Epi-Σûre
La statistique et l'épidémiologie par l'exemple
Recueil et analyse de données

jamovi Stats. Open. Now.

Tutoriel pour les épidémiologistes débutants

Bernard Branger, 2026, NANTES

https://www.episure.fr/

Introduction

Les sciences de la santé ont besoin d'analyse de données fiables, diverses et facilement accessibles pour des personnes non spécialisées. L'appel à des épidémiologistes et des statisticiens confirmés n'est pas toujours possible, et, dans tous les cas, nécessite de bien connaître ses propres données et de se cantonner à ses objectifs.

Le présent tutoriel s'adresse aux médecins, pharmaciens, sages-femmes, dentistes, infirmiers… qui recueillent des données pour un travail de mémoire ou de recherche, qui ont eu quelques bases (d'ailleurs souvent oubliées) en statistiques et qui veulent s'accaparer leurs données en vue de l'écriture et de la présentation de leur travail.

Nous n'envisagerons pas ici la partie "recueil" et "saisie" des données évoquées dans un texte à part. Nous ferons le postulat que les données sont disponibles en format tableur comme Excel® ou d'autres logiciels comme Libre Office Calc®, sur tout support Windows, Mac ou Linux.

jamovi® est utilisé ici car il est performant, gratuit, disponible pour Windows, Mac, Linux et Chrome OS et répond aux nécessités des travaux statistiques pour un mémoire ou une thèse habituelle. D'autres logiciels existent (voir www.episure.fr), gratuits comme "R" qui est d'abord (très) difficile, ou EpiData® plus ancien ou **Epi Info™ du CDC** sous Windows. D'autres sont "cliquables" avec ou sans lignes de commande ou syntaxe mais souvent payants comme Stata®, SPSS® d'IBM, MedCalc®.

Le présent tutoriel concerne **jamovi** téléchargé en octobre 2026. Des techniques statistiques plus élaborées nécessitent de toute manière l'aide d'un statisticien qui utilisera son logiciel propre

Télécharger Jamovi (toutes plates-formes)

Dans un moteur de recherche, chercher https://www.jamovi.org/download.html, choisir son support Windows, MacOS, Linux ou même ChromeOS. Pour Windows, télécharger un fichier *.msix ou *.exe[1] pour Windows ("Solid" 2.7.38 ou "Current" 28.4.0), ou le format pour les autres supports.

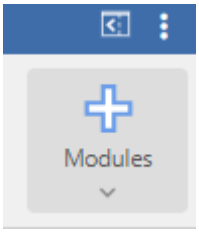
Cliquer sur le fichier téléchargé et suivre l'installation proposée. Sur PC, le fichier qui lancera **jamovi** est `jamovi.exe` par exemple dans ce type de chemin sous Windows : `C:\Program Files\jamovi 2.7.16.0\bin\jamovi.exe`

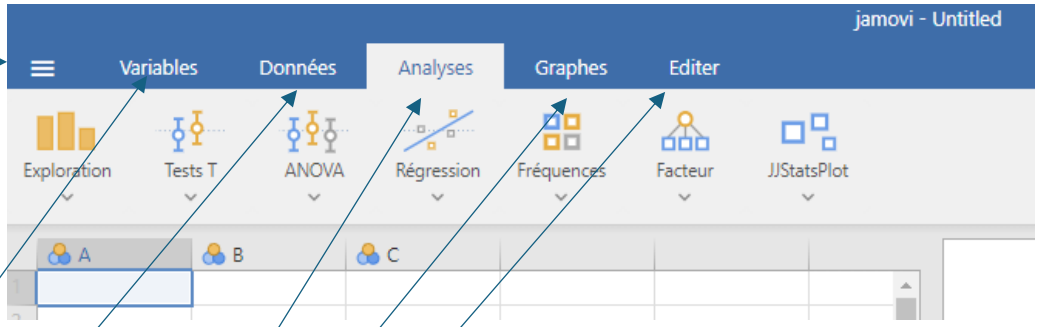
Sur certains ordinateurs Windows, il se peut que les téléchargements soient bloqués : il s'agit du mode S de Windows. Pour désactiver le mode S de Windows (qui restreint l'installation aux applications du Microsoft Store), ouvrez les Paramètres Windows via le menu Démarrer ou le raccourci Touche Windows + I, puis allez dans Système > Activation et cliquez sur Aller au Store sous la section « Mode S » avant de sélectionner `Obtenir`

Ce tutoriel n'est pas un cours de statistique ou d'épidémiologie. Il est destiné à utiliser un logiciel. Des options et des analyses supplémentaires sont disponibles dans **jamovi** *et ne sont pas évoquées ici comme certaines conditions d'utilisation ou de vérification d'hypothèse ou d'adéquation de modèles (en analyse multivariée en particulier).*

[1] Les fichiers ont tous un "format" avec un nom et une extension après un point comme *.xslx pour Excel ou *.docx pour Word ou *.pptx pour PowerPoint. Elles n'apparaissent pas toujours dans les panneaux de recherche, mais peuvent être visibles dans tous les supports. Pour Windows 11, Explorateur */ Afficher / Afficher / Extension de noms de ficher.*

L'ouverture permet de visualiser la partie supérieure de la fenêtre du logiciel avec ses **onglets ou modules** qui, déroulés, montrent des menus pour chaque onglet. Ici, le menu de l'onglet "Analyses" :



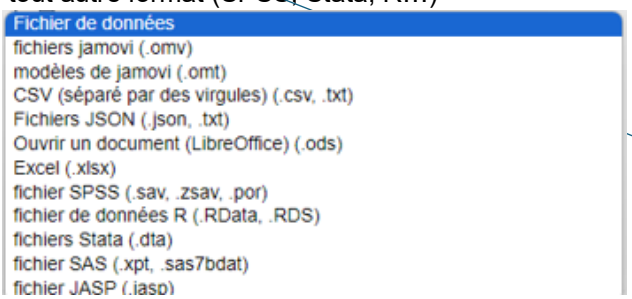


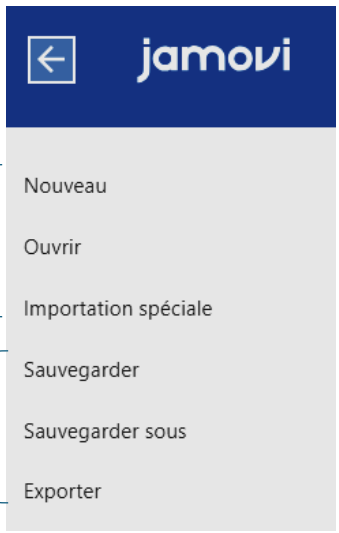
- **Maniement des variables** : nom, catégories, codages, transformations, calculs…
- **Données** : aspect des données sous formes de tableur avec colonnes pour les variables et lignes pour les observations
- **Analyses statistiques avec (ici) 7 modules** : il en existe plusieurs dizaines d'autres : Voir la croix + en haut à droite
- **Graphes divers** à construire et à exporter éventuellement pour être modifiés
- **Traitement de texte** pour modifier les sorties des analyses
- **Onglet d'import/ ouverture et sauvegarde** des fichiers de données, qui se déroule ainsi :

Modules par défaut ou téléchargés

= Ouvrir un fichier de l'ordinateur : le plus souvent un fichier de tableur. On peut aussi importer un fichier texte (en *.csv par exemple), ou un fichier SPSS ou "R"… On peut aussi réaliser une saisie directement.

On peut sauvegarder ou exporter les fichiers au format **jamovi** (*.omv), ou selon tout autre format (SPSS, Stata, R…)





Le fichier d'exemple

Ce sont de vraies données concernant des **déclenchements d'accouchements pour des motifs maternels ou fœtaux**. Certains accouchements se terminent par "voie basse" (VB) (voies naturelles, dites souvent « succès ») ou par césarienne (CS) (dite souvent « échec »). Ce ne sont que des singletons. L'objectif de l'étude est de déterminer les facteurs associés ou causaux qui conduisent à la césarienne. Il y a 903 observations et 59 variables. Les résultats ont été publiés (en français). Voir *Branger B, Dochez V, Gervier S, Winer N. [Cesarean after labor induction: Risk factors and prediction score]. Césarienne après déclenchement : facteurs de risques et score de prédiction. Gynecol Obstet Fertil Senol. 2018 May;46(5):458-465. doi: 10.1016/j.gofs.2018.03.008.*

Pour importer les fichiers d'exemple : ouvrir le site Internet de Episure : https://www.episure.fr/formations, cliquer sur ***Fichiers sur les déclenchements à télécharger (fichier ZIP) et*** copier les fichiers d'Excel en *.xlsx (à dézipper) vers un endroit, sur votre ordi, que vous retrouverez facilement (Bureau par exemple) et le **dézipper** dans un endroit également facilement trouvable. **Le présent tutoriel concerne essentiellement le fichier decl_temoins.xslx, et dans quelques cas le fichier declenchement_2025.xslx ou decl_cohorte.xslx.**

Questionnaire de recueil des informations sur les déclenchements

N° Fiche …….. N° Accouchement …….

Issue : 1. Cas (CS) ☐ 2. VB témoins ☐ 3. VB cohorte (cahier acc) ☐

Date accouchement (dd/mm/yyyy) …… Heure (hh/mn) ……

Gestité (0 pour nulligeste) …. Parité (1 pour primipare) ……

Date naissance mère …….. Age mère (par différence ou directement) ……..

Date prévue d'accouchement …… Age gestationnel (par différence ou directement) : …..SA + .. Jours

Nombre de fœtus : …… Présentation : 1. Céphalique ☐ 2. Siège ☐

Accouchement

Mode de déclenchement : 1. Maturation ☐ 2. Déclenchement ☐ 3. Les deux ☐

Motif de déclenchement : 1. Grossesse prolongée ☐ 2. Fœtal ☐ 3. Maternel ☐
4. Obstétrical ☐ 5. Plusieurs motifs ☐ 6. Convenance ☐

Motif fœtal : 1. ARCF ☐ 2. Patho foetale ☐ 3. Gémellaire ☐ 4. Baisse des MAF ☐
5. RCIU/PAG ☐ 6. Macrosomie ☐ 7. Cassure courbe croissance ☐ 8. MFIU ☐

Motif maternel : 1. HTA ☐ 2. Pathologie ☐ 3. Thrombopénie ☐ 4. Fenêtre thérapeutique ☐ 5. Métro ☐

Motif obstétrical : 1. RPM ☐ 2. Cholestase ☐ 3. Diabète ☐ 4. Allo-imm ☐ 5. Chorio-amniotite ☐
6. Oligamios ☐ 7. Hydramnios ☐ 8. Doppler anomalies ☐

Motifs autres …… RPDE > 12 h : 1. Oui ☐ 2. Non ☐

Mode accouchement : 1. Voie basse ☐ 2. Instrumental ☐ 3. Césarienne ☐

Instrument : 1. Forceps ☐ 2. Ventouse ☐ 3. Spatule ☐ 4. Plusieurs ☐

Cause césarienne : 1. Itérative ☐ 2. Fœtale ☐ 3. Maternelle ☐
4. Obstétricale ☐ 5. Plusieurs causes ☐ 6. Convenance ☐

Nouveau-Né

Poids de naissance (g)…….. Sexe : 1. Garçon ☐ 2. Fille ☐

Apgar 1 minute …… Apgar 5 minutes … pH artériel …… pH veineux …….

Compléments dans les dossiers

Utérus cicatriciel : 1. Oui ☐ 2. Non ☐

Poids mère …….. Taille mère …… IMC ……..

Prise de poids pendant la grossesse …… Bishop initial ……

[…]

Cause carienne en cours de travail : 1. ARCF ☐ 2. Stagnation ☐ 3. Echec ☐ 4. Non engagement ☐
5. ARCF + stagnation 6. Autre ☐

Date déclenche …….. Heure déclenche ……

(calcul de délai déclenche = heure acc - heure déclenche (avec correction en cas de franchissement de minuit)

Méthode déclenchement codage de 1 à 14 (pas mis ici)

Valider les données

On peut sauter cette partie si le fichier est contrôlé et la maitrise des variables affirmée.

Passer page 10 Décrire la population, décrire les variables

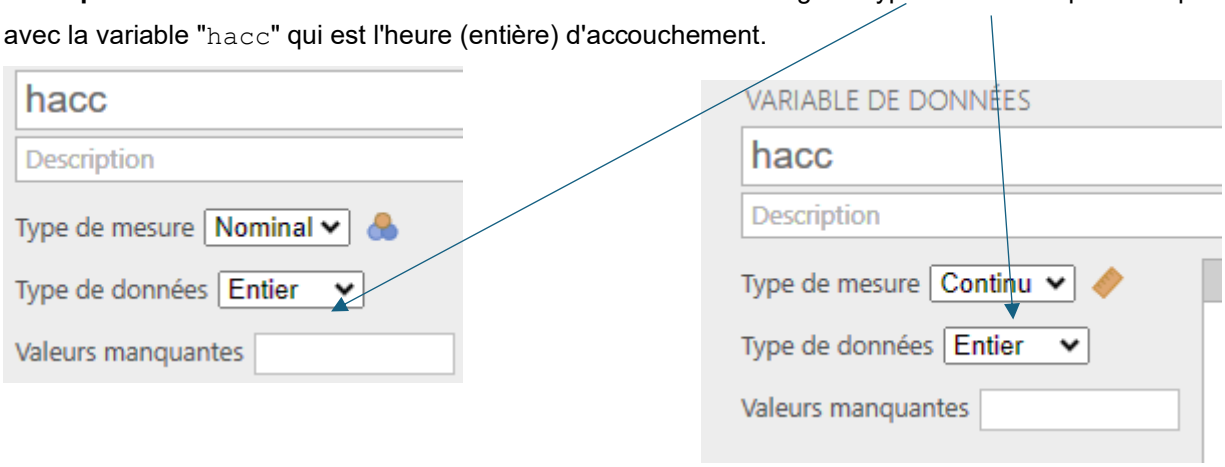
Valider les données c'est comprendre les variables et maitriser toute l'information qu'elles contiennent. C'est une étape indispensable, fastidieuse sur les gros fichiers. Fichier `decl_temoins.xlsx`

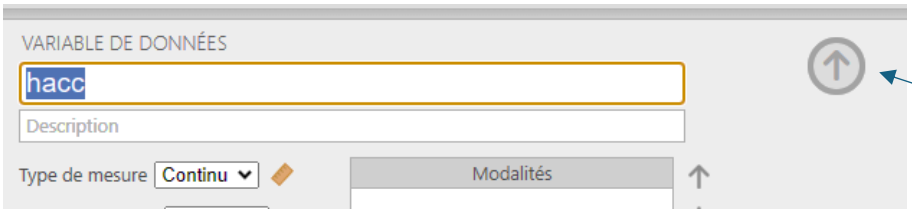
- Formats des variables

En épidémiologie, on distingue les catégories (ou formats) de variables suivantes (avec des noms qui varient selon les livres et les logiciels) :

1. Les **variables nominales** dites aussi : "qualitatives", ou "catégorielles", ou "factor" sous "R", ou "codées" en nombres entiers. L'analyse en fera des dénombrements avec des effectifs et des fréquences, et des % appelés taux, pourcentages ou proportions…
2. Les **variables quantitatives** dites aussi continues, ou "d'échelle" ou "scale" ou "int" ou 'float' sous "R" avec des chiffres entiers ou avec virgules. Les cellules vides (données manquantes) ne doivent pas être mises à "0", mais réellement vides. L'analyse en fera des moyennes, médiantes et valeurs de dispersion comme l'écart-type, ou des classes avec un ou plusieurs seuils. Toujours vérifier les effectifs.
3. Les **variables ordinales** qui peuvent être traitées soit comme des variables nominales (avec des analyses de fréquences), soit comme des variables d'échelle (avec des moyennes). Comme exemple, on peut citer la parité en obstétrique qui peut être analysée en qualitatif avec la proportion de primipares ou en quantitatif avec la moyenne des enfants…. Ou encore un score de gravité de symptômes ou de maladies.
4. **Les variables autres** : variables de date (sujets délicats si le format n'est pas maitrisé à la saisie), variables de texte…

Lors de la validation, chaque variable (en colonnes) a un format retenu par **jamovi** : on peut changer **en cliquant sur le nom de la variable en tête de colonnes** et changer le type de mesure : par exemple avec la variable "`hacc`" qui est l'heure (entière) d'accouchement.

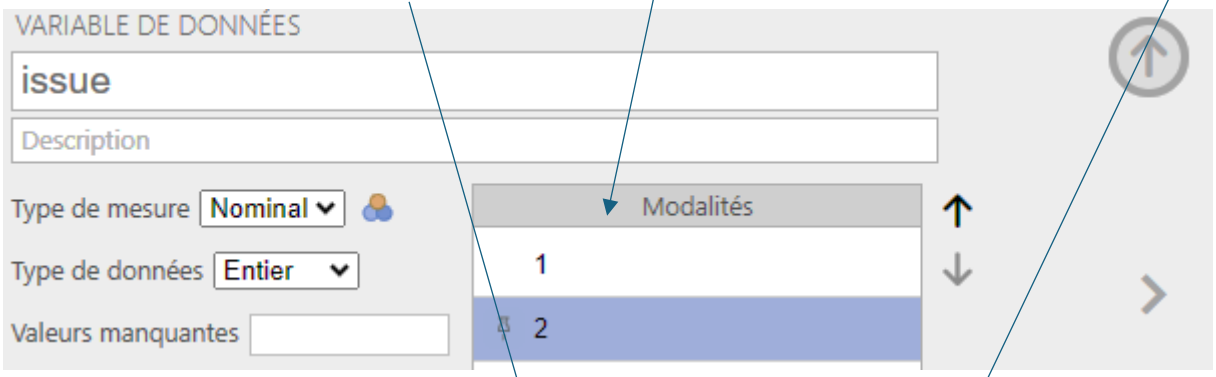


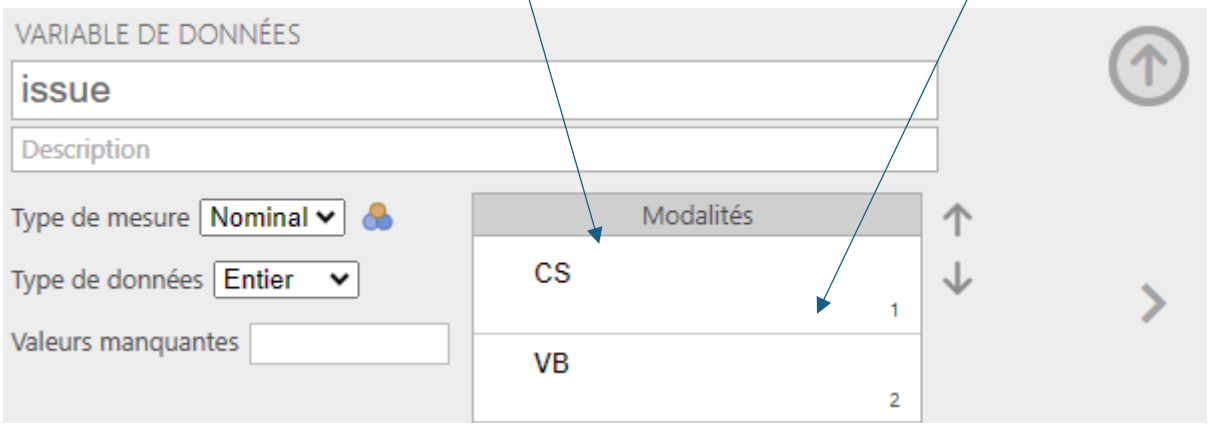


Pour revenir au tableur, cliquez sur la flèche :

- Donner des noms de labels aux variables nominales (codées)

La variable `Issue` a été saisie et codée 1 ou 2 ou 3. Ces codes correspondant à 1 = "Césarienne", 2 = "Voie basse témoin" et 3 = "Voie basse cohorte". On peut ainsi, pour ne pas se tromper, donner un "label" à chaque codage qui cependant sera traité en sous-main et 1/2/3. Ecrire "CS" à la place de "1"





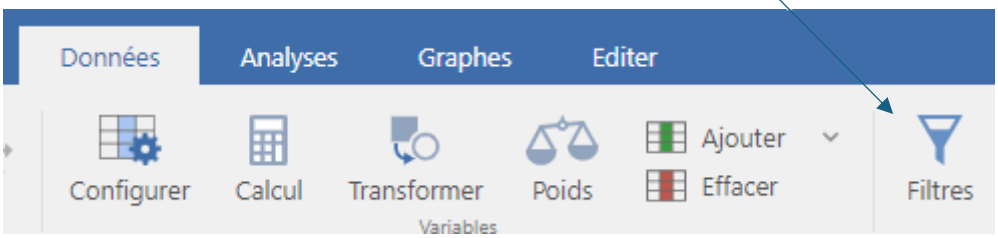
Revenir au tableur avec 

On n'a pas intérêt dans les fichiers longs à donner des noms de labels à toutes les variables nominales. Seulement les variables de jugement et les autres importantes ou complexes à coder.

Question importante : en cas de réponse binaire, lors de la saisie dans un tableur, ne pas coder "Yes" et "No", ni "Y" ou "N" car ce sont des lettres. Préférer deux possibilités : soit "1" pour oui et "2" pour non : ce sera plus commode pour les tableaux 2 x 2, soit "0" pour non et "1" pour oui : c'est ennuyeux pour les tableaux (mais rattrapable) et commode pour les analyses multivariées, mais qqfois les "0" disparaissent à l'import…. Garder le principe de 1 et 2 ; on peut toujours changer dans JAMOVI après.

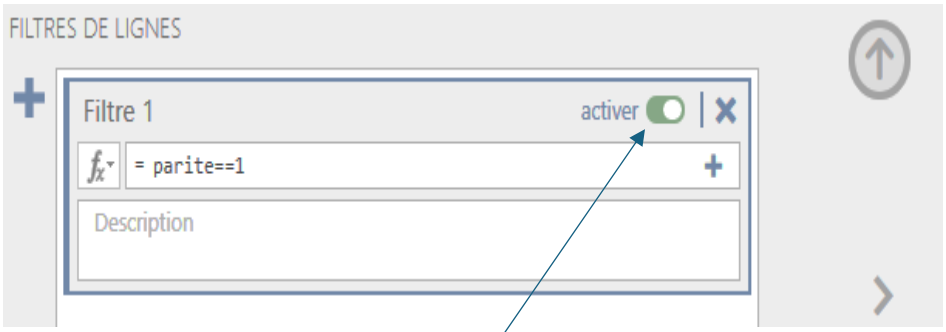
- Sélectionner des lignes (certaines valeurs de variables)

Dans le fichier "decl_temoins.xlsx", on veut ne sélectionner que les "primipares" à partir de la variable "parite" codée par un chiffre (avec 1 pour les primipares). Dans l'onglet "Données", cliquer sur l'icone "Filtre"



Dans la fenêtre, écrire `parite==1` avec deux signes égal ==. On peut aussi mettre des < ou > ou >= ou <=

Le fichier s'affiche avec une nouvelle colonne appelée `Filtre` qui coche les bonnes lignes avec ✓ et les non-sélectionnées avec X



	Filtre 1
1	✓
2	✓
3	✓
4	✓
5	✓
6	✓
7	✓
8	✓
9	✓
10	✓
11	✓
12	✓
13	✓
14	✓
15	✓
16	✓
17	✓
18	✗
19	✓
20	✓
21	✓
22	✓
23	✗
24	✗

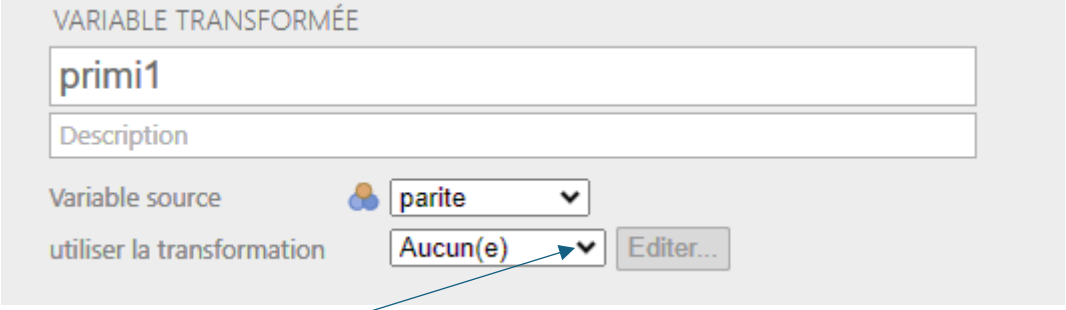
Tous les résultats qui suivront ne concerneront que les primipares. Pour revenir à l'ensemble du fichier, décocher (désactiver) ici

- Transformer des variables

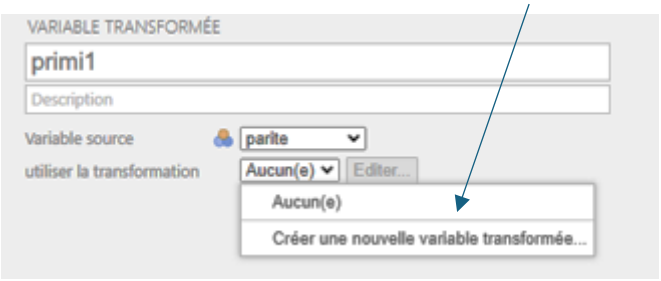
Recoder une variable vers une autre nouvelle

Pour séparer les primipares des multipares, on peut aussi créer une nouvelle variable appelée `primi1`. On veut en extraire une variable correspondant aux primipares à partir de `parite`=1 codée `primi1`=1 versus multipares (`parite`>1, codée `primi1`=2).

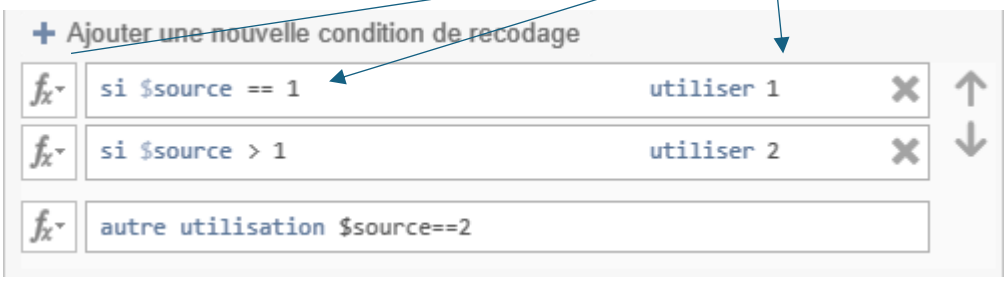
Processus ➔ Onglet **Données** - Cliquer sur la *variable* `parite` en tête de colonnes – Cliquer sur Transformer : une nouvelle variable est créée de type `parite(1)`. On renomme cette nouvelle variable, `primi1`, et on utilise donc la variable source `parite`.



Cliquer sur et cliquer sur *Créer une nouvelle variable transformée* qui affiche la fenêtre suivante.

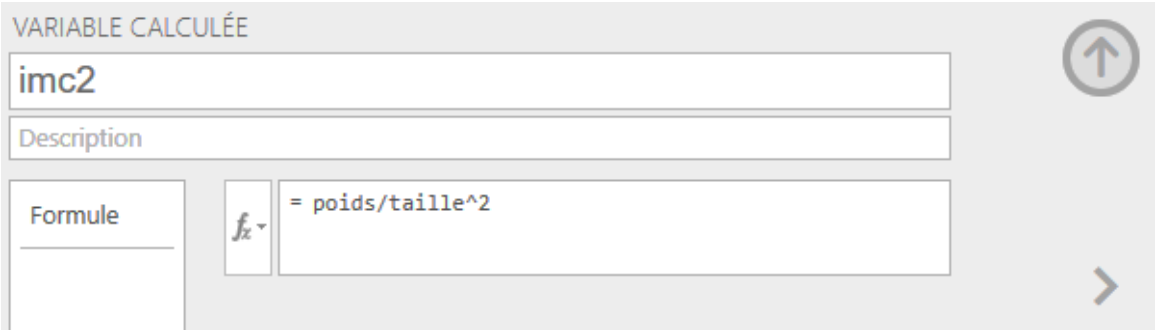


On utilise alors le menu de transformation de la manière suivante : mettre deux signes '==' pour l'égalité, ou >, < ou <= ou >=. Voir pour ajouter une nouvelle condition. Le mot `source` correspond à `parite`



Créer une variable à partir d'autres variables : calcul IMC = poids/taille²

On peut aussi créer une variable avec un calcul établi à partir d'autres variables, comme IMC (BMI).



Créer une variable par différence de temps (heures, jours ou années)

Pour la version téléchargée 2.7.38, on ne peut pas faire de différence de date ni d'heure sous format commun (dit sexagésimal comme 8 h 30, ou 15 mn 30) : les faire dans le tableur de saisie avant d'importer le fichier. Si les heures sont en décimales (8.5 pour 8h30), on peut calculer une différence de temps, en se méfiant des différences des heures qui franchissent minuit. Ainsi, une femme qui est

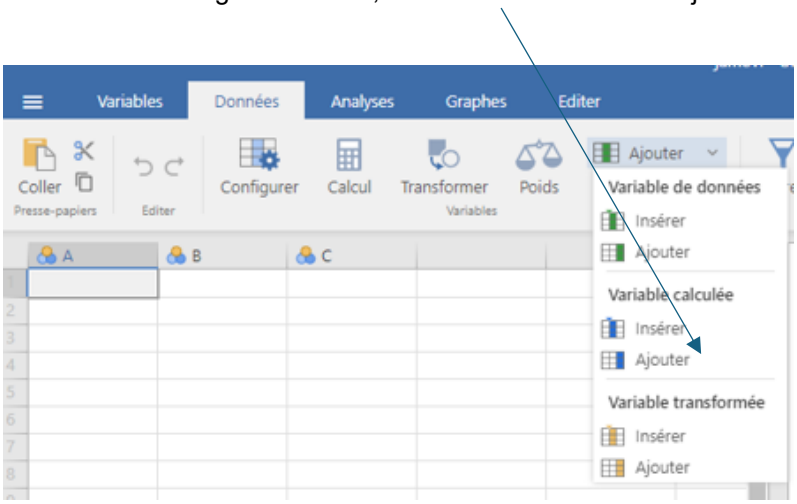
déclenchée à 20 h et qui accouche le lendemain à 5 h du matin, la différence est de 5 h – 20 h = - 15 h ; il faut en fait ajouter 24 h, soit 24 h + (-15 h) = 9 h.

Etablir des seuils pour transformer une variable quantitative en qualitative

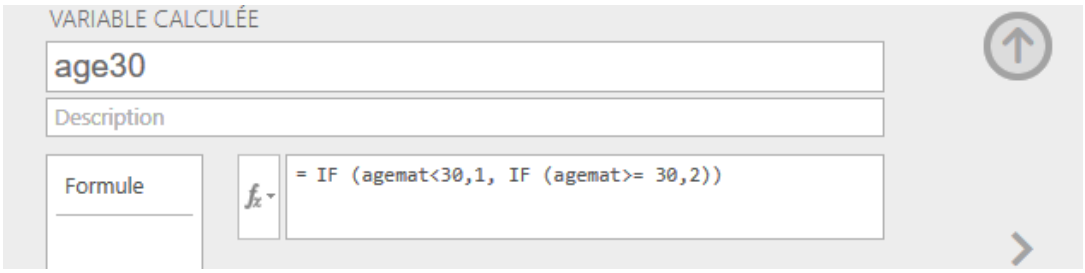
Lors de l'élaboration du questionnaire, et à la saisie, il est déconseillé de fixer des catégories pour les variables quantitatives, car c'est une perte d'information et les classes choisies peuvent ne pas correspondre à d'autres articles qui auraient pris d'autres seuils. **Il faut saisir toujours une variable quantitative comme telle (âge, poids, taille pour calcul IMC en continu),** ou faire des différences de dates pour les âges en quantitatif par exemple.

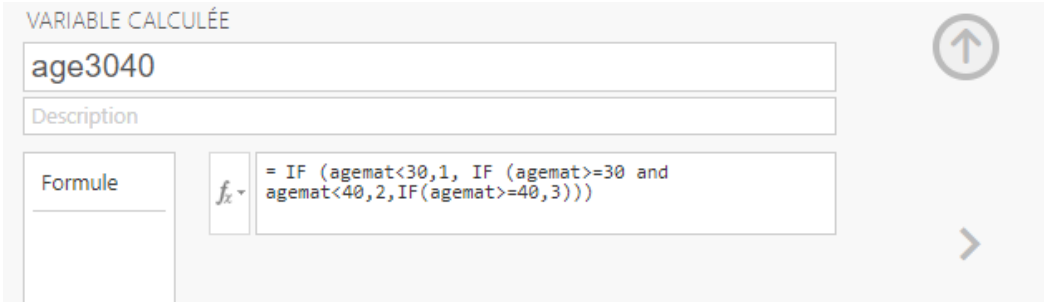
On dispose de la variable `agemat` en années ; on veut déterminer un seuil pour séparer les femmes jeunes (< 30 ans) (codées en 1) de femmes plus âgées (≥ 30 ans) (codées en 2). On peut aussi faire plus de deux classes avec des seuils à 30 et 40 ans.

Processus ➔ Onglet Données, Icône Variable calculée/ ajouter



Ecrire selon la syntaxe suivante le seuil de 30 ans pour l'âge :





Où trouver de l'aide dans jamovi ?

Dans la version 2.7 présentée, il n'y pas d'onglet d'aide.

Sur internet et avec un logiciel d'intelligence artificielle, on trouve facilement des réponses aux questions. Attention, certaines aides concernent des versions anciennes qui ne s'appliquent pas à cette version de 2026. On peut ainsi consulter en PDF :

https://hal.science/hal-02335912v1/file/Manuel%20Jamovi%20fr.pdf

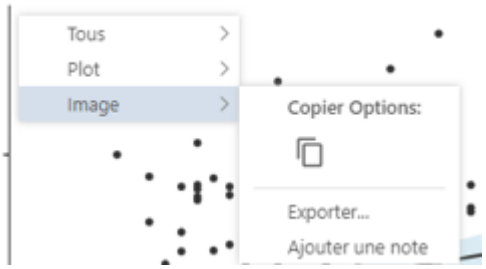
https://jmeunierp8.github.io/ManuelJamovi/

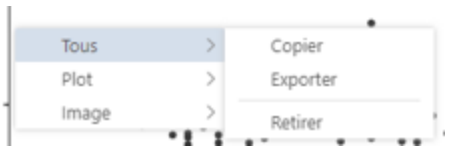
De nombreuses vidéos sont également disponibles.

Principes de manipulation

Dans la suite du document, les **onglets** à utiliser sont en haut de l'écran avec des menus déroulantes, les **fenêtres de commandes** s'affichent à gauche de l'écran, et les **résultats** sont présentés à droite, au fur et à mesure des commandes.

On peut exporter ou copier les résultats sous forme d'images ou PDF (clic droit sous Windows). Pour supprimer tous les résultats, cliquer sur un résultat, puis cliquer sur `Tous` et cliquer sur `retirer`.





Quand on se met au niveau d'une fenêtre de résultats, la fenêtre de commandes est présentée, et on peut vérifier ou modifier si besoin par exemple.

Une fois le fichier validé, il faut le sauver au format **jamovi** sous forme de *.omv. Cliquer sur les trois traits en haut à gauche de l'écran et sauvegarder directement, ou sauvegarder sous… à l'endroit voulu.

Notes personnelles

……

……

……

……

……

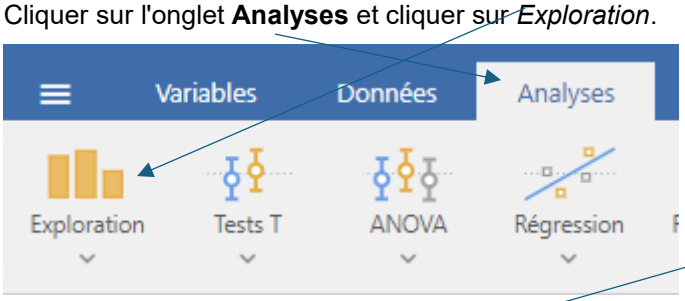
……

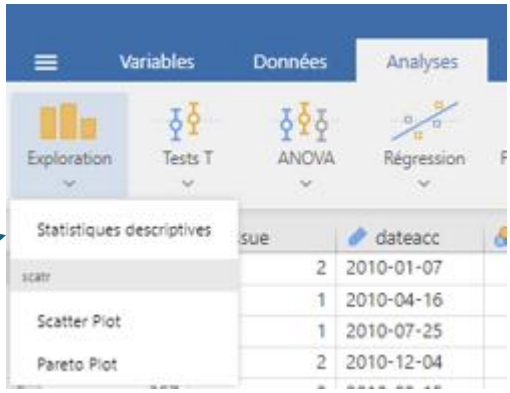
Décrire la population, décrire les variables

- Variables nominales ou ordinales ➔ effectifs et fréquences

Les variables nominales ou codées sont analysées avec des effectifs et des proportions.

Cliquer sur l'onglet **Analyses** et cliquer sur *Exploration*.

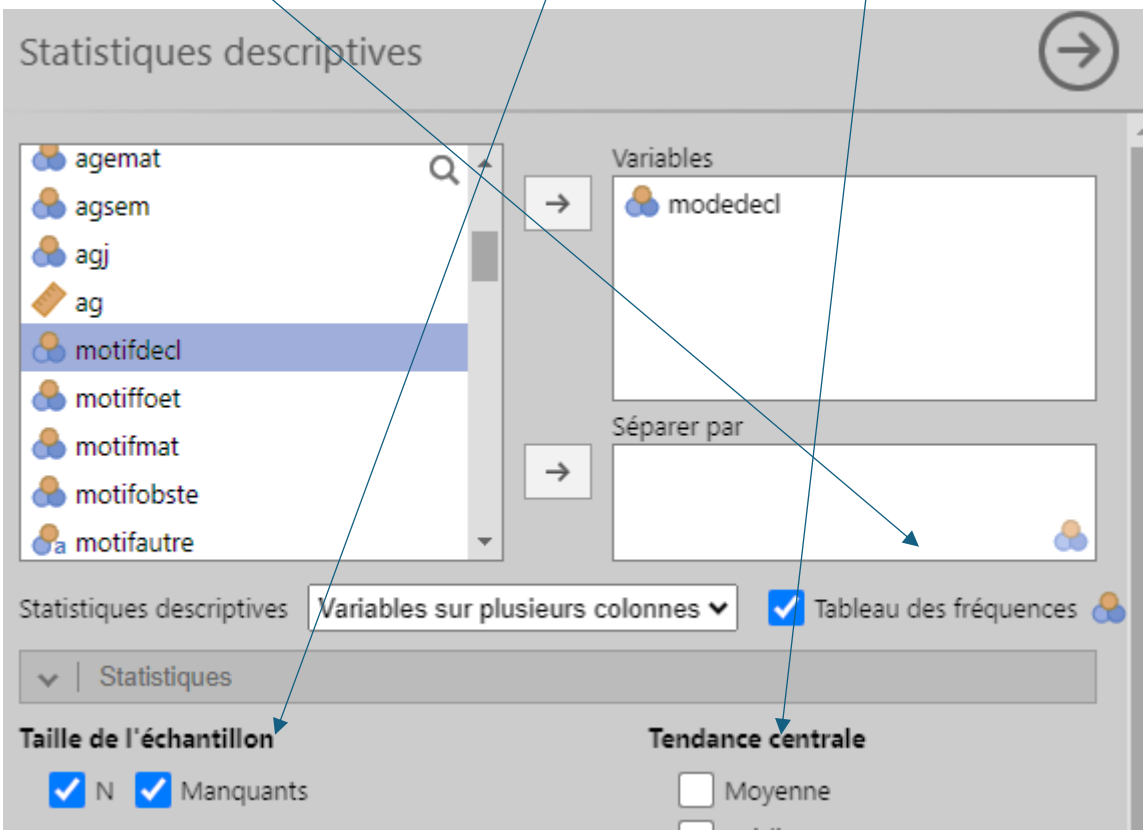




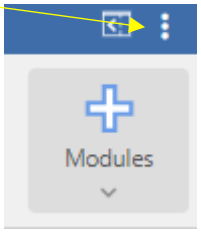
Puis cliquer sur *Statistiques descriptives* :

La fenêtre affiche les variables et on passe les variables que l'on veut analyser de gauche à droite.

Cocher tableau de fréquences, ainsi que N et Manquants. Les autres cases doivent être décochées +++ (la moyenne et autres calculs d'une variable nominale n'ont pas de sens)



Les résultats s'affichent dans la fenêtre de droite. NB : Pour avoir deux chiffres après la virgule, il faut cliquer sur les trois petits points et cocher dans format numérique : 4 chiffres significatifs





Statistiques descriptives

	modedecl
N	295
Manquants	0

Fréquences

Fréquences de modedecl

modedecl	Comptes	% du Total	% cumulés
1	120	40,68 %	40,68 %
2	149	50,51 %	91,19 %
3	26	8,81 %	100,00 %

Si on veut des codes plus explicites avec des labels, voir ➔ Donner des noms aux labels". On peut ainsi avoir le même tableau :

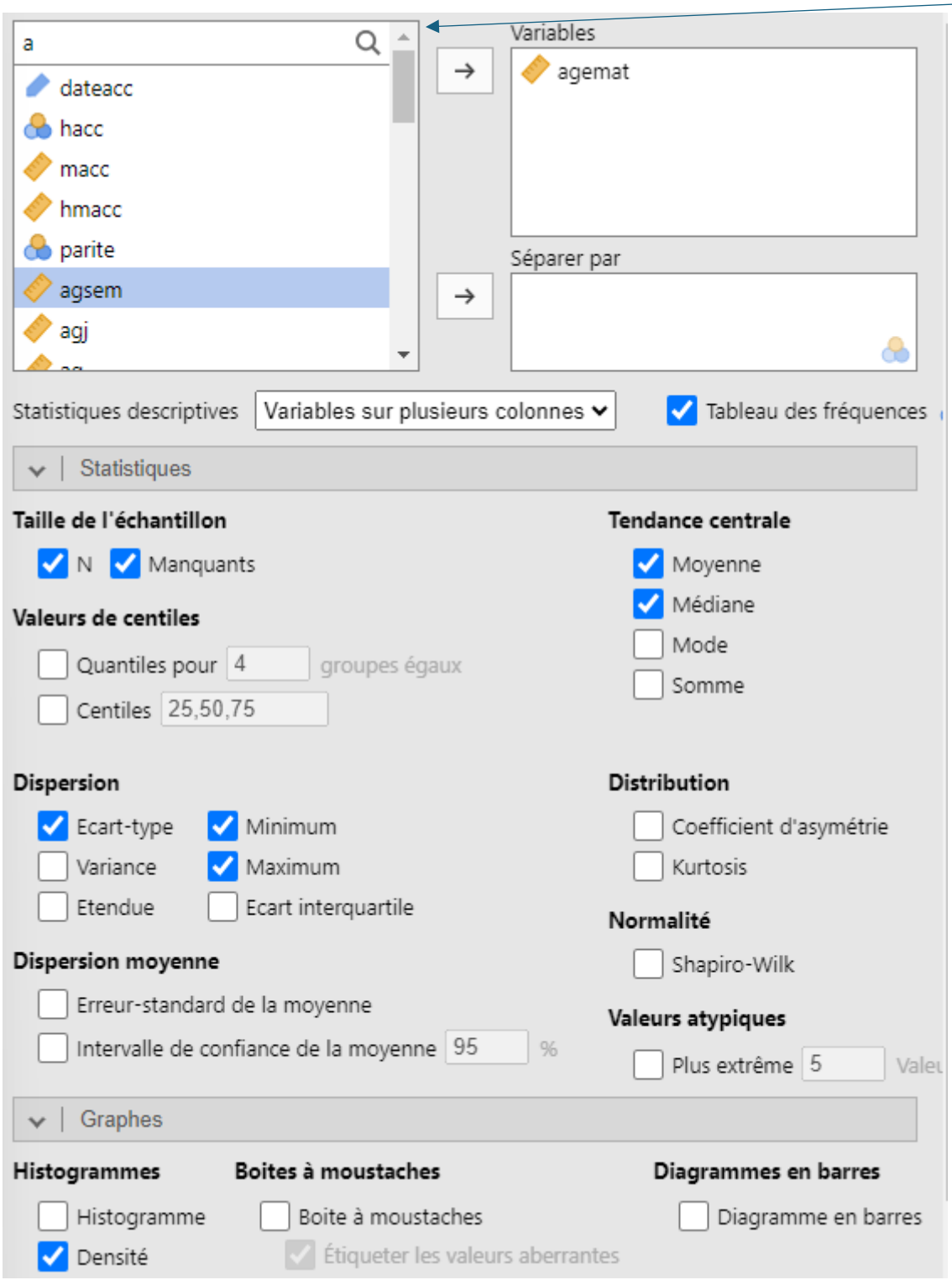
Fréquences

Fréquences de modedecl

modedecl	Comptes	% du Total	% cumulés
Prosta	120	40,68 %	40,68 %
Ocyto	149	50,51 %	91,19 %
Les deux	26	8,81 %	100,00 %

- Variables d'échelle (quantitatives) ➔ valeurs centrales et dispersion

De la même manière ; Onglet **Exploration** / *Statistiques descriptives* / et développer *Statistiques.*
Cocher moyenne et autres valeurs résumé, et écart-type et autres valeurs de dispersion. On peut aussi développer l'onglet *Graphes*. On analyse ici l'âge des mères.
On peut, dans les fichiers avec de nombreuses variables, chercher les premières lettres avec la loupe

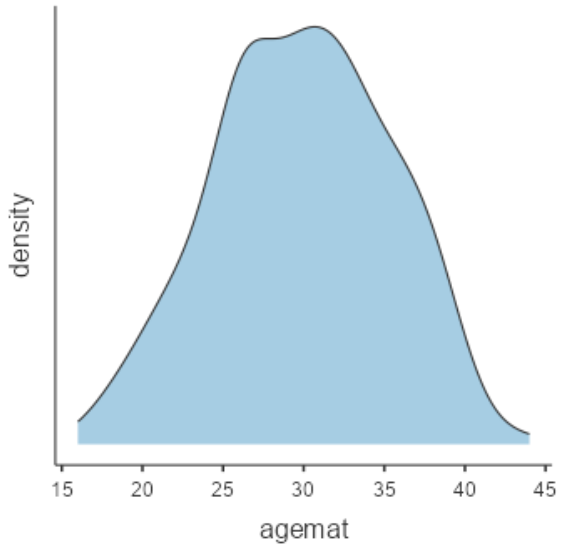


. Les résultats peuvent être en colonnes :

Statistiques descriptives

	agemat
N	295
Manquants	0
Moyenne	29,93
Médiane	30
Déviation standard	5,440
Ecart interquartile	8,000
Etendue	28
Minimum	16
Maximum	44
25-ième percentile	26,00
50-ième percentile	30,00
75-ième percentile	34,00

Exemple de graphe en densité



Ou en lignes

Statistiques descriptives

	N	Manquants	Moyenne	Médiane	Ecart-type	Ecart interquartile	Etendue	Minimum	Maximum
agemat	295	0	29,93	30	5,440	8,000	28	16	44

Les autres cases à cocher concernent les tests de normalité des distributions des variables quantitatives.

Distribution

☑ Coefficient d'asymétrie

☑ Kurtosis

Normalité

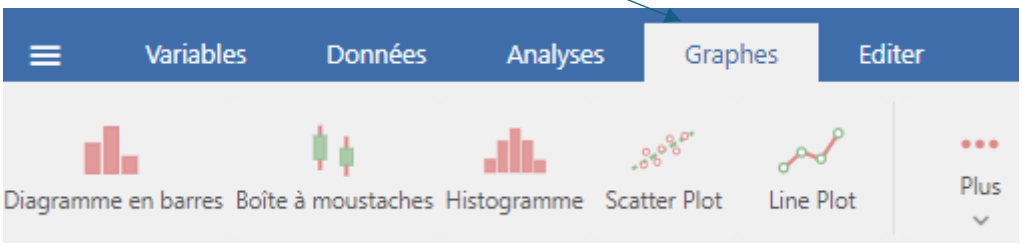
☑ Shapiro-Wilk

Coefficient d'asymétrie	−0,09
Asymétrie de l'erreur-standard	0,14
Kurtosis	−0,53
Kurtosis de l'erreur-standard	0,28
W de Shapiro-Wilk	0,99
Valeur p de Shapiro-Wilk	,041

La loi normale comme modèle d'analyse des données en santé est symétrique avec une forme déterminée (centre et extrêmes). Les tests **d'asymétrie** (on dit aussi skewness) déterminent si les extrêmes (les "queues" de distribution) sont négatives (vers la gauche) comme ici (mais faiblement) ou positives. Le **Kurtosis** détermine si la courbe est plate comme ici (négative) ou pointue (si positive). Enfin, le test de Shapiro-Wilk (voisin du test de Kolmogorov-Smirnov) teste la **normalité** dans l'ensemble : ici avec un "p" à 0.041 on peut dire la répartition de l'âge maternel n'a pas tout à fait une distribution normale.

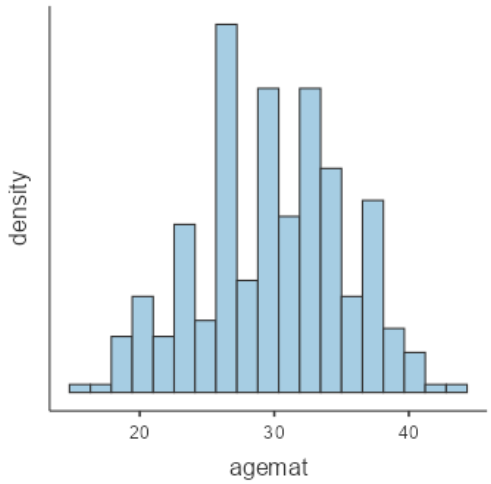
- Graphes descriptifs

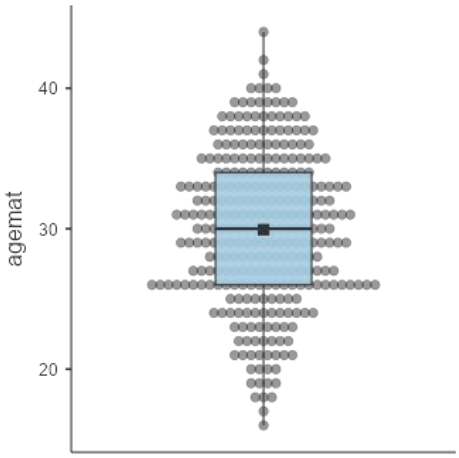
On peut afficher des graphes soit dans **Analyses** *Exploration* / *Statistiques descriptives* à partir de "Plots", soit avec la version 2.7 dans un onglet spécifique **Graphes,** soit avec un module spécifique comme **JJStatsPlot** (voir en fin de document) à télécharger.



Il n'est pas d'usage de montrer dans les articles (quelquefois dans les mémoires ou les thèses) des graphes en barres pour les variables nominales, ni les histogrammes. On montre quelquefois des box-plots (dites dans **jamovi** boîte à moustaches ce qui n'est pas tout à fait pareil… avec ou sans les points pour les valeurs). La couleur bleue est demandée dans le menu des trois points à droite de l'écran.



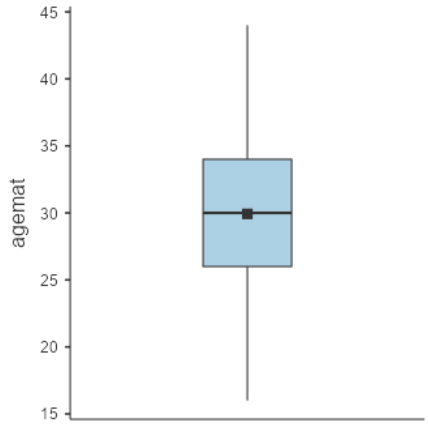




A quoi correspondent les traits des box-plots ?

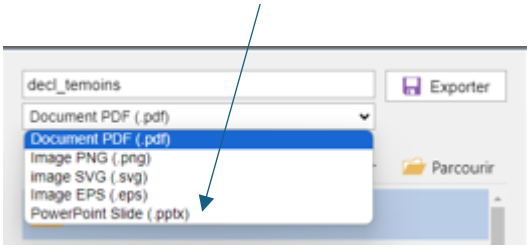
- trait du milieu : médiane (50 % de l'ordre des valeurs)
- Trait supérieur de la boîte : 3ème quartile (75 %)
- Trait inférieur : 1er quartile (25 %)
- Fin du trait vertical supérieur : médiane + 1.5 x (écart -interquartile (valeurs 75 % - 25 % (IQR en anglais))
- Fin du trait vertical inférieur : médiane – 1.5 x (écart -interquartile (valeurs 75 % - 25 % (IQR en anglais))

On peut trouver des valeurs extrêmes en dehors des traits.



L'onglet Graphes permet de modifier les graphes de manière intéressante : on laisse les étudiants découvrir les modalités de modification.

On peut aussi exporter les graphes en différents formats (clic droit sur le graphe) : format *.pdf, ou image *.png ou sous forme d'une diapo dans PowerPoint où on peut le modifier à loisir.

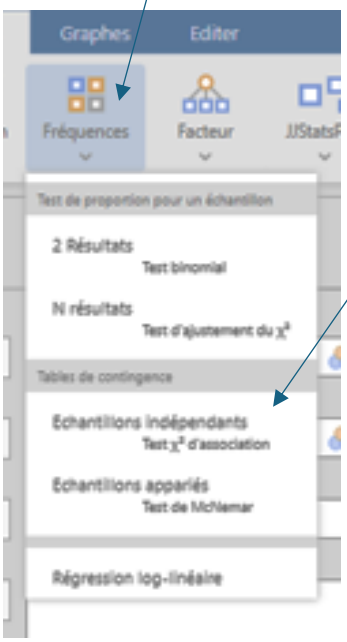


Comparer les données – Tests statistiques

- La variable de jugement est qualitative à deux modalités

La ou les variables d'exposition sont qualitatives : tableau 2 x 2 ou 2 x n

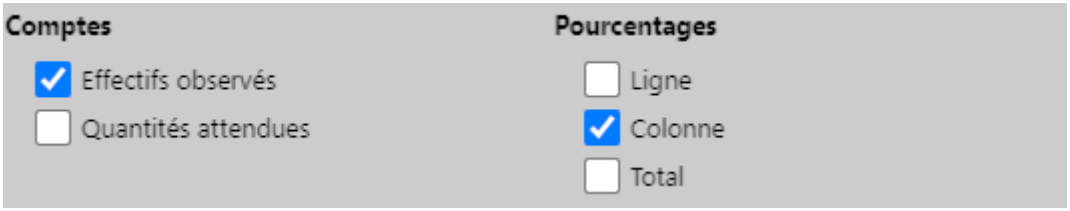
Onglet **Fréquences** (mal nommé) et cocher *Echantillons indépendants* (test du χ^2)



On va croiser le critère de jugement (issue en CS ou VB) en colonnes et la primiparité en lignes.

Cocher le test du χ^2, et éventuellement le test de Fisher en cas de petits effectifs, et ici l'odds ratio (c'est une enquête cas-témoins) avec la différence des proportions et les intervalles de confiance des proportions.





Tables de contingence

Tables de contingence

primipare		issue: CS	issue: VB	Total
Primi	Observé	113	55	168
	% par colonne	77,40 %	37,16 %	57,14 %
Multi	Observé	33	93	126
	% par colonne	22,60 %	62,84 %	42,86 %
Total	Observé	146	148	294
	% par colonne	100,00 %	100,00 %	100,00 %

On compare donc les proportions de primipare en cas de césarienne (CS) et de voie basse (VB). Proportions en colonnes +++

Tests χ^2

	Valeur	ddl	p
χ^2	48,58	1	< ,001
N	294		

Mesures comparatives

		Intervalles de confiance à 95%	
	Valeur	Borne inf	Supérieur
Différence entre deux proportions	0,4107 [a]	0,3062	0,5153
Rapport des cotes (odds ratio)	5,790	3,472	9,656

[a] Lignes comparées

Le test du χ^2 vaut 48.58 (limite attendue > 3.84) et montre une p-value < 0.001, avec une différence de proportions de 41.07 % (IC95 % : 30.62 % - 51.53 %). La différence est très significative avec p< 0.001. L'odds ratio permet de quantifier la différence avec OR = 5.79 (3.47 – 6.65) (prendre deux décimales). NB : L'intervalle de confiance de la différence n'est pas montrée ici (certains articles préfèrent désormais ce résultat au "p"). Voir éventuellement module "esci" (fin document) avec le résultat suivant (IC 95 % de la différence ne comprend pas 0) :

Proportion Difference

			95% CI	
Outcome variable	issue Effect	*P*	LL	UL
primipare	P_{Primi}: VB	0,372	0,298	0,452
	P_{Primi}: CS	0,774	0,699	0,834
	P_{Primi}: VB – CS	−0,402	−0,500	−0,294

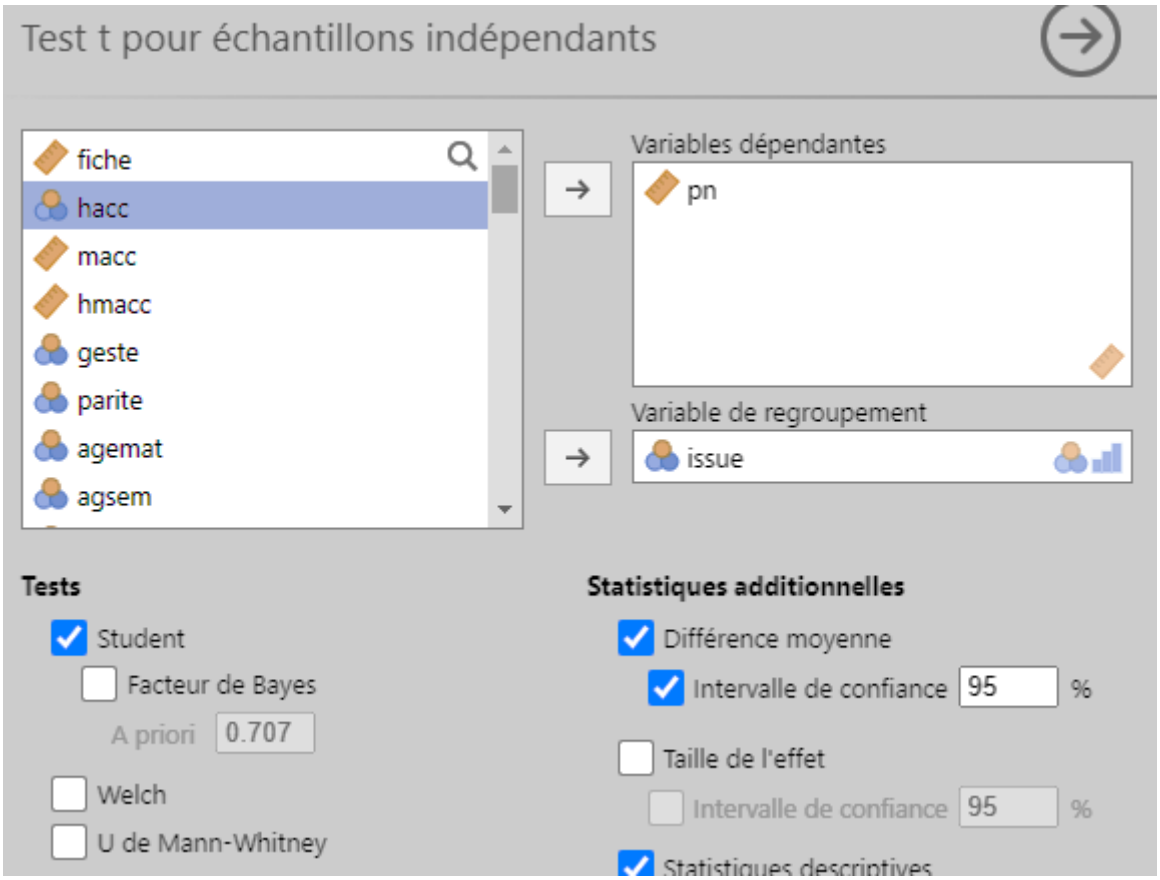
La ou les variables d'exposition sont quantitatives

Cliquer sur l'onglet Tests T et cliquer sur test T pour échantillons indépendants



Les variables d'intérêt en quantitatif sont dites "variables dépendantes". On compare donc le poids de naissance `pn` selon l'issue de l'accouchement CS ou VB.

Cocher test de Student, et éventuellement Welch (calcul voisin de Student) et éventuellement pour les petits effectifs U de Mann-Whitney qui est un équivalent du test de Wilcoxon.



Test t pour échantillons indépendants

Test t pour échantillons indépendants

		Statistique	ddl	p	Différence moyenne	Différence d'erreur standard	Intervalle de confiance à 95% Borne inf	Supérieur
pn	t de Student	−2,34 [a]	293,00	,020	−166,24	70,95	−305,88	−26,59

Note. H_a $\mu_{CS} \neq \mu_{VB}$

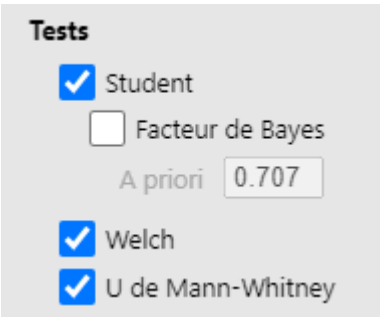
[a] Le test de Levene est significatif (p<0.05), suggérant une violation de la condition d'égalité des variances

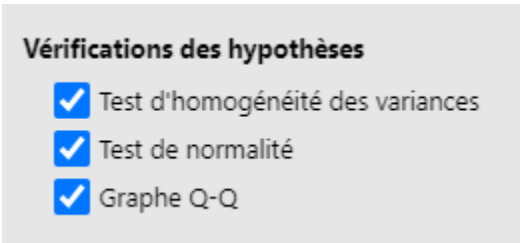
Statistiques descriptives des groupes

	Groupe	N	Moyenne	Médiane	Ecart-type	Erreur standard
pn	CS	147	3165,71	3215,00	676,01	55,76
	VB	148	3331,95	3332,50	534,96	43,97

Les nouveau-nés nés par CS pèsent 3 166 g (arrondi, pas besoin de décimales) et ceux nés par VB pèsent 3 332 g. La différence est de 166 g (27 – 305 g en valeur absolue, ne comprenant pas 0 g), le test de t vaut 2.34 avec 295 – 2 = 293 degrés de liberté avec un p = 0.02, donc p<0.05, significatif. A noter que sur le plan médical, il est contre-intuitif d'observer que les nouveau-nés par voie basse pèsent plus lourd que les nouveau-nés nés par césarienne (il y a des explications) : ne pas regarder seulement le "p" ou la différence, mais **LE SENS DE LA DIFFERENCE.**

Le test de t de Student suppose la normalité des distributions (habituelle en cas de grands effectifs comme ici), et l'égalité des variances dans les deux échantillons. On peut tester la comparaison des variances :





Test d'homogénéité des variances (test de Levene)

	F	ddl	ddl2	p
pn	7,765	1	293	,006

Note. Une valeur p faible suggère une violation de la condition d'égalité des variances

Statistiques descriptives des groupes

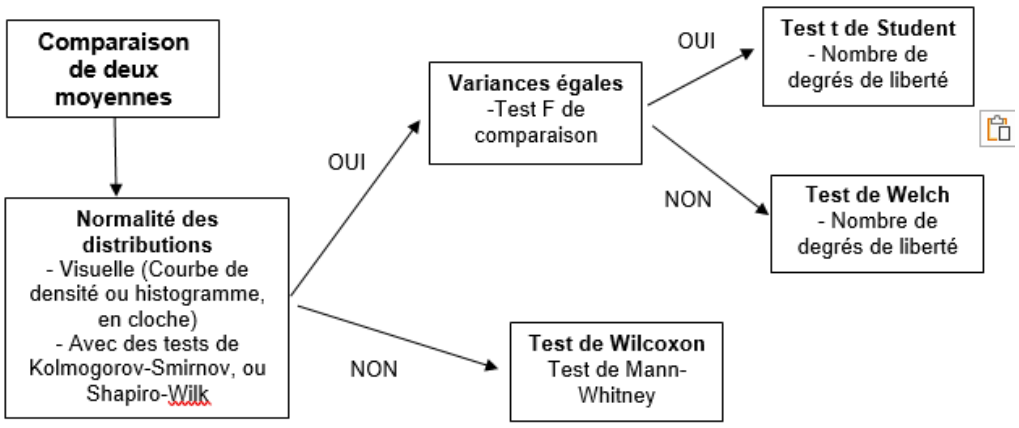
	Groupe	N	Moyenne	Médiane	Ecart-type	Erreur standard
pn	CS	147	3166	3215	676,0	55,76
	VB	148	3332	3333	535,0	43,97

Test de normalité (Shapiro-Wilk)

	W	p
pn	0,9883	,018

Note. Une valeur p faible suggère une violation de la condition de normalité

Les variances sont significativement différentes (p = 0.006) et donc il faut plutôt utiliser le test de Welch qui ne fait cette hypothèse et montre aussi un résultat significatif (p=0.02). De plus, le test de normalité montre que la répartition des `pn` n'est pas normale (p=0.018). Dans ces conditions, on devrait sans doute utiliser le test de Mann-Whitney qui n'est pas significatif (p= 0.062, soit > 0.05)…
Voir ci-après les principes d'utilisation des tests de comparaisons de deux moyennes.



- La variable de jugement est quantitative

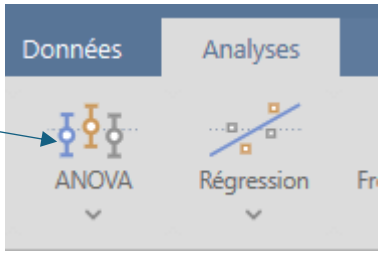
La ou les variables d'exposition sont qualitatives à deux modalités

Voir le test de t et semblable ci-dessus page 20 : les variables d'exposition sont quantitatives.

La ou les variables d'exposition sont qualitatives à plus de deux modalités

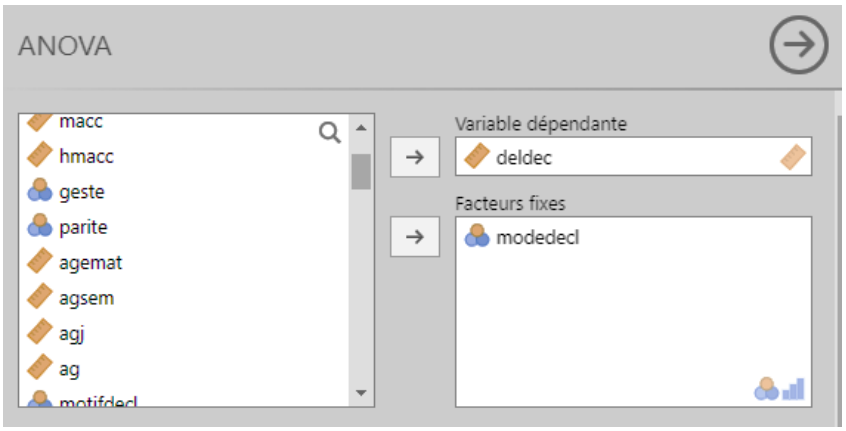
La méthode statistique consiste à comparer une variable quantitative comme la durée du déclenchement (en heures) (`deldec`) avec les modalités de déclenchement codée en 3 catégories (1-Prosta, 2-Ocytocine, 3- les deux) (`modedecl`) ou bien avec trois catégories d'âge maternel /30 et 40 ans (`age30_40`)

Onglet **Analyses** / *ANOVA*



On introduit la variable dépendante avec le délai et le mode de déclenchement et le facteur fixe avec le mode de déclenchement qui doit être en nominal ou ordinale (pas échelle).

On coche les tests d'égalité des variances et les tests dits "pos-hoc" qui permet de comparer les moyennes deux à deux. On peut également coccher les résultats sous forme de graphes.



ANOVA - deldec

	Somme des carrés	ddl	Carrés moyens	F	p
modedecl	25490	2	12744,8	97,32	< ,001
Résidus	37323	285	131,0		

Le résultat se lit sur la ligne de `modedecl` : le Test de F est de 97.32 avec 2 degrés de liberté (3 modalités – 1) pour 285 (on dit dans les tables un F^2_{285} dont le seuil à 0.05 est de 6.81) et une p-value < 0.001. La comparaison post-hoc montre chaque catégorie 1,2 ou3 est différente des autres. A noter que la répartition de la variable `deldec` ne suit pas une loi normale et que les variances ne sont pas homogènes, et demande peut-être à être réservé sur le plan statistique pur (on peut s'en sortir mais pas dans ce tuto)

ANOVA - deldec

	Somme des carrés	ddl	Carrés moyens	F	p
modedecl	25374	2	12686,8	96,29	< ,001
Résidus	37813	287	131,8		

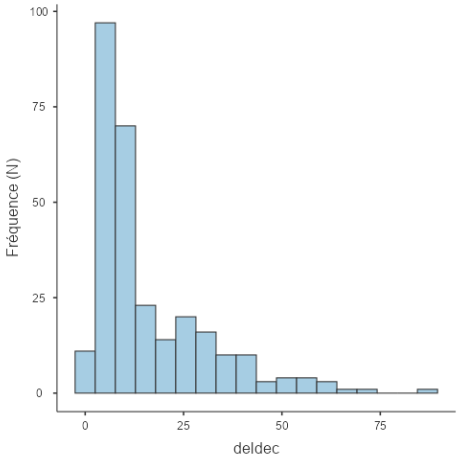
Vérifications des hypothèses

Test d'homogénéité des variances (test de Levene)

F	ddl1	ddl2	p
59,98	2	287	< ,001

Test de normalité (Shapiro-Wilk)

Statistique	p
0,8922	< ,001



Tests post hoc

Comparaisons post hoc - modedecl

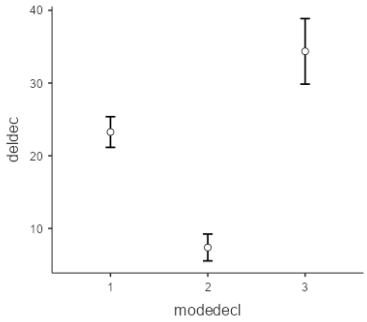
Comparaison			Différence moyenne	Erreur standard	ddl	t	p_{tukey}	$p_{bonferroni}$
modedecl		modedecl						
1	-	2	15,87	1,424	285,0	11,148	< ,001	< ,001
	-	3	−11,10	2,527	285,0	−4,391	< ,001	< ,001
2	-	3	−26,97	2,473	285,0	−10,905	< ,001	< ,001

Note. Comparaisons basées sur les moyennes marginales estimées

Le graphe et l'analyse des moyennes dans l'onglet **Exploration** / *Statistiques descriptives* montrent les délais d'accouchement selon les produits utilisés pour le déclenchement. L'utilisation de prostaglandines correspond à une durée plus longue car le col est fermé et doit être maturé, tandis que l'utilisation de l'ocytocine s'adresse aux cols déjà ouverts. L'utilisation des deux produits successifs correspondant aux délais les plus longs… On peut voir la différence des écart-types (donc des variances).

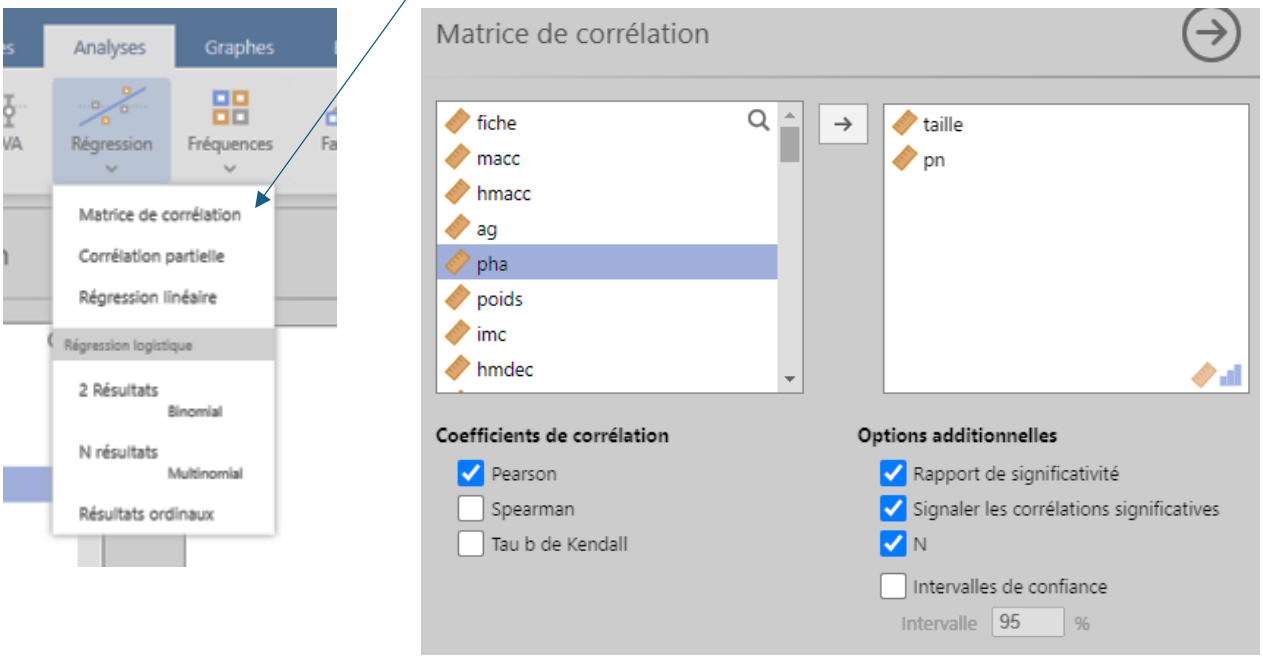
Statistiques descriptives

	modedecl	N	Manquants	Moyenne	Ecart-type
deldec	1	114	6	23,262	15,632
	2	149	0	7,388	4,727
	3	25	1	34,360	16,336



La ou les variables d'exposition sont quantitatives

Dans l'onglet **Analyses**, cocher *Matrice de corrélation*



Matrice de corrélation

Matrice de corrélation

		taille	pn
taille	r de Pearson	—	
	ddl	—	
	valeur p	—	
	N	—	
pn	r de Pearson	0,18 **	—
	ddl	283	—
	valeur p	,002	—
	N	285	—

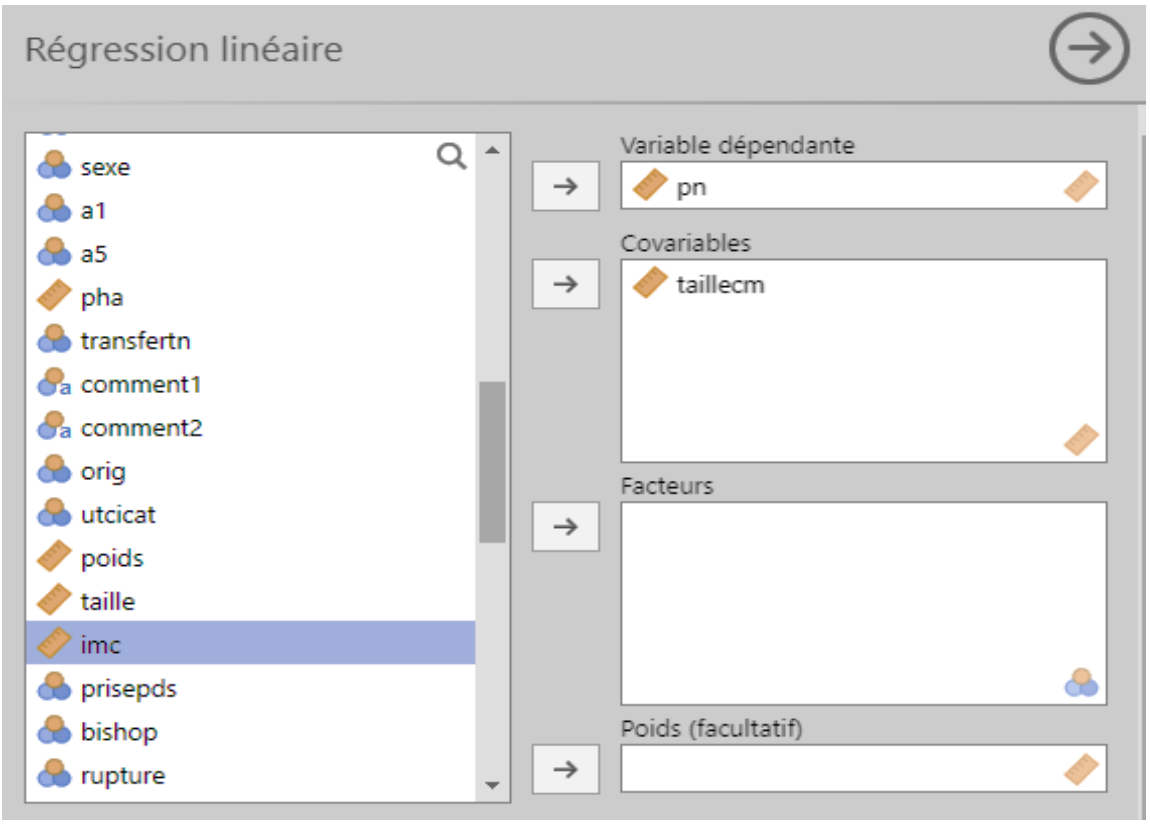
Note. * p < .05, ** p < .01, *** p < .001

Le coefficient de corrélation vaut 0.18 : il est positif c'est-à-dire que plus la taille de la mère augmente, plus le poids du nouveau-né augmente. La p-value vaut p = 0.002 veut dire que le coefficient est différent de 0 (il varie théoriquement de -1 à +1 en passant par 0). Cependant il est assez faible montrant qu'il y a d'autres facteurs qui peuvent expliquer le poids de naissance.

En cas de petits effectifs, préférer le ρ de Spearman (rho) qui s'interprète de la même manière.

Régression linéaire simple

On garde les mêmes variables et on réalise une analyse telle que la variable dépendante (ou à expliquer) est le poids de naissance `pn`, et la variable explicative est la `taille` de la mère.



Régression linéaire

Mesures de l'ajustement du modèle

Modèle	R	R^2
1	0,18	0,03

Note. Models estimated using sample size of N=285

Coefficients du modèle - pn

Prédicteur	Estimation	Erreur standard	t	p
Ordonnée à l'origine	505,87	869,90	0,58	,561
taillecm	16,78	5,33	3,15	,002

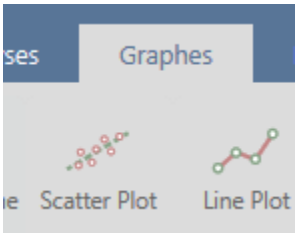
On retrouve le coefficient de corrélation R = 0.18. Le R^2 vaut 0.03 et veut dire que 3 % de la variance du poids de naissance sont expliqués par la taille de la mère (c'est faible).

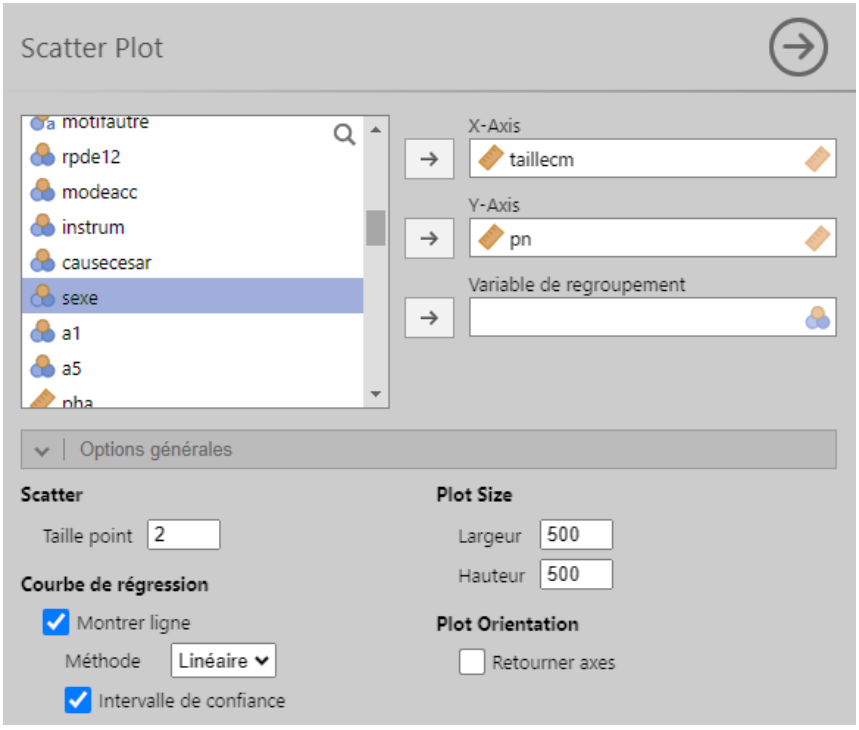
La courbe de la droite de régression s'écrit donc :

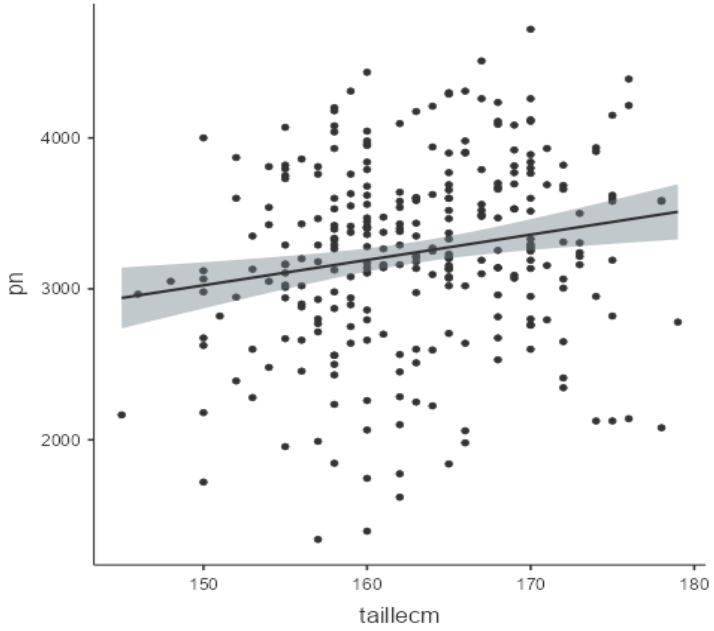
Poids de naissance (g) = 16.78 * taille de la mère (cm) + 505.87

Le coefficient 16.78 est la pente de la droite et elle est différente de 0 avec p =0.002. On peut dire aussi que pour 1 cm de la mère en plus, le nouveau-né pèse 16.78 g de plus (sous l'hypothèse de la linéarité).

On peut dresser un graphe *scatter plot* avec la droite de régression tracée.



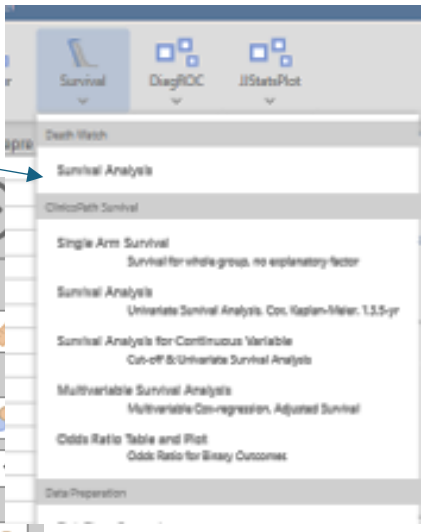




Autres méthodes statistiques

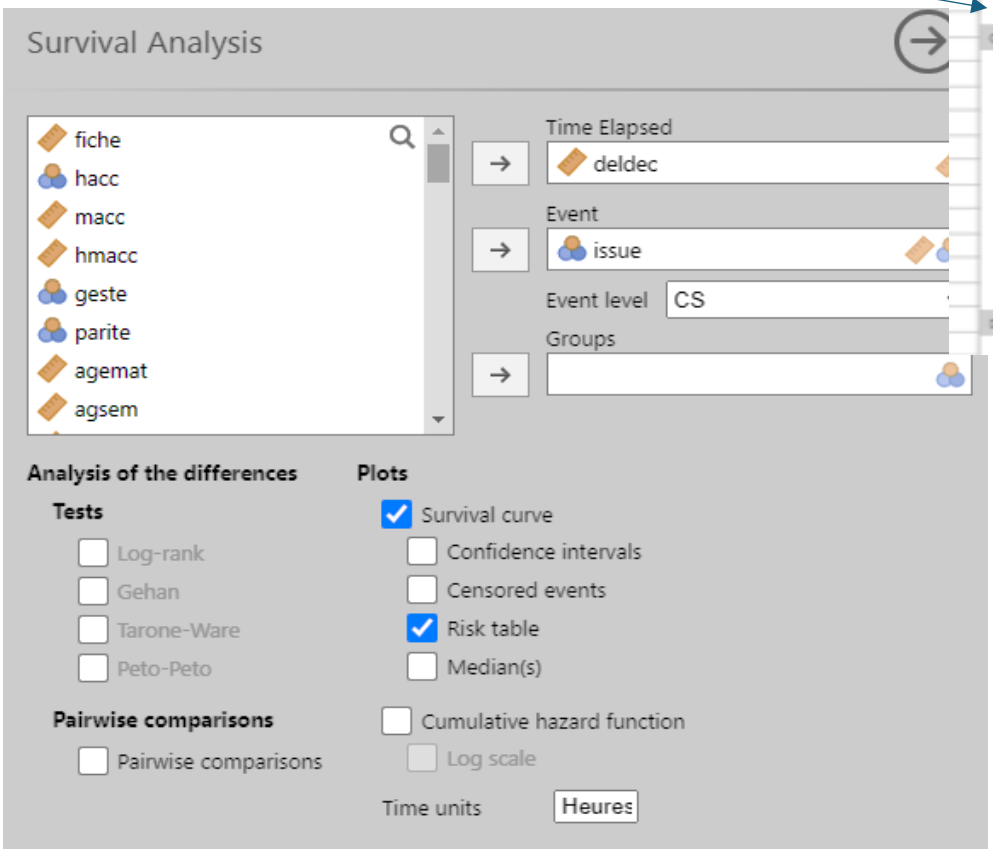
Courbes de survie.

Ajouter le module "jsurvival" (+ dans le coin droit) (voir fin document).



Survie globale, simple

Dans l'onglet **Survival**, cliquer sur *Survival Analysis*



Ici, on prend comme variable de temps le délai entre le moment du déclenchement et l'heure de l'accouchement, en heures (`deldec`) dans *Time Elapsed.* Le critère habituel Event (décédé/vivant) est ici la variable `issue` CS/VB dans *Event*, le "décès" étant vu comme la césarienne (CS). On coche Courbe de survie et table des événements (CS), et préciser les unités de temps en heures.

Survival Analysis

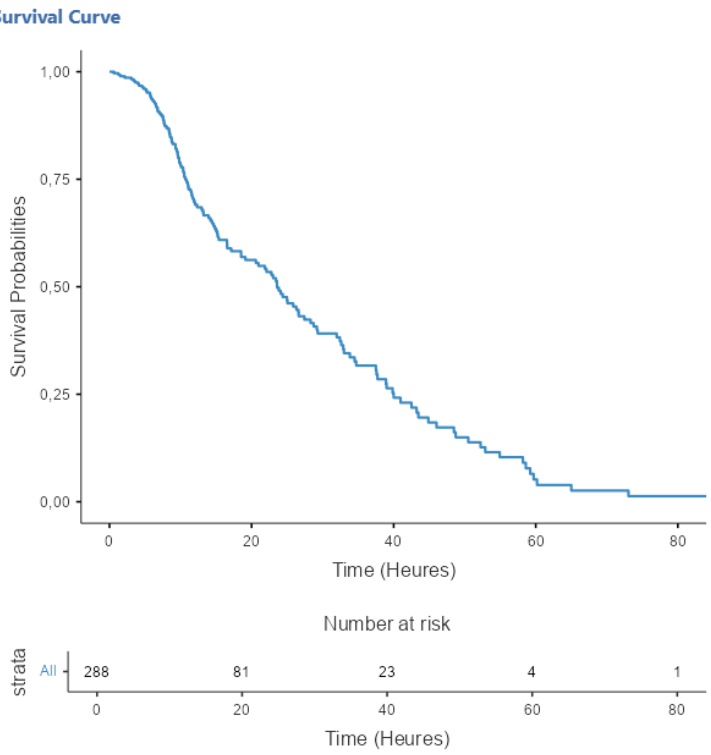
Events Summary

N	Censored	Observed Events	Expected Events
288	147	141	.

Median Estimates

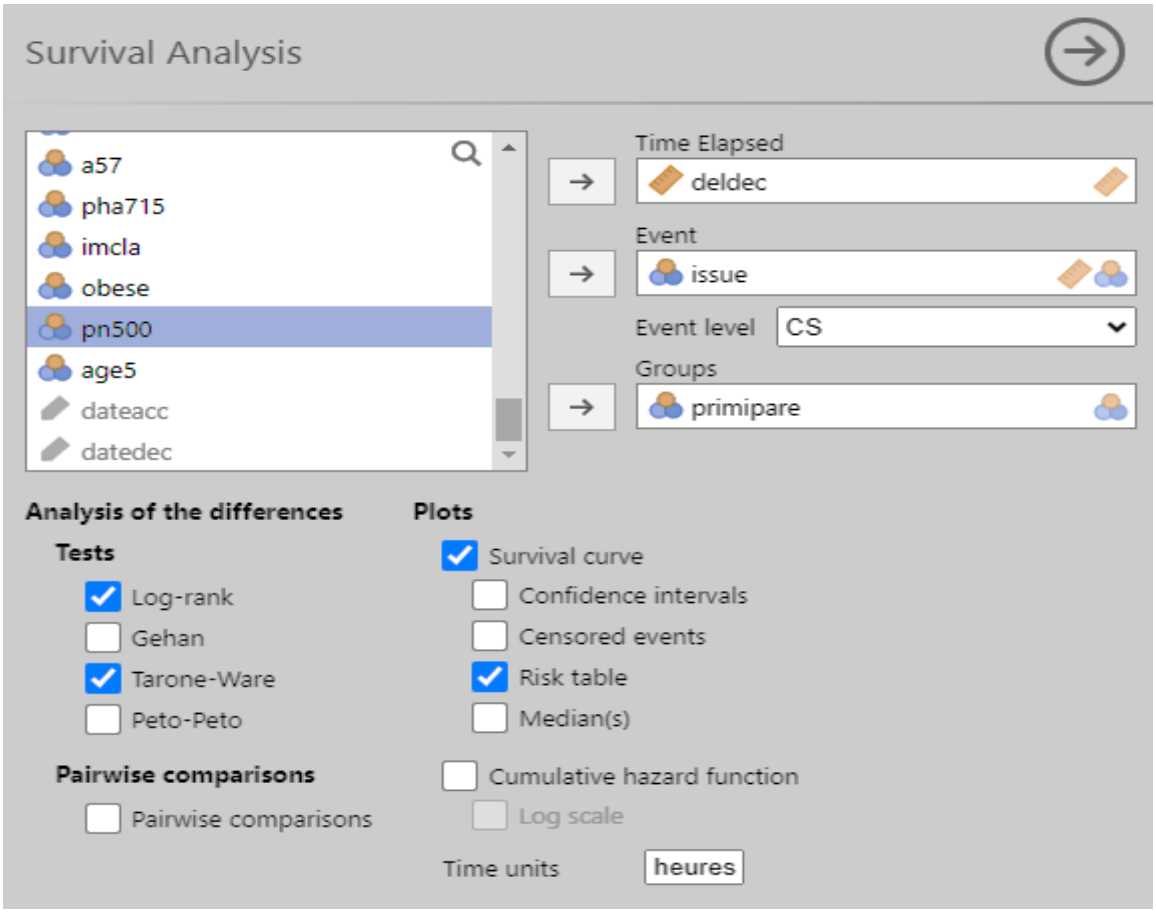
Median	Lower	Upper
23,58	19,13	28,74

La médiane de survie est de 23.58 heures. Le graphe n'est pas modifiable dans les menus, comme mettre les axes des X avec 6 heures ou 12 heures comme graduations. La table de risque est plus précise et souvent demandée dans les articles publiés (montrant souvent les faibles effectifs de queue de courbe).



"Survie" (durée accouchement) selon la primiparité (/ Multiparité) – Log-Rank

Il s'agit de comparer la survie selon que la femme est primipare ou multipare. Il faut ajouter la variable "Groups" avec `primipare codée` en 2 modalités.



Survival Analysis

Events Summary

	N	Censored	Observed Events	Expected Events
Primi	164	54	110	115,18
Multi	123	93	30	24,82

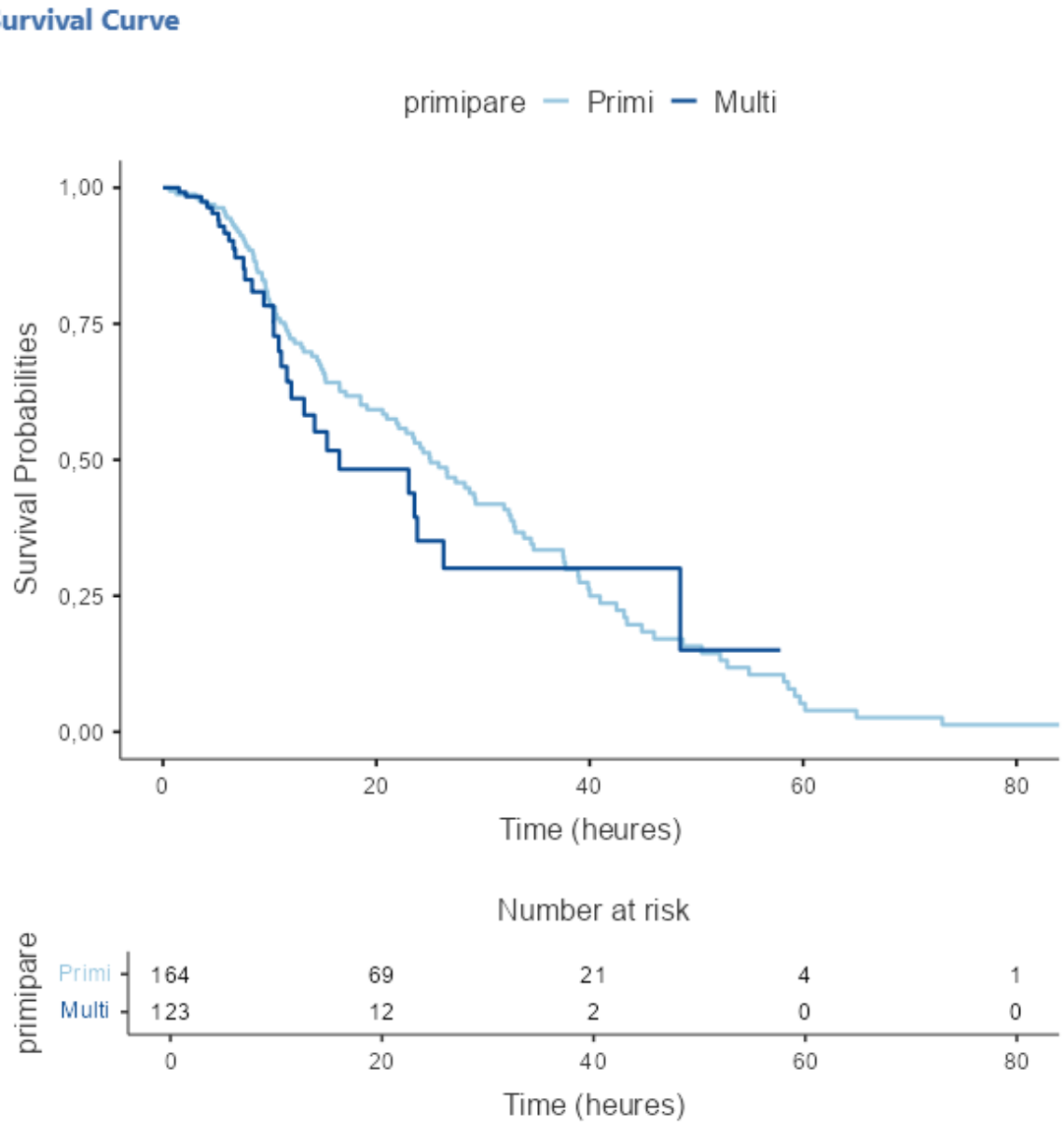
Median Estimates

	Median	Lower	Upper
Primi	25,05	20,97	32,43
Multi	16,53	12,03	NaN

Analysis of the differences

Test	χ^2	df	p
Log-rank	1,410	1	,235
Tarone-Ware	1,464	1	,226

Les médianes de "survie"(durée accouchement) sont de 25 heures pour les primipares et de 16 heures 30 pour les multipares. Le test de comparaison du Log-Rank vaut 1.410 avec p = 0.235 (non significatif). D'ailleurs les courbes se croisent rendant difficile l'interprétation à partie de 40 heures. Le test de Tarone-Ware est une autre manière de calculer la différence (pas dans ce tutoriel).



Autre exemple avec la variable "utérus cicatriciel" `utcicat` codé en oui (1) et 2 (non) uniquement chez les multipares. La différence est significative < 0.001 avec des médianes respectives de 8 heures pour les atctd d'utérus cicatriciel et 26 heures pour les autres.

	N	Censored	Observed Events	Expected Events
1	18	4	14	3,525
2	269	143	126	136,475

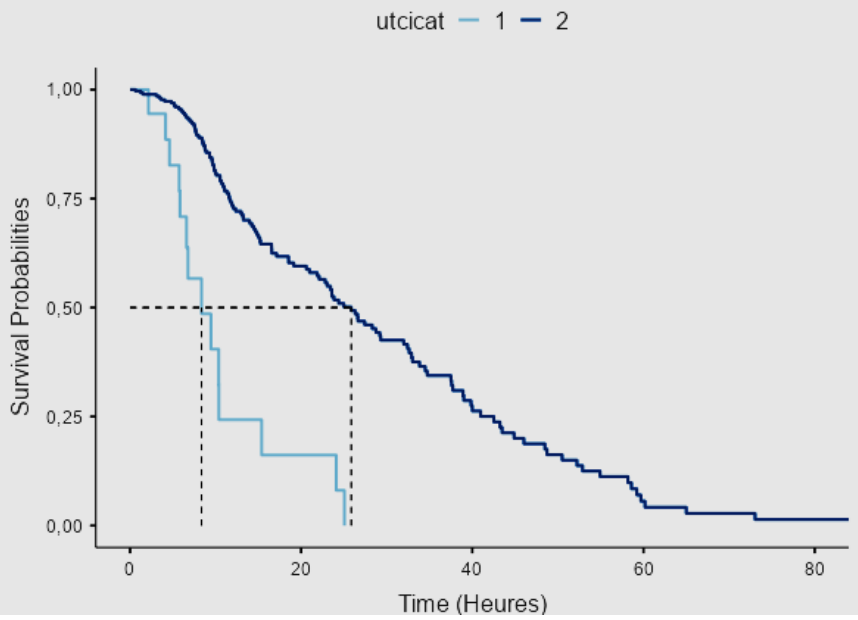
Median Estimates

	Median	Lower	Upper
1	8,350	6,600	NaN
2	25,850	22,100	32,43

Analysis of the differences

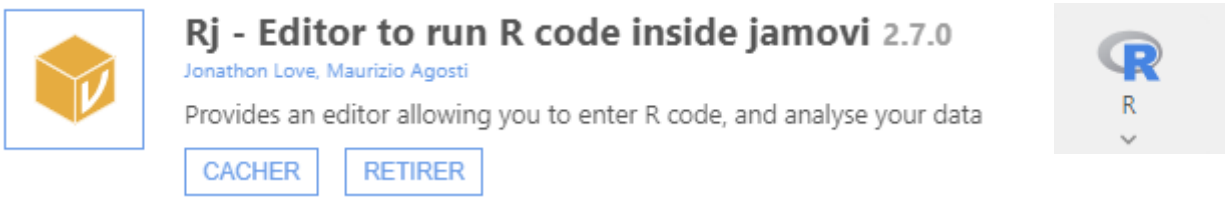
Test	χ^2	df	p
Log-rank	32,49	1	< ,001
Tarone-Ware	22,67	1	< ,001

Survival Curve



Modèle de Cox avec variables nominales et continues. Notion de HR

En cours de réalisation dans **jamovi …** Sinon, pour les pro des stats, on peut passer par R avec le module suivant, avec des lignes de commande.

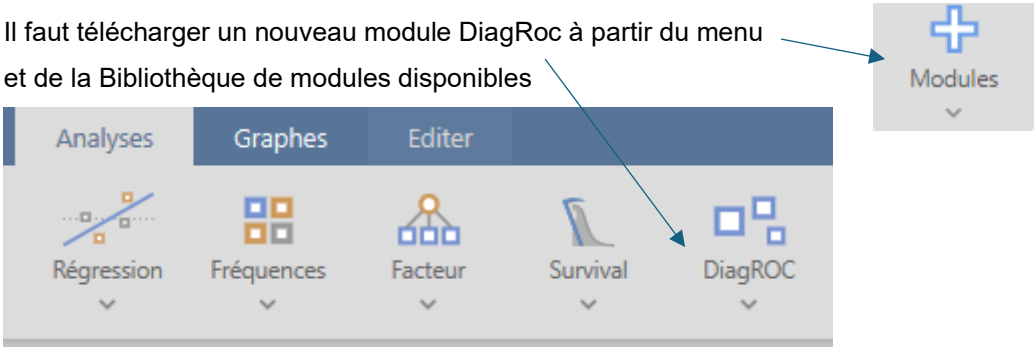


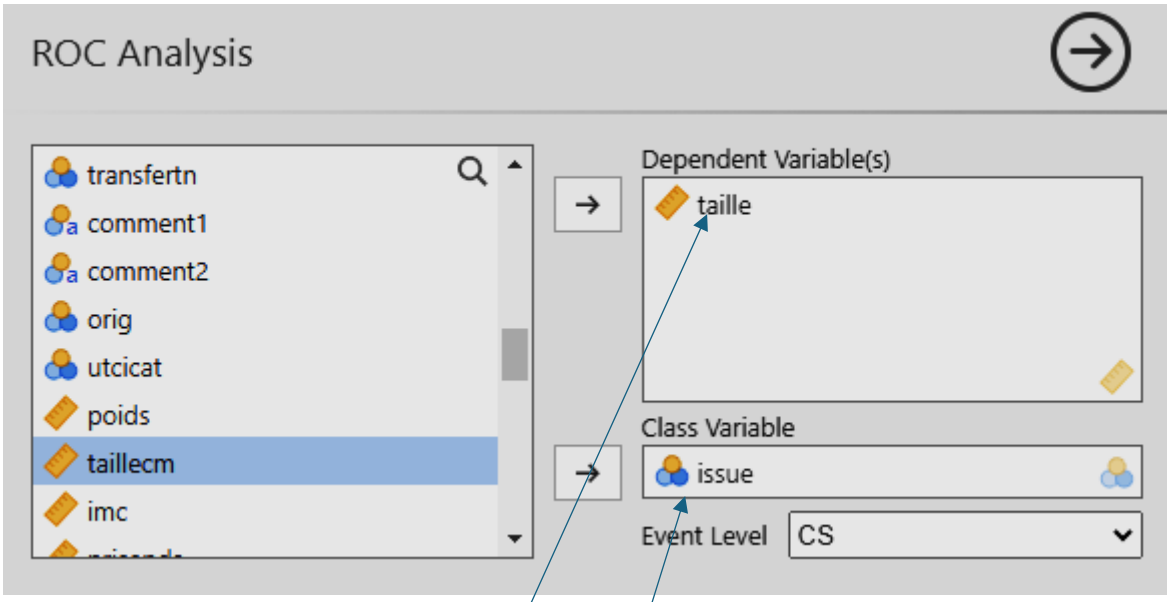
Voir le tutoriel de "R" d'Epi-Sûre https://www.episure.fr/tutoriels

Analyse des tests diagnostiques ou de dépistage – Courbes ROC

Il faut télécharger un nouveau module DiagRoc à partir du menu

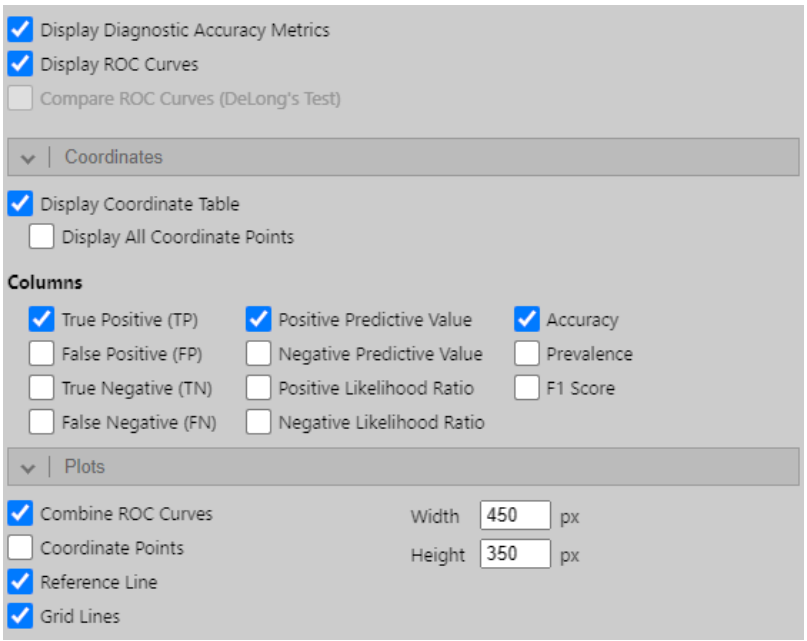
et de la Bibliothèque de modules disponibles





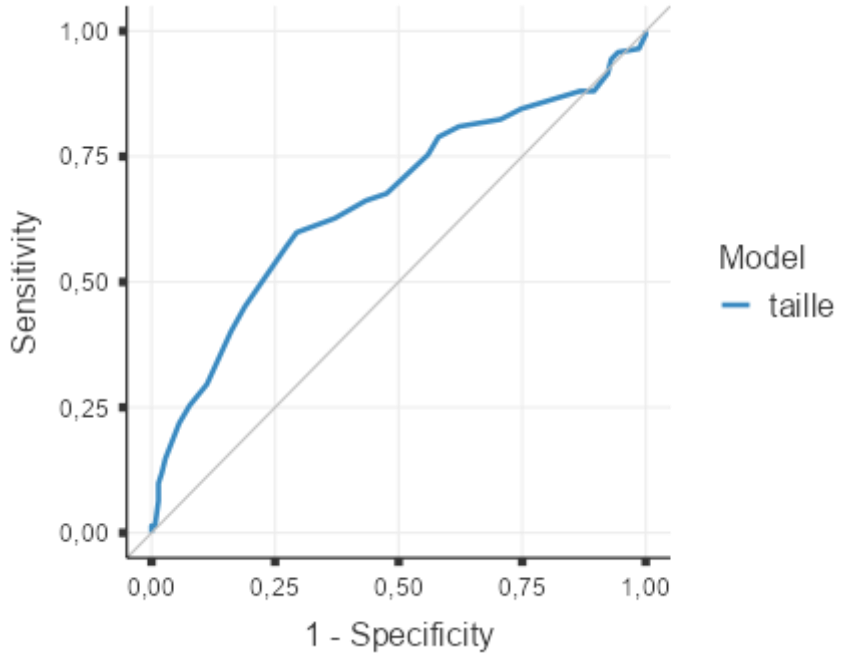
La variable dépendants est une variable continue pour laquelle on va déterminer le meilleur seuil (ici la taille) pour prédire la "class variable" pour une variable de maladie/bien portant ou équivalent(ici césarienne versus voie basse).

Les autres cases à cocher sont les suivantes (non limitatives selon ce que l'on veut) :



La courbe ROC (Receiver Operating Curve) ou courbe de performance en français s'affiche avec la diagonale qui correspond à la surface sous la courbe de 0.5. Plus la courbe est tirée vers le coin supérieur gauche, plus la surface sous la courbe (AUC : Area Under Curve) est grande, plus le signe est discriminant.

Combine ROC Curves



ROC Curve Summary

	AUC	Std. Error	95% Confidence Interval Lower	95% Confidence Interval Upper	p
taille	0,6610	0,03254	0,5972	0,7247	< ,001

L'aire sous la courbe vaut 0.66 et est significativement différente de 0.5, mais ce nombre dans l'absolu est assez faible (exiger > 0.80). En prenant quelques seuil (ou cut-off), on obtient le sensibilité, la spécificité, les vrais positifs (TP true positive), la valeur prédictive positive (PPV), les biens classés (accurarcy) et un indice de Younden (Se + Sp – 100). C'est à la taille de 1 m 61 que le seuil paraît intéressant.

taille

Cut-off Coordinates - taille

Cut-off	Sensitivity	Specificity	TP	PPV	Accuracy	Youden's Index
1,470	1,41 %	100,00 %	2	100,00 %	50,88 %	0,01408
1,615	59,86 %	70,63 %	85	66,93 %	65,26 %	0,30489
1,665	78,87 %	41,96 %	112	57,44 %	60,35 %	0,20831
1,745	95,77 %	5,59 %	136	50,18 %	50,53 %	0,01369

Note. Positive if test ≤ cut-off (direction chosen to maximize AUC).

Contingency Table - taille

		issue		
		CS	VB	Total
taille	≤1.615	85	42	127
	>1.615	57	101	158
	Total	142	143	285

Note. Based on optimal cut-off (Youden's Index)

La table avec un seuil de 1 m 615 permet de vérifier les différents indices.

Diagnostic Accuracy - taille

		95% Confidence Interval	
	Result	Lower	Upper
Sensitivity	59.86 %	51.31 %	67.99 %
Specificity	70.63 %	62.44 %	77.94 %
Positive Likelihood Ratio	2,0381	1,5286	2,7173
Negative Likelihood Ratio	0,5683	0,4529	0,7131
Prevalence	49.82 %	43.87 %	55.78 %
Positive Predictive Value	66.93 %	60.28 %	72.96 %
Negative Predictive Value	63.92 %	58.54 %	68.98 %
Accuracy	65.26 %	59.42 %	70.78 %

Chez les femmes "patites" (taille < 1 m 615), la VPP de 66.93 % permet de dire que l'on peut prédire 66.93 % de césarienne, tandis que pour les femmes "grandes" la probabilité de césarienne est de 100 – VPN = 36.08 %. Les bien classés (accuracy) sont 85 + 101 = 186 = 65.26 %.

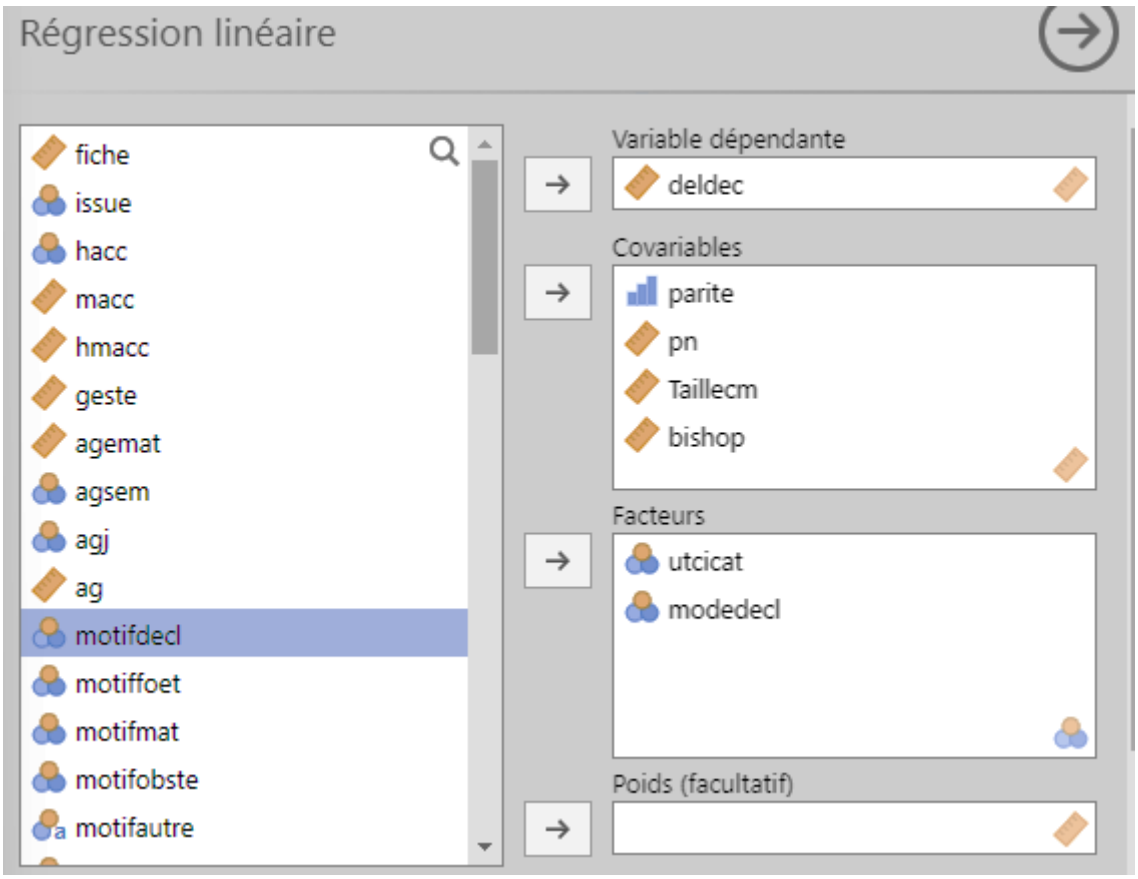
Analyse multivariée

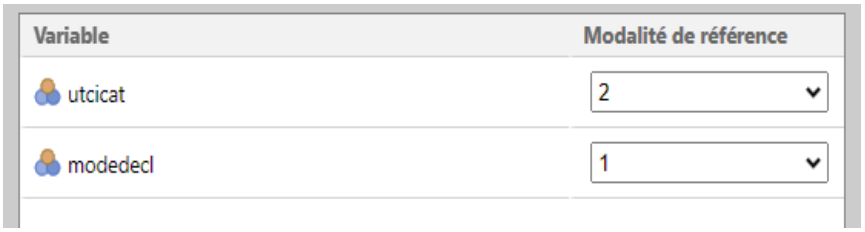
Le critère de jugement est quantitatif : Régression linéaire multiple

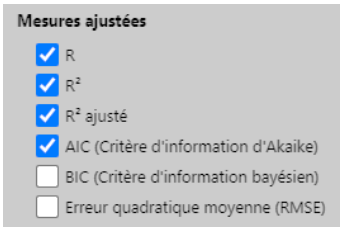
On se propose d'analyser les facteurs en lien avec le délai d'accouchement (temps en heures entre le moment du déclenchement et le moment de l'accouchement) appelé `deldec`. **On analyse ici des facteurs d'exposition quantitatifs** comme le poids du nouveau-né, la parité, la taille de la mère ou le bishop, mais aussi, sous certaines conditions, des variables nominales comme utérus cicatriciel ou le mode de déclenchement.



Cliquer sur **Analyses**, puis sur *Régression*, et *Régression linéaire*. La variable dite dépendante (à expliquer) est la variable `deldec`, et on introduit des covariables – quantitatives – et des facteurs de type nominal. Pour éviter la colinéarité – c'est-à-dire le lien trop fort entre variables – il faut éviter d'introduire des variables trop liées comme âge gestationnel et poids de naissance ou taille et IMC par exemple (il faut choisir). Dans l''exemple, le lien entre score de bishop et mode de déclenchement est fort :à explorer pour n'en garder qu'un seul +++







Pour les variables nominales, choisir la valeur de référence et cocher le R² pour évaluer la part explicative du modèle retenu. Dans ce tutoriel, il n'est pas prévu d'étudier les interactions entre variables telles que `utcicat*modedecl` par exemple (les retirer de la proposition systématique).

Régression linéaire

Mesures de l'ajustement du modèle

Modèle	R	R^2	R^2 ajusté	AIC (Critère d'information d'Akaike)
1	0,6830	0,4665	0,4527	2140

Note. Models estimated using sample size of N=279

Le R² ajusté sur le modèle est à 0.45, soit 45 % de la variabilité du délai d'accouchement est expliqué par toutes les variables retenues.

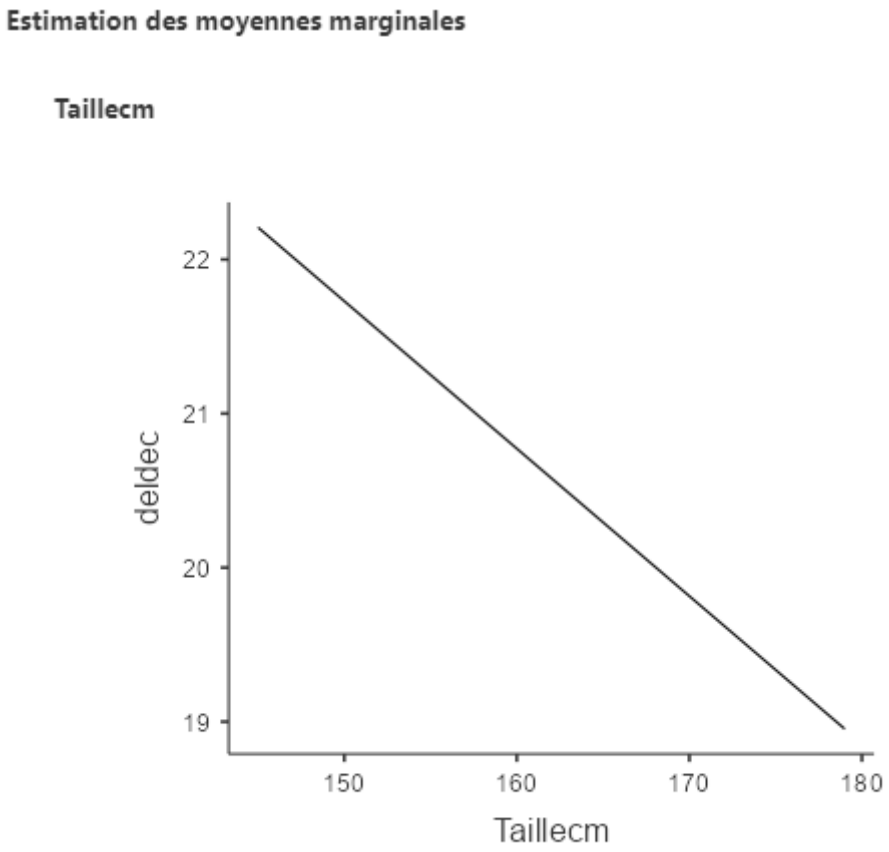
Coefficients du modèle - deldec

Prédicteur	Estimation	Erreur standard	t	p
Ordonnée à l'origine[a]	45,450843	16,397691	2,7718	,006
parite	−2,337942	0,719808	−3,2480	,001
pn	0,002080	0,001168	1,7809	,076
Taillecm	−0,095726	0,101898	−0,9394	,348
bishop	−2,736212	0,543637	−5,0332	< ,001
utcicat:				
2 – 1	−0,386855	2,829102	−0,1367	,891
modedecl:				
2 – 1	−7,157137	2,204748	−3,2462	,001
3 – 1	12,622427	2,502260	5,0444	< ,001

[a] Représente le niveau de référence

Il faut regarder non seulement la "p-value" mais aussi le sens des coefficients (estimation). Quand le coefficient est négatif comme pour la parité, la taille de la mère et le bishop, cela veut dire que plus le délai d'accouchement est long, moins la parité, la taille ou le score de Bishop sont importants (intuitif). On peut dire également que pour chaque cm de taille de la mère, le délai d'accouchement diminue de 0.09 heure (soit 5 minutes…) sous une hypothèse linéaire (hypothèse non traitée dans ce tuto).

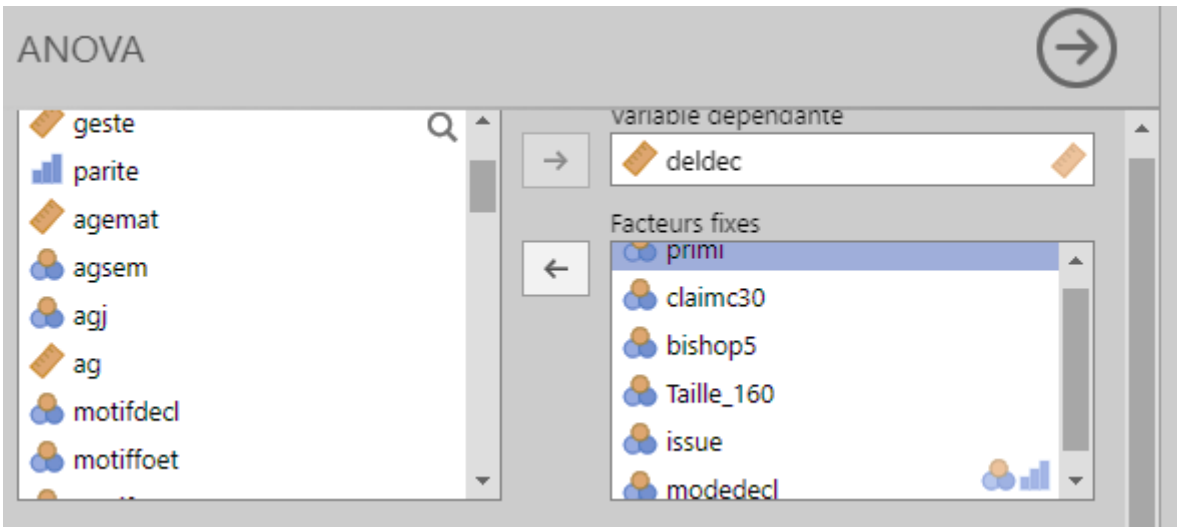
Pour les variables nominales, c'est plus difficile à interpréter selon la valeur de référence : pour l'utérus cicatriciel, la valeur "1" (présent) est un facteur qui diminue la durée de l'accouchement (intuitif aussi).



Le graphe précédent montre **qu'en moyenne** et en tenant compte des autres variables, le délai d'accouchement diminue quand la taille de la mère augmente.

Le critère de jugement est quantitatif : ANOVA multivariée

On analyse ici des **variables d'exposition qualitatives** (ordinales ou nominales)



ANOVA

ANOVA - deldec

	Somme des carrés	ddl	Carrés moyens	F	p
primi	1255,9	1	1255,9	10,323	,001
claimc30	155,2	1	155,2	1,276	,260
bishop5	759,3	1	759,3	6,241	,013
Taille_160	294,2	1	294,2	2,418	,121
issue	712,1	1	712,1	5,853	,016
modedecl	6414,2	2	3207,1	26,360	< ,001
Résidus	32727,8	269	121,7		

Sont significatifs : primipares, bishop5 (deux codages bishop< 5 et bishop≥5), issue (CS et VB), et le mode de déclenchement, mais pas la taille avec un seuil de 1 m 60.

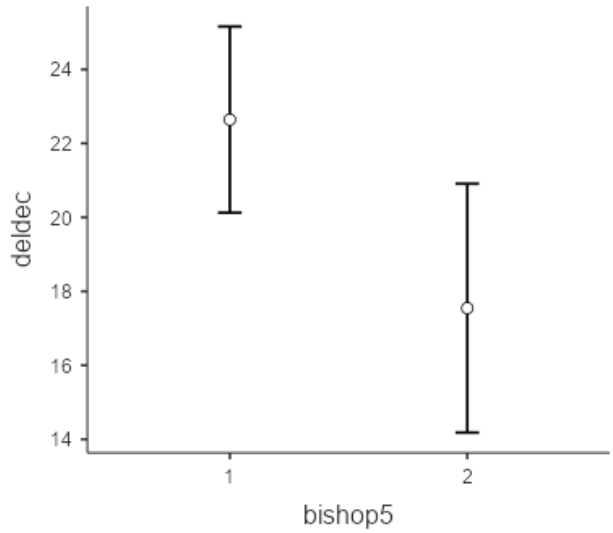
Tests post hoc

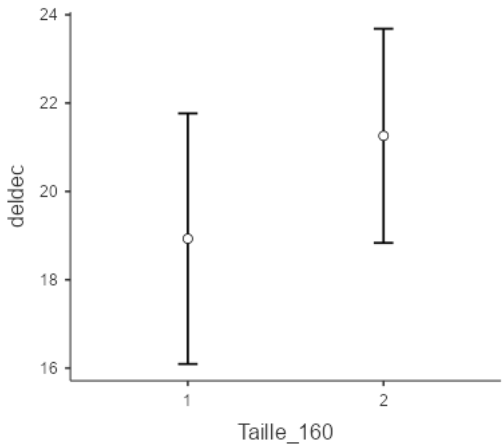
Comparaisons post hoc - modedecl

Comparaison			Différence moyenne	Erreur standard	ddl	t	Ptukey	Pbonferroni
modedecl		modedecl						
1	-	2	8,249	2,203	269,0	3,745	< ,001	< ,001
	-	3	−12,516	2,516	269,0	−4,975	< ,001	< ,001
2	-	3	−20,765	2,860	269,0	−7,260	< ,001	< ,001

Note. Comparaisons basées sur les moyennes marginales estimées

Entre les 3 modes de déclenchement, les différences sont toutes significatives : + 8 heures pour les prosta, - 12 heures pour l"ocytocine et – 20 heures pour les deux (intuitif).



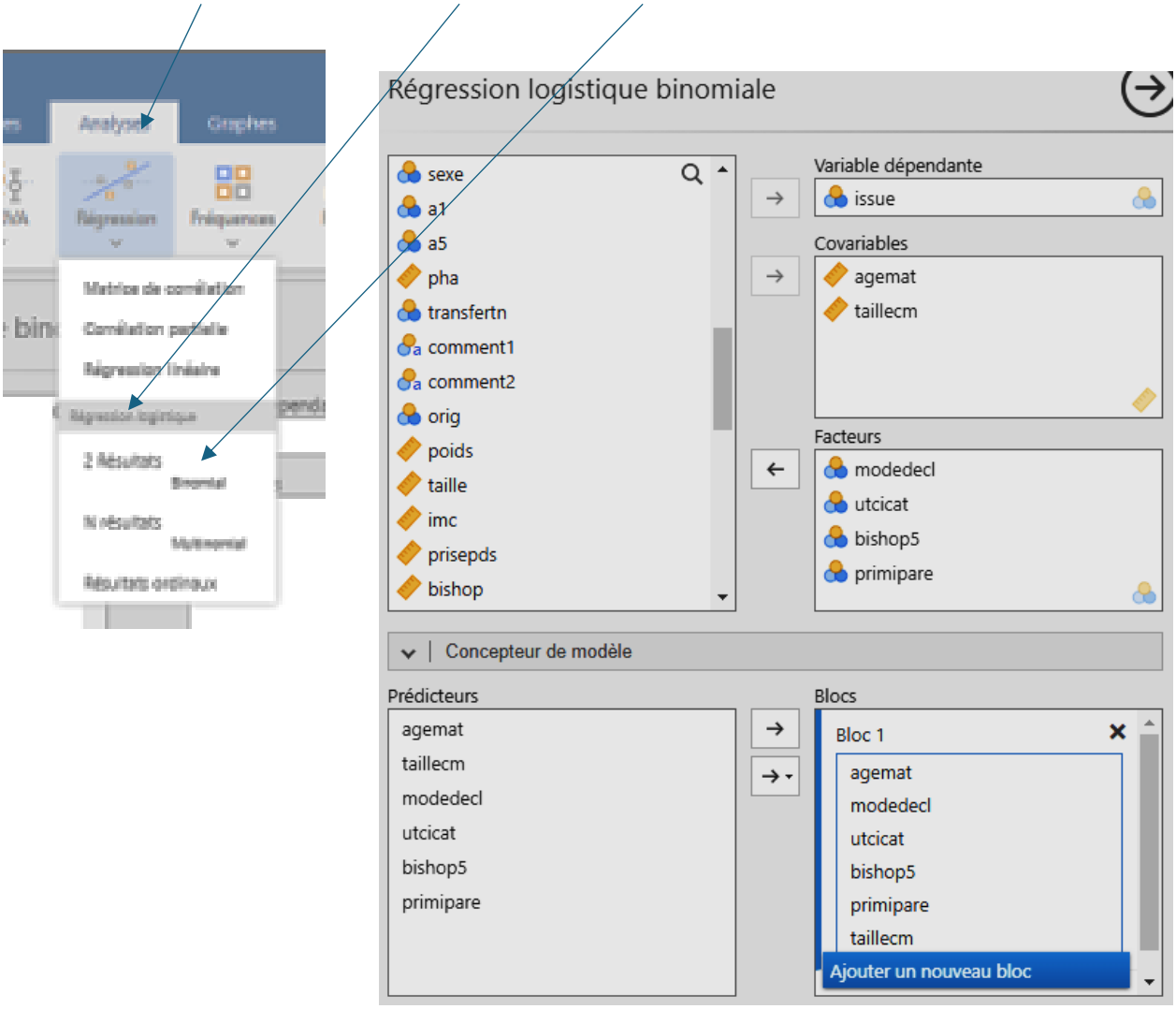


Ces deux graphes d'exemple montrent les différences de délais d'accouchement (en heures) selon le bishop ou la taille de la mère en tenant compte des autres facteurs.

Le critère de jugement est qualitatif (nominal) binaire: Régression logistique

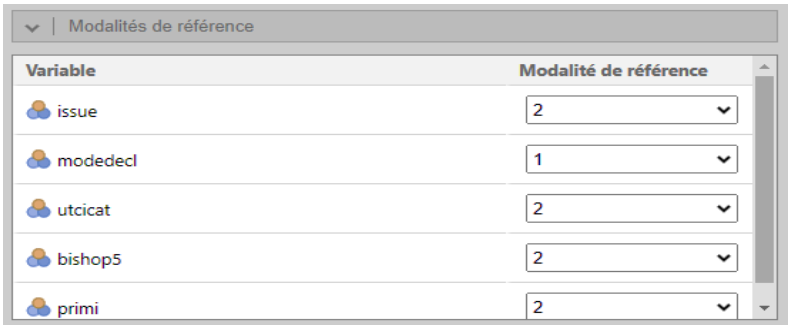
On se propose d'analyser les facteurs qui sont liés à l'issue en césarienne (`issue`=1) ou en voie basse (`issue` = 2). On raisonne en facteurs liés à la CS, c'est-à-dire 1 par rapport à 2.

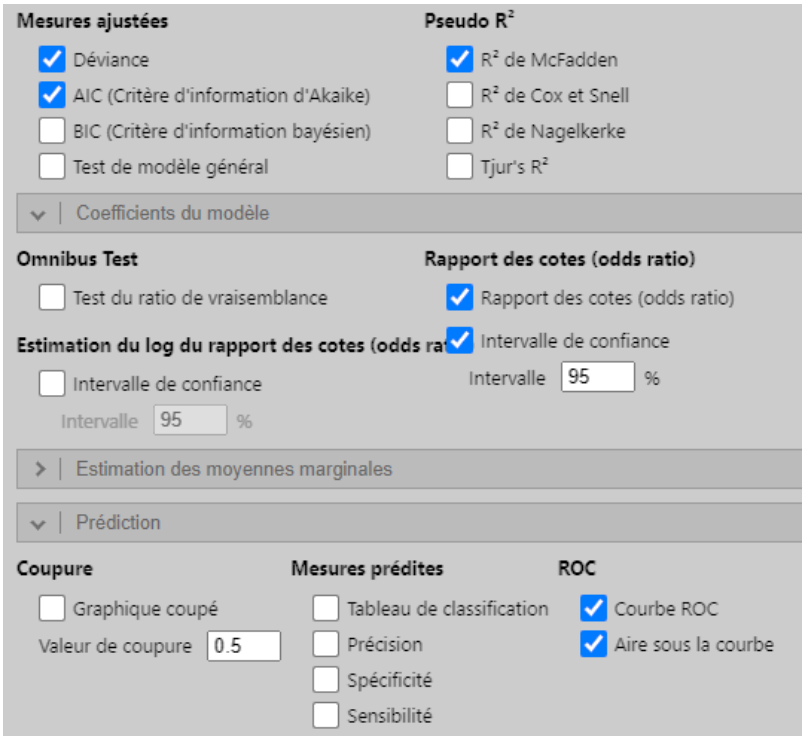
Cliquer sur **Analyses**, puis *Régression logistique*, puis *2 résultats/ Binomial*



On introduit le critère de jugement comme variable dépendante, ainsi que les variables quantitatives comme covariables et les variables nominales comme facteurs explicatifs.

On coche ensuite les R^2 et les odds ratio avec leurs intervalles de confiance à 95 %. On peut aussi évaluer le caractère prédictif du modèle retenu avec la courbe ROC et sa valeur sous la forme d'aire sous la courbe (Aera Under Curve AUC).





Coefficients du modèle - issue

Prédicteur	Estimation	Erreur standard	Z	p	Rapport des cotes (odds ratio)	Intervalle de confiance à 95%	
						Borne inf	Supérieur
Ordonnée à l'origine	11,59303	3,92623	2,9527	,003	108339,5158	49,2920	2,381e+8
agemat	0,01823	0,03045	0,5985	,549	1,0184	0,9594	1,0810
modedecl:							
2 – 1	−0,06994	0,46597	−0,1501	,881	0,9324	0,3741	2,3241
3 – 1	0,58723	0,56910	1,0319	,302	1,7990	0,5897	5,4884
utcicat:							
1 – 2	2,70353	0,67343	4,0146	< ,001	14,9324	3,9893	55,8928
bishop5:							
1 – 2	1,50621	0,45012	3,3463	< ,001	4,5096	1,8664	10,8963
primipare:							
Primi – Multi	2,15307	0,37830	5,6914	< ,001	8,6113	4,1026	18,0751
taillecm	−0,08867	0,02339	−3,7909	< ,001	0,9151	0,8741	0,9581

Note. L'estimation représente le log des cotes (odds) de "issue = CS" vs. "issue = VB"

Les odds ratio des facteurs significatifs (p<0.05) sont faciles à interpréter : pour le bishop < 5 versus bishop ≥ 5, l'OR ajusté (ORa) vaut 4.51 (0.87 – 1.89 en arrondis à 2 décimales) soit le bishop bas est 4.51 fois plus associé à la CS qu'à la VB. Idem pour primipare (ORa = 8.61), et utérus cicatriciel (ORa = 14.93).

Pour les facteurs introduits en quantitatif (covariables), le raisonnement est plus complexe : si l'ORa > 1, comme l'âge maternel en années, chaque année d'âge augmente le risque de césarienne de 0,018, soit 1.8 %, tandis que pour la taille de la mère en cm, avec un ORa < 1, chaque cm de taille diminue le risque de césarienne de (1 – 0.91 = 0.09) soit 9 %.

Note sur la performance du modèle retenu :

Mesures de l'ajustement du modèle

Modèle	Déviance	AIC (Critère d'information d'Akaike)	R^2_{McF}
1	278,8	294,8	0,2944

Note. Models estimated using sample size of N=285

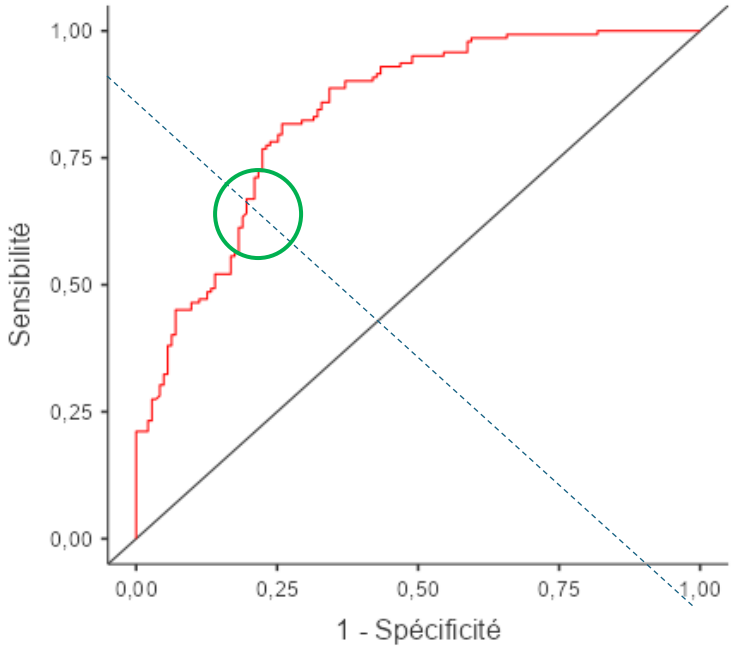
La déviance est calculée à partir de la "vraisemblance" du modèle (Likelihood en anglais), c'est-à-dire la capacité du modèle (de l'équation logistique) à être au plus près des données réelles. Comme c'est une probabilité faible et entre 0 et 1, on transforme le calcul de la vraisemblance (V) selon la formule :

Déviance = - 2 * Ln (V).

Plus le résultat est faible, plus le modèle est adéquat : on peut ainsi comparer plusieurs modèles avec des variables différentes ou des codages différents (variable quantitative au lieu de qualitative par exemple). La déviance suit une loi de χ^2.

AIC (Akaike Information Criterion) suit la même logique selon la formule : AIC = 2*k – 2*Ln(V) avec la même interprétation, k étant le nombre de paramètres. Dans d'autres logiciels, on trouve aussi le BIC (Bayesian Information Croterion) : BIC = k*Ln(n) – 2*Ln(V), n étant le nombre d'observations.

Courbe ROC



Prédiction

Mesures prédites
Aire sous la courbe
0,8374

Note. La valeur de coupure est réglée sur 0.5

L'ensemble des critères retenus séparent bien les facteurs prédictifs de CS et VB avec une aire sous la courbe de 83.7 % ce qui "mathématiquement" important.

Le logiciel propose d'explorer la taille qui sépare le mieux les CS des VB : il propose la valeur avec un Se de ≈ 75 % et une Sp de ≈ 1 – 0.25 = 75 % aussi. Le tableau ci-dessous permet de se rendre compte des enjeux d'une décision a priori des modes d'accouchement : lorsque l'on prédit une VB, la femme accouche par VB dans 107 cas, soit 74.83 % des femmes prédites, mais 36 ont une césarienne au cours du travail (c'est un moindre mal). A l'opposé, quand on prédit une césarienne (l'idée est que la CS pourrait être programmée sans subir le travail) 113 femmes ont bien une césarienne, soit 79.58 % des femmes prédites, mais 29 ont quand même une voie basse ! Ce serait très "embêtant" de faire d'emblée une césarienne alors que quelques femmes finiraient d'accoucher par VB…

Tableau de classification – ...

	Prédit		
Observé	VB	CS	% correct
VB	107	36	74,83
CS	29	113	79,58

Note. La valeur de coupure est réglée sur 0.5

Mesures prédites

Précision	Spécificité	Sensibilité	Aire sous la courbe
0,7719	0,7483	0,7958	0,8374

Note. La valeur de coupure est réglée sur 0.5

Les onglets et modules supplémentaires qui peuvent servir

Onglets par défaut

En italiques, les lignes utilisées dans ce tutoriel



- **Exploration** : *Statistiques descriptives* et *scatter plot*. Pareto plot (graphes en barre des causes et courbe cumulative utilisé en études de qualité ou de performance selon la règles des 80-20 %)



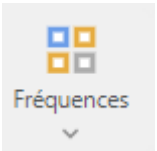
- **Test-T** : pour *deux échantillons indépendants*, pour des échantillons appariés (en pharmacologie par exemple), et test T pour un seul échantillon / 0



- **ANOVA** : *ANOVA simple*, ANOVA pour mesures répétées, MANOVA et ANCOVA pour des analyses multivariées. Equivalent d'une ANOVA en non-paramétrique pour échantillons indépendants (Kruskall-Wallis) ou à mesures répétées (Friedman).



- **Régression** : *Corrélations simples* et partielles, *R linéaire* et *R logistique binomiale* et multinomiale ou ordinale (critère de jugement à plus de 2 modalités)



- **Fréquences** : tests pour un échantillon, *tests pour tableau 2x2 avec χ^2 pour échantillons indépendants* ou appariés (test de MacNemar), régression log-linéaire (analyses de plusieurs variables qualitatives ensemble)



- **Facteur** : analyse exploratoire des liens entre les variables. Etude de qualité de vie…..

Modules possibles

Avec la version 2.7, 54 modules sont disponibles : ont été présentés les modules comme **ROC Analysis** ou **Survival** ou **Rj** pour ouvrir le logiciel R avec des lignes de commande. L'auteur de ce tutoriel ne les a pas tous testés… Certains analysent les données téléchargées, d'autres permettent l'entrée de statistiques déjà constituées. On peut remarquer :

- Module pour les **tests non-paramétriques**
- Méthode de **Bland et Altman** pour comparer deux techniques de mesure (par exemple glycémie au doigt et glycémie au laboratoire).
- **Méta-analyse** pour faire la synthèse d'études similaires
- **Randomizer** pour tirage au sort dans les études (voir aussi https://www.openepi.com/Menu/OE_Menu.htm assez facile sur le Net)
- Plusieurs modules de graphes (Plots) dont **jjstatplots** et **flexplots**
- **bmTests** pour des statistiques dites robustes, non-paramétriques
- **esci** : statistiques avec intervalle de confiance
- Calcul de **puissance** (**power**) et de taille d'échantillon

Table des matières

INTRODUCTION **3**

TELECHARGER JAMOVI (TOUTES PLATES-FORMES) **3**

LE FICHIER D'EXEMPLE 4

VALIDER LES DONNEES **6**

- FORMATS DES VARIABLES 6

- DONNER DES NOMS DE LABELS AUX VARIABLES NOMINALES (CODEES) 7

- SELECTIONNER DES LIGNES (CERTAINES VALEURS DE VARIABLES) 8

- TRANSFORMER DES VARIABLES 8

Recoder une variable vers une autre nouvelle *8*

Créer une variable à partir d'autres variables : calcul IMC = poids/taille² *9*

Créer une variable par différence de temps (heures, jours ou années) *9*

Etablir des seuils pour transformer une variable quantitative en qualitative *10*

DECRIRE LA POPULATION, DECRIRE LES VARIABLES **12**

- VARIABLES NOMINALES OU ORDINALES ➔ EFFECTIFS ET FREQUENCES 12

- VARIABLES D'ECHELLE (QUANTITATIVES) ➔ VALEURS CENTRALES ET DISPERSION 14

- GRAPHES DESCRIPTIFS 16

COMPARER LES DONNEES – TESTS STATISTIQUES **17**

- LA VARIABLE DE JUGEMENT EST QUALITATIVE A DEUX MODALITES 17

La ou les variables d'exposition sont qualitatives : tableau 2 x 2 ou 2 x n *17*

La ou les variables d'exposition sont quantitatives *20*

- LA VARIABLE DE JUGEMENT EST QUANTITATIVE 23

La ou les variables d'exposition sont qualitatives à deux modalités *23*

La ou les variables d'exposition sont qualitatives à plus de deux modalités *23*

La ou les variables d'exposition sont quantitatives *25*

AUTRES METHODES STATISTIQUES **28**

COURBES DE SURVIE. 28

Survie globale, simple *28*

"Survie" (durée accouchement) selon la primiparité (/ Multiparité) – Log-Rank *30*

Modèle de Cox avec variables nominales et continues. Notion de HR *32*

ANALYSE DES TESTS DIAGNOSTIQUES OU DE DEPISTAGE – COURBES ROC 33

ANALYSE MULTIVARIEE **36**

LE CRITERE DE JUGEMENT EST QUANTITATIF : REGRESSION LINEAIRE MULTIPLE 36

LE CRITERE DE JUGEMENT EST QUANTITATIF : ANOVA MULTIVARIEE 38

LE CRITERE DE JUGEMENT EST QUANTITATIF (NOMINAL) BINAIRE: REGRESSION LOGISTIQUE 40

LES ONGLETS ET MODULES SUPPLEMENTAIRES QUI PEUVENT SERVIR **45**