

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

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.*

```
GET SAS DATA='C:\Documents and Settings\Mes documents\jugement.sas7bdat'
```

```
REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT RESM /METHOD=ENTER  
SCOM.
```

Indique  
l'emplacement  
du fichier de  
données nommé  
« jugement »

### Régression

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

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

Remarques		
Entrée	Résultat obtenu	09-déc.-2010 13:44:46
	Commentaires	
	Ensemble de données actif	Ensemble_de_données1
	Filtrer	<aucune>
	Poids	<aucune>
	Scinder fichier	<aucune>
Gestion des valeurs manquantes	N de lignes dans le fichier de travail	342
	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Les statistiques sont basées sur des observations ne contenant aucune valeur manquante pour toute variable utilisée.

Syntaxe		REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT RESM /METHOD=ENTER SCOM.	Syntaxe qui indique la variable dépendante
Ressources	Temps de processeur	0:00:00.000	Syntaxe qui indique la variable indépendante
	Temps écoulé	0:00:00.032	
	Mémoire requise	2996 octets	
	Mémoire supplémentaire requise pour les diagrammes résiduels	0 octets	

Variables introduites/supprimées <sup>b</sup>				Variable indépendante introduite dans le modèle
Modèle	Variables introduites	Variables supprimées	Méthode	
1	SCOM <sup>a</sup>	.	Entrée	

a. Toutes variables requises saisies.

b. Variable dépendante : RESM

Récapitulatif des modèles					Indique les coefficients de détermination (R-deux) non ajusté et les coefficients de détermination ajusté (R-deux ajusté)
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	
1	,664 <sup>a</sup>	,440	,439	1,522	

a. Valeurs prédites : (constantes), SCOM

ANOVA <sup>b</sup>						
Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	619,672	1	619,672	267,582	,000 <sup>a</sup>
	Résidu	787,381	340	2,316		
	Total	1407,053	341			

a. Valeurs prédites : (constantes), SCOM

b. Variable dépendante : RESM

Coefficients<sup>a</sup>

Modèle	Coefficients non standardisés		Coefficients standardisés		
	A	Erreur standard	Bêta	t	Sig.
1 (Constante)	1,830	,304		6,023	,000
SCOM	,116	,007	,664	16,358	,000

a. Variable dépendante : RESM

Le coefficient 0.116 est positif et fort, il indique que le jugement de l'enseignant augmente lorsque le score des élèves est plus élevé.

## Modèle agrégé: régression sur les données agrégées:

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

```
REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT mresm /METHOD=ENTER  
mscom.
```

## Régression

mresm= jugement moyen de l'enseignant de la classe

Remarques		
Entrée	Résultat obtenu	28-oct.-2010 10:51:21
	Commentaires	
	Données	C:\Users\Desktop\donnees stats\jugement.sav
	Ensemble de données actif	Ensemble_de_données1
	Filtrer	<aucune>
	Poids	<aucune>
	Scinder fichier	<aucune>
Gestion des valeurs manquantes	N de lignes dans le fichier de travail	342
	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Les statistiques sont basées sur des observations ne contenant aucune valeur manquante pour toute variable utilisée.

Syntaxe		REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT mresm /METHOD=ENTER mscom.
Ressources	Temps de processeur	0:00:00.000
	Temps écoulé	0:00:00.010
	Mémoire requise	3060 octets
	Mémoire supplémentaire	0 octets
	requise pour les diagrammes résiduels	

[Ensemble\_de\_données1] C:\Users\Desktop\donnees stats\jugement.sav

**Variables introduites/supprimées<sup>b</sup>**

Modèle	Variables introduites	Variables supprimées	Méthode
1	mscom <sup>a</sup>	.	Entrée

Variable indépendante introduite dans le modèle

**Récapitulatif des modèles**

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,519 <sup>a</sup>	,270	,268	,68649

Indique le coefficient de détermination (R-deux) non ajusté et le coefficient de détermination ajusté (R-deux ajusté)

**ANOVA<sup>b</sup>**

Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	59,161	1	59,161	125,534	,000 <sup>a</sup>
	Résidu	160,233	340	,471		
	Total	219,393	341			

Coefficients<sup>a</sup>

Modèle	Coefficients non standardisés		Coefficients standardisés		
	A	Erreur standard	Bêta	t	Sig.
1 (Constante)	3,476	,285		12,216	,000
mscom	,077	,007	,519	11,204	,000

Indique les coefficients associés à la constante et aux variables indépendantes ainsi que la probabilité critique associée à l'impact significatif du niveau moyen de la classe sur le jugement de l'enseignant.

## 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.*

```
SPLIT FILE OFF. REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT RESM /METHOD=ENTER SCOM mscom.
```

## Régression

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

Remarques		
Entrée	Résultat obtenu	28-oct.-2010 10:51:21
	Commentaires	
	Données	C:\Users\Desktop\donnees stats\jugement.sav
	Ensemble de données actif	Ensemble_de_données1
	Filtrer	<aucune>
	Poids	<aucune>
	Scinder fichier	<aucune>
Gestion des valeurs manquantes	N de lignes dans le fichier de travail	342
	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Les statistiques sont basées sur des observations ne contenant aucune valeur manquante pour toute variable utilisée.

Syntaxe		REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT RESM /METHOD=ENTER SCOM mscom.
Ressources	Temps de processeur	0:00:00.000
	Temps écoulé	0:00:00.009
	Mémoire requise	3316 octets
	Mémoire supplémentaire requis pour les diagrammes résiduels	0 octets

[Ensemble\_de\_données1] C:\Users\Desktop\donnees stats\jugement.sav

Variables introduites/supprimées			
Modèle	Variables introduites	Variables supprimées	Méthode
1	mscom, SCOM <sup>a</sup>	.	Entrée

Variables indépendantes introduites dans le modèle

Récapitulatif des modèles				
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,676 <sup>a</sup>	,457	,453	1,502

Indique le coefficient de détermination (R-deux) non ajusté et le coefficient de détermination ajusté (R-deux ajusté)

ANOVA <sup>b</sup>					
Modèle		Somme des carrés	ddl	Moyenne des carrés	D
1	Régression	642,494	2	321,247	142,439
	Résidu	764,559	339	2,255	
	Total	1407,053	341		

Sig. ,000<sup>a</sup>



Coefficients<sup>a</sup>

Modèle	Coefficients non standardisés		Coefficients standardisés		
	A	Erreur standard	Bêta	t	Sig.
1 (Constante)	3,565	,622		5,727	,000
SCOM	,128	,008	,730	16,156	,000
mscom	-,054	,017	-,144	-3,181	,002

Indique les coefficients associés à la constante et aux variables indépendantes ainsi que la probabilité critique associée à l'impact significatif du score de début de CE2 en mathématiques et à l'impact significatif et négatif du score moyen de la classe aux évaluations nationales de début de CE2 dans cette discipline, sur le jugement de l'enseignant. L'effet de contexte indique que le jugement du maître, à score individuel donné, est plus sévère lorsque le niveau de la classe est élevé.