

Vous trouverez le fichier de données *ses_v98_500.rda* dont vous avez besoin pour ce TP sur la page dédiée à ce cours du site web <http://www.andreberchtold.com>. Ce fichier contient des données concernant 500 étudiants de l'université de Genève.

Exercice 1 Recodage et transformation de variables

1. La variable *DOMMERE* représente le domicile de la mère de l'étudiant. Recoder cette variable en un facteur *DOMMERE3* ne contenant que 3 catégories : 1 pour le canton de Genève, 2 pour le reste de la Suisse et 3 pour l'étranger. N'oubliez pas de donner un nom à chacun des niveaux du facteur que vous avez créé. Examinez ensuite la distribution de *DOMMERE3*, puis construisez une table de contingence pour ces deux variables.
2. La variable *ANNENAI3* représente l'année de naissance de l'étudiant. Recoder cette variable en un facteur *ANNENAI4* contenant 4 catégories : 1 (< 1972), 2 (1972-1976), 3 (1977-1978), 4 (1979-1981). N'oubliez pas de donner un nom à chacun des niveaux du facteur que vous avez créé. Examinez ensuite et comparez la distribution de la variable numérique *ANNENAI3* et celle du facteur *ANNENAI4*.
3. Ecrivez un script permettant de construire les 3 facteurs suivants et d'examiner leur distribution :

Variable d'origine	Facteur	Description	Catégories
<i>PROFPERE</i>	<i>PROFPERE5</i>	profession du père	1 ouvrier, employé 2 artisan, commerçant, cadre moyen 3 cadre supérieur, prof. libérale 4 sans profession 5 chômeur
<i>DATIMMAT</i>	<i>DATIMMAT3</i>	date immatriculation	1 avant 97 2 97 3 98
<i>DIPLSEC</i>	<i>DIPLSEC3</i>	type de dipl. secondaire	1 classique, latine 2 moderne, scientifique, économ. 3 autre

Exercice 2 Sélection de données

Il est souvent nécessaire de créer un nouveau fichier de données (dataframe) comme un sous-ensemble d'un fichier de données existant.

1. Visualiser le fichier de données à l'aide de la commande View().
2. Créer un nouveau fichier de données ne comportant que les étudiants de sciences sociales.
3. Créer un nouveau fichier de données ne comportant que les étudiants dont l'année de naissance (variable ANNENNAIS) est comprise entre 77 et 79.
4. Créer un nouveau fichier de données ne comportant que les étudiants de sciences sociales de sexe féminin et dont la nationalité (variable NATIONAL) est Genève, Vaud ou Valais.

Exercice 3 Création de variables

La construction d'une nouvelle variable peut souvent faire intervenir des informations provenant de plusieurs autres variables.

1. Créer une nouvelle variable nommée FG codée 1 pour les étudiants de sexe féminin et de nationalité "Genève" et zéro pour les autres.
2. Créer une nouvelle variable nommée DOM2 codée 1 pour les étudiants dont le domicile du père (DOMPERE) est identique à celui de la mère (DOMMERE) et zéro pour les autres. Attention : En cas de données manquantes sur DOMPERE et/ou DOMMERE, la nouvelle variable doit être codée NA.