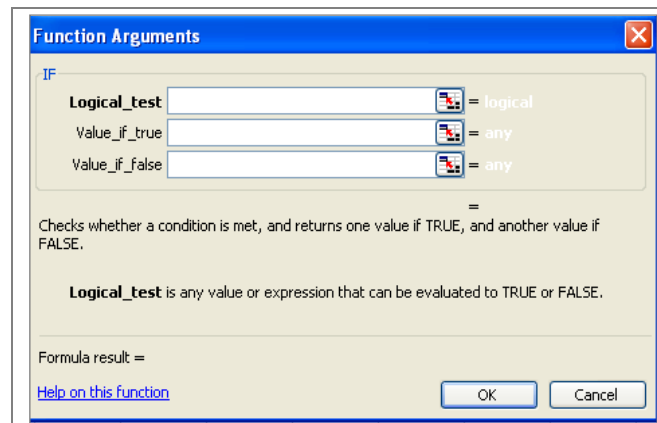
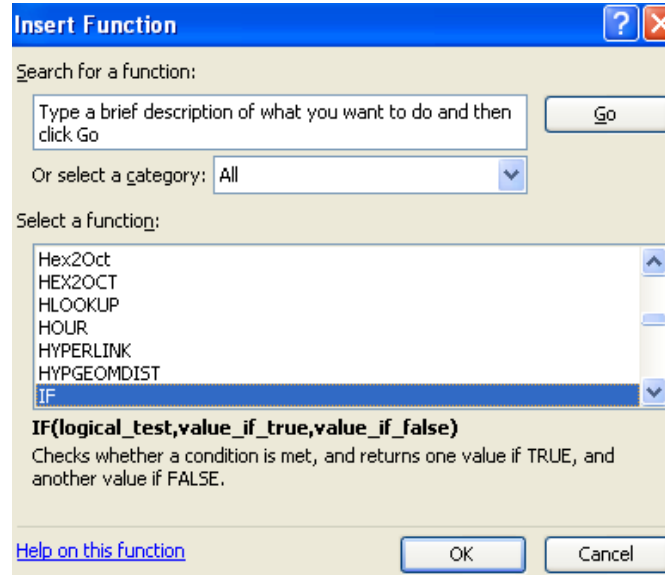
La fonction IF (si) pour vérifier une condition

- vérifiez la condition spécifiée dans le problème en utilisant la fonction **IF(SI)**;
- cette fonctionne **IF(Si)** retourne une valeur si la condition spécifiée est vraie et une autre valeur si la condition est fausse ;
- **la syntaxe de la fonction IF** a trois arguments: **logical_test, value_if_true, value_if_false**

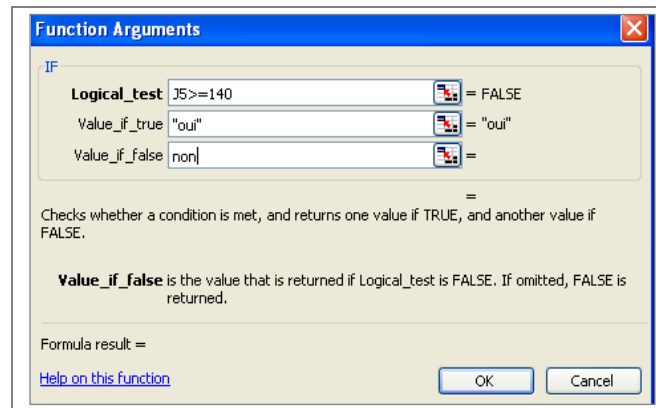
Logical_test est la condition spécifiée dans le problème ;

Value_if_true est la valeur qui est obtenue si la condition est vraie.
Value_if_false est la valeur qui est obtenue si condition est fausse.

- cherchez dans le tab **Formulas** → **Insert Functions**  la fonction **IF** ;



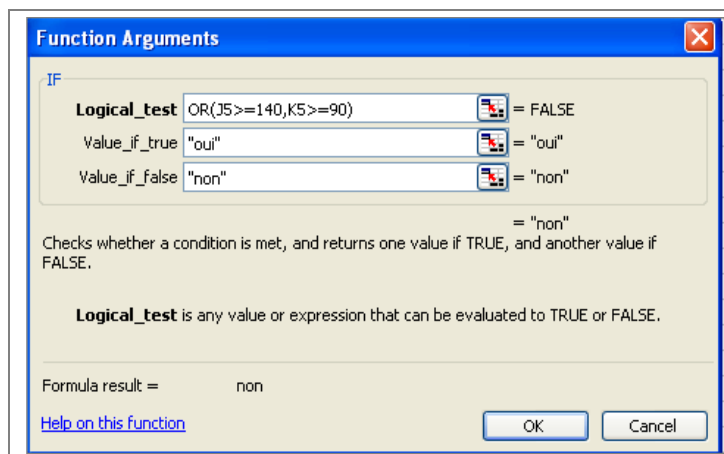
- **Logical test**= la référence de la cellule contenant la valeur d'une variable pour un sujet qui est plus grande, plus petite ou égale a une certain valeur.
 - A2>200, B3<160, D7 = "oui"**Value_if_true** = Oui (ou autre catégorie) est la réponse attendue si la condition est vraie ;
Value_if_false =Non (ou autre catégorie) est la réponse attendue si la condition est fausse ;



Pour des conditions logiques plus complexe vous pouvez utiliser les fonctions OR (OU), AND (ET).

La fonction **OR** (OU) pour la situation dans laquelle deux ou plus des conditions sont reliées entre elles avec l'opérateur logique ou. La syntaxe de la fonction est OR(condition 1, condition 2, condition 3, ...) Voir suivant

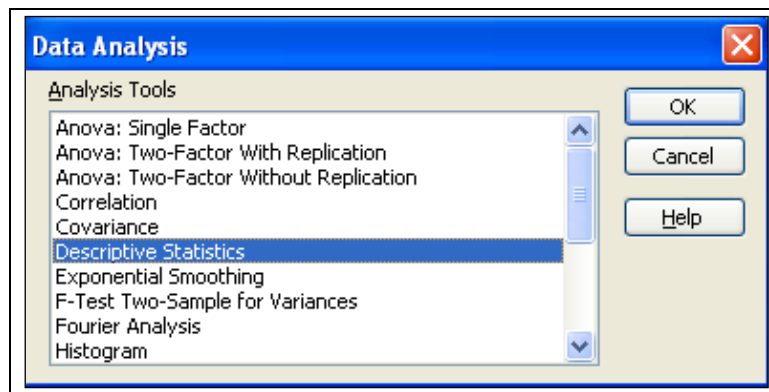
La fonction **AND** (ET) pour la situation dans laquelle deux ou plus des conditions sont reliées entre elles avec l'opérateur logique et. La syntaxe de la fonction est AND(condition 1, condition 2, condition 3, ...)



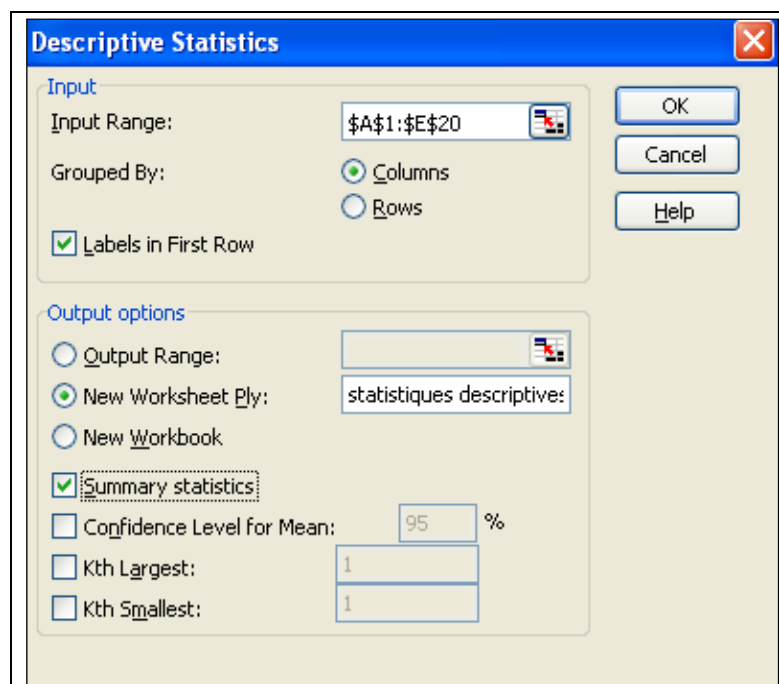
Calcul les paramètres de statistiques descriptives:

Calcule simultanément de statistiques descriptives pour toutes les variables quantitatives continues;

- utilisez le tab Data-Data Analysis ;
- si l'option **Data Analysis** n'apparaît pas installer le (voir le tp antérieur) ;
- dans la boîte de dialogue de Data Analysis, sélectionnez les statistiques descriptives et cliquez sur OK.



- le calcul des statistiques descriptives se fait en suivant l'exemple indiqué ci-dessous:



Note :

Input Range: =les références des cellules des variables dont les statistiques descriptives doivent être analysées.

Grouped By : indique si les données introduites sont organisées en lignes ou en colonnes.

Labels in First Row/Labels in First Column (Labels en première file / Labels dans la première colonne) : Si la première ligne de votre plage d'entrée contient des étiquettes ou les noms des variables, sélectionnez les « Labels in First Row ».

Si vous cochez :


Output Range (Plage de sortie) : les résultats du calcul seront affichés sur la même page du fichier dans une cellule vide sélectionnée ;

New Worksheet Ply (Nouvelle feuille) : une nouvelle feuille des résultats

New Workbook (Nouveau fichier d'Excel) : un nouveau fichier d'Excel des résultats

Summary statistics (Sommaire des statistiques) : les plus importantes statistiques ;

Installation du module Data Analysis pour l'analyse statistique

- l'installation d'analyse des données (**Data Analysis**) se fait à l'aide des options Excel dans le menu File -  -Add Ins-Manage Excel Add-Ins – Click sur Go. Dans la fenêtre, cochez l'option: Analysis ToolPack-OK;
- dans le tab **Data** se trouve le **Data Analysis** ;

Création des tableaux

Création des tableaux de fréquence (en utilisant COUNTIF)

- Un exemple de tableau de fréquences pour la variable Paresthésie est la suivante :

Paresthésie	Nombre
oui	67
non	12
Total	79

- Vous commencez d'écrire dans des cellules vides Paresthésie, oui, non, total et Nombre

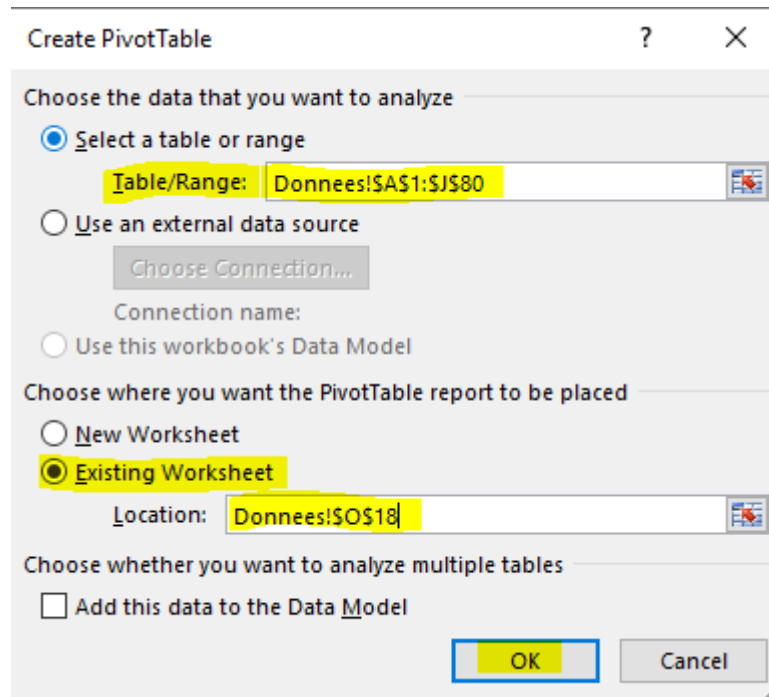
Paresthésie	Nombre
oui	
non	
Total	

- Utilisez la fonction **COUNTIF** (voir avant) pour calculez les fréquences absolues (nombre de sujets avec oui/non dans cet exemple); Vous devez sélectionner les cellules de la colonne avec les données concernant la Paresthésie, et comme critère oui, respectivement non.
- Utilisez la fonction **SUM** pour calculer le total.

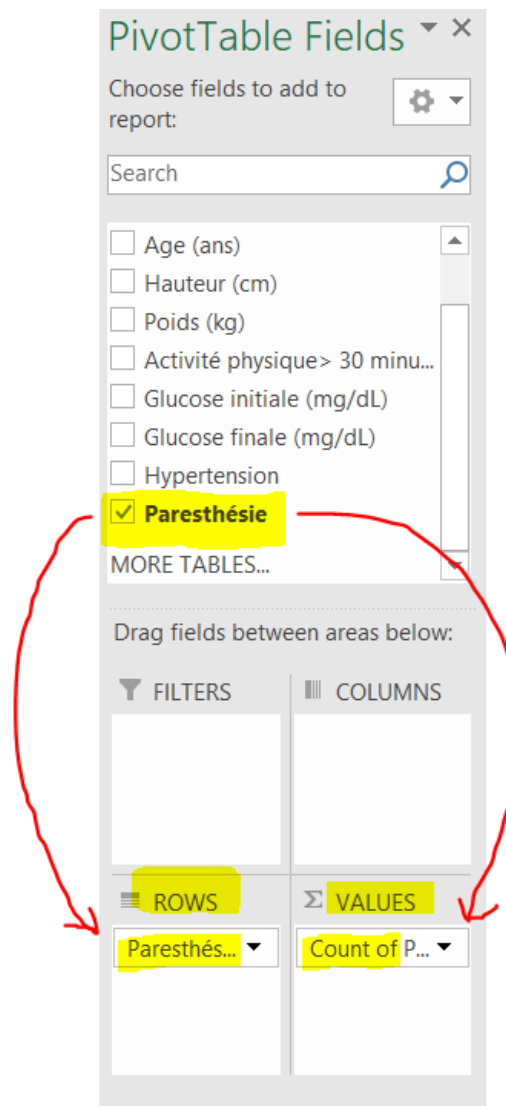
Création des tableaux de fréquence (en utilisant PivotTable):

- **Sélectionnez toutes les données** incluant le titre des colonnes

- Insérez depuis le tab **Insert** , **Tables**, click sur 



- Dans la fenetre dans **Table/Range** vous devez observer que toutes les donnees ont ete selectionne, sinon vous pouvez les selectionner ici.
- Vous pouvez mettre le tableau de contingence sur une nouvelle feuille de calcul en selectionnant **New Worksheet**
- Ou vous pouvez mettre le tableau de contingence sur la meme feuille de calcul en selectionnant **Existing Worksheet**. Puis dans **Location** vous devez selectionner une cellule pour la position du tableau sur la feuille.
- A la fin – click sur le bouton **OK**



- Déplacez la variable sur les lignes - **ROWS**
- Déplacez la variable, encore une fois dans la boîte **Values**
- Vérifiez qu'il a compte les cas (dans la boîte **Values** il doit écrire **Count of ...**)
- Vous allez obtenir un **tableau des fréquences**:

Row Labels	Count of Paresthésie
non	67
oui	12
Grand Total	79

- Parfois on a besoin de déplacer la ligne Oui dans la première ligne (ex le facteur de risque présent sur la première ligne et le facteur de risque absent sur la deuxième) – **click droit** sur Oui, puis sélection **Move**, et **Move “Oui” to Beginning**

Graphiques - generalitees

La création d'un graphique

Pour créer une graphique, on doit sélectionner le tableau qui contient les données, en incluant les nommes des lignes et des colonnes. Puis dans l'onglet **Insert** (Insérer), la section

Charts (Graphiques), click sur le bouton type de graphique qui correspond à votre type des données, à votre but, et votre choix esthétique.

Changer la couleur des éléments graphiques (colonnes, lignes,...)

Pour changer les couleurs sélectionnez une seule barre/colonne.. (click sur une, et après encore un click), puis click droit et sélectionnez dans la barre d'outils le bouton **Shape fill** (remplir avec couleur) et enfin click sur la couleur que vous désirez.

Effacer des éléments graphiques (titre, légende, ...)

Pour effacer le titre, ou la légende – on sélection l'objet avec le souris et on frappe la touche **Del** (effacer).

Ajouter un titre à un graphique

Pour ajouter un titre à un graphique, sélectionnez le graphique, puis dans l'onglet **Chart tools** (des outils graphiques), l'onglet **Layout** (disposition), vous pouvez ajouter le titre avec **Chart title** (le titre du graphique), en sélectionnant **Above chart** ou **Centered overlay title**.

Ajouter la définition de l'axes du graphique

Pour ajouter la définition de l'axe horizontale, sélectionnez le graphique, puis dans l'onglet **Chart tools** (des outils graphiques), l'onglet **Layout** (disposition), **Axis Titles** (les titres des axes), **Primary horizontal axis** (l'axe horizontale primaire), **Title below axis** (le titre en dessous du graphique).

Pour ajouter la définition de l'axe verticale, sélectionnez le graphique, puis dans l'onglet **Chart tools** (des outils graphiques), l'onglet **Layout** (disposition), **Axis Titles** (les titres des axes), **Primary vertical axis** (l'axe verticale primaire), l'une des suivantes options **Rotated title** (le titre avec rotation – le plus utilisée), **Vertical title** (le titre est mis vertical), et **Horizontal title** (le titre est mis horizontal).

Changer les marqueurs dans les graphiques Nouage des points et Lignes

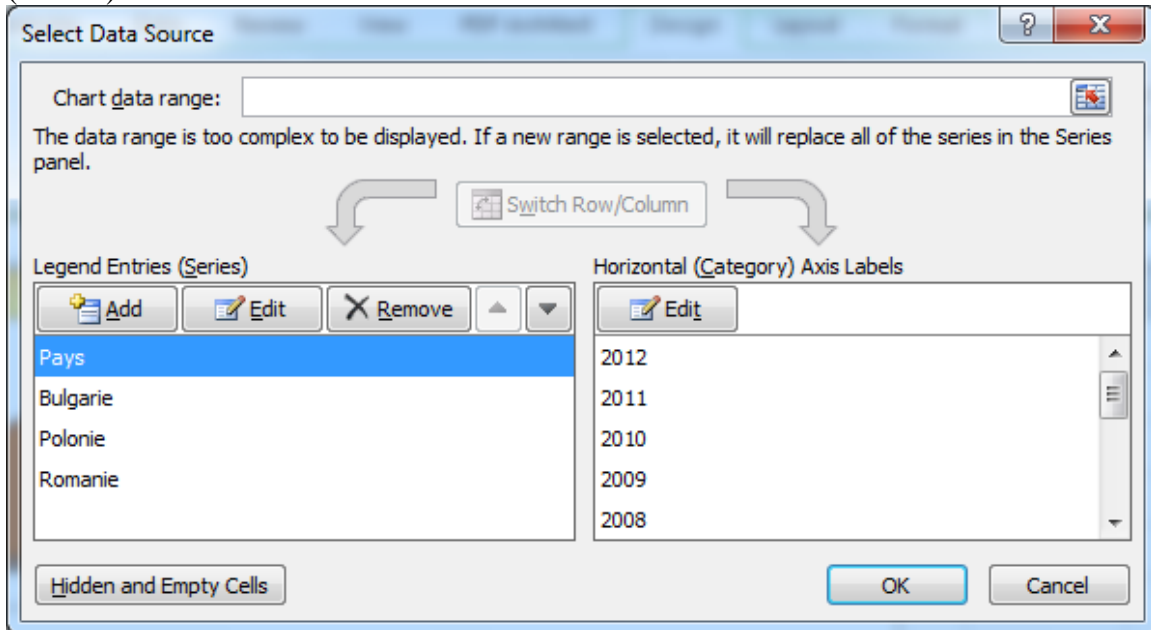
Pour changer les marqueurs dans les graphiques Nouage des points et Lignes, sélectionnez toutes les marqueurs avec un seul click sur un marqueur. Puis avec un click droit sélectionnez **Format Data Series** (formatez la série des données). Dans la fenêtre qui s'ouvre dans la section Marker Options (les options des marqueurs) pour **Marker type** (type de marqueur), sélectionnez **Built-in** (disponible).

Changer les intervalles des valeurs pour les axes d'un graphique (Scatter)

Pour changer les intervalles des valeurs pour les axes d'un graphique, comme un nouage des points, sélectionnez le graphique, et dans l'onglet Chart Tools, puis Layouts, dans la section **Axes**, click sur le bouton **Axes** et sélectionnez l'axe d'intérêt : **Primary Horizontal Axis** (l'axe horizontale) ou **Primary Vertical Axis** (l'axe verticale). Dans le menu sélectionnez **More Primary Horizontal/Vertical Options**. Une nouvelle fenêtre va apparaître. Dans cette fenêtre, dans la section **Axis Options**, changez le **Minimum**, ou le **Maximum**, dans **fixed** (fixe) au lieu de **Auto** (automatique). Puis écrivez la valeur pour bien contenir les valeurs dans le graphique.

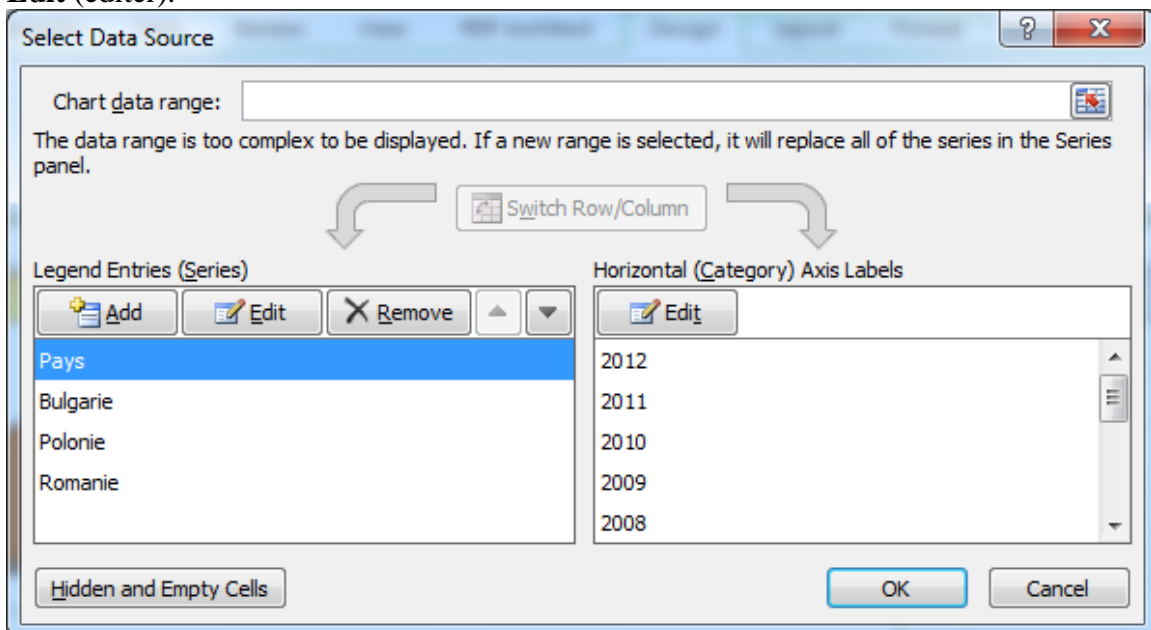
Effacer une série des données dans un graphique (Line (ligne), ...).

Pour effacer une série des données dans un graphique, sélectionnez le graphique, click droit, et dans le menu contextuel sélectionnez **Select Data Source** (sélectionnez la source des données). Dans la fenêtre qui s'ouvre dans la liste **Legend Entries (Series)** (les séries des données), sélectionnez la série des données à éliminer et cliquez sur le bouton **Remove** (effacer).



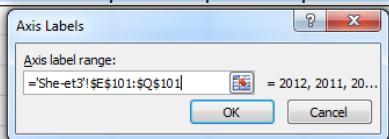
Ajoutez les étiquetés sur l'axe pour un graphique de type Line (ligne)

Pour afficher les étiquetés sur l'axe pour un graphique de type Line (ligne), sélectionnez le graphique, click droit, et dans le menu contextuel sélectionnez **Select Data Source** (sélectionnez la source des données). Dans la fenêtre qui s'ouvre dans la liste **Horizontal (Category) Axis Labels** (les étiquetés de l'axe horizontale), cliquez sur le bouton **Edit** (éditer).



Une nouvelle fenêtre vous permet de sélectionner les cellules qui contient les étiquetés de l'axe.

2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001
18	41	39	44	58	88	247	1968	474	252	805	1655
6263	0	4196	7586	4598	22890	20668	7946	4857	10588	40518	84419
20812	3494	350	343	1746	2958	3563	6801	47444	120377	11079	5076



Montrer le nombre des sujets sur les colonnes (autre entités) d'un graphique.

Pour montrer le nombre des sujets sur les colonnes d'un graphique, on sélectionne la colonne d'intérêt (deux click sur la colonne), ou on sélectionne toutes les colonnes (un seul click), et on fait un click droit. Dans le menu on choisit **Add data labels** (ajouter les étiquetés)

Joindre les colonnes (effacer l'espace entre les colonnes)

Pour joindre les colonnes dans un graphique de type colonnes, on fait un click sur une des colonnes pour sélectionner toutes les colonnes. Puis avec un click droit on sélectionne **Format data series** (formater la série des données). Une nouvelle fenêtre s'ouvre, dans laquelle, dans la première section, **Series options** (les options du série), on met **Gap width** (espace entre les colonnes) à 0%.

Montrer les pourcentages sur un graphique

Pour montrer les pourcentages sur le graphique, on doit changer le type du graphique après on a sélectionné le graphique, dans la section **Chart Layouts**, par exemple l'option **100% Stacked Bar**.

Changer la position de la légende

Pour changer la position de la légende – au droit, en dessous du graphique, on peut suivre la méthode automatique, après on a sélectionné le graphique, dans la section **Chart Layouts**, on doit identifier une graphique qui répond a notre besoins. Après on a fait la choix on peut déplacer la légende avec le souris.

Le choix du type du graphique en fonction des types des variables et but !!!

Description d'une seule variable

Qualitative

- camembert (sectoriel – **Pie**)
- **Colonnes (Column)** (si les noms des catégories ne sont pas très longues)
- **Barre (Bar)** (si les noms des catégories sont très longues)

Quantitative

- **Histogramme**, polygone des fréquences, graphique des moyennes, box and whiskers (boite a moustaches), graphique des quantiles

La relation entre deux variables

Qualitative

- **Column** (Clustered Column/ Stacked Column/ 100% Stacked column),
ou **Bar** (Clustered Bar / Stacked Bar / 100% Stacked Bar)

Quantitative

- **Scatter** (nouage des points)

Une variable quantitative en fonction d'une variable qualitative

Si les donnees sont normale distribuees

- Graphique des moyennes (aved deviation standard)
- Graphique Colonnes avec barre d' erreur

Si les donnees sont normale distribuees

- Graphique box plot ou wiskers/boite a moustaches (boite a moustaches)

L'évolution dans le temps d'une variable qualitative ou quantitative

- **Line** (Ligne)

La relation entre trois variables quantitatives

- **Bubble** (nouage des spheres)
- Nouage des points tridimensionel

Une variable qualitative en fonction des intervalles d'une variable quantitative

- **Area** (Surface)

Création des graphiques (la technique)




Graphique d'une seule variable qualitative (en utilisant COUNTIF):

- Vous avez besoin d'un tableau de fréquences de la forme suivante :

Paresthésie	Nombre
oui	67
non	12
Total	79

- Utilisez la fonction **COUNTIF** pour calculez les fréquences absolues (nombre de sujets avec oui/non dans cet exemple);
- Sélectionnez les noms des catégories avec les fréquences absolues seulement

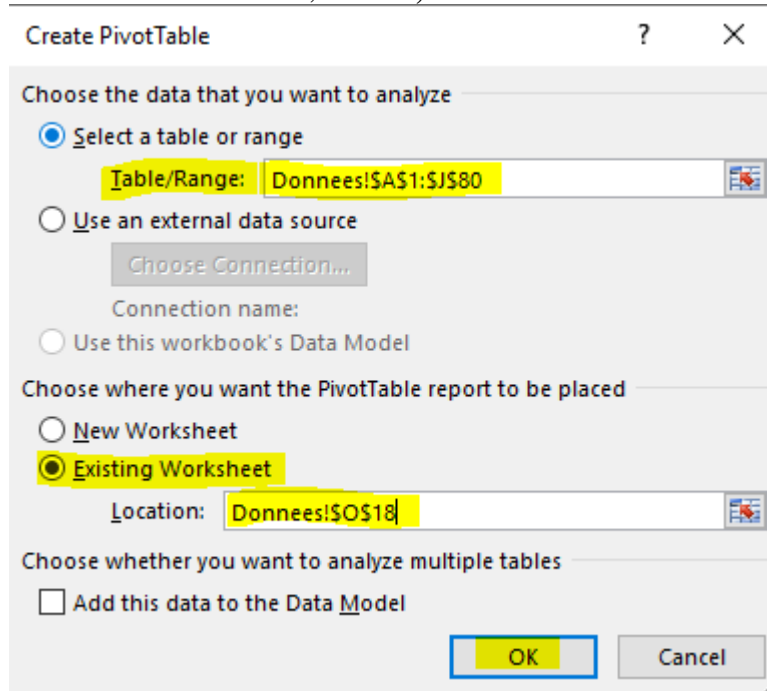
Paresthésie	Nombre
oui	67
non	12
Total	79

- puis dans l'onglet **Insert** , section **Charts**, click sur  pour insérer des graphiques de type Colonnes (**Column**) ou Barres (**Bar**), ou  pour insérer des graphiques de type camembert (**Pie**)

Graphique d'une seule variable qualitative (en utilisant PivotTable):

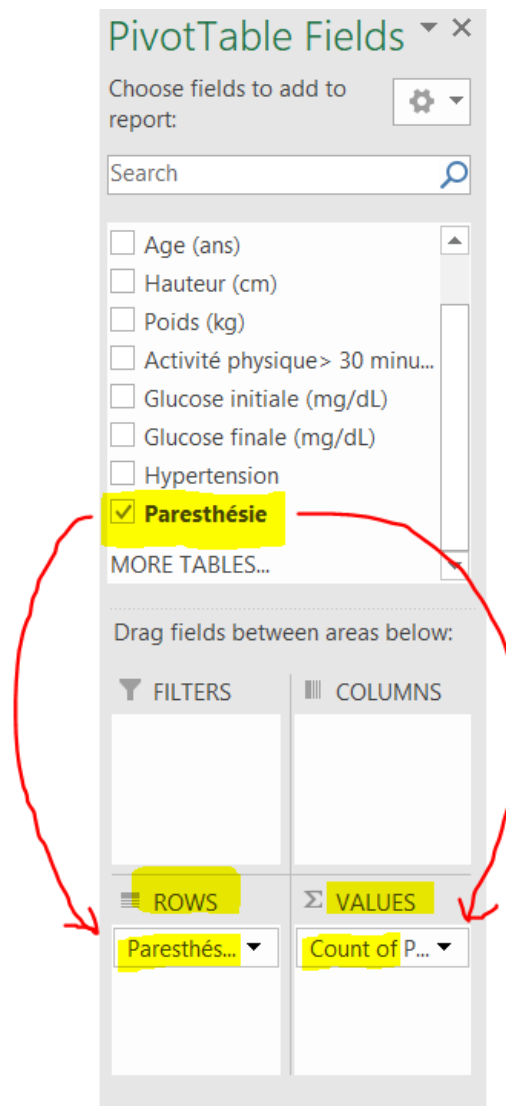
- Sélectionnez toutes les données incluant le titre des colonnes

- Insérez depuis le tab **Insert** , **Tables**, click sur 



■




- Dans la fenêtre dans **Table/Range** vous devez observer que toutes les données ont été sélectionnées, sinon vous pouvez les sélectionner ici.
 - Vous pouvez mettre le tableau de contingence sur une nouvelle feuille de calcul en sélectionnant **New Worksheet**
 - Ou vous pouvez mettre le tableau de contingence sur la même feuille de calcul en sélectionnant **Existing Worksheet**. Puis dans **Location** vous devez sélectionner une cellule pour la position du tableau sur la feuille.
 - A la fin – click sur le bouton **OK**
-





- Déplacez la variable sur les lignes - **ROWS**
- Déplacez la variable, encore une fois dans la boîte **Values**
- Vérifiez qu'il a compte les cas (dans la boîte **Values** il doit écrire **Count of ...**)
- Vous allez obtenir un **tableau des fréquences**:

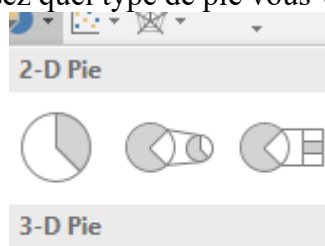
Row Labels	Count of Paresthésie
non	67
oui	12
Grand Total	79


- Parfois on a besoin de déplacer la ligne Oui dans la première ligne (ex le facteur de risque présent sur la première ligne et le facteur de risque absent sur la deuxième) – **click droit** sur Oui, puis sélection **Move**, et **Move “Oui” to Beginning**
- puis **insérez un graphique des colonnes ou des barres, ou un graphique camembert**
 - pour cela vous devez faire un click à l'intérieur du tableau de contingence antérieur créé

- puis dans l'onglet **Insert**  , section **Charts**, click sur  pour insérer des graphiques de type Colonnes (**Column**) ou Barres (**Bar**), ou  pour insérer des graphiques de type camembert (**Pie**)

Graphique sectoriel (camembert – Pie)

- Sélectionnez les données du tableau des fréquences antérieur créé – voir avant (les 4 ou plus cellules du votre tableau sans le titre du tableau – seulement les noms des catégories et les nombres des sujets dans chaque catégorie) ;
- cherchez dans l'onglet **Insert**  , section **Charts**, click sur  ;
- choisissez quel type de pie vous voulez et puis appuyez sur l'option



- Chart Layout :  ; On préfère les représentations **2-D**
- la forme finale du graphique doit avoir :
 - le titre** : » Distribution des dans l'échantillon étudié”,
 - la légende**
 - les pourcentages**

Graphique des colonnes

- Sélectionnez les données du tableau des fréquences antérieur créé – voir avant (les 4 ou plus cellules du votre tableau sans le titre du tableau – seulement les noms des catégories et les nombres des sujets dans chaque catégorie) ;

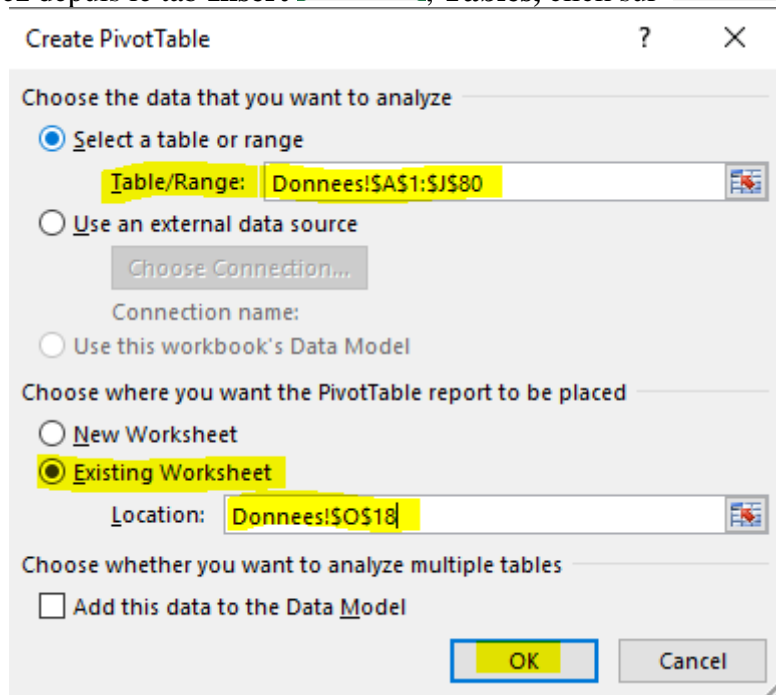


- cherchez le tab: **Insert**→ **Column**
- Choisissez quel type de graphique des colonnes vous voulez et puis appuyez sur l'option Chart Layout
- la forme finale du graphique doit avoir :
 - le titre** : » Distribution des dans l'échantillon étudié”,
 - les titres des axes**

Graphique pour la relation entre deux variables qualitatives :

- Utilisez PivotTable
 - Sélectionnez toutes les données incluant le titre des colonnes

- Insérez depuis le tab **Insert**, **Tables**, click sur

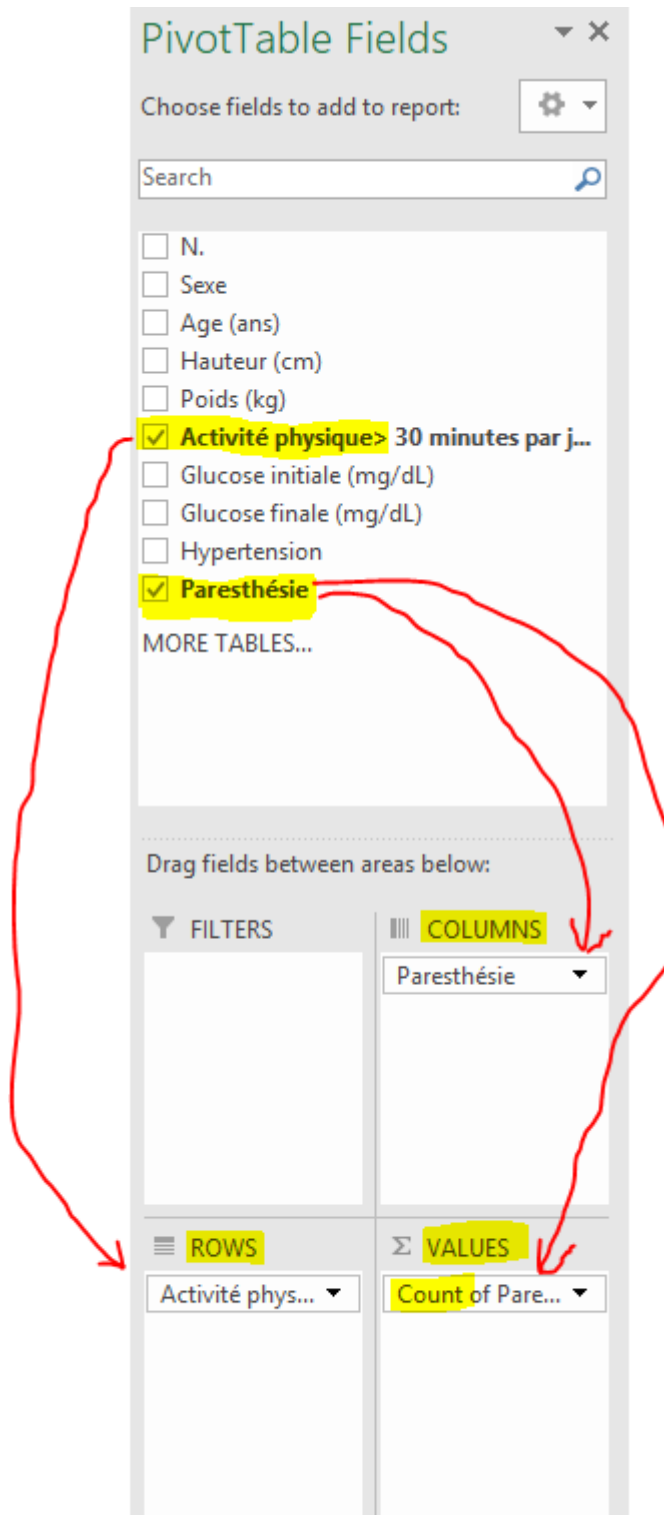


Dans la fenetre dans **Table/Range** vous devez observer que toutes les donnees ont ete selectionne, sinon vous pouvez les selectionner ici.

Vous pouvez mettre le tableau de contingence sur une nouvelle feuille de calcul en selectionnant **New Worksheet**

Ou vous pouvez mettre le tableau de contingence sur la meme feuille de calcul en selectionnant **Existing Worksheet**. Puis dans **Location** vous devez selectionner une cellule pour la position du tableau sur la feuille.

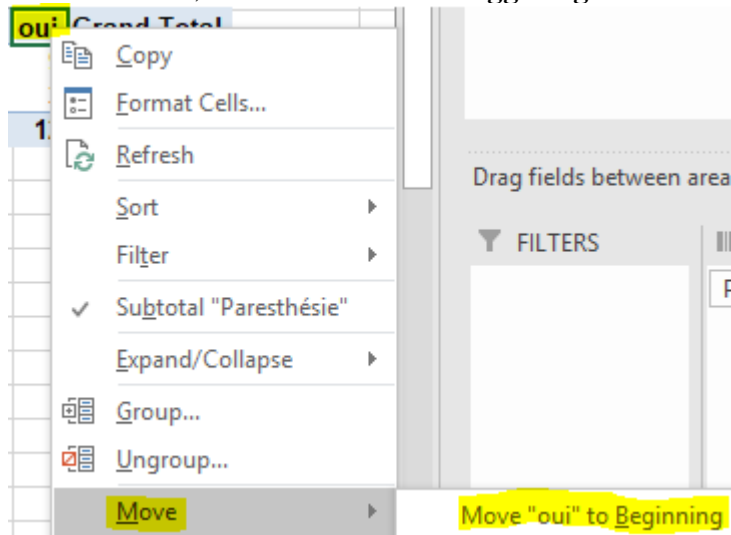
A la fin – click sur le bouton **OK**



- Déplacez la variable dépendante (qui est influencée/l'effet) (ex. maladie, résultat d'un traitement, complication, ...) sur les colonnes - **COLUMNS**
- Déplacez la variable indépendante (qui influence/la cause supposée) (ex. traitement, facteur de risque, nouvel test diagnostique, variable de regroupement) sur les lignes - **ROWS**
- Déplacez une des deux variables dans la boîte **Values**
- Vérifiez qu'il a compte les cas (dans la boîte **Values** il doit écrire **Count of ...**)
- Vous allez obtenir un tableau de contingence :



Count of Paresthésie	Column Labels		
Row Labels	non	oui	Grand Total
non	32	9	41
oui	35	3	38
Grand Total	67	12	79

- Parfois on a besoin de déplacer la colonne Oui dans la première colonne (ex les malades sur la première colonne, et les non malades sur la deuxième) – **click droit** sur Oui, puis sélection **Move**, et **Move “Oui” to Beginning**





- Parfois on a besoin de déplacer la ligne Oui dans la première ligne (ex le facteur de risque présent sur la première ligne et le facteur de risque absent sur la deuxième) – **click droit** sur Oui, puis sélection **Move**, et **Move “Oui” to Beginning**
- Vous allez obtenir un tableau similaire à :

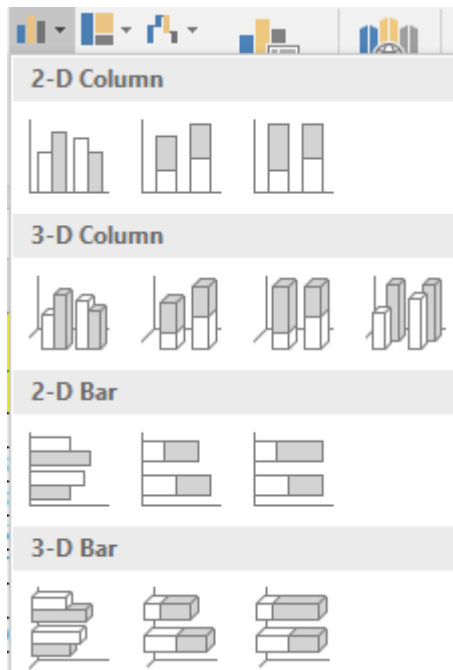
Count of Paresthésie	Column Labels		
Row Labels	oui	non	Grand Total
oui	3	35	38
non	9	32	41
Grand Total	12	67	79

- puis **insérez un graphique des colonnes ou des barres**
- pour cela vous devez faire un click à l’intérieur du tableau de contingence antérieur créé
- puis dans l’onglet **Insert** , section **Charts**, click sur , puis sélectionnez un des graphiques – soit **colonnes (Column)** soit **barres (Bar)**, soit

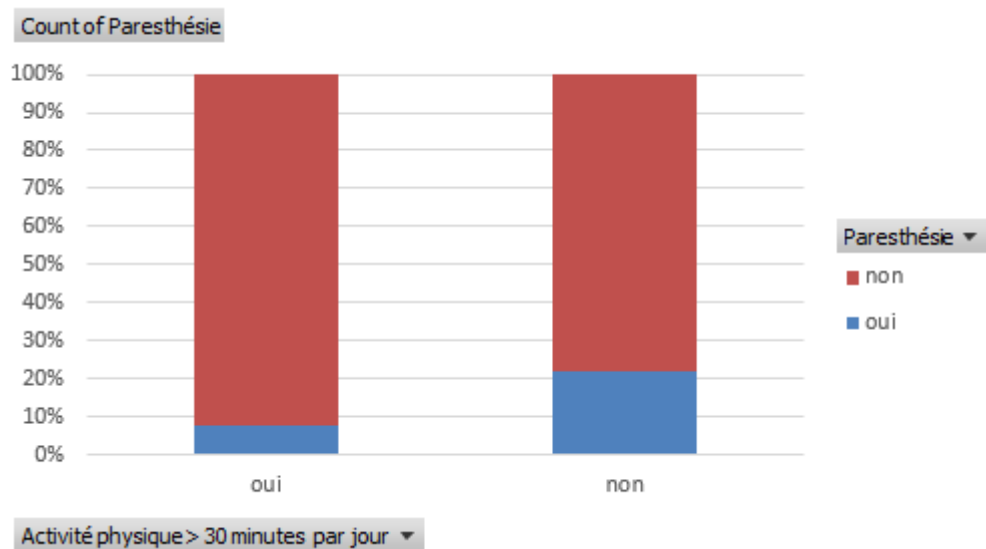
empilé () ou non, soit **2d** (recommande) ou **3d**, soit en **valeur absolues**

(  - utile pour voir les effectifs du chaque group) soit en

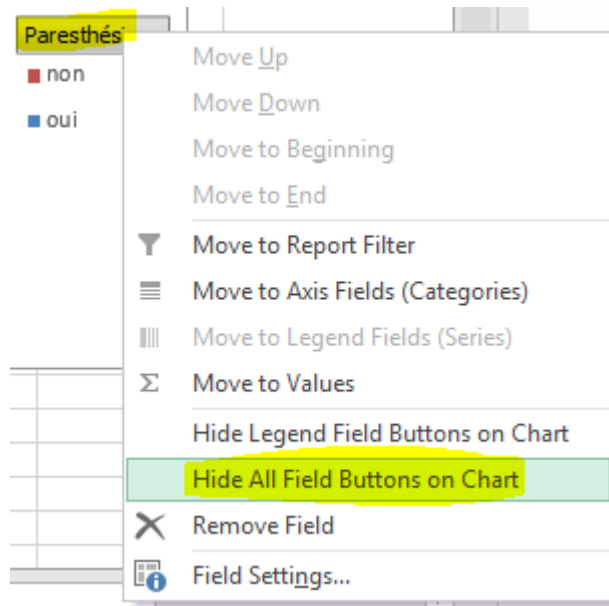
pourcentages (100% stacked columns  - utile pour comprendre mieux les différences)



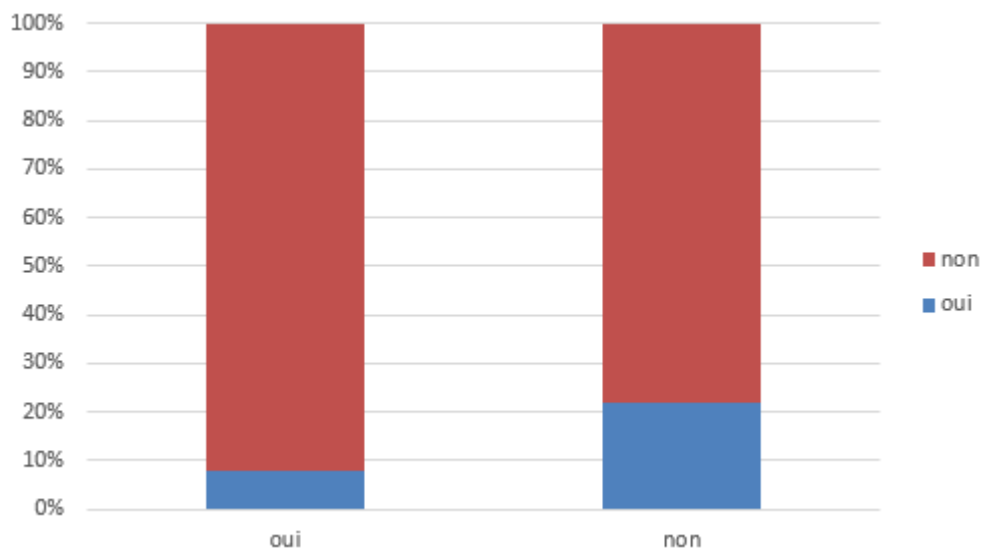
Voici un graphique de type 100% stacked columns



Pour enlever les boutons depuis un graphique réalisé avec **PivotTable**, on fait un clic droit sur l'un des boutons du graphique, et on va choisir **Hide all field buttons on Chart**





Comme ça on va avoir un graphique comme le suivant:

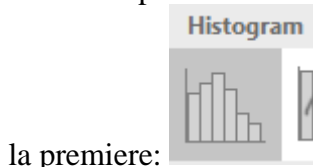


Graphique pour une seule variable quantitative continue ou discrète

Création d'un histogramme (variante simple)

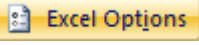
Sélectionnez la colonne (incluant le titre de la colonne)

Insérez depuis le tab **Insert** , section **Charts**, click sur , puis selectionnez



la première:

Création d'un histogramme (variante difficile – si la variante simple ne marche pas) (sans préciser les classes) :

- utilisez le tab Data→Data Analysis (s'il n'existe pas, installez-le en choisissant les options : menu File -  →Add Ins→Manage Excel Add-Ins →Go→Analysis ToolPack→OK);

- Parmi les options de l'analyse de données se trouve l'**Histogramme** :
- Dans la boîte de dialogue de l'histogramme on choisit :
- **Input Range**: =toutes les valeurs de l'âge avec le titre ;
- **Bin Range** : Rien ! (Comme ça il va calculer automatiquement les intervalles de classe)
- **Labels** : Sélectionnez les étiquettes si votre sélection des cellules contient le titre.
- **Output options** : sélectionnez l'option de sortie dans une nouvelle feuille de calcul. Le résultat sera placé dans une nouvelle feuille appelée **Histogramme** ;
- **Chart Output** : Excel vous va réaliser une graphique (l'histogramme)
- changez le format de l'histogramme :
 - Agrandissez le graphique et supprimez l'espace entre les barres verticales en faisant le clic droit de souris sur l'une des barres verticales et l'option : **FORMAT Data Séries** → modifiez la valeur de la largeur à 0 (Gap width).
 - Effacez la légende ;

Création d'un histogramme (variante difficile – on ne le demande pas) ayant un certain nombre de classes :

- Dans votre feuille de calcul, écrivez le tableau suivant :

Minimum	
Maximum	
amplitude	
Grandeur de la classe	

- Remplissez ce tableau en utilisant :
- la fonction **MIN** pour la valeur minimale de besoins énergétiques;
- la fonction **MAX** pour la valeur maximale de besoins énergétiques;
- la différence entre le minimum et le maximum pour calculer l'amplitude ;
- la longueur de la classe est donnée par sa formule : **=Amplitude / (Numéro de classes - 1)**
- renommez une nouvelle colonne par « **Intervalles de classe** »
- dans cette nouvelle colonne, commencez l'écriture des intervalles par la valeur minimale calculée antérieurement, par exemple :

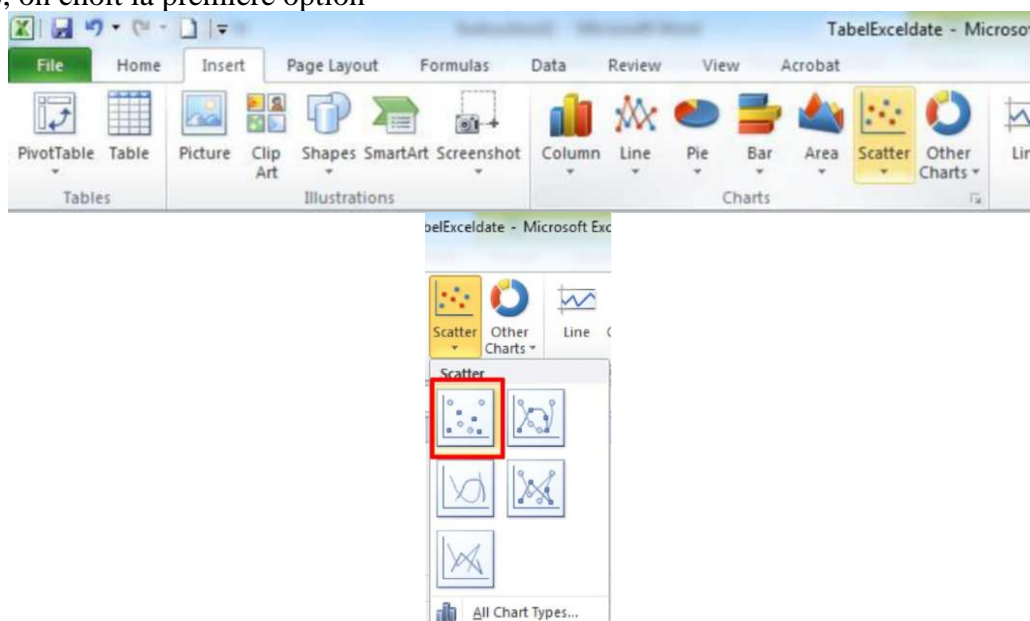
L
Intervalles de classe
écrivez la valeur minimale obtenue
=valeur minimale+longueur de la classe
=valeur minimale+2*longueur de la classe
=valeur minimale+3*longueur de la classe
↓
valeur maximale

- utilisez le tab Data→Data Analysis (s'il n'existe pas, installez-le en choisissant les options : menu File - **Excel Options** →Add Ins→Manage Excel Add-Ins →Go→Analysis ToolPack→OK);
- parmi les options de l'analyse de données se trouve l'**Histogramme** :
- dans la boîte de dialogue de l'histogramme on choisit :
- **Input Range**: =toutes les valeurs de l'âge avec le titre ;
- **Bin Range** := les groupes d'années avec le titre
- **Labels** : Sélectionnez les étiquettes parce que chaque colonne contient le titre.
- **Output options**: sélectionnez l'option de sortie dans une nouvelle feuille de calcul. Le résultat sera placé dans une nouvelle feuille appelée **Histogramme** ;
- **Chart Output**: Excel vous va réaliser une graphique (l'histogramme)
 - changez le format de l'histogramme :
 - agrandissez le graphique et supprimez l'espace entre les barres verticales en faisant le click droit de souris sur l'une des barres verticales et l'option : **FORMAT Data Séries**→ modifiez la valeur de la largeur à 0 (Gap width).
 - écrivez les 4 classes sur l'axe Ox en modifiant les valeurs du tableau de fréquence ;
 - affichez sur les colonnes du graphique les fréquences calculées ;
 - Effacez la légende ;

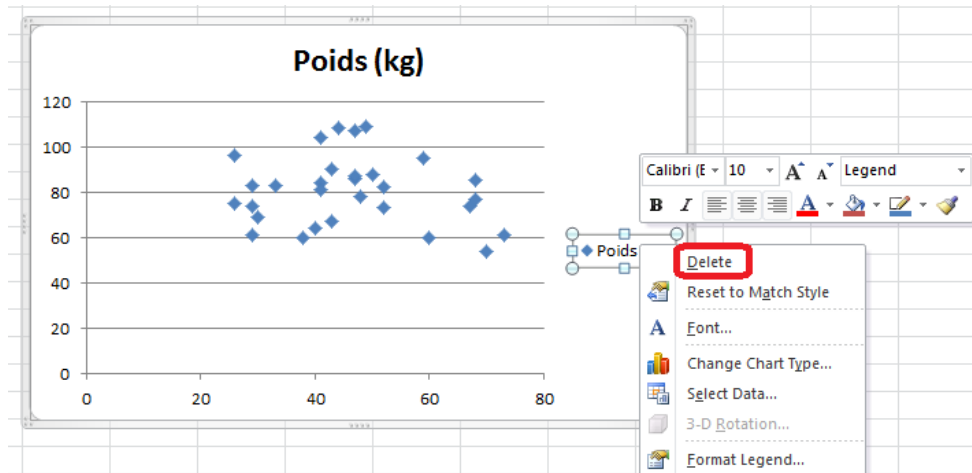
Graphique des deux variables quantitatives – nuage des points - scatter

Vous pouvez le réaliser directe dans la feuille avec les données, ou vous pouvez copiez les colonnes des deux variables dans une nouvelle feuille de calcul. La première colonne sera la variable qui sera représentée sur l'axe horizontale, la deuxième colonne sera la variable qui sera représentée sur l'axe verticale.

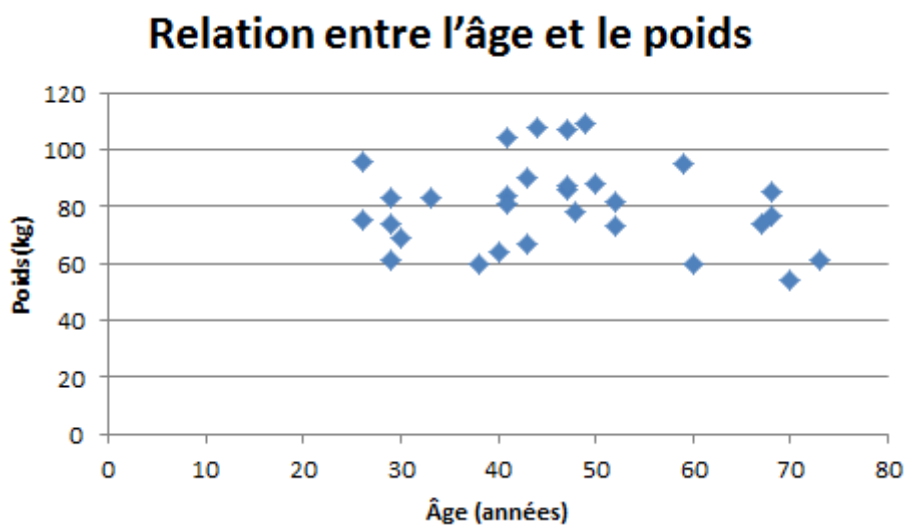
Sélectionnez les deux colonnes. Si les deux colonnes ne sont pas jointes, on sélectionne la première colonne et puis on maintient la touche CTRL foncée en sélectionnant la deuxième colonne. Dans l'onglet Insert, la section Chart (Graphiques), le bouton Scatter (Nuage des points), on choisit la première option



On obtient le graphique. On efface la légende avec clic droit sur la légende et on choisit l'option Delete (Effacer)

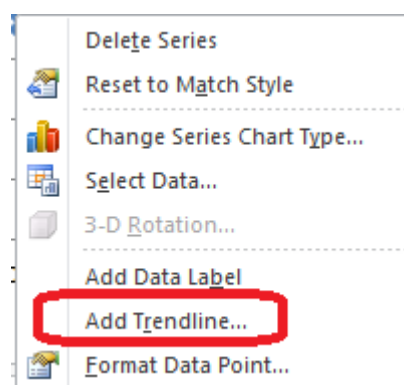


On ajoute les définitions des axes et on modifie le titre.

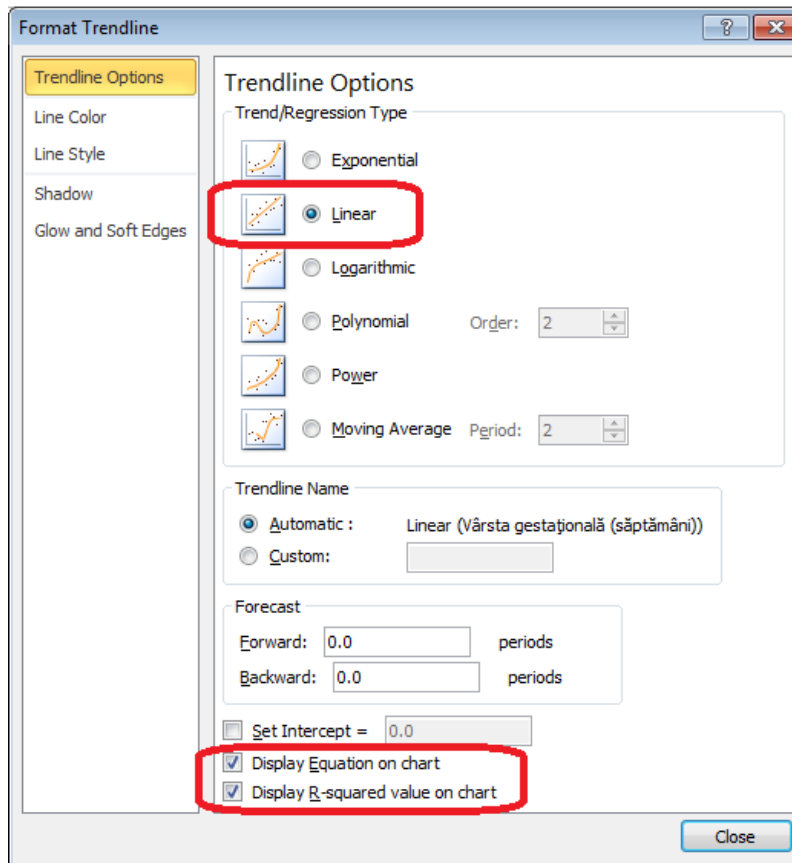


Ajouter la droite de régression sur le graphique

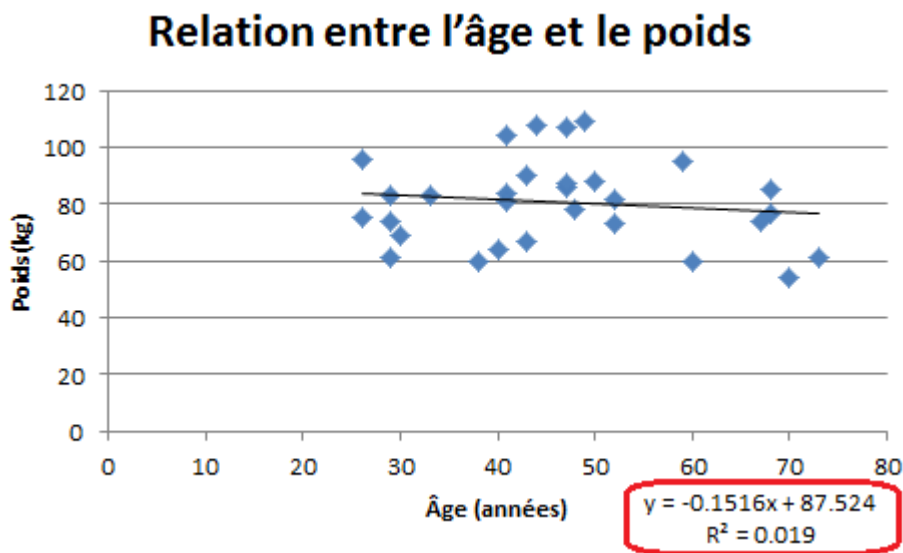
Clic droit sur n'importe quel point du nuage des points, et sélectionnez **Add Trendline** (Ajouter une ligne de tendance)



Dans la fenêtre qui s'ouvre sélectionnez Linear (Linéaire) – pour afficher la droite de régression linéaire. Puis vous pouvez sélectionner **Display Equation on chart** – pour afficher l'équation de la droite de régression et **Display R-squared value on chart** – pour afficher le coefficient de détermination.

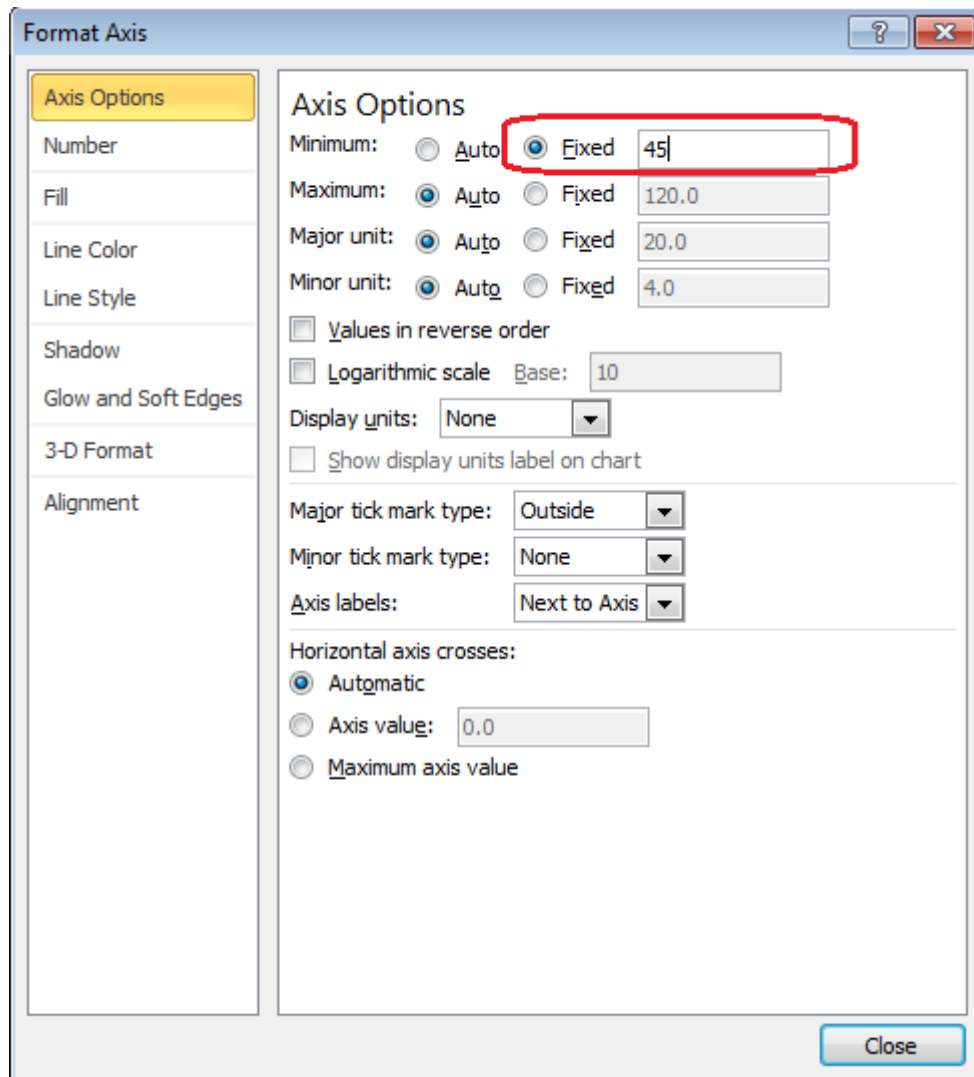


On clique sur le bouton Close (Fermer), et après le repositionnement de l'équation et du coefficient de détermination (R^2) on obtient :

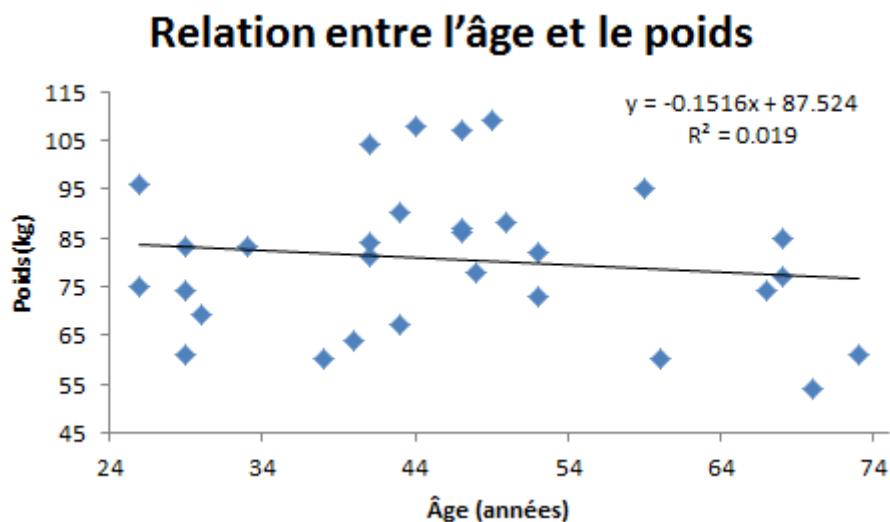


La modification de la valeur minimale et maximale d'affichage pour les axes du graphique nouage des points.

Pour modifier ces valeurs pour l'axe verticale on fait double clic sur les valeurs de l'axe verticale et on obtient la fenêtre suivante :



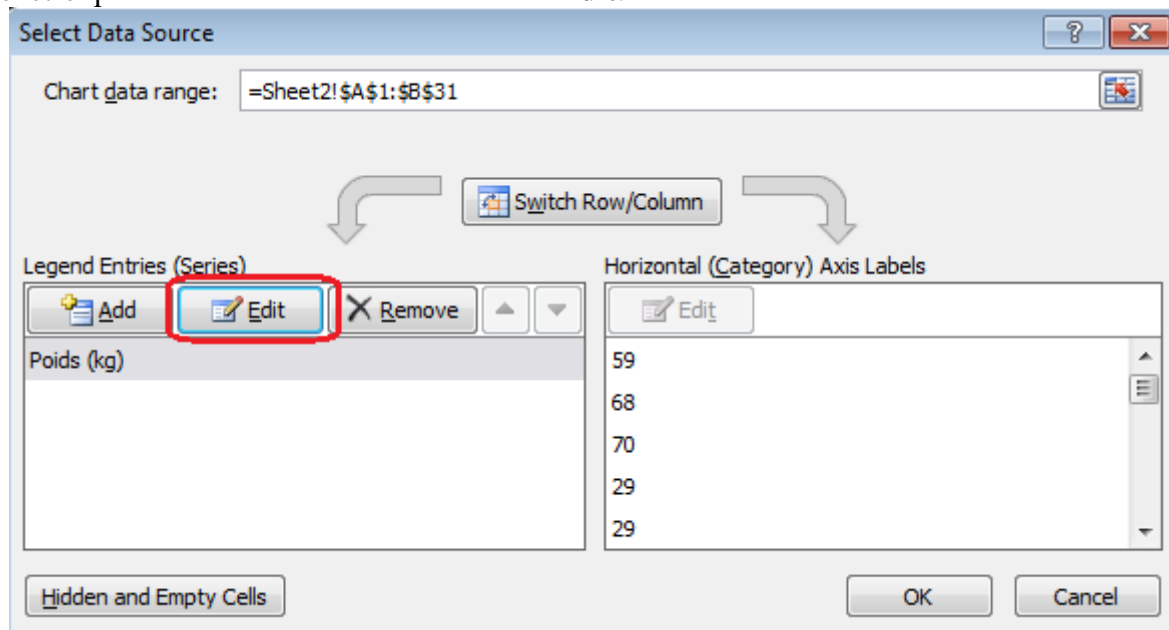
Pour modifier la valeur minimale pour **Minimum**, on choisit **Fixed** (Fixée – manuelle) et on met une valeur proche aux plus bas points sur le graphique. On procède d’une manière similaire pour la valeur maximale et aussi pour l’axe horizontale. On peut obtenir un graphique comme le suivante :



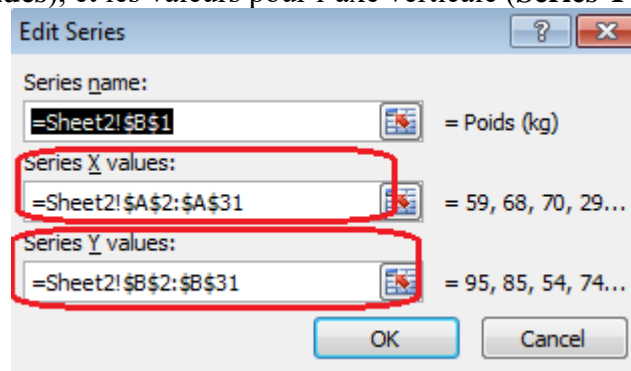
Les lignes horizontales ont été effacées en le sélectionnant avec un click, et puis avec clic droit on choisit Delete (Effacer)

Le changement des axes entre eux

Clic droit sur le graphique. On choisit l'option **Select Data** (Sélectionnez les données). Dans la fenêtre qui s'ouvre on fait clic sur le bouton **Edit**.



On va obtenir une nouvelle fenêtre dans laquelle on peut sélectionner les valeurs pour l'axe horizontale (**Seriex X values**), et les valeurs pour l'axe verticale (**Series Y values**).



Graphique boîte à moustaches (box and whiskers) – pour la relation entre une variable quantitative et une variable qualitative

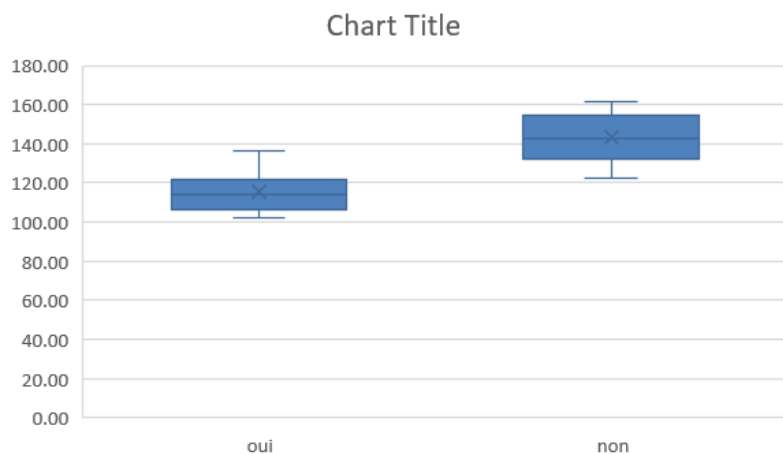
Important! La colonne qui a la variable qualitative qui représente les groupes à comparer doit être à la gauche de la colonne qui a la variable quantitative.

Sélectionnez les deux colonnes. Si les deux colonnes ne sont pas jointes, on sélectionne la première colonne et puis on maintient la touche CTRL foncée en sélectionnant la deuxième

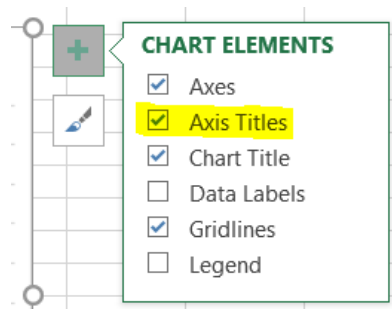
F	G	H	I
Activité physique > 30 minutes par jour	Glucose initiale (mg/dL)	Glucose finale (mg/dL)	
oui	160.69	115.11	nc
non	176.84	156.87	nc
oui	170.84	128.33	nc
oui	178.25	134.62	nc
non	149.55	129.72	nc
oui	153.67	106.18	ou
oui	148.85	104.95	nc
non	162.74	142.48	nc

Insérez depuis le tab **Insert** , section **Charts**, click sur , puis selectionnez dans

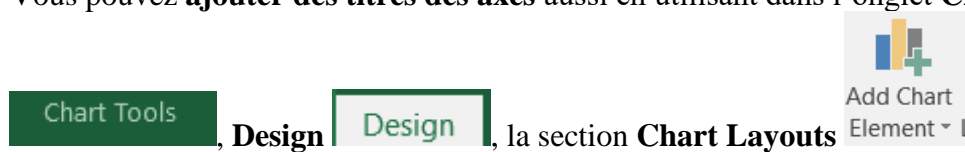
Box and whiskers:

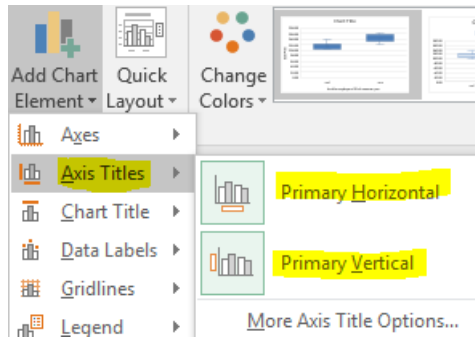


Puis **ajoutez des titres des axes**, soit en cliquant sur le sign + pour ajouter des elements pour le graphique



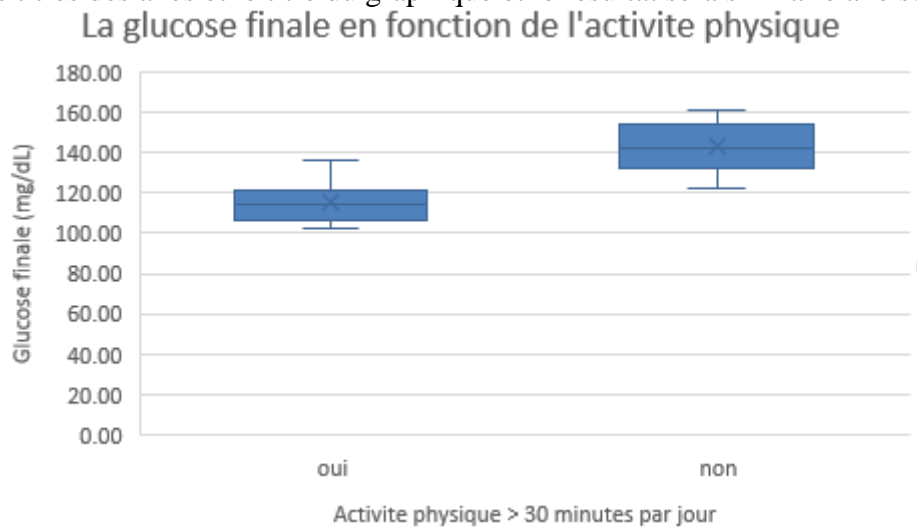
Vous pouvez **ajouter des titres des axes** aussi en utilisant dans l'onglet **Chart Tools**





Puis on selection dans **Axis Titles: Primary Horizontal** (pour avoir un titre sur l'axe horizontale), et **Primary Vertical** (pour avoir un titre sur l'axe verticale)

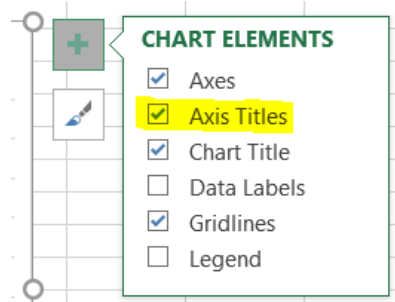
On ecrit les titres des axes et le titre du graphique et le resultat sera similaire a le suivant:



Modifications des graphiques (ajouter des titres, légende ...)

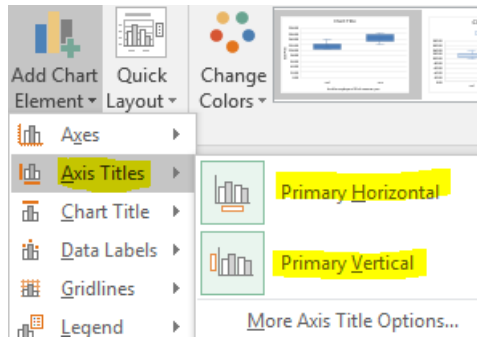
Ajouter les titres des axes

Variante 1 - en cliquant sur le sign + pour ajouter des éléments pour le graphique



Variante 2 - Vous pouvez **ajouter des titres des axes** aussi en utilisant dans l'onglet **Chart**

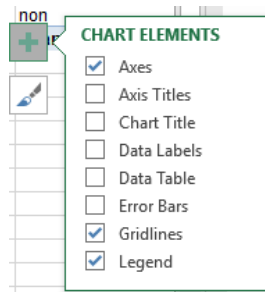
Tools **Chart Tools**, **Design** **Design**, la section **Chart Layouts** 



Puis on selection dans **Axis Titles: Primary Horizontal** (pour avoir un titre sur l'axe horizontale), et **Primary Vertical** (pour avoir un titre sur l'axe verticale)

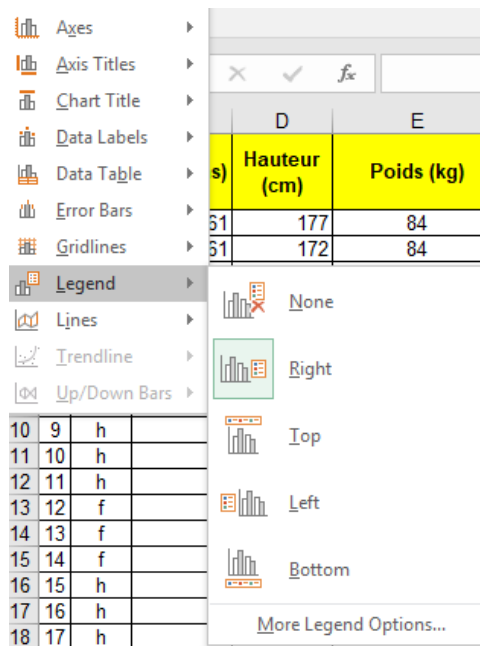
Ajouter la légende

Variante 1 - en cliquant sur le sign + pour ajouter des éléments pour le graphique et en choisissant **Legend**



Variante 2 - Vous pouvez **ajouter des titres des axes** aussi en utilisant dans l'onglet **Chart**

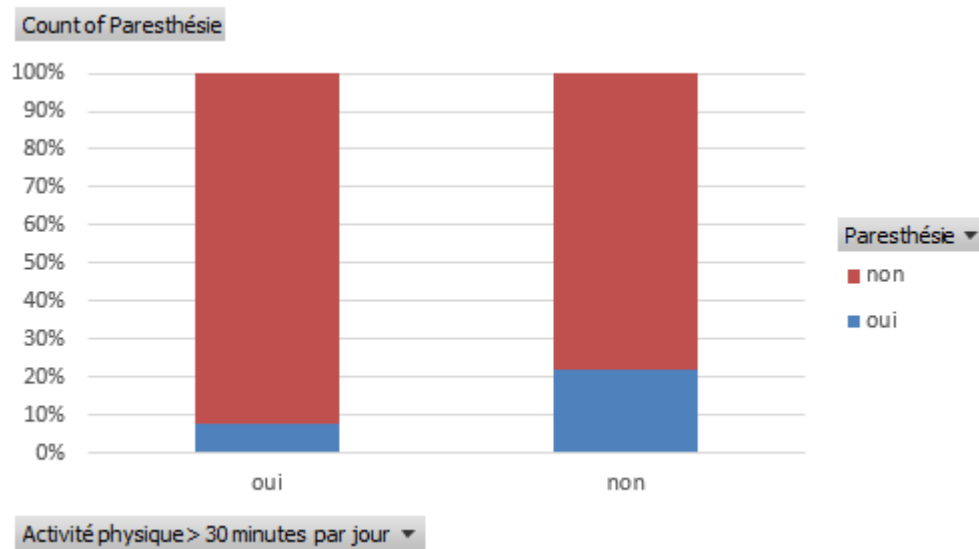
Tools **Chart Tools**, **Design** **Design**, la section **Chart Layouts** **Add Chart Element** et **Legend**



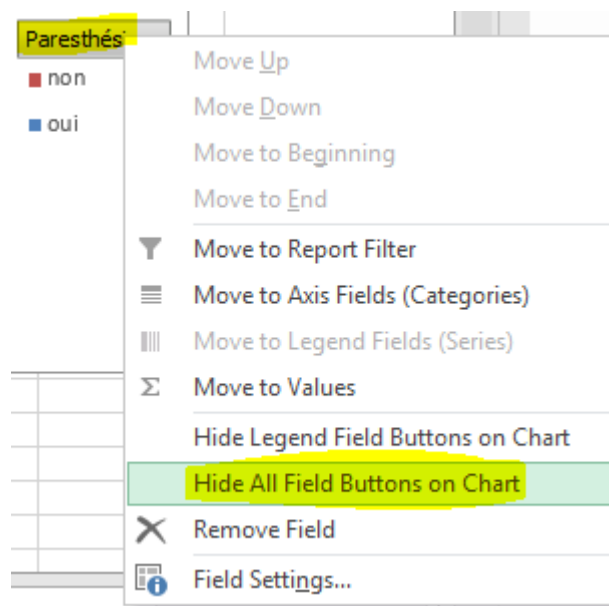
La vous pouvez choisir la position de la légende (Right, Top, Left, Botto, ou aucune – None), ou des options plus avancées avec **More Legend Options...**

Pour enlever les boutons depuis un graphique réalisé avec PivotTable

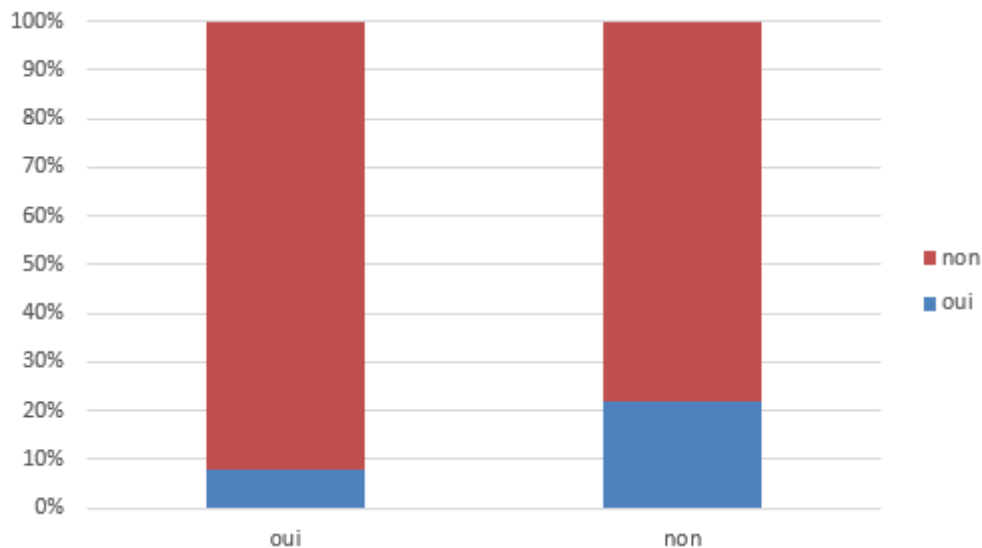
Voilà un graphique avec des boutons :



Pour enlever les boutons depuis un graphique réalisé avec **PivotTable**, on fait un clic droit sur l'un des boutons du graphique, et on va choisir **Hide all field buttons on Chart**



Comme ça on va avoir un graphique comme le suivant:



Interprétations

Les types des variables:

Qualitatives (non métriques):

= représentent des catégories ou des classes et n'impliquent pas les quantités d'un attribut ou une caractéristique;

Les catégories (classes) sont *exhaustives* (à chaque sujet doit correspondre une classe) et mutuellement *incompatibles* (en ayant une classe particulière, toutes les autres classes sont exclus);

Les types des variables qualitatives :

- dichotomique*: deux catégories (e.g. malade : oui/non ; facteur de risque : présent/absent ; avec/sans carie dentaire)
- nominale*: plusieurs catégories (>2), sans ordre (e.g. couleur des cheveux, type des dents, histologie d'un cancer)
- Ordinale*: plusieurs catégories (>2), avec ordre (e.g. sévérité de la maladie : cancer stade I, IIA, IIB, IIIA, IIIB, IV; niveau du duleur: sans, faible, moderee, forte)

Quantitative (métriques)

Les types des variables quantitatives :

- discrète*: seulement nombres entières (e.g. le nombre d'enfants dans une famille ; nombre des incisifs perdues ; nombre d'interventions chirurgicales ; nombre des jours de hospitalisation);
- continue*: nombres réels, avec virgule, même si on ne l'utilize pas (e.g. poids, hauteur du poche de la gencive);

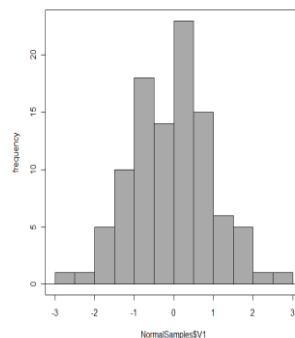
Variables de survie

Le temps qui s'écoule entre l'introduction d'un sujet dans l'étude et l'apparition d'un évènement (décès, guérison, métastase, adénopathie, Un, perte d'un implant, perte d'un obturation, défection d'un prothèse) e.g. le temps de survie jusqu'à le décès)

L'évaluation de normalité des données en utilisant les indicateurs de tendance centrale, d'aplatissement et d'asymétrie.

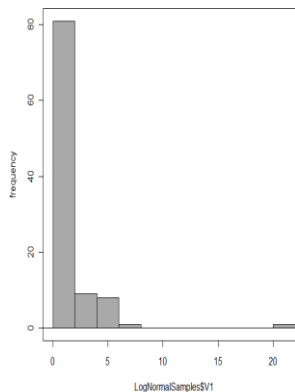
Les suivantes situations indiquent des distributions normales ou non normales. Mais ils ne peuvent pas garantir ça. Ils donnent des suggestions. C'est mieux d'utiliser les graphiques pour évaluer la normalité des données.

Indications de normalité : la moyenne est approximative égale à la médiane. Le coefficient d'aplatissement et le coefficient d'asymétrie sont proches à 0 (dans l'intervalle $-1 - 1$). Si l'une des deux (coefficient d'aplatissement et le coefficient d'asymétrie) est en dehors de l'intervalle $-1 - 1$, on considère la distribution comme non normale – cette règle est plus bonne que l'égalité approximative de la moyenne avec la médiane. Voir en bas un histogramme d'une variable avec distribution approximative normale.



Exemple interprétation du graphique: *l'histogramme a un forme de chapeau, symétrique, ce qui est un suggestion que les données sont normale distribuées*

Indications de non normalité : la moyenne est approximative différente de la médiane. Le coefficient d'aplatissement **ou/et** le coefficient d'asymétrie sont différente de 0 (>1 ou <-1). Voir en bas un histogramme d'une variable avec distribution non normale.



Exemple interprétation du graphique: *l'histogramme n'a pas un forme de chapeau, il y a une forte asymétrie a la droite (asymétrie positive/ le graphique a une queue a la droite), ce qui est un suggestion que les données ne sont pas normale distribuées*

Exemples:

- Dans un échantillon la moyenne du poids est 68 kg, et la médiane du poids est 68,1 kg. Évaluez la normalité du poids. Les valeurs du poids semblent suivre une distribution normale (semblent être normale distribuées), parce que la moyenne est très proche a la médiane.
- Dans un échantillon le coefficient d'asymétrie du poids est 0,3, et le coefficient d'aplatissement du poids est -0,4. Évaluez la normalité du poids. Les valeurs des

triglycérides semblent suivre une distribution normale (semblent être normale distribuées), parce que le coefficient d'asymétrie, et le coefficient d'aplatissement sont les deux dans l'intervalle $[-1,1]$ (proche à 0).

- Dans un échantillon la moyenne des triglycérides est 160 mg/dL, et la médiane des triglycérides est 190 mg/dL. Évaluez la normalité des triglycérides. Les valeurs des triglycérides ne semblent pas à suivre une distribution normale (ne semblent pas être normale distribuées), parce que la moyenne est très différente de la médiane.
- Dans un échantillon le coefficient d'asymétrie des triglycérides est 0,3, et le coefficient d'aplatissement des triglycérides est 1,4. Évaluez la normalité du poids. Les valeurs des triglycérides ne semblent pas suivre une distribution normale (ne semblent pas être normale distribuées), parce que le coefficient d'aplatissement n'appartient pas à l'intervalle $[-1,1]$ (distante de 0).
- Dans un échantillon le coefficient d'asymétrie des triglycérides est -2,3, et le coefficient d'aplatissement des triglycérides est 0,4. Évaluez la normalité des triglycérides. Les valeurs des triglycérides ne semblent pas suivre une distribution normale (ne semblent pas être normale distribuées), parce que le coefficient d'asymétrie n'appartient pas à l'intervalle $[-1,1]$ (distante de 0).

Interprétation des mesures de tendance centrale, emplacement, dispersion, ...

Les mesures de tendance centrale sont les suivants: la moyenne, la médiane, le mode, la moyenne géométrique, la moyenne harmonique, la valeur la centrale.

Les mesures de l'emplacement sont: les quartiles, les percentiles, les déciles

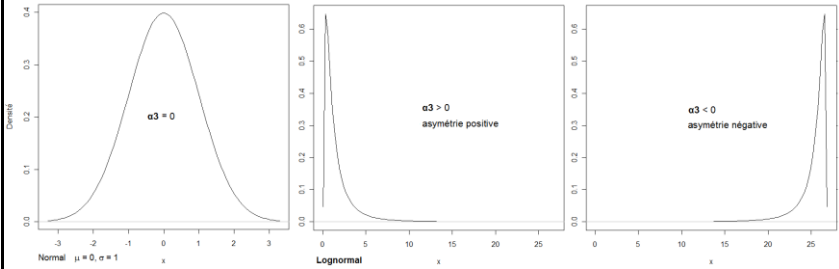
Les mesures de dispersion: Amplitude, Intervalle interquartile, la variance, l'écart type, le coefficient de variation

L'asymétrie – skewness (=coefficient d'asymetrie).

L'aplatissement - kurtosis (=coefficient d'aplatissement).

Moyenne arithmétique	La valeur au milieu des données si la distribution des données est normale
Médiane	L'observation qui divise la distribution sur deux partis égaux ; Voir aussi en bas l'interprétation de la quartile 2. La moyenne est approximative égale à la médiane si la distribution est normale
Mode	La valeur de l'observation associée à la fréquence la plus élevée. Si la fonctionne donne comme résultat NA ça signifie qu'il n'y a pas une mode. Parfois on peut avoir des distributions avec plusieurs valeurs modales. La fonctionne va offrir comme résultat la première valeur modale.
Amplitude	La différence entre la plus grande et la plus petite valeur dans la série des données
Intervalle interquartile	La différence entre le quartile 3 et le quartile 1 dans la série des données.
Écart type (déviaton standard)	Est la racine carrée de la variance. Montre la dispersion des données autour de la moyenne. Pour une série des données avec une distribution normale (gaussienne): Dans l'intervalle $m \pm 1$ DS on trouve ~68 % de la population. Dans l'intervalle $m \pm 2$ DS on trouve ~95 % de la population. Dans l'intervalle $m \pm 3$ DS on trouve 99,7 % de la population.

Erreur standard	Est utilisée dans le calcul de l'intervalle de confiance de 95% autour de la moyenne (pour une distribution normale), aussi est impliqué dans l'inférence statistique.
Quartiles	Il y a trois quartiles: Q1, Q2, Q3 le 1er quartile sépare les 25% inférieurs des données 25% des valeurs sont $\leq Q1$, 75% sont $\geq Q1$ le 2e quartile est la médiane de la série (50%) 50% des valeurs sont $\leq Q2$ / médiane, 50% sont $\geq Q2$ / médiane le 3e quartile sépare les 75% inférieurs des données 75% des valeurs sont $\leq Q3$, 25% sont $\geq Q3$
Percentiles	le 1er percentile sépare le 1 % inférieurs des données; le 95e percentile sépare les 95 % inférieurs des données;
Coefficient d'aplatissement - Kurtosis	<p><i>proche a 0 [-1, 1]</i> indique une distribution <i>mesokurtique</i>. Si la valeur appartient à l'intervalle [-1, 1] pourrait être considéré que les données ont une distribution mesokurtique.</p> <p>un coefficient d'aplatissement positif (>0) correspond à une distribution <i>pointue</i> (<i>leptokurtique depuis >1</i>) et un coefficient d'aplatissement négatif (<0) correspond à une distribution <i>aplatie</i> (<i>platikurtique depuis < -1</i>),</p> <p>Ex. coef aplatissement = 0,5 => <i>il est proche a 0</i> (appartient à l'intervalle [-1, 1]) dont il y a une distribution mesokurtique ; on observe qu'il est positif, donc il est un peu pointue</p> <p>Ex. coef aplatissement = - 0,7 => <i>il est proche a 0</i> (appartient à l'intervalle [-1, 1]) dont il y a une distribution mesokurtique ; on observe qu'il est negative, donc il est un peu aplati</p> <p>Ex. coef aplatissement = 1,3 => <i>il n'est pas proche a 0</i> (<i>n'appartient pas à l'intervalle [-1, 1]</i>) et on observe qu'il est positif, donc il y a une distribution leptokurtique - la distribution est un pointue</p> <p>Ex. coef aplatissement = - 2,4 => <i>il n'est pas proche a 0</i> (<i>n'appartient pas à l'intervalle [-1, 1]</i>) et on observe qu'il est négative, donc il y a une distribution platikurtique - la distribution est un aplatie</p>
	<p>The figure contains three side-by-side plots of probability density functions. The first plot, titled 'distribution mesokurtique', shows a standard bell-shaped curve with a peak at 0 and a label $\alpha_4 = 0$. The second plot, titled 'distribution leptokurtique', shows a narrower, taller bell-shaped curve with a peak at approximately 15 and a label $\alpha_4 > 0$. The third plot, titled 'distribution platykurtique', shows a wider, shorter bell-shaped curve with a peak at approximately 15 and a label $\alpha_4 < 0$. All plots have 'Densité' on the y-axis and 'x' on the x-axis. A legend at the bottom left indicates 'Normal: $\mu = 0, \sigma = 1$'.</p>

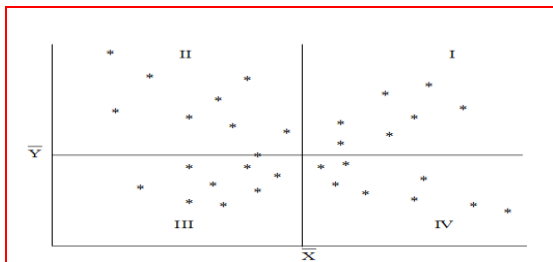
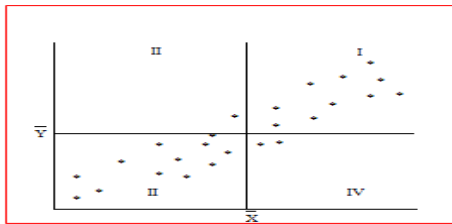
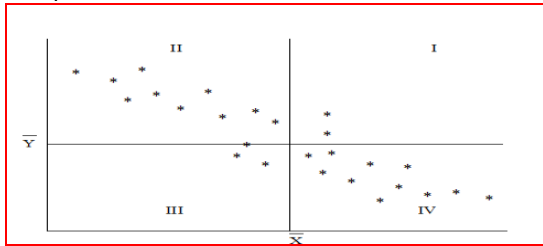
<p>Coefficient d'asymétrie - Skewness</p>	<p>Un coefficient positif (>0) indique une distribution étalée vers la droite ; une asymétrie à la droite</p> <p>Un coefficient négatif (<0) indique une distribution étalée vers la gauche ; une asymétrie à la gauche</p> <p>Si la valeur appartient à l'intervalle [-1, 1] pourrait être considéré que les données ont une distribution symétrique (du point de vue de l'asymétrie). Sinon la distribution est asymétrique</p> <p><i>Ex. coef d'asymétrie = 0,5 => il est proche à 0</i> (appartient à l'intervalle [-1, 1]) dont il y a une distribution symétrique ; on observe qu'il est positif, donc la distribution est un peu étalée vers la droite</p> <p><i>Ex. coef d'asymétrie = - 0,7 => il est proche à 0</i> (appartient à l'intervalle [-1, 1]) dont il y a une distribution symétrique ; on observe qu'il est négative, donc la distribution est un peu étalée vers la gauche</p> <p><i>Ex. coef d'asymétrie = 1,3 => il n'est pas proche à 0</i> (n'appartient pas à l'intervalle [-1, 1]) et on observe qu'il est positif, donc la distribution est étalée vers la droite ou on dit qu'il y a une asymétrie à la droite</p> <p><i>Ex. coef d'asymétrie = - 2,4 => il n'est pas proche à 0</i> (n'appartient pas à l'intervalle [-1, 1]) et on observe qu'il est négative, donc la distribution est étalée vers la gauche ou on dit qu'il y a une asymétrie à la gauche</p> 
<p>Coefficient of variation</p>	<p>Pour l'évaluation de l'homogénéité d'une série des données</p> <p>Si CV <10% la série des données peut être considérée comme homogène;</p> <p>Si CV est comprise entre $\geq 10\%$ - $< 20\%$ la série des données peut être considérée comme relativement homogène;</p> <p>Si CV est entre $\geq 20\%$ à $< 30\%$ la série des données peut être considérée comme relativement hétérogène;</p> <p>Si CV $\geq 30\%$ la série des données peut être considérée comme hétérogène.</p>

Interprétation du graphique nuage des points

Interprétations du graphique – le sens/tendance (direction) de la relation: divisez le nuage de points en 4 cadrans à l'aide de la moyenne de X et la moyenne de Y : (voir les exemples en bas)

S'il existe une relation linéaire entre X et Y, les marqueurs seront en cadrans II et IV (tendance croissante/ pente ascendante/ pente positive/ lien (direct) proportionnel) ou I et III (tendance décroissante / pente descendante/ pente négative/ lien inversement proportionnelle).

Si les marqueurs sont uniformément dispersés dans l'ensemble des quatre cadrans, le graphique indique l'absence d'une relation linéaire entre X et Y.



Interprétation de la corrélation

Le coefficient de corrélation – r

Montre la direction/sens/tendance et l'intensité/force de la corrélation;

Direction/sens/tendance :

- > 0 tendance croissante/ pente ascendante/ lien direct proportionnel/ corrélation positive
- < 0 tendance décroissante/ pente descendante/ lien inversement proportionnel/ corrélation négative
- approximatif 0 - aucune tendance

Interprétation de l'intensité/force/degré/importance de la corrélation (on préfère le mot corrélation ici, même si association/liens/relation peut être utilisé) avec les règles empiriques de Colton [Colton T. Statistics in Medicine. Little Brown and Company, New York, NY 1974]:

(-0.25 et 0,25)

=> Une relation négligeable ou aucune corrélation linéaire entre les variables

[0.25 et 0.50) ou [-0.25 et -0.50)

=> un degré de corrélation faible/acceptable

[0.50 et 0.75) ou [-0.50 et -0.75)

=> un degré de corrélation modérée à bonne

[0.75 et 1] ou [-0.75 et -1]

=> une corrélation très bonne vers excellente

Interprétation du coefficient de détermination

- r^2 - réponds à la question suivante: quel est le pourcentage de variation de Y qui peut être expliquée par la relation linéaire entre X et Y? exemple: $r^2 = 0.0027 \rightarrow 0,27\%$ de la variation de l'IMC pourrait être expliquée par la relation linéaire entre l'IMC et l'âge.

Interprétation des coefficients d'une équation de régression linéaire simple

La droite de régression $Y(X)$: $a_0 + a_1 X$

a_0 = est l'ordonnée à l'origine

a_1 = la pente de la droite de régression.

Interprétation de pente - a_1

Pour chaque changement d'une unité de mesure de X, Y va être modifiée avec a_1

Exemple : Le poids (kg) = 0.65 * hauteur (cm) – 62.6 kg. La pente a_1 est 0.65. Pour chaque centimètre en plus de la hauteur le poids va augmenter avec 0.65 kilogrammes.

La pente – indique la

- > 0 tendance croissante/ pente ascendante/ pente positive/ lien direct proportionnel

- < 0 tendance décroissante/ pente descendante/ pente négative/ lien inversement proportionnel

- \approx 0 - aucune tendance

- plus a_1 est grand plus la relation est forte

- plus a_1 est proche du 0, plus la relation est faible

Intervalles de confiance

L'interprétation de l'Intervalle de confiance:

Interprétation simplifiée : l'intervalle de confiance à 95% pour un paramètre statistique est l'intervalle des valeurs dans lequel on peut être 95% sûrs que le «vrai» paramètre de la population cible se retrouve.

Exemples :

- ***Pour une moyenne*** : la vraie valeur de la moyenne du poids dans la population cible est situé entre 64 kg et 83 kg, avec une probabilité de 95%.
- ***Pour une fréquence*** : la vraie valeur de la fréquence des fumeurs dans la population cible est situé entre 20% et 30%, avec une probabilité de 95%.
- ***Pour une différence entre deux moyennes*** : la vraie valeur de la différence entre les moyennes du poids des ceux qui ont reçu un traitement A et un traitement B pour la perte du poids dans la population cible est situé entre -2 kg et +4 kg, avec une probabilité de 95%.
- ***Pour une différence entre deux fréquences*** : la vraie valeur de la différence entre les fréquences des réactions secondaires des ceux qui ont reçu un traitement avec aspirine et un traitement avec paracétamol pour la douleur dans la population cible est situé entre 3% et 7%, avec une probabilité de 95%.

Interprétation correcte : l'Intervalle de confiance à 95 % pour un paramètre d'une population est un intervalle pour lequel environ 95 % des intervalles construits de manière similaire sur des échantillons pris au hasard dans la population d'intérêt, contiendront le paramètre recherché (le vrai paramètre de la population).

Le calcul des intervalles de confiance

Notations :

σ - écart-type de la population

S, S_1, S_2 - l'écart type de l'échantillonnage

s, s_1, s_2 - l'écart type d'échantillon

m, m_1, m_2 - la moyenne arithmétique de l'échantillon

n, n_1, n_2 - nombre de sujets

f, f_1, f_2 - fréquence observée

$\alpha = 1 -$ niveau de confiance (par exemple 95, 90) / 100 ex. pour un niveau de de confiance 95%, $\alpha = 0,05$

$$S^2 = \frac{n}{n-1} s^2$$

Le calcul de l'écart type dans Excel

Pour calculer l'écart type descriptive de l'échantillon on peut utiliser la fonction STDEVP (ou STDEV.P) . **Pour calculer l'écart type d'échantillonnage** on peut utiliser soit la fonction STDEV (ou STDEV.S) soit le **Data Analysis – Descriptive Statistics**.

Formules de calcul

Les intervalles de confiance pour : (les données ont une distribution normale)

La moyenne d'une variable (caractéristique) quantitative

Pour une variance / écart-type (σ) connu dans la population :

$$(a) \left(m - Z_{critique} ES; m + Z_{critique} ES \right) = \left(m - Z_{critique} \frac{\sigma}{\sqrt{n}}; m + Z_{critique} \frac{\sigma}{\sqrt{n}} \right)$$

Les valeurs critiques de Z a un niveau de confiance de :

90% Z = 1,64

95% Z = 1,96

99% Z = 2,58

Pour une variance / écart-type inconnu dans la population, des petites ou grands échantillons

$$(b) \left(m - t_{critique} ES; m + t_{critique} ES \right) = \left(m - t_{critique} \frac{S}{\sqrt{n}}; m + t_{critique} \frac{S}{\sqrt{n}} \right)$$

Le **t critique** se calcule avec la fonction TINV(alpha, nombre de sujets - 1) par exemple pour une intervalle de confiance de 95% pour un échantillon de 200 sujets, alfa est 0.05 (1-0.95), donc =TINV(0.05, 200-1)

La différence des deux moyennes

$$c) \left((m_1 - m_2) - t_{critic_1} ES; (m_1 - m_2) + t_{critic_1} ES \right) \text{ ou}$$

$$ES = \sqrt{\frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2}}$$

Le **t critique** se calcule avec la fonction TINV(alpha, nombre de sujets dans les deux échantillons - 2) par exemple pour une intervalle de confiance de 95% pour un échantillon 1 de 200 sujets et un échantillon 2 de 100 sujets, alfa est 0.05 (1-0.95), donc =TINV(0.05, 200 + 100 - 1)

Une fréquence relative d'une caractéristique qualitative: np >= 10 et nq >= 10

$$(d) \left(f - Z_{critique} ES; f + Z_{critique} ES \right) = \left(f - Z_{critique} \sqrt{\frac{f(1-f)}{n}}; f + Z_{critique} \sqrt{\frac{f(1-f)}{n}} \right)$$

Les valeurs critiques de Z a un niveau de confiance de :

90% Z = 1,64

95% Z = 1,96

99% Z = 2,58

La différence des deux fréquences relatives

$$(e) \left((f_1 - f_2) - Z_{critique} ES; (f_1 - f_2) + Z_{critique} ES \right) \text{ ou}$$

$$ES = \sqrt{\frac{f_1 \times (1 - f_1)}{n_1} + \frac{f_2 \times (1 - f_2)}{n_2}}$$

Les valeurs critiques de Z a un niveau de confiance de :

90% Z = 1,64

95% Z = 1,96

99% Z = 2,58

Tests statistiques

Les étapes d'un test statistique

Étape 1 : Définitions des hypothèses (voir le chapitre suivant pour des exemples plus utiles)

- **L'hypothèse nulle: H₀**
 - Il n'y a pas une différence statistiquement significative
 - (entre 2 traitements, entre deux groups)
 - en ce qui concerne une caractéristique
 - Il n'y a pas un relation/lien/association statistiquement significative
 - (entre 2 caractéristiques: Facteur de risque - maladie)
- **L'hypothèse alternative: H₁ (négation du H₀)**
 - Il y a une différence statistiquement significative (entre 2 traitements, entre deux groups) en ce qui concerne une caractéristique
 - Il y a un relation/lien/association statistiquement significative
 - (entre 2 caractéristiques: Facteur de risque - maladie)

Étape 2.

- Le calcul de la *statistique appropriée* du test
 - (paramètre du test)
 - exprimant la différence entre les éléments comparés.

- Qui suit une loi de distribution si l'hypothèse nulle est vraie.

Étape 3.

- Sélectionner le niveau de signification (alfa) pour le test statistique, ou la valeur α = la probabilité de rejet incorrect de H_0 quand elle est vraie

Valeur traditionnelle: $\alpha = 0,05$, (5% erreur) ! On va utiliser cette valeur pour tous les tests dans nos cours et travaux pratiques.

Étape 4.

- Déterminer la/les *valeur/s critique/s* de la statistique du test :
 - trouvée dans la table de la distribution statistique t, Z, Chi, ... en fonction d' α et de nombre de degrés de liberté (s'il faut)
 - ou à l'aide des logiciels statistiques (ex. Excel, avec Data Analysis l'offre pour chaque test)
- on détermine
 - la **région critique** ou **région de rejet (RR)**
 - la **région de non rejet RnR=R-RR**.
 - Pour un test non directionnel:
 - Pour le test **Z**, ou **t**
 - $RR = (-\infty, - \text{valeur critique}] \cup [\text{valeur critique}, +\infty)$
 - $RnR = (-\text{valeur critique}, \text{valeur critique})$
 - Pour le test **F** ou **Chi 2** :
 - $RR = [\text{valeur critique}, +\infty)$
 - $RnR = [0, \text{valeur critique})$

Étape 5.

- Calculer la valeur de la statistique / paramètre du test

Étape 6 : la décision statistique en fonction de la région critique :

- Si Z_0 est dans **RR** (région du rejet/critique)
 - on accepte **H_1** et on rejette **H_0**
- Si Z_0 est dans **RnR** (région de non rejet)
 - on reste avec **H_0** , on ne peut pas rejeter **H_0**

Étape 6. la décision statistique en fonction de la p-valeur:

Définition de la p-valeur := la probabilité d'obtenir, quand H_0 est supposée vraie, un résultat pour la statistique du test plus extrême que le résultat observé. Dans ce TP le résultat signifie la différence entre les statistiques (moyennes) observées sur deux échantillons (mais il peut être la différence entre deux fréquences...)

- Si $p\text{-value} \leq \alpha (0,05) \Rightarrow$ on accepte **H_1** et on rejette **H_0**
- Si $p\text{-value} > \alpha (0,05) \Rightarrow$ on reste avec **H_0** \Rightarrow mais ça ne signifie pas que H_0 est vraie !!! Donc on va affirmer : on ne peut pas dire qu'il y a une différence statistiquement significative entre / on ne peut pas dire qu'il y a une lien/association/relation statistiquement significative entre

Exemples des hypothèses statistiques

Exemple général

- L'hypothèse nulle: H_0
 - Il n'y a pas une différence statistiquement significative entre ≥ 2 groupes (ex. un traitement [ibuprofène vs. placebo] ou un facteur de risque [présent vs. absent]) en ce qui concerne la moyenne/ médiane/ variance/

- fréquence ... d'une caractéristique (ex résultat du traitement: la fréquence du guérison [oui vs. non] ou la moyenne de la température)
- **Il n'y a pas** un **relation/lien/association/dépendance/corrélation statistiquement significative** entre 2 caractéristiques/variables: (ex. un Facteur de risque [présent vs. absent] – une maladie [oui vs. non] , ou un traitement [ibuprofène vs. placebo] – le résultat du traitement (ex: la fréquence du guérison [oui vs. non] ou la moyenne de la température))
 - L' hypothèse **alternative**: H_1 (négation du H_0)
 - **Il y a** un **différence statistiquement significative** entre $2/\geq 2$ groups en ce qui concerne la **moyenne/ médiane/ variance/ fréquence ... d'une caractéristique**
 - **Il y a** un **relation/lien/association/dépendance/corrélation statistiquement significative** entre 2 caractéristiques/variables

Exemple pour le test t (Student) pour échantillons indépendants

- L' hypothèse **nulle**: H_0
 - **Il n'y a pas** un **différence statistiquement significative** entre 2 groups (ex. un traitement [ibuprofène vs. placebo]) en ce qui concerne la **moyenne d'une caractéristique** (ex la moyenne de la température)
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - **Il y a** un **différence statistiquement significative** entre $=2$ groups en ce qui concerne la **moyenne ... d'une caractéristique**

Exemple pour le test t (Student) pour échantillons dépendants (appariées)

- L' hypothèse **nulle**: H_0
 - **Il n'y a pas** un **différence statistiquement significative** entre la **moyenne d'une caractéristique** (ex la moyenne de la température) **mesurée avant et après une intervention thérapeutique** (ex. ibuprofene)
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - **Il y a** une **différence statistiquement significative** entre la **moyenne d'une caractéristique** (ex la moyenne de la température) **mesurée avant et après une intervention thérapeutique** (ex. ibuprofene)

Exemple pour le test F pour les variances pour échantillons indépendants

- L' hypothèse **nulle**: H_0
 - **Il n'y a pas** un **différence statistiquement significative** entre 2 groups (ex. un traitement [ibuprofène vs. placebo]) en ce qui concerne la **variance d'une caractéristique** (ex la variance de la température)
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - **Il y a** un **différence statistiquement significative** entre $=2$ groups en ce qui concerne la **variance ... d'une caractéristique**

Exemple pour le test Khi-deux, ou Fisher exact

- Question possible : Est-ce que il y a une association entre le facteur de risque (présent/absent) et la maladie (présente/absente)?

- L' hypothèse **nulle**: H_0
 - Il n' y a pas de relation* statistiquement significative entre le facteur de risque et la maladie
 - Il n' y a pas d' association* statistiquement significative entre le facteur de risque et la maladie
 - Il n' y a pas de lien* statistiquement significative entre le facteur de risque et la maladie
 - Il n' y a pas de dépendance * statistiquement significative entre le facteur de risque et la maladie
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans maladie en ce qui concerne la fréquence du facteur de risque
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans exposition au facteur de risque en ce qui concerne la fréquence de la maladie.
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - Il y a un relation statistiquement significative entre le facteur de risque et la maladie
 - Il y a un association statistiquement significative entre le facteur de risque et la maladie
 - Il y a un lien statistiquement significative entre le facteur de risque et la maladie
 - Il y a un dépendance statistiquement significative entre le facteur de risque et la maladie
 - Il y a un différence statistiquement significative entre ceux avec ou sans maladie en ce qui concerne la fréquence du facteur du risque
 - Il y a un différence statistiquement significative entre ceux avec ou sans exposition au facteur de risque en ce qui concerne la fréquence de la maladie.
- Question possible : Est-ce que le traitement (oui/non) et le guérison (oui/non) sont liées?
 - L' hypothèse **nulle**: H_0
 - Il n' y a pas de relation statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de association statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de lien statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de dépendance statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans traitement en ce qui concerne la fréquence du guérison
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans guérison en ce qui concerne la fréquence du traitement.
 - L' hypothèse **alternative**: H_1 (négation du H_0)
 - Il n' y a pas de relation statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de association statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de lien statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de dépendance statistiquement significative entre le traitement et le guérison
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans traitement en ce qui concerne la fréquence du guérison
 - Il n' y a pas de différence statistiquement significative entre ceux avec ou sans guérison en ce qui concerne la fréquence du traitement.

Exemple pour le test statistique pour le coefficient de corrélation Pearson ou Spearman

- L' hypothèse **nulle**: H_0
 - il n'y a pas d'une différence statistiquement significative entre le coefficient de corrélation (des variables x et y) et 0
 - il n'y a pas d'une corrélation linéaire statistiquement significative entre les variables x et y
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - il y a une différence statistiquement significative entre le coefficient de corrélation (des variables x et y) et 0
 - il y a une corrélation linéaire statistiquement significative entre les variables x et y

Exemple pour le test statistique pour le pente de la régression linéaire

- L' hypothèse **nulle**: H_0
 - il n'y a pas une différence statistiquement significative entre le coefficient (la pente) de la variable x et 0
 - le coefficient (la pente) de la variable x n'est pas statistiquement significative
- L' hypothèse **alternative**: H_1 (négation du H_0)
 - il y a une différence statistiquement significative entre le coefficient (la pente) de la variable x et 0
 - le coefficient (la pente) de la variable x est statistiquement significative

Exemples de créations des hypothèses statistiques avec le choix des hypothèses en fonction du p

P<0.05

T.1.1. Énoncez les hypothèses nulle et alternative, pour le test qui vérifie s'il y a une différence entre les étudiants dans des familles de type nucléaire ou commun en ce qui concerne niveau moyenne d'anxiété (voir dans le tableau les lignes pour Family Type) :

- H_0 (hypothèse nulle): il n'y a pas une différence statistiquement significative entre les étudiants dans des familles de type nucléaire ou commun en ce qui concerne la moyenne du niveau d'anxiété.
- H_1 (hypothèse alternative - test bilatéral): il y a une différence statistiquement significative entre les étudiants dans des familles de type nucléaire ou commun en ce qui concerne la moyenne du niveau d'anxiété.

T.1.2. Écrivez le nom du test utilisé pour la comparaison:

- test t de Student pour échantillons indépendants (voir independent sample t-test dans le tableau).

T.1.3. Ecrivez la valeur du P du test (voir la colonne P)

- $p=0.004$

T.1.4. Interpréter du point de vue statistique le résultat du test statistique, et argumentez votre réponse

- il y a une différence statistiquement significative entre les étudiants dans des familles de type nucléaire ou commun en ce qui concerne la moyenne du niveau d'anxiété
- parce que $p=0.004$ est plus petit que le niveau de signification de 0.05

P>0.05

Voici un exemple des demandes avec des reposes déjà écrits

T.3.1. Énoncez les hypothèses nulle et alternative, pour le test qui vérifie s'il y a une différence entre les étudiants depuis le debut de l'année et à la fin de l'année en ce qui concerne niveau moyenne d'agrabilité (voir dans le tableau la ligne pour Agreableness) :

- H0 (hypothèse nulle): il n'y a pas une différence statistiquement significative entre les étudiants depuis le debut de l'année et à la fin de l'année en ce qui concerne niveau moyenne d'agrabilité.
- H1 (hypothèse alternative - test bilatéral): il y a une différence statistiquement significative entre les étudiants depuis le debut de l'année et à la fin de l'année en ce qui concerne niveau moyenne d'agrabilité.

T.3.2. Écrivez le nom du test utilisé pour la comparaison:

- test t de Student pour échantillons dépendants/appariés (voir paired t-test dans le tableau).

T.3.3. Ecrivez la valeur du P du test (voir la colonne P)

- $p=0.502$

T.3.4. Interpréter du point de vue statistique le résultat du test statistique, et argumentez votre réponse

- on ne peut pas dire qu'il y a une différence statistiquement significative entre les étudiants depuis le debut de l'année et à la fin de l'année en ce qui concerne niveau moyenne d'agrabilité.
- parce que $p=0.502$ est plus grand que le niveau de signification de 0.05

Comment faire le choix entre des tests statistiques avec échantillons indépendants ou dépendants (appariés)

- **Echantillons dépendants (appariés)** : ou chaque sujet de l'échantillon a un correspondant avec des caractéristiques similaires dans l'autre échantillon
 - Comparaison entre des **jumeaux**
 - Comparaison sur les mêmes sujets **observée répétée** dans le temps (ex. **avant et après** le traitement)
 - Comparaison entre des **échantillons appariés (matched** – en anglais), par les chercheurs (pour chaque sujet cas on a un sujet témoin avec les mêmes caractéristiques – âge, sexe, et autres facteurs de confusion)
 - Comparaison **sur le même sujet** (on compare la **partie gauche et droit** sur la peau traitée avec des produits dermatologiques, ou on compare la partie antérieure et postérieure du scalp)
 - Comparaison des **deux méthodes de mesure** ou **diagnostiques** utilisées **sur les mêmes sujets** (eg. On compare une tensiomètre classique avec un tensiomètre digital sur les mêmes sujets)
- **Echantillons indépendants** : ou chaque sujet de l'échantillon n'a pas un correspondant avec des caractéristiques similaires dans l'autre échantillon (on le déduit, en excluant les situations des échantillons appariés)

Le choix du test t - Student

Le test Student :

- **Pour échantillons indépendants**

- les observations sont indépendantes à l'intérieur des groupes et d'un groupe à l'autre
 - Ex. on compare des sujets qui ont reçu un traitement hypocholestérolémiant et des sujets qui ont reçu un placebo
 - Il y a deux types:
 - Avec **variances égales - t-Test: Two-Sample Assuming Equal Variances**
 - Avec **variances inégales - t-Test: Two-Sample Assuming Unequal Variances**
 - **La chois sera faite avec le test F pour comparer les variances** (voir en bas).
- **Pour échantillons appariés – t-Test : Paired Two Sample for Means**
- la situation des mesures répétées sur les mêmes sujets.
 - Ex. on mesure la fréquence cardiaque d'un patient avant et après la prise d'un bêta bloquant
 - on compare des jumeaux
 - on compare des observations faites sur la partie gauche et droit d'un sujet (test dermatologique)

Réaliser les tests statistiques dans Excel avec Data Analysis

Pour réaliser les tests statistiques avec Data Analysis les données doit être ordonnées en ascendant en fonction de la variable de regroupement (ex. Si on doit comparer le traitement avec le placebo, on doit trier la variable Traitement). Pour cela cliquez sur le titre de la variable de regroupement (ex. traitement), et dans l'onglet Home, section **Editing : Sort & Filter** la commande **Sort A-Z**.

Le test Student - Pour échantillons indépendants

(on procède d'une manière identique si les variances sont égales ou non)

N'oubliez pas de trier les données avant !!! Pour cela cliquez sur le titre de la variable de regroupement (ex. traitement), et dans l'onglet **Home**, section **Editing : Sort & Filter** la commande **Sort A-Z**, ou mieux, sélectionnez toutes les données, et dans l'onglet **Home**, section **Editing : Sort & Filter** la commande Custom Sort, et sélectionnez la colonne pour laquelle vous voulez trier les données.

Dans **Data Analysis** on choisit : **t-Test: Two-Sample Assuming Equal Variances**

Dans la fenêtre on sélection les données comme dans l'exemple suivant (l'exemple est différent dès notre données) :

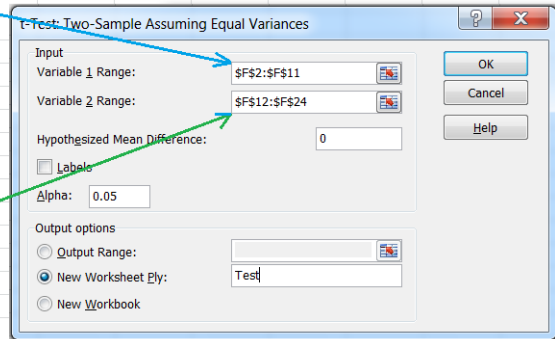
Variable 1 Range: les valeurs de la variable quantitative pour le premier group

Variable 2 Range: les valeurs de la variable quantitative pour le deuxième group

Alpha: la probabilité d'erreur de type I pour le test statistique. Utilisez la valeur par défaut de 5%.

Click sur le bouton **OK**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Tratament	Picioare umflate	Sex	Varsta (ani)	Hispanic	Tensiune arteriala sistolica initiala (mmHg)	Tensiune arteriala sistolica finala (mmHg)	Obezitate	Dureri de cap							
1																
2	nebivolol	da	F	50	da	165	135	da	da							
3	nebivolol	da	F	43	da	157	120	da	da							
4	nebivolol	da	F	52	da	158	144	da	da							
5	nebivolol	da	F	31	da	150	160	da	da							
6	nebivolol	da	F	36	da	165	159	da	da							
7	nebivolol	da	F	39	da	169	167	da	da							
8	nebivolol	da	F	44	da	146	106	da	da							
9	nebivolol	da	F	45	da	155	136	da	da							
10	nebivolol	nu	F	50	da	145	157	da	da							
11	nebivolol	nu	F	53	da	175	123	da	da							
12	placebo	nu	F	39	da	159	129	da	da							
13	placebo	nu	F	33	da	161	113	da	da							
14	placebo	nu	F	57	da	138	111	da	nu							
15	placebo	nu	F	44	da	161	125	da	nu							
16	placebo	nu	F	46	da	167	128	da	nu							
17	placebo	nu	F	51	da	144	107	da	nu							
18	placebo	nu	F	53	da	145	131	da	nu							
19	placebo	nu	F	43	da	128	134	da	nu							
20	placebo	nu	F	42	da	145	142	da	nu							
21	placebo	nu	F	39	da	163	134	da	nu							
22	placebo	nu	F	31	da	144	135	da	nu							
23	placebo	nu	F	50	da	136	134	da	nu							
24	placebo	nu	F	44	da	165	161	da	nu							



On obtienne un tableau comme le suivant

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	Variable 2
Mean	158.4457346	150.503302
Variance	99.5457539	160.8658677
Observations	10	13
Pooled Variance	134.585819	
Hypothesized Mean Difference	0	
df	21	
t Stat	1.627652338	
P(T<=t) one-tail	0.059255249	
t Critical one-tail	1.720742903	
P(T<=t) two-tail	0.118510498	
t Critical two-tail	2.079613845	

Interprétation des résultats :

t Critical two-tail – représente la valeur critique avec lequel on peut créer les régions du rejet et d'acceptance. Ex. Ici la région du rejet pour le test t bidirectionnel est : $(-\infty, -2.079]$ U $[2.079, +\infty)$. La région d'acceptance est : $(-2.079, 2.079)$

t-Stat – représente la statistique du test t. Avec elle on va décider l'hypothèse à choisir, avec la région du rejet. (ex. 1,62)

P(T<=t) two-tail – représente la valeur du p pour les tests bidirectionnels. Avec elle on va décider l'hypothèse à choisir, avec le niveau de signifiante alpha.

Le test Student - Pour échantillons dépendants

N'oubliez pas de trier les donnes avant !!! Pour cela cliquez sur le titre de la variable de regroupement (ex. traitement), et dans l'onglet **Home**, section **Editing** : **Sort & Filter** la commande **Sort A-Z**, ou mieux, sélectionnez toutes les données, et dans

l'onglet **Home**, section **Editing** : **Sort & Filter** la commande Custom Sort, et sélectionnez la colonne pour laquelle vous voulez trier les données.

Dans **Data Analysis** on choisit : **t-Test : Paired Two Sample for Means**

Dans la fenêtre on sélectionne les données comme dans l'exemple suivant (l'exemple est différent de nos données) :

Variable 1 Range: les valeurs de la variable quantitative pour le premier group (ex. avant le traitement – initiale)

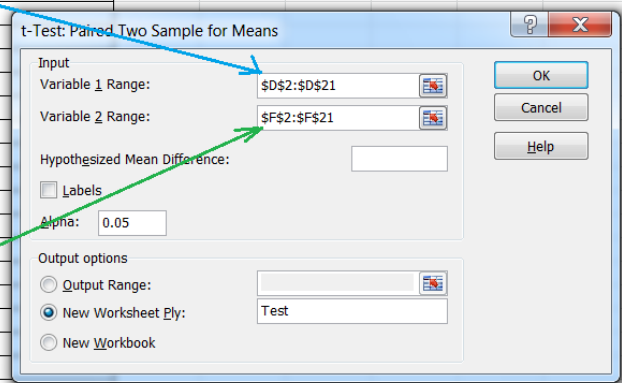
Variable 2 Range: les valeurs de la variable quantitative pour le deuxième group (ex. après le traitement – finale)

Alpha: la probabilité d'erreur de type I pour le test statistique. Utilisez la valeur par défaut de 5%.

Click sur le bouton **OK**

1

	A	B	C	D	E	F	G	H	I	J	K	L	M
	ID	Milieu	Maladie cardiovasculaire	Glycémie initiale (mg/dl)	Traitement	Glycémie finale (mg/dl)	Reponse au traitement						
1													
2	4	u	oui	148.55	hypoglycemique	78.06							
3	13	r	non	148.96	hypoglycemique	105.73							
4	15	u	oui	122.70	hypoglycemique	100.20							
5	19	r	non	181.77	hypoglycemique	98.18							
6	21	u	oui	143.67	hypoglycemique	123.92							
7	24	u	non	132.24	hypoglycemique	96.16							
8	26	r	non	152.29	hypoglycemique	114.55							
9	29	r	oui	114.39	hypoglycemique	103.93							
10	31	u	oui	134.96	hypoglycemique	134.37							
11	32	u	non	183.28	hypoglycemique	64.92							
12	34	r	oui	156.95	hypoglycemique	110.36							
13	41	r	oui	131.48	hypoglycemique	79.25							
14	44	r	oui	139.03	hypoglycemique	113.99							
15	46	u	non	158.86	hypoglycemique	115.45							
16	48	u	non	165.80	hypoglycemique	128.32							
17	50	r	oui	149.75	hypoglycemique	108.23							
18	51	r	oui	121.95	hypoglycemique	127.24							
19	55	u	oui	150.19	hypoglycemique	125.78							
20	56	u	non	177.08	hypoglycemique	115.71							
21	57	r	non	126.97	hypoglycemique	130.84							
22	1	u	oui	148.24	placebo	127.54							
23	2	u	non	125.54	placebo	177.23							
24	3	u	oui	139.47	placebo	121.96							
25	5	r	non	157.61	placebo	126.86							

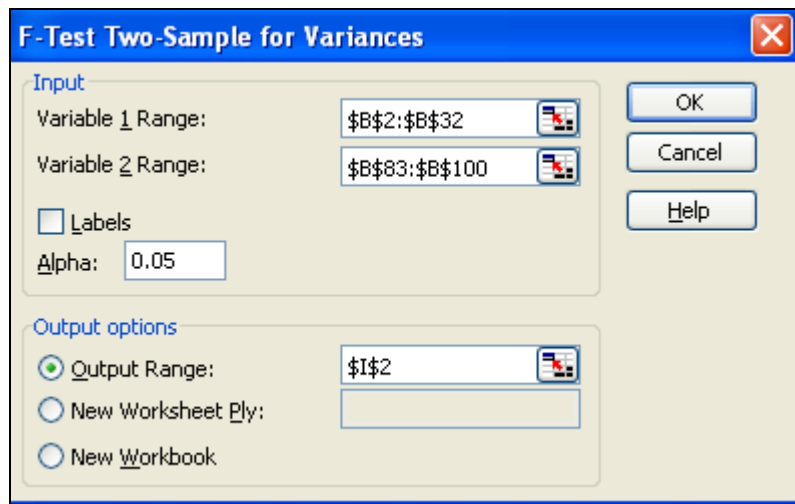


Interprétation des résultats : voir le test pour les échantillons indépendants

Le test F pour comparer les variances

N'oubliez pas de trier les données avant !!! Pour cela cliquez sur le titre de la variable de regroupement (ex. traitement), et dans l'onglet **Home**, section **Editing** : **Sort & Filter** la commande **Sort A-Z**, ou mieux, sélectionnez toutes les données, et dans l'onglet **Home**, section **Editing** : **Sort & Filter** la commande Custom Sort, et sélectionnez la colonne pour laquelle vous voulez trier les données.

Le test Fisher se trouve dans le menu: **TOOLS-DATA ANALYSIS-F-test Two sample for Variances**



Variable 1 Range: les valeurs de la variable quantitative pour le premier group

Variable 2 Range: les valeurs de la variable quantitative pour le deuxième group

Alpha: la probabilité d'erreur de type I pour le test statistique.

Utilisez la valeur par défaut de 5%.

Output options: la cellule ou commencer de mettre les résultats

Cliquez sur le bouton **OK**.

Interprétation de résultats:

Mean: moyenne pour le premier et pour le deuxième group

Variance: variance pour le premier et pour le deuxième group

Observations: nombre des sujets par group.

df: le degré de liberté ($n_1 - 1, n_2 - 1$)

F: le paramètre calculé est 1.74

P-value one-tail: la probabilité unilatérale.

- Si $p > 0.05$, le résultat n'est pas statistiquement significative - on ne peut pas dire que les variances pour les deux groupes sont inégales, donc on va utiliser le test Student pour échantillons indépendants avec variation égales
- Si $p \leq 0.05$, le résultat est statistiquement significative - les variances pour les deux groupes sont inégales, donc on va utiliser le test Student pour échantillons indépendants avec variation inégales

Réaliser le test Khi-deux dans Excel

On commence en réalisant le **tableau de contingence observée** entre les deux variables qualitatives avec PivotTable

Dans la même feuille de calcul on va créer le **tableau théorique**. On recopie le tableau observé, avec Copier + Coller (sans formatage, seulement les valeurs). On efface les valeurs du contenu du tableau, mais on préserve les totaux. Puis on calcule les fréquences théoriques en utilisant les formules indiquées ci-dessous :

	F	G	H
1	Tableau de contingence theorique		
2		Nebivolol	Placebo
3	Femmes		215
4	Hommes		426
5	Total	427	214

	A	B	C	D	E	F	G	H	I
1	Tableau de contingence observe				Tableau de contingence theorique				
2		Nebivolol	Placebo	Total			Nebivolol	Placebo	Total
3	Femmes	120	95	215		Femmes	143	72	215
4	Hommes	307	119	426		Hommes	284	142	426
5	Total	427	214	641		Total	427	214	641
6									
7									
8						Valeur critique du Khi carré	3,84		
9						Niveau de signification (alfa)	0,05		
10						df (Degré de liberté)	1,00		
11						p-valeur			
12						Statistique de test	16,63		

On determinez la **p-valeur** en utilisant la **fonction** prédéfinie **CHITEST** où Actual_range représente les frequences observées et Expected_range représente les frequences attendues.

On peut **calculer la statistique du test** en utilisant la formule suivante:

$$\chi^2 = \sum_{i=1}^{r \times c} \frac{(f_i^o - f_i^t)^2}{f_i^t}$$

ou χ^2 est la statistique du test, r = nombre de lignes dans un tableau de

contingence, c = nombre de colonnes dans un tableau de contingence, f_i^o =fréquences observées, f_i^t =fréquences théoriques;

Pour le test statistique on considere un niveau de signification egal a 5% et la **valeur critique** égale à 3,84 (si les deux variables sont dichotomiques). La **region du rejet** est [valeur critique, + infini)

Probabilités

Théorie

a) Si A et B sont incompatibles (ils ne peuvent pas se produire simultanément), alors :

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

b) $\Pr(\emptyset) = 0$.

c) Si $A \subseteq B$ alors $\Pr(A) \leq \Pr(B)$.

d) Pour tous les événements aléatoires A et B, nous avons l'égalité:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) .$$

e) Si deux événements A et B sont indépendants en probabilité, alors :

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

f) Si A est le contraire du B, $\Pr(A) = 1 - \Pr(B)$.

Probabilités conditionnels - Risque relatif (RR)

- On doit créer un tableau de contingence (ex : pour les variables Genre et le Test Standard: en utilisant PivotTable (tableau croisé dynamique). Dans le champ « row » mettez le facteur (genre) et dans le champ "colonne" mettez la maladie (Test Standard))
- le calcul du risque relatif conformément à sa formule: $RR = \Pr(M+ | F+) / \Pr(M+ | F-)$ ou $M = \{\text{maladie}\}$ et $F = \{\text{facteur étudié}\}$
- les deux probabilités conditionnelles de la formule du RR peuvent être approximé par les fréquences relatives :

	M+	M-	Total
F+	a	b	a + b
F-	c	d	c + d
Total	a + c	b + d	n

$M+$ = malade ; $M-$ = sans la maladie ; $F+$ = être exposé au facteur F ; $F-$ = n'être pas exposé au facteur F

$$\text{Donc le RR estimé} = \frac{a/(a+b)}{c/(c+d)}$$

Interpretation

Si $RR = 1 \Rightarrow$ le risque d'avoir la maladie est le même dans les deux groups de sujets (les sujets exposés et non exposés au facteur F)

Si $RR > 1 \Rightarrow$ le risque d'être malade est plus élevé chez un patient exposé au facteur F (par exemple un $RR=3$ déterminé dans l'échantillon signifie qu'un patient exposé a un risque de maladie 3 fois supérieur à celui d'un patient non-exposés)

Si $RR < 1 \Rightarrow$ le risque d'être malade est plus diminué chez les sujets exposés au facteur F

Probabilités conditionnels - Tests diagnostiques

- On doit créer le tableau de contingence entre les variables (ex : Test 1 et le Test Standard en utilisant l'option PivotTable (tableau croisé dynamique). Dans le champ « row » mettez le test (Test1 par exemple) et dans le champ "colonne" mettez le résultat de l'autopsie.)
- Trier ascendant les catégories de la variable Test 1 pour obtenir un tableau de contingence ayant la forme affiché au-dessous :

	Test Standard (malade)	Test Standard (sans maladie)	Total

Test 1+	a	b	a + b
Test1 -	c	d	c + d
Total	a + c	b + d	n

$T+$ = test positif ; $T-$ = test négatif

- Calculez, à côté du tableau croisé dynamique les indicateurs suivants :

Sensibilité (Se) = la probabilité d'avoir un test positif si la personne choisie au hasard est malade (la maladie est identifiée par le test standard=résultat positif à l'autopsie) :

$$Se = \Pr(T+ / M+) \approx a / (a + c)$$

Spécificité (Sp) = probabilité d'avoir un test négatif sachant que la personne choisie au hasard n'est pas malade :

$$Sp = \Pr(T^- / M^-) \approx d / (b + d)$$

Les éléments de décision pour le médecin

Valeur prédictive positive (VPP) = la capacité du test à identifier correctement parmi les patients dont le test est positif, ceux qui ont réellement la maladie= la probabilité d'avoir la maladie, sachant que le test est positif.

$$VPP = \Pr(M^+ | T^+) \approx a / (a + b)$$

Valeur prédictive négative (VPN) = la capacité du test à identifier correctement parmi les patients dont le test est négatif, ceux qui n'ont pas réellement la maladie= la probabilité de n'avoir pas la maladie, sachant que le test est négatif.

$$VPN = \Pr(M^- | T^-) \approx d / (c + d)$$