

MATH 60604

Modélisation statistique

§ 6b - Données corrélées

HEC Montréal
Département de sciences de la décision

Exemple: mobilisation au travail

On considère des données corrélées par groupe.

- Une grande entreprise a procédé à une collecte de données auprès de ses employés à l'aide d'un questionnaire.
- La variable réponse est la **mobilisation au travail**.
- L'échelle de mobilisation au travail est définie comme étant la somme de trois items du questionnaire, à savoir
 - Je partage plusieurs des valeurs de l'entreprise.
 - Je me sens loyal à l'entreprise.
 - Je suis fier de dire aux gens pour quelle entreprise je travaille.

mesurés sur une échelle de Likert à cinq points allant de fortement en désaccord (1) à fortement en accord (5).

- Les données sont inspirées de l'étude
Lee, H.-J. et Peccei, R. (2007). Organizational-Level Gender Dissimilarity and Employee Commitment. British Journal of Industrial Relations, 45, 687-712.

- Les données `mobilisation.sas7bdat` contiennent les variables
 - `nunite`: nombre d'employé dans l'unité (département).
 - `idunite`: identification de l'unité (département) où travaille l'employé.
 - `idemploye`: identifiant de l'employé à l'intérieur de l'unité.
 - `anciennete`: ancienneté de l'employé (en années).
 - `sexe`: sexe de l'employé, soit homme (0) ou femme (1).
 - `agegest`: âge du gestionnaire de l'unité (en années).
 - `mobilisation`: score à l'échelle de mobilisation.
- Une **source possible de corrélation** entre les observations est **l'unité d'appartenance**.
- Il est en effet probable que la mobilisation soit affectée par l'unité d'appartenance à cause de facteurs tels le climat de travail, le type de travail effectué par les employés et ainsi de suite.

Dans les données `mobilisation`, le nombre d'employé par unité varie. Par exemple, l'unité 1 a neuf employés (trois femmes et six hommes).

Obs.	nunite	idunite	employe	anciennete	sexe	mobilisation	agegest
1	9	1	1	16	0	8	40
2	9	1	2	18	0	6	40
3	9	1	3	17	1	7	40
4	9	1	4	16	0	8	40
5	9	1	5	13	0	9	40
6	9	1	6	3	0	13	40
7	9	1	7	10	0	9	40
8	9	1	8	4	1	14	40
9	9	1	9	13	1	10	40

Il y a deux types de variables qui peuvent être incluses dans le modèle:

1. celles qui sont fixes pour tous les individus de l'unité (`nunite` et `agegest`)
2. celles qui changent selon les individus (`anciennete` et `sexe`).

- Dans le cas des données *vengeance*, le groupe était l'individu et on a considéré seulement des variables explicatives fixes dans le temps, à part bien sûr la variable temps elle-même.
- Ici, il y a 100 unités en tout et 1016 observations dans le fichier.
- Le but de l'étude est d'évaluer l'impact du sexe, de l'ancienneté, de la taille de l'unité et de l'âge du gestionnaire sur la mobilisation.
- Il faut tenir compte de la possible corrélation intra-unité. Il n'y a pas d'ordre naturel pour les observations au sein d'une unité (contrairement aux données *vengeance* qui contenait des données répétées dans le temps).

- On considère le modèle d'équicorrélation intra-unité pour la corrélation.
 - Cela signifie que la corrélation (conditionnelle) pour chaque paire d'observations (d'employés) est la même au sein de l'unité.

Code SAS pour le modèle linéaire avec erreurs équicorrélées

```
proc mixed data=modstat.mobilisation method=reml;  
class idunite;  
model mobilisation = sexe anciennete agegest  
          nunite / solution;  
repeated / subject=idunite type=cs r=1 rcorr=1;  
run;
```

Matrice de covariance intra-unité (unité 1)

Matrice R estimée pour idunite 1

Ligne	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
2	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
3	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
4	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448
5	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448
6	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448
7	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448
8	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448
9	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709

Valeur estimée du paramètre de covariance			Test du rapport de vraisemblance du modèle nul		
Param. de cov.	Sujet	Estimation	DDL	khi-2	Pr > khi-2
CS	idunite	0.2448			
Residual		1.1261	1	79.49	<.0001

- L'estimé du paramètre de covariance du modèle d'équicorrélation est $\hat{\tau} = 0,2448$ et il est significativement différent de zéro. Cela suggère une association positive entre la mobilisation des travailleurs au sein de l'unité, après avoir pris en compte l'effet des variables explicatives.
- L'estimé de la corrélation entre deux travailleurs intra-unité est $\hat{\rho} = 0,1785$.

Solution pour effets fixes					
Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > t
Intercept	13.7633	0.3955	97	34.80	<.0001
sexe	0.5622	0.06835	914	8.23	<.0001
anciennete	-0.4722	0.006015	914	-78.50	<.0001
agegest	0.01929	0.006801	97	2.84	0.0056
nunite	0.006470	0.02019	97	0.32	0.7493

- Trois des variables explicatives sont significatives : **sexe**, **anciennete** et **agegest**:
 - Les femmes sont plus mobilisées que les hommes en moyenne.
 - Plus l'employé a de l'ancienneté, moins il est mobilisé.
 - Plus le gestionnaire est âgé, plus l'employé est mobilisé.
- Par contre, la taille de l'unité n'est pas significative, au-delà des autres variables.

Exemple: mobilisation au travail

- Il pourrait être intéressant d'incorporer un effet unité dans le modèle.
- En ajoutant un effet fixe au niveau du groupe, on perd en revanche la possibilité d'estimer les effets des variables qui sont fixes au niveau du groupe.
- Ça voudrait dire qu'on ne pourrait pas inclure les variables `agegest` et `nunite`.
- Il existe une façon de quand même incorporer un effet groupe, tout en conservant la possibilité d'avoir des variables fixes au niveau du groupe.
- Il s'agit d'utiliser des **effets aléatoires** plutôt que des effets fixes.
- De plus, nous verrons que les effets aléatoires servent aussi à modéliser la structure de covariance.