



Examen Data Exploration - ING1 MAN

2 feuilles R/V manuscrites autorisées, Calculatrice autorisée

Date : 18 décembre 2025

Durée : 1h

Nombre de pages : 2

Deux rendus sont attendus : le code réalisé sur RStudio et un document écrit présentant vos réponses (explications), suivis d'une présentation orale.

◇ ◇ ◇

1 Partie I : Statistiques (5pt)

(L'étudiant peut effectuer les calculs soit à la main, soit à l'aide de RStudio)

Un enseignant étudie les notes que ses élèves ont eu à son dernier examen

Notes	5	6	7	9	10	12	13
Effectifs	1	1	5	8	7	6	4

- (1.5pt) Rentrer les données dans Rstudio.
- (1.5pt) Donner les valeurs de la moyenne, le première quartile, le troisième quartile, la variance, l'écart inter-quartile et l'étendue.
- (1pt) L'enseignant décide d'ajouter un point à tout le monde. Quel(s) indicateur(s) ne sont pas impacté(s) ? Expliquer.
- (1pt) L'élève ayant eu 5 bénéficie d'un tier-note. Quel(s) indicateur(s) ne sont pas impacté(s) ? Expliquer.

2 Partie II : Data exploration

2.1 Présentation de données

Considérons les données des valeurs nutritionnelles de 34 légumes enregistrées dans un data-frame sur R nommé *tab*, dont voici un extrait pour 9 légumes :

legumes (Pour 100g)	energie (kcal)	eau (g)	fibres (g)	glucides(g)	protides(g)	lipides(g)	famille
...
Petit pois	80	74	6	12.3	6	0.7	graine
Pissenlit	40	85.5	3.5	5.7	2.7	0.7	feuilles
Poireau	27	90.5	3.5	4	2	0.3	feuilles
Poivron	21	91	2	3.5	1.1	0.3	fruit
Pomme de terre	85	77	2.1	19	2	0.1	racine
Potiron	20	92.8	1.3	4.1	0.8	0.1	fruit
Radis	15	94.5	1.5	3	0.6	0.1	tige
Salsifis	30	79	4	5	1.4	0.4	racine
Topinambour	31	79	7.6	4.5	2.4	0.4	racine

Légende : kcal : kilocalorie. g : gramme

La dernière colonne concerne une classification selon la partie de la plante que l'on consomme.

2.2 Importer les données et statistiques univariées (4pt)

1. (0.5pt) Lire le jeu de données à l'aide de l'instruction : `tab = read.table(...)`.
2. (1pt) Combien y a-t-il d'observations ? Combien y a-t-il de variables ?
3. (1.5pt) Choisir une variable quantitative et réaliser une représentation graphique appropriée de sa distribution. Avec la même variable, tracer également une boîte de Tukey (boxplot). Est-ce qu'il y a des individus atypiques ?
4. (1pt) Nommer la variable qualitative et réaliser une représentation graphique appropriée.

2.3 Statistiques bivariées

2.3.1 Quantitative \times Quantitative (3pt)

À l'aide de la fonction `lm` avec les variables `energie` et `glucides`, répondre aux questions suivantes :

1. (0.5pt) Tracer nuage de point de ces deux variables. Deviner relation entre elles. Donner le coefficient de corrélation entre elles.
2. (1pt) Donner la droite de régression de `energie` en fonction de `glucides`.
3. (0.5pt) Tracer la droite sur graphique.
4. (0.5pt) Le modèle est-il valide ? Expliquer.
5. (0.5pt) Pour une légume dont valeur `glucides` est 12, donner la valeur prédite de `energie`.

2.3.2 Quantitative \times Qualitative (2pt)

À l'aide de la fonction `lm` avec les variables `energie` et `type`, répondre aux questions suivantes :

1. (1pt) Quelles sont la variance inter-classe et la variance intra-classe ?
2. (1pt) Existe-t-il une relation entre les `type` et la variable `energie` ?

2.4 Effectuer une ACP (6pt)

Réaliser une ACP avec la fonction `PCA` du package `FactoMineR`, en ajoutant la variable `type` comme variable supplémentaire. Le package `factoextra` peut être utilisé pour l'affichage des résultats.

1. (0.5pt) Réaliser une ACP
2. (1pt) Combien d'axes au total ? Combien d'axes doit-on retenir pour l'étude ? Justifier votre réponse.
3. (0.5pt) Combien de groupes de variables observe-t-on ?
4. (1.5pt) Quelles variables contribuent le plus à la construction de l'axe 1 ? de l'axe 2 ? Donnez la réponse sous forme numérique et ajoutez un graphique pour faciliter la visualisation.
5. (2pt) Quelle est la contribution moyenne d'une observation ? Certaines observations ont-elles une contribution anormalement importante sur l'axe 1 et l'axe 2 ? Donnez la réponse sous forme numérique et ajoutez un graphique pour faciliter la visualisation.
6. (0.5pt) Interpréter les caractéristiques du `maïs`