

Instructions pour résoudre les travaux pratiques

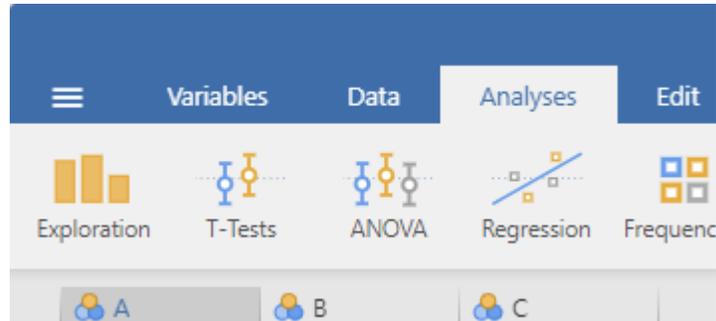
Instructions d'utilisation Jamovi pour l'analyse des données	3
Importation de bases de données	3
Vérifier le type de données.....	3
Modification de l'ordre des catégories pour une variable qualitative	4
Description d'une variable qualitative – tableau de fréquences, graphique à colonnes	4
Analyse entre 2 variables qualitatives – test Khi deux, Fisher, tableau de contingence, OR, RR, RA (RAR), graphique colonnes.....	6
Tableau de contingence.....	9
Résultats des tests Khi deux et de Fisher exacte	9
Indicateurs médicaux de la force de l'association	10
Graphique colonne	10
Test Student pour deux groupes indépendants	11
Évaluation des conditions d'application du test Student – normalité.....	11
Réalisation du test de Student pour des échantillons indépendants	17
Choisir entre le test de Student à variance égales et le test de Welch – Variance inégales	19
Statistiques descriptives	20
Numérique	20
Graphique des moyennes	20
Tests diagnostiques.....	21
Analyse ROC.....	26
Tableaux de valeurs limites et statistiques associées.....	28
Graphique de la courbe ROC.....	29
Comparaison des courbes ROC par des tests statistiques	30
Analyse de survie (graphique Kaplan Meyer, test log-rank, régression Cox – HR)	31
La médiane du temps de survie	33

La probabilité de survie a différents moments.....	34
La courbe de survie Kaplan-Meier, avec le test log-rank.....	35
Méta-analyse	36
Installation d'un module d'analyse supplémentaire (par exemple pour l'analyse ROC).....	40
Instructions d'utilisation de EpiInfo et Excel pour l'analyse des données.....	41
Travailler avec Epi Info.....	41
Téléchargement du logiciel <i>Epi Info</i> (pour utilisation à domicile)	41
Ouverture du programme Epi Info	41
Ouverture de l'outil de l'analyse statistique.....	41
L'Importation d'un fichier Excel dans Epi Info	43
Sélectionnez le dossier ou les résultats de l'analyse seront enregistrée.....	46
Montrer les données.....	47
Sélection d'un sous ensemble des données (analyser des sous-groupes)	48
Activation du module <i>Data Analysis</i> dans <i>Microsoft Excel</i>	49
Statistiques descriptives	49
Données qualitatives (catégoriques):	49
Données quantitatives:.....	53
Description individuelle des variables quantitatives	53
Description d'un lien potentiel entre deux variables quantitatives	55
Données de survie:	57
Analyse des données.....	58

Instructions d'utilisation Jamovi pour l'analyse des données

Importation de bases de données

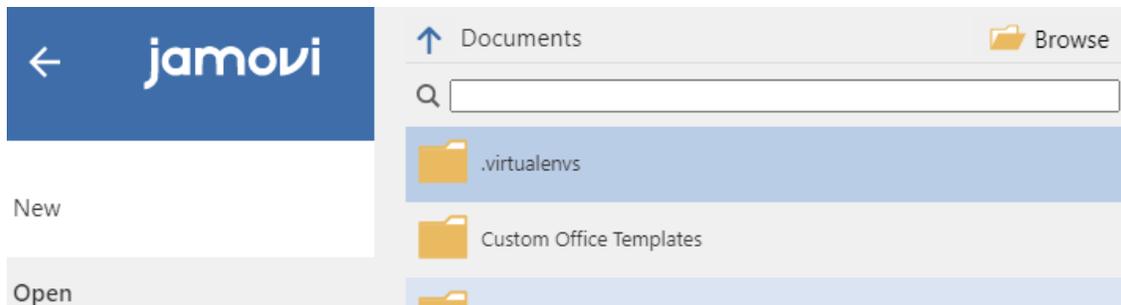
Appuyez sur l'icône du menu **hamburger**. ☰



Sélectionnez ensuite l'option **Ouvrir (Open)**



Recherchez ensuite le dossier contenant la base de données en appuyant sur la touche **Parcourir (Browse)**, sélectionnez le fichier souhaité et appuyez sur la touche **Ouvrir (Open)** pour l'importer.



Vérifier le type de données

Avant de commencer l'analyse, il est nécessaire de vérifier si le programme a correctement classé les variables dans la base de données. La variable **Activité** est une variable qualitative et les 4 autres variables sont quantitatives. Pour vérifier le type d'une variable, nous donnons **Double-cliquer** sur **Titre de la colonne** (nom de la variable). Dans le cas de la variable **Activité**, on s'assure que dans la zone de sélection **Type de mesure (Measure type)** l'option est sélectionnée **Nominal**. Pour les autres variables quantitatives, nous veillons à ce que l'option soit sélectionnée dans la zone de sélection **Type de mesure Continu (Continuous)**.

DATA VARIABLE

Activitate

Description

Measure type **Nominal** 

Data type **Text**

Missing values

Levels
da
nu

Retain unused levels in analyses

DATA VARIABLE

CHIT1

Description

Measure type **Continuous** 

Modification de l'ordre des categories pour une variable qualitative

Dans le champ **Levels**, cliquez sur une categorie, et puis cliquez sur une **fleche** pour modifier l'ordre.

DATA VARIABLE

Aspirine

Description

Measure type **Nominal** 

Data type **Text**

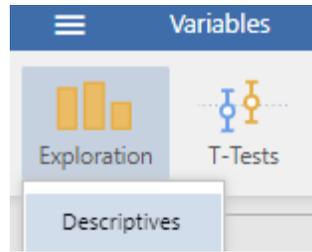
Missing values

Levels
Non
 Oui

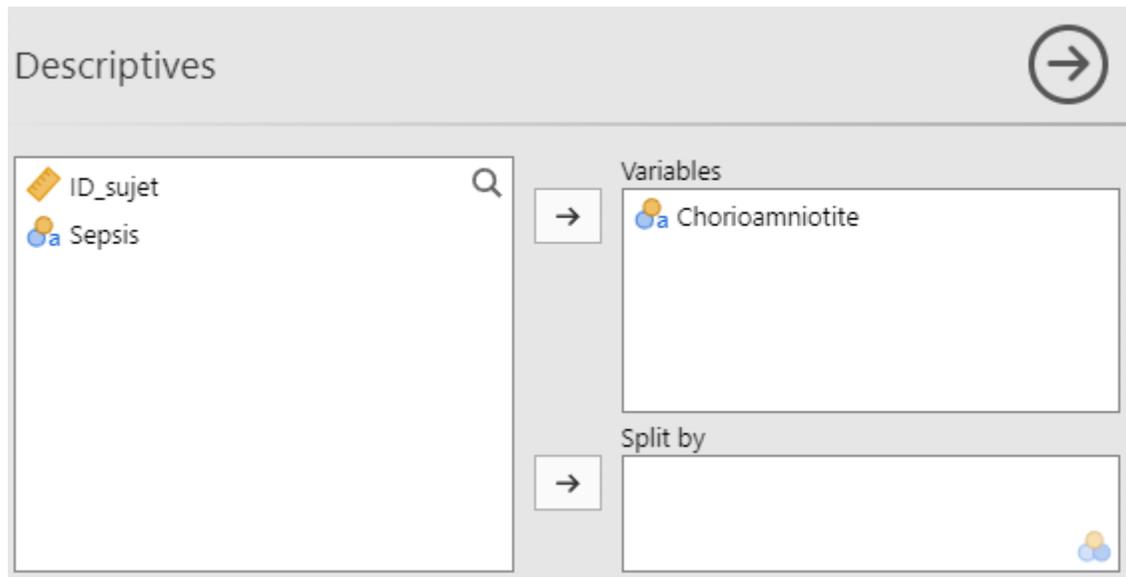
Retain unused levels in analyses

Description d'une variable qualitative – tableau de fréquences, graphique à colonnes

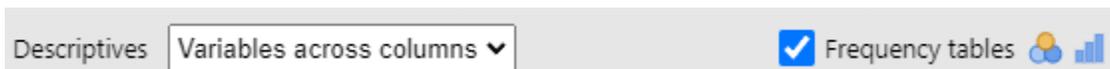
Dans l'onglet **Analyses**, le bouton **Exploration**, choisissez l'option **Descriptives**.



La **variable d'intérêt** (p. ex. **maladie**, ou **facteur de risque**), se déplace en appuyant sur la flèche à côté du champ **Variables**.



Pour afficher le tableau des fréquences, cochez l'option Tableaux des fréquences (**Frequency tables**).

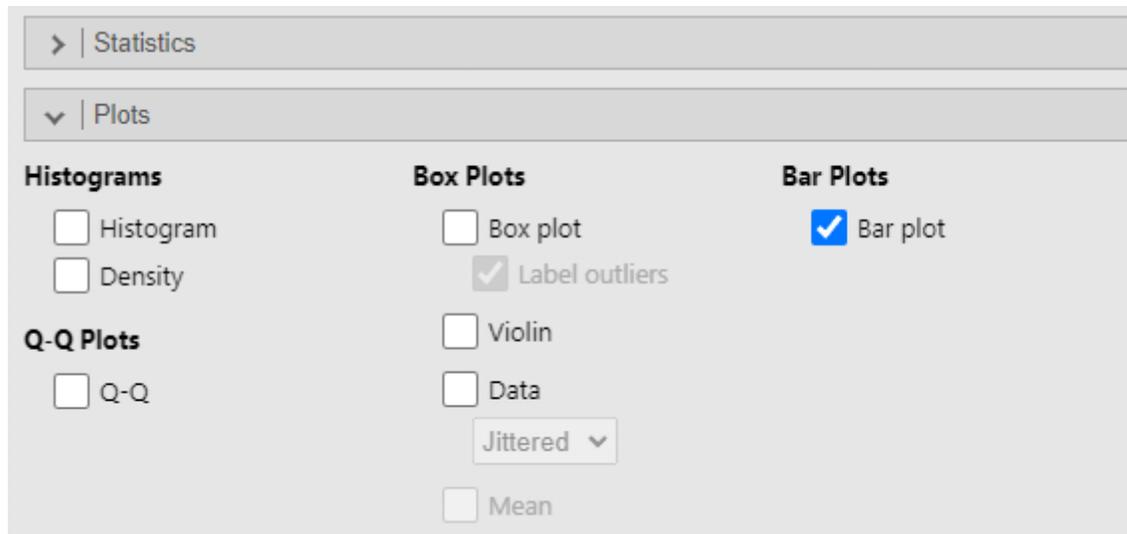


On obtient un tableau de fréquences similaire à celui ci-dessous :

Frequencies

Frequencies of Chorioamniotite			
Chorioamniotite	Counts	% of Total	Cumulative %
Avec	57	26.4 %	26.4 %
Sans	159	73.6 %	100.0 %

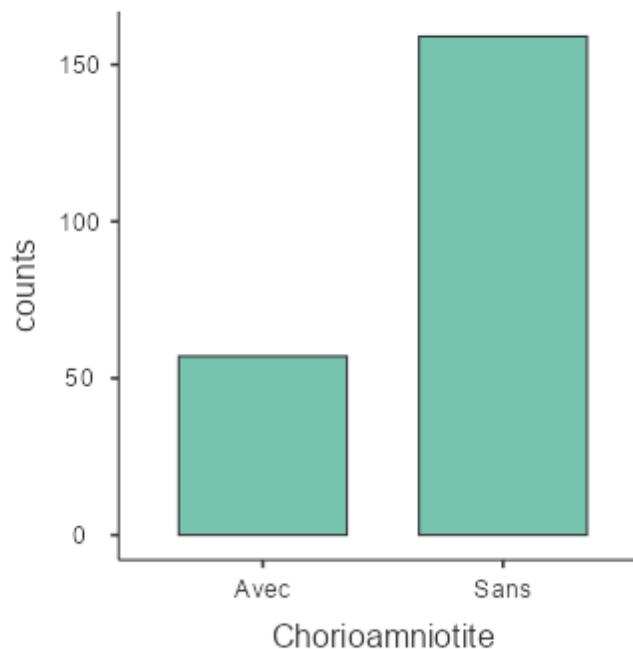
Pour afficher un graphique colonnes, cochez l'option **Graphique à barres (Bar plot)**.



Un graphique colonnes est obtenu :

Plots

Chorioamniotite



Analyse entre 2 variables qualitatives – test Khi deux, Fisher, tableau de contingence, OR, RR, RA (RAR), graphique colonnes

Dans l'onglet **Analyses**, le module **Fréquences (Frequencies)** on choisit l'option **Echantillons indépendants (X^2 test d'association) (Independent Samples (X^2 test of association))**.

The screenshot shows a software interface with a menu bar (Variables, Data, Analyses, Edit) and a toolbar with icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, and Surv. Below the toolbar is a data table with columns 'NrCrt.', 'Sex', and 'Aspirina'. A dropdown menu is open, listing various statistical tests under categories like 'One Sample Proportion Tests', 'Contingency Tables', and 'Log-Linear Regression'.

	NrCrt.	Sex	Aspirina
1	1	m	nu
2	2	m	nu
3	3	m	da
4	4	f	nu
5	5	m	nu
6	6	f	da
7	7	m	da
8	8	m	nu
9	9	f	nu
10	10	f	nu
11	11	m	nu
12	12	m	nu

Statistical tests available in the dropdown menu:

- One Sample Proportion Tests
 - 2 Outcomes: Binomial test
 - N Outcomes: χ^2 Goodness of fit
- Contingency Tables
 - Independent Samples: χ^2 test of association
 - Paired Samples: McNemar test
- Log-Linear Regression

Variable de regroupement (p. ex. **facteur de risque**, ou **traitement**), se déplace en appuyant sur la flèche à côté du champ **Lignes (Rows)**.

La variable qui vous intéresse (p. ex. **maladie**, ou **complication**, ou **résultat du traitement**), se déplace en appuyant sur la flèche en regard du champ **Colonnes (Columns)**.

Comme options, nous recommandons les choix suivants : Tests : X^2 , **test exact de Fisher** ; Mesures comparatives (2x2 uniquement) : **Odds ratio**, **Risque relatif**, Risque attribuable/**Différence de proportion**, **Intervalles de confiance** ; Dénombrements : **Dénombrements observés**, **Dénombrements attendus/théoriques** ; Pourcentages : sur les lignes (**ligne**) ; Tracés : colonne (**Bar Plot**), sur l'axe Y (**Y-Axis**) : pourcentages, sur les lignes (**à l'intérieur des lignes**) ; **Type de barre**, empilée ; sur l'axe des abscisses (**axe des abscisses**) : lignes (**lignes**).

Contingency Tables



- NrCr.
- Sexe
- D1Ulceration (mm)
- D2Ulceration (mm)
- D1-D2 (mm)

Rows:

Columns:

Counts (optional):

Layers:

Statistics

Tests

- χ^2
- χ^2 continuity correction
- Likelihood ratio
- Fisher's exact test
- z test for difference in 2 proportions

Hypothesis

- Group 1 \neq Group 2
- Group 1 $>$ Group 2
- Group 1 $<$ Group 2

Comparative Measures (2x2 only)

- Odds ratio
- Log odds ratio
- Relative risk
- Difference in proportions
- Confidence intervals

Interval %

Compare

▼ | Cells

Counts

Observed counts

Expected counts

Percentages

Row

Column

Total

▼ | Plots

Plots

Bar Plot

Y-Axis

Counts

Percentages within rows ▼

Bar Type

Side by side

Stacked

X-Axis

Rows

Columns

Tableau de contingence

		Contingency Tables		
		Guerison		Total
Aspirine		Oui	Non	
Oui	Observed	58	142	200
	Expected	38.0	162	200
	% within row	29.0 %	71.0 %	100.0 %
Non	Observed	18	182	200
	Expected	38.0	162	200
	% within row	9.0 %	91.0 %	100.0 %
Total	Observed	76	324	400
	Expected	76.0	324	400
	% within row	19.0 %	81.0 %	100.0 %

Résultats des tests Khi deux et de Fisher exacte

La valeur de p, correspondant aux tests, se trouve dans la colonne p.

χ^2 Tests			
	Value	df	p
χ^2	26.0	1	< .001
Fisher's exact test			< .001
N	400		

Indicateurs médicaux de la force de l'association

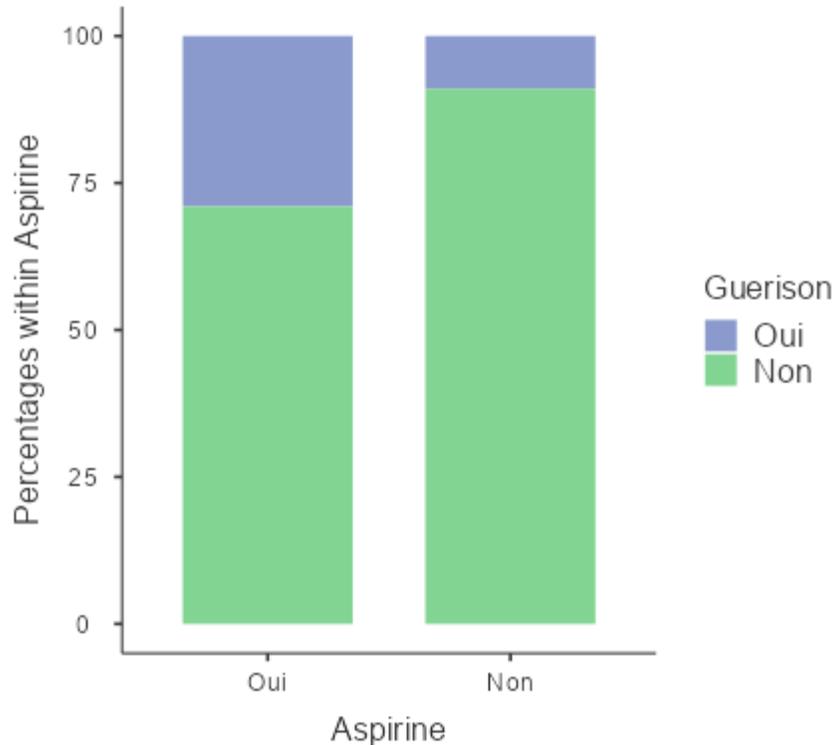
Les résultats des indicateurs médicaux sont présentés dans le tableau Mesures comparatives (Comparative Measures), où l'estimateur ponctuel se trouve dans la colonne **Valeur (Value)**, et les extrémités de l'intervalle de confiance à 95 % se trouvent dans les **colonnes Inférieur (Lower), Supérieur (Upper) (intervalles de confiance à 95 % (95% Confidence Intervals))**.

Comparative Measures			
	Value	95% Confidence Intervals	
		Lower	Upper
Difference in 2 proportions	0.200*	0.126	0.274
Odds ratio	4.13	2.33	7.32
Relative risk	1.28*	1.16	1.41

* Rows compared

Graphique colonne

Enfin, nous avons la représentation graphique de l'association entre les deux variables qualitatives.



Test Student pour deux groupes indépendants

Évaluation des conditions d'application du test Student – normalité

Le test Student ne peut pas être appliqué que si les données des deux groupes comparés suivent une distribution normale. Différentes méthodes peuvent être utilisées pour évaluer la normalité : indicateurs de distribution des données : coefficient d'asymétrie, coefficient de aplatisement ; tests de normalité ; graphiques (histogramme, graphique des quantiles).

Les indicateurs de distribution des données peuvent être utilisés pour en savoir plus sur la normalité des données. Si le coefficient d'asymétrie (en anglais : skewness) et le coefficient d'aplatisement (en anglais kurtosis) (les deux coefficients) sont compris dans l'intervalle (-1 ; 1), il est suggéré que les données suivent une distribution normale ; si le coefficient d'asymétrie ou le coefficient d'aplatisement (l'un des coefficients) est en dehors de l'intervalle (-1 ; 1), nous avons une suggestion que les données ne suivent pas une distribution normale. Exemple dans le tableau : pour la variable D2Ulceration (mm) – les données suggèrent qu'il n'y a pas une distribution normale (asymétrie : oui = 4,30, non = 2,54 ; aplatisement : oui = 20,2 ; non = 6,58) ; pour la variable D1-D2 (mm) – les données suggèrent qu'il y a une distribution normale (asymétrie : oui = -0,0718, non = 0,119 ; aplatisement : oui = -0,480 ; non = -0,403).

Une option pour évaluer la normalité des données consiste à utiliser des tests statistiques pour évaluer la normalité. Si l'échantillon a moins de 50 observations, nous regardons le résultat du test de Shapiro-Wilk, s'il a plus de 50 observations, nous regardons le

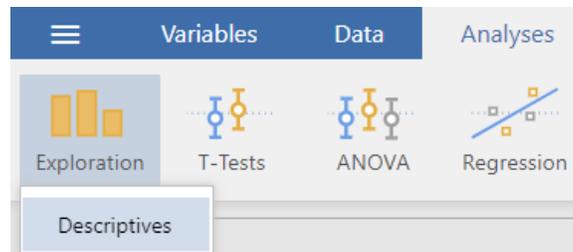
résultat du test de **Kolmogorov-Smirnov** (s'il est disponible). Le résultat d'intérêt est la **valeur p** du test (p de Shapiro-Wilk). Si la valeur **p** est **inférieure à 0,05**, on a une **suggestion** que les données **ne suivent pas une distribution normale**, **sinon** si la valeur **p** est **supérieure à 0,05**, on a une **suggestion** que les données suivent une **distribution normale**. Exemple dans le tableau : pour la variable D2 ulcération (mm) – les données suggèrent qu'il n'y a pas une distribution normale (oui : $p < 0,001$, non : $p < 0,001$) ; pour la variable D1-D2 (mm) – les données suggèrent qu'il y a une distribution normale (oui : $p = 0,547$, non : $p = 0,512$).

Une autre **façon d'évaluer la normalité** des données est d'utiliser des graphiques, tels que l'histogramme ou le graphique des quantiles.

Interprétation de graphiques d'histogrammes. Si le graphique suit une **forme en cloche approximativement et relativement symétrique**, on a une **suggestion** que les données suivent une **distribution normale**, **sinon** nous avons une suggestion que les données **ne suivent pas une distribution normale**. Exemple dans l'image : pour la variable D2Ulceration (mm) – suggestion qu'il n'y a pas une distribution normale; pour la variable D1-D2 (mm) – suggestion qu'il y a une distribution normale.

Interprétation du graphique des quantiles (QQ plot). Si la **tendance** des observations est qu'elles sont disposées **loin de la ligne continue** (qui représente une distribution normale), alors c'est une **suggestion** que les données **ne suivent pas une distribution normale**, **sinon**, si la tendance est que les observations sont proches de la ligne continue, c'est une **suggestion** que les données suivent une **distribution normale**.

L'évaluation de la normalité se fera dans l'**onglet Analyse**, le menu **Exploration**, la commande **Descriptives**.



Dans la section **Variables**, les variables quantitatives pour lesquelles la normalité doit être évaluée (par exemple Déulcération (mm), D1-D2 (mm)) sont déplacées. Dans la section **Split by**, la variable qualitative est déplacée, qui identifie les groupes à comparer (par exemple l'Aspirine).

Descriptives →

NrCr.

Sexe

Guerison

D1Ulceration (mm)

→

Variables

D2Ulceration (mm)

D1-D2 (mm)

Descriptives Variables across columns ▼

Frequency tables

Split by

Aspirine

→

Dans la section **Statistics** on doit cocher **Skewness, Kurtosis, Shapiro-Wilk.**

Statistics

Sample Size

N Missing

Percentile Values

Cut points for equal groups

Percentiles

Dispersion

Std. deviation Minimum

Variance Maximum

Range IQR

Mean Dispersion

Std. error of Mean

Confidence interval for Mean %

Central Tendency

Mean

Median

Mode

Sum

Distribution

Skewness

Kurtosis

Normality

Shapiro-Wilk

Outliers

Most extreme values

Dans la section **Plots**, on doit cocher **Histogram** și **Q-Q**.

Plots

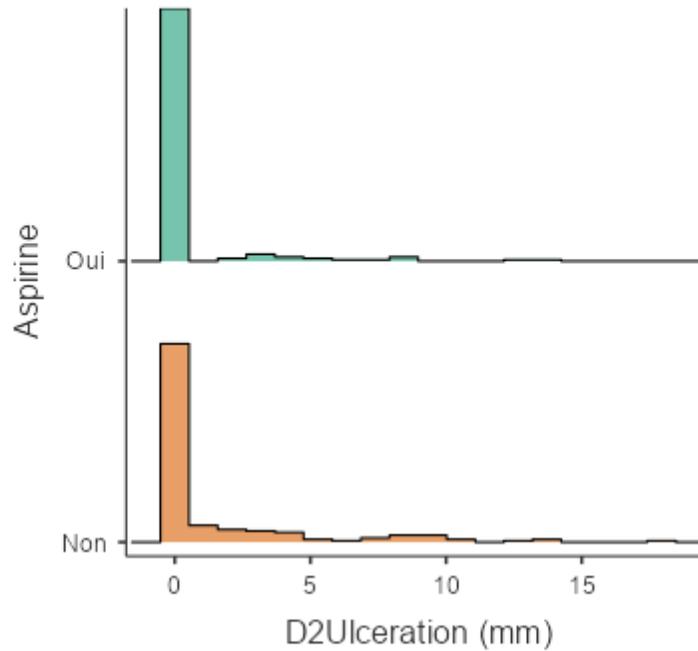
Histograms	Box Plots	Bar Plots
<input checked="" type="checkbox"/> Histogram	<input type="checkbox"/> Box plot	<input type="checkbox"/> Bar plot
<input type="checkbox"/> Density	<input checked="" type="checkbox"/> Label outliers	
Q-Q Plots	<input type="checkbox"/> Violin	
<input checked="" type="checkbox"/> Q-Q	<input type="checkbox"/> Data	
	Jittered ▾	
	<input type="checkbox"/> Mean	

Un tableau avec des statistiques sur chaque variable est obtenu, pour chaque groupe :

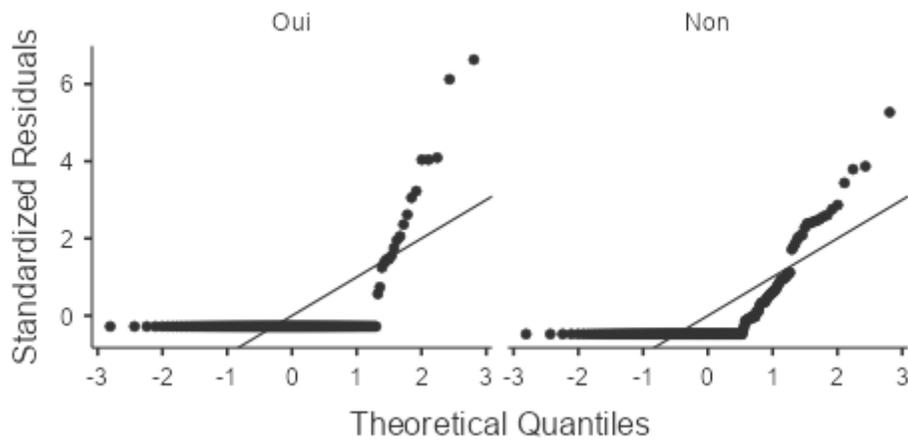
Descriptives			
	Aspirine	D2Ulceration (mm)	D1-D2 (mm)
N	Oui	200	200
	Non	200	200
Missing	Oui	0	0
	Non	0	0
Mean	Oui	0.547	23.7
	Non	1.49	22.3
Median	Oui	0.00	23.6
	Non	0.00	21.9
Standard deviation	Oui	1.97	3.85
	Non	3.12	4.18
Minimum	Oui	0.00	13.8
	Non	0.00	12.6
Maximum	Oui	13.6	32.9
	Non	17.9	33.2
Skewness	Oui	4.30	-0.0718
	Non	2.54	0.119
Std. error skewness	Oui	0.172	0.172
	Non	0.172	0.172
Kurtosis	Oui	20.2	-0.480
	Non	6.58	-0.403
Std. error kurtosis	Oui	0.342	0.342
	Non	0.342	0.342

L'**histogramme** d'une variable pour laquelle les données **ne suivent pas une distribution normale (asymétrie marquée à droite, leptokurtique marquée)** :

D2Ulceration (mm)

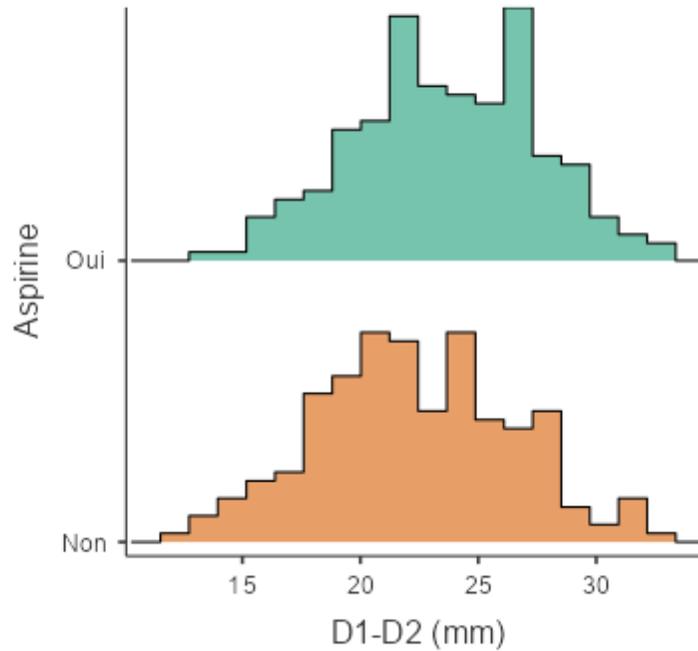


Graphique Q-Q plot (graphique des quantiles) pour une variable pour laquelle les données **ne suivent pas une distribution normale (les observations sont loin de la droite normale)** :

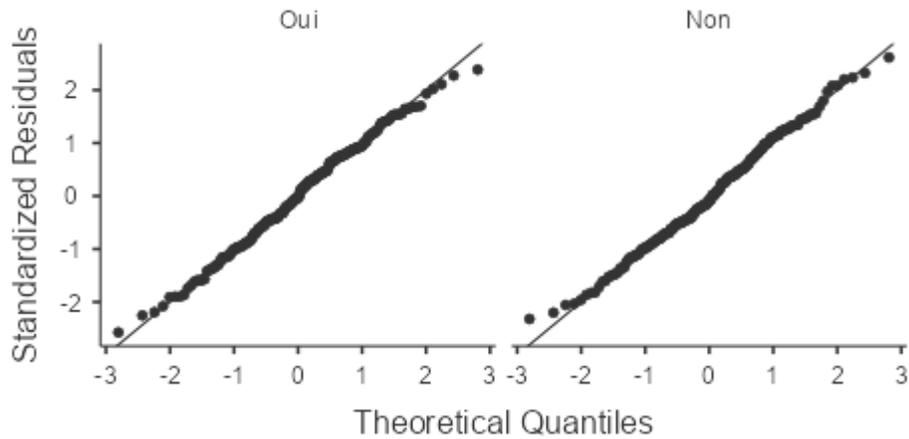


Le graphique de type **histogramme** d'une variable pour laquelle les données **suivent une distribution à peu près normale (graphique en forme de cloche, relativement symétrique)** :

D1-D2 (mm)

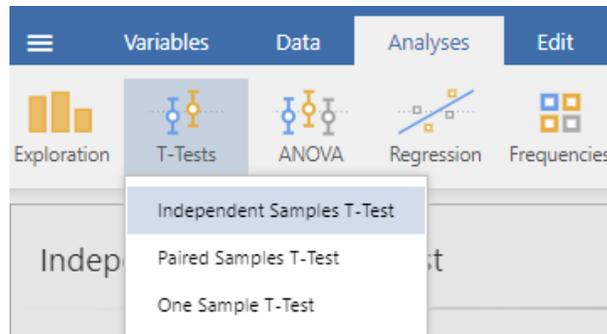


Graphique **Q-Q plot** (graphique des quantiles) d'une variable pour laquelle les données **suivent une distribution normale** (les observations sont proches de la droite normale) :



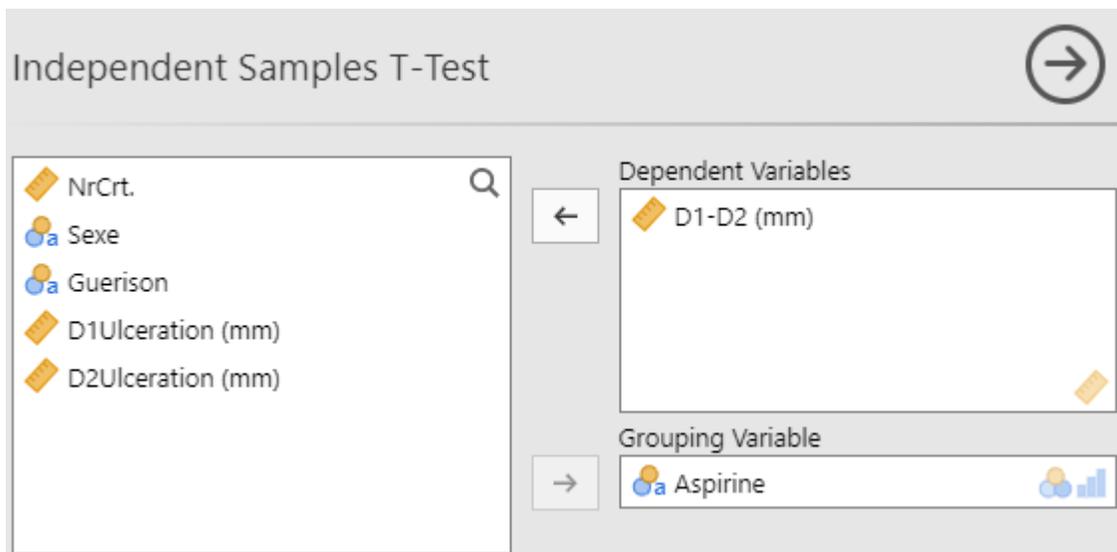
Réalisation du test de Student pour des échantillons indépendants

Dans l'onglet **Analyses**, choisissez l'option Test T d'échantillons indépendants (**Independent Samples T-Test**).



La variable qualitative qui **identifie les groupes comparés** se déplace en appuyant sur la flèche a cote du champ **Variable de regroupement (Grouping Variable)**.

La ou les variables **quantitatives** qui vous intéressent se déplacent en appuyant sur la flèche a cote du champ **Variable dépendante (Dependent Variable)**.



Cochez les options suivantes :

1. dans la section **Tests : Student, Welch**
2. Statistiques **complémentaires (Additional Statistics)**: **Différence moyenne (Mean difference)**, **Intervalle de confiance (Confidence interval)**, **Descriptifs (Descriptives)**, **Graphiques descriptifs (Descriptive plots)**.
3. Vérification **des hypothèses (Assumption Checks)**: **Test d'homogénéité (Homogeneity test)**.

<p>Tests</p> <p><input checked="" type="checkbox"/> Student's</p> <p><input type="checkbox"/> Bayes factor</p> <p>Prior <input type="text" value="0.707"/></p> <p><input checked="" type="checkbox"/> Welch's</p> <p><input type="checkbox"/> Mann-Whitney U</p> <p>Hypothesis</p> <p><input checked="" type="radio"/> Group 1 ≠ Group 2</p> <p><input type="radio"/> Group 1 > Group 2</p> <p><input type="radio"/> Group 1 < Group 2</p> <p>Missing values</p> <p><input checked="" type="radio"/> Exclude cases analysis by analysis</p> <p><input type="radio"/> Exclude cases listwise</p>	<p>Additional Statistics</p> <p><input checked="" type="checkbox"/> Mean difference</p> <p><input checked="" type="checkbox"/> Confidence interval <input type="text" value="95"/> %</p> <p><input type="checkbox"/> Effect size</p> <p><input type="checkbox"/> Confidence interval <input type="text" value="95"/> %</p> <p><input checked="" type="checkbox"/> Descriptives</p> <p><input checked="" type="checkbox"/> Descriptives plots</p> <p>Assumption Checks</p> <p><input checked="" type="checkbox"/> Homogeneity test</p> <p><input type="checkbox"/> Normality test</p> <p><input type="checkbox"/> Q-Q plot</p>
--	---

Choisir entre le test de Student à variance égales et le test de Welch – Variance inégales

Dans le cas **du test de Student pour les groupes indépendants**, il existe **deux variantes**, pour la situation où des **variances égales** ou **inégales** sont **supposées**. Pour choisir entre les deux, des tests peuvent être utilisés **pour comparer les variances**. Dans le tableau **Tests d'homogénéité des variances (Homogeneity test)**, nous examinons **la valeur de p** obtenue avec **le test de Levene**. Si **la valeur de p** est **inférieure à 0,05**, il s'agit d'une **suggestion** que les groupes **ont des variances inégales**, **sinon** il s'agit d'une **suggestion** que les groupes ont **des variances égales**.

Assumptions

Homogeneity of Variances Test (Levene's)

	F	df	df2	p
D1-D2 (mm)	1.02	1	398	0.314

Note. A low p-value suggests a violation of the assumption of equal variances

Si **les variances** sont supposées égales, nous choisissons les résultats correspondants pour **le test de Student**, sinon nous choisissons les résultats correspondants pour le **test de Welch** dans le

tableau **Test T des échantillons indépendants (Independent Samples T-Test)**. Dans la colonne **P**, nous avons la **valeur p** du test, dans la colonne **Différence moyenne (Mean difference)**, nous avons la **différence entre les moyennes des deux groupes**, et dans **Inférieur (Lower), Supérieur (Upper) (intervalle de confiance à 95 % (95% Confidence Interval))**, il y a l'**intervalle de confiance à 95 %** associé à la différence entre les moyennes des groupes.

Independent Samples T-Test

		Statistic	df	p	Mean difference	SE difference	95% Confidence Interval	
							Lower	Upper
D1-D2 (mm)	Student's t	3.44	398	< .001	1.38	0.402	0.591	2.17

Note. $H_a: \mu_{Oui} \neq \mu_{Non}$

Statistiques descriptives

Numérique

Les statistiques descriptives sont disponibles sous forme numérique dans le **tableau Descriptifs du groupe (Group Descriptives)**. Le **nombre de sujets** est indiqué dans la colonne **N**, la **moyenne** dans la colonne **Moyenne (Mean)**, et l'**écart-type** dans la colonne **SD**.

Group Descriptives

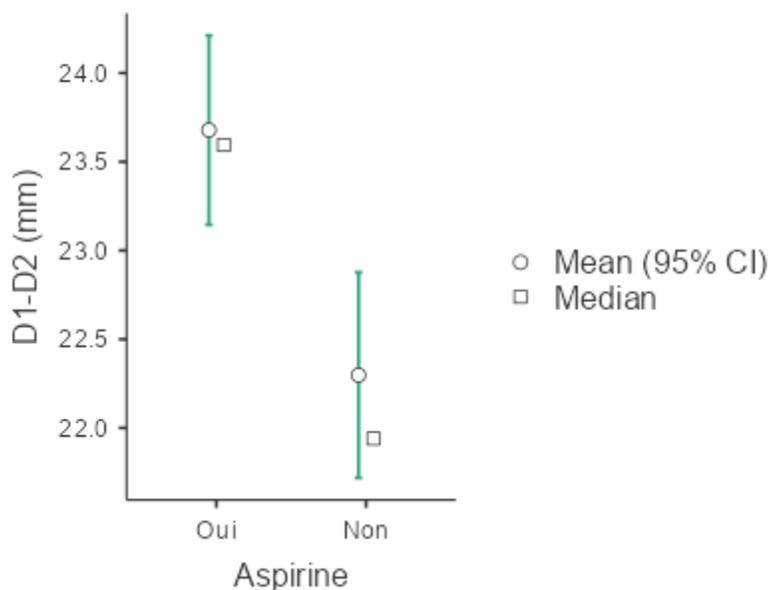
		Group	N	Mean	Median	SD	SE
D1-D2 (mm)	Oui		200	23.7	23.6	3.85	0.272
	Non		200	22.3	21.9	4.18	0.296

Graphique des moyennes

Un graphique des moyennes **est également fourni**, où le cercle représente la moyenne, les barres d'erreur représentent l'intervalle de confiance à 95 % et le carré représente la médiane.

Plots

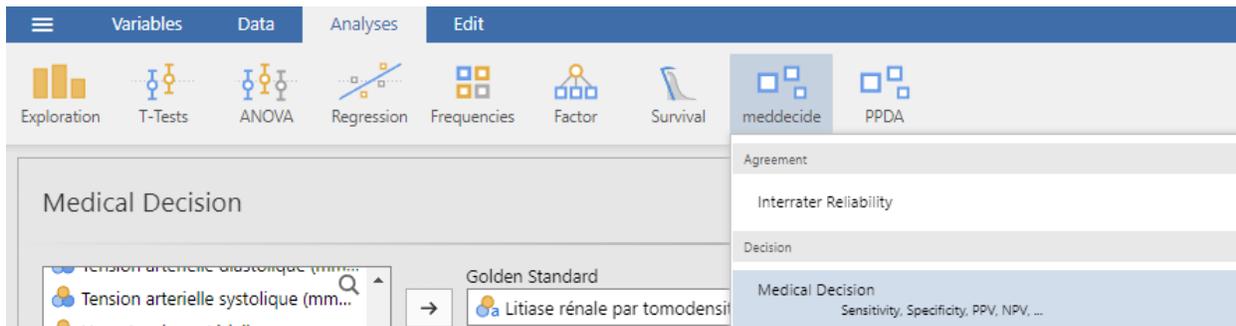
D1-D2 (mm)



Tests diagnostiques

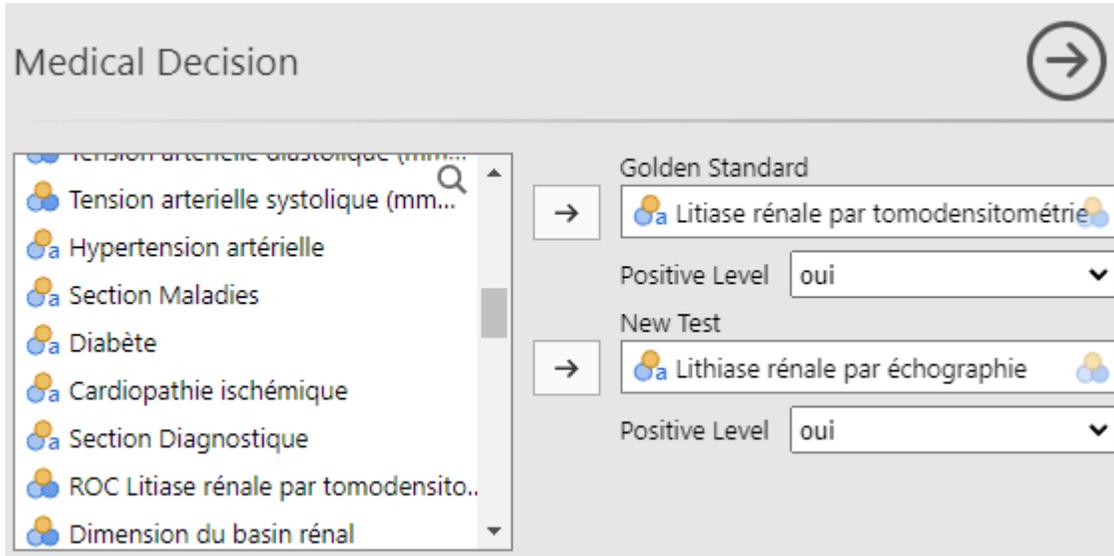
Assurez-vous que le module **Meddecide – Functions for medical decision in ClinicoPath** est installée. Voir le chapitre **Installation d'un module d'analyse supplémentaire**.

Dans l'onglet **Analyses**, appuyez sur le bouton **Meddecide**, dans le menu, sélectionnez l'option Medical Decision (Sensitivity, Specificity, PPV, NPV).

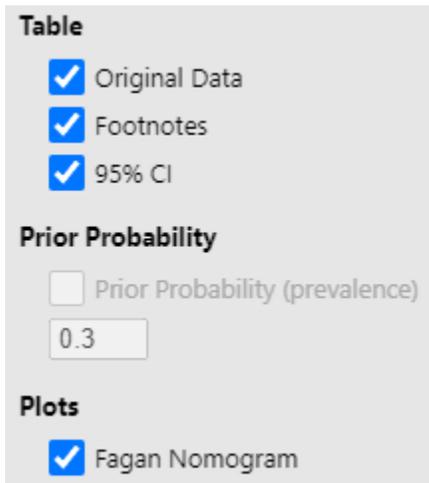


Sélectionnez la variable qui représente le **test standard** (par exemple, Lithiase rénale par tomodynamisme) et appuyez sur le **bouton fléché** à côté du champ **Golden Standard**. Précisez quelle catégorie de la variable représente la présence de la maladie, comme indiquée par le test standard (ici oui), dans la liste **Positive Level**. **Sélectionnez** la variable qui représente le **nouveau test** (par exemple, Lithiase rénale par échographie) et appuyez sur le **bouton fléché** à côté du champ **New Test**. Précisez

quelle catégorie de la variable représente un résultat positif du nouveau test (ici oui) , dans la liste **Positive Level**.



Selectionnez les suivantes options : dans la section Table : Original Data, Footnotes, 95% CI ; dans la section Plots : Fagan Nomogram.



Original Data

		Lithiase rénale par tomodensitométrie	
Lithiase rénale par échographie		non	oui
non		157	6
oui		12	40

Lithiase rénale par échographie Lithiase rénale par tomodensitométrie n

Lithiase rénale par échographie Lithiase rénale par tomодensitométrie n

1	non	non	157
2	non	oui	6
3	oui	non	12
4	oui	oui	40

**Vérifiez si les données sont bien mis dans le tableau en bas, par rapport au tableaux en haut !
S'il y a des problèmes changez l'option Positive Level.**

Recoded Data for Decision Test Statistics

	Gold Positive	Gold Negative	Total
Test Positive	40	12	52
Test Negative	6	157	163
Total	46	169	215

En bas il y a les statistiques des totaux des sujets, par différent catégories:

	n
Totalundefined	215 ^a
Diseasedundefined	46 ^b
Healthyundefined	169 ^d
Positive Testundefined	52 ^e
Negative Testundefined	163 ^f
True Testundefined	197 ^g
Wrong Testundefined	18 ^h

^a Total Number of Subjects

^b Total Number of Subjects with Disease

- ^d Total Number of Healthy Subjects
- ^e Total Number of Positive Tests
- ^f Total Number of Negative Tests
- ^g Total Number of True Test Results
- ^h Total Number of Wrong Test Results

En bas il y a les **statistiques classiques des tests diagnostiques, avec leur intervalle de confiance 95%**.

	Ratios
Sensitivityundefined	87.0% ^a
Specificityundefined	92.9% ^b
Accuracyundefined	91.6% ^d
Prevalenceundefined	21.4% ^e
Positive Predictive Valueundefined	76.9% ^f
Negative Predictive Valueundefined	96.3% ^g
Post-test Disease Probabilityundefined	76.9% ^h
Post-test Health Probabilityundefined	96.3% ⁱ
Positive Likelihood Ratioundefined	12.2
Negative Likelihood Ratioundefined	0.140

^a Sensitivity (True Positives among Diseased)

^b Specificity (True Negatives among Healthy)

^d Accuracy (True Test Result Ratio)

^e Disease Prevalence in this population

^f Positive Predictive Value (Probability of having disease after a positive test using this experimental population)

^g Negative Predictive Value (Probability of being healthy after a negative test using this experimental population)

^h Post-test Probability of Having Disease (Probability of having disease after a positive test using known Population Prevalence)

ⁱ Post-test Probability of Being Healthy (Probability of being healthy after a negative test using known Population Prevalence)

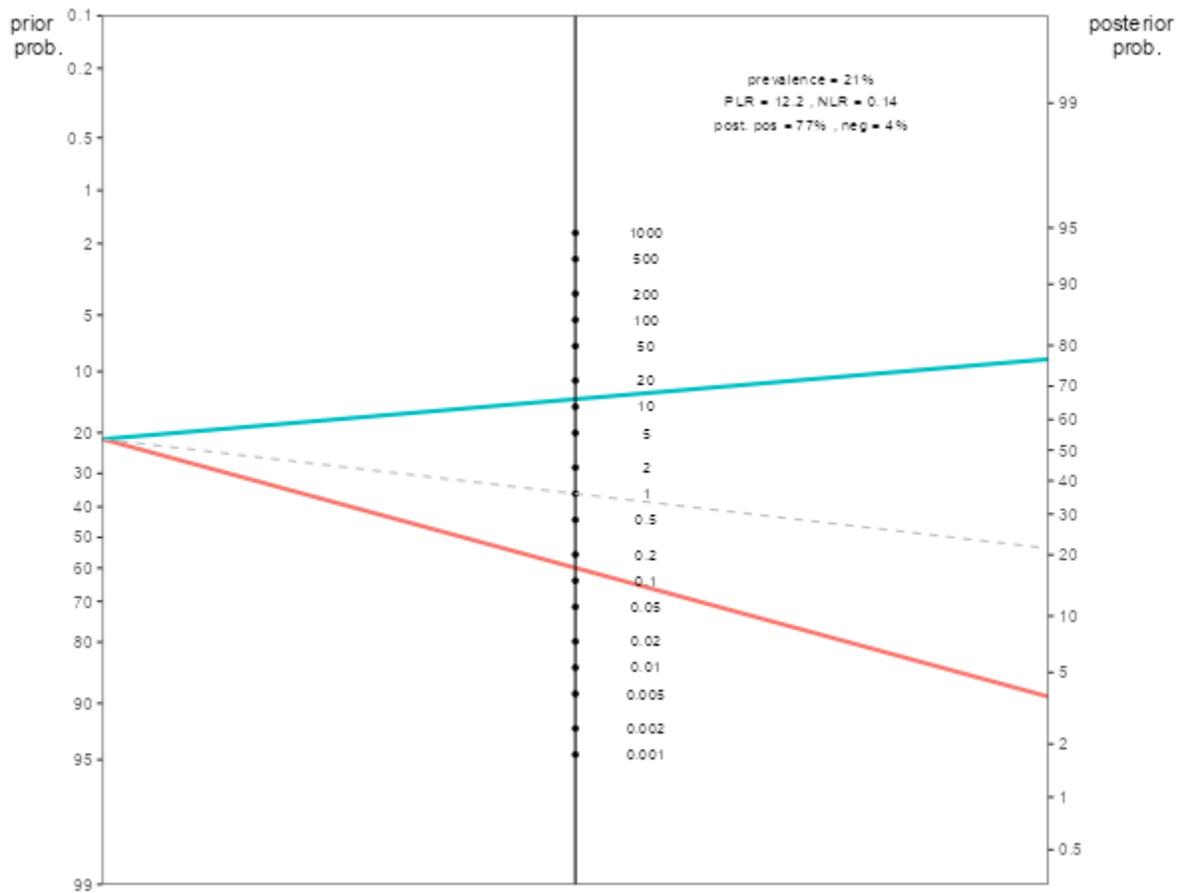
Decision Statistics	Estimate	95% Confidence Interval	
		Lower	Upper
Apparent prevalence	24.2%	18.6%	30.5%
True prevalence	21.4%	16.1%	27.5%
Test sensitivity	87.0%	73.7%	95.1%
Test specificity	92.9%	87.9%	96.3%

Diagnostic accuracy	91.6%	87.1%	95.0%
Positive predictive value	76.9%	63.2%	87.5%
Negative predictive value	96.3%	92.2%	98.6%
Proportion of false positives	7.1%	3.7%	12.1%
Proportion of false negative	13.0%	4.9%	26.3%
False Discovery Rate	23.1%	12.5%	36.8%
False Omission Rate	3.7%	1.4%	7.8%

Decision Statistics	Estimate	95% Confidence Interval	
		Lower	Upper
Diagnostic odds ratio	87.222	30.8399	246.684
Number needed to diagnose	1.252	1.0949	1.622
Youden's index	0.799	0.6167	0.913
Likelihood ratio of a positive test	12.246	7.0183	21.369
Likelihood ratio of a negative test	0.140	0.0665	0.296

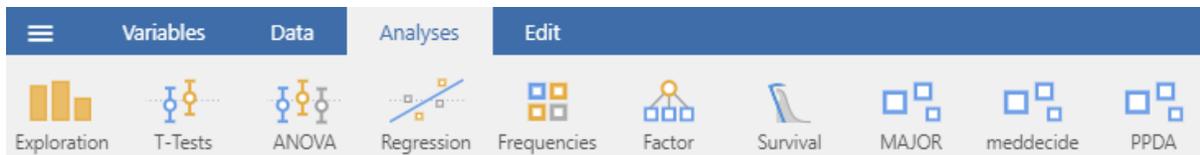
En bas il y a la **nomogramme Fagan**, basée sur la prévalence observée dans l'échantillon, en imaginant un résultat du test positive, et un autre négatif.

Fagan nomogram



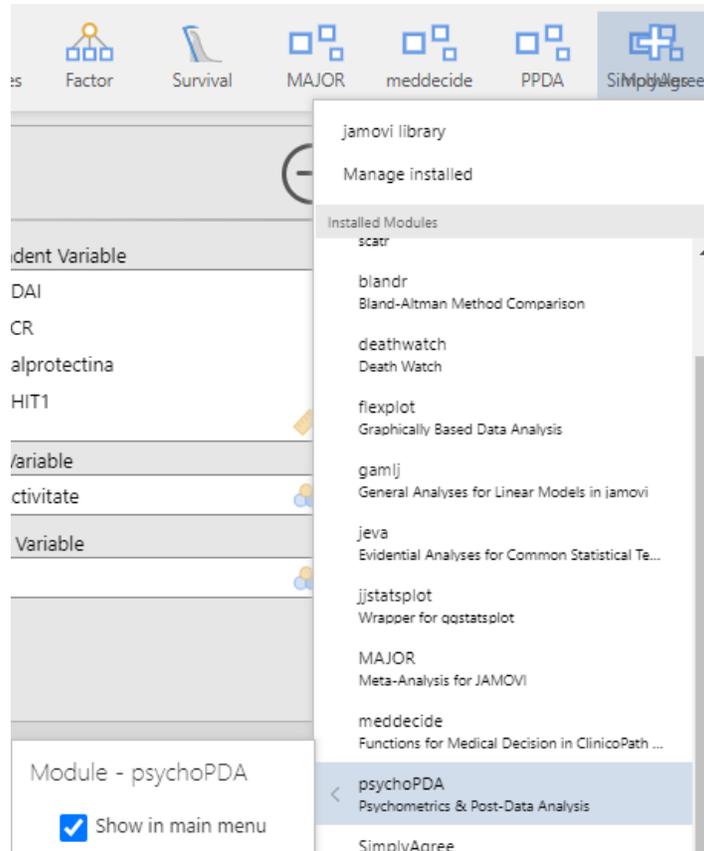
Analyse ROC

Dans l'onglet **Analyses**, appuyez sur le bouton **PPDA**.



Si

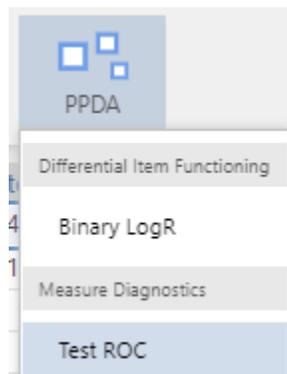
vous ne le trouvez pas, il se peut qu'il ne s'affiche pas. Dans l'onglet **Analyses**, à **droite**, appuyez sur **+** Modules et vérifiez s'il n'est pas présent dans la liste des modules installés.



Vous pouvez vérifier que le module est visible dans l'onglet **Analyses**, en cliquant sur l'option **Afficher dans le menu principal (Show in main menu)**.

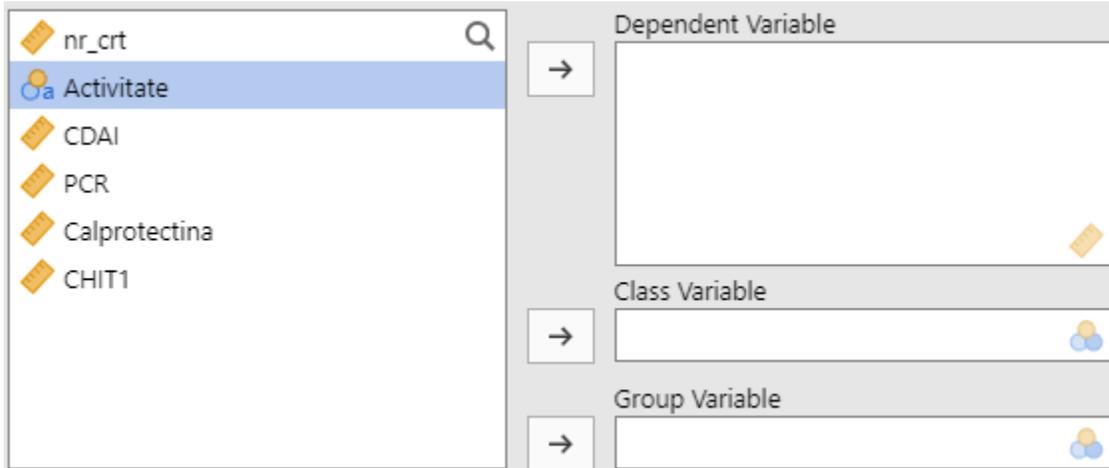
S'il n'est pas installé, suivez les étapes indiquées dans le chapitre Installer un **modul d'analyse supplémentaire**.

En appuyant sur le module PPDA, sélectionnez l'option Test ROC dans le menu

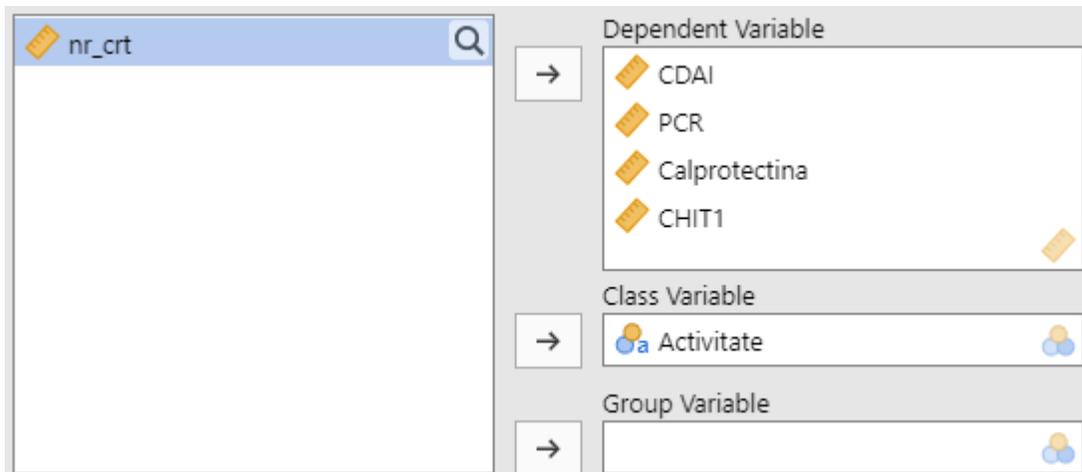


Sélectionnez la variable qui représente le **test standard** (par exemple, Activité) et appuyez sur le **bouton fléché** à côté du champ **Variable de classe (Class Variable)**. Sélectionnez **les variables**

quantitatives qui représentent les tests d'intérêt **et appuyez sur le bouton fléché à côté du** champ Variable dépendante (Dependent Variable).



À la fin, vous obtiendrez quelque chose de similaire à l'image ci-dessous :



Tableaux de valeurs limites et statistiques associées

Le programme a déjà effectué l'analyse ROC et vous présente sur les tableaux de droite pour différentes valeurs seuil (**Cutpoint**), la sensibilité (**Se**), la spécificité (**Sp**), valeur prédictive positive (**VPP**) et négatif (**VPN**), l'indice de Youden (**L'index de Youden**), l'aire sous la courbe ROC (**AUC**).

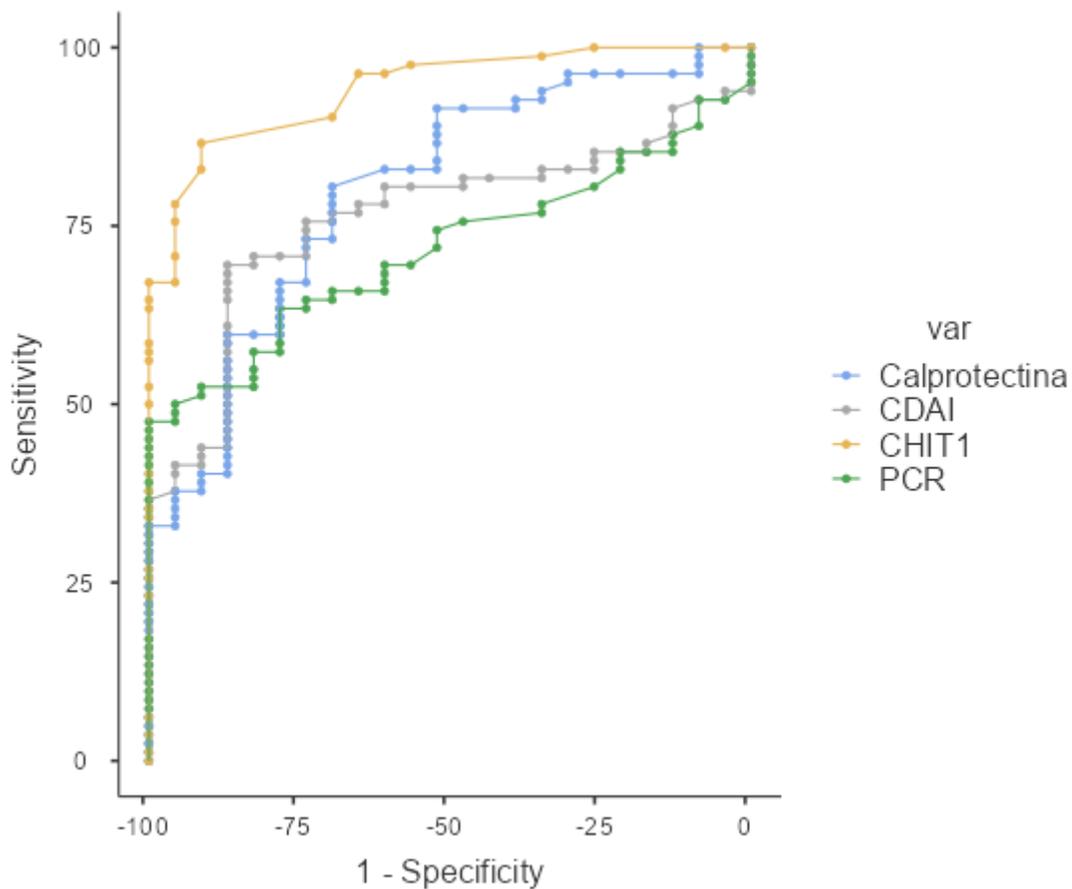
Ainsi la **meilleure valeur seuil**, dont l'indice de Youden a la valeur la plus élevée, est de 210, avec une sensibilité et une spécificité associées de 80,49 % et 69,57 %, respectivement.

Scale: Calprotectina

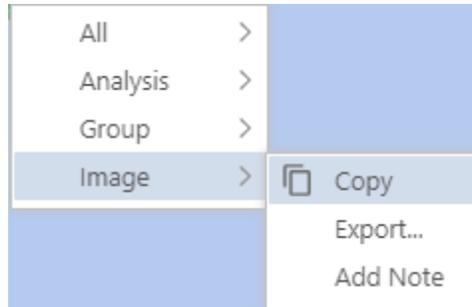
Cutpoint	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Youden's index	AUC	Metric Score
210	80.49%	69.57%	90.41%	50%	0.501	0.801	1.50
215	79.27%	69.57%	90.28%	48.48%	0.488	0.801	1.49
220	78.05%	69.57%	90.14%	47.06%	0.476	0.801	1.48
221.6	76.83%	69.57%	90%	45.71%	0.464	0.801	1.46
240	75.61%	69.57%	89.86%	44.44%	0.452	0.801	1.45
280	73.17%	73.91%	90.91%	43.59%	0.471	0.801	1.47
300	71.95%	73.91%	90.77%	42.5%	0.459	0.801	1.46
400	67.07%	78.26%	91.67%	40%	0.453	0.801	1.45
596	59.76%	86.96%	94.23%	37.74%	0.467	0.801	1.47
600	58.54%	86.96%	94.12%	37.04%	0.455	0.801	1.45

Graphique de la courbe ROC

Ci-dessous est montré le graphique avec la courbe ROC (caractéristique de fonctionnement du récepteur).

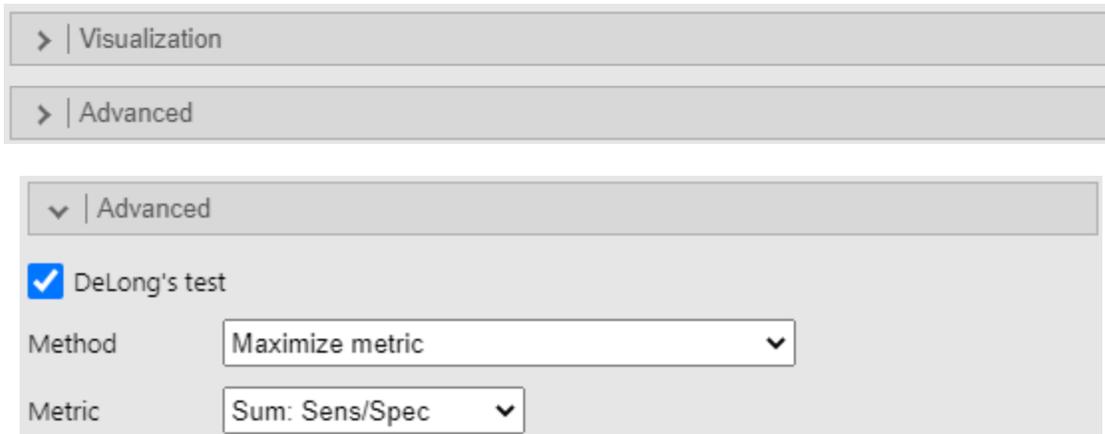


Pour copier l'image, cliquez avec le bouton droit de la souris et sélectionnez **Copier l'image (Image / Copy)**



Comparaison des courbes ROC par des tests statistiques

Pour effectuer **des tests statistiques comparant les courbes ROC** entre elles, cliquez sur le bouton > à côté de **Avancé (Advanced)**, puis sélectionnez l'option **Test de De Long (DeLong's test)**.



Les résultats se présentent comme suit :

DeLong Test of Difference between AUCs

Estimated AUC's:

	AUC	SD(Hanley)	P(H0: AUC=0.5)	SD(DeLong)	P(H0: AUC=0.5)
1	0.762	0.050	0.000	0.048	0.000
2	0.717	0.055	0.000	0.049	0.000
3	0.801	0.045	0.000	0.050	0.000
4	0.942	0.022	0.000	0.022	0.000

Pairwise comparisons:

	AUC Difference	CI(lower)	CI(upper)	P.Value	Correlation
1 vs. 2	0.045	-0.059	0.149	0.400	0.400
1 vs. 3	-0.039	-0.145	0.066	0.465	0.409
1 vs. 4	-0.179	-0.266	-0.092	0.000	0.406
2 vs. 3	-0.084	-0.187	0.019	0.110	0.437
2 vs. 4	-0.224	-0.318	-0.131	0.000	0.268
3 vs. 4	-0.140	-0.228	-0.052	0.002	0.455

Overall test:

p-value = 1.28e-06

Le premier tableau présente des estimations **des aires sous la courbe ROC** pour chaque test diagnostique, un écart-type et un **test de signification statistique pour une courbe ROC** pour chaque variable sélectionnée dans l'ordre dans lequel elles sont sélectionnées. Ainsi, pour le score d'activité clinique CDAI, l'aire sous la courbe ROC est de 0,717 et le résultat est statistiquement significatif P (H_0 : $ASC = 0,5$) inférieur à 0,05 (valeur du $p < 0.001$). **L'hypothèse nulle** du test pour une **courbe ROC** est la suivante : l'aire sous la courbe ROC pour CDAI, par rapport à l'examen histopathologique, n'est pas statistiquement significativement différente de 0,5. **L'hypothèse alternative** du test est la suivante : l'aire sous la courbe ROC pour le CDAI par rapport à l'examen histopathologique est statistiquement significativement différente de 0,5.

Dans le deuxième tableau, il y a des comparaisons entre les tests effectués deux par deux, montrant la différence entre les surfaces des courbes ROC (différence AUC), avec un intervalle de confiance de 95%, une valeur de test statistique (**P.Value**). Par exemple, si l'on compare le score d'activité clinique CDAI (1) avec la protéine C-réactive (2), le résultat n'est pas statistiquement significatif ($p = 0,400$). **Hypothèse nulle** du **test de comparaison de deux courbes ROC** : Il n'y a pas de différence statistiquement significative entre la précision diagnostique du score d'activité clinique CDAI et la protéine C-réactive, mesurée par AUC, avec l'examen histopathologique comme standard. **Hypothèse alternative** du test de comparaison de deux courbes ROC : Il existe une différence statistiquement significative entre la précision diagnostique du score d'activité clinique CDAI et la protéine C-réactive, mesurée par AUC, ayant comme standard l'examen histopathologique.

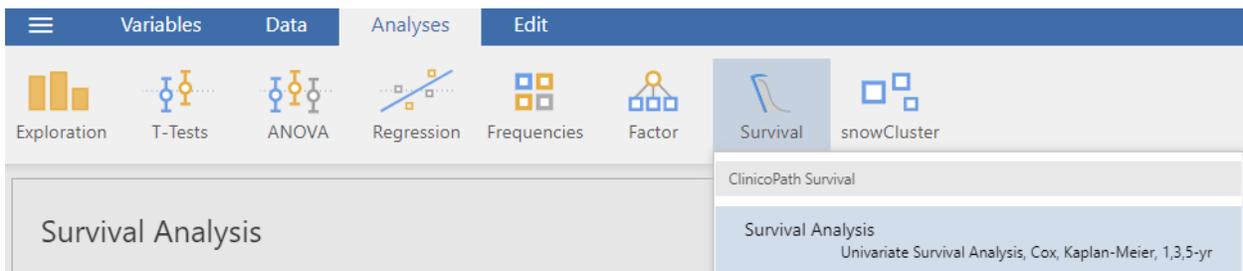
Enfin, le résultat d'un test statistique global comparant toutes les courbes ROC entre elles est présentée.

Analyse de survie (graphique Kaplan Meyer, test log-rank, regression Cox - HR)

Assurez-vous que le module **jsurvival - Survival Module of ClinicoPath** for Jamovi 0.0.2. Voir le chapitre **Installation d'un module d'analyse supplémentaire**.

Vérifiez dans l'onglet **Data**, que le type des variables est bien choisi, et que l'ordre des catégories pour chaque variable est comme vous préférez.

Dans l'onglet **Analyses**, cliquez sur le bouton **Survival**, et sélectionnez dans le menu la commande **Survival Analysis (Univariate Survival Analysis, Cox, Kaplan-Meier, 1,3,5-yr)**.



Déplacez la variable qui indique le **temps de survie** avec un clic sur la flèche à côté du champ **Time Elapsed**. Déplacez la variable qui indique le **statut de censure/l'évènement** avec un clic sur la

flèche a cote du champ critère de jugement (**Outcome**). Assurez vous que la **catégorie** qui indique **l'évènement** est sélectionne dans la liste **Event Level** (ex. 1). Déplacez la variable de regroupement avec un clic sur la flèche a cote du champ variable explicative (**Explanatory Variable**) (ex. Type, Traitement, Facteur de risque).

Survival Analysis

ID
Traitement

Time Elapsed
Temp_mois

Outcome
Censure

Event Level 1

Explanatory Variable
Type_histologique

> | Advanced Elapsed Time Options

> | Analysis with Multiple Outcomes

> | Single Arm Survival

> | Pairwise Comparisons

∨ | Plots

Plots

Survival Plot

Dans la section **Plots**, cochez l'option graphique de survie (**Survival Plot**).

Vous pouvez modifier les options du graphique : de montrer la valeur du p (**p-value**), le tableau avec le nombre des sujets au risque (**risktable** – utile pour montrer aussi la legende), le cas avec du censure (**censored**), les limites sur les axes.

▼ | Plots

Plots

Survival Plot

KMunicate-Style Plot

Cumulative Events

Cumulative Hazard

Plot Arguments

Plot End Time

Time Interval

95% CI

risktable

censored

Pour savoir la probabilité de survie à différents moments, vous pouvez indiquer ces moments dans la section tableau de survie (**Survival table**), le champ valeurs seuils (**Cutpoints**).

▼ | Survival Tables

Survival table

Cutpoints

La médiane du temps de survie

Dans la section des résultats de synthèse des résultats concernant la médiane de survie (**Median Survival Summary and Table**), on peut trouver dans le texte avant le tableau les médianes de survie avec des intervalles de confiance 95% dans le texte avant le tableau. Aussi, dans le tableau, dans la colonne **Median** on trouve la médiane par group, et l'intervalle de confiance 95% dans les colonnes Inferieur et Supérieur Intervalle de confiance 95% (**Lower, et Upper 95% Confidence interval**). Par exemple pour le groupe avec histologie à petites cellules la médiane de survie est 2 mois.

Median Survival Summary and Table - Type_histologique

When Type_histologique is Petites cellules, median survival is 2 [1 - 2, 95% CI] months.
When Type_histologique is Squameux, median survival is 4 [3 - 10, 95% CI] months.

Median Survival Table: Levels for Type_histologique

Levels	Records	Events	rmean	se_rmean	Median	95% Confidence Interval	
						Lower	Upper
Type_histologique=Petites cellules	48	45	2.72	0.497	2.00	1.00	2.00
Type_histologique=Squameux	35	31	7.76	1.601	4.00	3.00	10.00

Regression Cox – hazard ratio

Dans le tableau avec les résultats de la régression Cox (**Cox regression summary and table**), vous pouvez trouver le hazard ratio avec l'intervalle de confiance 95%, dans la colonne **HR (univariable)**. Pour savoir quel group est le **groupe de référence**, dans la colonne **HR (Univariable)** vous cherchez le groupe qui montre **sign -**. Dans l'exemple plus bas, on peut observer que le groupe de référence est le type histopathologique a petites cellules. Donc le HR est obtenu par la division du hazard des participants avec un cancer avec histologie Squameux, a le hazard des participants avec un cancer avec histologie a petites cellules.

Cox Regression Summary and Table - Type_histologique

When Type_histologique is Squameux, there is 0.43 (0.26-0.72, p=0.001) times risk than when Type_histologique is Petites cellules.

Cox Table- Type_histologique

Explanatory	Levels	all	HR (Univariable)
Type_histologique	Petites cellules	48 (57.8)	-
	Squameux	35 (42.2)	0.43 (0.26-0.72, p=0.001)

La probabilité de survie a différents moments

La probabilité de survie a différents moments (par exemple dans la colonne temps (**Time**), 12 mois) peut être trouve dans le tableau Sommaire de survie et tableau (**Survival Summary and Table**), dans la colonne Survie (**Survival**), et l'intervalle de confiance 95% dans les colonnes Inferieur et Superieur Intervalle de confiance 95% (**Lower, et Upper 95% Confidence interval**), mais aussi en format texte, avant le tableau. Par exemple pour le groupe avec histologie a petites cellules la probabilité de survie a 12 mois est 5,9%.

1, 3, 5-yr Survival Summary and Table - Type_histologique

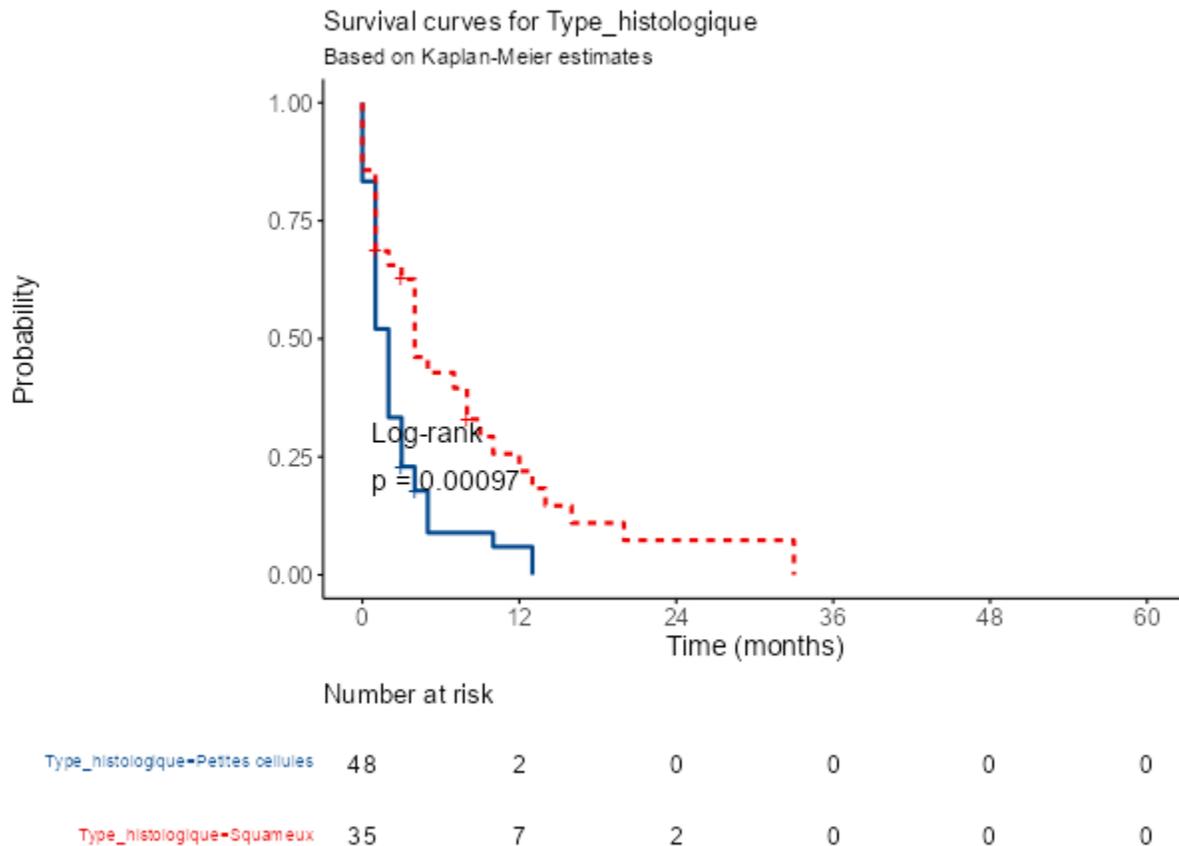
When Type_histologique=Petites cellules, 12 month survival is 6% [1.6%-22%, 95% CI].
 When Type_histologique=Squameux, 12 month survival is 22% [11.1%-43%, 95% CI].

1, 3, 5 year Survival - Type_histologique

Levels	time	Number at Risk	Number of Events	Survival	95% Confidence Interval	
					Lower	Upper
Type_histologique=Petites cellules	12	2	43	5.9 %	1.6 %	21.6 %
Type_histologique=Squameux	12	7	25	22.0 %	11.1 %	43.3 %

La courbe de survie Kaplan-Meier, avec le test log-rank

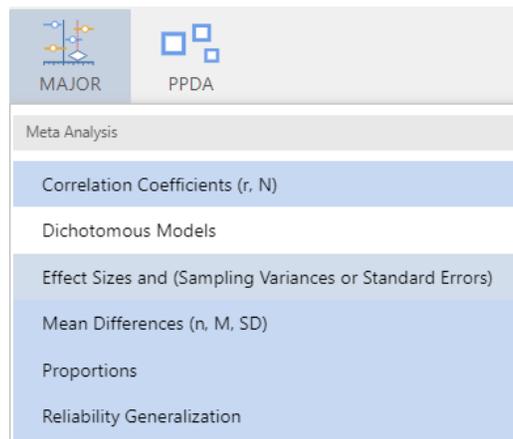
La courbe de survie Kaplan-Meier, avec la valeur du p du test log-rank, et le tableau avec le nombre des sujets au risque, peut être trouve dans la section graphique de survie (**Survival Plot**). Dans le tableau avec le nombre des sujets au risque, on peut observer au début de l'étude il y a 48 sujets vivants (a risque de faire l'évènement – le décès) dans le groupe avec histologie a petites cellules, et 35 sujets dans le groupe avec histologie squameus. A 12 mois il reste seulement 2 sujets vivants (a risque de faire l'évènement – le décès) dans le groupe avec histologie à petites cellules, et 7 sujets dans le groupe avec histologie squameus.



Méta-analyse

Méta-analyse pour comparer deux groupes en termes de données dichotomiques

On s'assure que le module **MAJOR** soit installé. S'il n'est pas installé, suivez les instructions du chapitre suivant Installation d'un module d'analyse supplémentaire (par exemple pour l'analyse ROC). On s'assure également que le type de variables contenant le nombre d'événements, respectivement le nombre total de sujets, soit de type Continu. Dans le module **MAJOR**, nous choisissons l'option **Dichotomous Models**.



Déplacez la variable indiquant le nombre d'événements dans le groupe expérimental/intervention/exposé dans le champ **Number of Incidents in Experimental Group**; la variable indiquant le nombre d'événements dans le groupe témoin/non-exposé dans le champ **Number of Incidents in Control Group**; la variable indiquant le nombre total de participants dans le groupe expérimental/d'intervention/exposé dans le champ **Total Sample Size for Experimental Group** (nombre de sujets dans le groupe expérimental); la variable indiquant le nombre total de participants dans le groupe témoin ou non exposé dans le champ **Total Sample Size for Control Group**; la variable indiquant le nom de l'étude dans le champ **Study Label**.

Dans la section **Model Options** (Options du modèle), sélectionnez **Model Estimator** (Estimateur du modèle): **Restricted Maximum-Likelihood**; pour **Model measure** (l'estimateur d'intérêt d'association), choisissez **Logs Odds Ratio** (pour les études cas témoins), **Log risk ratio** ou **Risk difference** (pour les études de cohorte). Cochez **Back-Transform Log Odds Ratio to Odds Ratio**. Le reste des options peut être laissé tel qu'il est. Si vous le souhaitez, en cas d'hétérogénéité réduite vous choisissez la méthode **Fail-Safe N, Rosenberg**.

Dichotomous Models →

Type d'étude Q

→ Number of Incidents in Experimental Group
Evénements chez les exposés

→ Number of Incidents in Control Group
Evénements chez les témoins

→ Total Sample Size for Experimental Group
Total exposés

→ Total Sample Size for Control Group
Total témoins

→ Study Label
Nom d'étude

→ Moderator (optional)

Model Options

Model estimator Restricted Maximum-Likelihood

Model measures Log odds ratio

Moderator type No Moderator

Back-Transform Log Odds Ratio to Odds Ratio

L'estimateur de l'association

Si l'option **Log odds ratio** est sélectionnée comme estimation de l'association, et si l'option **Back-Transform Log Odds Ratio to Odds Ratio** est cochée, dans le tableau du même nom, vous trouverez la valeur de l'estimateur ponctuel de l'OR, ainsi que l'intervalle de confiance à 95 %.

Back-Transform Log Odds Ratio to Odds Ratio		
Odds Ratio	CI Lower Bound	CI Upper Bound
1.542	1.135	2.096

Dans le tableau **Random-Effects Model**, vous trouverez la valeur estimée de l'association logarithme odds ratio/log risque ratio/différence de risque dans la colonne **estimate**, avec l'intervalle de confiance dans les deux dernières colonnes. La valeur du p pour la signification statistique se trouve dans la colonne p.

Random-Effects Model (k = 10)

	Estimate	se	Z	p	CI Lower Bound	CI Upper Bound
Intercept	0.433	0.157	2.77	0.006	0.126	0.740

Note. Tau² Estimator: Restricted Maximum-Likelihood

Hétérogénéité statistique des études

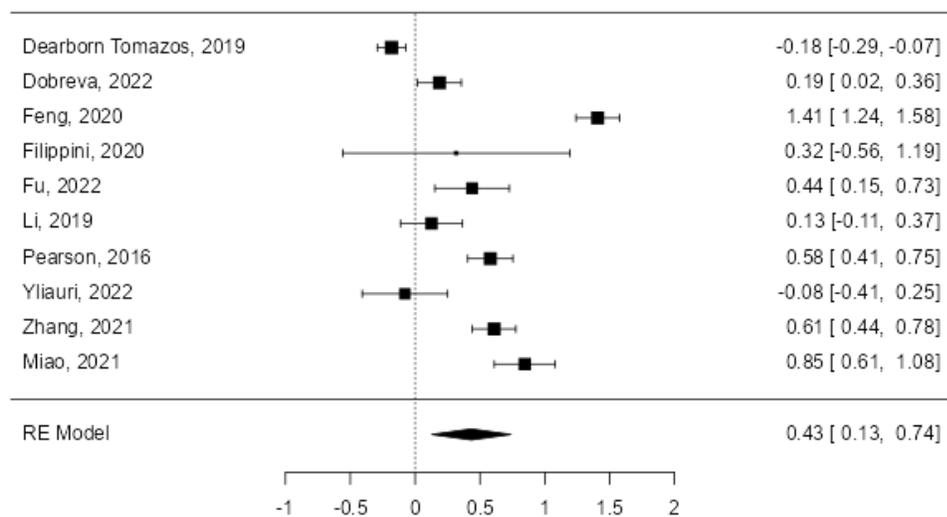
Dans le tableau **Heterogeneity Statistics**, l'indice d'incohérence se trouve dans la colonne I² et la valeur du p du test associé se trouve dans la colonne p.

Heterogeneity Statistics

Tau	Tau ²	I ²	H ²	R ²	df	Q	p
0.471	0.222 (SE= 0.1147)	95.71%	23.304	.	9.000	288.885	< .001

Graphique foret (forest plot)

Dans la section **forest plot**, vous trouverez le graphe homonyme. Pour les estimateurs du log odds ratio, log risk ratio, les valeurs logarithmiques de OR ou RR sont entrées dans le graphique. L'avantage est la symétrie des intervalles de confiance. L'inconvénient est qu'il est plus difficile à les interpréter. La valeur de 0 est équivalente à un OR ou à un RR de 1. Les valeurs de graphe supérieures à 0 pour OR/RR correspondent à des valeurs OR/RR supérieures à 1 et les valeurs inférieures à 0 correspondent à des valeurs inférieures à 1.



Le biais de publication

Les statistiques

Le résultat du test de régression pour l'asymétrie du graphique en entonnoir (**Regression Test for Funnel Plot Asymetry**), la valeur p dans la section **Publication Bias Assessment** (évaluation du biais de publication), peuvent être utilisés pour évaluer l'erreur de publication systématique.

Publication Bias Assessment

Fail-Safe N Analysis (File Drawer Analysis)

Fail-safe N	p
593.000	< .001

Note. Fail-safe N Calculation Using the Rosenthal Approach

Rank Correlation Test for Funnel Plot Asymmetry

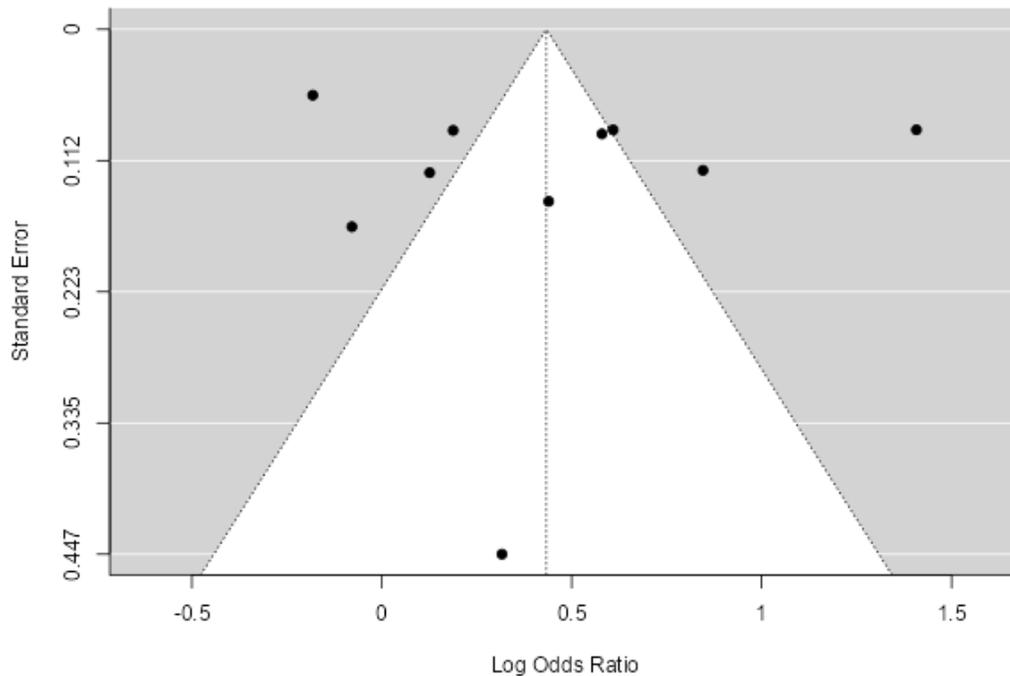
Kendall's Tau	p
-0.156	0.601

Regression Test for Funnel Plot Asymmetry

Z	p
-0.306	0.760

Le graphique en entonnoir

Le graphique du tracé de l'entonnoir se trouve dans la section **Funnel plot**.

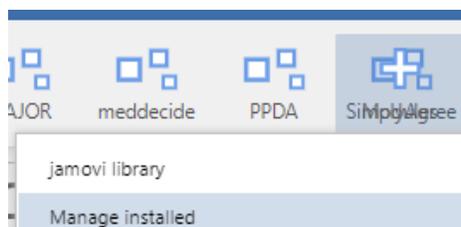


Installation d'un module d'analyse supplémentaire (par exemple pour l'analyse ROC)

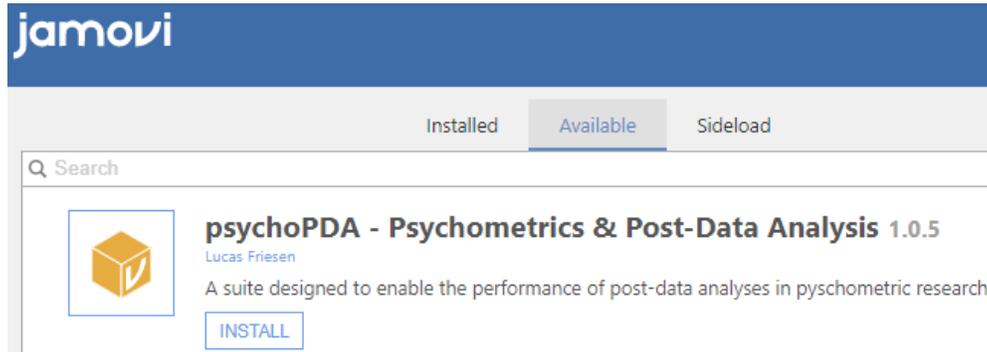
Dans l'onglet Analyses, à droite, appuyez sur + Modules et vérifiez s'il n'est pas présent dans la liste des modules installés.



Pour installer des modules d'analyse en plus de ceux par défaut, sélectionnez l'onglet **Analyses**, puis sur le côté droit, cliquez sur le bouton + **Modules**, puis sélectionnez **Gérer les modules installés (Manage installed)**.



Sélectionner l'onglet **Disponible (Available)**, faites défiler jusqu'à ce que vous trouviez le module souhaité (par exemple psychoPDA) et appuyez sur la touche **INSTALLER (INSTALL)**.



Revenez à la fenêtre d'analyse en appuyant sur le bouton fléché :



Instructions d'utilisation de EpiInfo et Excel pour l'analyse des données

Travailler avec Epi Info

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

Utilisez un navigateur Internet pour naviguer

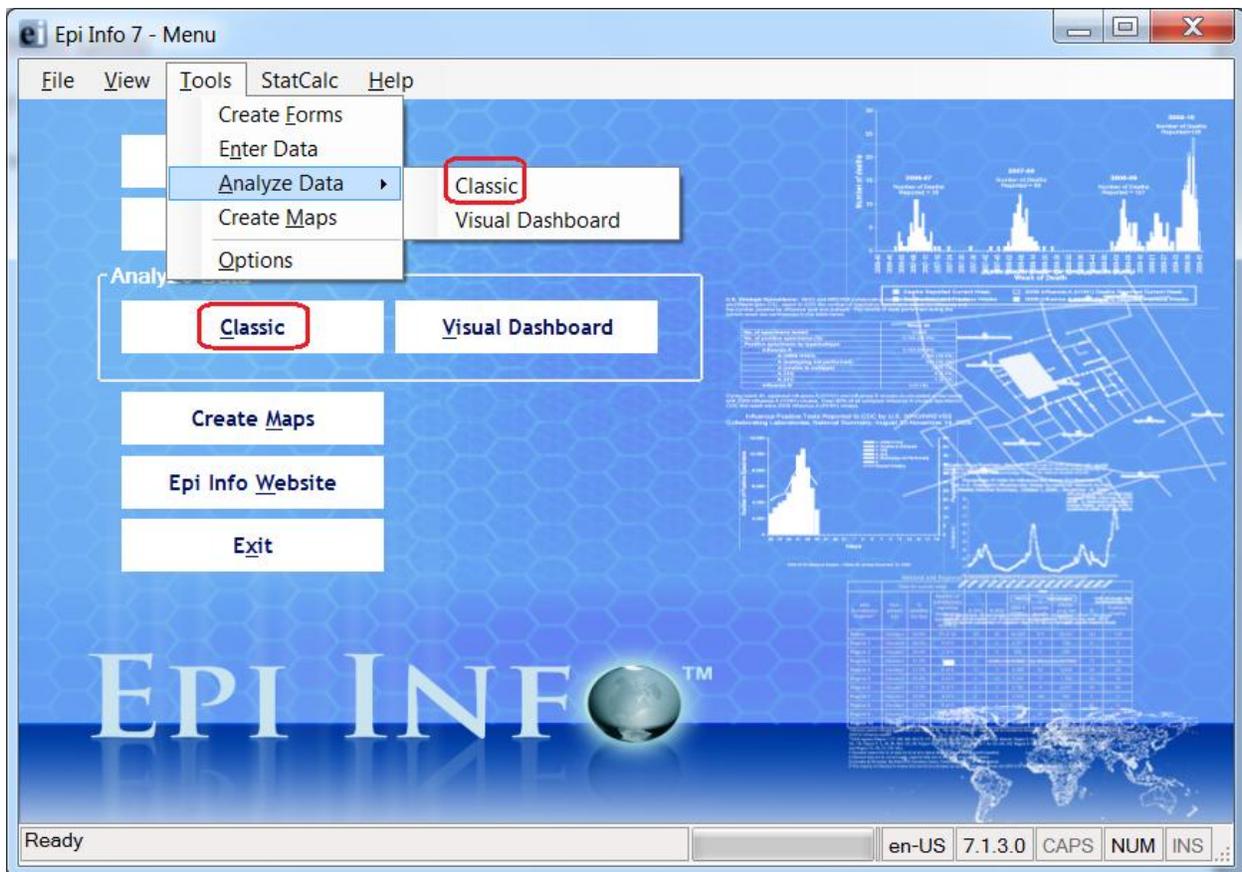
vers: <http://www.cdc.gov/epiinfo/installation.htm> et suivez les instructions d'installation.

Ouverture du programme *Epi Info*

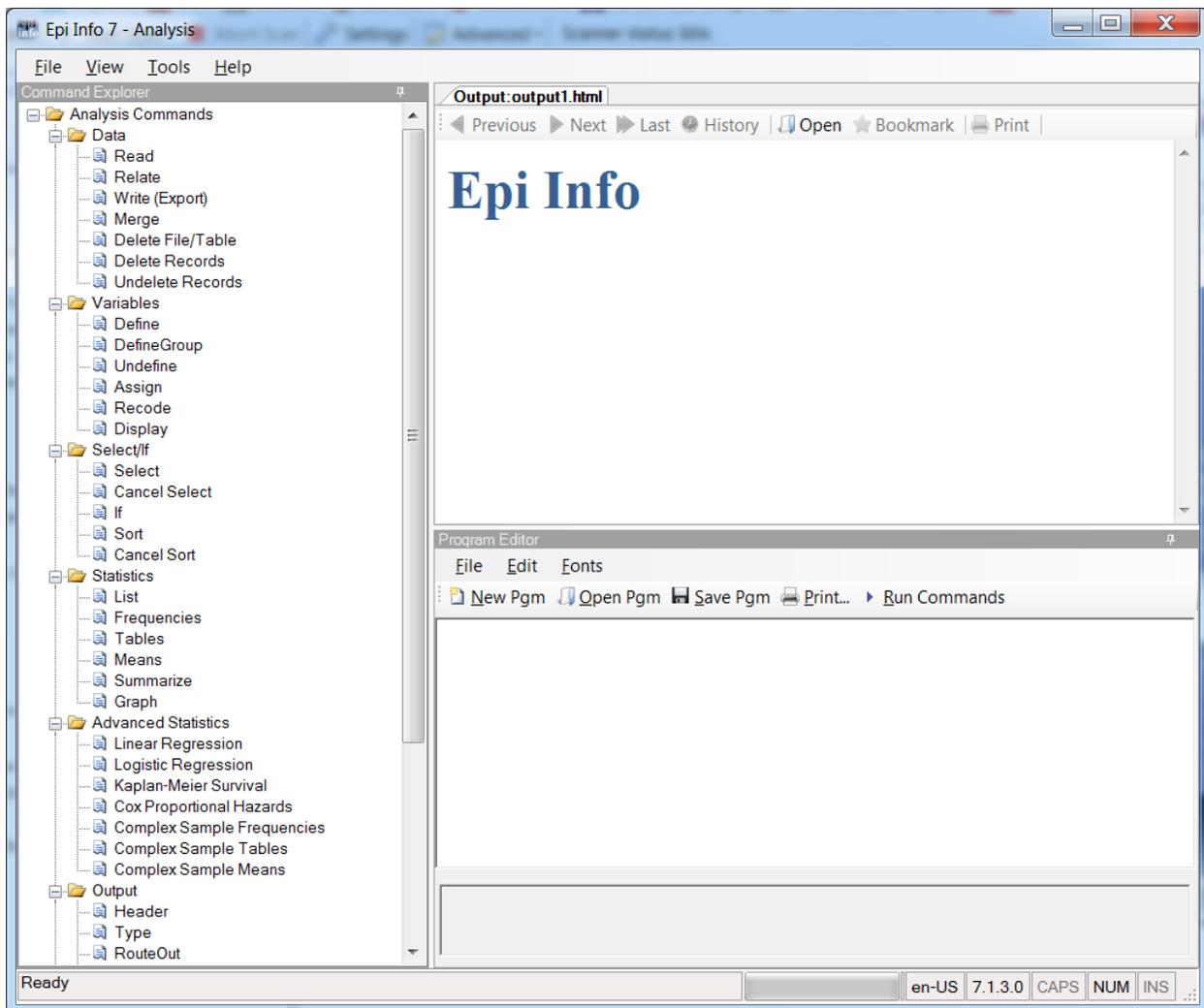
Le bouton **Démarrer**, raccourci **Epi Info™ 7** ou **Démarrer, All Programs, CDC, Epi Info 7, Epi Info™ 7**

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).



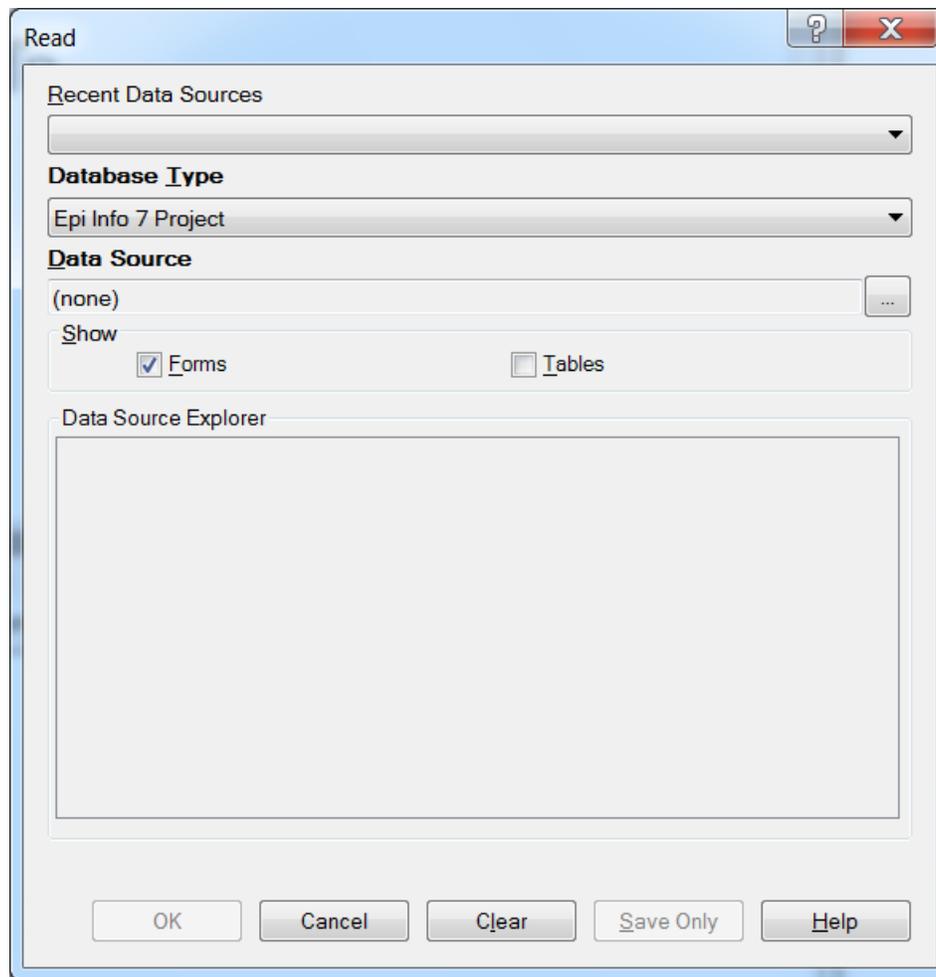
La fenêtre pour l'analyse des données:



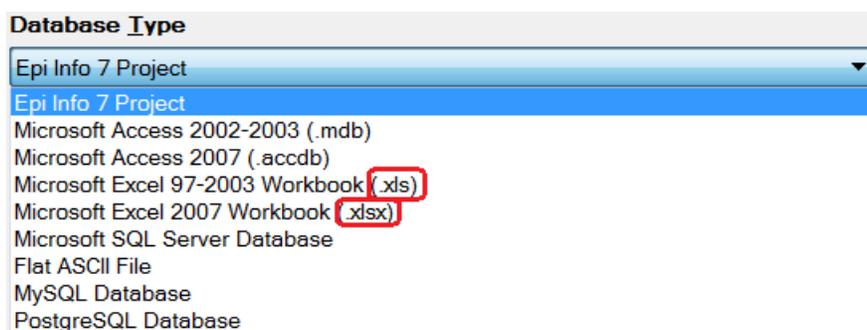
L'Importation d'un fichier Excel dans Epi Info

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.

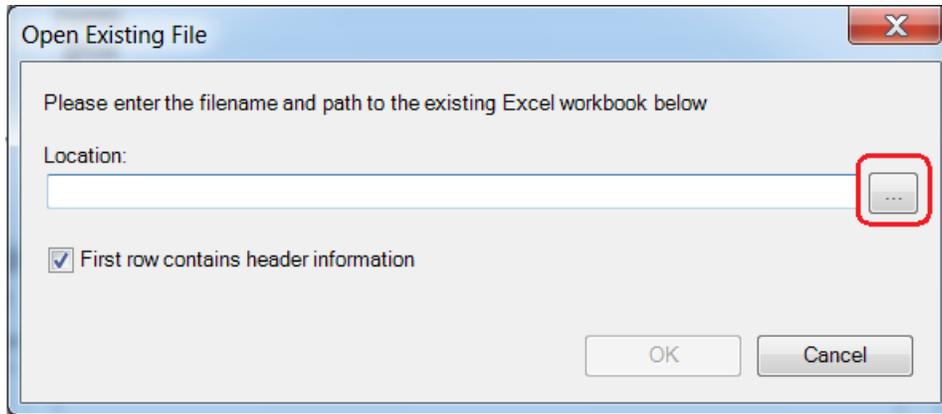
Dans la section des commandes (**Command Explorer** - le côté gauche de la fenêtre) on choisit dans **Analysis Commands**, la section des données : **Data** - la commande **Read** (Lire). Ouvrez la fenêtre d'importation des données. Choisissez le type de fichier à importer.



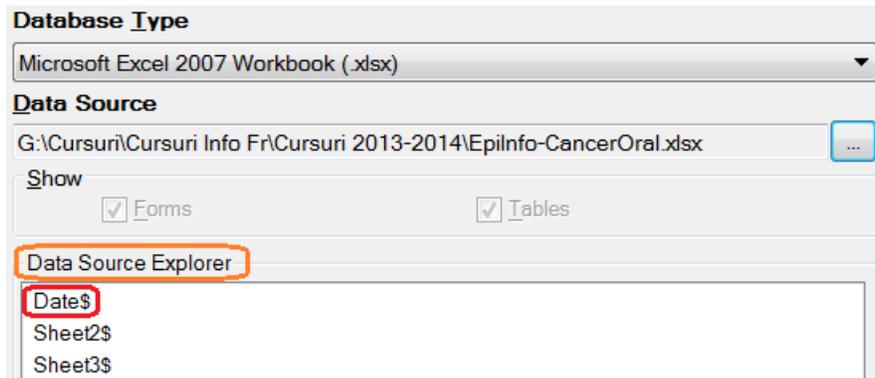
Si le fichier **Excel** a l'extension **.xlsx**, on choisit **Microsoft Excel 2007 Workbook (.xlsx)**, si le fichier a l'extension **.xls**, on choisit **Microsoft Excel 97-2003 Workbook (.xls)**.



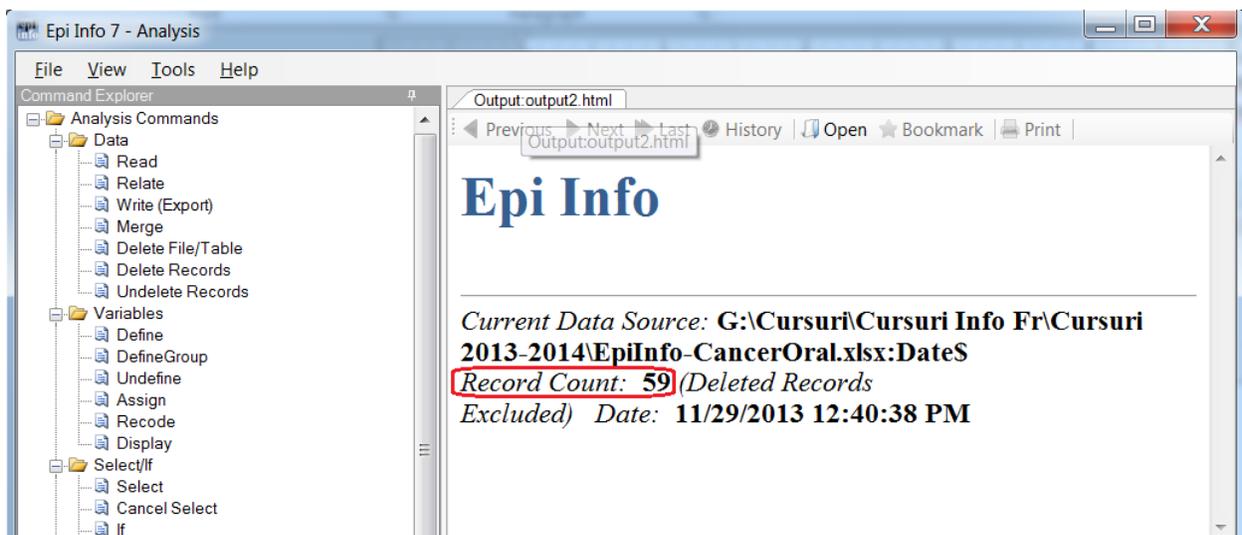
Après on choisit le fichier d'intérêt pour importation en appuyant le bouton  a cote de **Data Source**. Dans la fenêtre qui s'ouvre appuyez le bouton  a cote de **Location** (l'emplacement des fichiers), et cherchez le fichier d'intérêt. Après on a sélectionné le fichier on peut faire click sur le bouton **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.

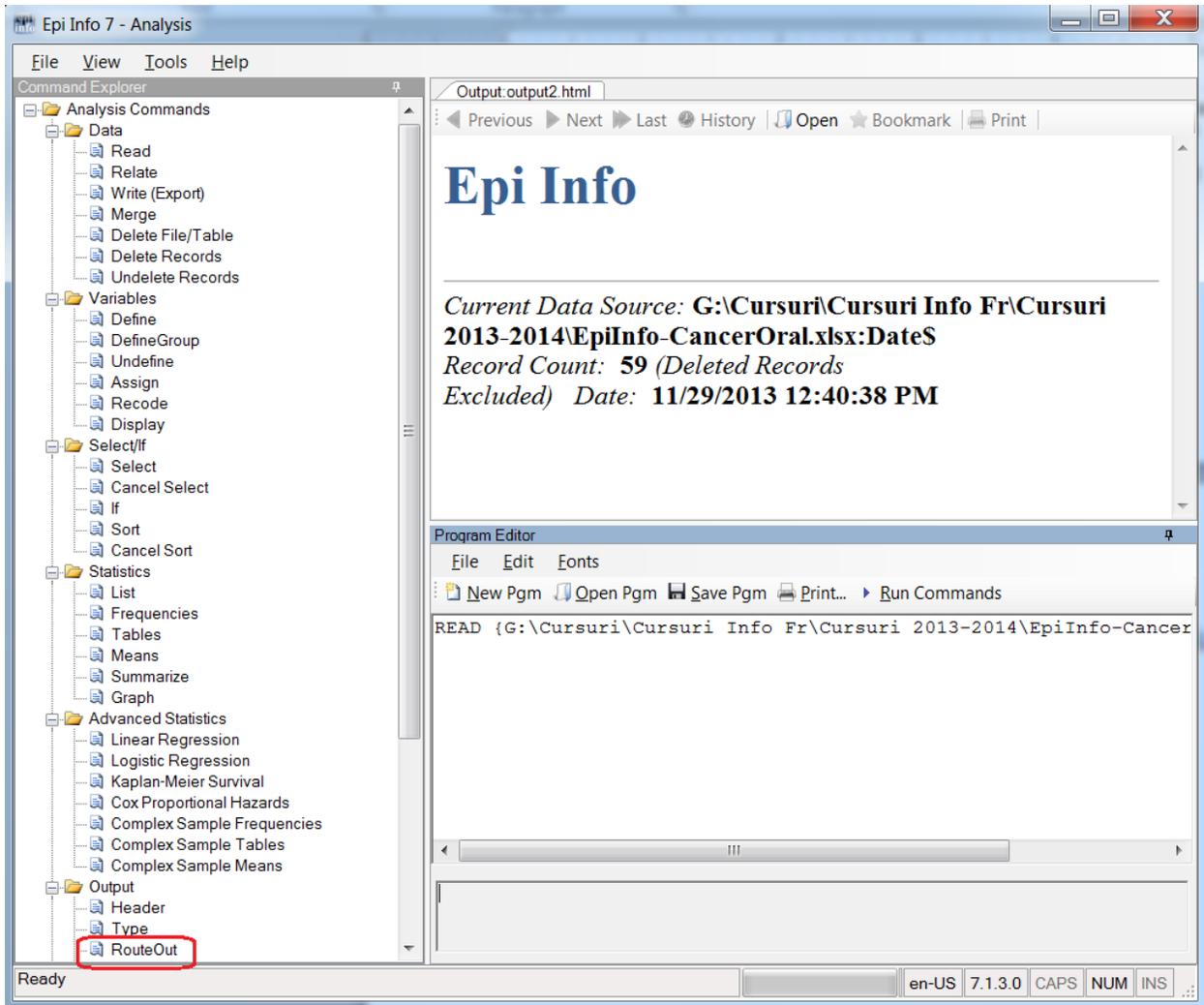


Après l'importation, Epi Info montre dans la zone des résultats (**Output**), le nom du fichier importée, et le nombre des enregistrements trouve dans le fichier (**Record Count**).

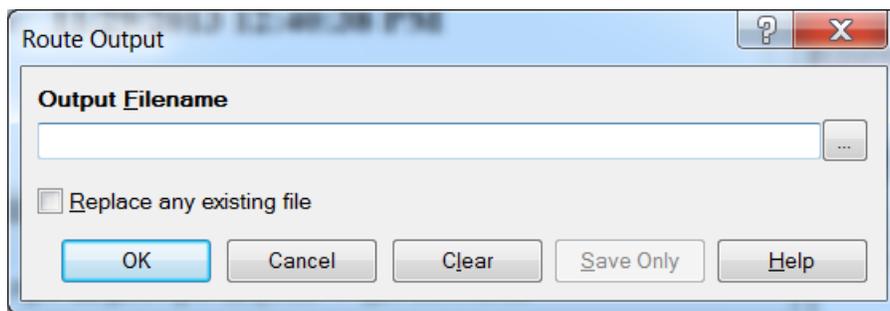


Sélectionnez le dossier ou les résultats de l'analyse seront enregistrée

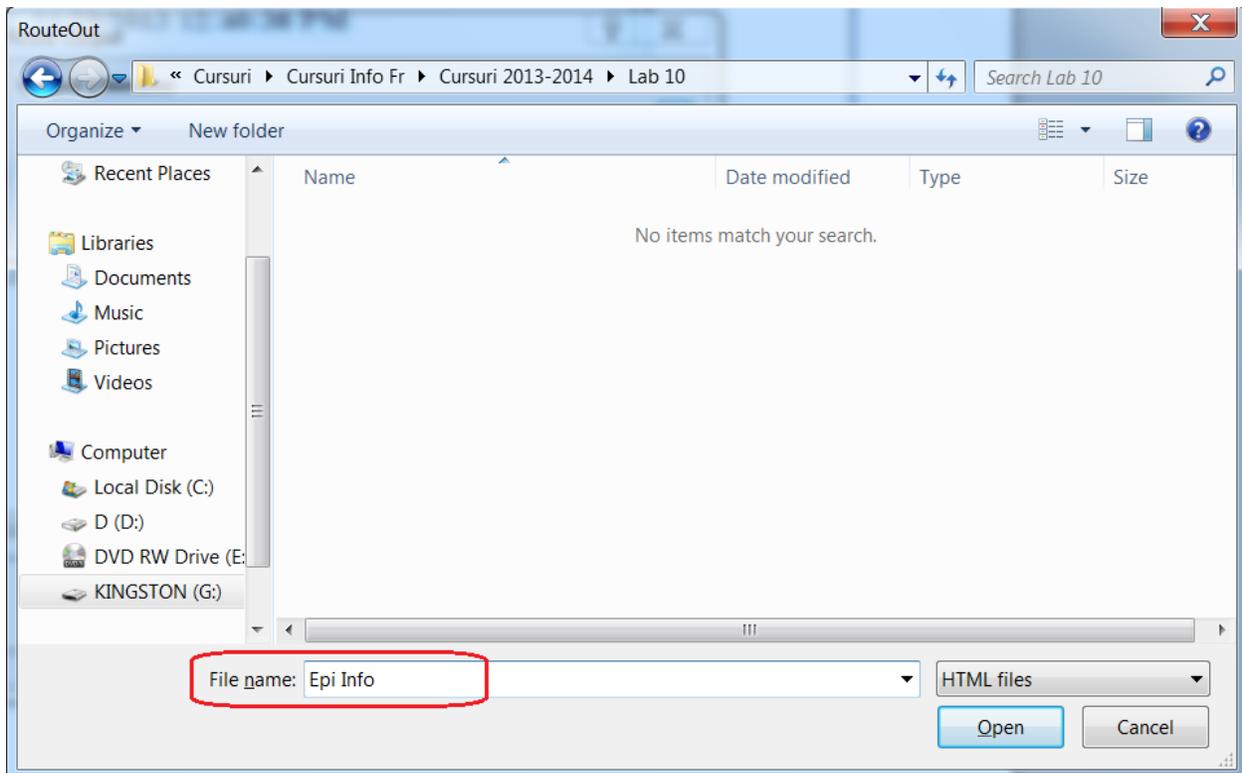
Optionnelle ! Dans la liste des commandes (**Command Explorer**) de la fenêtre de l'analyse de l'Epi Info, la section **Output** on choit **RouteOut**.



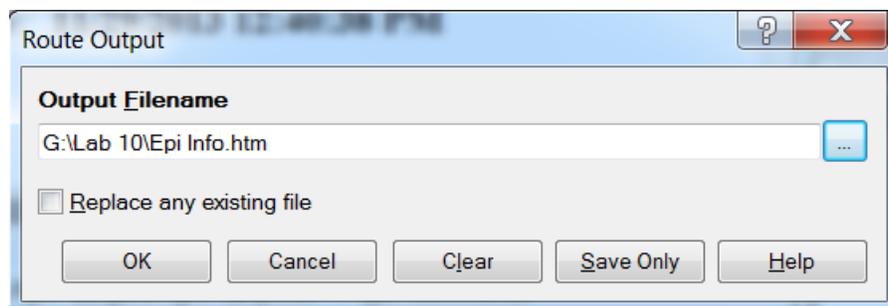
Dans la fenêtre qui s'ouvre cliquez sur le bouton  a cote de **Output Filename** (NOME du fichier des résultats)



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**.



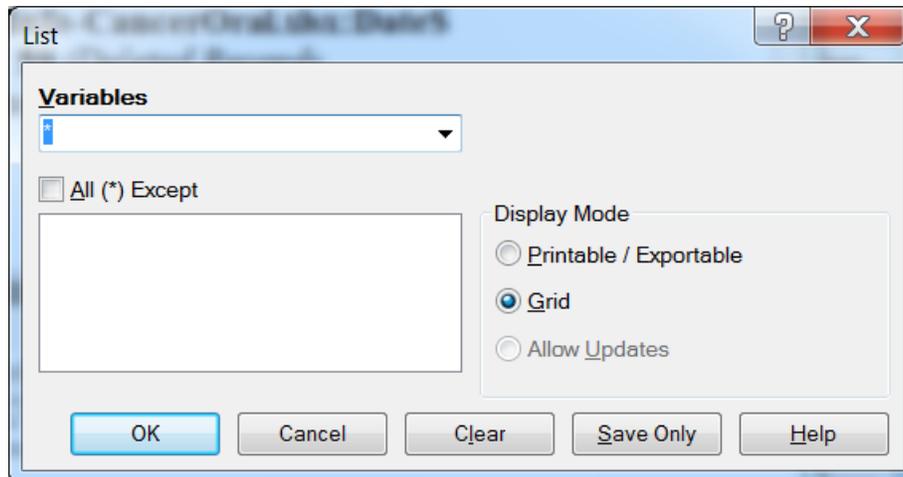
La fenêtre **Route Output** va montrer le nom de fichier et le dossier où l'analyse statistique sera enregistrée. Cliquez sur le bouton **OK**.



Montrer les données

Pour voir le contenu du fichier importé, dans la liste des commandes (**Command Explorer**) de la fenêtre d'analyse de l'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**.

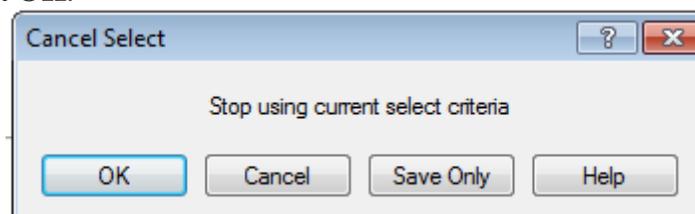


On vérifie si les données ont été bien importée. Puis on peut fermer la fenêtre.

	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:

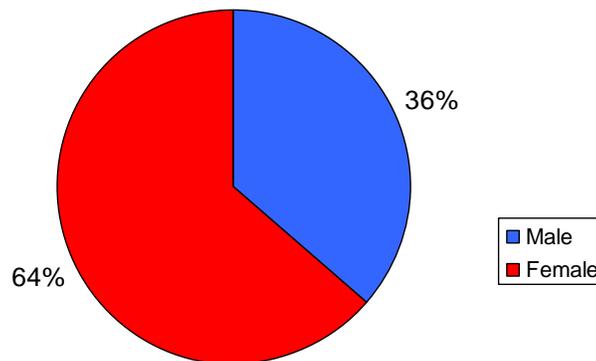


Figure 1. Répartition par sexe dans l'échantillon étudié

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?).

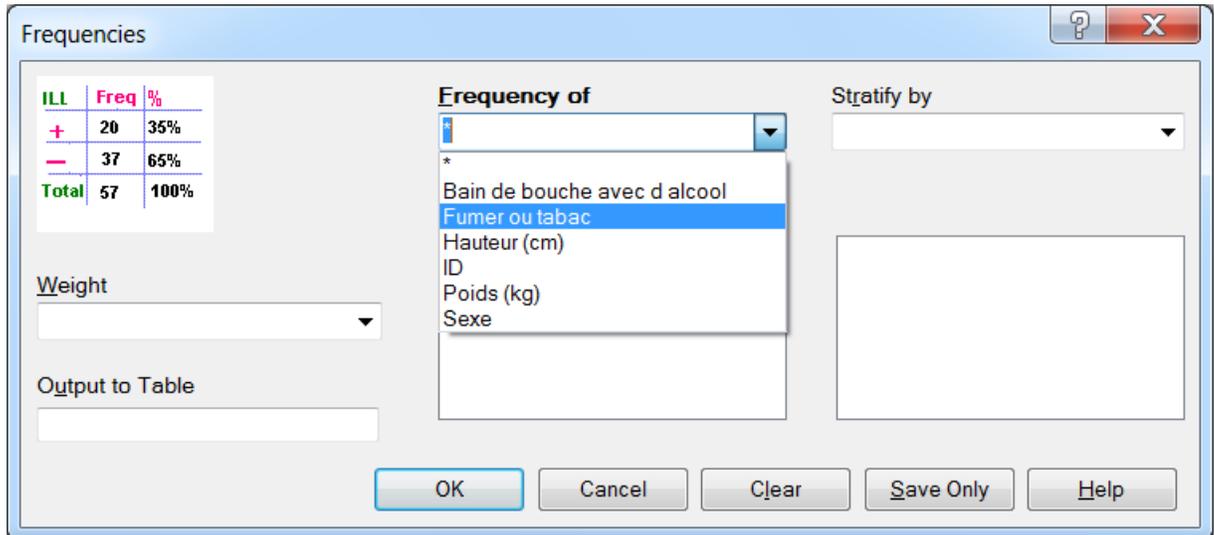
Si vous prévoyez d'utiliser le tableau comme une figure dans un document **Word**, effacez le titre du graphique dans **Excel**, mais n'oubliez pas d'étiqueter le graphique dans **Microsoft Word** comme vu dans l'exemple précédent: sélectionnez la figure, faites un clic droit sur cette sélection et utilisez **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**) du 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%	<div style="width: 23.73%; height: 10px; background-color: yellow;"></div>
M	45	76.27%	100.00%	<div style="width: 76.27%; height: 10px; background-color: yellow;"></div>
Total	59	100.00%	100.00%	<div style="width: 100%; height: 10px; background-color: orange;"></div>

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

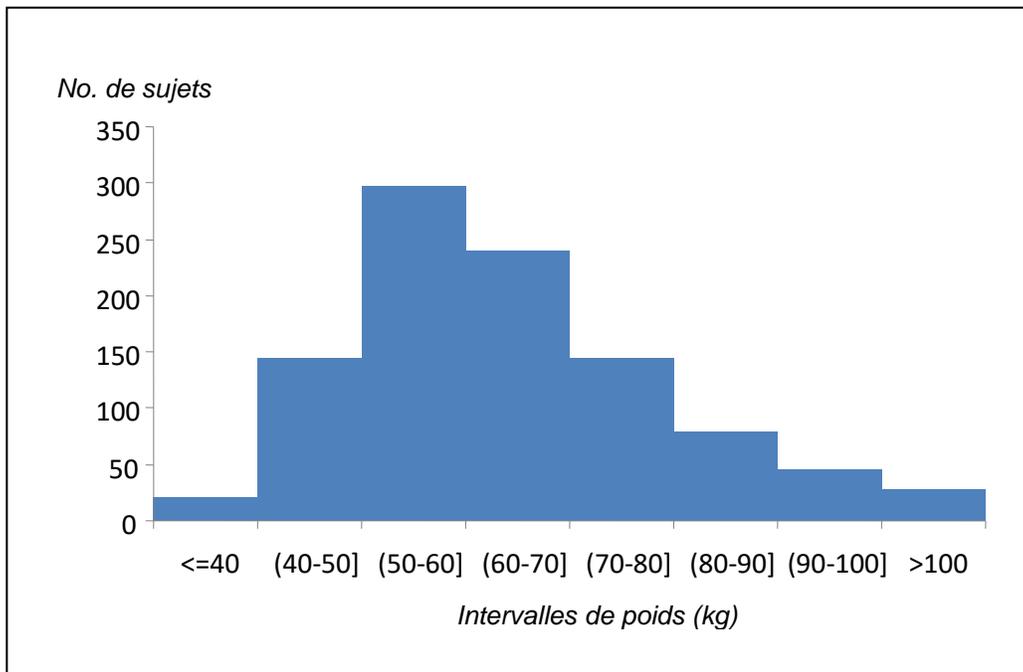


Figure 2. Histogramme des poids dans l'échantillon étudié

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:

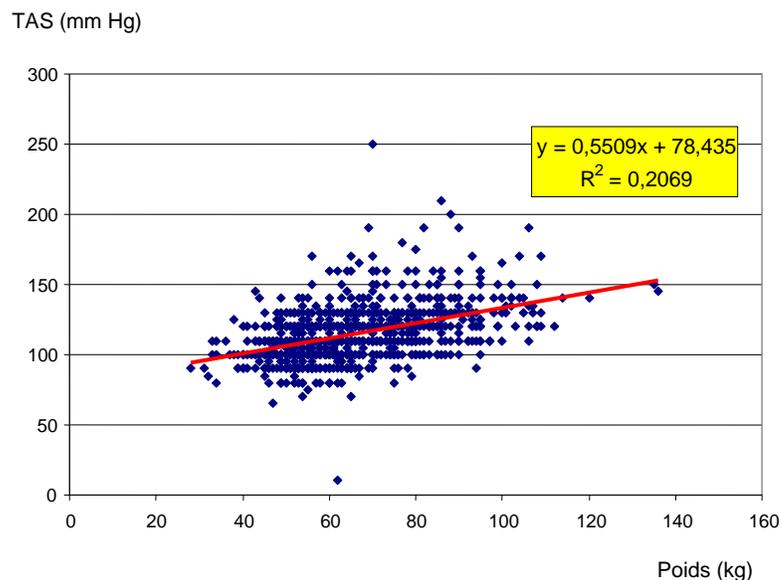


Figure 3. Relation entre le poids et la pression artérielle systolique pour les sujets inclus dans l'échantillon étudié

Données de survie:

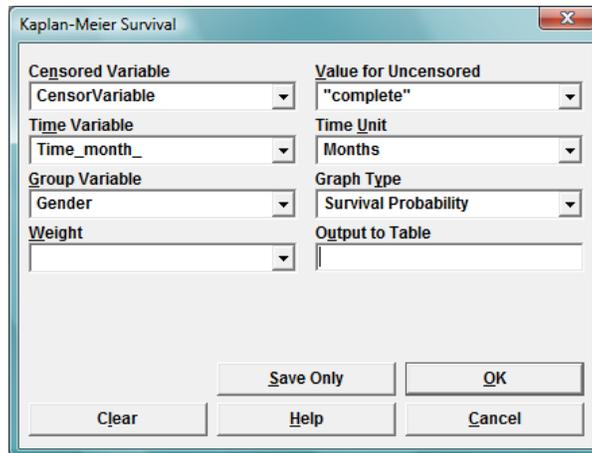
Médiane du temps de survie

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

Buttons: Clear, Save Only, Help, OK, 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.

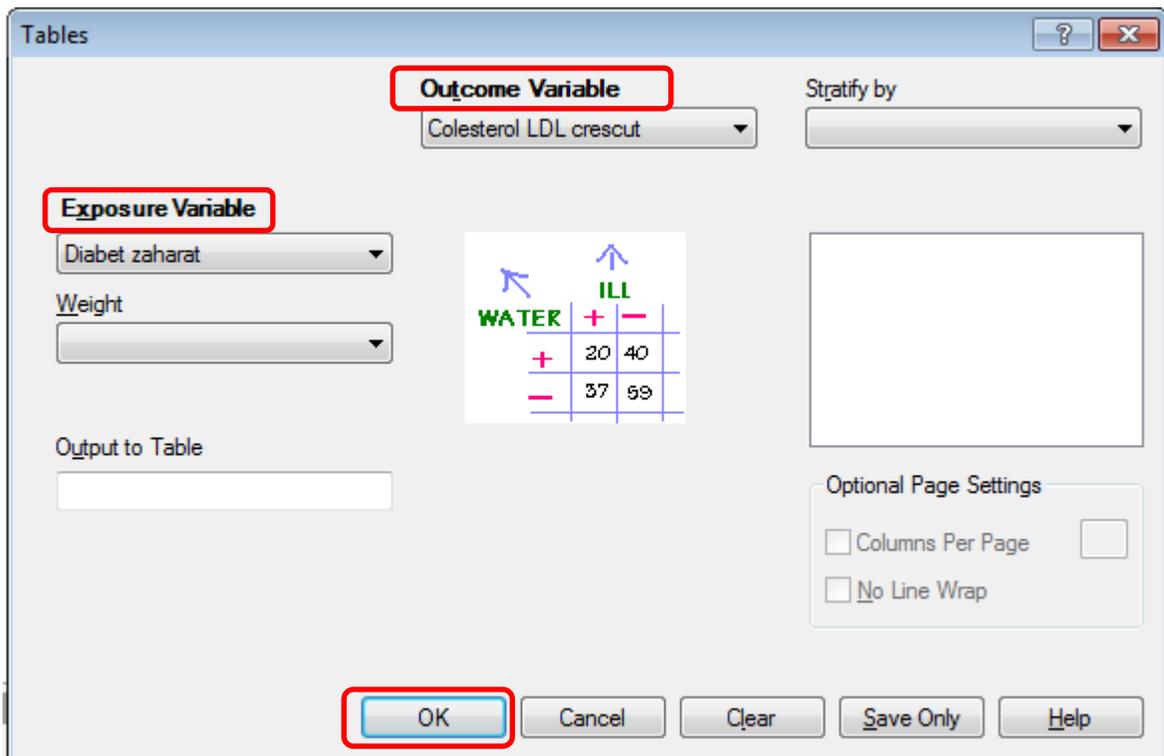
Si la valeur de p inclut la lettre E suivie d'un chiffre négatif, cela signifie en réalité une très faible (i.e. très significative) valeur de p (par ex. $p = 3,22342E-6 = 3,22342 \times 10^{-6} = 0,0000032234$).

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)

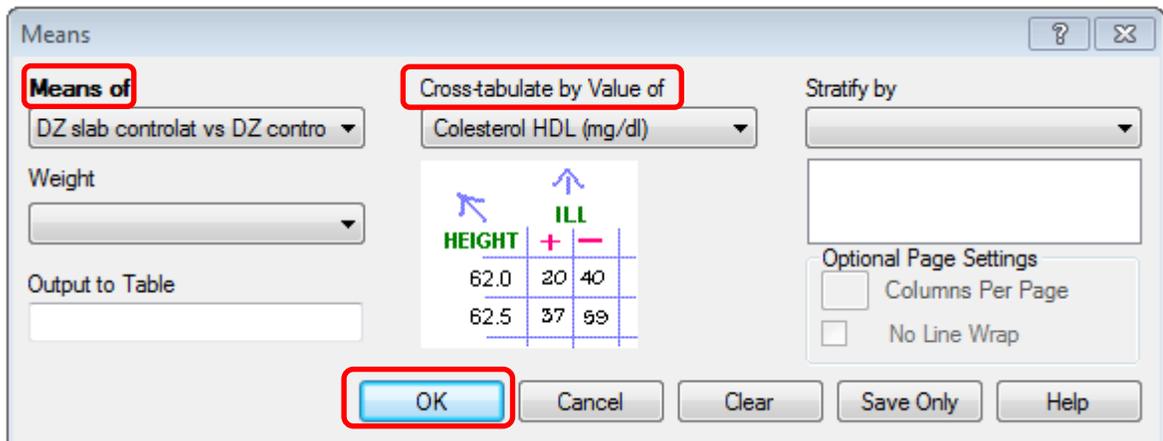
3. Des tests statistiques (on utilise le p bidirectionnel – 2-tailed p):
- Le test chi carre non corrigée (Chi-square – uncorrected)

- 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

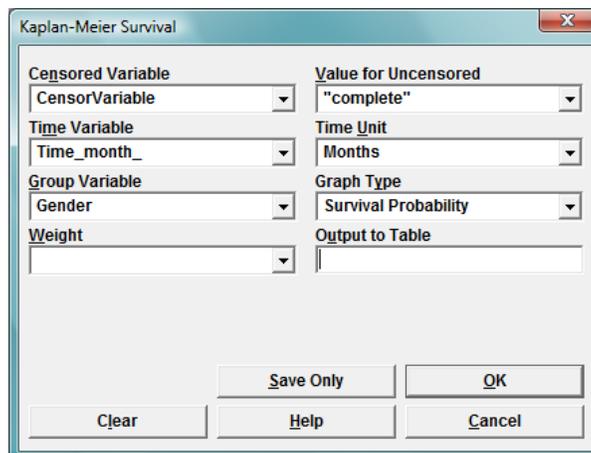
Degrees of freedom = 1

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:



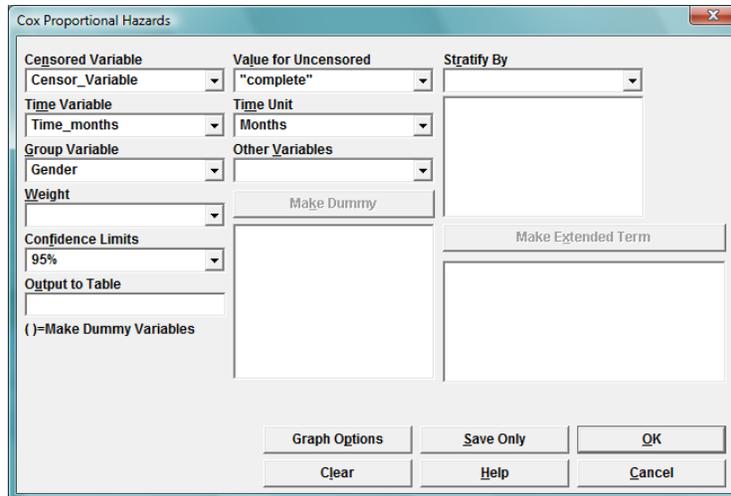
Changez **Group Variable** comme nécessaire pour votre comparaison.

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.

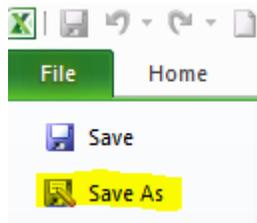
Pour **Input X Range** sélectionnez la plage de cellules contenant les variables indépendantes (x_1, x_2, \dots, x_n) de votre échantillon, ceux que vous souhaitez utiliser afin de prédire la variable dépendante (**y**) en utilisant une régression linéaire multiple. Si les variables indépendantes ne 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.

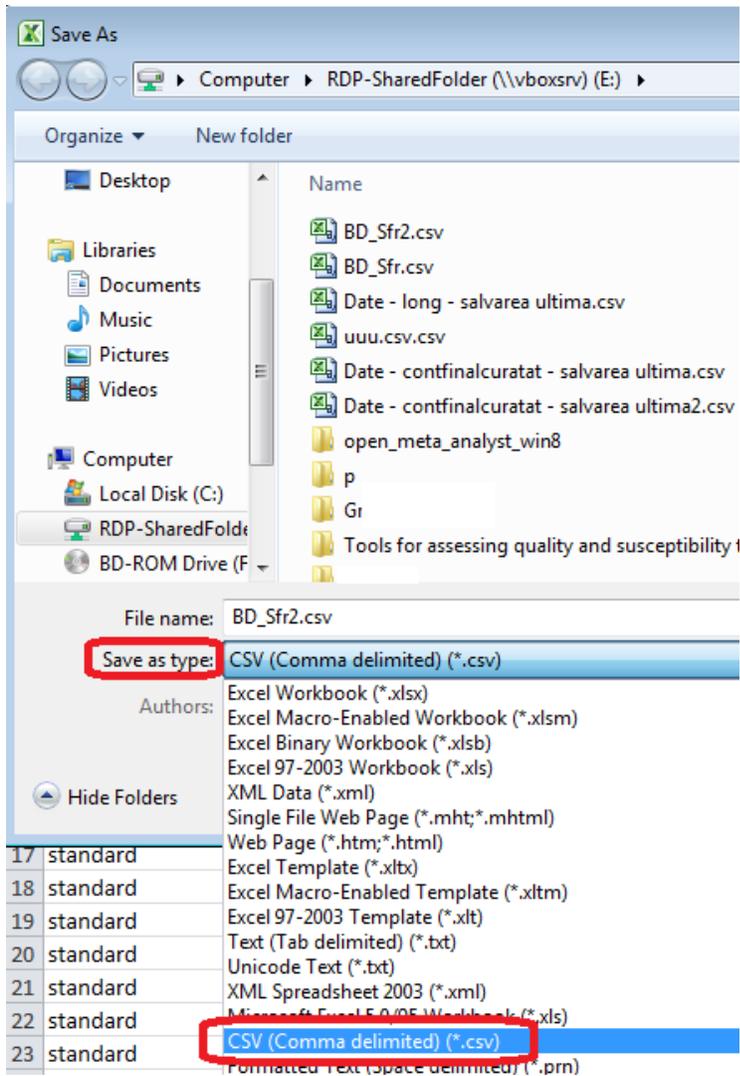
Réaliser un graphique Kaplan-Meyer et un test Log-rank pour l'analyse de la survie sur Internet

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**

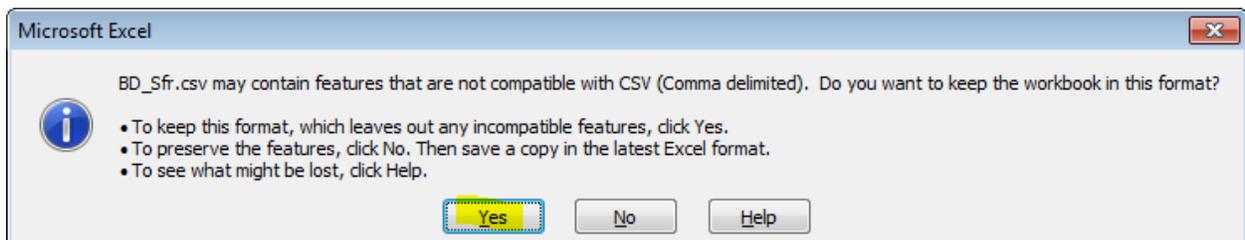


On va choisir dans: **Save as type: CSV (Comma delimited) (*.csv)**



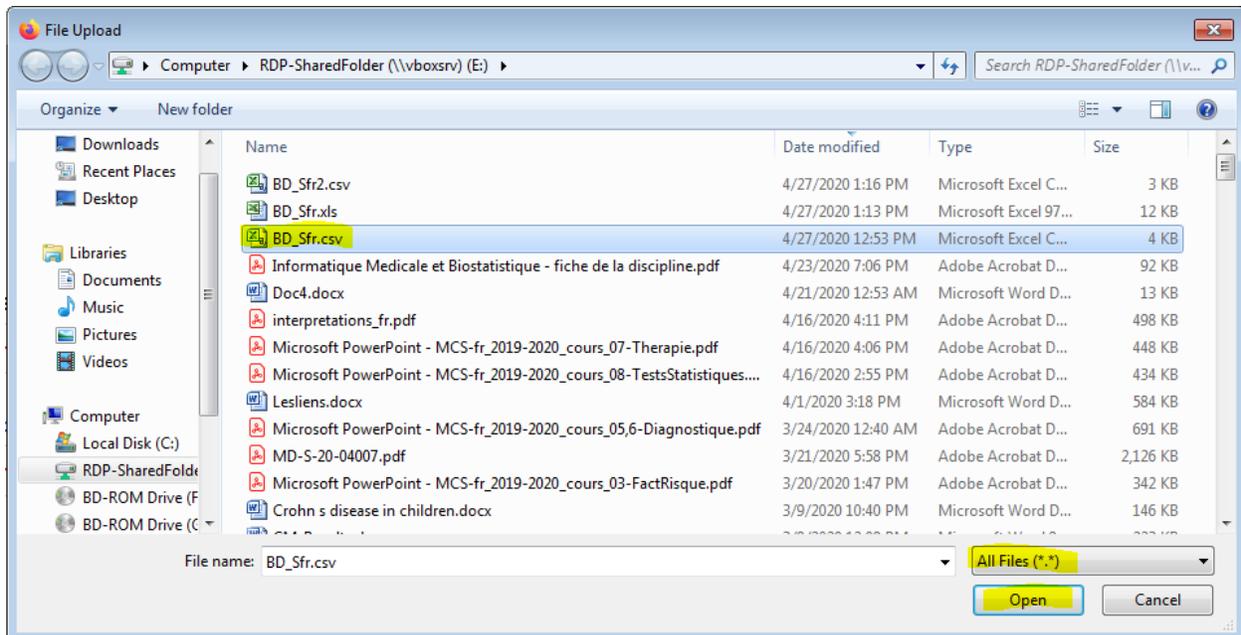
Puis on va faire clic sur le bouton **Save**.

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 inférieure 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 remplacer 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

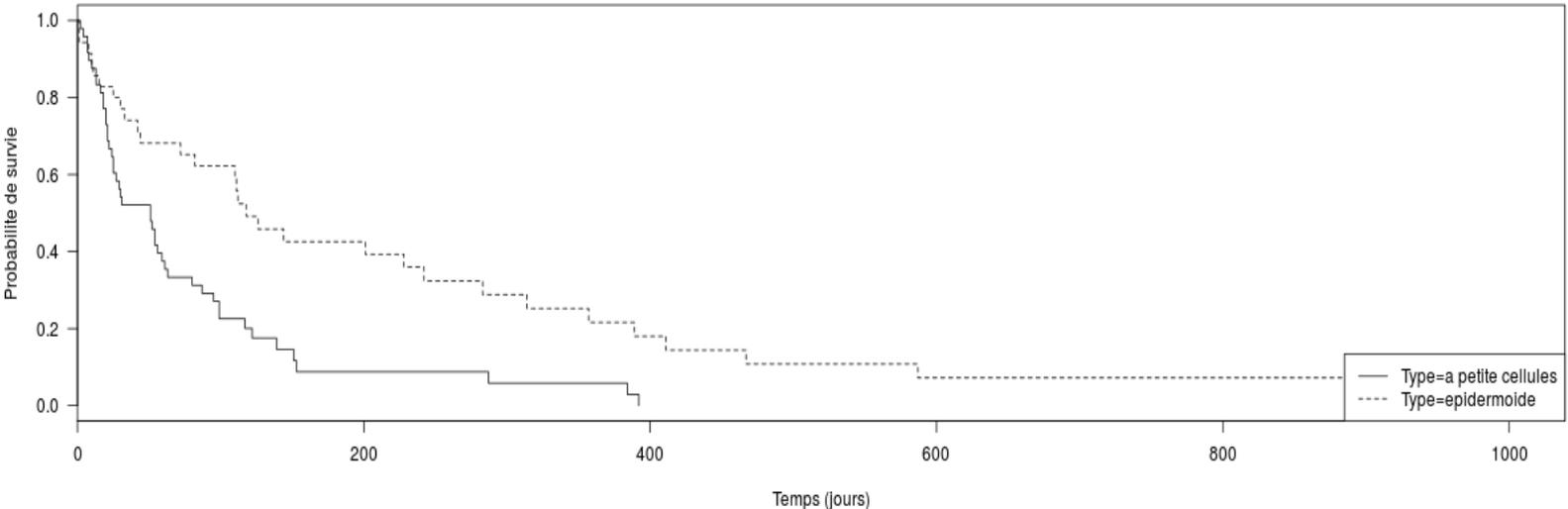
Legend Position

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

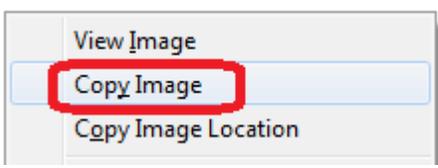
A la fin, on va fair **clie** sur le bouton **Run Survival Analysis**

Run Survival Analysis

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

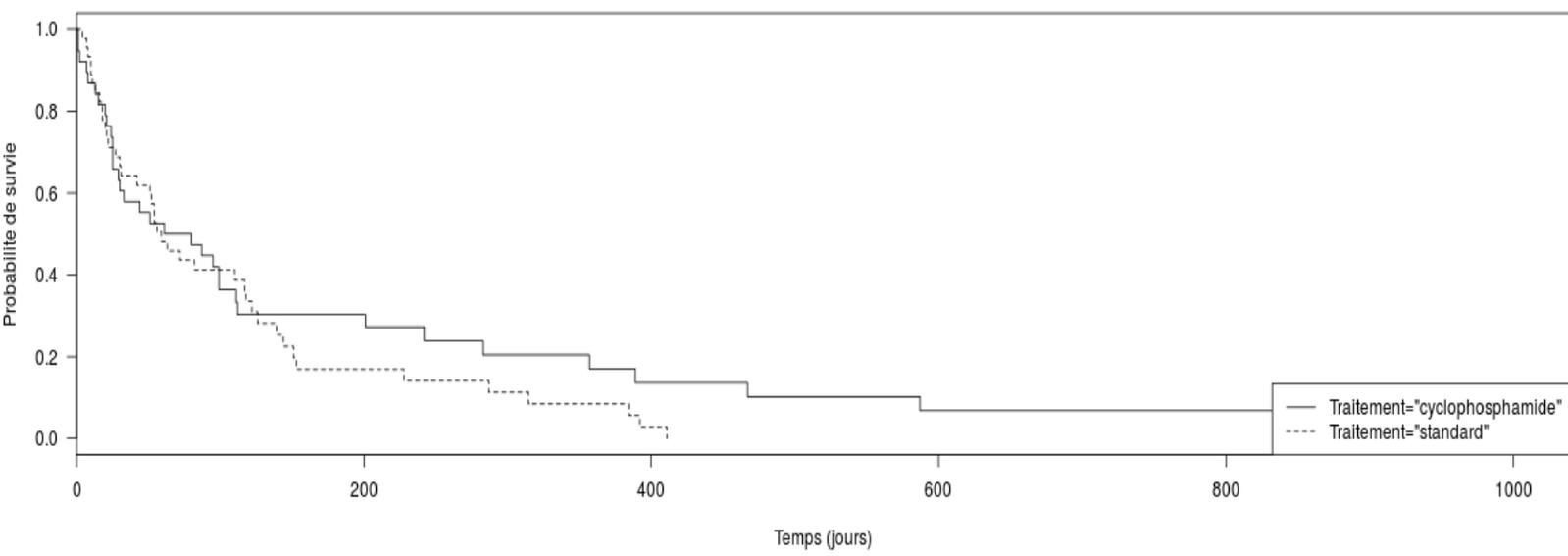


On peut **copier** le graphique avec un **clik droit** du souris sur l'image, en utilisant l'option **Copy Image**



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://merse.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
```