

Instructions pour résoudre les travaux pratiques

Travailler avec Epi Info	3
Téléchargement du logiciel <i>Epi Info</i> (pour utilisation à domicile)	3
Ouverture du programme Epi Info	3
Ouverture de l'outil de l'analyse statistique	3
L'Importation d'un fichier Excel dans Epi Info	4
Sélectionner le dossier ou les résultats de l'analyse seront enregistrés	7
Montrer les données	8
Sélection d'un sous ensemble des données (analyser des sous-groupes)	9
Activation du module <i>Data Analysis</i> dans <i>Microsoft Excel</i>	10
Statistiques descriptives	10
Données qualitatives (catégoriques):	10
Tableaux de fréquences	10
Diagrammes circulaires (Camembert / Pie)	11
La réalisation des tableaux des fréquences – avec des intervalles de confiance dans Epi Info	12
Les tableaux de contingence	13
Le graphique colonnes associé à un tableau de contingence	13
Données quantitatives:	14
Description individuelle des variables quantitatives	14
Moyenne, médiane, écart type, intervalle de confiance 95% pour les moyennes	14
Tableau de fréquence et l'histogramme	14
Description d'un lien potentiel entre deux variables quantitatives	16
Scatter Chart (diagramme de dispersion)	16
Données de survie:	18

Médiane du temps de survie	18
Graphique de la probabilité de survie	18
Analyse des données.....	19
Effectuer un test de Student (t-test) dans Excel.....	19
Le calcul du tableau de contingence, Risque relatif (RR) et odds ratio (OR) dans Epi Info Exécution d'un test Chi- carré (χ^2) dans Epi Info.....	20
Comparaisons données quantitatives (Test student/ANOVA/ ...) dans Epi Info	22
Effectuer un test Log-rank pour l'analyse de la survie dans Epi Info	24
Exécution d'une régression de Cox et le calcul du Hazard Ratio (HR) dans Epi Info	24
Exécution d'une régression linéaire simple dans Microsoft Excel	25
Exécution d'une régression linéaire multiple dans Microsoft Excel	25
Réaliser un graphique Kaplan-Meier et un test Log-rank pour l'analyse de la survie sur Internet	26
Calculer le Hazard ratio (HR) pour l'analyse de la survie sur Internet	33

Travailler avec Epi Info

Téléchargement du logiciel *Epi Info* (pour utilisation à domicile)

Utilisez un navigateur Internet pour naviguer

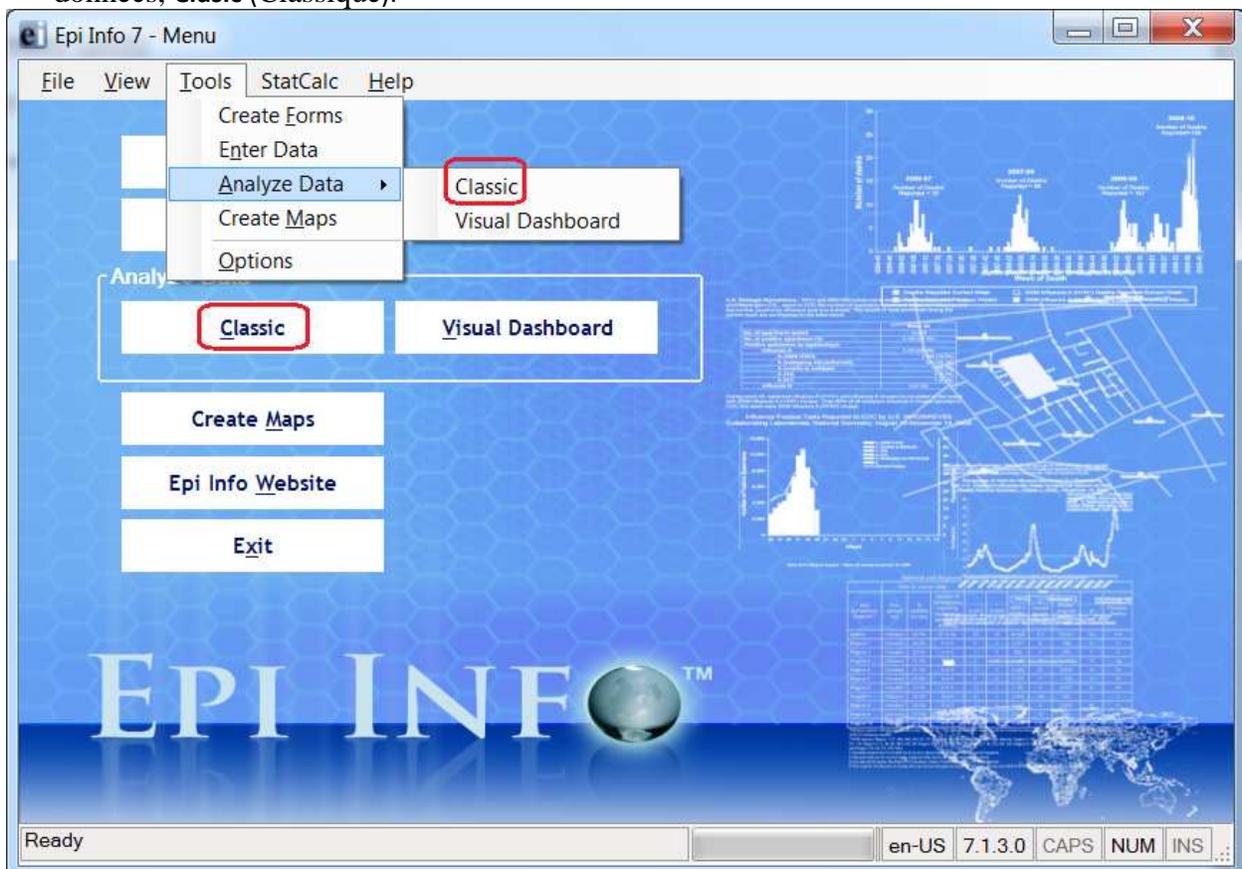
vers: <http://www.cdc.gov/epiinfo/installation.htm> et suivez les instructions d'installation.

Ouverture du programme *Epi Info*

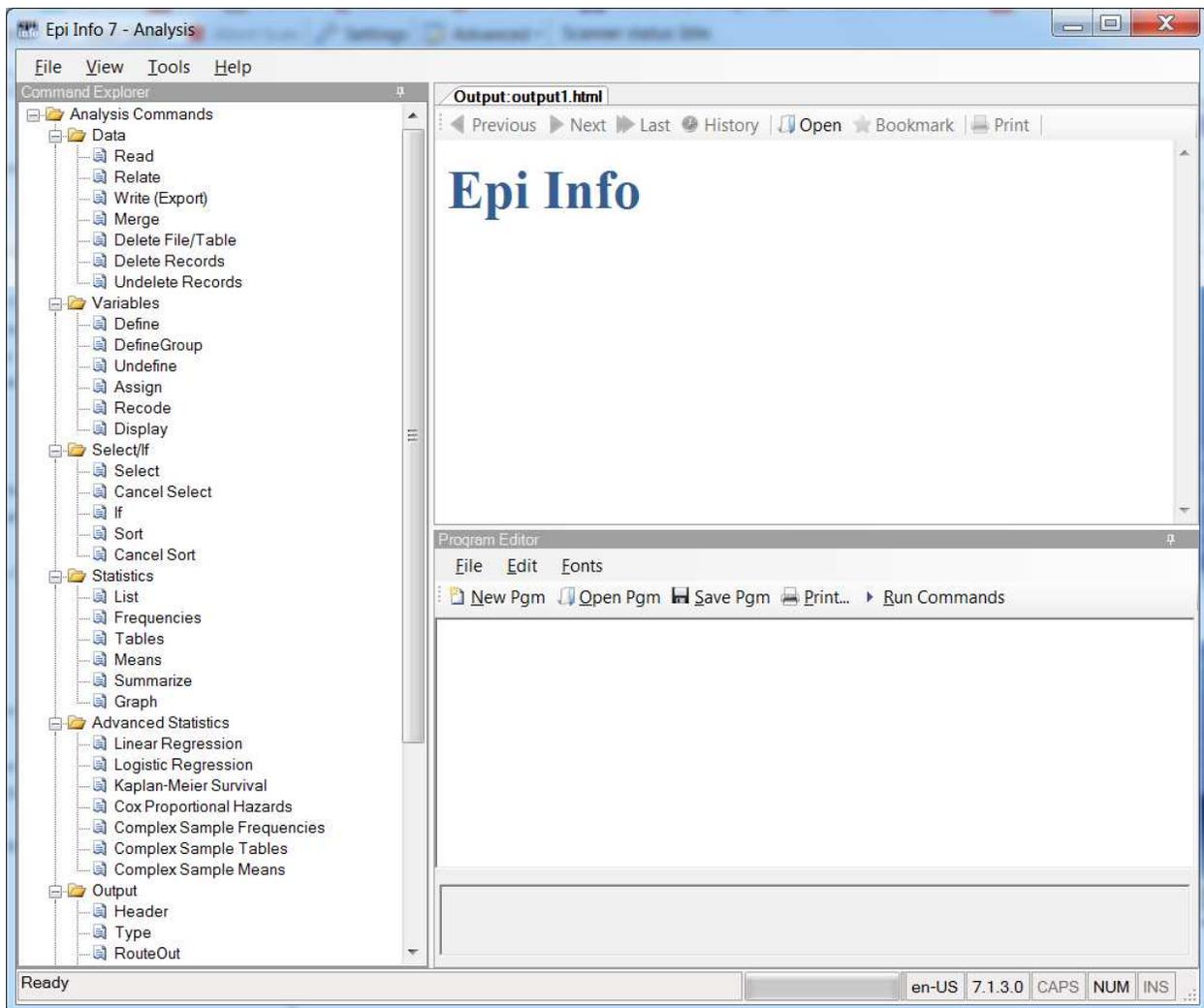
Le bouton **Démarrer**, raccourci **Epi Info™ 7** ou **Démarrer, All Programs, CDC, Epi Info 7, Epi Info™ 7**

Ouverture de l'outil de l'analyse statistique

Dans la fenêtre du programme, appuyez sur le bouton **Clasic** (Classique) dans la section **Analyze Data** (Analyser les données), ou entrer dans le menu **Tools** (Outils), **Analyze Data** analyser les données, **Clasic** (Classique).



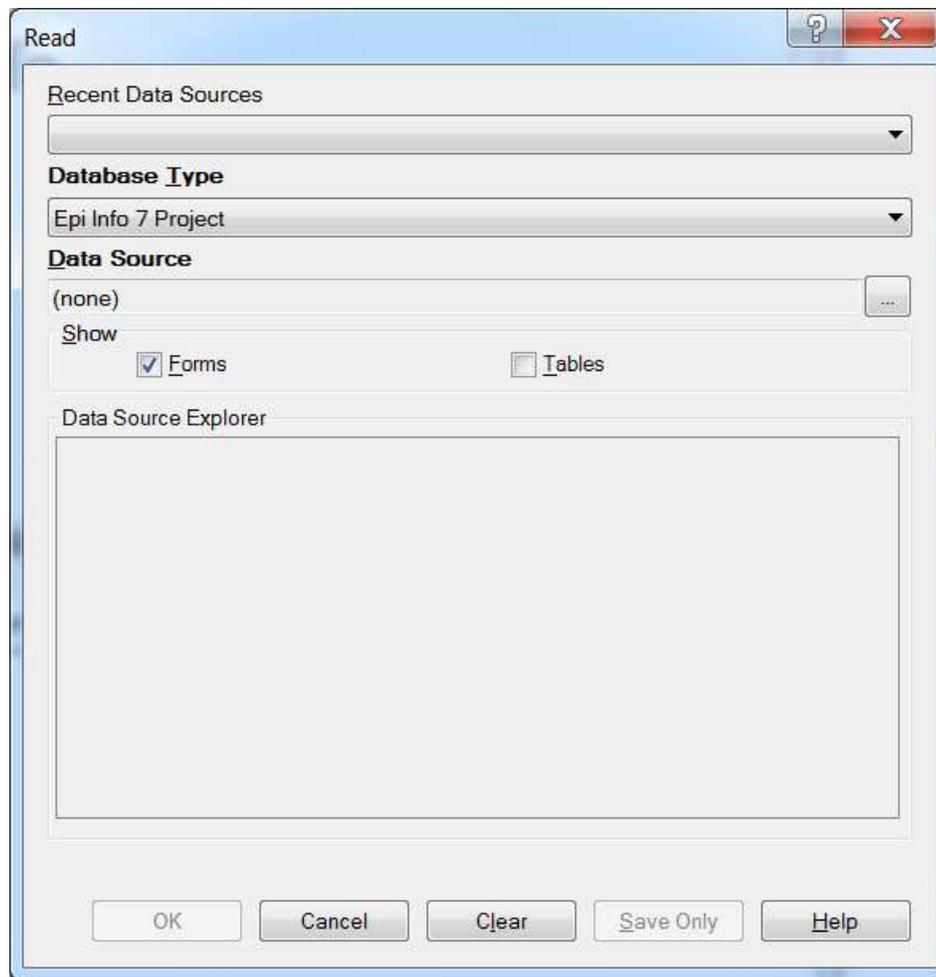
La fenêtre pour l'analyse des données:



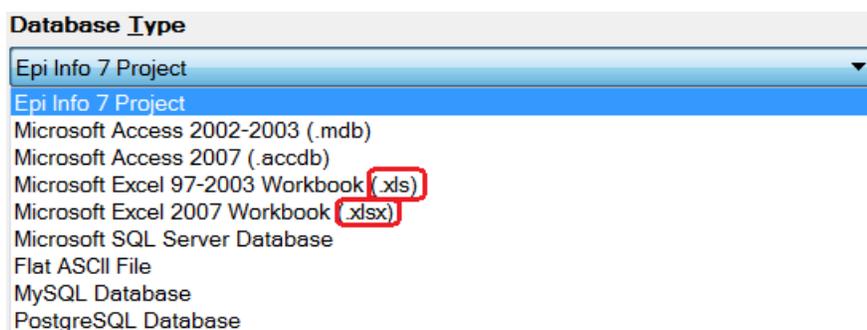
L'Importation d'un fichier Excel dans Epi Info

Pour importer un fichier Excel dans Epi Info, on fait trois opérations, en précisant le type de fichier à importer, le choix du fichier sur l'ordinateur et sélectionnez la feuille de calcul qui contient les données pour l'analyse.

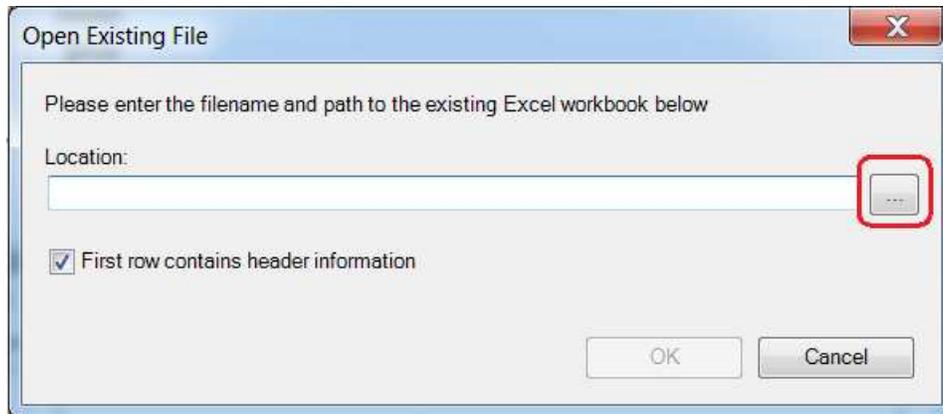
Dans la section des commandes (**Command Explorer** - le côté gauche de la fenêtre) on choisit dans **Analysis Commands**, la section des données : **Data** - la commande **Read** (Lire). Ouvrez la fenêtre d'importation des données. Choisissez le type de fichier à importer.



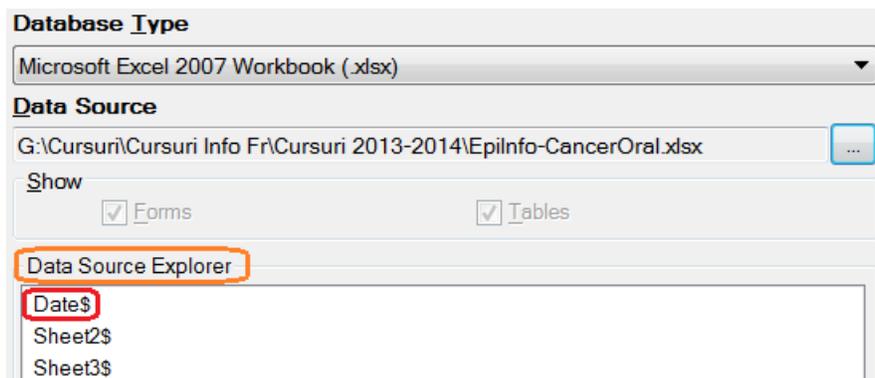
Si le fichier **Excel** a l'extension **.xlsx**, on choisit **Microsoft Excel 2007 Workbook (.xlsx)**, si le fichier a l'extension **.xls**, on choisit **Microsoft Excel 97-2003 Workbook (.xls)**.



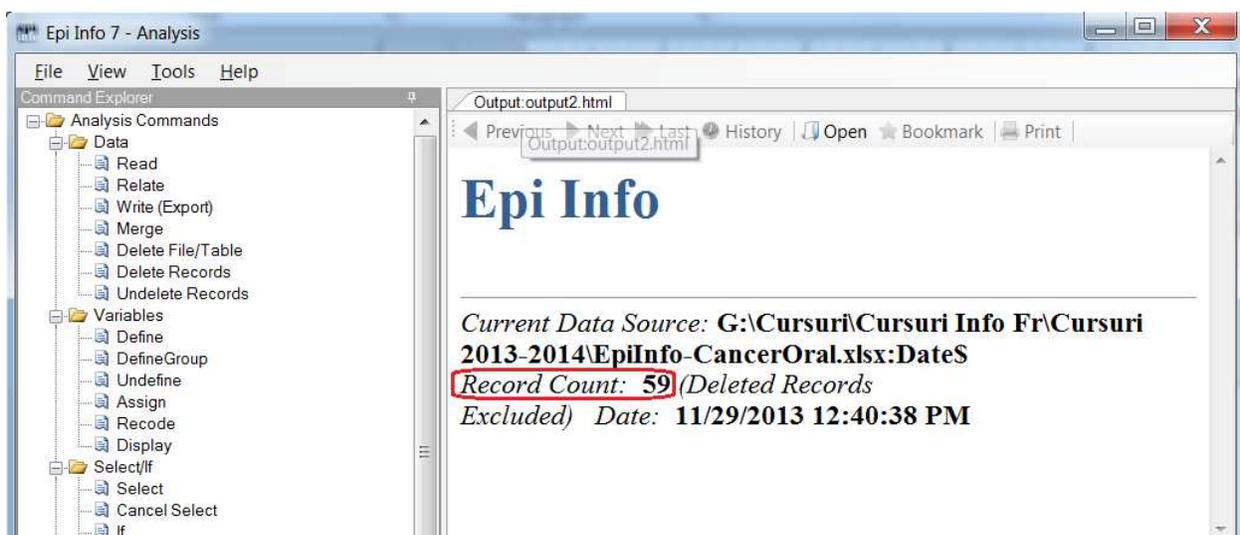
Après on choisit le fichier d'intérêt pour importation en appuyant le bouton  a cote de **Data Source**. Dans la fenêtre qui s'ouvre appuyez le bouton  a cote de **Location** (l'emplacement des fichiers), et cherchez le fichier d'intérêt. Après on a sélectionné le fichier on peut faire click sur le bouton **OK**, en laissant l'option **First row contains header information** (La première ligne contient les titres des colonnes) cochée.



Après ces opérations EpiInfo montre les feuilles des calculs contenue dans le fichier dans **Data Source Explorer** (Explorateur des sources des données). Sélectionnez la feuille de calcul qui contient les données pour l'analyse, et après appuyez le bouton **OK** pour finir l'importation.

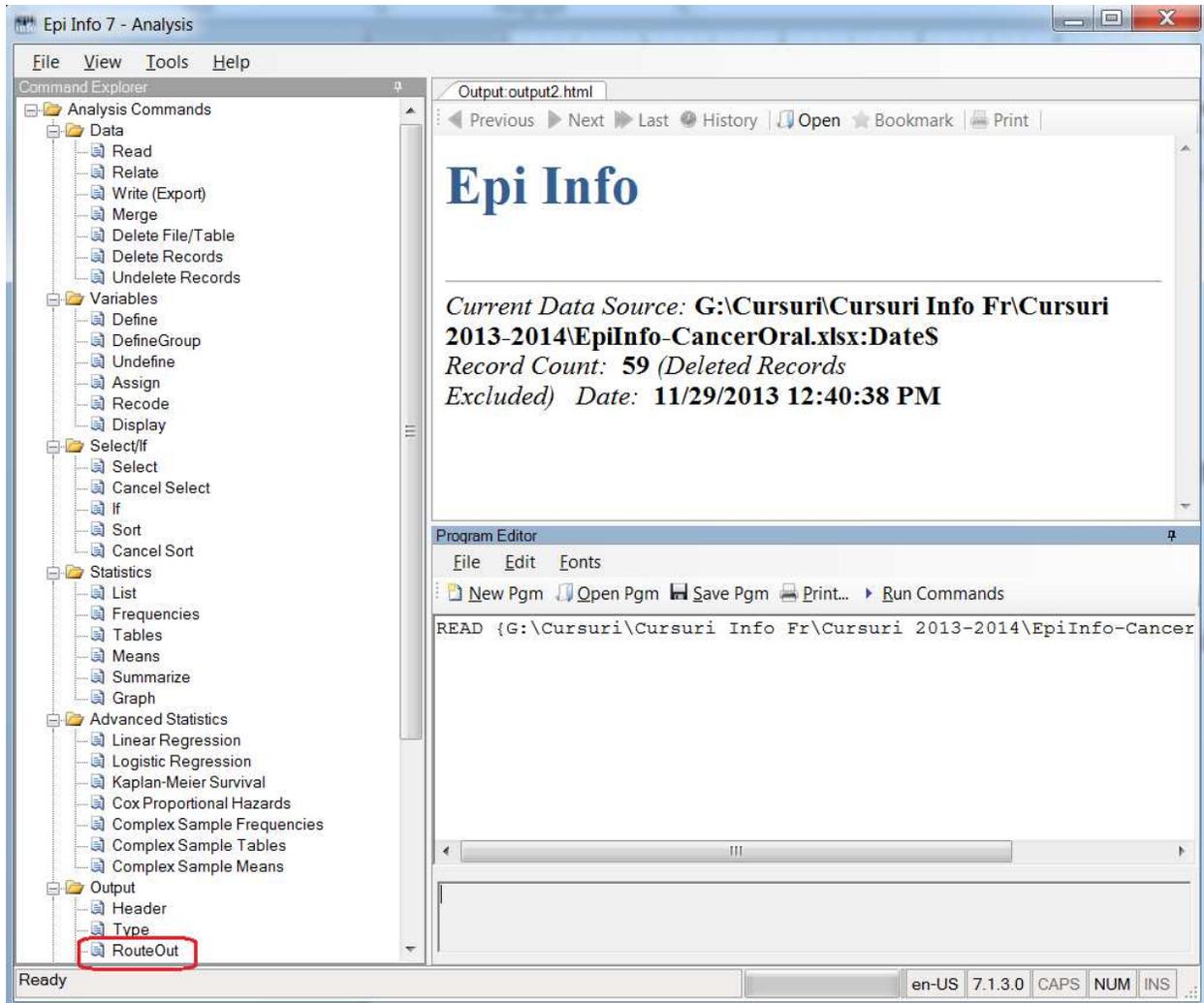


Après l'importation, Epi Info montre dans la zone des résultats (**Output**), le nom du fichier importée, et le nombre des enregistrements trouve dans le fichier (**Record Count**).

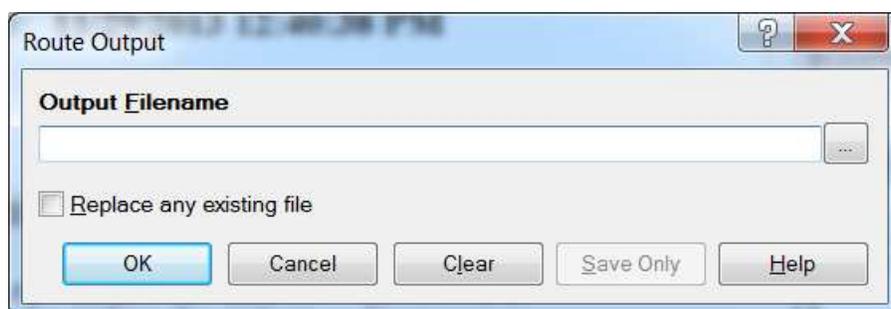


Sélectionnez le dossier ou les résultats de l'analyse seront enregistrée

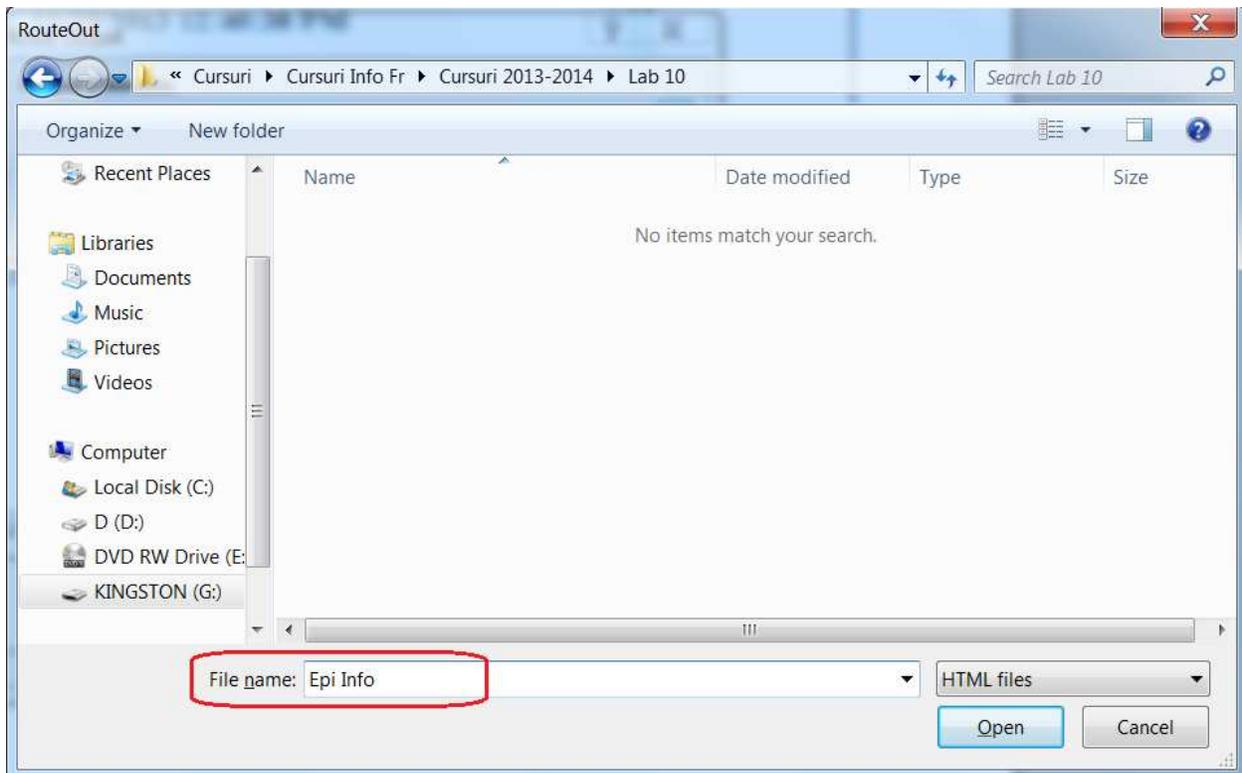
Optionnelle ! Dans la liste des commandes (**Command Explorer**) de la fenêtre de l'analyse de l'Epi Info, la section **Output** on choisit **RouteOut**.



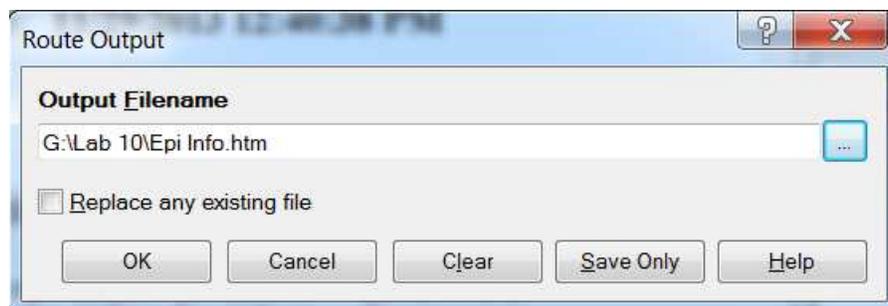
Dans la fenêtre qui s'ouvre cliquez sur le bouton  a cote de **Output Filename** (Noms du fichier des résultats)



Dans la fenêtre qui s'ouvre on doit chercher le dossier ou on veut enregistrer les résultats de l'analyse statistique. Entrez dans le dossier, et dans la fenêtre dans le champ **File name** (nom du fichier) écrivez le nom de fichier pour l'analyse statistique (ex. **Epi Info**). Puis click sur le bouton **Open**.



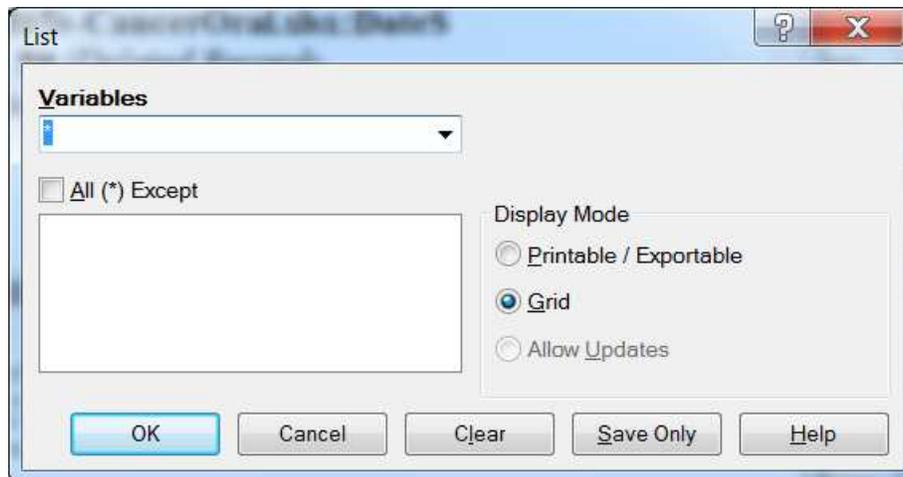
La fenêtre **Route Output** va montrer le nom de fichier et le dossier où l'analyse statistique sera enregistrée. Cliquez sur le bouton **OK**.



Montrer les données

Pour voir le contenu du fichier importé, dans la liste des commandes (**Command Explorer**) de la fenêtre d'analyse de Epi Info, la section **Statistics** (Statistique) on choisit la commande **List**.

Dans la fenêtre qui s'ouvre on laisse les options sans les changer et on fait click sur le bouton **OK**.

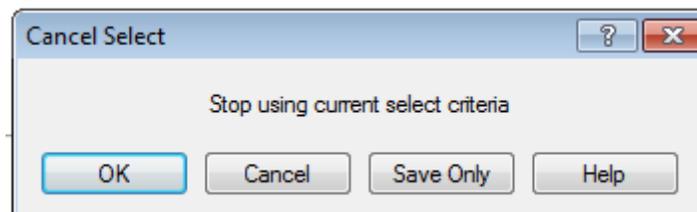


On vérifie si les données ont été bien importée. Puis on peut fermer la fenêtre.

ID	Sexe	Bain de bouche avec d alcool	Fumer ou tabac	Poids (kg)	Hauteur (cm)
1	H	Oui	Oui	84.0026120...	200.196383...
2	H	Oui	Non	60.2310990...	189.682870...
3	H	Oui	Oui	90.8516528...	182.657414...
4	F	Non	Non	66.6379889...	168.004268...
5	H	Oui	Oui	73.2323394...	166.644630...
6	H	Non	Oui	58.3638790...	163.924002...
7	H	Non	Oui	63.2351781...	198.272438...
8	F	Non	Oui	105.403508...	188.736623...

Sélection d'un sous ensemble des données (analyser des sous-groupes)

Pour comparer seulement deux groupes, bien que il y a trois ou plus, nous sélectionnons un sous-ensemble de données avec commande **Select** dans la section **Select/If**. Si vous avez déjà fait une sélection, on doit annuler la sélection précédente, avec la commande **Cancel Select** et **OK**.



La fenêtre affiche une condition logique de sélection. La variable de regroupement est la variable Groupe. Dans la fenêtre **Select** depuis **Available variables** et les boutons adjacents, vous pouvez sélectionner les variables puis effectuez votre sélection avec des opérateurs logiques (OR – ou, AND - et)

Dans **Select Criteria** on doit écrire (Groupe=1) OU (Groupe=2). Ça signifie de sélectionner pour analyse seulement des sujets dans le groupe 1 ou du groupe 2. Appuyez sur **OK**.

Activation du module *Data Analysis* dans *Microsoft Excel*

Cliquez sur une cellule vide, puis cliquez sur **Add-ins** dans le menu **Tools**. Cochez la case en regard de **Analysis ToolPack**, puis cliquez sur OK. Choisissez une autre cellule vide, puis recherchez **Data Analysis** dans le menu **Tools**.

Si **Data Analysis** ne figure pas dans le menu **Tools**, tout en étant coché dans **Add-Ins**, décochez la case dans **Add-Ins** et répétez la procédure ci-dessus.

Statistiques descriptives

Données qualitatives (catégoriques):

Tableaux de fréquences

Utilisez la **fonction COUNTIF** dans *Microsoft Excel* pour compter combien de fois chaque valeur prise par une variable apparaît dans la base de données (sa fréquence absolue).

Par exemple, pour savoir combien de personnes de sexe féminin (codé avec F dans le fichier) figurent dans l'échantillon, une cellule vide à l'emplacement futur du tableau de fréquences devrait contenir une formule analogue à =COUNTIF (A2: A58, "F"), Si les données concernant les sexes sont enregistrés dans les cellules A2 à A58. Un tableau de fréquence correcte devrait ressembler à ceci:

Tableau 1. Répartition par sexe dans l'échantillon étudié

Sexe	Nombre de sujets
Masculin	20
Féminin	37
Total	57

Notez que n'importe quel tableau doit être marqué sur le dessus de celui-ci, en utilisant un titre clair et précis. Sélectionnez le tableau, faites un clic droit sur cette sélection et choisissez **Caption** en vue de l'étiquette du tableau. Les étiquettes en rangée et de colonne doit être visible et facilement compréhensible par le lecteur, sans qu'il soit nécessaire de rechercher des explications supplémentaires afin de comprendre le contenu du tableau.

Diagrammes circulaires (Camembert / Pie)

Suivez les instructions ci-dessus pour créer un tableau de fréquence en utilisant **COUNTIF**.

Sélectionnez uniquement les cellules contenant les fréquences absolues et leurs étiquettes (ne choisissez pas le totale ou l'étiquette de colonne). Utilisez **Insert - Chart** et sélectionnez **Pie**. Cliquez sur **Next**. Dans la fenêtre **Chart Options** cliquez sur l'onglet **Data Labels** et cochez **Percentage**. Poursuivre et terminer l'assistant graphique. *Un diagramme circulaire correct devrait ressembler à ceci:*

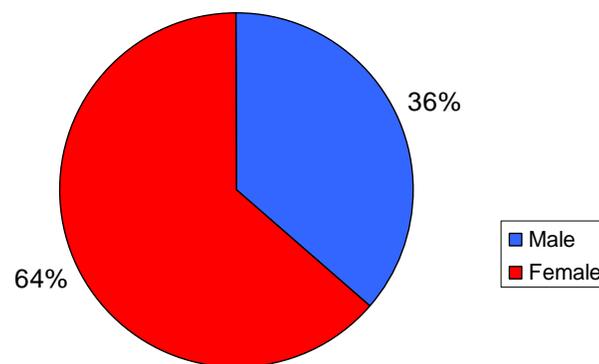


Figure 1. Répartition par sexe dans l'échantillon étudié

Notez que les graphiques doivent être étiquetés à l'aide des pourcentages visibles.

Si vous prévoyez d'utiliser le tableau dans une présentation **PowerPoint**, assurez-vous de l'étiqueter sur le dessus et d'utiliser un titre clair et précis (quoi, comment et lesquels sujets ont été représentés?).

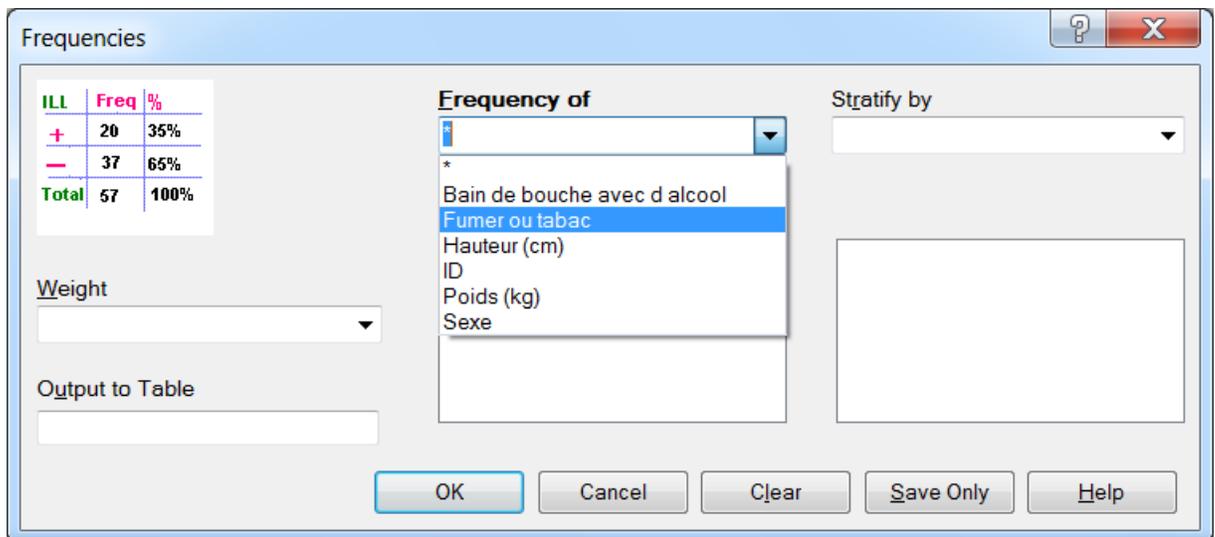
Si vous prévoyez d'utiliser le tableau comme une figure dans un document **Word**, effacez le titre du graphique dans **Excel**, mais n'oubliez pas d'étiqueter le graphique dans **Microsoft Word** comme vue dans l'exemple précédent: sélectionnez la figure, faites un clic droit sur cette sélection et utilisez **Caption**.

Toutes les étiquettes et les entrées de légende doivent être visible et facilement compréhensible par le lecteur, sans le besoin de recherche pour plus d'explications afin de comprendre le contenu de la figure.

La réalisation des tableaux des fréquence – avec des intervalles de confiance dans Epi Info

Dans la liste des commandes (**Command Explorer**) de la fenêtre d'analyse du Epi Info, la section **Statistics** (Statistique) on choisit la commande **Frequencies** (Fréquence).

Dans la fenêtre qui s'ouvre on choisit la variable pour laquelle on veut réaliser le tableau de fréquence dans la liste de **Frequency of**, et on fait click sur le bouton **OK**.



Dans la fenêtre des résultats (**Output**), les tableau des fréquences et les intervalles des confiance pour chaque catégorie des variables qualitatives sont affichée.

FREQ Gen

GEN	Frequency	Percent	Cum. Percent	
F	14	23.73%	23.73%	
M	45	76.27%	100.00%	
Total	59	100.00%	100.00%	

95% Conf Limits

F 13.62% 36.59%

M 63.41% 86.38%

Les tableaux de contingence

Dans **Microsoft Excel**, sélectionnez une cellule contenant des données. Ensuite, cliquez dans la barre de menu **Data – Pivot Table – Pivot Chart Report**. Travaillez votre chemin à travers l'assistant et obtenez une nouvelle feuille contenant un tableau croisé vide et une liste de champs.

Glissez-déposez le champ représentant un facteur pronostique (le facteur de risque, le nouveau test diagnostique ou le nouveau traitement, selon le scénario donné de la recherche) à la zone étiquetée **Drop Row Fields Here**. Glissez-déposez le champ représentant un résultat (la maladie, le test diagnostique de référence ou la réponse au traitement, selon le scénario donné) à la zone étiquetée **Drop Column Fields Here**. Finalement, faire glisser et déposer l'un des champs qui servait auparavant à la zone étiquetée **Drop Data Fields Here**.

Renommer les étiquettes de rangée et de colonne de sorte qu'ils soient facilement compréhensibles par le lecteur, sans le besoin de recherche pour plus d'explications afin de comprendre le contenu du tableau (*par exemple si le sexe masculin était codée comme m, renommez l'étiquette de ligne correspondant: masculin*)

Faites un clic droit sur une étiquette de ligne et de l'option **Order** corrigez, si nécessaire, l'ordre des lignes dans votre tableau de contingence. Faites un clic droit sur une étiquette de colonne et de l'option **Order**, corrigez, si nécessaire, l'ordre des colonnes dans votre tableau de contingence.

Après avoir inséré le tableau de contingence dans votre document **Word**, n'oubliez pas de l'étiqueter à l'aide de **Caption** avec un titre correct.

Le graphique colonnes associé à un tableau de contingence

Après avoir créé un tableau de contingence, sélectionnez **Insert - Chart** dans la barre de menu.

Pour masquer les boutons de tableau faites un clic droit sur le bouton **Count of** et sélectionnez **Hide Pivot Chart Field Buttons**.

Pour afficher les étiquettes de fréquence, cliquez à droite la zone de graphique vide vers le coin supérieur gauche, sélectionnez **Chart Options** et, dans l'onglet **Data Labels**, cochez **Percentage** ou **Value**.

Puis, passez à l'onglet **Titles** et définir des titres clairs et précis pour vos axes du graphique, y compris les unités de mesure entre parenthèses, si nécessaire.

Après avoir inséré le graphique dans votre document **Word**, n'oubliez pas de l'étiqueter à l'aide de **Caption**, en utilisant un titre correct.

Données quantitatives:

Description individuelle des variables quantitatives

Moyenne, médiane, écart type, intervalle de confiance 95% pour les moyens

Dans *Microsoft Excel*, utilisez la commande **Tools – Data Analysis – Descriptive Statistics** pour calculer simultanément les paramètres les plus importants pour certaines variables descriptives quantitatives.

Dans la fenêtre **Descriptive Statistics** cochez les options **Summary Statistics** et **Confidence Level for Mean**.

Pour trouver la limite inférieure de l'intervalle de confiance de 95%, calculez **Mean** (moyenne) moins **Confidence Level (95%)**.

Pour trouver la limite supérieure de l'intervalle de confiance de 95%, calculez **Mean** plus **Confidence Level (95%)**.

Tableau de fréquence et l'histogramme

Dans *Microsoft Excel*, utilisez la commande **Tools – Data Analysis – Descriptive Statistics** pour calculer la valeur minimale, maximale et l'amplitude de la variable quantitatif désigné.

Choisissez une taille d'intervalle de fréquence (bin) commode pour la variable d'intérêt (un numéro rond représentant la taille pour chaque des 7-10 intervalles qui couvriraient l'amplitude entière de la variable).

Étiquetez une colonne vide comme "**Bin Nom de la variable (unités de mesure)**" sur la même feuille de calcul comme la variable d'intérêt.

En dessous de cette étiquette, insérez la valeur du minimum + la taille choisie antérieur pour un intervalle de fréquence (bin)

Pour compléter la colonne contenant les valeurs bin pour votre variable, utilisez **Edit – Fill – Series** (sélectionnez les options: in Columns, la valeur Step size = bin, Stop value = valeur maximale - bin).

Maintenant, utilisez **Tools – Data Analysis – Histogram**:

Pour **Input Range** sélectionnez la plage de cellules contenant la variable quantitative pour laquelle vous souhaitez tracer un tableau de fréquence et l'histogramme. Pour **Bin Range** sélectionnez la colonne nouvellement créée. Dans les deux cas, a inclure les étiquettes de colonne dans votre sélection et cochez **Labels**.

En **New Worksheet Ply** écrivez un nom évocateur pour la feuille de calcul qui va contenir le tableau de fréquence et l'histogramme pour votre variable.

Afin d'afficher l'histogramme, vous devez sélectionner **Chart Output**.

Après avoir appuyé sur le bouton OK, le tableau de fréquences et l'histogramme s'affiche sous une forme brute, inachevée.

Afin d'être compréhensible, tant le tableau de fréquence et l'histogramme nécessitent des ajustements:

1. remplacez la limite supérieure de chaque intervalle montré dans le tableau inachevé avec l'intervalle de fréquence correspondant ;
2. supprimez la légende du graphique depuis son information est redondante et ne prend que l'espace ;
3. supprimez le titre **Histogram**, puisque vous étiquettera la figure dans **Microsoft Word**, en utilisant **Caption** ;
4. redimensionnez la zone de graphique afin d'avoir une vue claire sur votre histogramme ;
5. éliminez les espaces entre les colonnes par un clic droit sur n'importe quelle colonne et en utilisant **Format Data Series - Options** ajustez **Gap Width** ;
6. vérifiez le contenu et la taille de toutes les étiquettes afin de vous assurer que votre histogramme est facilement compréhensible par toute personne qui lit votre travail.

Un tableau de fréquence correct et un histogramme correct devraient ressembler à ceci:

Tableau 2. Répartition du poids dans l'échantillon étudié

Intervalles de poids (kg)	Nombre de sujets
<=40	21
(40-50]	144
(50-60]	297
(60-70]	240
(70-80]	145
(80-90]	79
(90-100]	45
>100	29

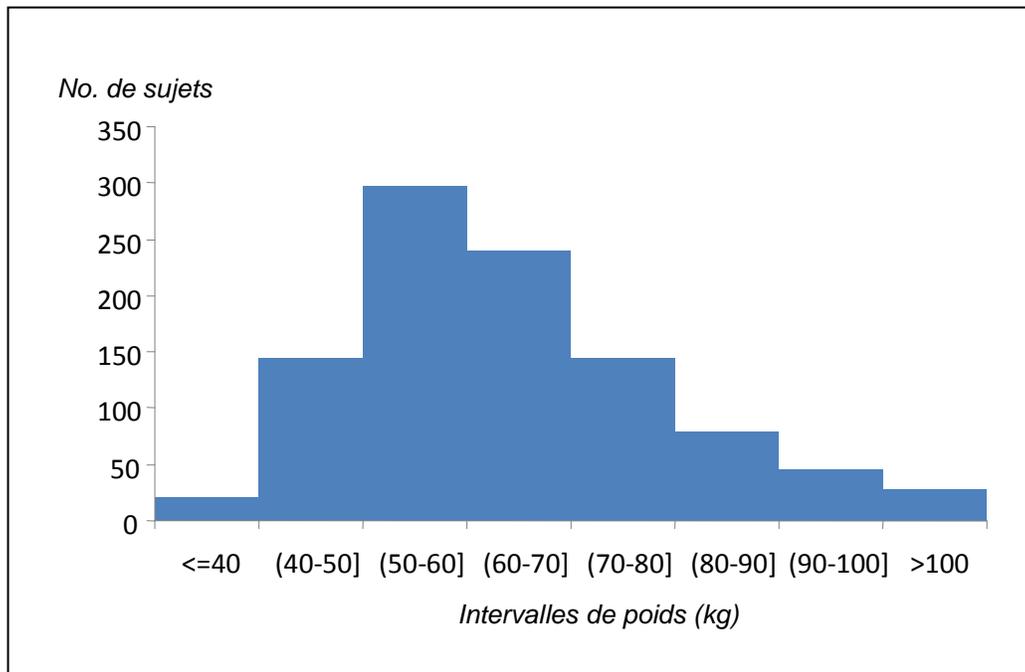


Figure 2. Histogramme des poids dans l'échantillon étudié

Description d'un lien potentiel entre deux variables quantitatives

Scatter Chart (diagramme de dispersion)

Dans **Microsoft Excel**, sélectionnez les colonnes contenant les deux variables quantitatives, y compris leurs étiquettes. Ensuite, sélectionnez **Insert – Chart** et choisissez **XY (Scatter)**.

Passer à l'étape 3 de l'assistant graphique et définir des titres corrects pour les deux axes X et Y. N'oubliez pas de préciser les unités de mesure après le titre de chaque axe, entre parenthèses.

Cliquez ensuite sur l'onglet **Legend** et décochez la case **Show legend**, puisque aucune information utile ne vient d'une légende où l'on étudie seulement deux variables à la fois.

Dans l'onglet **Titles** écrire le titre exact de chaque axe, y compris les unités de mesure entre parenthèses.

Si vous prévoyez d'utiliser le graphique dans une présentation **PowerPoint**, assurez-vous d'utiliser un titre clair et précis sur le graphique (quoi, comment et pour quels sujets on a représenté?).

Si vous prévoyez d'utiliser le graphique comme une figure dans un document **Word**, effacez le titre du graphique dans **Excel**, mais n'oubliez pas d'étiqueter le graphique dans **Microsoft Word**: sélectionnez la figure, faites un clic droit sur cette sélection et utilisez **Caption**.

Toutes les étiquettes doivent être visibles et facilement compréhensible par le lecteur, sans le besoin de recherche pour plus d'explications afin de comprendre le contenu de la figure.

Après avoir terminé les étapes guidées par l'assistant graphique, cliquez droit sur n'importe quel point du nuage de points et sélectionnez **Add Trendline**. La tendance la plus commune des nuages de données est linéaire. Dans l'onglet **Options** cochez **Display equation on chart** et **Display R-squared value on chart**.

Pour mettre en évidence la ligne de tendance (trendline) en utilisant une couleur contrastante, cliquez droit sur la ligne de tendance et utilisez **Format Trendline**.

Pour mettre en évidence les étiquettes de la ligne de tendance en utilisant une couleur contrastante, cliquez droit sur la boîte de l'étiquette et utilisez **Format Data Labels**.

Un diagramme de dispersion correcte devrait ressembler à ceci:

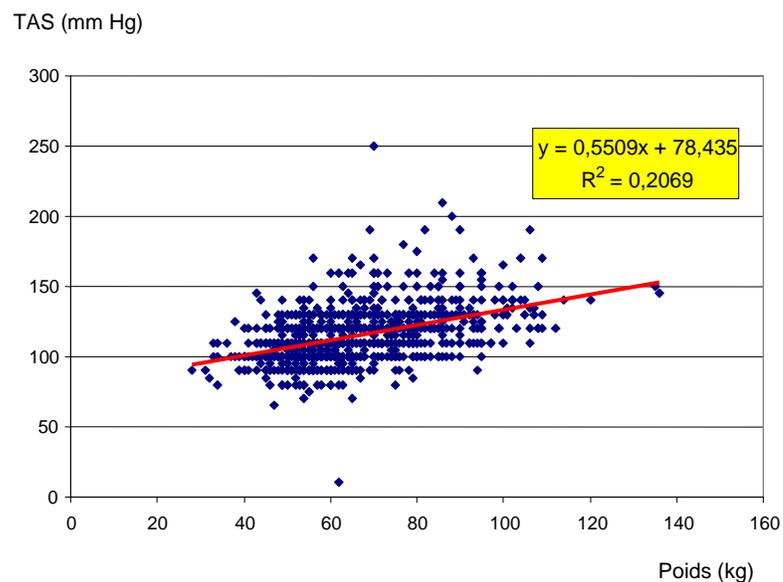


Figure 3. Relation entre le poids et la pression artérielle systolique pour les sujets inclus dans l'échantillon étudié

Données de survie:

Médiane du temps de survie

Dans **Microsoft Excel**, utilisez la commande **Tools – Data Analysis – Descriptive Statistics**. Dans la fenêtre **Descriptive Statistics** cochez **Summary Statistics**.

Graphique de la probabilité de survie

Dans le module **Analysis d'Epi Info** cliquez sur **Kaplan-Meyer Survival** dans le panneau de gauche.

Renseignez la boîte de dialogue selon le modèle ci-dessous:

Censored Variable	Value for Uncensored	
CensorVariable	"complete"	
Time Variable	Time Unit	
Time_month_	Months	
Group Variable	Graph Type	
Gender	Survival Probability	
Weight	Output to Table	
Clear	Save Only	OK
Help	Cancel	

Changez **Group Variable** comme nécessaire pour votre comparaison.

Analyse des données

Effectuer un test de Student (t-test) dans Excel

Avant d'effectuer le test, vous devez trier vos données selon les groupes que vous souhaitez comparer. (par exemple, si vous voulez comparer les valeurs de cholestérol des hommes avec les valeurs de cholestérol des femmes, vous devez trier vos données par sexe).

Pour trier vos données, cliquez sur n'importe quelle cellule à l'intérieur de votre plage de données, puis utilisez **Data – Sort**.

Si les groupes que vous souhaitez comparer sont indépendants (par exemple la comparaison des valeurs de cholestérol des femmes sur ceux des hommes), utilisez la commande **Tools – Data Analysis – t-Test: Two-Sample Assuming Unequal Variances**.

Si les groupes que vous souhaitez comparer sont appariés (par exemple la comparaison des valeurs de cholestérol des mêmes sujets, avant et après traitement), utilisez la commande **Tools – Data Analysis – t-Test: Paired Two Sample for Means**.

Dans la fenêtre de test, sélectionnez pour **Variable 1 Range** les cellules contenant les variables quantitatives correspondant au premier groupe (par exemple les valeurs de cholestérol initial pour les femmes) et pour **Variable 2 Range** les cellules contenant les variables quantitatives correspondant au deuxième groupe (par exemple les valeurs de cholestérol initial pour les hommes), sans sélectionner l'étiquette de colonne. Faites attention de ne pas sélectionner la variable de regroupement (par exemple, *sexe*) au lieu de la variable quantitative correspondante que vous souhaitez comparer.

Puisque l'hypothèse nulle (H_0) de votre raisonnement stipule l'absence de différence entre les valeurs moyennes de la variable, d'introduire 0 dans la zone **Hypothesized mean difference**.

Donnez un titre à la feuille qui contiendra les résultats du test, en entrant un nom évocateur dans la zone **New Worksheet Ply**.

Immédiatement après avoir appuyé sur OK, renommez les étiquettes génériques **Variable 1** et **Variable 2** en utilisant des étiquettes suggestives: a inclure des informations concernant les variables quantitatives que vous avez comparées, et la variable de groupement qui les différencie. Cela vous permettra d'interpréter facilement les résultats de votre test plus tard.

La valeur *two-tailed p-value* rendue par le test révèle la **signification statistique** de la différence entre les valeurs moyennes des groupes comparés.

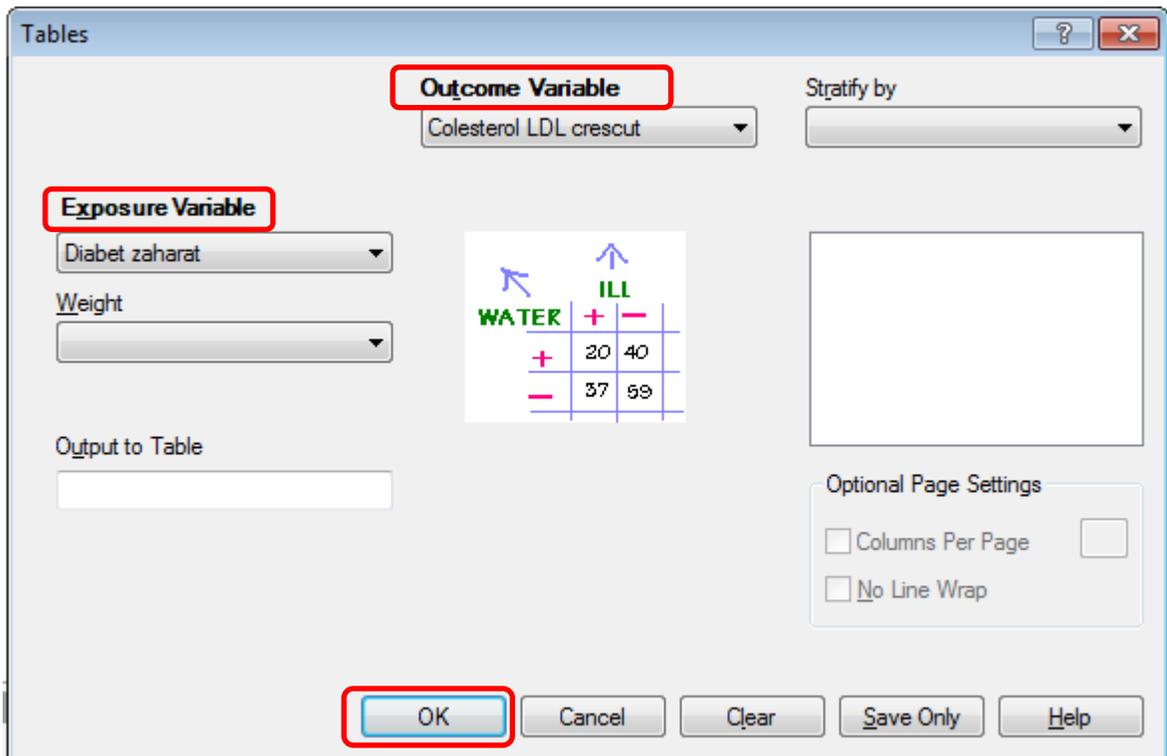
Si la valeur de p inclut la lettre E suivie d'un chiffre négatif, cela signifie en réalité une très faible (i.e. très significative) valeur de p (par ex. $p = 3,22342E-6 = 3,22342 \times 10^{-6} = 0,0000032234$).

Les résultats du test comprennent également les valeurs moyennes des variables comparées. En les soustrayant, vous serez en mesure d'évaluer la différence entre les valeurs moyennes, en évaluant ainsi la **signification clinique** de cette différence.

Le calcul du tableau de contingence, Risque relatif (RR) et odds ratio (OR) dans Epi Info Exécution d'un test Chi-carré (χ^2) dans Epi Info

Dans la zone des commandes (**Command Explorer**) de la fenêtre d'analyse de Epi Info, la section **Statistics** on choisit la commande **Tables**.

Dans la fenêtre ouverte choisissez la variable de regroupement dans la liste **Exposure Variable** (par exemple, le facteur de risque, traitement, etc.), et la variable résultat de la liste **Outcome Variable** (la variable résultat attendu (par exemple la maladie suspectée d'être un résultat du facteur de risque, l'amélioration de la santé suspecté d'être un résultat du traitement, etc.), puis appuyez sur le bouton **OK**.



Dans la fenêtre avec les résultats (Output), on a:

1. **Le tableau de contingence:**

	Colesterol LDL crescut		
DIABET ZAHARAT	da	nu	Total
da	40	42	82
Row%	48.78%	51.22%	100.00%
Col%	85.11%	56.00%	67.21%
nu	7	33	40
Row%	17.50%	82.50%	100.00%
Col%	14.89%	44.00%	32.79%
TOTAL	47	75	122
Row%	38.52%	61.48%	100.00%
Col%	100.00%	100.00%	100.00%

2. **Des indicateurs qui quantifient la relation entre les deux variables (ex. Risk Difference – montre la différence de la réalisation du résultat entre les deux groupes de la variable de regroupement):**

Selon le type de collecte de données utilisées dans votre scénario de recherche, interprétez seulement l'indicateur approprié (**RR**, **RD** ou **OR**) et son **Intervalle de confiance 95%** affiché à droite de **l'estimation ponctuelle** de chaque indicateur.

Dans la plupart des cas, la valeur **two-tailed p-value** rendue par le test **Chi-square - uncorrected** montre la **signification statistique** de la différence entre les distribution de fréquence dans les groupes comparés. Pourtant, parfois, quand une ou plusieurs fréquences attendues sont inférieures à 5, un message s'affichera ci-dessous les résultats du test, vous invitant à interpréter la valeur p-value rendue par le test **exact de Fisher**.

Single Table Analysis

	Point Estimate	95% Confidence Interval		
		Lower	Upper	
PARAMETERS: Odds-based				
Odds Ratio (cross product)	4.4898	1.7831	11.3049	(T)
Odds Ratio (MLE)	4.4367	1.8082	11.9542	(M)
		1.6804	13.2787	(F)
PARAMETERS: Risk-based				
Risk Ratio (RR)	2.7875	1.3725	5.6611	(T)
Risk Difference (RD%)	31.2805	15.2896	47.2714	(T)

3. **Des tests statistiques (on utilise le p bidirectionnel – 2-tailed p):**
- Le test chi carre non corrigée (**Chi-square – uncorrected**)

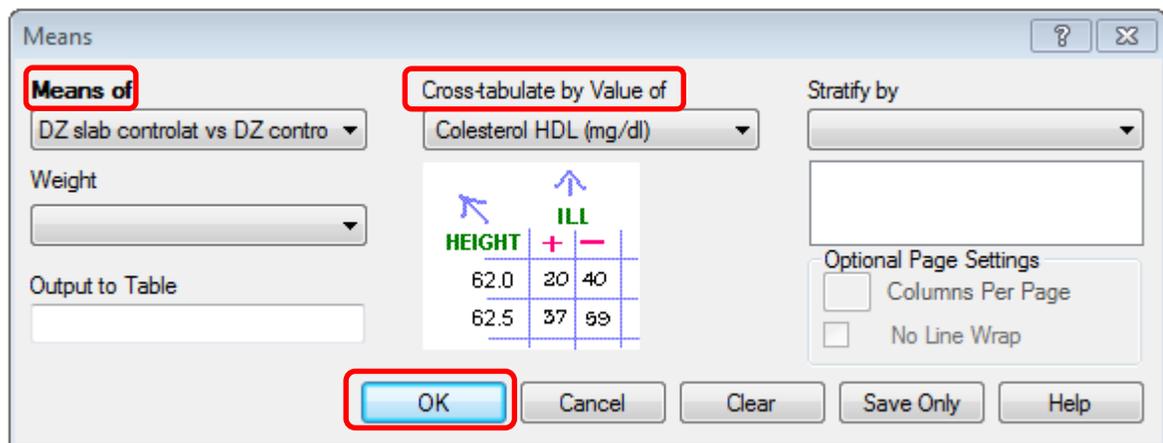
b. Le test chi carré avec la correction Yates (**Chi-square – corrected (Yates)**)

c. Le test exact Fisher (**Fisher Exact**)

STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi-square - uncorrected	11.1076		0.0008608989
Chi-square - Mantel-Haenszel	11.0166		0.0009041693
Chi-square - corrected (Yates)	9.8261		0.0017216883
Mid-p exact		0.0003773484	
Fisher exact		0.0006285038	0.0007946499

Comparaisons données quantitatives (Test student/ANOVA/...) dans Epi Info

Dans la zone des commandes (**Command Explorer**) de la fenêtre d'analyse de Epi Info, la section **Statistics** on choisit la commande **Means**.



Dans la fenêtre ouverte choisissez la variable quantitative de la liste **Means of**, et la variable de regroupement dans la liste **Cross-tabulate by Value of**, puis appuyez sur le bouton **OK**.

Dans la fenêtre avec les résultats (**Output**), on a:

1. **La statistique descriptive par groups:**

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev	
DZ controlat	45.0000	2502.0000	55.6000	112.4727	10.6053	
DZ slab controlat	37.0000	1932.0000	52.2162	77.8408	8.8227	
	Minimum	25%	Median	75%	Maximum	Mode
DZ controlat	39.0000	48.0000	55.0000	62.0000	80.0000	54.0000
DZ slab controlat	36.0000	46.0000	52.0000	58.5000	76.0000	41.0000

2. Le résultat du test t (Student) pour comparer les moyennes des deux échantillons indépendants avec variances égales (Pooled), ou inégales (Unequal):

T-Test

	Method	Mean	95% CL Mean	Std Dev
Diff (Group 1 - Group 2)	Pooled	3.3838	-0.9633 7.7309	9.8432
Diff (Group 1 - Group 2)	Satterthwaite	3.3838	-0.8867 7.6543	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	80	1.55	0.1253
Satterthwaite	Unequal	79.98	1.58	0.1188

3. Le résultat du test ANOVA pour comparer les moyennes des 3 ou plus d'échantillons indépendants avec variances égales:

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	232.49071	1	232.49071	2.39957
Within	7751.07027	80	96.88838	
Total	7983.56098	81		

P-value = 0.12532

4. Le résultat du test Bartlett pour comparer les variances des deux échantillons indépendants:

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 1.30103 df=1 P value=0.25403

A small p-value (e.g., less than 0.05 suggests that the variances are not homogeneous and that the ANOVA may not be appropriate.

5. Les résultats des tests non paramétriques pour comparer deux échantillons indépendants (test Mann-Whitney/Wilcoxon Two-Sample) ou plusieurs échantillons indépendants (test Kruskal-Wallis):

Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

Kruskal-Wallis H (equivalent to Chi square) = 1.4703

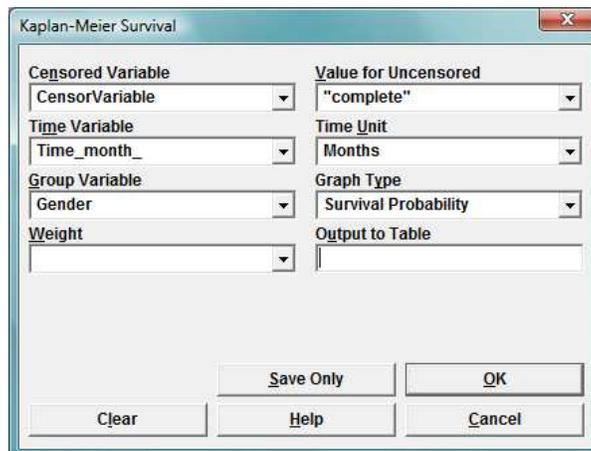
Degrees of freedom = 1

P value = 0.2253

Effectuer un test Log-rank pour l'analyse de la survie dans Epi Info

Dans le module **Analysis** du logiciel **Epi Info** cliquez sur **Kaplan-Meier Survival**, à partir du panneau de gauche.

Renseignez la boîte de dialogue comme le montre l'image ci-dessous:



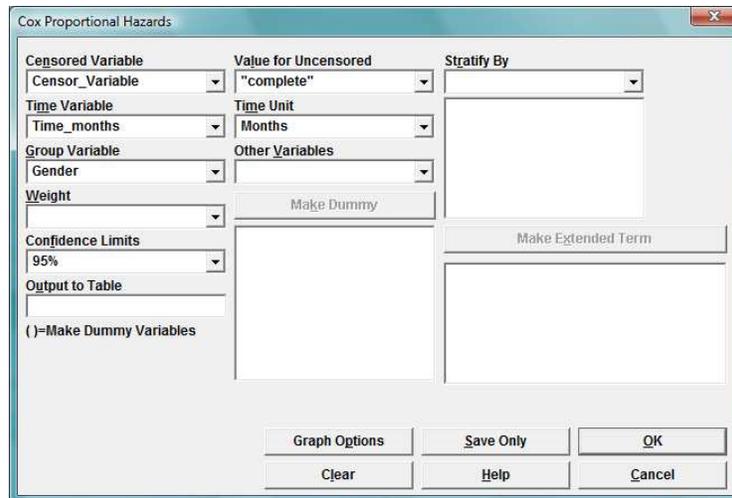
Changez **Group Variable** comme nécessaire pour votre comparaison.

Ci-dessous le diagramme de survie, vous trouverez le résultat du test Log-rank.

Exécution d'une régression de Cox et le calcul du Hazard Ratio (HR) dans Epi Info

Dans le module **Analysis** du logiciel **Epi Info** cliquez sur **Cox proportional hazard**, à partir du panneau de gauche.

Renseignez la boîte de dialogue comme le montre l'image ci-dessous:



Changez **Group Variable** comme nécessaire pour votre comparaison.

Le Hazard Ratio (HR), son intervalle de confiance 95%, et la signification statistique du modèle de régression de Cox seront énumérés ci-dessous le diagramme du modèle de régression.

Exécution d'une régression linéaire simple dans Microsoft Excel

Utilisez **Data Analysis – Regression** dans le menu **Tools** de **Microsoft Excel**.

Pour **Input Y Range**, sélectionnez la plage de cellules contenant la variable dépendante (**y**) dans votre échantillon, celui que vous voulez prédire à l'aide d'une régression linéaire simple.

Pour **Input X Range** sélectionnez la plage de cellules contenant la variable indépendante (**x**) de votre échantillon, celui que vous souhaitez utiliser afin de prédire la variable dépendante (**y**) en utilisant une régression linéaire simple.

Veillez à inclure dans la sélection les cellules contenant les étiquettes des deux variables et cochez la case **Labels**. Cochez également la case **Confidence Level** pour le CI 95% et entrez un nom évocateur pour la nouvelle feuille de calcul où votre régression linéaire sera sauvée.

Exécution d'une régression linéaire multiple dans Microsoft Excel

Utilisez **Data Analysis – Regression** dans le menu **Tools** de **Microsoft Excel**.

Pour **Input Y Range**, sélectionnez la plage de cellules contenant la variable dépendante (**y**) dans votre échantillon, celui que vous voulez prédire à l'aide d'une régression linéaire multiple.

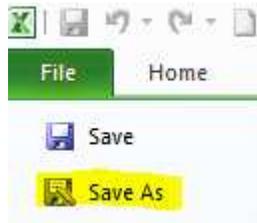
Pour **Input X Range** sélectionnez la plage de cellules contenant les variables indépendantes (x_1, x_2, \dots, x_n) de votre échantillon, ceux que vous souhaitez utiliser afin de prédire la variable dépendante (**y**) en utilisant une régression linéaire multiple. Si les variables indépendantes ne constituent pas une plage de cellules contiguës, coupez les variables isolées avant d'utiliser la commande **Regression**, et insérez-les dans des colonnes adjacentes pour former une plage de cellules contiguës.

Veillez à inclure dans la sélection les cellules contenant les étiquettes pour toutes les variables dépendantes et indépendantes et cochez la case **Labels**. Cochez également la case **Confidence Level** pour le CI 95% et entrez un nom évocateur pour la nouvelle feuille de calcul où votre régression linéaire sera sauvée.

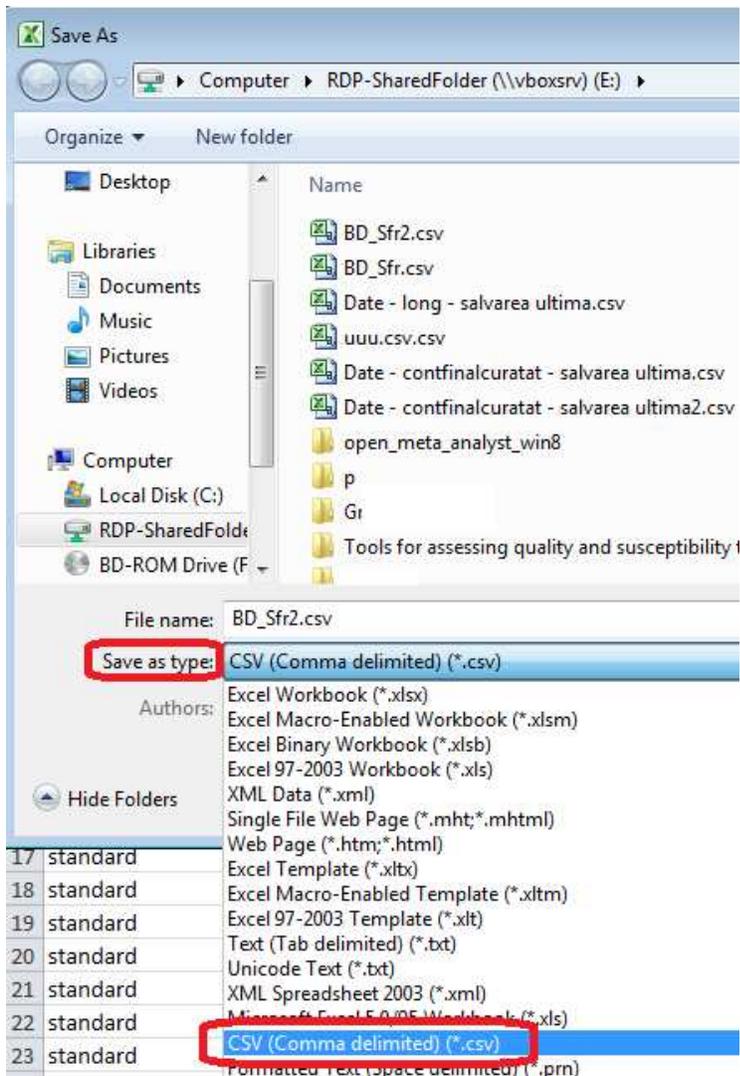
Réaliser un graphique [Kaplan-Meyer](#) et un [test Log-rank](#) pour l'analyse de la survie [sur Internet](#)

Le fichier doit être dans le format **.CSV** (comma separated values).

Si le fichier n'est pas dans ce format on peut le transformer en faisant clic dans le tab **File, Save As**

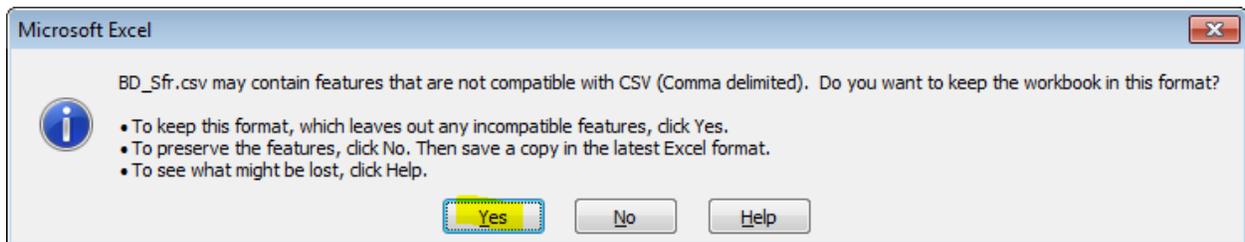


On va choisir dans: **Save as type: CSV (Comma delimited) (*.csv)**



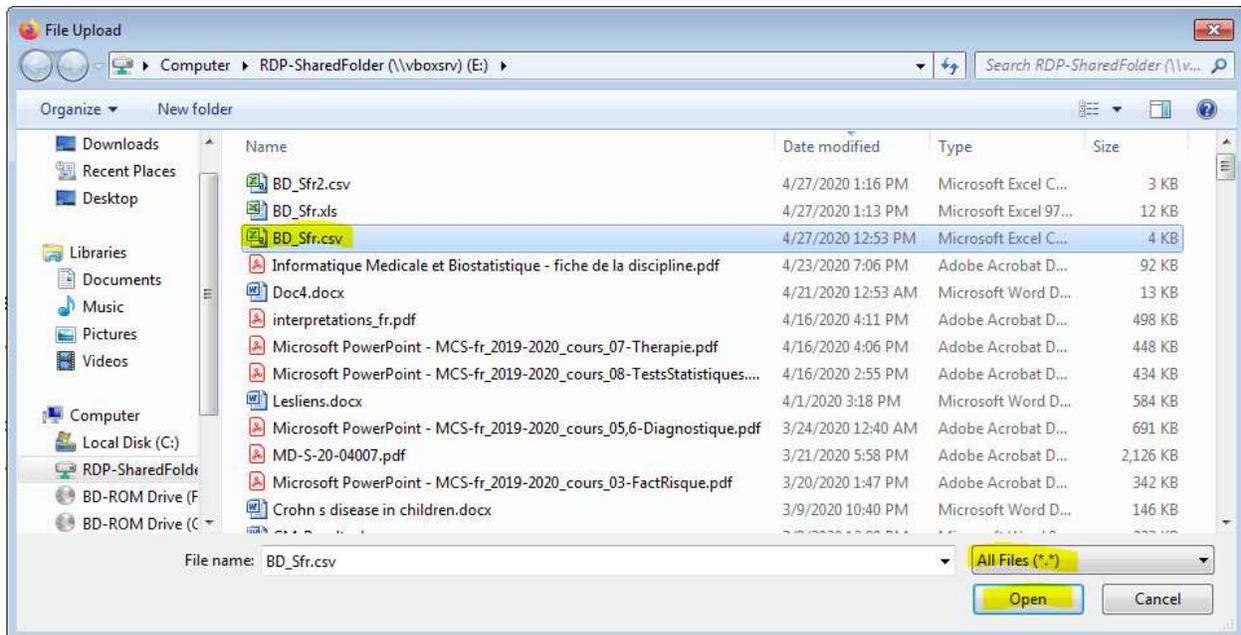
Puis on va faire clic sur le bouton **Save**.

Dans la fenêtre qui va s'ouvrir on va faire click sur le bouton **Yes**



Si le fichier est dans ce format, on va ouvrir le site : <https://merser.shinyapps.io/Survival/>

Sur la page on va faire clic sur le bouton **Browse**



Dans la fenêtre qui s'ouvre on va choisir le fichier. Si le fichier n'est pas visible, changez la sélection dans la partie inferieure droite, pour que la fenêtre montre toutes les types des fichiers (**All Files (« . »**)). Puis on va faire clic sur le bouton **Open**.

Dans la fenêtre du navigateur les données doit être importées, et il va montrer une image similaire a la suivante :

Survival Analysis

Write or copy/paste Your data into the area below.
The first line MUST include column names!

```
Traitement, Type, Temps, Censure
"standard", "a, petite, cellules", 30, 1
"standard", "a, petite, cellules", 384, 1
```

Parfois au lieu des noms que vous avez eu dans le fichier il va montrer: X.Traitement., X.Type., X.Temps., X.Censure. On peut faire l'analyse sans probleme.

Pour faire l'analyse on doit preciser quelle sont les variables. Dans le champ **Edit model (case sensitive)** on doit **remplacer time** avec la variable qui indique le temps de survie (**Temps**), puis on doit rempplacer event, avec la variable qui indique l'etet de censure (**Censure**), et on doit remplacer 1 avec la variable de regroupement, en fonction duquelle on va comparer la survie des patients (**Type** ou **Traitement**): (**Temps, Censure**) ~ **Type**

Edit model (case sensitive)

(time, event) ~ 1

Edit model (case sensitive)

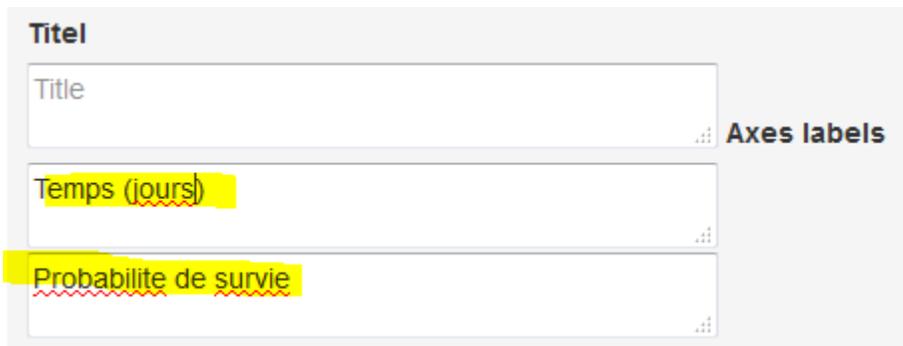
(Temps, Censure) ~ Type

Ou parfois on doit ecrire (X.Temps., X.Censure.) ~ X.Type. ou des autres noms en fonction des noms des variables das la fenetre.

Pour calculer en fonction du traitement la formule sera: **(Temps, Censure) ~ Traitement**

Puis on va **modifier les options** :

On va **changer les titres des axes** du graphique.



Titel

Title

Temps (jours)

Probabilite de survie

Axes labels

On peut choisir de ne pa montrer les intervalles de confiance, et de montrer la legende dans la partie inferieure droite

Confidence Interval

- None
- Plain
- Log
- Log-Log

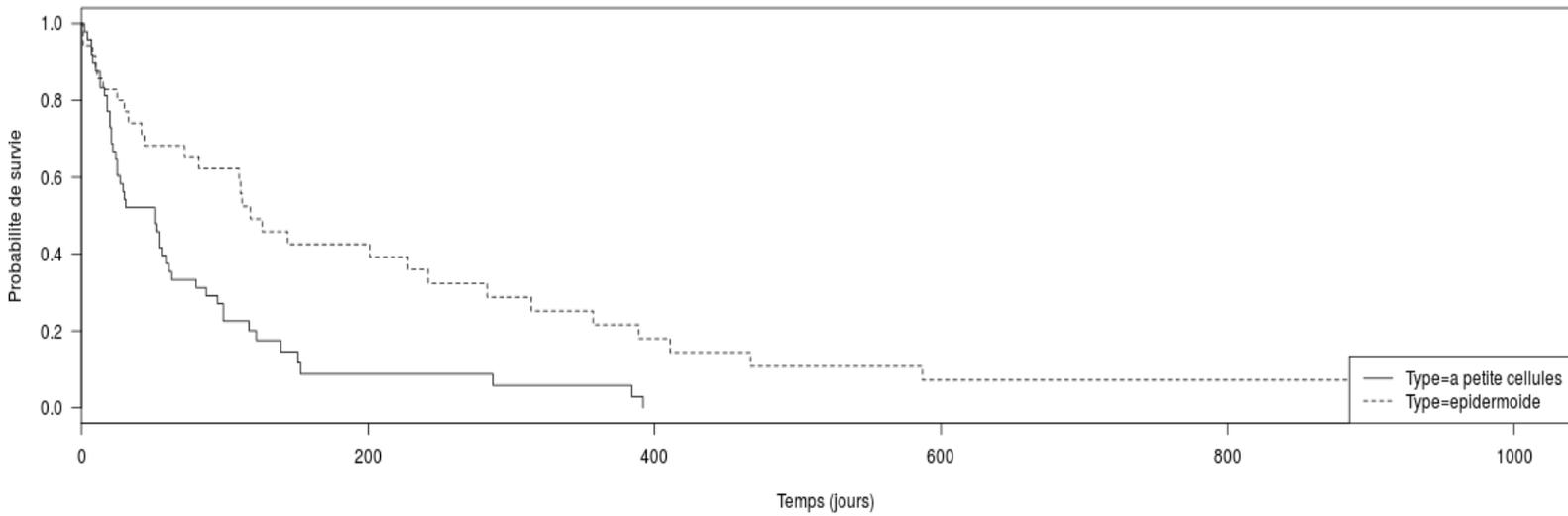
Legend Position

- None
- Bottom Right
- Bottom Left
- Bottom
- Top Right
- Top Left
- Top
- Right
- Left
- Center

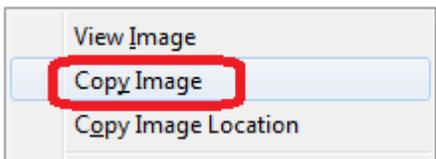
A la fin, on va fair **clie** sur le bouton **Run Survival Analysis**

Run Survival Analysis

On va trouver le graphique: (**Pour la survie en fonction du Type**)

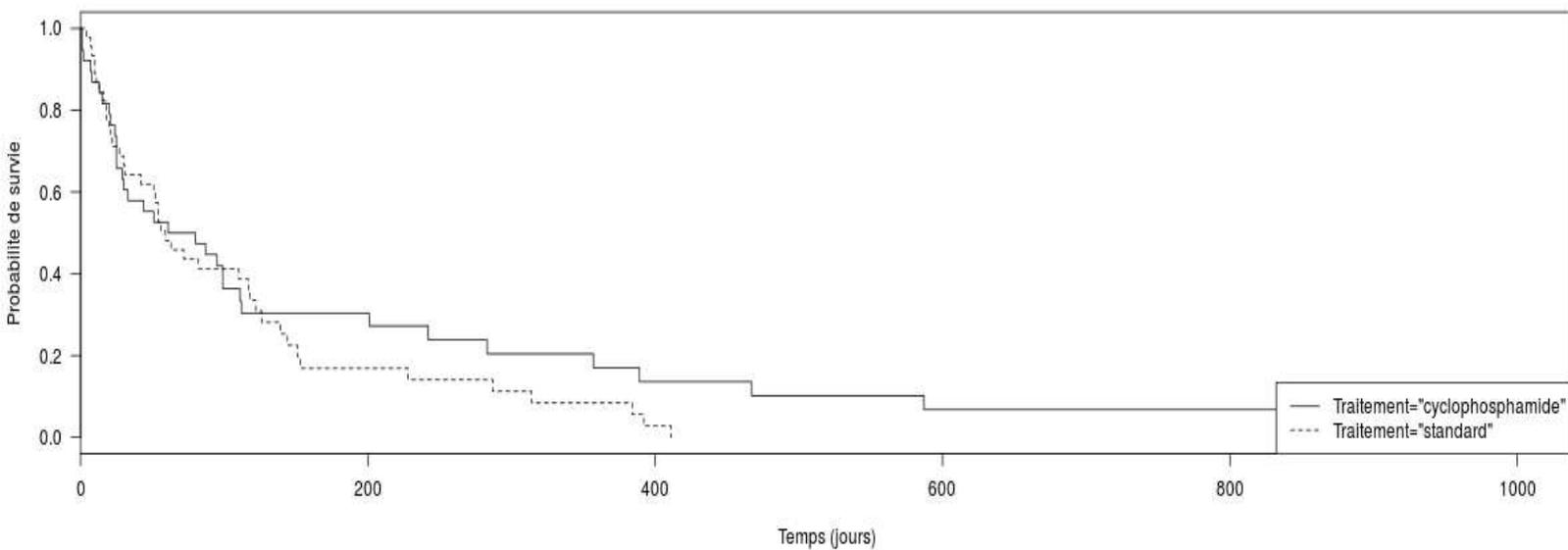


On peut **copier** le graphique avec un **clik droit** du souris sur l'image, en utilisant l'option **Copy Image**



Pour la survie en fonction du **Traitement** le graphique est le suivant :

KAPLAN-MEIER PLOT



Pour trouver la survie a 1 an, d'un group, on va regarder dans le tableau du group (ici Type= a petite cellules), pour chercher dans la colonne du temps (time) la valeur la plus proche du 365 (1 an). La probabilité de survie est dans la colonne survival : 0.058, donc 5.8%

SURVIVAL TABLE

```
Call: survfit(formula = Surv(Temps, Censure) ~ Type, conf.type = "none")
```

Type=a petite cellules

```
time n.risk n.event survival std.err
```

287	3	1	0.0585	0.0387
384	2	1	0.0293	0.0283
392	1	1	0.0000	NaN

Type=epidermoide

```
time n.risk n.event survival std.err
```

357	7	1	0.216	0.0757
389	6	1	0.180	0.0711

D'une maniere similaire on observe 0.216 pour le group epidermoide.

Pour trouver le resultat du **test Log-rank on** regarde apres le derniere tableau, la valeur du p (ici 0.0006)

LogRank Test

```
Call:
```

```
survdif(formula = Surv(Temps, Censure) ~ Type)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Type=a petite cellules	48	45	31.3	5.97	11.6
Type=epidermoide	35	31	44.7	4.19	11.6

```
Chisq= 11.6 on 1 degrees of freedom, p= 0.000669
```

Pour le test log-rank en fonction du traitement le resultat est le suivant:

LogRank Test

```
Call:
survdif(formula = Surv(Temps, Censure) ~ Traitement)

          N Observed Expected (O-E)^2/E (O-E)^2/V
Traitement="cyclophosphamide" 38      35      39.1      0.439      0.985
Traitement="standard"         45      41      36.9      0.466      0.985

Chisq= 1 on 1 degrees of freedom, p= 0.321
```

Calculer le Hazard ratio (HR) pour l'analyse de la survie sur Internet

On va importer le fichier .CSV (voir en haut), mais sur le lien :

<https://mercer.shinyapps.io/CoxReg/>

Dans le champ **Model**, on va remplacer les variables comme pour le test Log-rank (voir en haut)

Model

(Temps, Censure) ~ Type

Puis on va faire clic sur le bouton **Run Cox Regression Analysis**

Run Cox Regression Analysis

Dans le resultat, on va trouver **le hazard ratio (HR)**. Ici, le HR (le rapport du hazard **du groupe epidermoide**, a le hazard du groupe **a petite cellules**)=0.43 (l'estimateur ponctuel – exp(coef)). L'intervalle de confiance 95% est: 0.26 – 0.70 (**lower .95, upper .95**)

Cox Proportional Hazard Regression

REGRESSION RAPPORT

```
Call:
coxph(formula = Surv(Temps, Censure) ~ Type, ties = "efron")

n= 83, number of events= 76

              coef exp(coef) se(coef)      z Pr(>|z|)
Type"epidermoide" -0.8531    0.4261  0.2565 -3.326 0.00088 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Type"epidermoide"    0.4261      2.347    0.2578    0.7044

Concordance= 0.593 (se = 0.033 )
Rsquare= 0.13 (max possible= 0.998 )
Likelihood ratio test= 11.59 on 1 df,  p=7e-04
Wald test               = 11.06 on 1 df,  p=9e-04
Score (logrank) test = 11.61 on 1 df,  p=7e-04
```