

Fichiers de données

1 Consommation de protéines en Europe

Le fichier **Protéines_2** contient des informations relatives à la consommation de protéines dans 20 pays d'Europe au début des années 1970. Source : <http://lib.stat.cmu.edu/DASL/Datafiles/Protein.html>

Nous disposons de 9 variables donnant de la consommation journalière moyenne en grammes de protéines de 9 origines différentes (viande rouge, viande blanche, oeufs, lait, poisson, céréales, féculents, noix, fruit & légumes). Nous disposons également des 4 variables suivantes :

- Pays : Nom du pays.
- Europe : Variable dichotomique codée 0 pour les pays d'Europe de l'est et 1 pour ceux d'Europe de l'ouest.
- Animal : Pourcentage des protéines d'origine animale.
- Animal2 : Variable dichotomique codée 1 pour les pays dont au moins 50% des protéines sont d'origine animale et 0 sinon.

Ces données peuvent être utilisées pour comparer la consommation de protéines entre Europe de l'ouest et Europe de l'est, ainsi que pour évaluer la présence de relation entre les sources de protéines.

2 Composition des céréales

Le fichier **Céréales** contient des informations concernant 77 sortes de céréales pour le petit-déjeuner qui ont été analysées par le journal américain Consumer Reports. Une note comprise entre zéro et 100 a été attribuée à chaque sorte de céréales en fonction de ses qualités diététiques.

Le fichier contient les variables suivantes :

- Marque : Marque des céréales (1 : General Mills, 2 : Kelloggs, 3 : Autre).
- Calorie : Nombre de calories par portion.
- Protéine : Nombre de grammes de protéines par portion.
- Graisse : Nombre de grammes de graisses par portion.
- Sodium : Nombre de milligrammes de sodium par portion.
- Fibres : Nombre de grammes de fibres par portion.
- Carbone : Nombre de grammes de carbone par portion.
- Sucre : Nombre de grammes de sucre par portion.
- Potassium : Nombre de milligrammes de potassium par portion.
- Vitamine : % des recommandations en vitamines de la FDA.
- Rayon : Emplacement sur les rayons dans le magasin (1 : rayon du bas, 2 : mi-hauteur, 3 : rayon du haut).
- Note : Note globale attribuée par Consumer Reports.

Ces données peuvent être utilisées pour identifier les composantes de la note globale attribuée par Consumer Reports en fonction de la composition des céréales.

3 Tour des Dents du Midi

Le fichier **Tddm** contient des données relatives à l'édition 2000 du Tour des Dents du Midi (course à pied sur la distance du marathon). Une population de 102 coureuses et coureurs a participé à la course. Nous disposons d'un échantillon de taille 42 recueilli aléatoirement parmi les participants.

Le fichier contient les variables suivantes :

- Dossard : Numéro de dossard.
- Age : Age en années.
- Temps : Temps réalisé pour la course en heures.
- P_avant : Poids en kilos avant la course.
- P_après : Poids en kilos après la course.
- Taille : Taille en centimètres.
- Masse_g : Index de masse grasse.

Ces données peuvent être utilisées pour étudier la relation entre les caractéristiques physiologiques des participants et leur résultat.

4 Poids des bébés à la naissance

Le fichier **Poids_naissance** contient des données relatives au poids à la naissance de 189 bébés, ainsi que des informations concernant leurs mères. Parmi les bébés, 59 avaient un poids trop faible à la naissance (moins de 2500 grammes). Source : Hosmer and Lemeshow (2000) *Applied Logistic Regression : Second Edition*. Copyright John Wiley & Sons Inc.

Le fichier contient les variables suivantes :

- ID : Identificateur numérique unique de chaque bébé.
- Poids : Poids du bébé à la naissance en grammes.
- Poids_r : Poids recodé (0 : ≥ 2500 grammes, 1 : < 2500 grammes).
- Age : Age de la mère en années.
- Fumeuse : Mère fumeuse (0 : non, 1 : oui).

Ces données peuvent être utilisées pour déterminer si l'âge de la mère ou le fait qu'elle fume sont liés de façon significative avec la probabilité que le bébé ait un poids très faible à la naissance.

5 Santé

Le fichier **Santé** contient des informations concernant la santé pour un échantillon de 17030 personnes. Source : Hosmer and Lemeshow (2000) *Applied Logistic Regression : Second Edition*. John Wiley & Sons Inc.

Le fichier contient les variables suivantes :

- Age : Age en année.
- Sexe : Sexe (0 : femme, 1 : homme).
- Poids : Poids en livres.
- Taille : Taille en pouces.
- P_sys : Pression systolique.
- P_sys_r : Pression systolique recodée (0 : ≤ 140 , 1 : > 140).
- P_dia : Pression diastolique.
- Fumeur : Etre fumeur (1 : non, 2 : ancien, 3 : oui).

Ces données peuvent être utilisées pour identifier un lien entre le poids, le fait de fumer et la pression sanguine.

6 Santé dans les pays de l'OCDE

Le fichier **Santé_OCDE** contient des informations relatives à l'état de santé de la population dans 30 pays de l'OCDE. Source : OECD HEALTH DATA 2004, 1st edition.

Le fichier contient les variables suivantes :

- Pays : Nom du pays
- Région : Région du monde (1 : Europe de l'ouest, 2 : reste de l'Europe, 3 : reste du monde).
- Vie_f : Espérance de vie des femmes à la naissance en 2001.
- Vie_h : Espérance de vie des hommes à la naissance en 2001.
- Mort_infantile : Mortalité infantile pour 1000 naissances vivantes en 2001.
- Césariennes : Nombre de césariennes pour 1000 naissances vivantes en 2001.
- Age : Pourcentage de personnes de 65 ans et plus dans la population en 2001.
- Alcool : Consommation annuelle d'alcool en litres par habitant de plus de 15 ans en 2001.
- Tabac : Pourcentage de fumeurs réguliers (tous les jours) en 2001.
- Obésité : Pourcentage de personnes obèses en 2001.

Ces données peuvent être utilisées pour comparer les caractéristiques sanitaires de la population entre les 3 régions du monde considérées. On peut aussi étudier la relation entre l'espérance de vie à la naissance et les autres variables.

7 Santé à 18 ans

Le fichier **Santé_18** contient des informations relatives à la santé des jeunes de 18 ans en Suisse. Nous disposons d'un échantillon de 275 jeunes ayant eu 18 ans en 2004 et représentatifs de l'ensemble des jeunes de 18 ans vivant en Suisse. Source : Panel Suisse des ménages.

Le fichier contient les variables suivantes :

- Sexe : Sexe (1 : homme, 2 : femme).
- Santé : Santé en général (de 1 : très bonne, à 5 : très mauvaise).
- Activité : Nombre de jours par semaine avec une activité physique (de 0 à 7).
- Relation : Satisfaction dans les relations avec les autres (de 0 : pas du tout satisfait, à 10 : pleinement satisfait).
- Docteur : Consultation d'un docteur au moins une fois lors des 12 derniers mois (1 : oui, 2 : non).

Ces données peuvent être utilisées pour identifier certaines des raisons d'un mauvais état de santé.