

Corrigé de l'exercice 8

Statistique II – SP08

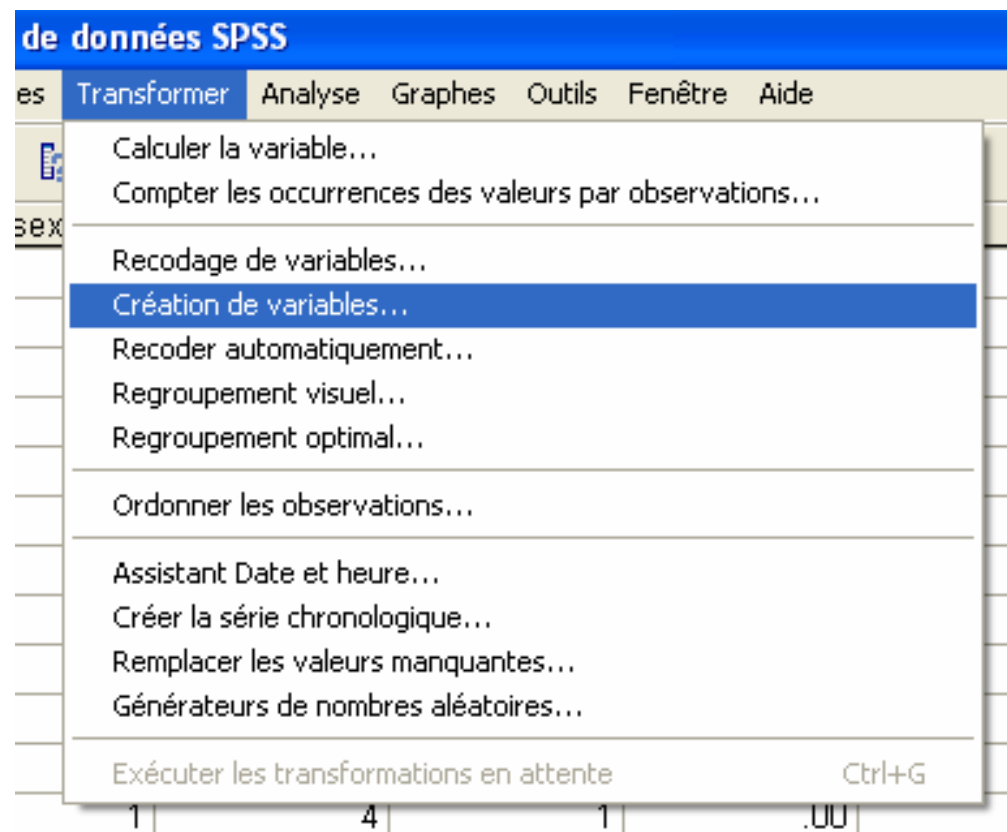
Enseignant: Jean-François Bickel

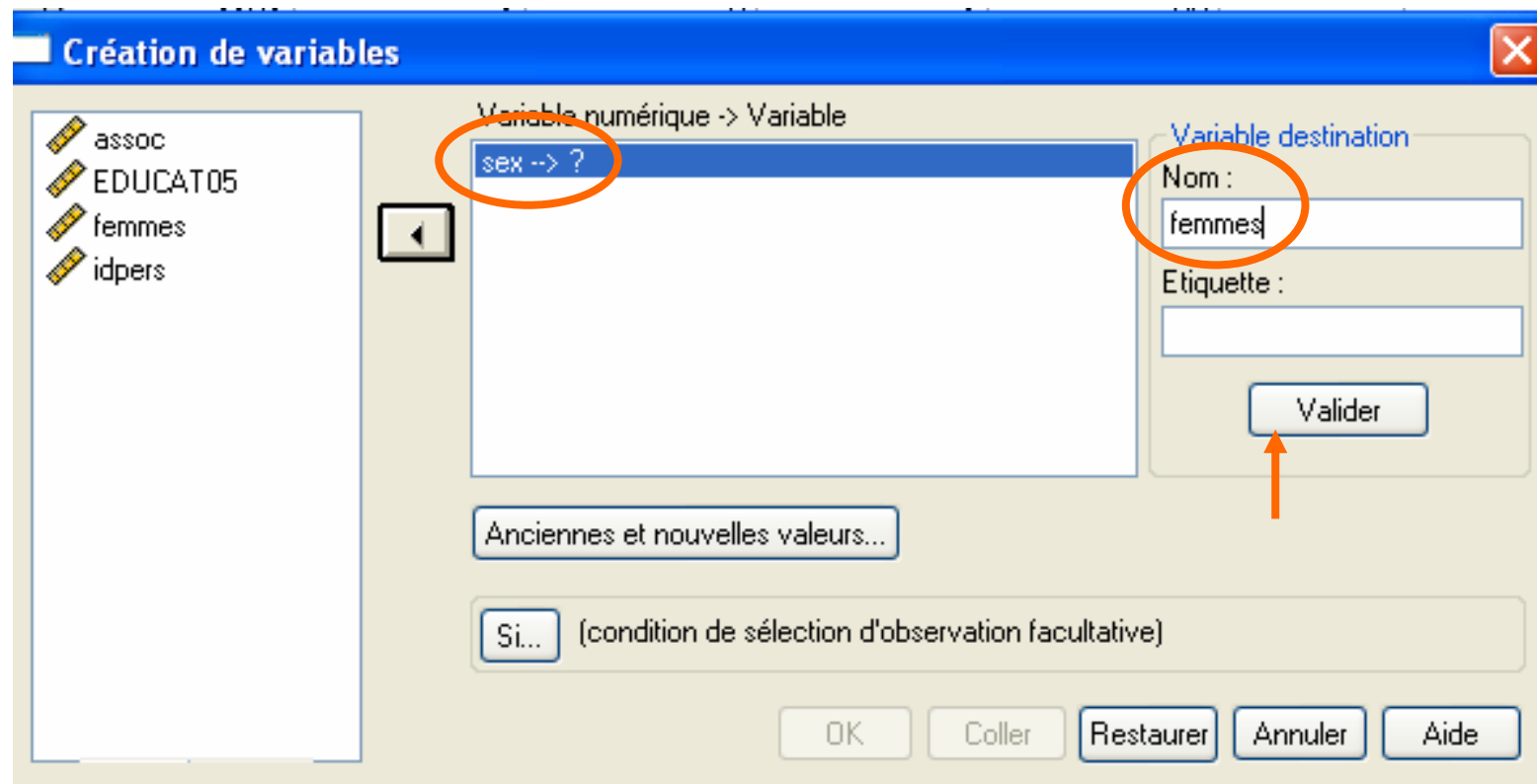
Contenu de l'exercice

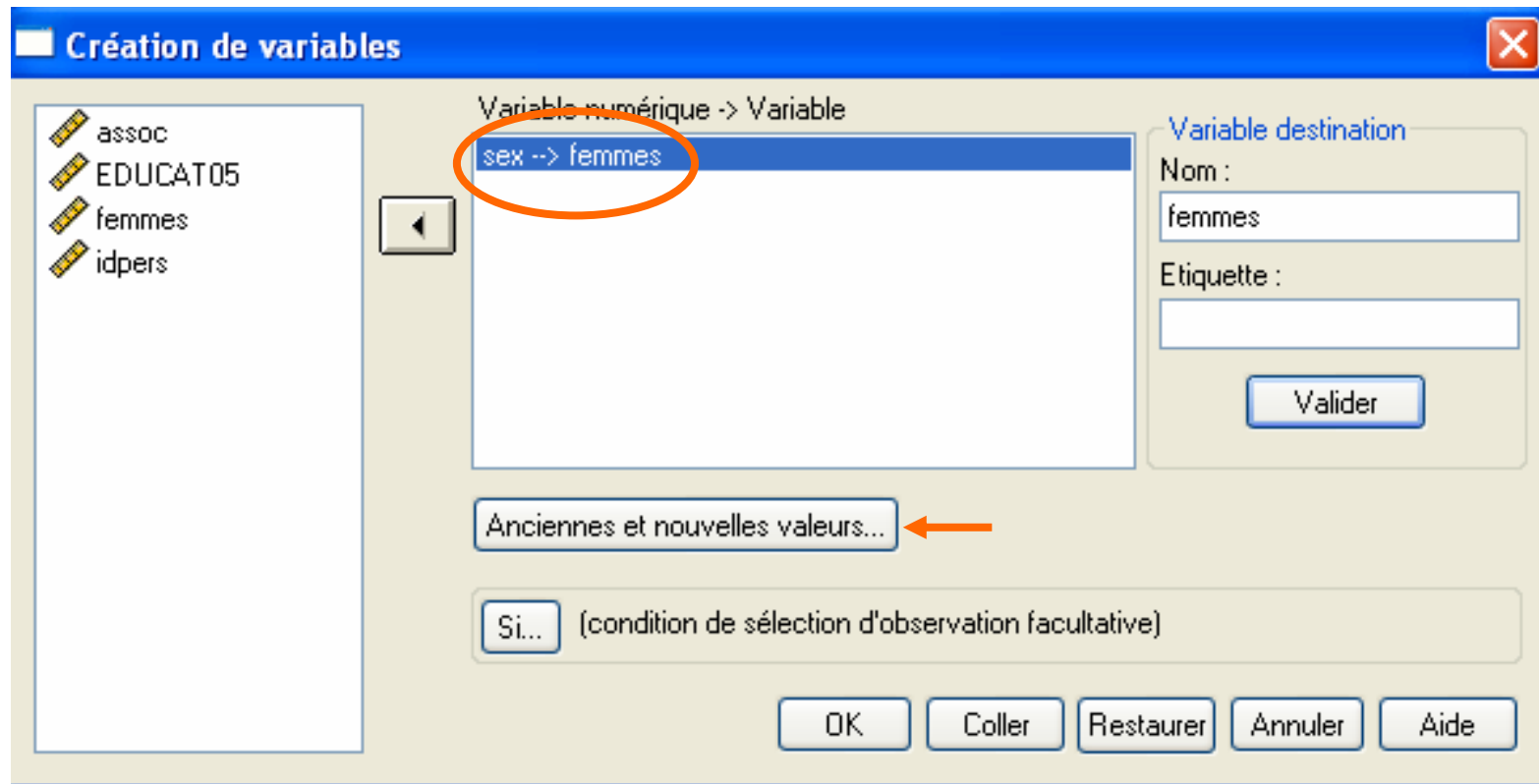
- Sur la base du fichier de données SHP_Ex8.sav, l'exercice consiste en l'analyse, au moyen d'une régression logistique, de la variable nommée assoc « est membre d'une association » en fonction des deux variables indépendantes sexe et éducation
- La variable assoc est codée avec 1=oui (est membre) et 0 =non (n'est pas membre)
- Comme toujours, l'accent est mis sur l'interprétation des résultats produits par SPSS et les conclusions substantielles que l'on peut en déduire

La variable sexe étant catégorielle, on la recode en variable dummy

1) Procédure via l'interface graphique de SPSS







Création de variables : Anciennes et nouvelles valeurs

Ancienne valeur

Valeur :

Manquante par défaut

Manquante par défaut ou spécifiée

Intervalle :

à

Intervalle, du MINIMUM à la valeur :

Intervalle, de la valeur au MAXIMUM :

Toutes les autres valeurs

Nouvelle valeur

Valeur :

Manquante par défaut

Copier les anciennes valeurs

Ancienne --> Nouvelle :

Variables destination sont des chaînes Longueur :

Convertir les chaînes numériques en nombres ('5'→5)

Création de variables : Anciennes et nouvelles valeurs

Ancienne valeur

Valeur :

Manquante par défaut

Manquante par défaut ou spécifiée

Intervalle :

à

Intervalle, du MINIMUM à la valeur :

Intervalle, de la valeur au MAXIMUM :

Toutes les autres valeurs

Nouvelle valeur

Valeur :

Manquante par défaut

Copier les anciennes valeurs

Ajouter → **Ancienne --> Nouvelle :**

Remplacer

Supprimer

Variables destination sont des chaînes Longueur :

Convertir les chaînes numériques en nombres ('5'→5)

Poursuivre **Annuler** **Aide**

Création de variables : Anciennes et nouvelles valeurs ✕

Ancienne valeur

Valeur :

Manquante par défaut

Manquante par défaut ou spécifiée

Intervalle :

à

Intervalle, du MINIMUM à la valeur :

Intervalle, de la valeur au MAXIMUM :

Toutes les autres valeurs

Nouvelle valeur

Valeur :

Manquante par défaut

Copier les anciennes valeurs

Ajouter

Remplacer


Supprimer

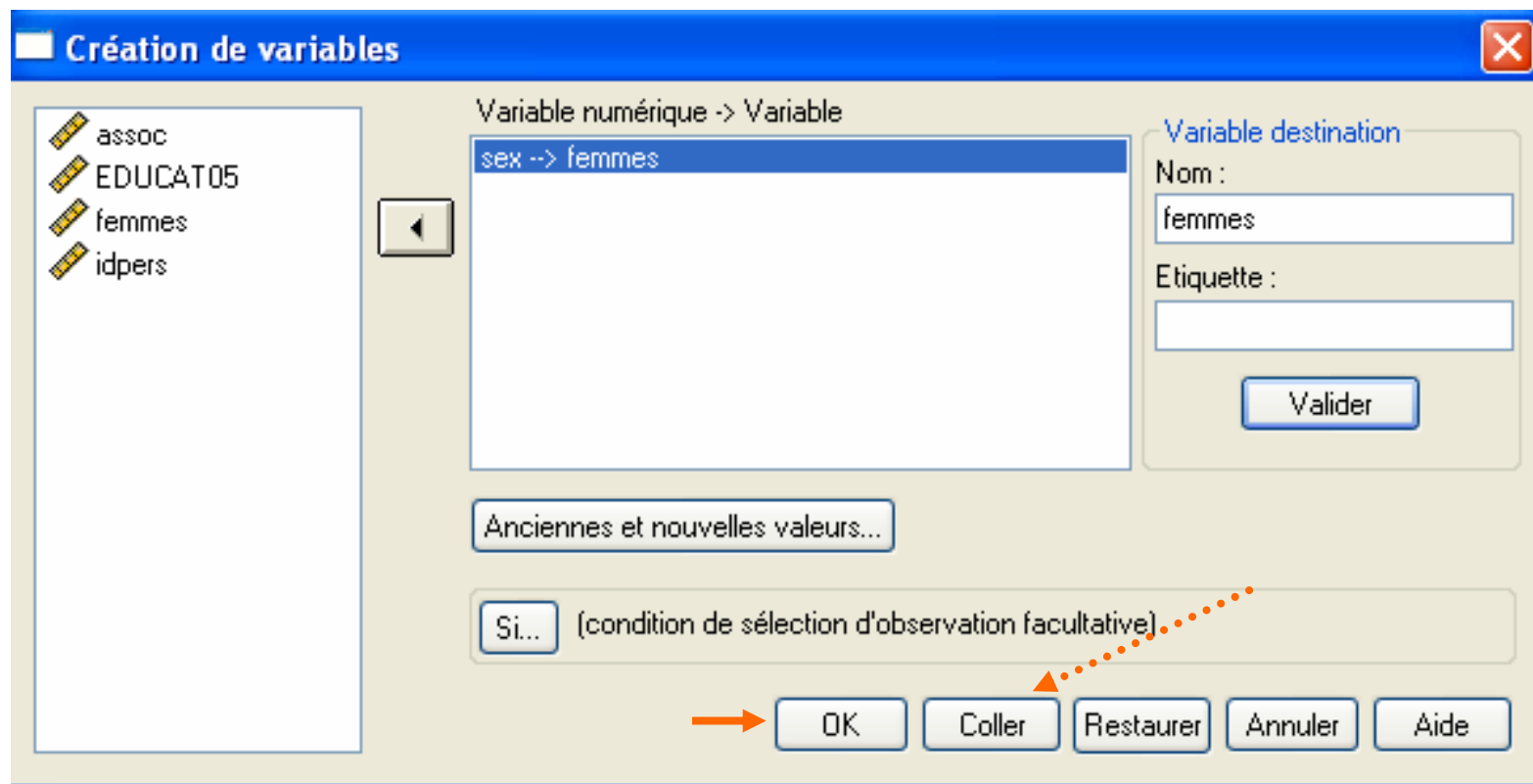
Ancienne --> Nouvelle :

```
1 --> 0
2 --> 1
```

Variables destination sont des chaînes Longueur :

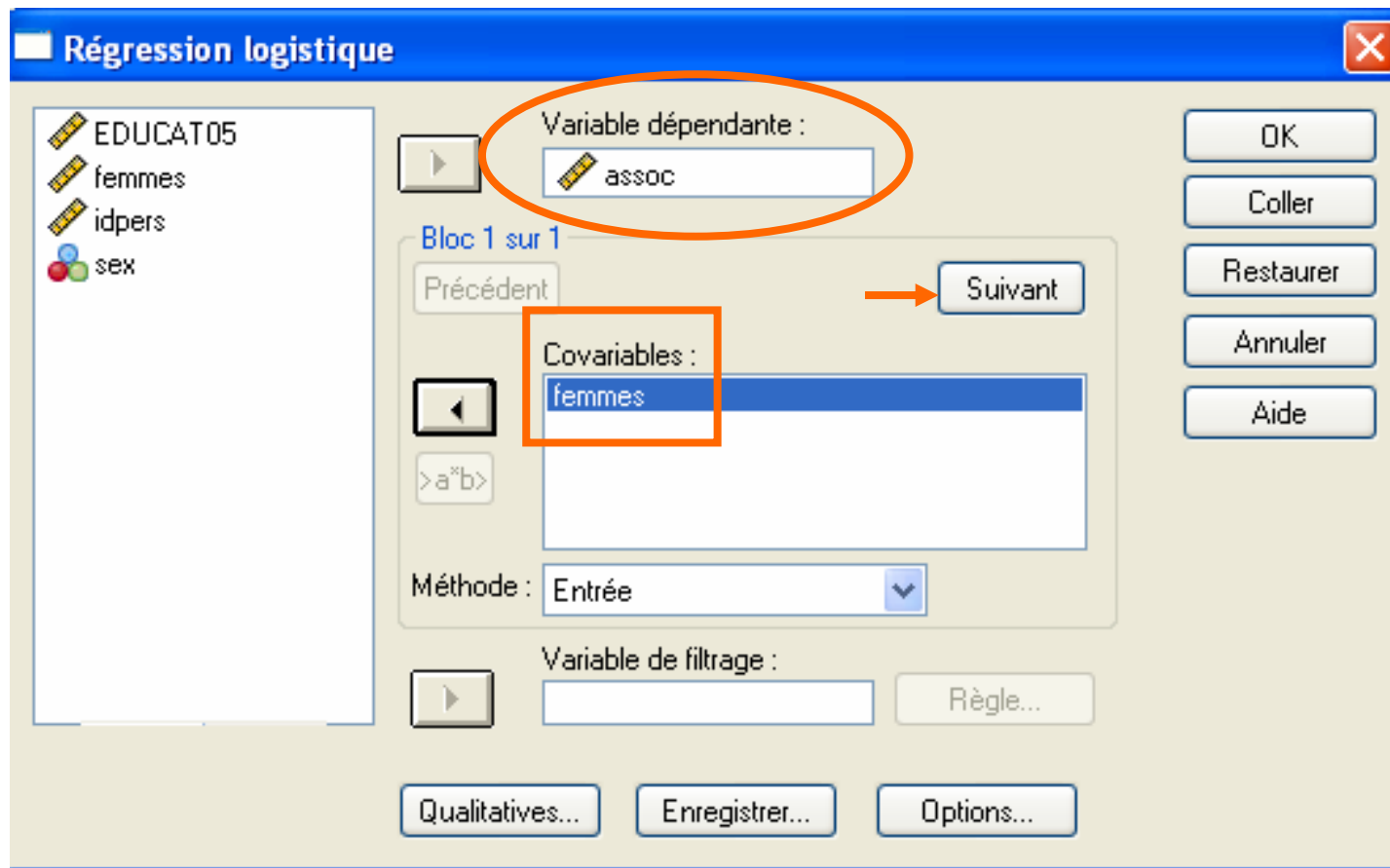
Convertir les chaînes numériques en nombres ('5'>5)

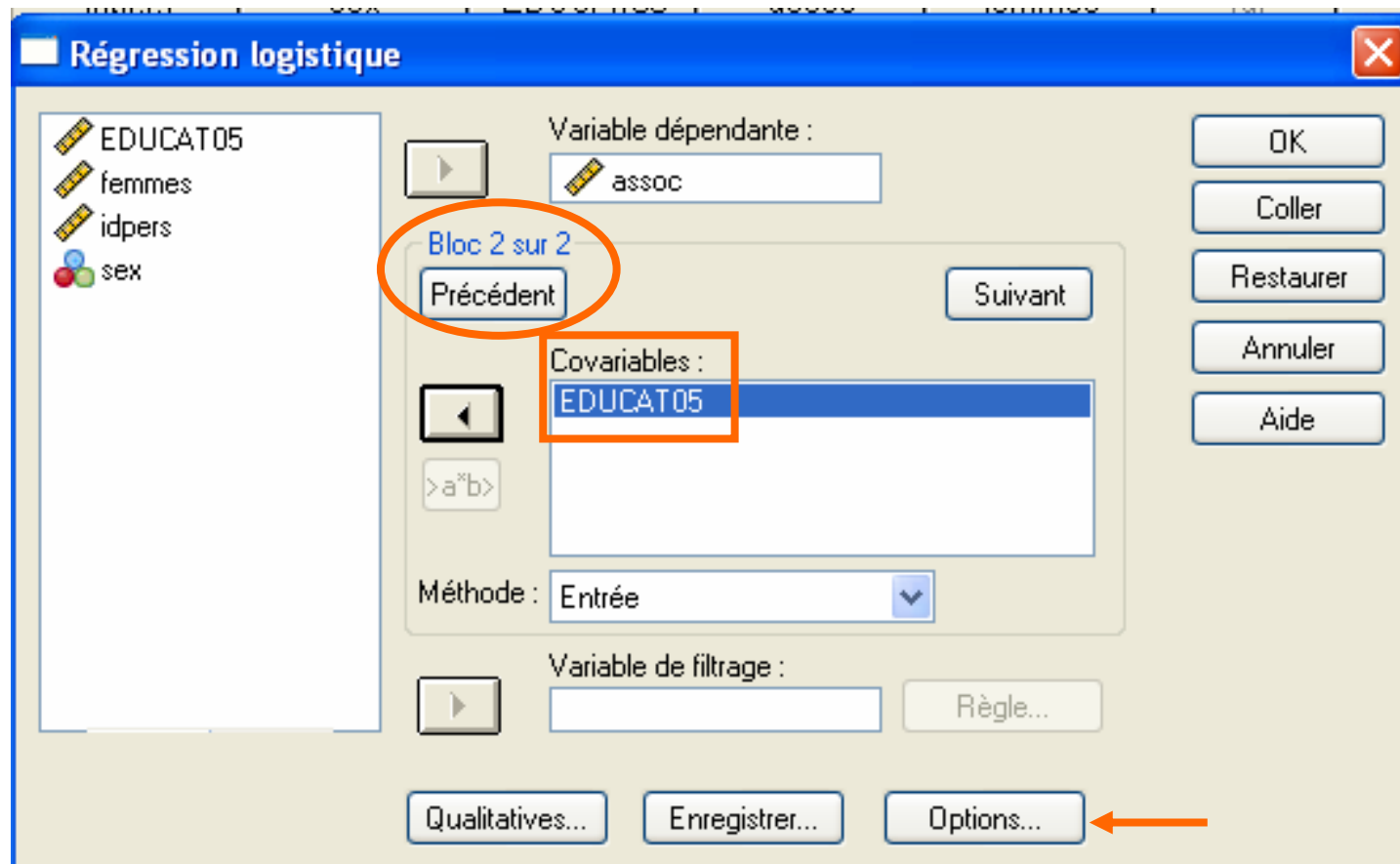




2) Procédure via la syntaxe

```
recode sex (1=0) (2=1) into femmes.  
exe.
```



Régression logistique : Options ✕

Diagrammes et statistiques

<input type="checkbox"/> Diagrammes de classement	<input type="checkbox"/> Corrélations des estimations
<input type="checkbox"/> Qualité d'ajustement de Hosmer-Lemeshow	<input type="checkbox"/> Historique des itérations
<input type="checkbox"/> Liste des résidus	<input checked="" type="checkbox"/> CI pour exp(B): <input type="text" value="95"/> %

Points atypiques à écarts type

Toutes les observations

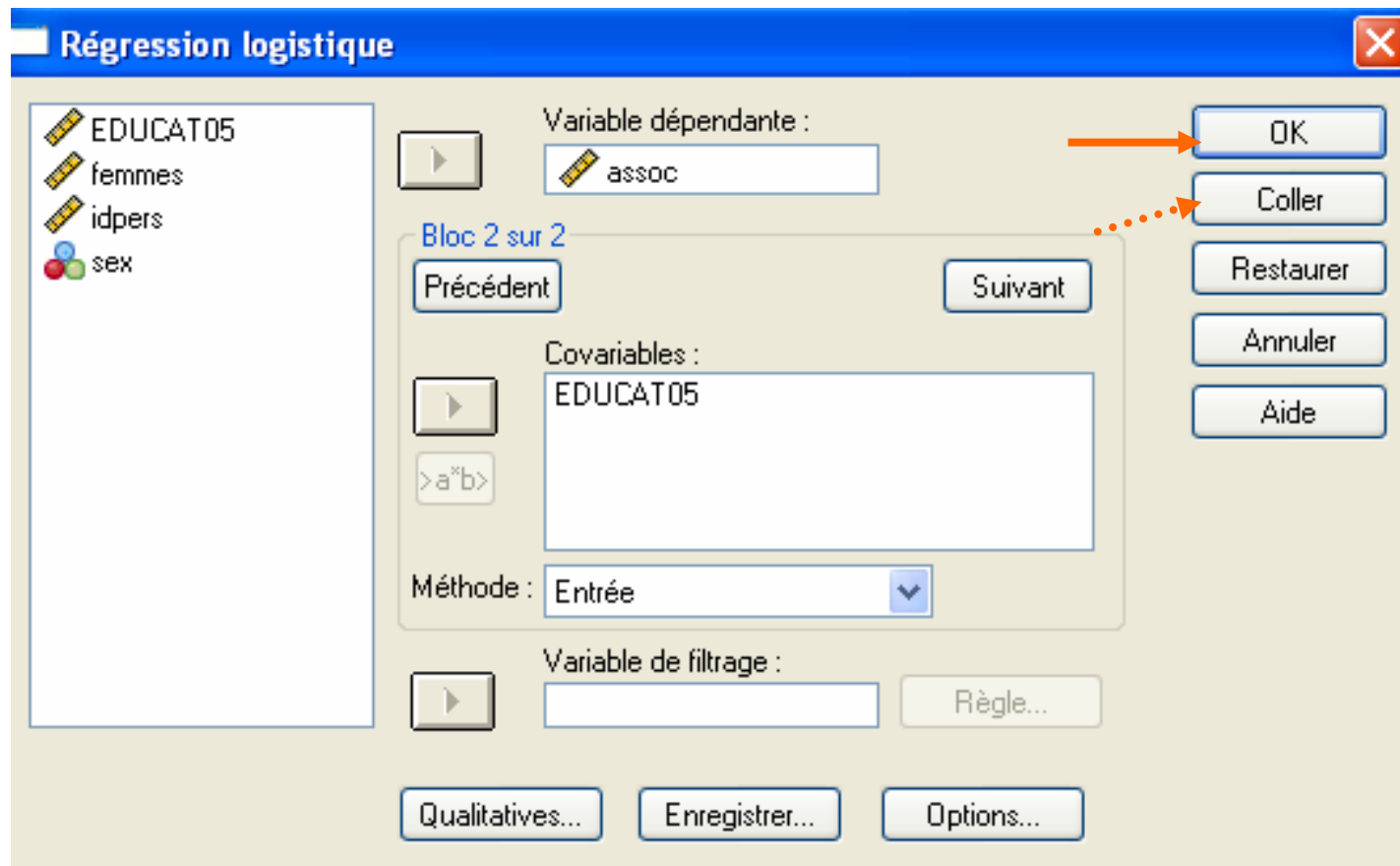
Afficher

A chaque pas Au dernier pas

Probabilité dans pas à pas

Entrée : <input type="text" value=".05"/>	Élimination : <input type="text" value=".10"/>	Césure du classement : <input type="text" value=".5"/>
		Maximum des itérations : <input type="text" value="20"/>

Inclure terme constant dans le modèle



2) Procédure via la syntaxe

```
logistic regression assoc  
  /method=enter age05  
  /method=enter femmes  
  /print=CI(95).
```

Etape 1
Sexe

Récapitulatif du modèle

Etape	-2log-vraisemblance	R-deux de Cox & Snell	R-deux de Nagelkerke
1	8851.217 ^a	.007	.010

a. L'estimation a été interrompue au numéro d'itération 3 parce que les estimations de paramètres ont changé de moins de .001.

Différence d'avec le modèle vide

Tests de spécification du modèle

		Khi-deux	ddl	Signif.
Etape 1	Etape	46.702	1	.000
	Bloc	46.702	1	.000
	Modèle	46.702	1	.000

Interprétation

- La valeur dite du « moins 2 log vraisemblance » (en anglais -2 log likelihood, parfois abrégé -2LL) peut être conçue comme une mesure de l'écart entre les valeurs de Y calculées à partir du modèle et les valeurs de Y telles qu'observées
- Ce qui importe n'est pas la valeur en soi de cette mesure mais la différence de valeurs entre deux modèles

- Pour savoir si cette différence est significative, on procède à un test du χ^2 , avec un degré de liberté égale à la différence du nombre de paramètres entre les deux modèles ainsi comparés

- Ici, le modèle de l'étape 1 avec le sexe comme unique variable indépendante est comparé au modèle vide (=sans variable indépendante)
- La valeur du « moins 2 log vraisemblance » pour le modèle de l'étape 1 est de 8851.217
- La différence de cette valeur avec celle calculée pour le modèle vide est de 46.702

- La significativité de cette différence fait l'objet d'un test du khi2, avec un degré de liberté correspondant à l'unique paramètre supplémentaire (la variable indépendante sexe ou plus précisément son coefficient) que comprend le modèle de l'étape 1 par rapport au modèle vide
- Le résultat du test indique une différence significative ($p < .001$)

- On en conclut que le modèle comprenant la variable indépendante sexe permet de mieux se rapprocher des valeurs de Y observées que le modèle vide
- Traduction: le fait d'être membre ou non d'une association varie en fonction du sexe de l'individu

Variables dans l'équation

		B	E.S.	Wald	ddl	Signif.	Exp(B)	IC pour Exp(B) 95.0%	
								Inférieur	Supérieur
Étape 1 ^a	femmes	-.344	.051	46.479	1	.000	.709	.642	.782
	Constante	.334	.038	78.106	1	.000	1.397		

a. Variable(s) entrées à l'étape 1 : femmes.

Interprétation

- Les femmes ont un odds d'être membre d'une association 0.71 fois plus petit que le odds des hommes (=catégorie de référence), ce dernier odds ayant par définition une valeur de 1
- De manière équivalente, on peut dire que les femmes ont une probabilité relative ou une propension 0.71 fois inférieure à celle des hommes d'être membre d'une association

- On peut également exprimer cet écart en %: dans ce cas on dira que la propension des femmes d'être membre d'une association est $(0.71-1) \times 100 = 29\%$ plus faible que celle des hommes
- Cet écart est très significatif: $p < .001$; l'intervalle de confiance pour l'exponentielle du coefficient du sexe ne comprend pas la valeur 1
- On peut donc conclure que l'adhésion aux associations est plus faible chez les femmes

Etape 1
Age

Récapitulatif du modèle

Etape	-2log-vrais emblance	R-deux de Cox & Snell	R-deux de Nagelkerke
1	8851.217 ^a	.007	.010

a. L'estimation a été interrompue au numéro d'itération 3 parce que les estimations de paramètres ont changé de moins de .001.

Etape 2
Sexe+Education

Récapitulatif du modèle

Etape	-2log-vrais emblance	R-deux de Cox & Snell	R-deux de Nagelkerke
1	8835.121 ^a	.010	.013

a. L'estimation a été interrompue au numéro d'itération 3 parce que les estimations de paramètres ont changé de moins de .001.

Différence
Etape 2 – Etape 1
(8851.217-8835.121=16.095)

Tests de spécification du modèle

		Khi-deux	ddl	Signif.
Etape 1	Etape	16.095	1	.000
	Bloc	16.095	1	.000
	Modèle	62.798	2	.000

Interprétation

- La valeur du « moins 2 log vraisemblance » pour le modèle de l'étape 2 (avec éducation) est de 8835.121
- Pour rappel, cette valeur pour le modèle de l'étape 1 (sans éducation) était de 8851.217
- La différence entre ces deux valeurs est donc de 16.095

- Pour déterminer si cette différence est significative, il est procédé à un test du khi2, avec 1 degré de liberté correspondant à l'unique variable indépendante introduite entre les 2 étapes, qui est l'éducation (ou plus précisément son paramètre, estimé par le coefficient de régression)
- Le résultat du test est positif ($p < .001$)

- On en conclut que le modèle comprenant les variables sexe et éducation permet de mieux se rapprocher des valeurs de Y observées que le modèle comprenant uniquement le sexe comme variable indépendante
- Traduction: le fait d'être membre ou non d'une association varie (également) en fonction du niveau d'éducation de l'individu

Variables dans l'équation

		B	E.S.	Wald	ddl	Signif.	Exp(B)	IC pour Exp(B) 95.0%	
								Inférieur	Supérieur
Étape 1 ^a	femmes	-.312	.051	37.009	1	.000	.732	.662	.810
	EDUCAT05	.035	.009	16.053	1	.000	1.035	1.018	1.053
	Constante	.140	.061	5.183	1	.023	1.150		

a. Variable(s) entrées à l'étape 1 : EDUCAT05.

Interprétation

- Pour chaque niveau d'éducation supplémentaire, le odds d'être membre d'une association est 1.04 fois plus grand (ou augmente de 1.04)
- De manière équivalente, on peut dire que la probabilité relative (ou propension) d'être membre une association est 1.04 fois plus forte quand le niveau d'éducation croît d'une unité

- On peut également exprimer cet écart en %: pour chaque niveau d'éducation additionnel, le odds (ou probabilité relative ou propension) d'être membre d'une association s'accroît de $(1.04-1) \times 100 = 4\%$
- Cet écart est significatif: $p < .001$; l'intervalle de confiance pour l'exponentielle du coefficient de l'éducation ne comprend pas le 1
- On peut donc conclure que l'adhésion aux associations augmente avec l'éducation

Nota Bene

- Dans le cas d'une variable indépendante de type quantitatif (comme ici l'éducation), il peut être intéressant de ne pas (seulement) considérer les écarts de la variable dépendante en fonction de la variation d'une unité de la variable indépendante (ce que l'on a fait ici)
- On peut ainsi considérer les écarts de Y en fonction d'une variation de 5, 10, etc. unités de la variable indépendante

- Pour cela, il faut multiplier la valeur du coefficient de régression par le nombre d'unité que l'on souhaite considérer, puis élever le résultat à l'exponentielle
- Ainsi par exemple, l'odds (ou probabilité relative ou propension) d'être membre d'une association est $\exp[.035 \times 5] = 1.19$ fois plus fort lorsque le niveau d'éducation augmente de 5 unités
- Ou encore, cette fois exprimé en %, est $(1.19 - 1) \times 100 = 19\%$ plus élevé