

## LE MODÈLE LINÉAIRE GÉNÉRAL ET LA RÉGRESSION MULTIPLE APPLIQUÉS À L'ANALYSE DE VARIANCE

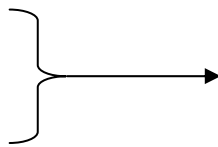
### Données

Source : Enquête de Statistique Canada sur les dépenses des familles en 1992  
 (1900 observations pour le Québec)

Variables :

1. Composition du ménage

- Personnes seules
- Couples sans enfants
- Couples avec enfants
- Familles monoparentales
- Autres ménages



Enfants : d'âge quelconque,  
 jamais mariés et  
 qui vivent avec leurs parents

2. Nombre d'enfants de moins de 16 ans (i.e. de 15 ans ou moins) : 0 ; 1 ; 2 ; 3 ; 4 ou plus
3. Âge de la personne de référence
4. Revenu du ménage après impôt
5. Variation nette de l'actif et du passif
6. Sécurité

#### VARIABLE DÉPENDANTE

- *EPARGNE* = Variation nette de l'actif et du passif + Sécurité

#### VARIABLES INDÉPENDANTES

- *REVAPIMP* = Revenu du ménage après impôt
- Âge
- Composition du ménage

#### STRUCTURE DE L'ÉCHANTILLON (nombre d'observations)

Composition du ménage	Âge de la personne de référence				Total
	Moins de 35	35-45	45-65	65 et plus	
Personnes seules	101	77	132	142	452
Couples sans enf.	95	49	177	133	454
Couples avec enf.	163	267	239	26	695
Familles mono.	30	71	48	10	159
Autres sans enf.	24	17	41	21	103
Autres avec enf.	8	16	12	1	37
<b>Total</b>	<b>421</b>	<b>497</b>	<b>649</b>	<b>333</b>	<b>1900</b>

Noter : Plan d'échantillonnage ni « équilibré », ni bi-proportionnel.

## VARIABLES INDÉPENDANTES D'ÂGE

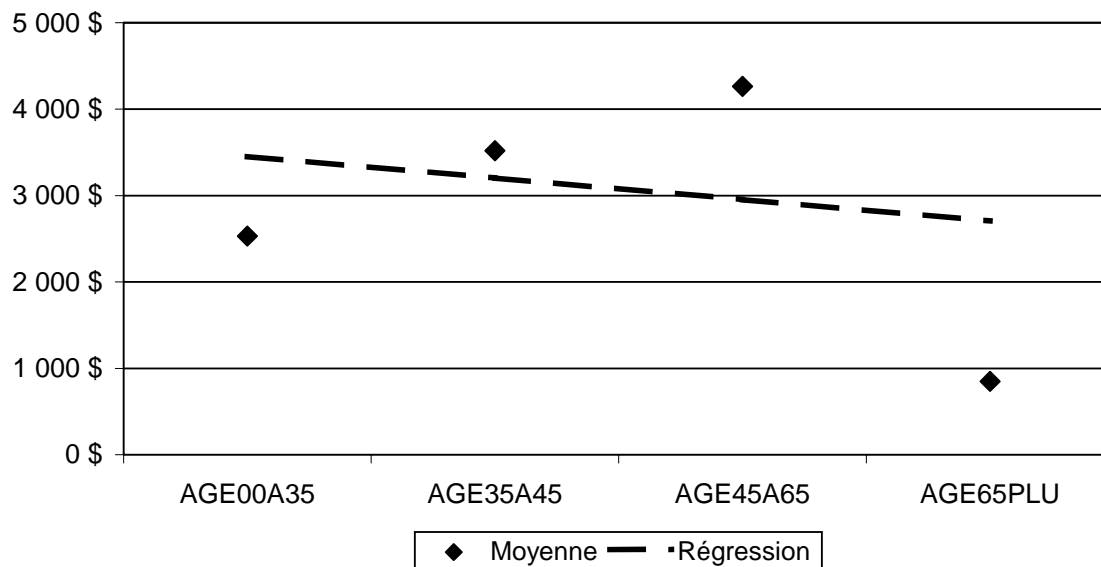
*GROUPE* = Groupe d'âge de la personne de référence

Âge de la personne de référence	Valeur de la variable <i>GROUPE</i>
moins de 35 ans	1
35 ans ou plus et moins de 45 ans	2
45 ans ou plus et moins de 65 ans	3
65 ans et plus	4

La variable *GROUPE* est une **variable ordinale d'ordre faible** (voir 1-1)

**Pourquoi faut-il remplacer *GROUPE* par des variables dichotomiques ?**

### VALEUR MOYENNE DE L'ÉPARGNE PAR GROUPE D'ÂGE ET RÉGRESSION LINÉAIRE sur *GROUPE*



Réponse : **Pour ne pas imposer une relation linéaire contraire aux faits.**

*GROUPE* est donc remplacée par des variables dichotomiques :

1.  $AGE00A35 = 1$  si  $GROUPE = 1$  (âge < 35) ; = 0 autrement
2.  $AGE35A45 = 1$  si  $GROUPE = 2$  (âge  $\geq 35$  et < 45) ; = 0 autrement
3.  $AGE45A65 = 1$  si  $GROUPE = 3$  (âge  $\geq 45$  et < 65) ; = 0 autrement
4.  $AGE65PLU = 1$  si  $GROUPE = 4$  (âge  $\geq 65$ ) ; = 0 autrement

## VARIABLES INDÉPENDANTES DE COMPOSITION DU MÉNAGE

- Composition du ménage : variable catégorique polytomique  
**À remplacer par des variables dichotomiques**  
(pour les mêmes raisons que *GROUPE*)
- Nombre d'enfants de moins de 16 ans

### Deux approches

#### A. 5 VARIABLES DICHOTOMIQUES (VOIR SCHÉMA)

#### B. 3 VARIABLES DICHOTOMIQUES :

1. *SEULMONO* = 1 s'il s'agit d'une personne seule ou d'une famille monoparentale ;  
*SEULMONO* = 0 autrement
2. *AUTRE* = 1 si le ménage appartient à la catégorie «Autres» ; *AUTRE* = 0 autrement
3. *ENFANTS* = 1 si le ménage compte au moins un enfant ; *ENFANTS* = 0 autrement

Enfants = d'âge quelconque, jamais mariés et qui vivent avec leurs parents,  
SAUF pour les ménages «Autres» : enfants = de moins de 16 ans

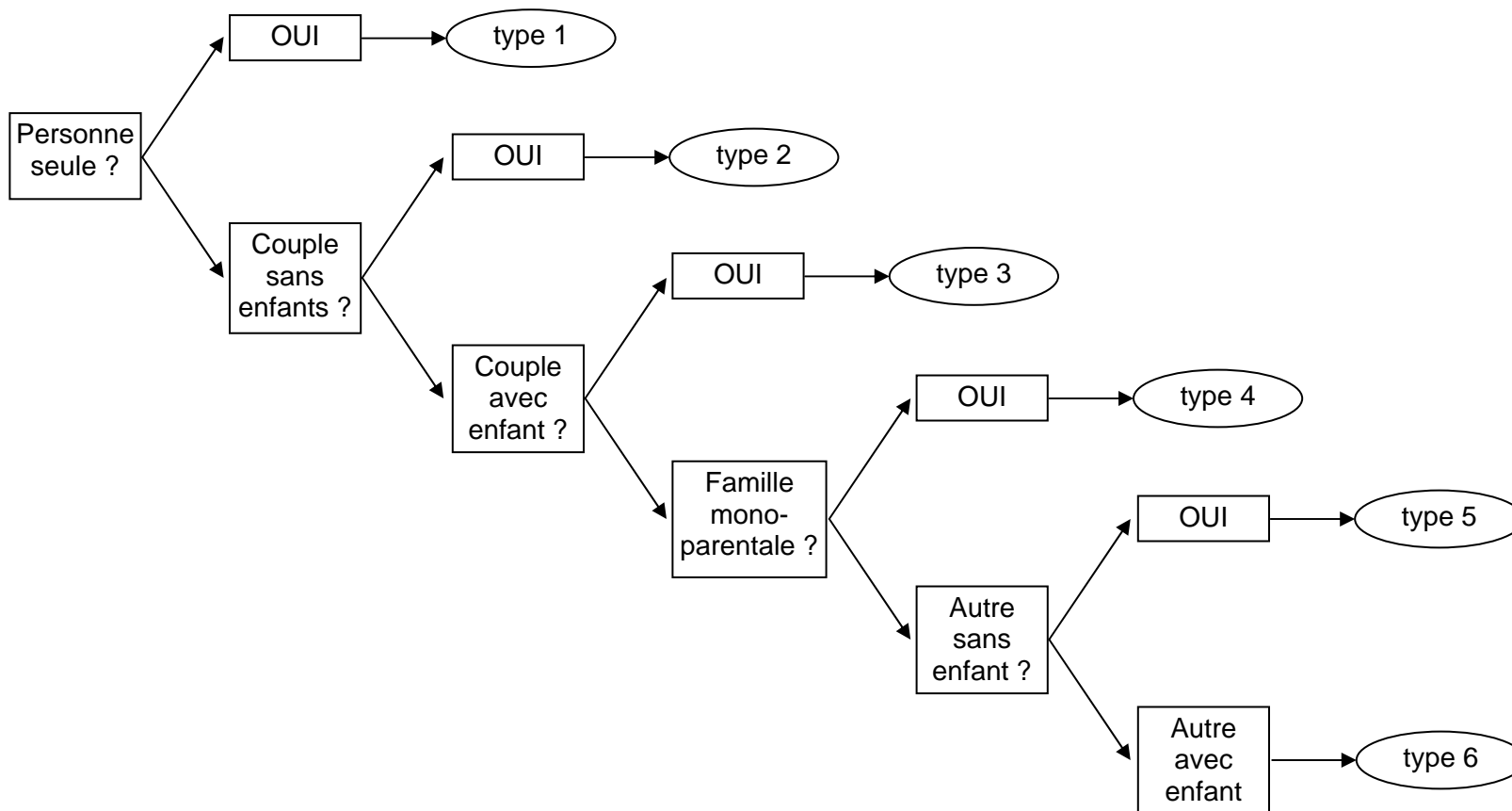
En combinant ces trois variables dichotomiques, on obtient la classification suivante :

Composition du ménage	Nombre d'enfants	Valeur de la variable <i>SEULMONO</i>	Valeur de la variable <i>AUTRE</i>	Valeur de la variable <i>ENFANTS</i>
Personnes seules	0	1	0	0
Couples sans enf.	0	0	0	0
Couples avec enf.	> 0	0	0	1
Familles mono.	> 0	1	0	1
Autres ménages	0	0	1	0
	> 0	0	1	1

**Cette manière de procéder IMPOSE une certaine cohérence au modèle**

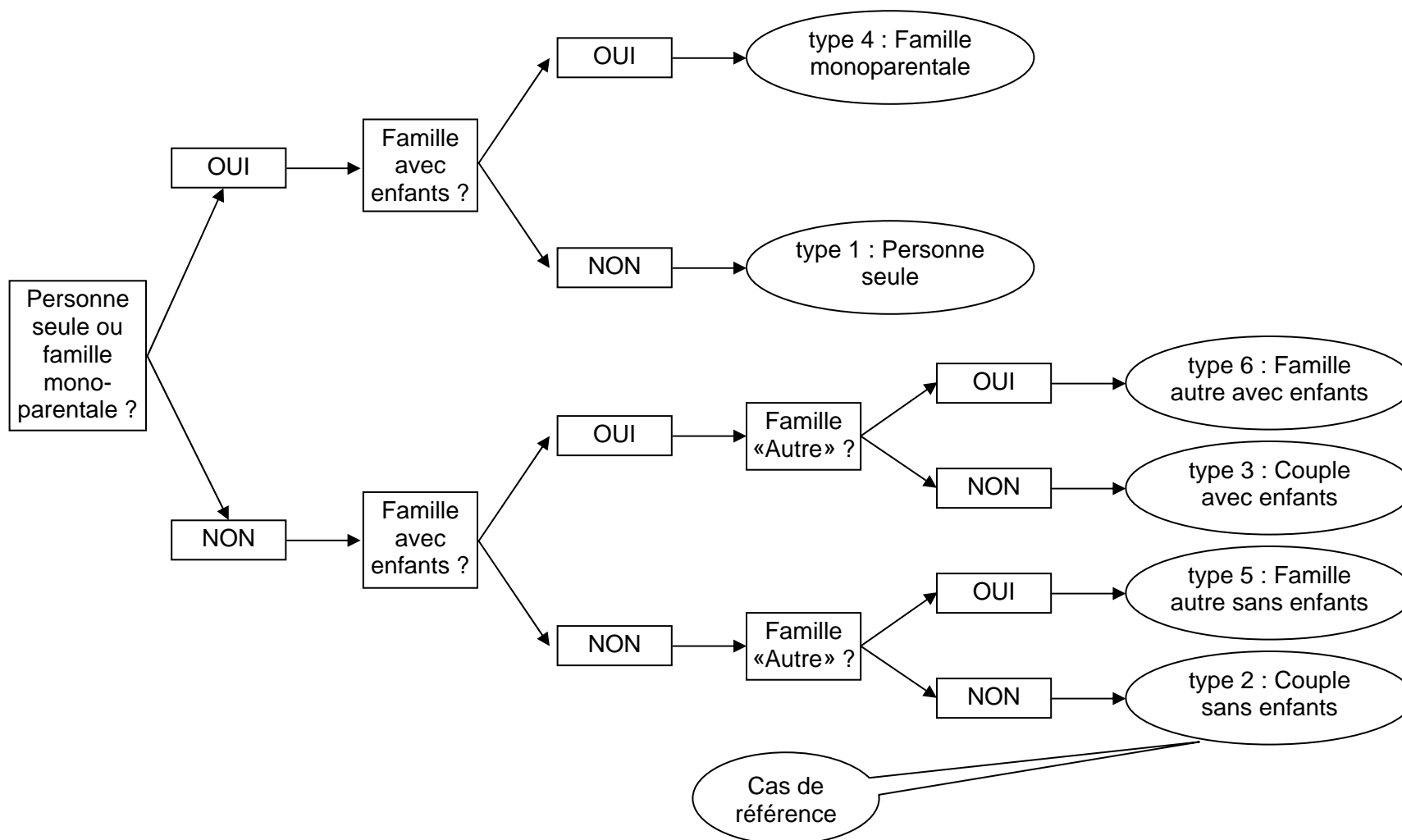
## DEUX ARBRES DE CLASSIFICATION À L'AIDE DE VARIABLES DICHOTOMIQUES (1)

### I – Classification à 5 variables



## DEUX ARBRES DE CLASSIFICATION À L'AIDE DE VARIABLES DICHOTOMIQUES (2)

### II – Classification à 3 variables



## REDONDANCE PARMIS LES VARIABLES INDÉPENDANTES

### 1. Variable dichotomique

Exemple : présence ou absence d'enfants

- $AVECENFANTS = 1$  si le ménage compte au moins un enfant ;  
 $AVECENFANTS = 0$  autrement
- $SANSENFANTS = 0$  si le ménage compte au moins un enfant ;  
 $SANSENFANTS = 1$  autrement

L'une de ces variables est-elle de trop ? **OUI !**

**Pourquoi ?**

- **Logiquement :**
  - si  $AVECENFANTS = 1$ , alors  $SANSENFANTS = 0$  (et vice-versa)
  - si  $AVECENFANTS = 0$ , alors  $SANSENFANTS = 1$  (et vice-versa)
- **Par rapport à l'hypothèse H4 de Gauss-Markov :**
  - $AVECENFANTS + SANSENFANTS = 1$  (constante du modèle)
  - Il y a multicollinéarité parfaite !

### 2. Variable polytomique

Exemple : *GROUPAGE*

$AGE00A35 = 1$  si  $GROUPAGE = 1$  (âge < 35) ; = 0 autrement

$AGE35A45 = 1$  si  $GROUPAGE = 2$  (âge ≥ 35 et < 45) ; = 0 autrement

$AGE45A65 = 1$  si  $GROUPAGE = 3$  (âge ≥ 45 et < 65) ; = 0 autrement

$AGE65PLU = 1$  si  $GROUPAGE = 4$  (âge ≥ 65) ; = 0 autrement

- **Logiquement :**
  - si on connaît la valeur de 3 des 4, on peut déduire la valeur de la 4<sup>ème</sup>
- **Par rapport à l'hypothèse H4 de Gauss-Markov :**
  - $AGE00A35 + AGE35A45 + AGE45A65 + AGE65PLU = 1 = \text{CONSTANTE}$
  - Il y a multicollinéarité parfaite !

Il y a une variable de trop.

#### Cas de référence

Après élimination de la redondance, le **cas de référence** est celui qui correspond aux variables dichotomiques supprimées.

## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE

### RÉSULTATS – MODÈLE SANS EFFETS D'INTERACTION

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_3 SEULMONO + \beta_4 AUTRE + \beta_5 ENFANTS + \beta_6 AGE00A35 + \beta_7 AGE45A65 + \beta_8 AGE65PLU$$

#### Cas de référence

- Pas une personne seule ou monoparentale :  $SEULMONO = 0$
- Pas un ménage « Autre » :  $AUTRE = 0$
- Pas d'enfants :  $ENFANTS = 0$
- Catégorie d'âge 35-45 :  $AGE00A35 = 0$ ,  $AGE45A65 = 0$  et  $AGE65PLU = 0$

C'est-à-dire : un **couple sans enfants** dont la personne de référence a **35-45** ans.

Variable	Description	Symbole	Coefficient estimé	t de Student	Probabilité critique
<i>CONSTANTE</i>		$\beta_1$	-7727	-11,062	0,0001
<i>REVAPIMP</i>	Revenu du ménage après impôt	$\beta_2$	0,340	28,468	0,0001
<i>ENFANTS</i>	Présence d'enfants	$\beta_5$	-2260	-4,937	0,0001
<i>SEULMONO</i>	Pers.seule ou monop.	$\beta_3$	1903	3,834	0,0001
<i>AUTRE</i>	Ménage «Autre»	$\beta_4$	-2578	-3,309	0,0010
<i>AGE00A35</i>	Âge 00-35	$\beta_6$	258	0,444	0,6574
<i>AGE45A65</i>	Âge 45-65	$\beta_7$	419	0,796	0,4263
<i>AGE65PLU</i>	Âge 65+	$\beta_8$	875	1,322	0,1862

$n = 1900$

$R^2 = 0,33$

Interprétation du modèle sans interaction

Composition du ménage	Nombre d'enfants	Groupe d'âge de la personne de référence	SEULMON	AUTRE	ENFANTS	AGE00A35	AGE45A65	AGE65PLU	EPARGNE prédite par le modèle
			0						
Personnes seules	0	< 35	1	0	0	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + 0 + \beta_6 + 0 + 0$
		≥ 35 et < 45	1	0	0	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + 0 + 0 + 0 + 0$
		≥ 45 et < 65	1	0	0	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + 0 + 0 + \beta_7 + 0$
		≥ 65	1	0	0	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + 0 + 0 + 0 + \beta_8$
Couples sans enf.	0	< 35	0	0	0	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + 0 + \beta_6 + 0 + 0$
		≥ 35 et < 45	0	0	0	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + 0 + 0 + 0 + 0$
		≥ 45 et < 65	0	0	0	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + 0 + 0 + \beta_7 + 0$
		≥ 65	0	0	0	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + 0 + 0 + 0 + \beta_8$
Couples avec enf.	> 0	< 35	0	0	1	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + \beta_5 + \beta_6 + 0 + 0$
		≥ 35 et < 45	0	0	1	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + \beta_5 + 0 + 0 + 0$
		≥ 45 et < 65	0	0	1	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + \beta_5 + 0 + \beta_7 + 0$
		≥ 65	0	0	1	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + 0 + \beta_5 + 0 + 0 + \beta_8$
Familles mono.	> 0	< 35	1	0	1	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + \beta_5 + \beta_6 + 0 + 0$
		≥ 35 et < 45	1	0	1	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + \beta_5 + 0 + 0 + 0$
		≥ 45 et < 65	1	0	1	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + \beta_5 + 0 + \beta_7 + 0$
		≥ 65	1	0	1	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 + 0 + \beta_5 + 0 + 0 + \beta_8$
Autres ménages	0	< 35	0	1	0	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + 0 + \beta_6 + 0 + 0$
		≥ 35 et < 45	0	1	0	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + 0 + 0 + 0 + 0$
		≥ 45 et < 65	0	1	0	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + 0 + 0 + \beta_7 + 0$
		≥ 65	0	1	0	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + 0 + 0 + 0 + \beta_8$
	> 0	< 35	0	1	1	1	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + \beta_5 + \beta_6 + 0 + 0$
		≥ 35 et < 45	0	1	1	0	0	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + \beta_5 + 0 + 0 + 0$
		≥ 45 et < 65	0	1	1	0	1	0	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + \beta_5 + 0 + \beta_7 + 0$
		≥ 65	0	1	1	0	0	1	$\beta_1 + \beta_2 \text{ REVAPIMP} + 0 + \beta_4 + \beta_5 + 0 + 0 + \beta_8$

- Pas d'enfants : ENFANTS = 0



## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE INTERPRÉTATION DES RÉSULTATS

$$\begin{aligned} EPARGNE &= \beta_1 + \beta_2 REVAPIMP + \beta_3 SEULMONO + \beta_4 AUTRE \\ &\quad + \beta_5 ENFANTS + \beta_6 AGE00A35 + \beta_7 AGE45A65 + \beta_8 AGE65PLU \\ EPARGNE &= -7727 + 0,34 REVAPIMP + 1903 SEULMONO + (-2578) AUTRE \\ &\quad + (-2260) ENFANTS + 258 AGE00A35 + 419 AGE45A65 + 875 AGE65PLU \end{aligned}$$

### Exemples

$$REVAPIMP = 20\ 000 \$$$

- Couple sans enfants, 35-45 (**cas de référence**)

$$EPARGNE = \beta_1 + \beta_2 \times 20\ 000 = -7727 + 0,340 \times 20\ 000 = -927$$

- Couple avec enfants, < 35 ans

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_5 ENFANTS + \beta_6 AGE00A35$$

$$EPARGNE = -7727 + 0,34 \times 20\ 000 + (-2260) ENFANTS + 258 AGE00A35 = -2929$$

$$EPARGNE = -927 + (-2260) + 258 = -2929$$

- Personne seule, sans enfants, > 65 ans

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_3 SEULMONO + \beta_8 AGE65PLU$$

$$EPARGNE = -7727 + 0,34 \times 20\ 000 + 1903 SEULMONO + 875 AGE65PLU = 1851$$

$$EPARGNE = -927 + 1903 + 875 = 1851$$

**Les coefficients des variables dichotomiques représentent des différences  
par rapport à ce cas de référence**

## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE VARIABLES D'INTERACTION

### INTERACTIONS ENTRE VARIABLES DICHOTOMIQUES

11.  $MONOMONO = ENFANTS \times SEULMONO$
12.  $AUTRENFA = ENFANTS \times AUTRE$
13.  $ENFA0035 = ENFANTS \times AGE00A35$
14.  $ENFA4565 = ENFANTS \times AGE45A65$
15.  $ENFA65PL = ENFANTS \times AGE65PLU$
16.  $AUTA0035 = AUTRE \times AGE00A35$
17.  $AUTA4565 = AUTRE \times AGE45A65$
18.  $AUTA65PL = AUTRE \times AGE65PLU$
19.  $SOLA0035 = SEULMONO \times AGE00A35$
20.  $SOLA4565 = SEULMONO \times AGE45A65$
21.  $SOLA65PL = SEULMONO \times AGE65PLU$

### INTERACTIONS ENTRE UNE VARIABLE CONTINUE ET DES VARIABLES DICHOTOMIQUES

22.  $REVENFAN = REVAPIMP \times ENFANTS$
23.  $REVSELMO = REVAPIMP \times SEULMONO$
24.  $REVAUTRE = REVAPIMP \times AUTRE$

## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE

### RÉSULTATS – MODÈLE AVEC EFFETS D'INTERACTION

#### Modèle

$$\begin{aligned}
 EPARGNE = & \beta_1 + \beta_2 \text{ REVAPIMP} + \beta_3 \text{ SEULMONO} + \beta_4 \text{ AUTRE} + \beta_5 \text{ ENFANTS} \\
 & + \beta_6^* \text{ AGE00A35} + \beta_7 \text{ AGE45A65} + \beta_8 \text{ AGE65PLU} \\
 & + \gamma_1^* \text{ MONOMONO} + \gamma_2^* \text{ AUTRENFA} \\
 & + \gamma_3^* \text{ ENFA0035} + \gamma_4 \text{ ENFA4565} + \gamma_5^* \text{ ENFA65PL} \\
 & + \gamma_6^* \text{ AUTA0035} + \gamma_7^* \text{ AUTA4565} + \gamma_8^* \text{ AUTA65PL} \\
 & + \gamma_9^* \text{ SOLA0035} + \gamma_{10} \text{ SOLA4565} + \gamma_{11} \text{ SOLA65PL} \\
 & + \alpha_1^* \text{ REVENFAN} + \alpha_2 \text{ REVSELMO} + \alpha_3 \text{ REVAUTRE}
 \end{aligned}$$

L'astérisque identifie les coefficients dont la valeur estimée n'est pas significativement différente de zéro.

#### Résultats de l'estimation après élimination des variables dont les coefficients ne sont pas significativement différents de zéro

Variable	Description	Symbole	Coefficient estimé	Erreur type	Probabilité critique
CONSTANTE		$\beta_1$	-10487	729	0,0001
REVAPIMP	Revenu du ménage après impôt	$\beta_2$	0,400	0,013	0,0001
ENFANTS	Présence d'enfants	$\beta_5$	-1927	561	0,0006
SEULMONO	Pers.seule ou monoparentale	$\beta_3$	8233	1000	0,0001
AUTRE	Ménage «Autre»	$\beta_4$	7969	1771	0,0001
AGE45A65	Âge 45-65	$\beta_7$	1767	779	0,0234
AGE65PLU	Âge 65+	$\beta_8$	1513	770	0,0497
<b>Variables d'interaction</b>					
ENFA4565	ENFANTS × AGE45A65	$\gamma_4$	-1506	897	0,0932
SOLA4565	SEULMONO × AGE45A65	$\gamma_{10}$	-1983	1008	0,0494
SOLA65PL	SEULMONO × AGE65PLU	$\gamma_{11}$	-1996	1138	0,0796
REVSELMO	REVAPIMP × SEULMONO	$\alpha_2$	-0,211	0,030	0,0001
REVAUTRE	REVAPIMP × AUTRE	$\alpha_3$	-0,297	0,046	0,0001

$n = 1900$  ;  $R^2 = 0,36$

## LES COEFFICIENTS DES VARIABLES D'INTERACTION S'INTERPRÈTENT COMME DES DIFFÉRENCES

Exemple : Couple sans enfants 35-45 et couple avec enfants 45-65

$$REVAPIMP = 40\ 000 \$$$

- Couple avec enfants 35-45

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_5 ENFANTS$$

$$EPARGNE = -10487 + 0,4 \times 40\ 000 + (-1927) ENFANTS = 3586$$

- Couple sans enfants 45-65

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_7 AGE45A65$$

$$EPARGNE = -10487 + 0,4 \times 40\ 000 + 1767 AGE45A65 = 7280$$

- Couple avec enfants 45-65

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_5 ENFANTS + \beta_7 AGE45A65 + \gamma_4 ENFA4565$$

$$EPARGNE = -10487 + 0,4 \times 40\ 000 + (-1927) ENFANTS + 1767 AGE45A65 + (-1506)$$

$$ENFA4565 = 3847$$

$$\text{Différence p/r couple av. enf. 35-45} = 3847 - 3586 = 261 = \beta_7 + \gamma_4$$

$$\text{Différence p/r couple sans enf. 45-65} = 3847 - 7280 = -3433 = \beta_5 + \gamma_4$$

## INTERACTION ENTRE UNE VARIABLE CATÉGORIQUE ET UNE VARIABLE CONTINUE

Exemple : Couple sans enfants 35-45 et personne seule 35-45

$REVAPIMP = 20\ 000\ \$$  et  $REVAPIMP = 40\ 000\ \$$

- Couple sans enfants 35-45 (**cas de référence**)

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP$$

$$\begin{aligned} EPARGNE = -10487 + 0,4 REVAPIMP &= -2487 \text{ si } REVAPIMP = 20\ 000\ \$ \\ &= 5513 \text{ si } REVAPIMP = 40\ 000\ \$ \end{aligned}$$

- Personne seule 35-45

$$EPARGNE = \beta_1 + \beta_2 REVAPIMP + \beta_3 SEULMONO + \alpha_2 REVSELMO$$

$$EPARGNE = \beta_1 + (\beta_2 + \alpha_2) REVAPIMP + \beta_3 SEULMONO$$

$$EPARGNE = -10487 + 0,4 REVAPIMP + 8233 SEULMONO + (-0.211) REVSELMO$$

$$EPARGNE = -10487 + (0,4 - 0.211) REVAPIMP + 8233 SEULMONO$$

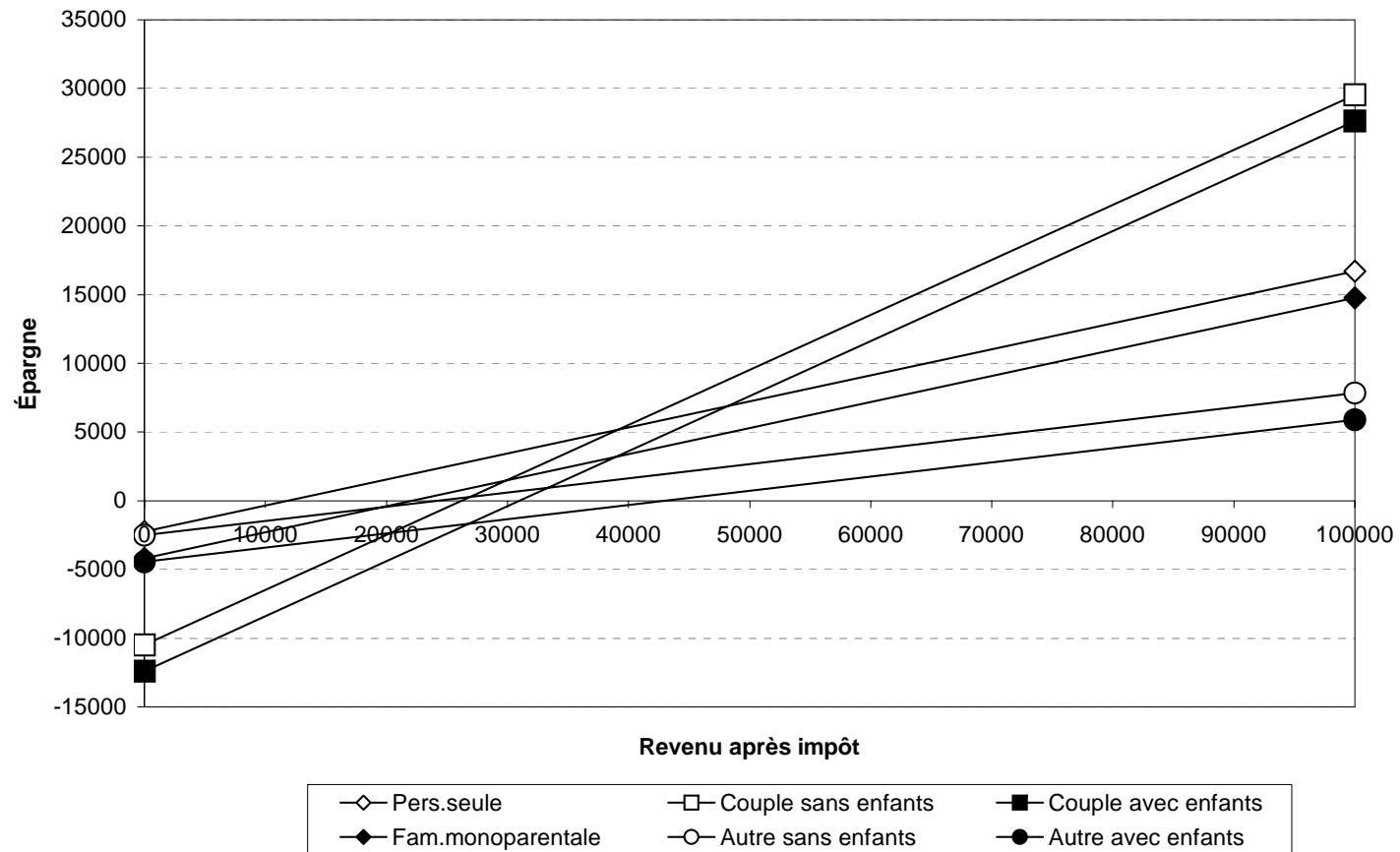
$$\begin{aligned} EPARGNE = -10487 + 0,189 REVAPIMP + 8233 SEULMONO \\ &= 1526 \text{ si } REVAPIMP = 20\ 000\ \$ \\ &= 5306 \text{ si } REVAPIMP = 40\ 000\ \$ \end{aligned}$$

$$\begin{aligned} \text{Différence p/r couple sans enfant} &= 4013 \text{ si } REVAPIMP = 20\ 000\ \$ \\ &= -207 \text{ si } REVAPIMP = 40\ 000\ \$ \\ &= 8233 SEULMONO + (-0.211) REVSELMO \end{aligned}$$

La différence varie avec le revenu  $REVAPIMP$ .

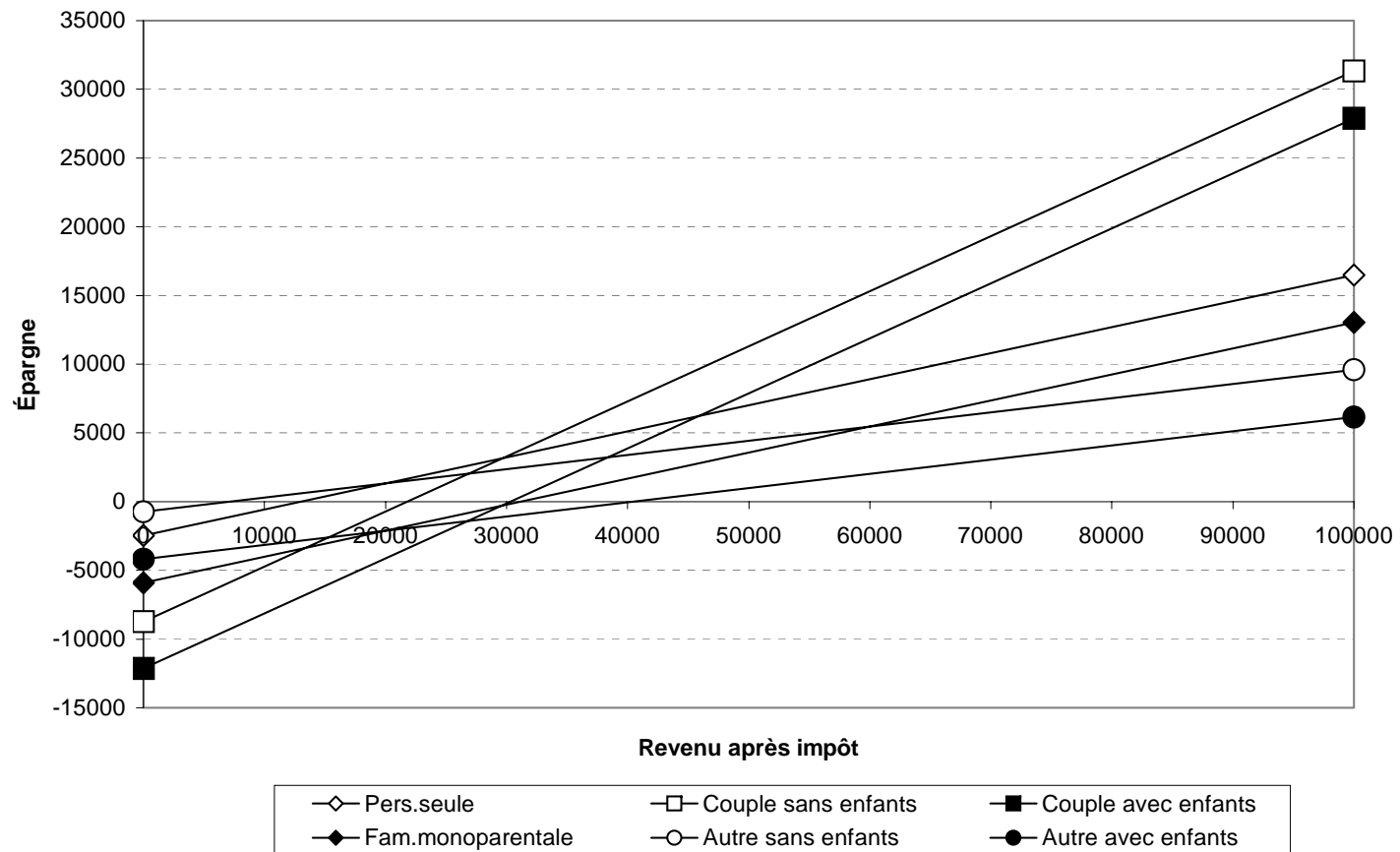
## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE RÉSULTATS – MODÈLE AVEC EFFETS D'INTERACTION

Épargne des moins de 45 ans



## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE RÉSULTATS – MODÈLE AVEC EFFETS D'INTERACTION

### Épargne des 45-65 ans



## LA RÉGRESSION MULTIPLE APPLIQUÉE À L'ANALYSE DE VARIANCE RÉSULTATS – MODÈLE AVEC EFFETS D'INTERACTION

### Épargne des 65 ans et plus

