

SPSS 22.0 : Commandes – Plan d'analyse – 27 pages

B. Branger - 2017 – Tél 06 32 70 33 80 6

Mail : branger44@gmail.com

Préalable : la saisie et l'exportation des données

La saisie conditionne la qualité des données +++

- Sources et méthodes :

- Questionnaire papier ou saisie directe sur ordinateur (difficile)
- Import d'une base de données (de service ou du PMSI par exemple)

- Un sujet, une question, une ou des réponses (uniques/exclusives ou multiples).

- Réponses codées en chiffres. Cas = 1 et témoins = 2 (ou respectivement 0 et 1).
- Oui=1, Non=2, ou Non=0, Oui=1, autre Oui = 2....
- Exemple de questionnaire

- Votre année de naissance - Code postal de votre domicile

- Votre nationalité : Française Etrangère

- Activité professionnelle

- Mère (grille ci-jointe, mettre un numéro)	1. agriculteur-trice 2. artisan, commerçant 3. profession libérale, cadre supérieur, ingénieur, professeur, médecin 4. profession intermédiaire, cadre moyen
- Père (grille ci-jointe, mettre un numéro)	5. employé(e) 6. ouvrier(e) 7. sans profession 8. autres

- Situation vis-à-vis de l'emploi

Mère (grille ci-jointe, mettre un numéro)	1. Avec un emploi 2. Etudiant.
Père(grille ci-jointe, mettre un numéro)	3. Au chômage 4. Autre situation

- Avez-vous eu une **chirurgie des seins** (réduction mammaire ou plastie) ? 0. Non 1. Oui

- Avez-vous été allaitée par votre propre **mère** ? 0. Non 1. Oui 2. Ne sais pas

- Nombre d'enfants que vous avez eus avant celui-ci :

- Si vous avez eu d'**autres enfants**, ont-ils été allaités :

0. Non, aucun 1. Oui, une partie 2. Oui, tous

Vous diriez que cette ou ces expériences d'allaitement :

1. Ont été un succès 2. Ont été un échec 3. Un peu des deux

- A **quel moment** avez-vous décidé d'allaiter ou de donner le biberon pour votre bébé actuel ?

0. Depuis toujours 1. Avant la grossesse 2. Pendant la grossesse

3. A la naissance 4. Ne sais pas

- Avez-vous suivi des cours de **préparation** à la naissance ? 0. Non 1. Oui

- Est-ce que l'on vous a **parlé** d'allaitement maternel pendant le cours de préparation ?

0. Non, pas du tout 1. Non, assez peu 2. Oui, un peu 4. Beaucoup

- Qui vous a influencé pour l'alimentation de votre bébé

conjoint amis entourage internet revues milieu médical

- Votre **poids** avant la grossesse

- Votre **taille**

- Eviter les textes, les commentaires. Si nécessaire, utiliser des mots-clés

- La saisie

- Sur EPIDATA : exemple. L'export se fait facilement vers SPSS.
- SPSS n'est pas adapté pour la saisie +++
- Sur tableur (Excel) : conditions
 - ✓ Les observations sont en lignes (les numéroter à partir de 1) et les variables sont en colonnes
 - ✓ 1^{ère} ligne : noms des variables (10 ou 15 lettres), pas de signes spéciaux (< % !....)
 - ✓ Format des cellules : chiffre, nombre avec ou sans décimales (contrôler le nombre de décimales), texte (à éviter, pas plus de 10 lettres sous forme de mots-clés), date de type européen dd/mm/yyyy. Ne pas mélanger dans la même colonne nombre et texte : par exemple coder dans les lignes successives 1 2 puis 3, c'est OK; mais si on met 1 + 2 : c'est du texte !
 - ✓ Un seul fichier pour les cas/témoins ou malades/non-malades : prévoir une colonne avec 1 et 2, ou 0 et 1 pour les différencier
 - ✓ Pas de couleur, ni italique : ne sert à rien
 - ✓ Si on a une quantité : saisir la quantité; on fera des classes APRES dans le logiciel de stat. Par exemple, pour l'âge, mettre 30 ans (soit, préférable saisir la date de naissance et la date d'enquête), mais 30 – 34 ans...
 - ✓ Transformer tous les systèmes non décimaux en système décimal
 - Semaines d'aménorrhée (sur une semaine) : 1 jour = 0.14 semaine. Formule : AG = semaines + (jour/7)
 - Heure, minutes : 1 heure 30 min = 1.5 heure. Formule : Durée = heure + (minutes/60)
 - Heures, jours : 36 heures = 1.5 jours. Formule : Durée = heures/24
 - Années, mois : 12 ans et 9 mois = 12.75. Formule : Age = années + (mois/12). OU années, mois, semaines, jours... Formule : Age en années = Age en jours/365.25 (pour les années bissextiles). Formule : Age en années = Age en mois /12. Formule : Age en mois = Age en jours/30.41

- L'export à partir d'Excel

SPSS peut importer tous les fichiers Excel en *.xls ou *.xlsx de Mac ou de PC sans trop de problème (sauf si les colonnes contiennent trop de texte quelquefois). Attention, aux imports des séparateurs des décimales : le mieux est de saisir avec des "," dans Excel qui feront des "." dans SPSS.

- ✓ *Pour Windows 10 : Panneau de configuration en format "Catégorie" (pas icônes) Modifier les formats de date, d'heure ou de nombre/ Paramètres supplémentaires/ modifier ",."*

Tableau I : Exemple de saisie sur tableur : minimum de texte, et des codes +++

N	Date	AG	Debut	Anes	Atcd	Previa	A1	Dateacc	PN	Duree	Sexe
1	11/11/2008	36.14	1	1	1	2	10	12/10/2017	2500	1.52	1
3	12/07/2005	40.42	3	2	2	2	9	15/02/2017	3100	1.25	2
4	17/03/2006	37.86	2	1	2	2	10	14/08/2017	2600	0.75	1

* Nom des variables en tête de colonne : pas + de 8 à 10 lettres, pas d'accent, pas d'espace, pas de " ? ". Nom de variable évocateur sans fioriture, sur une ligne.

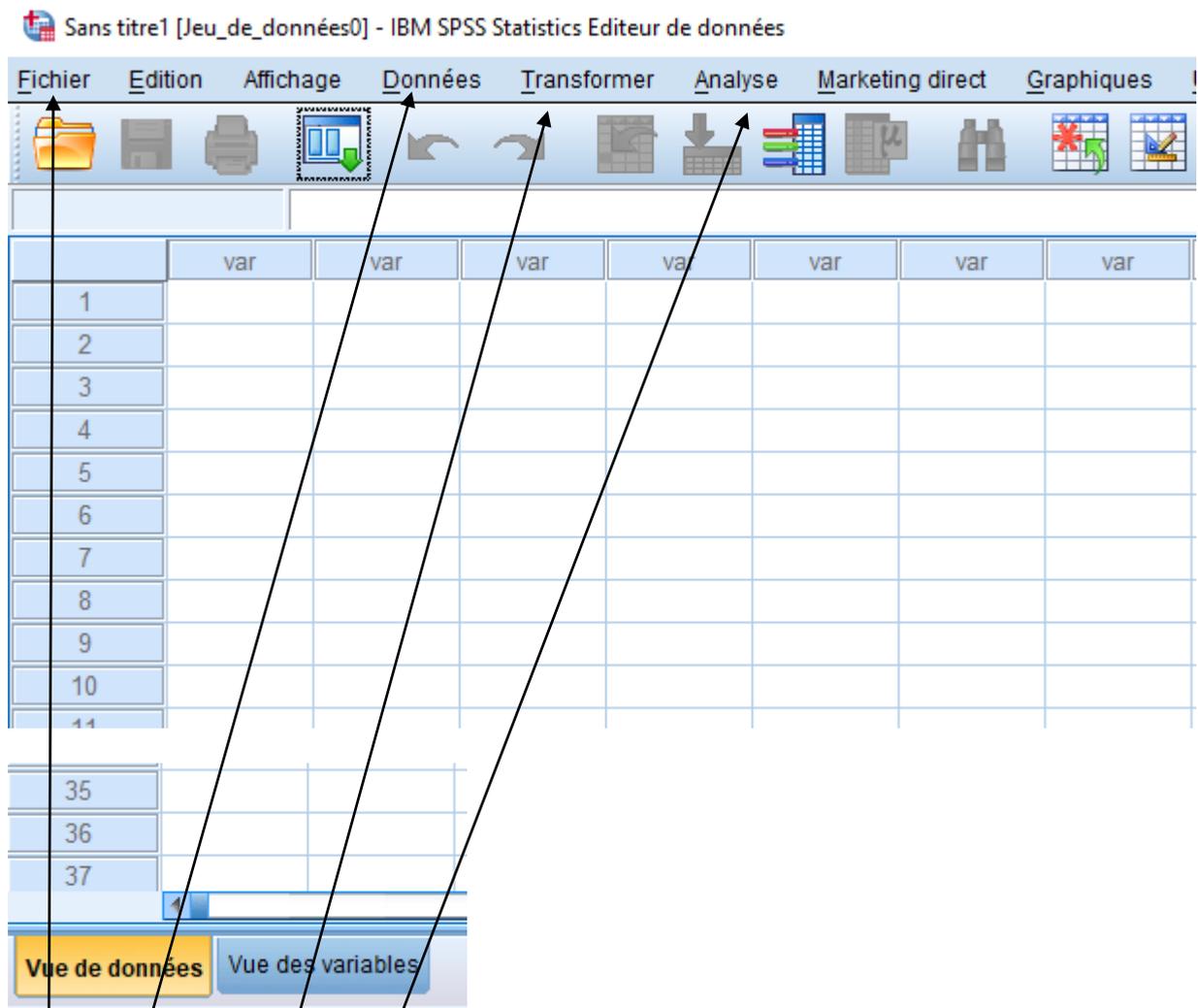
* 1 ligne par individu : champs des variables codés avec des chiffres le plus souvent. Nom des patients possibles (pour retrouver des dossiers par exemple). Surtout des codes avec chiffres +++

- Penser à

- Sauvegarder les données saisies sur tous les supports possibles (dont envoi de mails)
- Noter sur chaque fiche le numéro de saisie en informatique pour y revenir si nécessaire
- Détecter les erreurs de réponses : par exemple, la réponse est "non", et la case "si oui" est remplie.....

Présentation générale

Ecran d'accueil



Zone des données (format/tableur)

Ligne de commande :

- **Maniement des fichiers** : ouvrir (fichiers de tous types de formats), enregistrer (idem)
- **Données** : sélectionner des observations (homme, femmes, âge < 30 ans... etc)
- **Transformer les données** : type ou nom de variables, calculs de seuils ou de classes
- **Analyser les données** : décrire, comparer....

...

Zone des variables (deuxième onglet)

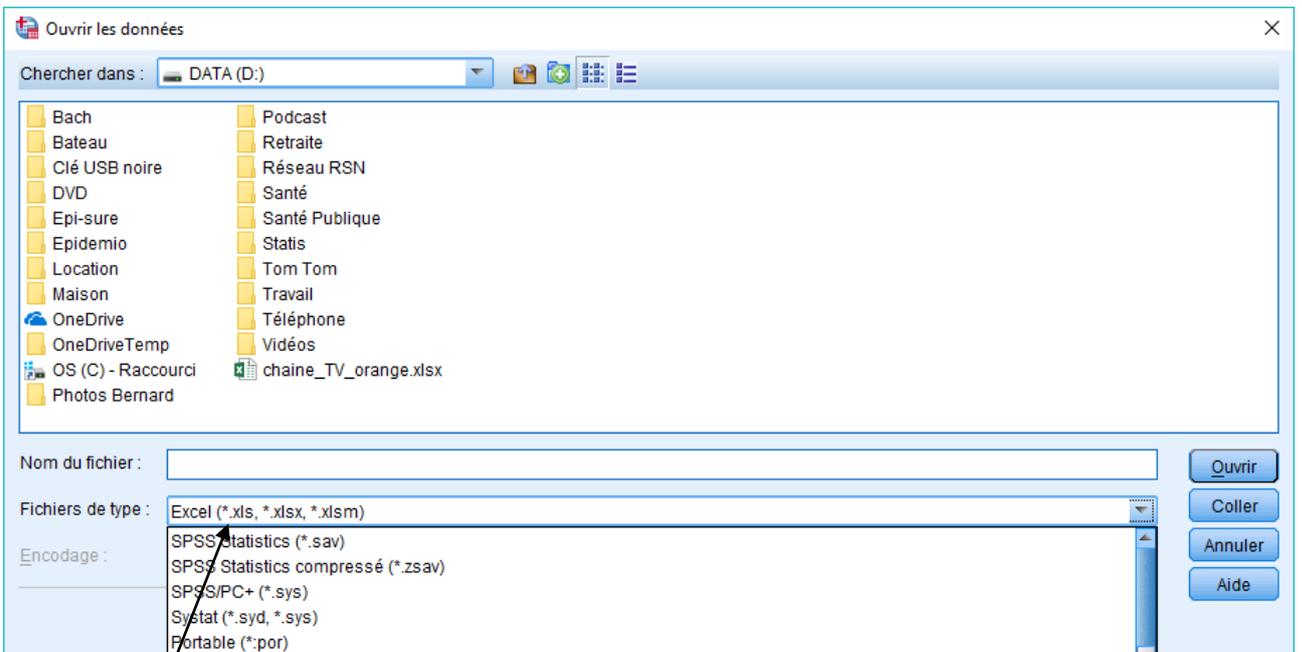
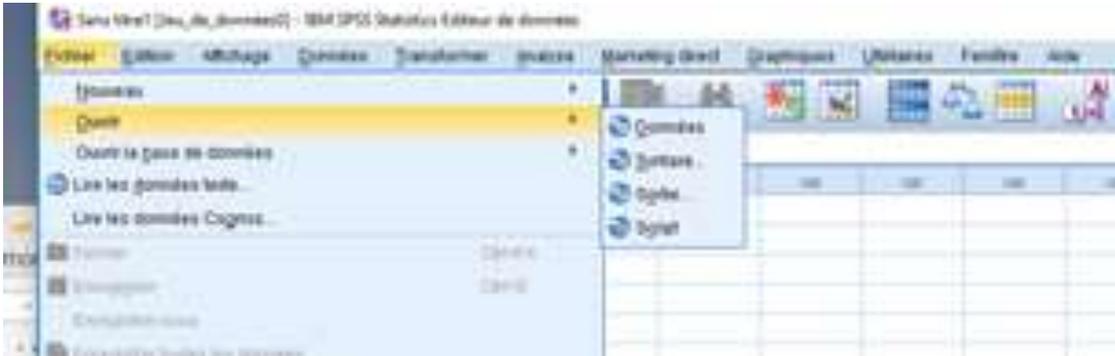
35	
36	
37	
38	
	1
Vue de données	Vue des variables

	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

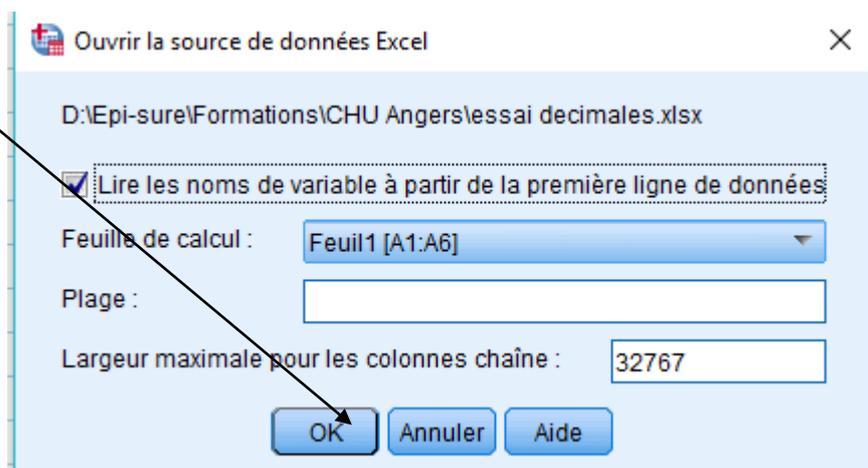
	Nom	Type	Largeur	Décimales	Libellé	Valeurs
1	Numero	Numérique	8	0		Aucun
2	Nom	Chaîne	8	0		Aucun
3	Prenom	Chaîne	8	0		Aucun
4	Date_naiss	Date	11	0		Aucun
5	Age_gesta	Numérique	4	2		Aucun
6	Poids	Numérique	3	0		Aucun
7	Taille	Numérique	3	2		Aucun
8	IMC	Numérique	4	2		Aucun
9	PN	Numérique	4	0		Aucun
10	pH	Numérique	3	2		Aucun
11	Apgar	Numérique	2	0		Aucun
12	Déclenchement	Numérique	8	2		{1,00, Propess}...

Ouvrir/ Importer des données

- Cliquer sur Fichier : Ouvrir/ Données
Chercher sur l'ordinateur le fichier de données.
 - Si sous format Excel : format *.xls ou *.xlsx.
 - Si sous format SPSS : format *.sav
 - Autre formats possibles : texte, Stata...
- Import de EPIDATA : ouvrir Syntaxe (cours complémentaire)

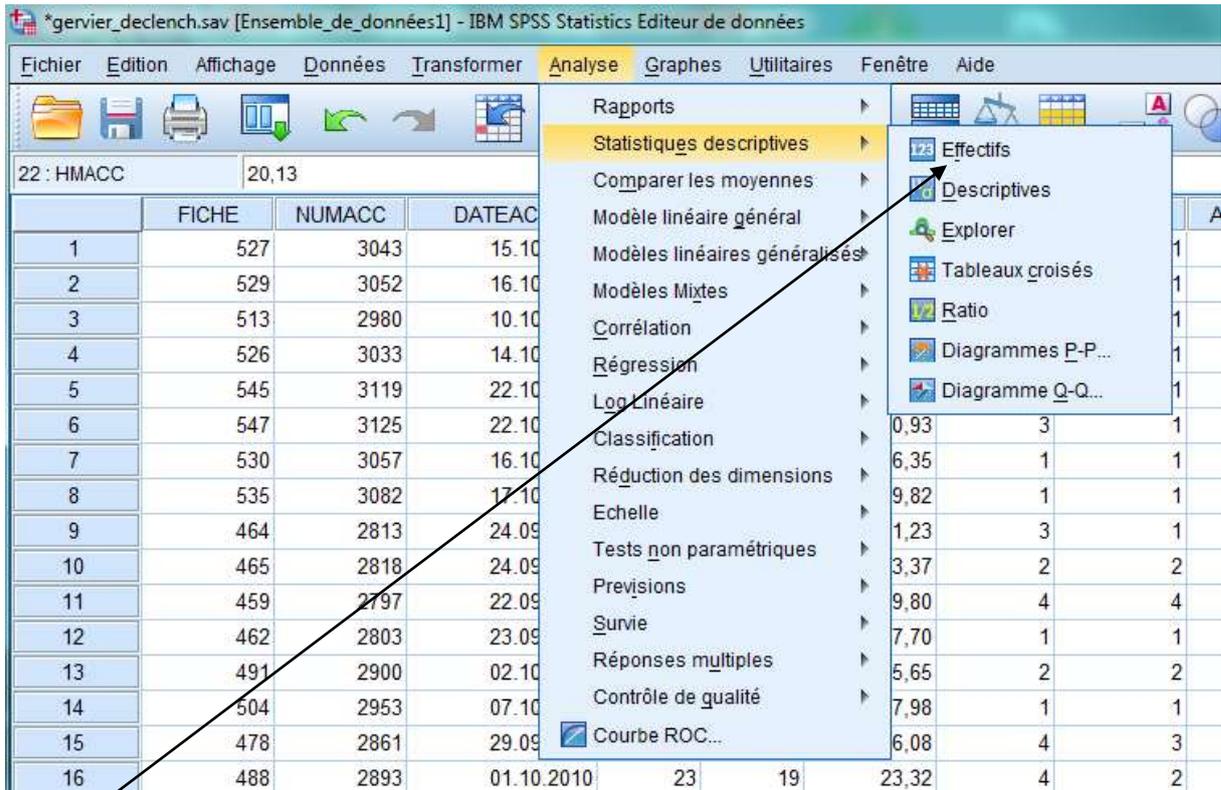


Fichiers Excel
Faire OK



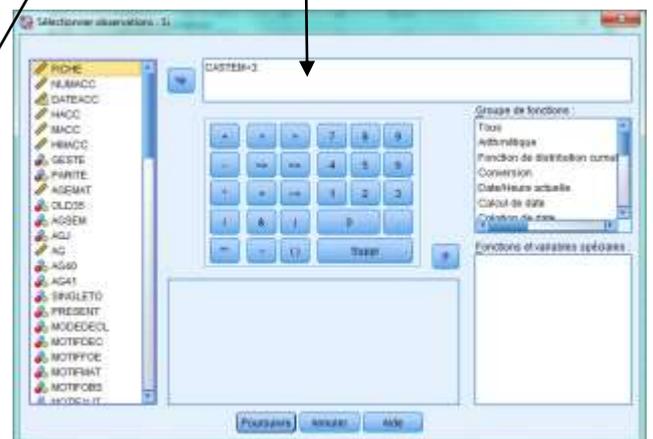
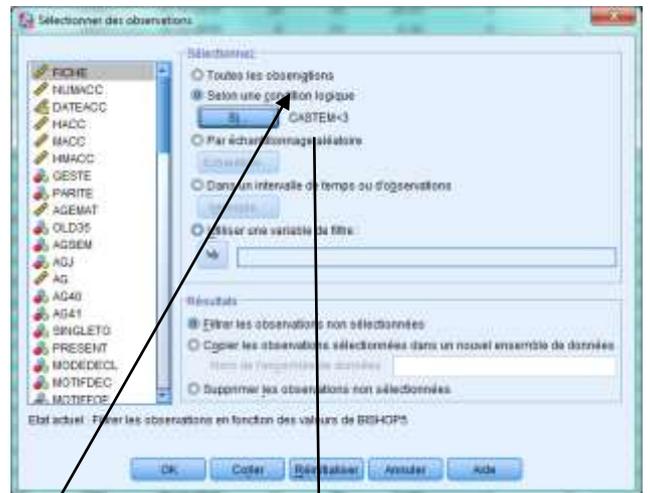
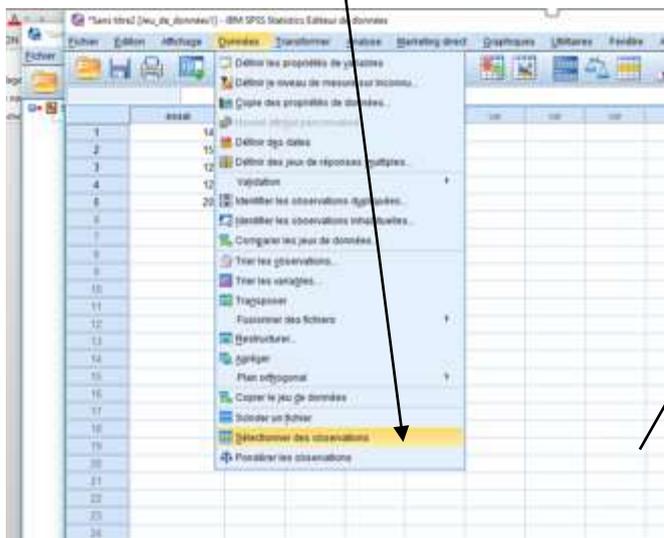
Analyses type

1. Décrire

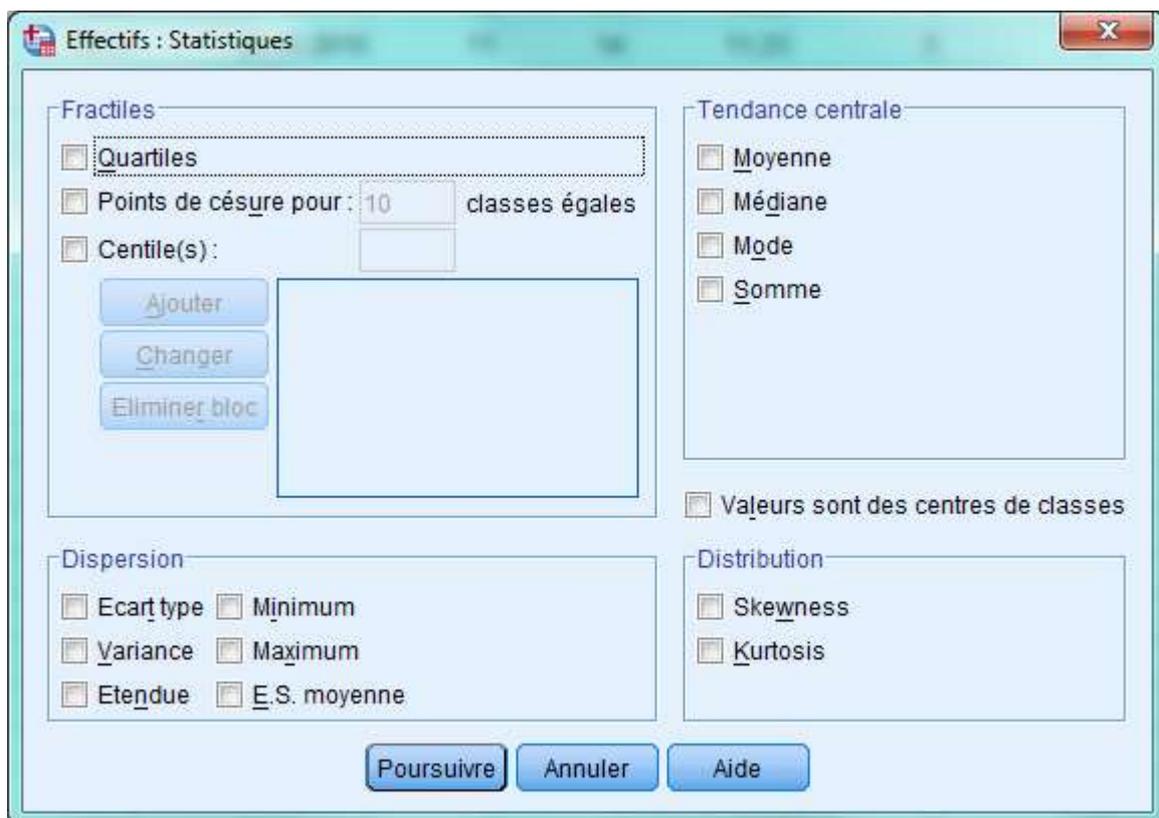


Effectifs

Sélectionner les observations

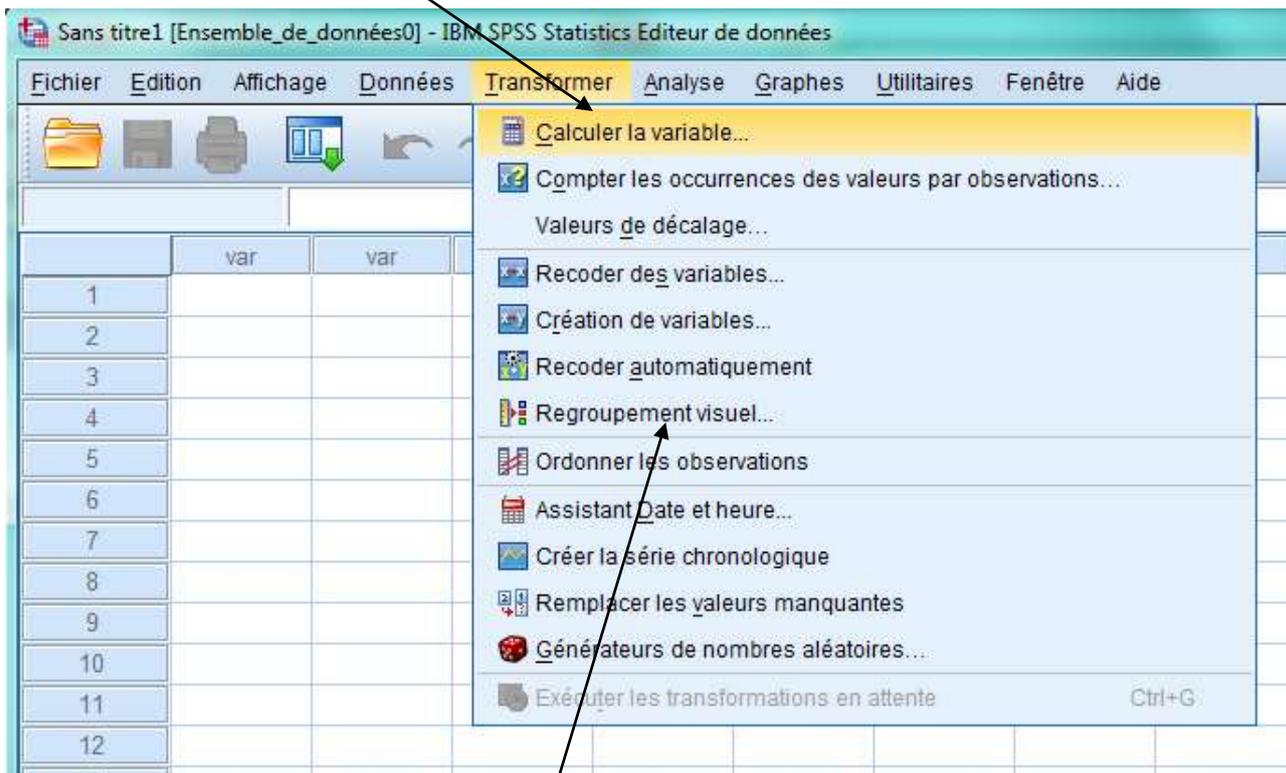


Effectifs (dénombrements)

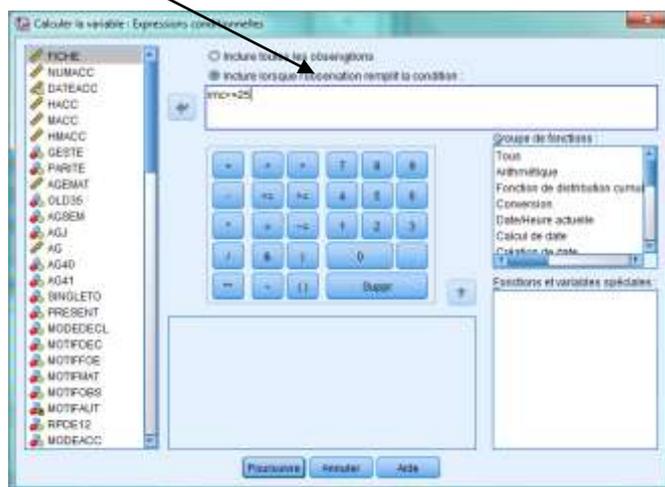
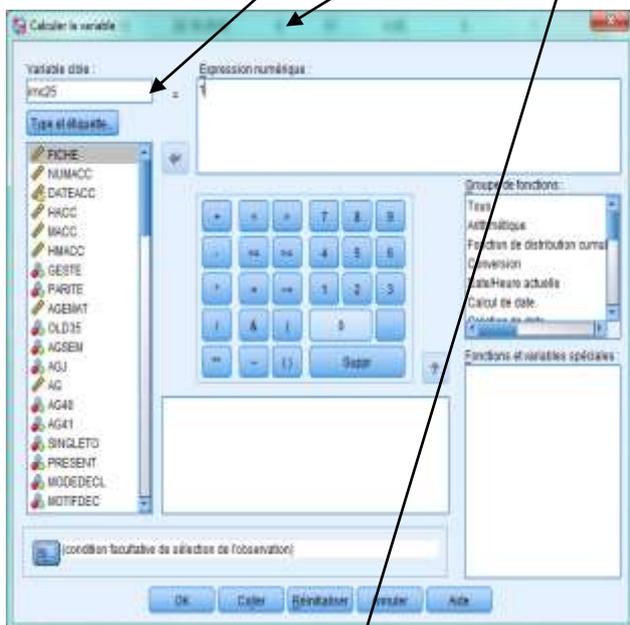


- S'il s'agit d'une **variable qualitative**, on utilise des % (en tenant compte ou non des manquants).
- S'il s'agit d'une **variable quantitative**, deux solutions :
 - = Décrire avec moyenne \pm écart-type avec éventuellement la médiane, l'étendue (range) avec minimum et maximum
 - = Faire des classes (deux avec un seul seuil, ou plus de deux avec autant de seuils que nécessaires). Voir l'onglet Transformer.

Déterminer des seuils, faire des classes



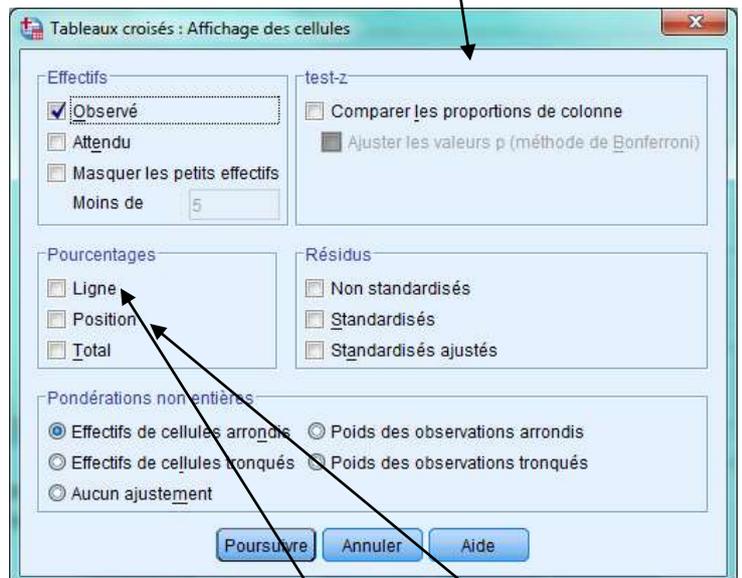
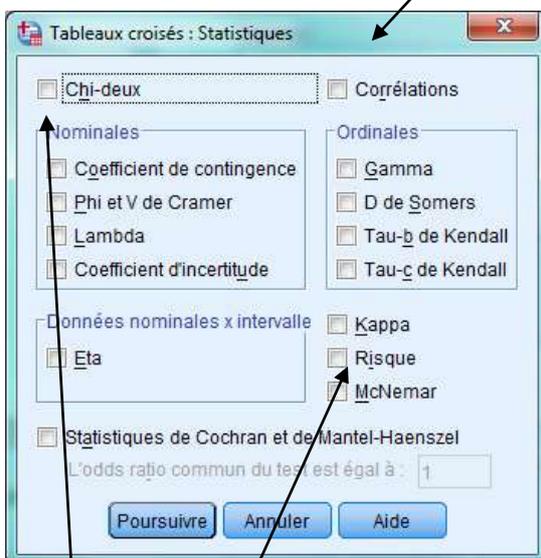
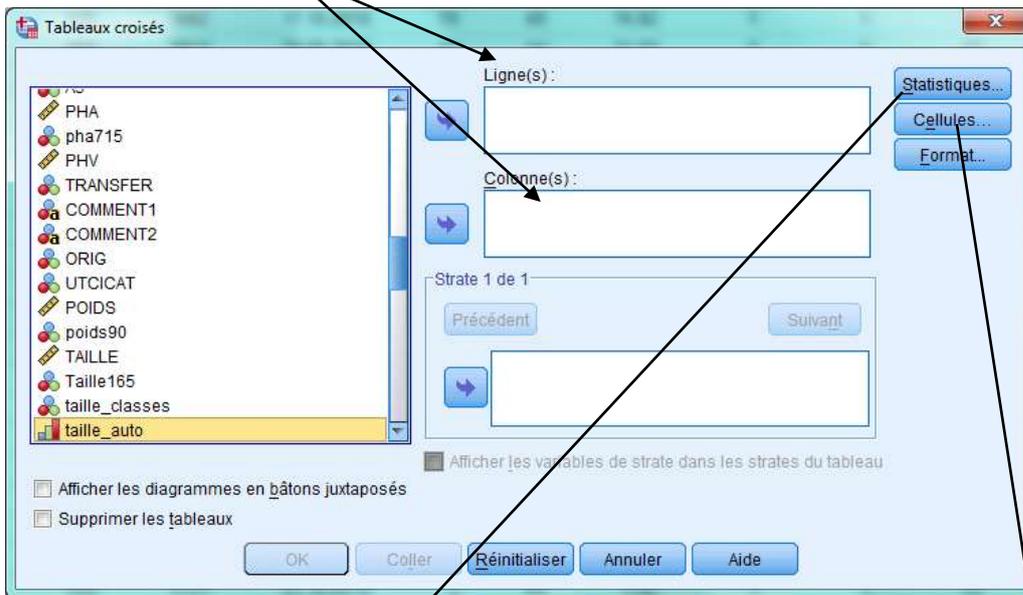
= Définir un seuil d'IMC à 25 : code= 1 si $IMC \geq 25$. Idem (refaire la manip) pour code = 2 si $IMC < 25$). Penser à cocher successivement sur Poursuivre puis OK après chaque sélection.



= Faire des classes "régulières" (poids de naissance par 500 g)

2. Comparer

= Analyse univariée : Tableaux croisés : comparaison de %



- Test du χ^2 (Fisher aussi si effectifs calculés < 5)
- Calcul des RR : c'est le rapport d'incidence de la maladie pour les exposés versus les non-exposés → donc RR en ligne
- Calcul des OR : c'est le ratio de deux odds (a : c / b : d) : soit = ad/bc → bien vérifier le sens +++. **Voir page 19 pour les détails des RR et des OR.**

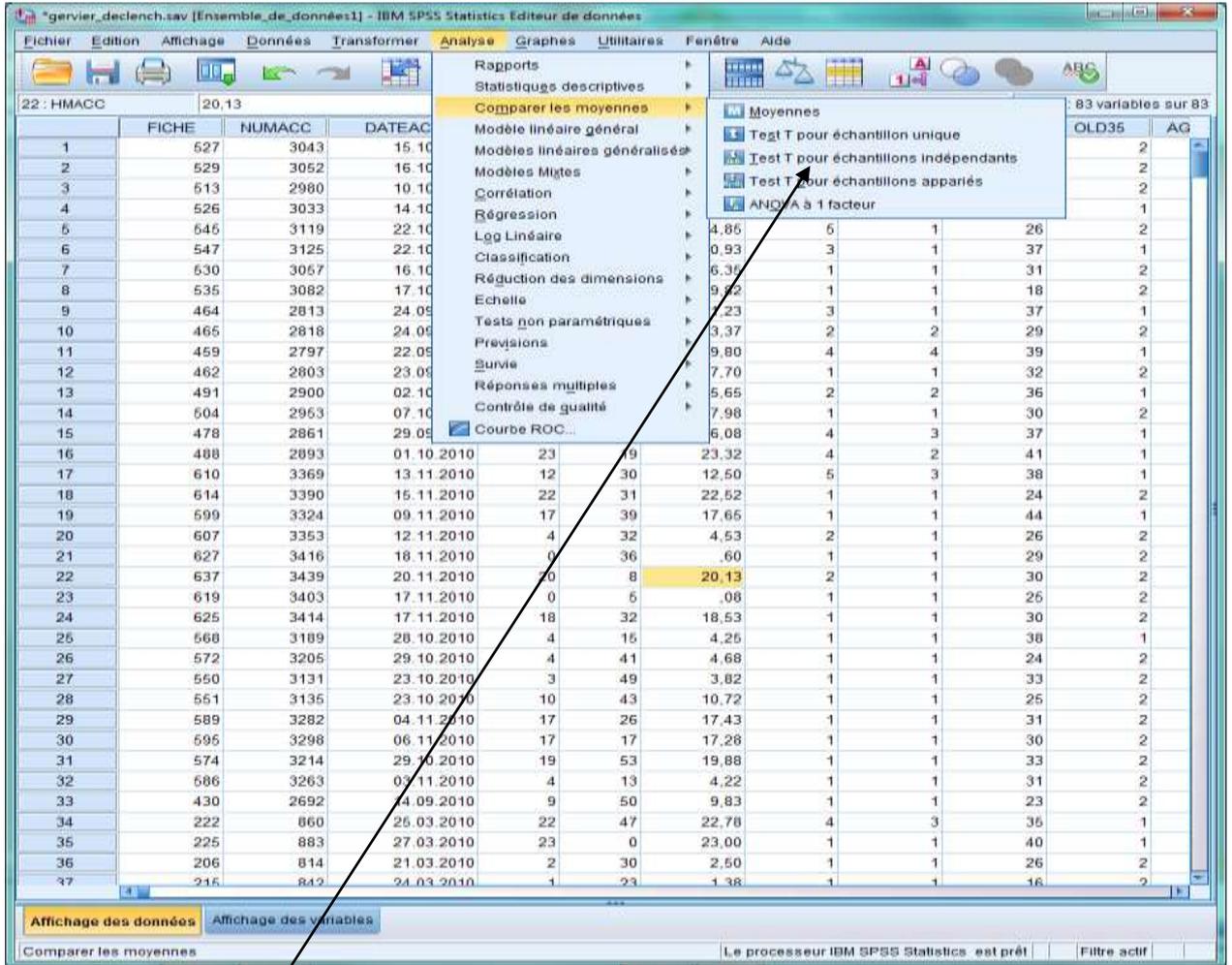
Pourcentage en lignes ou en colonnes

Attention au sens des codages : toujours par ordre croissant

	Malade + code = 1	Malade - (témoin) code = 2	Total
Exposé code = 1	a	b	
N exposé code = 2	c	d	
Total			

Si malade codage = 1 et non-malade codage = 0: nouvel OR ou RR= 1/OR ou 1/RR ancien (par exemple)

= Analyse univariée : Comparaison de moyennes (test t ou ANOVA)



Test de t standard



Préciser les codages : ici 1 et 2
 Pour comparaison de moyennes avec plus de 2 codages, voir ANOVA (infra)

Description des 3 moyennes

Descriptives									
AGEMAT Age mat									
	N	Moyenne	Ecart type	Erreur standard	Intervalle de confiance à 95 % pour la moyenne		Minimum	Maximum	
					Borne inférieure	Borne supérieure			
1	120	30,06	5,666	,517	29,03	31,08	19	44	
2	148	29,95	5,385	,443	29,08	30,83	16	41	
3	26	29,35	4,824	,946	27,40	31,29	17	37	
Total	294	29,94	5,442	,317	29,32	30,57	16	44	

ANOVA

AGEMAT Age mat					
	Somme des carrés	ddl	Carré moyen	F	Sig.
Intergroupes	10,872	2	5,436	,183	,833
Intragroupes	8665,145	291	29,777		
Total	8676,017	293			

Le « p » est là : ici non significatif : les 3 moyennes ne sont pas différentes

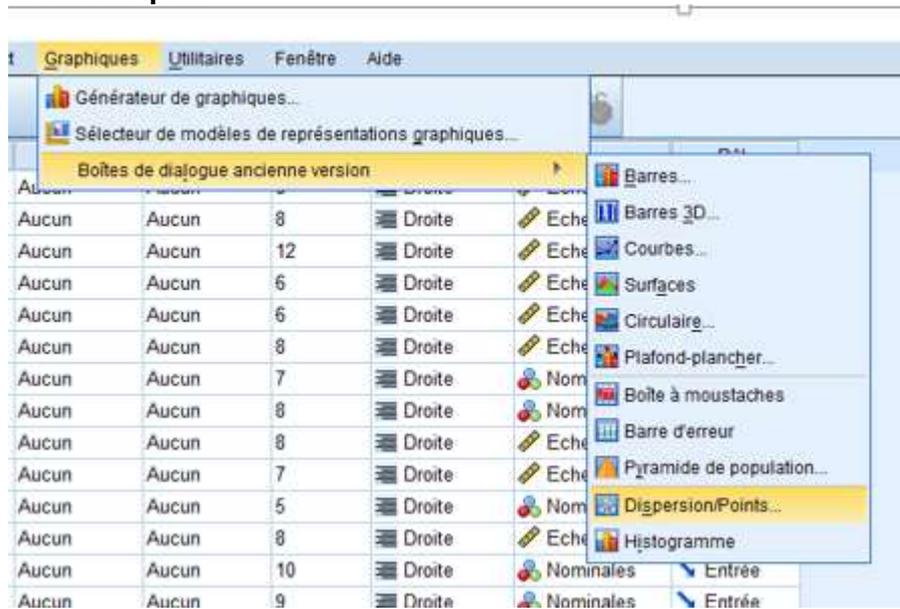
Si la différence entre les 3 moyennes est significative, on peut comparer les moyennes deux-à-deux selon des tests précis (nombreux et complexes) : cocher "Post-Hoc". Retenir le test de Bonferroni et celui de Scheffé.

Tests post hoc

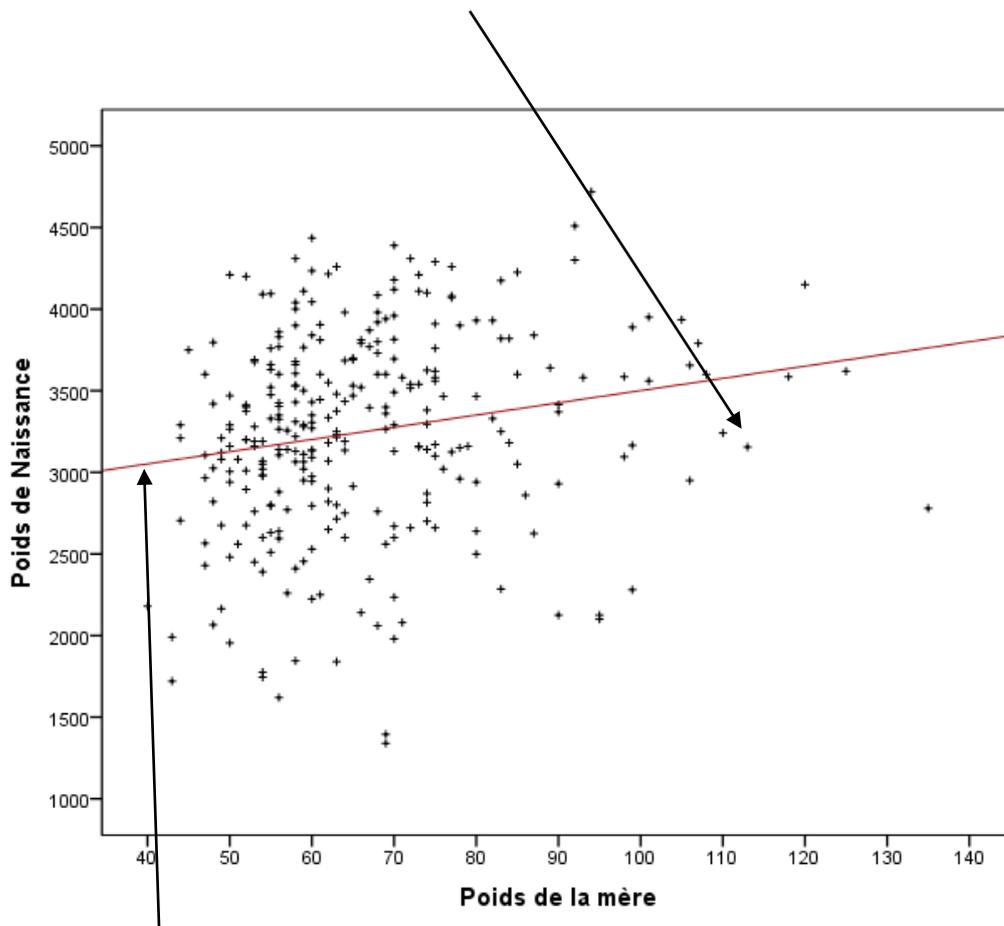
Comparaisons multiples :								
Variable dépendante: AGEMAT Age mat								
	(I) MODEDECL Mode decl	(J) MODEDECL Mode decl	Différence moyenne (I-J)	Erreur standard	Sig.	Intervalle de confiance à 95 %		
						Borne inférieure	Borne supérieure	
Scheffé	1	2	,106	,670	,988	-1,54	1,75	
		3	,712	1,180	,834	-2,19	3,62	
	2	1	-,106	,670	,988	-1,75	1,54	
		3	,607	1,160	,872	-2,25	3,46	
	3	1	-,712	1,180	,834	-3,62	2,19	
		2	-,607	1,160	,872	-3,46	2,25	
Bonferroni	1	2	,106	,670	1,000	-1,51	1,72	
		3	,712	1,180	1,000	-2,13	3,55	
	2	1	-,106	,670	1,000	-1,72	1,51	
		3	,607	1,160	1,000	-2,19	3,40	
	3	1	-,712	1,180	1,000	-3,55	2,13	
		2	-,607	1,160	1,000	-3,40	2,19	

= Analyse univariée : lien entre deux variables quantitatives

1. Scatter plot



Choisir la variable quantitative explicative en abscisse (en X) et celle à expliquer (en Y)
Dans le module "Graphe", insérer une courbe "d'ajustement au total". Chaque point est une paire de valeurs : exemple avec Poids \approx 108 kg et un PN de 2900 g. La courbe ascendante (pente >0) veut dire que plus le poids de la mère est élevé, plus le PN est important.



Droite dite "de régression" dont il faut déterminer l'équation (infra)

= **Analyse multivariée** (initiation)

Régression linéaire multiple ou logistique

	FICHE	NUMACC	DATEAC		GESTE	PARITE	AGEMAT	
1	527	3043	15.10.2010	23	1	1	33	
2	529	3052	16.10.2010	19	1	1	20	
3	513	2980	10.10.2010	12	2	1	26	
4	526	3033	14.10.2010	19	1	1	33	
5	545	3119	22.10.2010	23	1	1	33	
6	547	3125	22.10.2010	19	1	1	33	
7	530	3057	16.10.2010	12	1	1	33	
8	535	3082	17.10.2010	19	1	1	33	
9	464	2813	24.09.2010	12	1	1	33	
10	465	2818	24.09.2010	12	1	1	33	
11	459	2797	22.09.2010	12	1	1	33	
12	462	2803	23.09.2010	12	1	1	33	
13	491	2900	02.10.2010	12	1	1	33	
14	504	2953	07.10.2010	12	1	1	33	
15	478	2861	29.09.2010	12	1	1	33	
16	488	2893	01.10.2010	23	1	1	33	
17	610	3369	13.11.2010	12	5	3	38	

Régression logistique

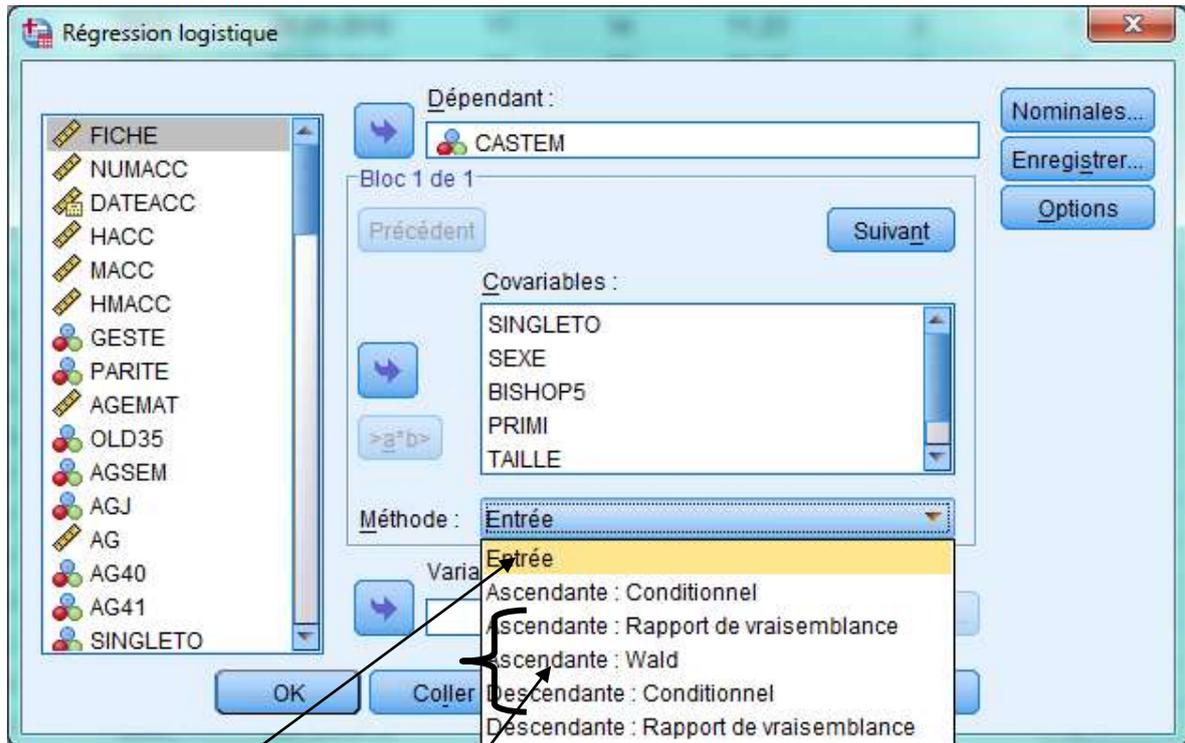
= Variable dépendante binaire : à expliquer : codage en binaire (1 versus 0, ou 2 versus 1)

= Variables explicatives : le nombre que l'on veut

→ Variables codées ou qualitatives : référence : le premier ou le dernier codage

→ Cocher CI pour exp(B)

A noter : le coefficient de l'équation de régression logistique est le log de l'OR. Donc pour passer du coefficient A (ou B) à un OR, le logiciel effectue la transformation par e^A (ou $\exp(A$ ou B)).



Méthode d'introduction des variables :

- Modèle complet (entrée)
- Modèle pas-à-pas ascendant : les variables les plus significatives d'abord
- Modèle pas-à-pas descendant : les variables les moins significatives sont éliminées

Résultats à lire

1. S'assurer des effectifs : une valeur manquante supprime l'observation
2. S'assurer du codage "dans le bon sens" de la variable dépendante (critère de jugement) : les OR seront calculés du "1" par rapport à "0"

Régression logistique

Récapitulatif de traitement des observations

Observations non pondérées ^a		N	Pourcentage
Observations sélectionnées	Incluses dans l'analyse	209	71,1
	Observations manquantes	85	28,9
	Total	294	100,0
Observations non sélectionnées		0	,0
Total		294	100,0

a. Si la pondération est active, consultez la table de classification pour connaître le nombre total d'observations.

Codage de variable dépendante

Valeur d'origine	Valeur interne
0	0
1	1

3. S'assurer du codage des variables qualitatives "dans le bons sans" : "1" par rapport à "0"

Codages des variables catégorielles

		Fréquence	Codage de paramètre	
			(1)	(2)
MODEDECL Mode decl	1	86	,000	,000
	2	103	1,000	,000
	3	20	,000	1,000
BISHOP5	1	116	1,000	
	2	93	,000	
UTCICAT	1	11	1,000	
	2	198	,000	
PRIMI	1	127	1,000	
	2	82	,000	

4. Lire le tableau des résultats

Variables de l'équation

	B	E.S	Wald	ddl	Sig.	Exp(B)	Intervalle de confiance 95% pour EXP(B)	
							Inférieur	Supérieur
Pas 1 ^a PRIM1(1)	2,259	,457	24,477	1	,000	9,574	3,912	23,431
BISHOP5(1)	1,437	,536	7,194	1	,007	4,206	1,472	12,018
IMC	,097	,034	8,364	1	,004	1,102	1,082	1,177
AGEMAT	-,007	,035	,041	1	,840	,993	,927	1,063
AG	-,076	,098	,599	1	,439	,927	,765	1,123
UTCICAT(1)	2,462	,788	9,756	1	,002	11,734	2,502	55,020
MODEDECL			1,759	2	,415			
MODEDECL(1)	-,153	,565	,074	1	,786	,858	,283	2,597
MODEDECL(2)	,825	,685	1,451	1	,228	2,282	,596	8,733
Constante	-1,429	3,913	,133	1	,715	,240		

a. Introduction des variables au pas 1 : PRIM1, BISHOP5, IMC, AGEMAT, AG, UTCICAT, MODEDECL.

"p" des critères : remarquer les $p < 0.05$

OR ajustés (Exp(B))

IC des OR

5. Comprendre la classification du modèle (prédiction / observation)

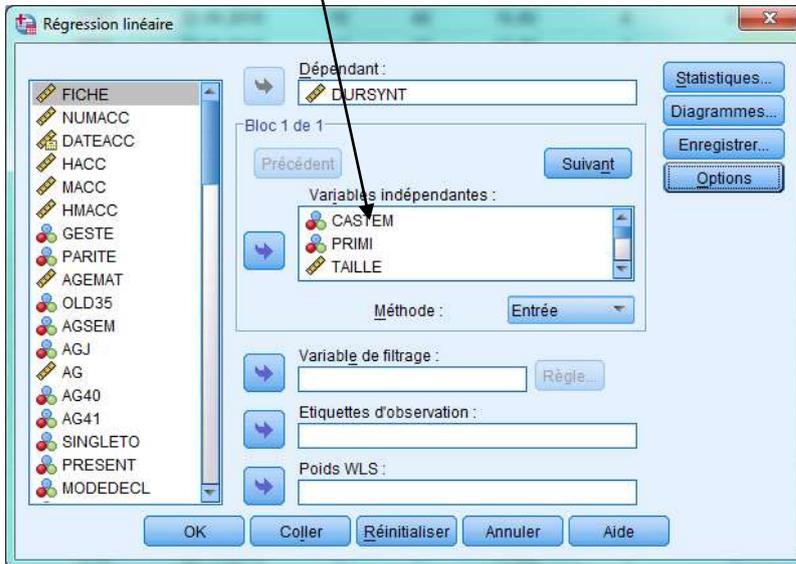
Table de classification^a

Observé	CASTEM	Prévisions		
		CASTEM		Pourcentage correct
		0	1	
Pas 1	0	74	25	74,7
	1	25	85	77,3
Pourcentage global				76,1

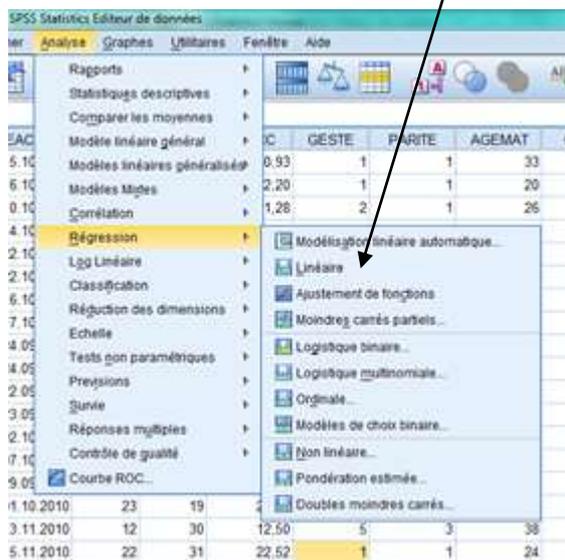
a. La valeur de coupe est ,500

= Régression linéaire multiple

- Variable à expliquer quantitative
- 1 seule variable (explicative) quantitative: régression linéaire simple ($Y = AX + B$)
- Plusieurs variables indépendantes : régression linéaire multiple ($Y = A_1X_1 + A_2X_2 + \dots + A_n X_n + Z$)



Même résultat que la droite de régression linéaire avec une seule variable dépendante. Donne en plus l'équation de la droite. Pour les graphes, voir Graphes/ Dispersion en points



Récapitulatif des modèles

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,365 ^a	,133	,105	18,44543

a. Prédicteurs : (Constante), IMC, AG, PARITE, BISHOP, PN

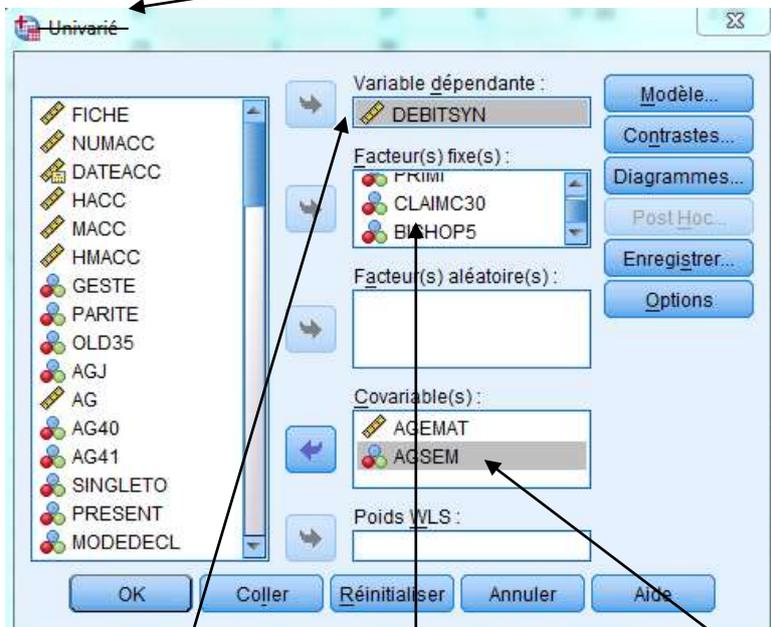
Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés		Sig.
		B	Ecart standard	Bêta	t	
1	(Constante)	28,924	40,912		,707	,481
	PARITE	-3,062	1,446	-,165	-2,118	,036
	AG	-,829	1,118	-,067	-,741	,460
	PN	,008	,004	,208	2,259	,025
	BISHOP	,272	,879	,025	,309	,758
	IMC	,879	,240	,285	3,662	,000

a. Variable dépendante : DEBITSYN debitsynth

- ANOVA avec plusieurs facteurs

Il existe toujours une variable quantitative à expliquer et on teste plusieurs facteurs explicatifs : soit des données codées, soit des données quantitatives (appelées co-variables)
 Pour SPSS 22.0, Analyse/ Modèle linéaire généralisé. En épidémiologie, ça s'appelle « multivarié ». Pour SPSS : « multivarié » veut dire plusieurs variables dépendantes (rare en épidémiologie)



Variable quantitative « à expliquer »

Variables qualitatives explicatives Variables quantitatives comme covariables

Voir le bouton « Options » pour la description des variables,

Voir le bouton « Modèle » pour analyser les facteurs sans interaction (ou avec, après l'étape « sans » si besoin ». Les « p » sont là : sont significatifs ($p < 0.05$) pour expliquer les doses de synto dans des accouchements déclenchés : primipare, et $IMC > 30$, mais le Bishop > 5 , pas l'âge maternel (en quantitatif, ce qui suppose que l'hypothèse de la relation était linéaire), ni l'âge gestationnel (idem).

Tests des effets inter-sujets

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Sig.
Modèle corrigé	5406,109 ^a	5	1081,222	3,346	,006
Ordonnée à l'origine	1755,106	1	1755,106	5,432	,021
PRIMI	2536,823	1	2536,823	7,851	,006
CLAIMC30	1839,652	1	1839,652	5,693	,018
BISHOP5	215,478	1	215,478	,667	,415
AGEMAT	26,421	1	26,421	,082	,775
AGSEM	232,324	1	232,324	,719	,397
Erreur	68178,659	211	323,122		
Total	451202,505	217			
Total corrigé	73584,768	216			

- **Modèle de Cox** : le critère est une durée de survie.

Les graphes avec SPSS : possibilité de mise en page performant. Voir aussi les graphes EXCEL

- Graphes possibles :

Univariés

- Pour les données qualitatives : graphes en barres (non jointives) verticales ou horizontales, camembert en principe interdit
- Pour les données quantitatives : histogramme (barres jointives) à éviter, box-plot
- Courbes de survie, courbes ROC

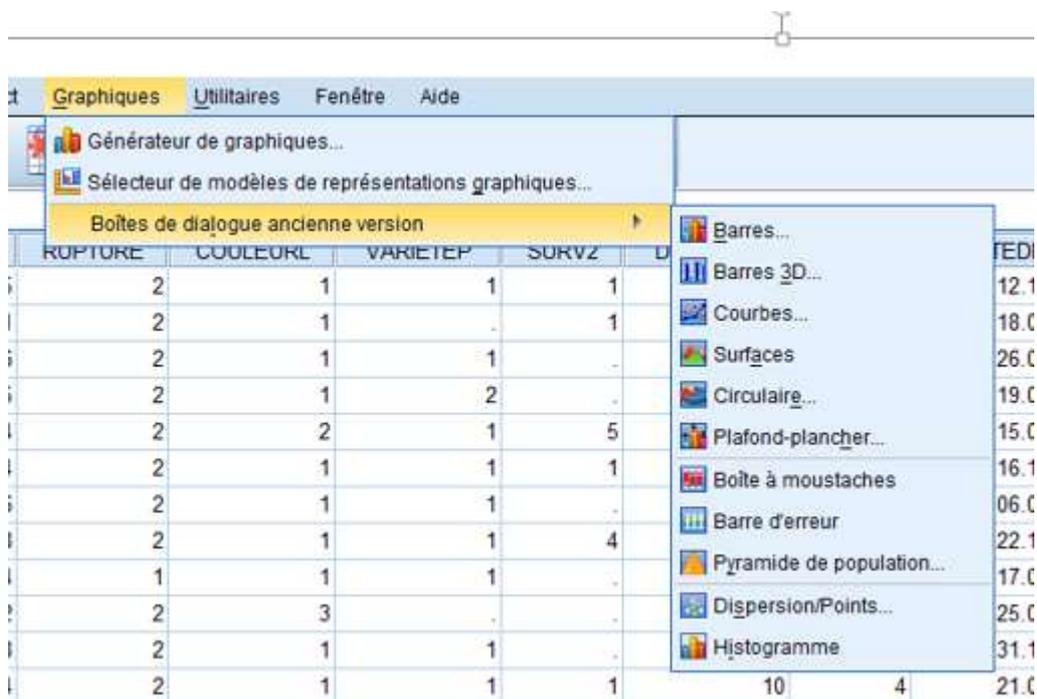
Bivariés

- Pour deux données qualitatives : Doubles barres
- Pour deux données quantitatives : Graphes en points avec la variable à expliquer en ordonnée; droite de régression
- Pour une donnée quantitative en fonction d'une variable qualitative : histogramme avec doubles barres
- En fonction du temps : temps en abscisse

- Graphes interdits

- En 3 D, en camembert (surtout le camembert 3D)
- Tout graphe lisible en plus de 10 secondes...

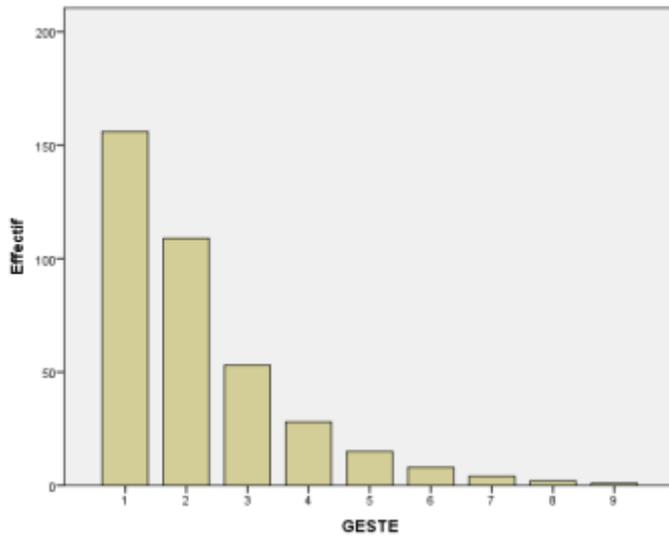
- Chemin



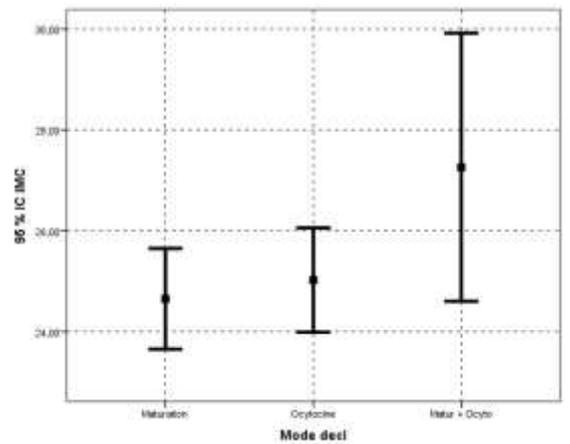
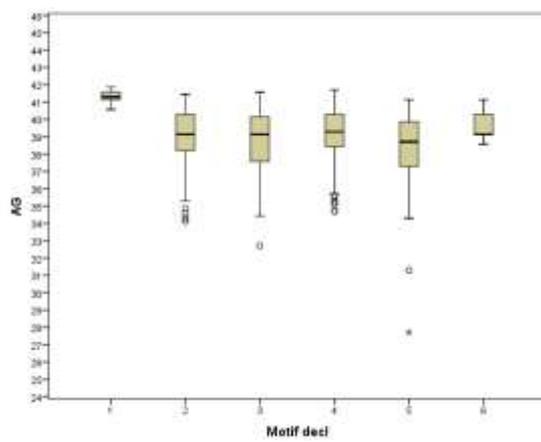
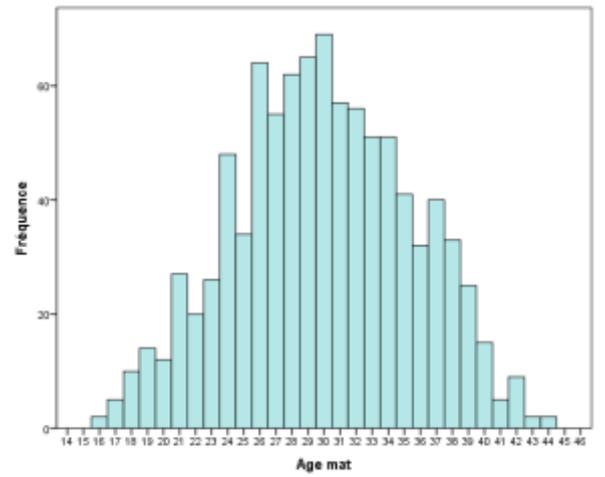
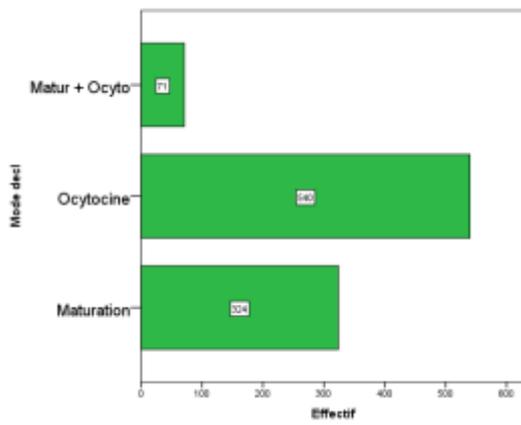
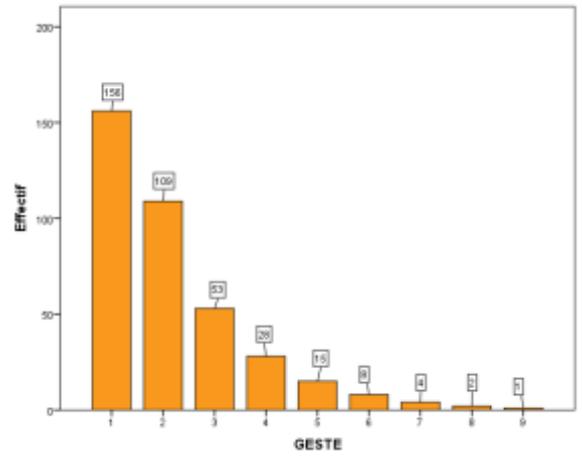
- Suivre les instructions
- Double clic sur le graphe pour modification → formation spécifique
- Copier dans Word au final

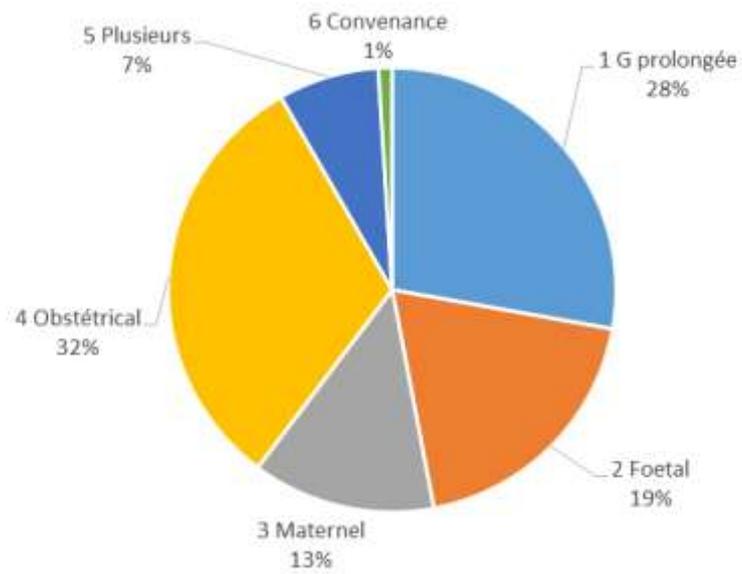
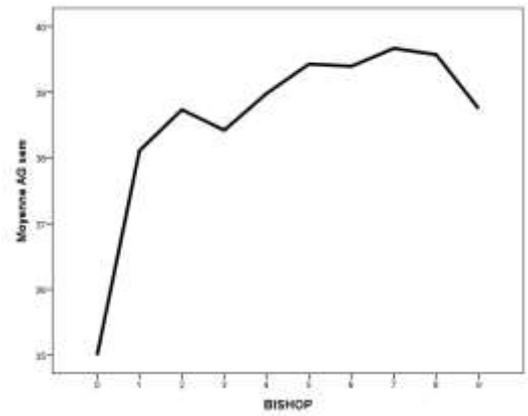
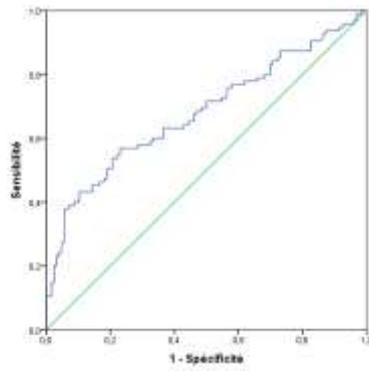
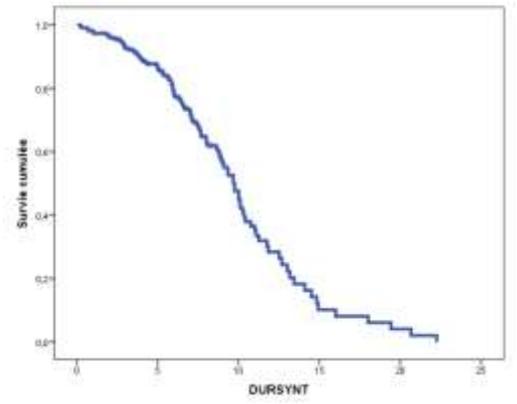
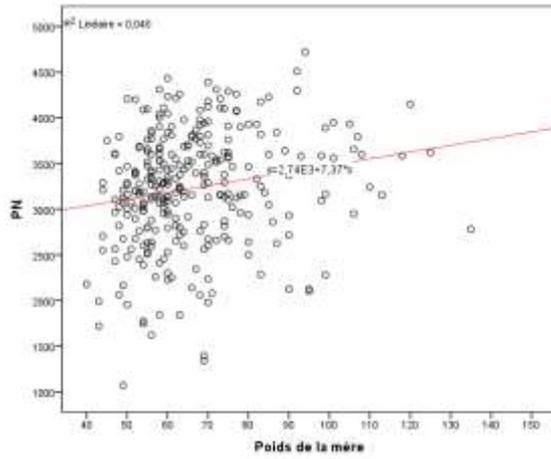
- Exemples

Graphe (brut)



Graphe (modifié)





=====

Risque relatif et odds ratio sous SPSS

= Tableau de base 2 x 2

Proportions "horizontales" = RR

	Malades ou Cas	Non-malades ou Témoins	Total des expositions/ non
Exposés	a	b	Total des exposés
Non-exposés	c	d	Total des non-exposés
	Total des malades ou cas	Total des non-malades ou témoins	Total population

Proportions "verticales" = OR

Prendre comme exemple : tabac et cancer du poumon. Publications de 1950 et 1956.

= Définitions

- Il s'agit d'une **enquête d'incidence** : on suit une population déterminée dans un temps déterminé. On peut parler de malades et de non-malades. On peut faire des rapports d'incidence = risque relatif
- Il s'agit d'une **enquête cas-témoins** : on sélectionne des cas, et on prend autant de témoins (ou 2 ou 3 fois plus quelquefois). On parle de cas et de témoins. On ne peut calculer d'incidence. On fait des rapports des expositions = odds ratio

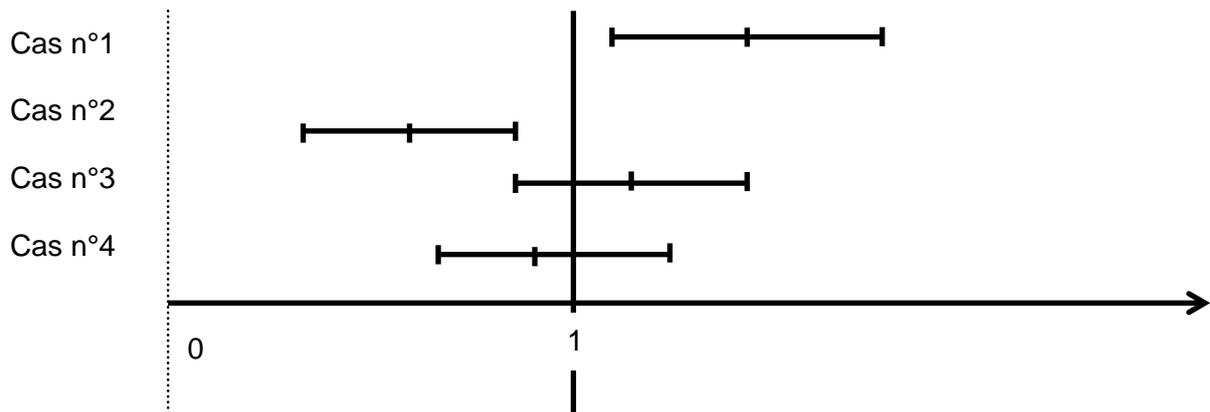
= Calculs théoriques

- **Risque relatif = RR = rapport d'incidence. Les proportions sont horizontales ++**
 - ✓ Incidence chez les exposés = Malades exposés / ensemble des exposés = $a / a+b$
 - ✓ Incidence chez les non-exposés : Malades non exposés / ensemble des non-exposés = $c / c + d$
 - ✓ $RR = (a/a+b) / (c/c+d)$
 - ✓ $RR > 1$: l'exposition est liée à la maladie; $RR < 1$: l'exposition est liée à l'absence de maladie (ex : sport, vaccins)
- **Odds ratio (ou rapport des cotes) = OR ou RC = rapport des expositions. Les proportions sont verticales ++**
 - ✓ Odd de l'exposition chez les cas : a / c .
 - Ce n'est pas la fréquence des expositions (qui serait $a / (a+c)$)
 - ✓ Odd de l'exposition chez les témoins : b / d
 - ✓ Rapport (ratio) des odds : $OR = (a/c) / (b/d) = ad / bc$
 - ✓ $OR > 1$: l'exposition est liée à la maladie; $OR < 1$: l'exposition est liée à l'absence de maladie (ex : sport, vaccins)

= Intervalle de confiance des RR ou OR (IC 95 %)

Sans entrer dans les calculs mathématiques, on peut dresser l'interprétation des RR ou OR (figure)

Figure 1 : Valeur et intervalle de confiance d'un RR : exemple fictif de 4 cas différents



Pour le cas n°1, le $RR > 1$ et significatif (c'est à dire significativement différent de 1). Pour le cas n°2, le $RR < 1$ et significatif. Pour le cas n°3, le $RR > 1$ et non significatif. Pour le cas n°4, le $RR < 1$, et non significatif.

= Précautions de l'utilisation du RR/OR

- La variable de maladie doit être **dichotomique** (en deux classes). Définition précise.
- La variable d'exposition doit être **dichotomique** : facile si cette présentation correspond à une variable naturellement à 2 classes (césariennes : oui, non). Des regroupements sont nécessaires dès que la variable est qualitative à plus de 2 classes (monitoring normal, anormal, non fait). Pour des variables quantitatives, des seuils ("cutpoints") doivent être trouvés : moyenne, médiane, référence admise (mères de plus de 35 ans, ou mortalité prématurée des adultes au-delà de 65 ans, nombre de cigarettes > 20 / jour..); toujours délicat, et le début des "petites bidouilles".
- **Les codages doivent correspondre au sens du tableau 2 x 2 : les cas/malades exposés doivent toujours être dans le "coin gauche" +++.** Les logiciels proposent le plus souvent le sens arithmétique ou alphabétique progressif, mais cet ordre peut être renversé dans les options. C'est pourquoi, le codage "1-Oui, 2-Non" est préférable à "0-Non, 1-Oui". Si la maladie ou l'exposition est codée "Oui/Non", ou "Yes/no", le "Non" apparaît en premier dans la 1^{ère} colonne ou la 1^{ère} ligne, ce qui est gênant.
- **RR ou OR significatif ne veut pas dire cause, mais association +++.** Toujours se méfier d'un autre facteur, lui causal, qui peut expliquer les deux variables (la première variable, celle du RR, est dite alors facteur de confusion). Exemple : les femmes qui préparent l'accouchement allaitent plus souvent que les autres ($RR=2$). Conclusion erronée : c'est la préparation qui favorise l'allaitement. En fait, les femmes qui vont allaiter, préparent plus que les autres. Un tiers facteur (milieu familial et culturel, antécédents...) explique à la fois la préparation et l'allaitement. C'est ce que l'on appelle **un facteur de confusion**, qui se lève à l'analyse multivariée+++

Exemple RR sous SPSS → Cohorte

Tableau croisé PRIM1 * CASTEM

			CASTEM		Total
			1	2	
PRIMI 1	Effectif	113	322	435	
	% dans PRIM1	26,0%	74,0%	100,0%	
2	Effectif	33	405	438	
	% dans PRIM1	7,5%	92,5%	100,0%	
Total	Effectif	146	727	873	
	% dans PRIM1	16,7%	83,3%	100,0%	

Tests du khi-deux

	Valeur	ddl	Sig. approx. (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-deux de Pearson	53,302 ^a	1	,000		
Correction pour continuité ^b	51,986	1	,000		
Rapport de vraisemblance	55,831	1	,000		
Test exact de Fisher				,000	,000
Association linéaire par linéaire	53,241	1	,000		
N d'observations valides	873				

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 72,75.

b. Calculée uniquement pour une table 2x2

Estimation du risque

	Valeur	Intervalle de confiance à 95%	
		Plus bas	Supérieur
Rapport des cotes pour PRIM1 (1 / 2)	4,307	2,845	6,519
Pour cohorte CASTEM = 1	3,448	2,395	4,964
Pour cohorte CASTEM = 2	,801	,753	,852
N d'observations valides	873		

$$RR = (113 / 435) / (33 / 438) = 26,0 \% / 7,5 \% = 3,49$$

IC RR =

Exemple OR sous SPSS → Cas-témoins

			CASTEM		Total
			1	2	
PRIMI	1	Effectif	113	55	168
		% dans CASTEM	77,4%	37,2%	57,1%
2	Effectif	33	93	126	
	% dans CASTEM	22,6%	62,8%	42,9%	
Total	Effectif	146	148	294	
	% dans CASTEM	100,0%	100,0%	100,0%	

Tests du khi-deux

	Valeur	ddl	Sig. approx. (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-deux de Pearson	48,584 ^a	1	,000		
Correction pour continuité ^b	46,955	1	,000		
Rapport de vraisemblance	50,191	1	,000		
Test exact de Fisher				,000	,000
Association linéaire par linéaire	48,419	1	,000		
N d'observations valides	294				

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 62,57.

b. Calculée uniquement pour une table 2x2

Estimation du risque

	Valeur	Intervalle de confiance à 95%	
		Plus bas	Supérieur
Rapport des cotes pour PRIMI (1 / 2)	5,790	3,472	9,656
Pour cohorte CASTEM = 1	2,568	1,881	3,507
Pour cohorte CASTEM = 2	,444	,349	,564
N d'observations valides	294		

OR = (113 : 33) / (55 : 93) = (113 x 93) / (55 x 33) = 5.79

OR IC à 95 %

=====