

Exemple 1

Couleur de la peau et capital social

Margaret L. Hunter. If You're Light You're Alright: Light Skin Color as Social Capital for Women of Color *Gender and Society*, 16, No. 2, (Apr., 2002), pp. 175-193.

L'hypothèse de travail dans cette étude est que pour des femmes américaines d'origine mexicaines et pour les femmes afro-américaines, la clarté de la peau est une forme de « capital social » qui peut servir à accroître des avantages relatifs sur le plan de l'éducation, du revenu et de l'état civil du mari. Trois variables (variables endogènes) servent à définir les avantages sociaux.

Variables endogènes

- Le niveau de scolarité atteint
- Le revenu
- Le niveau de scolarité du mari

La variable d'intérêt principal est la couleur de la peau et son effet sur les trois variables endo. L'auteure en a créé une variable quantitative sur une échelle allant de 1 (peau très foncée) à 5 (peau très claire). Elle a donc effectué une analyse de régression multiple avec pour variable endogène chacune des variables énumérées ci-dessus.

La variable exogène d'intérêt est la couleur de la peau. Mais les variables endogènes d'autres facteurs—dont la scolarité, le revenu et la scolarité du mari, entre autres—qui doivent être contrôlés. Chacun des modèles de régression multiple fait donc intervenir les variables exogènes suivantes.

Variables exogènes

- x_1 : Clarté de la peau (une mesure de la clarté de la peau sur une échelle de 1 à 5, 1 = moins claire, 5 = la plus claire)
- x_2 : Âge, en nombre d'années
- x_3 : État civil (=1 si mariée, 0 sinon)
- x_4 : Scolarité de la mère en nombre d'années, si connue
- x_5 : Scolarité de la mère inconnue (=1 si inconnue, 0 sinon)
- x_6 : Scolarité du père, en nombre d'années, si connue
- x_7 : Scolarité du père inconnue (=1 si inconnue, 0 sinon)
- x_8 : Milieu de vie (=1 si urbain, 0 sinon)
- x_9 : Pays de naissance (= 1 si États-Unis, 0 sinon)
- x_{10} : Anglais (une mesure quantitative de la connaissance de l'anglais)

Question 1 Prédiction de la scolarité chez les américaines d'origine mexicaine

Modèle de régression pour prédire le niveau de scolarité des femmes américaines d'origine mexicaine.
($n = 551$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	vp
x_1 : Peau	$\hat{\beta}_1$	0,281	0,119	0,0186
x_2 : Âge	$\hat{\beta}_2$	-0,078	0,010	0,0000
x_3 : État civil	$\hat{\beta}_3$	0,094	0,300	0,7541
x_4 : Scolarité de la mère	$\hat{\beta}_4$	0,248	0,048	0,0000
x_5 : Scolarité de la mère inconnue	$\hat{\beta}_5$	-0,797	0,456	0,0811
x_6 : Scolarité du père	$\hat{\beta}_6$	0,179	0,045	0,0001
x_7 : Scolarité du père inconnue	$\hat{\beta}_7$	-0,297	0,403	0,4615
x_8 : Milieu de vie	$\hat{\beta}_8$	0,372	0,357	0,2979
x_9 : Pays de naissance	$\hat{\beta}_9$	2,597	0,289	0,0000
x_{10} : Anglais	$\hat{\beta}_{10}$	0,403	0,123	0,0011

$R^2 = 0,47$, $vp = 0$.

- Vérifier les valeurs p présentées dans le tableau.
- Interpréter concrètement chacun des faits suivants :
 - $\hat{\beta}_1 < 0$
 - $\hat{\beta}_5 < 0$
 - $\hat{\beta}_9 < 0$.
- Tester l'hypothèse $H_{04} : \beta_4 = 0$. Exprimer votre conclusion en termes concrets.
- Tester l'hypothèse $H_{05} : \beta_3 = 0$. Compte tenu du résultat, y a-t-il une raison d'inclure l'état civil comme variable exogène dans la régression?
- L'auteur inclut dans sa régression un terme fixe (β_5) désignant les répondants ne connaissant pas le niveau de scolarité de la mère. Est-ce une bonne idée?
 - Supposons qu'une des répondantes la mère a 10 ans de scolarité répond qu'elle ne possède pas cette information. Montrer que cela entraîne une sous-estimation de plus de 3 années de la scolarité (par rapport à l'estimation qu'on aurait faite si on savait que la mère avait 10 ans de scolarité).
 - Montrer que le niveau de scolarité aurait été sous-estimé même si la mère n'avait qu'un an de scolarité.
 - Tester à 5 % l'hypothèse que $\beta_5 = 0$.

[Tout compte fait, ce terme (β_5) ne rend pas service. Mais l'exclure carrément de la régression serait problématique : quelle valeur donner à x_5 lorsqu'elle n'est pas connue? Ce problème de données manquantes est difficile à résoudre. Si le nombre de données manquantes est faible, la solution la plus facile consiste à ignorer tous les cas qui en contiennent. Sinon, il existe des techniques assez complexes qui consistent à remplacer les données manquantes par une valeur estimée qu'on appelle alors une *imputation*].

Question 2 *Prédiction de la scolarité chez les afro-américaines*Coefficients de régression pour prédire le niveau de scolarité des femmes afro-américaines ($n = 1265$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	pv
x_1 : Peau	$\hat{\beta}_1$	0,327	0,084	0,00010
x_2 : Âge	$\hat{\beta}_2$	-0,063	0,005	0,00000
x_3 : État civil	$\hat{\beta}_3$	0,596	0,160	0,00020
x_4 : Scolarité de la mère	$\hat{\beta}_4$	0,200	0,030	0,00000
x_5 : Scolarité de la mère inconnue	$\hat{\beta}_5$	-1,527	0,20	0,00000
x_6 : Scolarité du père	$\hat{\beta}_6$	0,048	0,029	0,09814
x_7 : Scolarité du père inconnue	$\hat{\beta}_7$	-0,359	0,178	0,04392
x_8 : Milieu de vie	$\hat{\beta}_8$	1,270	0,188	0,00000

$$R^2 = 0,37; vp = 0,00000000372$$

Question 3 *Prédiction du revenu chez les femmes américaines d'origine mexicaines*Coefficients de régression pour prédire le revenu des femmes américaines d'origine mexicaines ($n = 302$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	pv
x_1 : Peau	$\hat{\beta}_1$	57,66	156,4	0,7126
x_2 : Âge	$\hat{\beta}_2$	26,4	16,4	0,1085
x_3 : État civil	$\hat{\beta}_3$	180,3	381,3	0,6364
x_4 : Scolarité de la mère	$\hat{\beta}_4$	44,6	61,7	0,4704
x_5 : Scolarité de la mère inconnue	$\hat{\beta}_5$	-743,3	691,4	0,2800
x_6 : Scolarité du père	$\hat{\beta}_6$	-23,4	56,8	0,6807
x_7 : Scolarité du père inconnue	$\hat{\beta}_7$	-156,9	589,5	0,7903
x_8 : Milieu de vie	$\hat{\beta}_8$	119,0	469,4	0,8000
x_9 : Scolarité de la répondante	$\hat{\beta}_9$	246,2	59,0	0,0000
x_{10} : Pays de naissance	$\hat{\beta}_{10}$	-461,4	432,9	0,2874
x_{11} : Anglais	$\hat{\beta}_{11}$	20,6	165,2	0,9008

$$R^2 = 0,10; vp = 0,0011$$

Question 4 *Prédiction de la scolarité chez les femmes afro-américaines*Coefficients de régression pour prédire le revenu des femmes afro-américaines ($n = 1134$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	p
x_1 : Peau	$\hat{\beta}_1$	673,2	295,7	0,0235
x_2 : Âge	$\hat{\beta}_2$	84,8	17,9	0,0000
x_3 : État civil	$\hat{\beta}_3$	396,4	551,0	0,4720
x_4 : Scolarité de la mère	$\hat{\beta}_4$	29,45	105,0	0,7792
x_5 : Scolarité de la mère inconnue	$\hat{\beta}_5$	-207,2	715,8	0,7723
x_6 : Scolarité du père	$\hat{\beta}_6$	155,8	99,4	0,1173
x_7 : Scolarité du père inconnue	$\hat{\beta}_7$	-1131,0	615,1	0,0662
x_8 : Milieu de vie	$\hat{\beta}_8$	2580,5	654,6	0,0001
x_9 : Scolarité de la répondante	$\hat{\beta}_9$	1183,3	99,0	0,0000

 $n = 1134$; $R^2 = 0,19$; $vp = 0$.**Question 5** *Prédiction de la scolarité du mari chez les américaines d'origine mexicaine*Coefficients de régression pour prédire la scolarité du mari des femmes américaines d'origine mexicaines ($n = 366$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	pv
x_1 : Peau	$\hat{\beta}_1$	-0,052	0,170	0,7599
x_2 : Scolarité de la répondante	$\hat{\beta}_2$	0,623***	0,059	0,0000
x_3 : Scolarité de la mère	$\hat{\beta}_3$	-0,094	0,056	0,0941
x_4 : Scolarité de la mère inconnue	$\hat{\beta}_4$	-1,250*	0,585	0,0333
x_5 : Pays de naissance	$\hat{\beta}_5$	-0,165	0,431	0,7021
x_6 : Anglais	$\hat{\beta}_1$	0,437**	0,172	0,0115

 $R^2 = 0,40$; $vp = 0,00096$

Question 6 *Prédiction de la scolarité du mari chez les femmes afro-américaines*Coefficients de régression pour prédire la scolarité du mari des femmes afro-américaines ($n = 419$)

Variable exogène (x_j)	Coefficient	Valeur estimée $\hat{\beta}_j$	$\hat{\sigma}_{\hat{\beta}_j}$	pv
x_1 : Peau	$\hat{\beta}_1$	0,287	0,148	0,6532
x_2 : Scolarité de la répondante	$\hat{\beta}_2$	0,581	0,047	0,0000
x_3 : Scolarité de la mère	$\hat{\beta}_3$	0,142	0,047	0,0027
x_4 : Scolarité de la mère inconnue	$\hat{\beta}_4$	-0,766	0,300	0,0110

$$R^2 = 0,41; \nu p = 0.$$

Question 7 *Relation entre la couleur de la peau et l'état civil***7-1** *Relation entre la couleur de la peau et l'état civil parmi les américaines d'origine mexicaine*

Le tableau suivant présente la distribution conjointe du État civil et la couleur de la peau des femmes américaines d'origine mexicaine.

Distribution conjointe de la couleur de la peau et de l'état civil de 366 femmes américaines d'origine mexicaine.

	Très foncée	Foncée	Moyenne	Claire	Très claire	Total
Mariée	14	63	88	58	30	253
Célibataire	9	31	33	21	19	113
Total	23	94	121	79	49	366

- Déterminer les distributions conditionnelles de l'état civil étant donné la couleur de la peau, et interpréter.
- Déterminer les distributions conditionnelles de la couleur de la peau étant donné l'état civil, et interpréter.
- Déterminer si la relation observée est significative.

7-2 *Relation entre la couleur de la peau et l'état civil parmi les femmes afro-américaines*

Distribution conjointe de la couleur de la peau et de l'état civil de 366 femmes afro-américaines.

	Très foncée	Foncée	Moyenne	Claire	Très claire	Total
Mariée	8	36	71	23	4	142
Célibataire	21	81	122	44	9	277
Total	29	117	193	67	13	419

- Déterminer les distributions conditionnelles de l'état civil étant donné la couleur de la peau, et interpréter.
- Déterminer les distributions conditionnelles de la couleur de la peau étant donné l'état civil, et interpréter.
- Déterminer si la relation observée est significative.

Question 8 *Relation entre la clarté de la peau et l'état civil*

Nous reprenons ici la question soulevée à la question 7.1 (la relation entre la couleur de la peau et l'état civil) mais en traitant la couleur de la peau comme une variable quantitative (prenant des valeurs entières de 1 à 5)

a) *Américaines d'origine mexicaines*

Tester à l'aide de la loi de Student que la moyenne de la clarté de la peau est la même parmi les femmes mariées et les célibataires

b) *Afro-américaines*

Tester à l'aide de la loi de Student que la moyenne de la clarté de la peau est la même parmi les femmes mariées et les célibataires