



SEED

STATISTIQUE, ÉVALUATION
ÉCONOMIQUE, DATA-MANAGEMENT

Notice explicative du site pvalue.io à destination des internes du CHU de Lille

Version Mars 2023

Contact : biostatistique@chu-lille.fr

Table des matières

Etape 1 – Conformité de la base de données	3
Etape 2 – Importation de la base de données.....	4
Etape 3 – Définition des données.....	8
Etape 4 –Analyses statistiques	10
A – Description de l'ensemble de la population.....	10
B – Description +/- comparaison de la population selon des groupes.....	11
Etape 5 – Synthèse des résultats.....	14
Etape 6 – Interprétation des résultats	15
A. Descriptif global	15
B. Descriptif par groupe et comparaison (Analyses Univariates)	15
Paragraphe statistique	16
Pour aller plus loin.....	16

Etape 1 – Conformité de la base de données

Avant toute analyse, assurez-vous d'avoir un fichier conforme aux recommandations qui vous ont été fournies lors du contact avec l'unité de méthodologie – biostatistique.

Dirigez-vous sur le site : www.pvalue.io

Une fois votre inscription réalisée, la page d'accueil suivante s'ouvre :

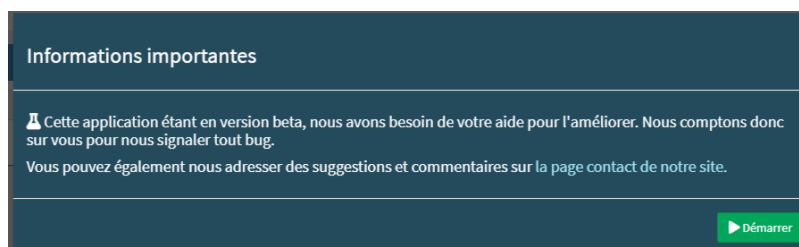


Si c'est votre première connexion sur le site, il vous sera demandé une adresse mail et un mot de passe. Aucune autre information ne sera nécessaire.

Etape 2 – Importation de la base de données

Sur la page d'accueil, cliquez sur « démarrer l'analyse » ; un chargement s'affichera. Patientez.

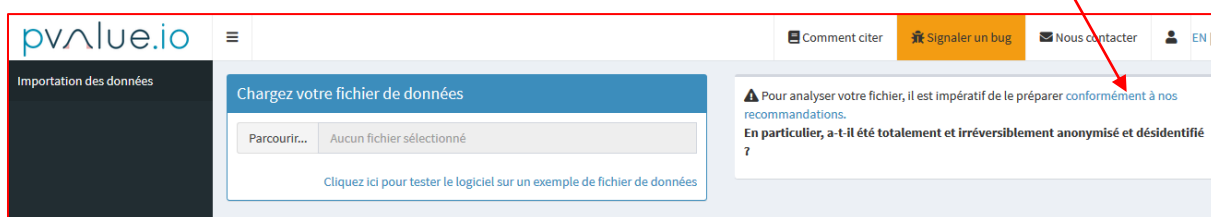
Vous devriez arriver sur une page avec un pop-up comme celui-ci :



Le site évoluant, certaines fonctionnalités non présentées dans ce document sont déconseillées. Pour les tests décrits dans la suite de ce document, les résultats ont été testés et validés.

IMPORTANT : De nouveau, avant le téléchargement de votre base de données sur le site, assurez-vous que votre fichier de données respecte **bien les recommandations** de l'application, sinon la qualité de votre analyse pourrait être compromise.

- Cliquez, à droite, sur « conformément à nos recommandations ».



En cliquant sur « conformément à nos recommandations », vous aurez une page listant 14 points à respecter avant le chargement de votre base de données. Nous allons vous les détailler.

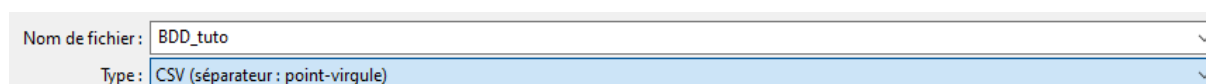
1. Si vous souhaitez réaliser une analyse de survie et que votre fichier comporte des dates, exportez-le en format Texte Unicode.

-> Les analyses de survie ne seront pas présentées dans ce document car ces analyses sont prises en charge par notre service. Mais si vous avez des dates dans votre fichier il est important de respecter le format JJ/MM/AAAA (exemple : 03/05/1999; 25/07/2006). Sinon **il est même préférable de supprimer toutes les dates de votre fichier** à importer.

2. et 3. Faites une copie de sauvegarde votre fichier. / Ouvrez votre fichier avec Microsoft Excel ou LibreOffice Calc

-> Votre fichier ne sera jamais modifié directement par le site, mais il reste toujours préférable d'enregistrer une copie de votre base de données. De préférence, privilégiez le format csv.

- Pour cela, ouvrez votre fichier dans Excel/Libre office, cliquez sur Fichier -> Enregistrer sous puis choisissez le format csv. Exemple ci-dessous.



4. Tous les patients doivent être dans le même tableau, sans aucune ligne vide.

-> Aucune ligne vide car chaque ligne est prise en compte.

5. Une ligne par patient, une caractéristique par colonne.

-> Faites attention aux commentaires laissés en fin de tableau et/ou aux calculs faits en bas de fichier (calcul d'effectif, moyenne, etc). Chaque colonne doit apporter une seule et unique information (exemple: évitez d'indiquer dans la même cellule la PAS et PAD comme suit : 127/79, il faut faire 2 colonnes distinctes. Evitez aussi les horaires/durée du type 11:35, 20:36, convertissez la colonne en heures OU en minutes).

6. Une ligne ne doit correspondre qu'à un patient.

-> Si des patients ont des mesures à différentes visites, créez autant de colonnes que de visites réalisées (exemple : FC_J1, FC_J2, FC_M3, FC_M6).

7. Les colonnes doivent avoir un intérêt pour l'analyse.

-> Au moment de la copie au format csv de votre base de données, profitez-en pour retirer les colonnes qui ne vous seront pas utiles pour les analyses. Vous gagnerez du temps pour les prochaines étapes.

8. Le fichier est irréversiblement anonymisé et désidentifié.

-> Pas de noms, prénoms, y compris les initiales, et éviter au maximum les dates. Rien ne doit permettre de retrouver un patient de votre fichier.

9. La première ligne doit être le nom de la variable.

-> La première ligne doit correspondre au nom des colonnes et la deuxième ligne à votre premier patient.

Comme présentez ici :

	A	B	C	D	E	F	G	H	I	J	K	L
1	DLR_PERINEE	ETAT_PERINEE	PRESENCE_ERAILLURE	PROFESSIONNEL	ATCD_DECHIRURE	NB_JOUR_TT	POIDS_NAISSANCE	DUREE_EFFORT	DELAI_DCEE	DUREE_TRAVAIL	AGE	BMI
2	1	0	1	4	0	5	3260	4	50	6	35	23
3	1	1	0		1	3	3600	2	0	2.5	33	
4	1	1	0		0	2	3920	6	120	4.5	30	30

10. Toutes les colonnes ont un titre unique.

-> Renommez ou supprimez les colonnes possédant le même nom.

11. Aucune colonne n'est le contraire d'une autre.

-> Si une variable a deux valeurs possibles (exemple : Sexe qui vaut Homme/Femme), inutile de faire une colonne Homme 0(non)/1(oui) et une colonne Femme 0(non)/1(oui), puisque le logiciel vous donnera l'effectif et pourcentage pour les deux modalités de la variable, donc si on conserve la colonne Homme, le nombre de 1 correspondra au nombre d'hommes dans votre population et le nombre de 0 correspondra au nombre de femmes.

12. Aucune unité de mesure ni de pourcentage dans une colonne numérique

-> N'indiquez rien d'autre que des valeurs numériques, si vous avez pour une même colonne différentes unités pensez à toutes les transformer en la même unité. En cas de valeurs indétectables sous un certain seuil, supprimez juste les signes "<" ou ">" qui ne seront pas reconnus et empêcheront le descriptif de la variable, conservez uniquement la valeur du seuil.

13. Les valeurs manquantes sont des cellules vides.

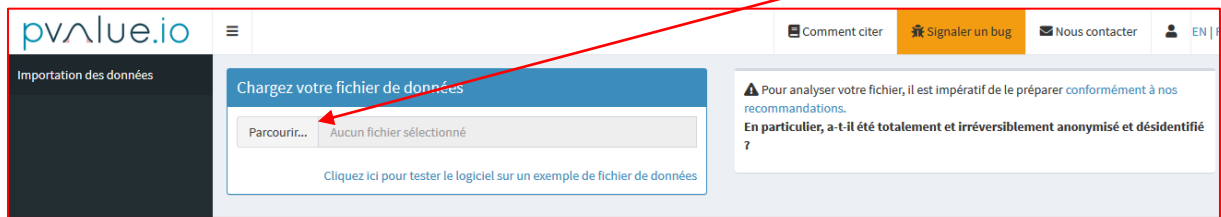
-> Pas de "0", "99...", "NA", "ND", "NR", "x" ou tout type de caractères spéciaux.

14. Les colonnes non numériques contiennent peu de valeurs différentes.

-> Si une variable possède un grand nombre de modalités différentes, par exemple 10 types de chirurgie, il n'est pas réaliste de faire un test statistique avec autant de modalités. Soit cette variable ne sert qu'à être décrite et dans ce cas on n'y touche pas, soit il faut recoder la variable en regroupant des modalités qui peuvent l'être, afin de réduire le nombre de modalités.

Remarque : Si une variable qualitative ne possède qu'une seule modalité, elle risque de ne pas être détectée par le site, il est préférable de la retirer de votre base de données et de la décrire à part par vous-même.

Une fois les recommandations respectées, vous pouvez désormais cliquer sur « Parcourir » afin de charger votre fichier.



Ici nous téléchargeons une base de données appelée « BDD_tuto.csv » :

Chargez votre fichier de données

Parcourir... BDD_tuto.csv

Upload complete

⚠ Pour analyser votre fichier, il est impératif de le préparer conformément à nos recommandations.
En particulier, a-t-il été totalement et irréversiblement anonymisé et désidentifié ?

Vos données

Afficher 5 éléments

	DLR PERINEE	ETAT PERINEE	PRESENCE ERAILLURE	PROFESSIONNEL	ATCD DECHIRURE	NB JOUR TT	POIDS NAISSANCE	DUREE EFFORT	DELAI DCEE
1	1	0	1	4	0	5	3260	4	50
2	1	1	0		1	3	3600	2	0
3	1	1	0		0	2	3920	6	120
4	1	0	0	3	0	5	3090	7	10
5	0	0	0	3	1	1	4180	21	60

Affichage de l'élément 1 à 5 sur 119 éléments

Précédent 1 2 3 4 5 ... 24 Suivant

1- Vérifiez que le nombre de patients est correct. Pour notre exemple, nous avons $n = 119$ patients.

2- Vérifiez que toutes vos colonnes ont été chargées. Pour cela, utiliser la barre de déplacement latérale.

Etape 3 – Définition des données

Par la suite, plusieurs fenêtres vont s'afficher. Le site va vous demander d'indiquer le type de données que vous lui fournissez. **Cette étape est importante**, les descriptifs et tests statistiques appliqués à vos données dépendront de ces informations.

Colonnes ambiguës

Les colonnes suivantes ne contiennent que des nombres entiers. Les nombres entiers peuvent avoir deux usages : être considérés comme tels (par exemple un nombre d'adénopathies) ou bien être utilisés comme des classes recodées (par exemple pas d'anomalie = 0, anomalie cérébrale = 1, anomalie cardiaque = 2, etc.). Par défaut, si la variable comporte 4 valeurs possibles ou moins, ces valeurs sont traitées comme des classes codées numériquement, et comme des nombres dans le cas contraire.

Ce comportement peut être aberrant dans certains cas, c'est pourquoi, **vous devez vérifier manuellement toutes les colonnes suivantes**.

💡 Si cela a du sens de dire que par exemple la moyenne ≈ 2.4 , alors choisir "Colonne numérique", sinon, choisir "Classes codées numériquement".

NB JOUR TT : 6 valeurs possibles (0, 1, 2, 3, 4, 5) Colonne numérique

PROFESSIONNEL : 4 valeurs possibles (1, 2, 3, 4) Classes codées numériquement

Pour les variables codées avec des chiffres entiers (0, 1, 2, etc.), indiquez s'il s'agit de colonnes numériques (par exemple, nombre de médicaments) ou s'il s'agit de classes codées numériquement (par exemple, Oui/Non ou Type de chirurgie).

Pour notre exemple, on indique que la variable *NB JOUR TT* est numérique et *PROFESSIONNEL* est une variable de classes codées numériquement.

Si vous avez beaucoup de données manquantes, une fenêtre vous indiquera les variables pour lesquelles le nombre de données manquantes est jugé trop important.

Colonnes avec trop de valeurs manquantes

Les colonnes suivantes sont trop souvent vides (à l'origine de plus de 20% de données manquantes). Vous pourrez les utiliser dans la plupart des cas si vous effectuez des **analyses descriptives**.

Si vous souhaitez réaliser des **analyses explicatives ou prédictives**, elles ne pourront pas être utilisées sur [pvalue.io](#). Si vous avez vraiment besoin de ces données, cela nécessite une expertise statistique ([nous contacter](#)).

PROFESSIONNEL

Dans notre base de données, la variable *PROFESSIONNEL* contient 55 données manquantes, soit 46%. Elle peut être gardée pour la suite mais des tests statistiques dessus seront à éviter.

Un récapitulatif des données non numériques vous sera fourni, vérifiez bien qu'aucune variable ne manque.

Colonnes non numériques

Les colonnes suivantes seront traitées comme non numériques. Cela peut être dû au fait qu'elles contiennent des caractères alphabétiques ou qu'elles ont peu de valeurs différentes (moins de 5). Veuillez vérifier que le nombre et le libellé des valeurs possibles sont corrects.

ATCD DECHIRURE : 2 classes (0 et 1)

DLR PERINEE : 2 classes (0 et 1)

ETAT PERINEE : 2 classes (0 et 1)

PRESENCE ERAILLURE : 2 classes (0 et 1)

PROFESSIONNEL : 4 classes (1, 2, 3 et 4)

Les variables traitées numériquement seront listées à la suite

Colonnes numériques

Les colonnes suivantes ont été reconnues comme numériques.

AGE

BMI

DELAI DCEE

DUREE EFFORT

DUREE TRAVAIL

NB JOUR TT

POIDS NAISSANCE

Si des dates sont présentes dans votre fichier, la dernière fenêtre vous les indiquera :



Dans notre exemple, aucune date n'est présente.

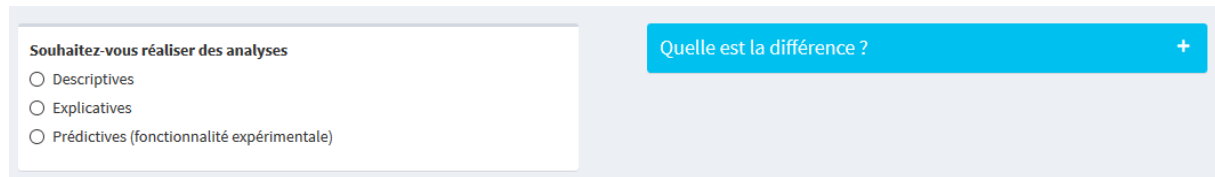
Si les étapes précédentes ont bien fonctionné, vous devriez avoir vu l'ensemble de vos colonnes présentes dans votre base. Dans notre exemple, nous avons:

- 5 colonnes non-numériques
- 7 colonnes numériques
- 0 date

Les (5+7+0) 12 colonnes de notre fichier sont bien définies.

Etape 4 –Analyses statistiques

Après avoir cliqué sur le bouton **Commencer les analyses statistiques**, une page s'ouvre et demande à l'utilisateur de renseigner le type d'analyse souhaité.



En cliquant sur le bouton **+**, une explication des différentes options vous est donnée. Dans la suite de ce tuto, seules les analyses "Descriptives" seront abordées. Les analyses explicatives et prédictives sont des analyses prises en charge par l'Unité SEED.

L'option "analyses descriptives" permet de :

- **faire une analyse descriptive globale** : toutes les variables de la base (ou une sélection de variables) seront décrites pour l'ensemble des sujets.

Par exemple : l'âge moyen de l'ensemble des patients accompagné de son écart-type ; le nombre et le pourcentage de femmes

- **faire des analyses descriptives par groupe** : toutes les variables de la base (ou une sélection de variables) seront décrites dans chacun des groupes d'intérêt.

Par exemple : si la variable groupe est le statut malade/non malade, on aura l'âge moyen (+/- écart-type) des patients malades ainsi que l'âge moyen (+/- écart-type) des patients non-malades

- **faire des tests de comparaisons de groupes** : les variables sélectionnées seront comparées entre les groupes d'intérêt avec les tests statistiques adaptés selon les types de variables (t-test de Student, test de Wilcoxon, test du Chi-2, ANOVA, etc).

Par exemple : si la variable groupe est le statut malade/non malade, le dosage en fer (variable quantitative) sera comparé entre les malades et les non malades à l'aide d'un t-test ou d'un test de Wilcoxon. La proportion de patients fumeurs (variable qualitative) sera comparée entre les malades et les non malades par un test du Chi-2 ou exact de Fisher.

A – Description de l'ensemble de la population

Je veux décrire l'ensemble de ma population (analyse descriptive globale)

- Le site pvalue.io a l'inconvénient de ne pas proposer cette option seule, il faut **obligatoirement spécifier une analyse par groupe pour obtenir le descriptif au global**.
- Suivez les instructions de la partie *Description de la population selon des groupes (variable qualitative)*, pour obtenir le descriptif global. Il suffit de renseigner n'importe quelle variable qualitative qui se trouve dans votre fichier (exemple : le sexe) ou créez-en une fictive, vous n'aurez pas à tenir compte des résultats de cette analyse par groupe si elle ne vous intéresse pas.

B – Description +/- comparaison de la population selon des groupes

Je veux décrire mes patients selon des groupes (analyse descriptive par groupe)

- Le site pvalue.io réalise un test statistique comparant chaque variable entre les groupes selon lesquels vous allez demander une description. Cependant, la comparaison de groupes n'est pas toujours pertinente ou nécessaire. Il est donc conseillé de ne pas tenir compte des résultats aux tests statistiques si vous n'en avez pas besoin.
- Si toutefois, la comparaison des variables entre vos groupes vous est utile, soyez vigilant sur les effectifs : **il est déconseillé de faire des tests statistiques sur des effectifs trop faible** (<8 sujets par groupe ou par modalité). Le site pvalue.io n'applique pas cette règle, à tort.

En cliquant sur « Descriptives », on obtient l'affichage suivant :

Cliquez ici pour effectuer des analyses en sous-groupe

Souhaitez-vous réaliser des analyses

☒ Descriptives

☐ Explicatives

☐ Prédictives (fonctionnalité expérimentale)

Quelle est la différence ? +

☐ Comparaison avant-après

Vous souhaitez comparer les caractéristiques des patients selon - ?

Dans notre exemple, nous souhaitons une description de la base de données en fonction de la présence/absence de douleur du périnée sur l'ensemble de nos variables.

Vous souhaitez comparer les caractéristiques des patients selon DLR PERINEE ?


Caractéristiques que vous souhaitez décrire (variables explicatives X)

<input checked="" type="checkbox"/> AGE	<input checked="" type="checkbox"/> DUREE EFFORT	<input checked="" type="checkbox"/> POIDS NAISSANCE
<input checked="" type="checkbox"/> ATCD DECHIRURE	<input checked="" type="checkbox"/> DUREE TRAVAIL	<input checked="" type="checkbox"/> PRESENCE ERAILLURE
<input checked="" type="checkbox"/> BMI	<input checked="" type="checkbox"/> ETAT PERINEE	<input checked="" type="checkbox"/> PROFESSIONNEL
<input checked="" type="checkbox"/> DELAI DCEE	<input checked="" type="checkbox"/> NB JOUR TT	
<input checked="" type="checkbox"/> Tout sélectionner/désélectionner		

Avant de lancer les analyses, il vous sera demandé si vous souhaitez réaliser une analyse de survie. Cochez « non » puisque seule la survenue d'une douleur nous intéresse et pas son délai de survenue. (Les analyses de survie sont des analyses particulières qui sont prises en charge par l'unité SEED)

Souhaitez-vous comparer des courbes de survie ? C'est le cas si DLR PERINEE est la mesure d'un évènement irréversible (exemple : décès, accouchement, infection par le VIH) et que vous avez mesuré le temps de suivi ou les dates d'inclusion et d'évènement. [Tutoriel vidéo.](#)

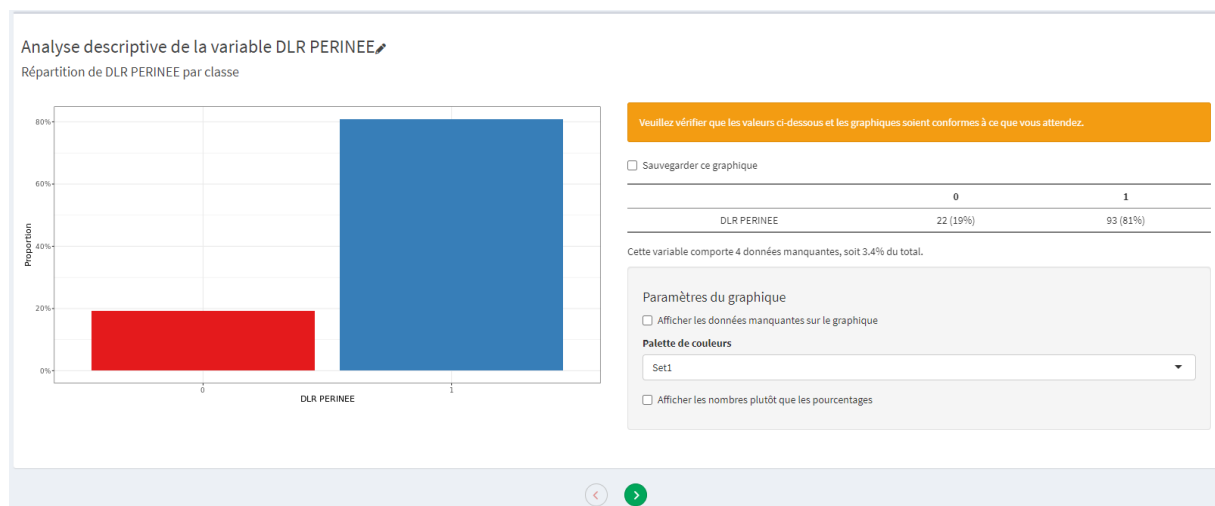
- ☒ Non, je ne souhaite pas réaliser d'analyse de survie
☐ Oui, l'évènement survient quand DLR PERINEE = 1
☐ Oui, l'évènement survient quand DLR PERINEE = 0

Cliquez ensuite sur le bouton vert 

Lorsque les analyses sont terminées, nous obtenons pour chaque variable une analyse descriptive globale ainsi qu'une analyse descriptive par groupe.

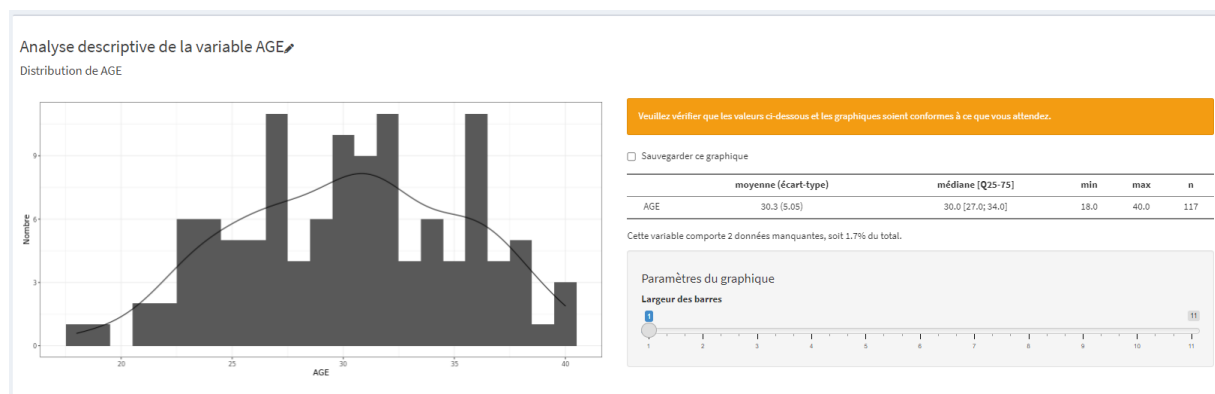
Ces analyses sont accompagnées de graphiques que l'on peut personnaliser (changement de couleur par exemple).

Les variables sont affichées une par une. Il faut cliquer sur les flèches (en bas des résultats) pour faire défiler le résultat des autres variables. (Nous verrons qu'il est possible de télécharger une synthèse de l'ensemble des résultats.)

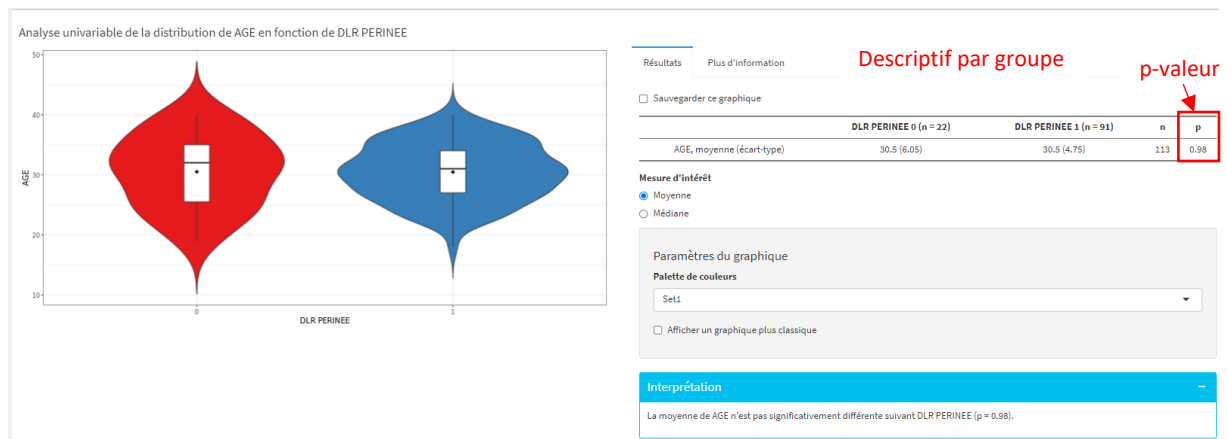


Résultats de l'analyse de la variable DLR_PERINEE (ici, il s'agit de notre variable "groupe d'intérêt"). Il s'agit d'une variable binaire pour laquelle nous obtenons, à droite, les effectifs et pourcentages des deux modalités. Un graphique apparaît sur la gauche : on peut modifier certains paramètres à l'aide des options disponibles sur la droite.

Vous pouvez faire défiler les résultats des différentes variables en utilisant les flèches en bas de page.

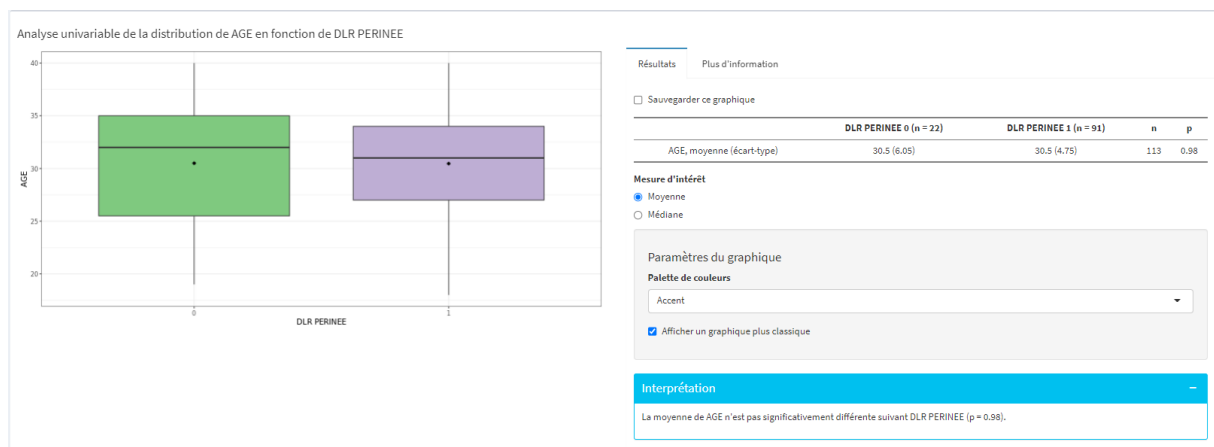


Analyse descriptive de la variable AGE : nous avons, dans un premier temps, un descriptif global de la variable.



Analyse descriptive de la variable AGE : nous avons, dans un second temps, un descriptif par groupe de la variable.

L'analyse par groupe d'une variable est accompagnée d'un test statistique (p-valeur) qui compare la variable entre les groupes concernés. Dans notre exemple, l'âge moyen est comparé entre les patients avec douleur du périnée et les patients sans douleur.



Analyse descriptive de la variable AGE : il s'agit de la précédente diapositive avec une modification des paramètres graphiques.

Etape 5 – Synthèse des résultats

Pour obtenir une synthèse de l'ensemble des analyses, il faut aller dans l'onglet "Espace de téléchargement" en bas de la page avec les résultats et cliquer sur "Télécharger les tableaux". L'ensemble des résultats est alors exporté dans un fichier Word.

◀
▶

Espace de téléchargement

☐ Tableaux en anglais

Télécharger les tableaux

Tables					
Analyses Descriptives					
Variables quantitatives					
	moyenne (écart-type)	médiane [Q25-75]	min	max	n
AGE	30.3 (5.05)	30.0 [27.0; 34.0]	18.0	40.0	117
BMI	24.5 (4.97)	24.0 [20.5; 27.0]	17.0	39.0	111
DELAI DCEE	45.8 (56.0)	22.0 [0; 60.0]	0	240	112
DUREE EFFORT	11.8 (10.5)	8.00 [4.00; 16.2]	0	55.0	112
DUREE TRAVAIL	6.52 (3.79)	6.00 [4.00; 9.00]	0	17.0	111
NB JOUR TT	2.36 (1.06)	2.00 [2.00; 3.00]	0	5.00	115
POIDS NAISSANCE	3317 (482)	3340 [3028; 3670]	2110	4390	111
Variables qualitatives					
	n (%)				
DLR PERINEE	0	22 (19%)			
	1	93 (81%)			
ATCD DECHIRURE	0	89 (75%)			
	1	30 (25%)			
ETAT PERINEE	0	79 (66%)			
	1	40 (34%)			
PRESENCE ERAILLURE	0	92 (77%)			
	1	27 (23%)			
PROFESSIONNEL	1	29 (46%)			
	2	14 (22%)			
	3	16 (25%)			
	4	4 (6.3%)			

Descriptif global

Analyses Univariable

En fonction de DLR PERINEE

	DLR PERINEE 0 (n = 22)	DLR PERINEE 1 (n = 93)	n	p	test
AGE, moyenne (écart-type)	30.5 (6.05)	30.5 (4.75)	113	0.98	Welch
BMI, moyenne (écart-type)	24.7 (4.64)	24.3 (4.90)	107	0.72	Welch
DELAI DCEE, médiane [Q25-75]	0 [0; 55.0]	30.0 [0; 80.0]	108	0.057	Mann-Whitney
DUREE EFFORT, médiane [Q25-75]	8.00 [4.00; 20.8]	7.00 [4.00; 16.0]	108	0.71	Mann-Whitney
DUREE TRAVAIL, moyenne (écart-type)	5.56 (3.29)	6.76 (3.93)	107	0.15	Welch
NB JOUR TT, médiane [Q25-75]	2.00 [1.00; 2.00]	2.00 [2.00; 3.00]	112	<0.01	Mann-Whitney
POIDS NAISSANCE, moyenne (écart-type)	3369 (563)	3315 (466)	107	0.69	Welch
ATCD DECHIRURE, n	0 16 (73%)	69 (74%)	85	0.89	Chi2
	1 6 (27%)	24 (26%)	30	-	-
ETAT PERINEE, n	0 11 (50%)	65 (70%)	76	0.076	Chi2
	1 11 (50%)	28 (30%)	39	-	-
PRESENCE ERAILLURE, n	0 17 (77%)	71 (76%)	88	0.93	Chi2
	1 5 (23%)	22 (24%)	27	-	-
PROFESSIONNEL, n	1 4 (57%)	22 (42%)	26	0.41	Fisher
	2 0 (0%)	14 (26%)	14	-	-
	3 3 (43%)	13 (25%)	16	-	-
	4 0 (0%)	4 (7.5%)	4	-	-

Descriptif par groupe et comparaisons

Etape 6 – Interprétation des résultats

A. Descriptif global

Les **variables quantitatives** sont décrites ici par la moyenne et l'écart-type, la médiane et l'intervalle interquartile (IQR) (i.e. 25e et 75e percentile), les valeurs minimale et maximale et le nombre de données renseignées (n).

L'écart-type et l'intervalle interquartile sont des paramètres de dispersion, ils sont indissociables de la moyenne pour le premier, de la médiane pour le second. Une moyenne seule ou une médiane seule ne donne pas suffisamment d'élément pour décrire une population.

Nous vous conseillons de décrire systématiquement l'ensemble de ces indicateurs (i.e. de garder le tableau descriptif qui vous sera fourni tel quel).

Les **variables qualitatives** sont décrites ici par la fréquence, n, et le pourcentage (%). Les pourcentages sont calculés sur le nombre de données renseignées. Par exemple, DLR_PERINEE est renseignée pour 22+93 patients, soit 115, les pourcentages de cette variable sont donc calculés sur 115.

B. Descriptif par groupe et comparaison (Analyses Univariées)

Les **variables quantitatives** : certaines sont décrites par la moyenne et l'écart-type, d'autres, par la médiane et l'IQR, le logiciel analyse la distribution et choisit le test le mieux adapté :

- Le **test de Welch** (équivalent au test t de Student) compare les moyennes de 2 groupes indépendants. Le logiciel suppose une distribution gaussienne de la variable.
- Le **test de Mann-Whitney** compare la distribution des rangs de la variable, le logiciel suppose une distribution non gaussienne. Dans ce cas, on décrira la variable par la médiane et l'IQR.

Les **variables qualitatives** : elles sont toutes décrites par la fréquence et le pourcentage.

- Le **test du Chi-2** compare les pourcentages de 2 ou plusieurs groupes indépendants.
- Le **test exact de Fisher** est utilisé lorsque les conditions d'application du Chi-2 ne sont pas réunies (souvent lors de petits effectifs). Son interprétation est similaire au Chi-2.

Quel que soit le test utilisé, une p-valeur sera fournie pour chaque variable. La p-valeur s'interprète par rapport au risque alpha de 5% :

- Si $p \geq 0.05$ alors on ne met pas en évidence de différence significative entre les groupes.
- Si $p < 0.05$ alors il y a une différence significative entre les groupes.

Dans notre exemple, seul le NB JOUR TT est significativement différent entre les 2 groupes. Il est plus faible dans le groupe DLR PERINEE 0 comparé au groupe DLR PERINEE 1 (le 25e percentile et le 75e percentile sont plus petits dans le groupe 0 que dans le groupe 1).

Pour **PROFESSIONNEL** un test a été réalisé. Cependant les effectifs dans le groupe 0 sont trop faibles, 4+3=7 patients au lieu de 22 (dû à des données manquantes), de ce fait, un test n'est pas recommandé

car il y a moins de 8 patients analysables au total dans un groupe ou une modalité (c'est le cas aussi de la modalité 4 de la variable PROFESSIONNEL = 4 patients)

Remarque : pour les variables qualitatives, un test est réalisé par variable et non par modalité.

Paragraphe statistique

Les paragraphes ci-dessous en italique sont à adapter selon les analyses réalisées, puis à inclure dans la partie méthodologie de votre manuscrit (thèse ou mémoire).

Les analyses statistiques ont été effectuées à l'aide de l'interface pvalue.io (Medistica., pvalue.io, une interface graphique au logiciel de statistique R pour les publications scientifiques médicales., 2019-22.).

Les variables qualitatives ont été décrites en termes d'effectifs et de pourcentages. Les variables quantitatives ont été décrites en termes de moyenne et d'écart-type et en termes de médiane et d'intervalle interquartile.

ET (si des tests statistiques ont été réalisés – comparaison de 2 groupes)

Les caractéristiques ont été comparées entre les deux groupes à l'aide d'un test du Chi-deux ou de Fisher exact (lorsque les conditions de validité du test du Chi-deux ne sont pas vérifiées) pour les variables qualitatives, à l'aide d'un test t de Student pour les variables continues gaussiennes et à l'aide d'un test U de Mann-Whitney pour les variables continues non gaussiennes.

ET/OU (si des tests statistiques ont été réalisés – comparaison de 3 groupes)

Les caractéristiques ont été comparées entre les trois groupes à l'aide d'un test du Chi-deux ou de Fisher exact (lorsque les conditions de validité du test du Chi-deux ne sont pas vérifiées) pour les variables qualitatives, à l'aide d'une ANOVA pour les variables quantitatives gaussiennes, et à l'aide d'un test de Kruskal-Wallis pour les variables quantitatives non gaussiennes.

ET (si des tests statistiques ont été réalisés)

Des tests bilatéraux ont été réalisés avec un niveau de significativité de 5%.

Pour aller plus loin...

Une chaine YouTube existe pour un tutoriel en vidéo, cliquez [ici](#).

Pour toute question, vous pouvez nous contacter via l'adresse mail suivante :

biostatistique@chu-lille.fr

Dans le but d'améliorer ce tutoriel et l'aide que nous pouvons vous apporter en amont de ces analyses, pensez à nous faire un retour de votre expérience via le court formulaire suivant :

[Lien du questionnaire](#)

Merci pour le temps que vous prendrez à y répondre.