

# Gestion et analyse d'une base de données

Formation DMG

04/07/2024

Arnaud Cugerone - [arnaud.cugerone@u-bordeaux.fr](mailto:arnaud.cugerone@u-bordeaux.fr)

Thomas Ferté - [thomas.ferte@u-bordeaux.fr](mailto:thomas.ferte@u-bordeaux.fr)

Paul Vanderkam - [paul.vanderkam@u-bordeaux.fr](mailto:paul.vanderkam@u-bordeaux.fr)

Linda Wittkop - [linda.wittkop@u-bordeaux.fr](mailto:linda.wittkop@u-bordeaux.fr)

# Récupérer la base de données

Moodle séminaire aide thèse :

<https://moodle.u-bordeaux.fr/course/view.php?id=6149>



Analyser ses données > Base de données petits poids de naissance

# Exemple fil rouge

## Description

Les données nous viennent de Hosmer et Lemeshow (2000). Appelées les données de faible poids de naissance (lbw), la variable de réponse est une variable binaire, `low_birth_weight`, qui indique si le poids de naissance d'un bébé est inférieur à 2500g ou supérieur.

## Source

Hosmer, D and S. Lemeshow (2000), Applied Logistic Regression, Wiley

Hilbe, Joseph M (2007, 2011), Negative Binomial Regression, Cambridge

University Press  
Hilbe, Joseph M (2009), Logistic Regression Models, Chapman & Hall/CRC

# Objectifs

## Objectif principal

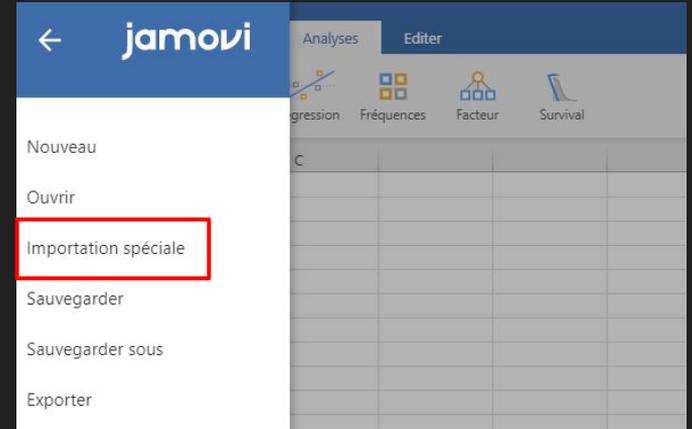
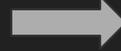
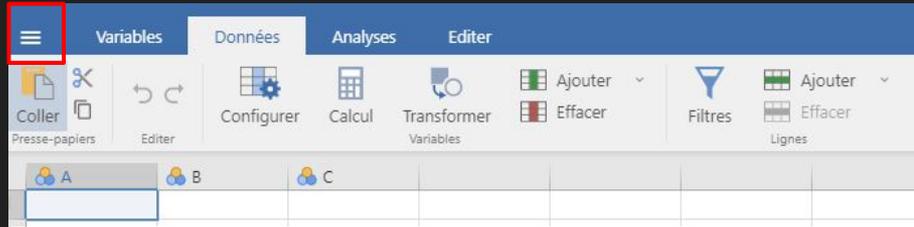
Évaluer l'association entre la consommation de tabac et le petit poids de naissance.

## Objectifs secondaires

- Évaluer l'association entre l'origine ethnique, l'âge de la mère, le poids de la mère, le nombre de fausses couches antérieures, l'hypertension artérielle et l'irritabilité utérine avec le petit poids de naissance en qualitatif.
- Idem avec le petit poids de naissance en quantitatif



# Importer ses données



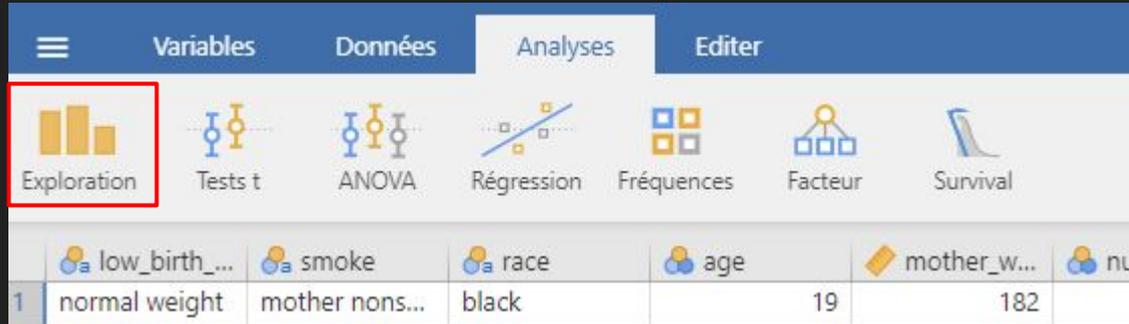
A screenshot of the Jamovi software interface showing a data table. The table has 10 columns and 15 rows of data. The columns are labeled: 'low\_birth...', 'smoke', 'race', 'age', 'mother\_w...', 'number\_o...', 'hypertens...', 'uterine\_irr...', 'number\_o...', and 'birth\_wei...'. The rows contain numerical and categorical data.

low_birth...	smoke	race	age	mother_w...	number_o...	hypertens...	uterine_irr...	number_o...	birth_wei...
normal weight	mother nons...	black	19	182	0	no hypertens...	uterine irritab...	0	2523
normal weight	mother nons...	other	33	155	0	no hypertens...	no irritability	3	2551
normal weight	history of mo...	white	20	105	0	no hypertens...	no irritability	1	2557
normal weight	history of mo...	white	21	108	0	no hypertens...	uterine irritab...	2	2594
normal weight	history of mo...	white	18	107	0	no hypertens...	uterine irritab...	0	2600
normal weight	mother nons...	other	21	124	0	no hypertens...	no irritability	0	2622
normal weight	mother nons...	white	22	118	0	no hypertens...	no irritability	1	2637
normal weight	mother nons...	other	17	103	0	no hypertens...	no irritability	1	2637
normal weight	history of mo...	white	29	123	0	no hypertens...	no irritability	1	2663
normal weight	history of mo...	white	26	113	0	no hypertens...	no irritability	0	2665
normal weight	mother nons...	other	19	95	0	no hypertens...	no irritability	0	2722
normal weight	mother nons...	other	19	150	0	no hypertens...	no irritability	1	2733
normal weight	mother nons...	other	22	95	0	history of hy...	no irritability	0	2750
normal weight	mother nons...	other	30	107	1	no hypertens...	uterine irritab...	2	2750



# Description des données

Analyses > Exploration > Statistiques descriptives



The screenshot shows the 'Analyses' menu in a software interface. The menu is divided into several sections: 'Variables', 'Données', 'Analyses', and 'Editer'. The 'Analyses' section is currently active and contains several options: 'Exploration', 'Tests t', 'ANOVA', 'Régression', 'Fréquences', 'Facteur', and 'Survival'. The 'Exploration' option is highlighted with a red box. Below the menu, a table of variables is visible, with columns for 'low\_birth...', 'smoke', 'race', 'age', 'mother\_w...', and 'nu'. The first row of data shows values for these variables: 'normal weight', 'mother nons...', 'black', '19', and '182'.

	low_birth_...	smoke	race	age	mother_w...	nu
1	normal weight	mother nons...	black	19	182	

# Exemple : Tabac en fonction du petit poids de naissance

Statistiques descriptives

Variables

- race
- age
- mother\_weight
- number\_of\_false\_or\_premature\_lab...
- hypertension
- uterine\_irritability
- number\_of\_physician\_visits\_in\_1st\_t...
- birth\_weight\_in\_grams

Variables

- smoke

Séparer par

- low\_birth\_weight

Statistiques descriptives

Variables sur plusieurs lignes  Tableau des fréquences

Statistiques

Taille de l'échantillon

N  Manquants

Tendance centrale

Moyenne  
 Médiane  
 Mode  
 Somme

Valeurs de centiles

Quantiles pour 4 groupes égaux  
 Centiles 25.50,75

Dispersion

Ecart-type  Minimum  
 Variance  Maximum  
 Etendue  Ecart interquartile

Distribution

Coefficient d'asymétrie  
 Kurtosis

Normalité

Shapiro-Wilk

Dispersion moyenne

Erreur-standard de la moyenne  
 Intervalle de confiance de la moyenne 95% Plus extrême 5 Valeurs

Valeurs atypiques

Graphes

Histogrammes

Boîtes à moustaches

Diagrammes en barres

Histogramme  Boîte à moustaches  Diagramme en barres

Statistiques descriptives

	low_birth_weight	N	Manquants	Minimum	Maximum
smoke	low birthweight baby	59	0		
	normal weight	130	0		

Fréquences

Fréquences de smoke

smoke	low_birth_weight	Quantités	% du Total	% cumulés
history of mother smoking	low birthweight baby	30	15.9 %	15.9 %
	normal weight	44	23.3 %	39.2 %
mother nonsmoker	low birthweight baby	29	15.3 %	54.5 %
	normal weight	86	45.5 %	100.0 %

Graphes

smoke

counts

low\_birth\_weight

- low birthweight baby
- normal weight

history of mother smoking mother nonsmoker

smoke

Exercice : décrivez le poids de la mère en fonction du petit poids de naissance. Réalisez le graphique adapté

# Exercice : décrivez le poids de la mère en fonction du petit poids de naissance. Réalisez le graphique adapté

### Statistiques descriptives

Variables: mother\_weight

Séparer par: low\_birth\_weight

Statistiques descriptives: Variables sur plusieurs lignes

Taille de l'échantillon:  N  Manquants

Valeurs de centiles:  Quantiles pour 4 groupes égaux  Centiles 25,50,75

Dispersion:  Ecart-type  Minimum  Variance  Maximum  Etendue  Ecart interquartile

Dispersion moyenne:  Erreur-standard de la moyenne  Intervalle de confiance de la moyenne 95%  Plus extrême 5 Valeurs

Tendance centrale:  Moyenne  Médiane  Mode  Somme

Distribution:  Coefficient d'asymétrie  Kurtosis

Normalité:  Shapiro-Wilk

Valeurs atypiques:  Plus extrême 5 Valeurs

Graphes:  Histogramme  Boîtes à moustaches  Diagrammes en barres

### Résultats

#### Statistiques descriptives

Statistiques descriptives

	low_birth_weight	N	Manquants	Minimum	Maximum	Centiles		
	low birthweight baby					25th	50th	75th
mother_weight	low birthweight baby	59	0	80	200	104	120	130
	normal weight	130	0	85	250	113	124	147

#### Graphes

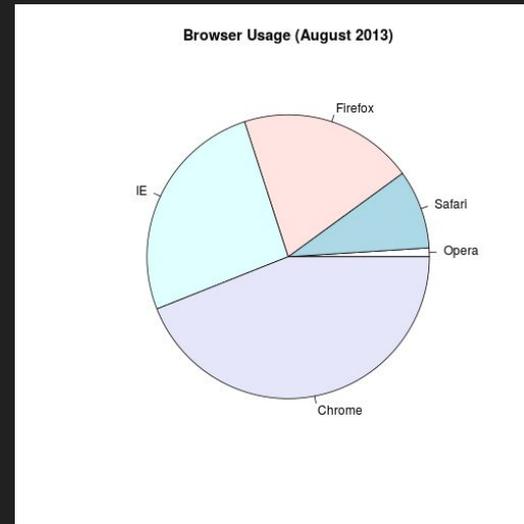
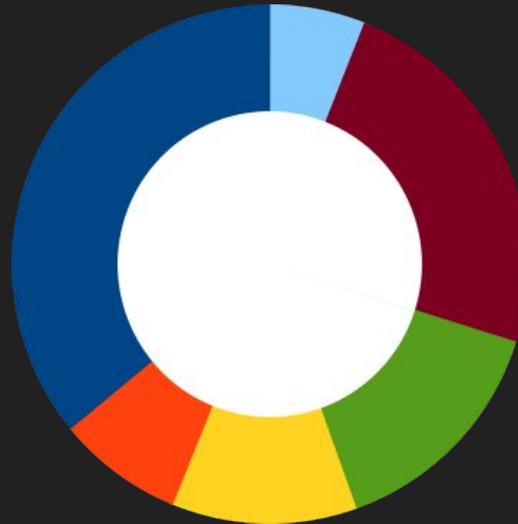
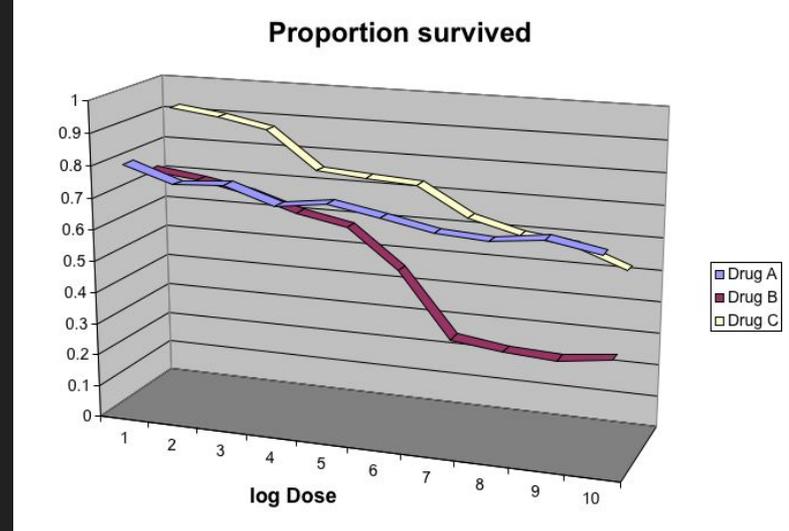
mother\_weight

low birthweight baby normal weight

low\_birth\_weight

# Ne pas faire

- Camemberts/Donuts
- Plots en 3D
- Un graphique et un tableau pour représenter la même chose
- Un graphique lorsque le tableau est plus parlant



# Bonnes idées

- Variable quantitative continue : histogramme
- Variable quantitative discrète : histogramme ou barplot
- Variable qualitative : souvent tableau > graphique
- Variable quantitative ~ Variable quantitative : nuage de points
- Variable quantitative ~ Variable qualitative : boîte à moustache
- Variable qualitative ~ Variable qualitative : souvent tableau > graphique

Test simples

# Choix du test

	<b>Variable qualitative</b>	<b>Variable quantitative</b>
<b>Variable qualitative</b>	Chi-2 / Fisher	Student / U de Mann Whitney
<b>Variable quantitative</b>	Student / U de Mann Whitney	Corrélation Spearman / Pearson
<b>Survie</b>	Log-rank	

# Évaluer l'association entre le statut tabagique et le petit poids de naissance

Analyses > Fréquences > Table de contingence > échantillons indépendants



# Évaluer l'association entre le statut tabagique et le petit poids de naissance

Si toutes les quantités attendues  $\geq 5$

=> **Chi-2**

Si au moins une quantité attendue entre 3 et 5

=> **Chi-2 avec correction**

**de continuité**

Si au moins une quantité attendue  $< 3$

=> **Fisher**

The screenshot shows the SPSS 'Tables de contingence' (Contingency Tables) dialog box. The 'Lignes' (Rows) field contains 'smoke' and the 'Colonnes' (Columns) field contains 'low\_birth\_weight'. Under 'Quantités (optionnel)' (Optional Quantities), the 'Quantités attendues' (Expected Quantities) checkbox is checked and highlighted with a red box. Under 'Pourcentages' (Percentages), the 'Total' checkbox is checked. The 'Effectifs observés' (Observed Counts) checkbox is also checked. The 'Statistiques' (Statistics) section is expanded, and the 'Cellules' (Cells) section is also expanded.

The right side of the screenshot shows the resulting contingency table and chi-square test results. The table is titled 'Tables de contingence' and has 'smoke' as the row variable and 'low\_birth\_weight' as the column variable. The columns are 'low birthweight baby', 'normal weight', and 'Total'. The rows are 'history of mother smoking' and 'mother nonsmoker'. The table shows observed counts and expected counts (Attendu) for each cell. The expected counts for the 'history of mother smoking' row are 23.1 and 50.9, and for the 'mother nonsmoker' row are 35.9 and 79.1. These expected counts are highlighted with red boxes. The total number of observations is 189.

smoke	low_birth_weight		Total
	low birthweight baby	normal weight	
history of mother smoking	Observé: 30 Attendu: 23.1	44 50.9	74 74.0
mother nonsmoker	Observé: 29 Attendu: 35.9	86 79.1	115 115.0
Total	Observé: 59 Attendu: 59.0	130 130.0	189 189.0

Tests  $\chi^2$

Valeur	
N	189

# Évaluer l'association entre le statut tabagique et le petit poids de naissance

Si toutes les quantités attendues  $\geq 5$

=> **Chi-2**

Si au moins une quantité attendue entre 3 et 5

=> **Chi-2 avec correction de continuité**

Si au moins une quantité attendue  $< 3$

=> **Fisher**

The screenshot shows a statistical software interface with the following sections:

- Variables:** birth\_weight\_in\_grams is selected.
- Statistiques:** Tests, Mesures comparatives (2x2 seulement), Hypothèse, Nominal, Ordinal, Cellules, Quantités, Pourcentages.
- Tests:**   $\chi^2$ ,  Correction de continuité du  $\chi^2$ ,  Ratio de vraisemblance,  Test exact de Fisher,  test z pour la différence entre deux proportions.
- Mesures comparatives (2x2 seulement):**  Rapport des cotes (odds ratio),  Log du rapport des cotes (odds ratio),  Risque relatif,  Différence entre les proportions,  Intervalles de confiance.
- Hypothèse:**  Groupe 1  $\neq$  Groupe 2,  Groupe 1  $>$  Groupe 2,  Groupe 1  $<$  Groupe 2.
- Nominal:**  Coefficient de contingence,  V de Phi et Cramer.
- Ordinal:**  Gamma,  Tau b de Kendall,  Mantel-Haenszel.
- Cellules:**  Cellules.
- Quantités:**  Effectifs observés,  Quantités attendues.
- Pourcentages:**  Ligne,  Colonne,  Total.

**Tables de contingence**

	smoke	low_birth_weight		Total
		low birthweight baby	normal weight	
history of mother smoking	Observé	30	44	74
	% par ligne	40.5 %	59.5 %	100.0 %
mother nonsmoker	Observé	29	86	115
	% par ligne	25.2 %	74.8 %	100.0 %
Total	Observé	59	130	189
	% par ligne	31.2 %	68.8 %	100.0 %

**Tests  $\chi^2$**

	Valeur	ddl	p
$\chi^2$	4.92	1	0.026
N	189		

**Mesures comparatives**

	Intervalles de confiance à 95%		
	Valeur	Borne inf	Supérieure
Différence entre deux proportions	0.153*	0.0161	0.290
Rapport des cotes (odds ratio)	2.02	1.08	3.78

\* Lignes comparées

**Références**

Évaluer l'association entre l'irritabilité utérine et le poids de naissance

# Évaluer l'association entre le statut tabagique et le petit poids de naissance

Si toutes les quantités attendues  $\geq 5$

=> **Chi-2**

Si au moins une quantité attendue entre 3 et 5

=> **Chi-2 avec correction**

**de continuité**

Si au moins une quantité attendue  $< 3$

=> **Fisher**

The screenshot shows a software interface for creating contingency tables. On the left, a list of variables includes 'uterine\_irritability' which is selected. The 'Lignes' (Rows) field contains 'hypertension' and the 'Colonnes' (Columns) field contains 'low\_birth\_weight'. Under 'Quantités (optionnel)', 'Quantités attendues' is checked and highlighted with a red box. Under 'Pourcentages', 'Ligne', 'Colonne', and 'Total' are unchecked. The right side displays the resulting contingency table with expected values highlighted in red.

**Tables de contingence**

hypertension	low_birth_weight		Total
	low birthweight baby	normal weight	
history of hypertension	Observé: 7	5	12
	Attendu: 3.75	8.25	12.0
no hypertension	Observé: 52	125	177
	Attendu: 55.25	121.75	177.0
Total	Observé: 59	130	189
	Attendu: 59.00	130.00	189.0

**Tests  $\chi^2$**

	Valeur
N	189

# Évaluer l'association entre le statut tabagique et le petit poids de naissance

Si toutes les quantités attendues  $\geq 5$

=> **Chi-2**

Si au moins une quantité attendue entre 3 et 5

=> **Chi-2 avec correction de continuité**

Si au moins une quantité attendue  $< 3$

=> **Fisher**

The screenshot shows the 'Tables de contingence' (Contingency Tables) dialog box in a statistical software package. The variables 'birth\_weight\_in\_grams' and 'hypertension' are selected. Under 'Statistiques' (Statistics), the 'Tests' section has 'Correction de continuité du  $\chi^2$ ' (Continuity correction for  $\chi^2$ ) checked. Under 'Mesures comparatives (2x2 seulement)' (Comparative measures (2x2 only)), 'Rapport des cotes (odds ratio)' (Odds ratio) and 'Différence entre les proportions' (Difference between proportions) are checked. The 'Hypothèse' (Hypothesis) is set to 'Groupe 1 ≠ Groupe 2' (Group 1 ≠ Group 2). The 'Intervalle' (Interval) is 95% and 'Comparer' (Compare) is set to 'lignes' (rows). Under 'Nominal' (Nominal), 'Coefficient de contingence' (Contingency coefficient) and 'V de Phi et Cramer' (Phi and Cramer's V) are unchecked. Under 'Ordinal' (Ordinal), 'Gamma', 'Tau b de Kendall', and 'Mantel-Haenszel' are unchecked. Under 'Quantités' (Quantities), 'Effectifs observés' (Observed counts) is checked. Under 'Pourcentages' (Percentages), 'Ligne' (Row) is checked. The 'Cellules' (Cells) section is empty.

The results pane on the right shows the 'Tables de contingence' (Contingency Tables) and 'Tests  $\chi^2$ ' (Chi-squared tests) sections. The contingency table shows the association between hypertension and low birth weight. The Chi-squared test results show a p-value of 0.076, which is highlighted with a red box.

hypertension	low_birth_weight		Total
	low birthweight baby	normal weight	
history of hypertension	Observé 7	5	12
	% par ligne 58.3 %	41.7 %	100.0 %
no hypertension	Observé 52	125	177
	% par ligne 29.4 %	70.6 %	100.0 %
Total	Observé 59	130	189
	% par ligne 31.2 %	68.8 %	100.0 %

	Valeur	ddl	p
Correction de continuité du $\chi^2$	3.14	1	0.076
N	189		

	Intervalles de confiance à 95%		
	Valeur	Borne inf	Supérieur
Différence entre deux proportions	0.290*	0.00265	0.576
Rapport des cotes (odds ratio)	3.37	1.02	11.1

\* Lignes comparées

Références

# Évaluer l'association entre le statut tabagique et le poids de naissance

Analyses > Test t > Test t pour échantillons indépendants



# Évaluer l'association entre le statut tabagique et le poids de naissance

Test t pour échantillons indépendants

Variables dépendantes  
birth\_weight\_in\_grams

Variable de groupage  
smoke

Tests  
 Student  
 Facteur de Bayes  
A priori 0.707  
 Welch  
 U de Mann-Whitney

Statistiques additionnelles  
 Différence moyenne  
 Intervalle de confiance 95 %  
 Taille de l'effet  
 Intervalle de confiance 95 %  
 Statistiques descriptives  
 Graphes descriptifs

Hypothèse  
 Groupe 1 ≠ Groupe 2  
 Groupe 1 > Groupe 2  
 Groupe 1 < Groupe 2

Valeurs manquantes  
 Exclure les cas analyse par analyse  
 Exclure les cas selon une liste

no hypertension	Observé	52	125	177
	% par ligne	29.4 %	70.6 %	100.0 %
Total	Observé	59	130	189
	% par ligne	31.2 %	68.8 %	100.0 %

Tests  $\chi^2$

	Valeur	ddl	p
Correction de continuité du $\chi^2$	3.14	1	0.076
N	189		

Mesures comparatives

	Valeur	Intervalle de confiance à 95%	
		Borne inf	Supérieur
Différence entre deux proportions	0.290*	0.00265	0.576
Rapport des cotes (odds ratio)	3.37	1.02	11.1

\* Lignes comparées

Test t pour échantillons indépendants

	Statistique	ddl	p	Différence moyenne	Différence d'erreur standard	Intervalle de confiance à 95%	
						Borne inf	Supérieur
birth_weight_in_grams	t de Student	-2.64	187	0.009	-283	107	-494 -71.7

Note: H<sub>0</sub>: H history of mother smoking = H mother nonsmoker

Statistiques descriptives des groupes

Variable	Groupe	N	Moyenne	Médiane	Ecart-type	Erreur standard
birth_weight_in_grams	history of mother smoking	74	2772	2776	660	76.7
	mother nonsmoker	115	3055	3100	752	70.2

N ≥ 30 dans les 2 groupes  
=> Student

N < 30 dans un des groupes  
=> U de Mann Whitney  
(= Wilcoxon)

Évaluer l'association entre l'hypertension et le poids de naissance

# Évaluer l'association entre l'hypertension et le poids de naissance

Test t pour échantillons indépendants

Variables dépendantes  
 birth\_weight\_in\_grams

Variable de groupage  
 hypertension

**Tests**

Student  
 Facteur de Bayes  
 A priori 0.707  
 Welch  
 U de Mann-Whitney

**Hypothèse**

Groupe 1 ≠ Groupe 2  
 Groupe 1 > Groupe 2  
 Groupe 1 < Groupe 2

**Valeurs manquantes**

Exclure les cas analyse par analyse  
 Exclure les cas selon une liste

**Statistiques additionnelles**

Différence moyenne  
 Intervalle de confiance 95 %  
 Taille de l'effet  
 Intervalle de confiance 95 %  
 Statistiques descriptives  
 Graphes descriptifs

**Vérifications des hypothèses**

Test d'homogénéité des variances  
 Test de normalité  
 Graphe Q-Q

Total	Observé % par ligne	59 31.2 %	130 68.8 %	189 100.0 %
-------	------------------------	--------------	---------------	----------------

Tests  $\chi^2$

	Valeur	ddl	p
Correction de continuité du $\chi^2$	3.14	1	0.076
N	189		

Mesures comparatives

	Valeur	Intervalles de confiance à 95%	
		Borne inf	Supérieur
Différence entre deux proportions	0.290 *	0.00265	0.576
Rapport des cotes (odds ratio)	3.37	1.02	11.1

\* Lignes comparées

## 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
birth_weight_in_grams	t de Student	-2.02	187	0.045	-435	216	-861 -9.62
	U de Mann-Whitney	774	0.117	-448			-979 113

Note. H<sub>a</sub>  $\mu$  history of hypertension ≠  $\mu$  no hypertension

Statistiques descriptives des groupes

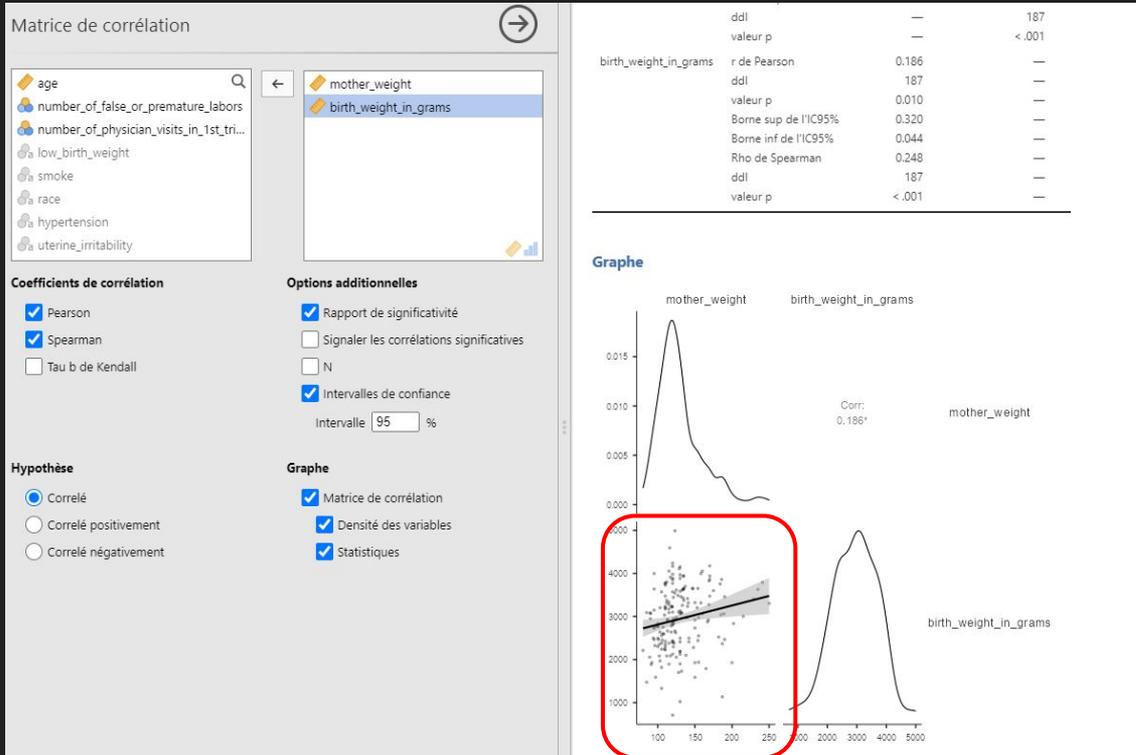
	Groupe	N	Moyenne	Médiane	Ecart-type	Erreur standard
birth_weight_in_grams	history of hypertension	12	2537	2495	917	265
	no hypertension	177	2972	2977	709	53.3

# Association entre poids de la mère et poids de l'enfant

Analyses > Régression > Matrices de corrélation



# Association entre poids de la mère et poids de l'enfant



Relation linéaire  
=> **Pearson**

Relation non linéaire (monotonique)  
=> **Spearman**

# Choix du test

	<b>Variable qualitative</b>	<b>Variable quantitative</b>
<b>Variable qualitative</b>	effectif attendu $\geq 5$ => <b>Chi-2</b> effectif attendu $\geq 3$ et $< 5$ => <b>Chi-2 + correction</b> effectif attendu $< 3$ => <b>Fisher</b>	effectif $\geq 30$ (dans chaque groupe) => <b>Student</b> effectif $< 30$ => <b>U de Mann Whitney (= Wilcoxon)</b>
<b>Variable quantitative</b>	effectif $\geq 30$ (dans chaque groupe) => <b>Student</b> effectif $< 30$ => <b>U de Mann Whitney (= Wilcoxon)</b>	relation linéaire => <b>Pearson</b> relation non linéaire (monotone) => <b>Spearman</b>
<b>Survie</b>	Log-rank	

# Taille d'étude avec Biostatgv

<https://biostatgv.sentiweb.fr/?module=etudes/sujets>

# Outcome qualitatif

Comparer 2 proportions binomiales

Calcul Aide

Nombre de sujets nécessaires : 2 proportions

Type de comparaison  
 d'une proportion observée à une proportion théorique  
 de 2 proportions observées

$\pi_1$  Proportion dans le groupe 1  valeur entre 0 et 1  
 $\pi_2$  Proportion dans le groupe 2  valeur entre 0 et 1

Risque de première espèce  $\alpha$   0.05 valeur entre 0 et 1  
Puissance 1 -  $\beta$   0.9 valeur entre 0 et 1

Nature du test  
 Bilatéral  Unilatéral

Réinitialiser Calculez

Résultats : Nombre de sujets nécessaires

Des résultats selon plusieurs méthodes sont disponibles

Proportion Observées (Arcsin approximation)

- Nombre total de sujet 214
- Nombre sujet dans le groupe 1 107
- Nombre sujet dans le groupe 2 107
- Alpha (erreur de type I) 0.025
- Puissance 0.9

epiR package 0.9-96

- Nombre total de sujet 218
- Nombre sujet dans le groupe 1 109
- Nombre sujet dans le groupe 2 109
- Puissance 0.9

Nombre de sujets nécessaire pour une proportion attendue d'événement dans les deux groupes de **20%** et **40%**, avec un puissance de **90%** (probabilité de mettre en évidence si elle existe) et un risque alpha de **5%** (probabilité de mettre en évidence une différence alors qu'elle n'existe pas)

# Outcome quantitatif

Comparer 2 moyennes observées

Calcul Aide

Saisie des paramètres

Moyenne du premier groupe  $\mu_1$

Moyenne du deuxième groupe  $\mu_2$

$d = |\mu_1 - \mu_2|$

Ecart type commun  $\sigma$

Risque de première espèce  $\alpha$   0.05 valeur entre 0 et 1

Puissance  $1 - \beta$   0.9 valeur entre 0 et 1

Nature du test  Bilatéral  Unilatéral

Résultats

Nombre de sujets nécessaires n (par groupe)

epiR package 0.9-96

- Nombre total de sujet 170
- Nombre sujet dans le groupe 1 85
- Nombre sujet dans le groupe 2 85

Nombre de sujets nécessaire pour mettre en évidence une différence moyenne de **10** entre les deux groupes, avec un écart-type de **20** (variabilité de la mesure) avec un puissance de **90%** (probabilité de mettre en évidence si elle existe) et un risque alpha de **5%** (probabilité de mettre en évidence une différence alors qu'elle n'existe pas)

Take home message

# Choix du test

	<b>Variable qualitative</b>	<b>Variable quantitative</b>
<b>Variable qualitative</b>	effectif attendu $\geq 5$ => <b>Chi-2</b> effectif attendu $\geq 3$ et $< 5$ => <b>Chi-2 + correction</b> effectif attendu $< 3$ => <b>Fisher</b>	effectif $\geq 30$ (dans chaque groupe) => <b>Student</b> effectif $< 30$ => <b>U de Mann Whitney (= Wilcoxon)</b>
<b>Variable quantitative</b>	effectif $\geq 30$ (dans chaque groupe) => <b>Student</b> effectif $< 30$ => <b>U de Mann Whitney (= Wilcoxon)</b>	relation linéaire => <b>Pearson</b> relation non linéaire (monotone) => <b>Spearman</b>
<b>Survie</b>	Log-rank	

Questions ?