

Exemple 6.1 : Illustration du problème de biais d'agrégation sur les données concernant le jugement

resm= jugement que portent les enseignants sur la valeur scolaire de leurs élèves en mathématiques

mresm= jugement moyen de l'enseignant de la classe

scom= scores obtenus par les élèves aux épreuves nationales de CE2 en mathématiques (performances scolaires effectives de début d'année)

mscom= score moyen de la classe aux évaluations nationales de début de CE2 en mathématiques

Modèle « total » : régression individuelle sur tous les individus sans tenir compte de l'appartenance aux différentes classes :

Ce modèle ne prend pas en compte l'environnement des individus. Il traite l'ensemble des individus comme un échantillon prélevé d'une population homogène, sans que leur appartenance à un quelconque environnement soit considérée.

Syntaxe SAS:

```
proc reg data=work.jugement;  
model resm=scom /stb;  
run;
```

L'option /stb indique que vous souhaitez spécifier les coefficients de régression standardisés.

23

24 Sorties SAS :

25	Procédure REG
26	Modèle : MODEL1
27	Variable dépendante : RESM
28	
29	Nombre d'observations lues 342
30	Nombre d'observations utilisées 342
31	
32	
33	Analyse de variance
34	
35	
36	Source
37	DDL
38	Somme des carrés
39	Moyenne quadratique
40	Valeur F
41	Pr > F
42	Modèle 1 619.67209 619.67209 267.58 <.0001
43	Erreur 340 787.38054 2.31583
44	Total sommes corrigées 341 1407.05263
45	
46	Root MSE 1.52178 R carré 0.4404
47	Moyenne dépendante 6.61404 R car. ajust. 0.4388
48	Coeff Var 23.00840
49	
50	Valeurs estimées des paramètres
51	
52	Variable DDL Valeur estimée des paramètres Erreur type Valeur du test t Pr > t Valeur estimée normalisée
53	Intercept 1 1.82980 0.30383 6.02 <.0001 0
54	SCOM 1 0.11614 0.00710 16.36 <.0001 0.66363

55 *Ligne 53* : La colonne « valeur estimée des paramètres » indique une estimation du coefficient
56 β_0

57 *Ligne 54* : La colonne « valeur estimée des paramètres » indique une estimation du coefficient
58 β_1

59 *Ligne 53* : La colonne « valeur estimée normalisée » indique une estimation du coefficient β_0
60 standardisé

61 *Ligne 54*: La colonne « valeur estimée normalisée » indique une estimation du coefficient
62 β_1 standardisé

63 Modèle agrégé: régression sur les données agrégées:

64 Ce modèle permet d'estimer l'influence de l'environnement des individus (école, classe,
65 entreprise...) Dans ce cas de figure, il n'y a qu'une seule observation par macro-unité et, par
66 conséquent, aucune variation des données individuelles.

67

68 Syntaxe SAS:

```
69
70 proc sort data=work.jugement;by classe;
71 proc means;by classe;
72 var resm scom;
73 output out=sortie
74 mean=mresm mscom;
75 run;
76 proc reg data=sortie;
77 model mresm=mscom/stb;
78 run;
```

cette syntaxe permet de créer un fichier temporaire composé uniquement de données agrégées, ici, le jugement moyen de l'enseignant par classe (mresm) ainsi que le score moyen pour chaque classe (mscom)

79 Sorties SAS :

80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109

Procédure REG
Modèle : MODEL1
Variable dépendante : mresm

Nombre d'observations lues17
Nombre d'observations utilisées17

Analyse de variance

Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	1	2.13330	2.13330	4.16	0.0593
Erreur	15	7.68416	0.51228		
Total sommes corrigées	16	9.81746			

Root MSE0.71574R carré0.2173

Moyenne dépendante6.61625R car. ajust.0.1651

Coeff Var10.81784

Valeurs estimées des paramètres

Variable	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr > t	Valeur estimée normalisée
Intercept	1	4.35518	1.12152	3.88	0.0015	0
mscom	1	0.05587	0.02738	2.04	0.0593	0.46615

110 *Ligne 85 :* Indique le nombre de classes sur lesquelles porte l'analyse. Cela confirme que
111 l'analyse porte sur des macro-unités

112 *Ligne 108 :* La colonne « valeur estimée des paramètres » indique une estimation du
113 coefficient β_0

114 *Ligne 109 :* La colonne « valeur estimée des paramètres » indique une estimation du
115 coefficient β_1

116 *Ligne 108 :* La colonne « valeur estimée normalisée » indique une estimation du coefficient
117 β_0 standardisé

118 *Ligne 109 :* La colonne « valeur estimée normalisée » indique une estimation du coefficient
119 β_1 standardisé

Le modèle contextuel : régression sur les données individuelles, avec introduction d'une variable agrégée :

Ce modèle consiste à intégrer simultanément la variable individuelle et la variable correspondante agrégée au niveau du groupe.

Syntaxe SAS:

```
proc reg data=work.jugement;
model resm=scom mscom;
run;
```

Sorties SAS :

Procédure REG						
Modèle : MODEL1						
Variable dépendante : RESM						
Nombre d'observations lues		342				
Nombre d'observations utilisées		342				
Analyse de variance						
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F	
Modèle	2	642.49361	321.24680	142.44	<.0001	
Erreur	339	764.55903	2.25534			
Total sommes corrigées	341	1407.05263				
Root MSE		1.50178	R carré	0.4566		
Moyenne dépendante		6.61404	R car. ajust.	0.4534		
Coeff Var		22.70592				
Valeurs estimées des paramètres						
Variable	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr > t	Valeur estimée normalisée
Intercept	1	3.56513	0.62249	5.73	<.0001	0
SCOM	1	0.12784	0.00791	16.16	<.0001	0.73046
mscom	1	-0.05368	0.01688	-3.18	0.0016	-0.14383

Ligne 134 : Indique le nombre d'élèves sur lequel porte l'analyse. Cela confirme que l'analyse porte sur des données individuelles

Ligne 156 : La colonne « valeur estimée des paramètres » indique une estimation du coefficient β_0

Ligne 157: La colonne « valeur estimée des paramètres » indique une estimation du coefficient β_1

Ligne 158: La colonne « valeur estimée des paramètres » indique une estimation du coefficient β_2

Ligne 156 : La colonne « valeur estimée normalisée » indique une estimation du coefficient β_0 standardisé

169 *Ligne 157* : La colonne « valeur estimée normalisée » indique une estimation du coefficient
170 β_1 standardisé

171 *Ligne 158* : La colonne « valeur estimée normalisée » indique une estimation du coefficient
172 β_2 standardisé