

Trois réponses honnêtes et une qui est seule de son côté

MCO naïfs, erreurs-types robustes aux grappes, constante aléatoire et agrégation au niveau école donnent des erreurs-types de 0,88, 1,46, 1,48 et 1,53 point. Trois d'entre elles concordent et la quatrième est celle que produit tout ajustement par défaut.

Pierre Robentz Cassion

Leçon 6 sur 8

Régression appliquée aux données de programme — Quelles covariables ont leur place

Ce que couvre cette leçon

- Le même coefficient, quatre erreurs-types
- Ce qu'ajoute une constante aléatoire
- Choisir parmi les trois
- Des prédicteurs à deux niveaux
- Trois niveaux, et quand s'arrêter
- Rapportez-le en entier
- La suite

Le même coefficient, quatre erreurs-types — En Python

```
import pandas as pd
import statsmodels.formula.api as smf

naive = smf.ols("rate ~ feeding_programme", data=per_student).fit()
robust = naive.get_robustcov_results(cov_type="cluster",
                                    groups=per_student["school_id"])

mixed = smf.mixedlm("rate ~ feeding_programme", data=per_student,
                   groups=per_student["school_id"]).fit()

by_school = (per_student.groupby(["school_id", "feeding_programme"])["rate"]
             .mean().reset_index())
aggregated = smf.ols("rate ~ feeding_programme", data=by_school).fit()
```

Le même coefficient, quatre erreurs-types — En R

```
library(lme4); library(sandwich); library(lmtest)

naive      <- lm(rate ~ feeding_programme, data = per_student)
robust     <- coeftest(naive, vcov = vcovCL, cluster = ~school_id)
mixed      <- lmer(rate ~ feeding_programme + (1 | school_id), data = per_student)
aggregated <- lm(rate ~ feeding_programme, data = by_school)
```

Le même coefficient, quatre erreurs-types

Approche	Coefficient	ET	t	n
MCO naïfs	+4,93 pts	0,88	5,63	1 200
ET robustes aux grappes	+4,93 pts	1,46	3,38	1 200
Constante aléatoire	+5,10 pts	1,48	3,44	1 200
MCO au niveau école	+5,21 pts	1,53	3,41	24

Le même coefficient, quatre erreurs-types

- Trois des quatre concordent et une non — Les coefficients s'étalent sur 0,3 point ; les trois erreurs-types honnêtes. . .
- C'est la forme à attendre — Le regroupement en grappes change rarement beaucoup une estimation et change couramment. . .

Ce qu'ajoute une constante aléatoire — En Python

```
print(mixed.summary())  
print(f"between-school variance: {float(mixed.cov_re.iloc[0, 0]):.5f}")  
print(f"residual variance:      {mixed.scale:.5f}")
```

Ce qu'ajoute une constante aléatoire — En R

```
VarCorr(mixed)
```

Ce qu'ajoute une constante aléatoire

Modèle	Variance inter-écoles	Variance résiduelle	CCI
Constante seule	0,00142	0,02042	0,065
+ cantine	0,00081	0,02042	0,038

Ce qu'ajoute une constante aléatoire

- La cantine explique 42,8 % de la variance inter-écoles et rien de la variance intra-école — ce qui est exactement ce. . .

Choisir parmi les trois

Employez	Quand	Coût
Agréger à la grappe	Peu de grappes ; l'exposition est de niveau grappe	Une grande école compte autant qu'une petite
ET robustes aux grappes	Beaucoup de grappes (30+) ; vous voulez le modèle individuel	Peu fiables en dessous d'une trentaine de grappes
Constante aléatoire	Vous voulez les composantes de variance, ou des prédicteurs aux deux niveaux	Suppose l'effet aléatoire non corrélé aux prédicteurs

Choisir parmi les trois

- Avec 24 grappes, agrégez ou ajustez la constante aléatoire, et dites laquelle — Le sandwich robuste est l'outil. . .
- Quand l'exposition ne varie qu'entre grappes, les trois convergent — ce que montre le tableau ci-dessus

Des prédicteurs à deux niveaux — En Python

```
two_level = smf.mixedlm(  
    "rate ~ feeding_programme + age_years + disability_reported",  
    data=d, groups=d["school_id"]).fit()  
print(two_level.summary().tables[1])
```

Des prédicteurs à deux niveaux — En R

```
lmer(rate ~ feeding_programme + age_years + disability_reported +  
      (1 | school_id), data = d)
```

Des prédicteurs à deux niveaux

- `feeding_programme` ne varie qu'entre écoles. `age_years` et `disability_reported` varient à l'intérieur — Un seul...
- Le mode d'échec est de mettre une exposition de niveau grappe dans un modèle sans terme de grappe — C'est la ligne...

Trois niveaux, et quand s'arrêter — En Python

```
# Two grouping levels: households nested within enumeration areas.  
smf.mixedlm("outcome ~ x", data=survey,  
            groups=survey["ea_id"], re_formula="1").fit()
```

Trois niveaux, et quand s'arrêter — En R

```
lmer(outcome ~ x + (1 | ea_id / household_id), data = survey)
```

Trois niveaux, et quand s'arrêter

- Ajoutez un niveau quand quelque chose est attribué ou mesuré à ce niveau et que vous en avez assez d'unités — Trois. . .
- N'ajoutez pas un niveau par souci d'ordre — Un niveau à quatre groupes ajoute un paramètre inestimable et le faux. . .

Rapportez-le en entier — Exemple (suite)

Attendance and school feeding, multilevel analysis

Linear mixed model, 1,200 students in 24 schools.

```
rate ~ feeding_programme + (1 | school_id)
```

Feeding programme +5.10 points 95% CI +2.20 to +8.00

Between-school SD 0.028 Residual SD 0.143

ICC 0.038 with the programme term; 0.065 without it, so the programme accounts for 43% of the between-school variance.

The naive single-level model gives +4.93 points with a standard error of 0.88 rather than 1.48. It is not reported: it treats 50 children in one school as 50 independent observations.

Cluster-robust standard errors on the single-level model (+4.93, SE 1.46)

Rapportez-le en entier — Exemple (suite)

and school-level aggregation (+5.21, SE 1.53) give the same answer. With 24 clusters the robust sandwich is at the edge of its assumptions and is reported as a check rather than as the headline.

Observational. Schools were not randomised into the programme.

Rapportez-le en entier

- Rapporter les trois fait quatre lignes et coupe court à l'objection évidente — que le résultat dépendrait de la méthode

- Cette leçon a accordé le modèle à la façon dont le *programme* a été attribué.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/regression-for-programmes/
reg-06-a-term-for-the-school`