

Corrigé de l'exercice 3

Jean-François Bickel

Statistique II – SP08

Syntaxe

```
REGRESSION
```

```
  /MISSING LISTWISE
```

```
  /STATISTICS COEFF OUTS CI R ANOVA
```

```
  /CRITERIA=PIN(.05) POUT(.10)
```

```
  /NOORIGIN
```

```
  /DEPENDENT I05WY
```

```
  /METHOD=ENTER AGE05 EDUCAT05 .
```

Résumé des modèles

<i>Modèles</i>	<i>R</i>	<i>R</i> ²	<i>R</i> ² _{ajusté}	<i>Erreur standard</i>
1) Age	.285	.081	.081	55147.3
2) Age + Formation	.450	.202	.202	51393.3

Anova des modèles

<i>Modèles</i>		<i>Sommes carrés</i>	<i>Ddl</i>	<i>F</i>	<i>Sign</i>
1	Régres.	1.13E+12	1	370.5	.000
	Résidu	1.27E+13	4184		
	Total	1.39E+13	4185		
2	Régres.	2.80E+12	2	530.6	.000
	Résidu	1.10E+13	4183		
	Total	1.39E+13	4185		

Coefficients des modèles

	<i>B</i>	<i>ES</i>	<i>Beta</i>	<i>T</i>	<i>Sign</i>	<i>CI</i>	
Const	8057	2836		2.8	.005	2496	13618
Age	1285	67	.29	19.3	.000	1154	1416
Const	-14839	2795		-5.3	.000	-20319	-9359
Age	890	64	.20	13.9	.000	764	1016
Educ	7110	282	.36	25.2	.000	6557	7663

- L'introduction de la variable du niveau d'éducation améliore de manière notable la qualité (fit) du modèle qui passe d'un R^2 de .08 à un R^2 de .20
- Une proportion substantiellement plus grande de la variabilité du revenu est expliquée lorsqu'on tient compte du niveau d'éducation (=une part importante de la variabilité de y dépend de l'éducation)

- Le test de F (tableau « ANOVA ») indique que l'hypothèse $\beta(\text{âge}) = \beta(\text{éducation}) = 0$ est extrêmement peu probable (la probabilité est inférieure à 0.1%)
- On peut donc conclure qu'au moins un des coefficients est différent de 0
- Et donc qu'au moins une partie de la variabilité de y est fonction des valeurs prises par une (ou les deux) variable(s) indépendante(s)

- Par rapport au premier modèle (exercice 1) dans lequel on ne considèrerait que la relation entre âge et revenu, ici cette relation est contrôlée par l'éducation
- Autrement dit, on considère la relation entre âge et revenu pour une valeur constante (ou fixe) d'éducation
- Ceci revient à mesurer la relation âge-revenu en omettant la part qui, dans cette relation, résulte du fait (potentiel) que l'éducation est associée (corrélée) à l'âge et au revenu

- On observe que le coefficient de l'âge est plus faible lorsqu'on contrôle par l'éducation ($B=890$) que sans contrôle ($B=1285$)
- Contrôler par l'éducation implique donc une réduction de la valeur du coefficient de 31% ; une part non négligeable de la relation entre âge et revenu est liée à l'association que l'éducation entretient avec ces deux variables

- Dans le même temps, il existe une relation positive significative ($B=890$; $t=13.9$; $p<.001$) entre âge et revenu, indépendamment de l'éducation
- Net du facteur éducation, chaque année d'âge supplémentaire équivaut un revenu moyen additionnel de 890 Frs, avec un intervalle de confiance situé entre 764 et 1016 Frs

- L'éducation est ici contrôlée par l'âge (au sens qu'on a dit)
- Indépendamment du facteur âge, chaque degré d'éducation supplémentaire « rapporte » un revenu moyen supplémentaire de 7110 Frs; l'augmentation « réelle » se situant, avec une probabilité de 95%, entre 6557 et 7663 Frs;
- Cet effet est très significatif ($t=25.2$, $p<.001$)

- L'examen des coefficients standardisés (Beta) montre que l'effet de la formation est comparativement plus élevé que celui de l'âge

- Il est aussi possible de calculer le revenu (prédit) pour une combinaison de valeurs des variables indépendantes
- Par exemple, le revenu moyen des personnes de 40 ans avec un diplôme universitaire est de:
$$-14839 + (890 \times 40) + (7110 \times 10) = 91861 \text{ Frs}$$