

# Quelques outils pour travailler sur des bases de données

---

Cours Sol 6210 - Analyse quantitative avancée

Claire Durand, 2011



# La création de variables agrégées

---

- Pourquoi? Il est parfois nécessaire
  - ▶ 1. De créer des variables qui sont par exemple, la moyenne des observations pour un groupe d'individus, le maximum d'une valeur pour un individu donné...
  - ▶ 2. D'ajouter ces variables au fichier ou de créer un fichier de variables agrégées.
- Fonction "aggregate" dans SPSS



# Aggregate : exemple 1

---

- Données de primes pour professeurs pour chaque année
  - ▶ A) Je désire avoir l'information suivante : le maximum et le minimum de primes qu'un professeur a obtenues durant les 10 ans
  - ▶ B) Je désire créer un fichier de niveau 2 qui comprend ces informations de même que la moyenne des primes – par professeur et par département. Je pourrais aussi ajouter l'information concernant la proportion de professeurs avec primes.. Ou titulaires...



# Aggregate examples -2011

---

- Données d'élèves par classe
- Je désire obtenir un fichier où,
  - ▶ pour chaque classe, j'aurai la moyenne de diverses variables relatives à l'attitude des élèves.
  - ▶ Pour chaque musicien, j'aurai le pourcentage de notes identifiées.
- Étapes:
  - ▶ 1) TRIER le fichier selon la variable d'agrégation.
  - ▶ 2) agréger



# Rectangulariser un fichier

---

- Pourquoi faire?
  - ▶ J'ai un fichier qui comprend, par exemple, une ligne par individu par année et la procédure demande que j'aie une ligne par individu avec toutes les variables sur la même ligne.
  - ▶ J'ai un fichier qui comprend une ligne par membre du ménage et j'ai besoin d'un fichier qui comprend une ligne par ménage avec les informations sur tous les membres du ménage, comme par exemple l'âge de chaque membre, sa scolarité, etc.



# Rectangulariser un fichier : exemple 1.

---

- J'ai un fichier comprenant une ligne par professeur par année. J'aimerais avoir un fichier comprenant une ligne par professeur avec la suite des primes à chaque année, i.e.
  - ▶ Passer de "prof1" "prime" pour les années 1 à 10
  - ▶ À prof1 prime1 prime2 prime3... prime10.
  - ▶ Note : il y aura autant de variables que le nombre de variables originales multiplié par le nombre d'années, ceci uniquement pour les variables qui changent à chaque année (et donc pas pour le sexe, par exemple).
- Procédure: restructurer "des observations aux variables"



# Exemples - 2011

---

- Pour chaque musicien, je veux une ligne où,
  - ▶ Pour chaque note, il y aura une variable qui indiquera la stratégie utilisée et la réussite.



# Créer un fichier personne-période

---

- Pourquoi faire?
  - ▶ Il s'agit de l'inverse de l'opération précédente. On a un fichier avec une ligne par personne ou par ménage, par exemple et on veut avoir un fichier personne-période ou ménage-personne c'est-à-dire un fichier où il y a plusieurs lignes par personne (une pour chaque période) ou plusieurs lignes par ménage (une pour chaque membre du ménage). C'est le genre de fichier que l'on veut pour l'analyse multi-niveaux au niveau 1.



# Créer un fichier personne-période

---

## Exemple 1.

- J'ai un fichier comprenant une ligne par professeur et pour chaque professeur une série de variables de type prime-année 1, prime-année 2, etc. J'aimerais avoir un fichier comprenant une ligne par professeur par année, i.e.
  - ▶ Passer de prof1 prime1 prime2 prime3... prime10
  - ▶ À "prof1" "année" "prime" pour les années 1 à 10
  - ▶ Note : il y aura autant de variables que le nombre de variables originales divisé par le nombre d'années, ceci uniquement pour les variables qui changent à chaque année (et donc pas pour le sexe, par exemple).
- Procédure : restructurer "cases to vars"



# Exemple 2011

---

- Faire l'inverse de l'opération précédente avec les musiciens et les notes.



# Joindre des fichiers : ajouter des observations (1)

---

- Pourquoi ? Quand on a plusieurs fichiers comprenant certaines variables communes – ca peut être des fichiers pour chaque année, pour chaque école, district, etc.
- On peut créer un nouveau fichier comprenant toutes les observations ce qui permet de faire des analyses conjointes et comparatives.
- Les variables doivent porter le même nom et être de même type, longueur, etc. Il peut être pertinent d'identifier la provenance des cas (variable population).



# Joindre des fichiers : ajouter des observations (2)

---

Comment?

- Procédure: Fusionner des fichiers - ajouter des observations. Automatiquement, la procédure fera apparaître les variables similaires (à condition qu'on leur ait donné le même nom et le même format).
- Le nouveau fichier comprend toutes les observations pour les variables similaires (et seulement pour les variables similaires).



# Joindre des fichiers : ajouter des variables (1)

---

- Pourquoi ? On voudrait ajouter à un fichier des informations qui proviennent d'une autre source.
  - ▶ Par exemple, si on a demandé les trois premiers caractères du code postal aux répondants à un sondage, on pourrait ajouter l'information provenant de Statistique Canada relativement aux pâtés de maison correspondant au code postal dans le fichier et utiliser ces informations comme prédicteurs dans une analyse.



# Joindre des fichiers : ajouter des variables (2)

---

## Comment?

- Dans SPSS, données - fusionner des fichiers - ajouter des variables.
  - Exemple: sondage pré-post électoral.
- Note: les deux fichiers ne doivent pas obligatoirement comprendre le même nombre d'observations. Le fichier final aura autant d'observations qu'il y en a en tout dans les deux fichiers.
- Note: le fichier fusionné pourrait avoir des informations d'un autre niveau comme les informations sur l'école de l'élève. Dans ce cas, il faut informer SPSS que l'un des fichiers "est une table".



# Passer d'un format à un autre

---

- On peut, directement avec plusieurs logiciels, lire des données écrites dans le format d'un autre logiciel (spss, stata, excel, etc.) . Toutefois, ce n'est pas toujours optimal, surtout à partir de Excel.
- On peut également utiliser un logiciel comme Stat transfer (pas très cher) qui fera automatiquement les opérations.

