

Exercice 1**Tout (ou presque) en un****16 points**

La tableau **X** (donné au verso) présente les données d'un fichier `cereal.mat`, présent sur votre bureau, contenant les mesures des ingrédients de 77 produits de céréales pour petit déjeuner. Sur chaque type de céréale, 8 variables ont été mesurées. Ces variables sont (dans l'ordre) :

Calories	la teneur en calories (le nombre pour un repas)
Carbo	la teneur en carbohydrates (en gramme pour un repas)
Cups	le nombre de tasse recommandé par repas
Fat	la teneur en graisse (en gramme pour un repas)
Fiber	la teneur en fibres alimentaires (en gramme pour un repas)
Potass	la teneur en potassium (en mg pour un repas)
Protein	la teneur en protéine (en gramme pour un repas)
Sugars	la teneur en sucre (en gramme pour un repas)

A chaque type de céréale a été associé son nom contenu dans la variable **Name**. Ainsi la première céréale est `Name(1) -> '100% Bran'`.

- Nous allons commencer par nous intéresser à la deuxième variable (**Carbo**) uniquement.
 - construisez un tableau avec les fréquences, les fréquences cumulées de la deuxième variable (**Carbo**).
 - tracez sa fonction de repartition empirique
 - déterminez graphiquement la médiane et les quartiles
 - en déduire la DIQ
 - dessinez la boîte à moustache associée sur une feuille
 - s'il y a une variable aberrante enlevez la et recalculez la médiane
- Nous allons maintenant nous intéresser au couple de variables 1 et 8 (**Calories** et **Sugars**).
 - représenter le nuage des points de ces deux variables à l'aide de votre ordinateur
 - trouvez graphiquement la médiane de Tuckey du nuage de points donné figure 1.
 - calculez la covariance et la corrélation de ce couple de variables
 - cette valeur de corrélation vous semble t'elle significativement importante ?
 - pourquoi la variable **Cv** n'a t'elle pas exactement la même valeur que `C(2,1)` ?

```
n = length(x1);
c1 = mean(x1); c8 = mean(x8);
Cv = x1'*x8/n - c1*c8
C = cov(x1,x8);
```
 - pourquoi la variable **R** a t'elle exactement la même valeur que `cc(8,1)` ?

```
vx1 = (x1 - c1)'*(x1 - c1)/n; vx8 = (x8 - c8)'*(x8 - c8)/n;
R = Cv/sqrt(vx1*vx8)
cc = corrccoef(X)
```
- Nous allons maintenant traiter du tableau **X** dans sa globalité.
 - centrez et réduisez les variables. Expliquez pourquoi cette étape est nécessaire avant d'effectuer une analyse en composantes principales (ACP)
 - calculez la matrice de variance covariance des données. Commentez.
 - calculez les composantes (ou facteurs ou directions) principales.
 - projetez les individus sur les deux composantes les plus informatives (si vous voulez mettre les noms des céréales associées c'est stylé)
 - la figure 2 donne la représentation des variables sur ces deux premières composantes. À la vue de cette figure, quelle interprétation donnez vous aux deux axes principaux. Quel est le rôle de la variable **Cups** ?

Exercice 2**Assimilation****4 points**

Soit m une moyenne connue et λ un facteur de confiance connu lui aussi. On souhaite connaître la valeur centrale d'un ensemble de points $x_i, i = 1, n$, sachant que cette valeur doit être proche de m avec une confiance λ . On modélise ce problème à travers une fonction coût J :

$$J(c) = \sum_{i=1}^n (x_i - c)^2 + \lambda(c - m)^2$$

1. Donnez l'expression de c qui minimise le coût J .
2. Application numérique : calculez c pour $m = 153$, $\lambda = 20$, $n = 10$ et $x_1, \dots, x_{10} = 140, 145, 151, 155, 140, 125, 132, 129, 163, 133$

x = [

70.0000	5.0000	0.3300	1.0000	10.0000	280.0000	4.0000	6.0000
120.0000	8.0000	-1.0000	5.0000	2.0000	135.0000	3.0000	8.0000
70.0000	7.0000	0.3300	1.0000	9.0000	320.0000	4.0000	5.0000
50.0000	8.0000	0.5000	0	14.0000	330.0000	4.0000	0
110.0000	14.0000	0.7500	2.0000	1.0000	-1.0000	2.0000	8.0000
110.0000	10.5000	0.7500	2.0000	1.5000	70.0000	2.0000	10.0000
110.0000	11.0000	1.0000	0	1.0000	30.0000	2.0000	14.0000
130.0000	18.0000	0.7500	2.0000	2.0000	100.0000	3.0000	8.0000
90.0000	15.0000	0.6700	1.0000	4.0000	125.0000	2.0000	6.0000
90.0000	13.0000	0.6700	0	5.0000	190.0000	3.0000	5.0000
120.0000	12.0000	0.7500	2.0000	0	35.0000	1.0000	12.0000
110.0000	17.0000	1.2500	2.0000	2.0000	105.0000	6.0000	1.0000
115.0000	13.0000	0.7500	3.0000	0	45.0000	1.0000	9.0000
110.0000	13.0000	0.5000	2.0000	2.0000	105.0000	3.0000	7.0000
110.0000	12.0000	1.0000	1.0000	0	55.0000	1.0000	13.0000
110.0000	22.0000	1.0000	0	0	25.0000	2.0000	3.0000
100.0000	21.0000	1.0000	0	1.0000	35.0000	2.0000	2.0000
110.0000	13.0000	1.0000	0	1.0000	20.0000	1.0000	12.0000
110.0000	12.0000	1.0000	1.0000	0	65.0000	1.0000	13.0000
110.0000	10.0000	0.5000	3.0000	4.0000	160.0000	3.0000	7.0000
100.0000	21.0000	1.0000	0	1.0000	-1.0000	3.0000	0
110.0000	21.0000	1.0000	0	1.0000	30.0000	2.0000	3.0000
100.0000	11.0000	0.7500	1.0000	2.0000	120.0000	2.0000	10.0000
100.0000	18.0000	0.7500	0	1.0000	80.0000	2.0000	5.0000
110.0000	11.0000	1.0000	1.0000	1.0000	30.0000	2.0000	13.0000
110.0000	14.0000	0.7500	0	1.0000	25.0000	1.0000	11.0000
100.0000	14.0000	0.8000	0	3.0000	100.0000	3.0000	7.0000
120.0000	12.0000	0.6700	2.0000	5.0000	200.0000	3.0000	10.0000
120.0000	14.0000	0.6700	0	5.0000	190.0000	3.0000	12.0000
110.0000	13.0000	0.7500	1.0000	0	25.0000	1.0000	12.0000
100.0000	11.0000	0.8800	0	0	40.0000	2.0000	15.0000
110.0000	15.0000	0.7500	1.0000	0	45.0000	1.0000	9.0000
100.0000	15.0000	0.8800	1.0000	3.0000	85.0000	3.0000	5.0000
110.0000	17.0000	0.2500	0	3.0000	90.0000	3.0000	3.0000
120.0000	13.0000	0.3300	3.0000	3.0000	100.0000	3.0000	4.0000
120.0000	12.0000	1.0000	2.0000	1.0000	45.0000	1.0000	11.0000
110.0000	11.5000	0.7500	1.0000	1.5000	90.0000	3.0000	10.0000
110.0000	14.0000	1.3300	0	0	35.0000	1.0000	11.0000
110.0000	17.0000	-1.0000	1.0000	1.0000	60.0000	2.0000	6.0000
140.0000	20.0000	0.7500	1.0000	2.0000	95.0000	3.0000	9.0000
110.0000	21.0000	1.5000	1.0000	0	40.0000	2.0000	3.0000
100.0000	12.0000	0.6700	2.0000	2.0000	95.0000	4.0000	6.0000
110.0000	12.0000	1.0000	1.0000	0	55.0000	2.0000	12.0000
100.0000	16.0000	-1.0000	1.0000	0	95.0000	4.0000	3.0000
150.0000	16.0000	-1.0000	3.0000	3.0000	170.0000	4.0000	11.0000
150.0000	16.0000	-1.0000	3.0000	3.0000	170.0000	4.0000	11.0000
160.0000	17.0000	0.6700	2.0000	3.0000	160.0000	3.0000	13.0000
100.0000	15.0000	1.0000	1.0000	2.0000	90.0000	2.0000	6.0000
120.0000	15.0000	0.6700	1.0000	0	40.0000	2.0000	9.0000
140.0000	21.0000	0.6700	2.0000	3.0000	130.0000	3.0000	7.0000
90.0000	18.0000	-1.0000	0	3.0000	90.0000	3.0000	2.0000
130.0000	13.5000	0.5000	2.0000	1.5000	120.0000	3.0000	10.0000
120.0000	11.0000	0.6700	1.0000	6.0000	260.0000	3.0000	14.0000
100.0000	20.0000	1.0000	0	1.0000	45.0000	3.0000	3.0000
50.0000	13.0000	1.0000	0	0	15.0000	1.0000	0
50.0000	10.0000	-1.0000	0	1.0000	50.0000	2.0000	0
100.0000	14.0000	0.5000	1.0000	2.0000	110.0000	4.0000	6.0000
100.0000	-1.0000	0.6700	2.0000	2.7000	110.0000	5.0000	-1.0000
120.0000	14.0000	0.7500	1.0000	5.0000	240.0000	3.0000	12.0000
100.0000	10.5000	0.5000	2.0000	2.5000	140.0000	3.0000	8.0000
90.0000	15.0000	0.5000	0	2.0000	110.0000	2.0000	6.0000
110.0000	23.0000	1.1300	0	0	30.0000	1.0000	2.0000
110.0000	22.0000	1.0000	0	0	35.0000	2.0000	3.0000
80.0000	16.0000	-1.0000	0	3.0000	95.0000	2.0000	0
90.0000	19.0000	0.6700	0	4.0000	140.0000	3.0000	0
90.0000	20.0000	0.6700	0	3.0000	120.0000	3.0000	0
110.0000	9.0000	0.7500	1.0000	1.0000	40.0000	2.0000	15.0000
110.0000	16.0000	1.0000	0	1.0000	55.0000	6.0000	3.0000
90.0000	15.0000	-1.0000	0	3.0000	90.0000	2.0000	5.0000
110.0000	21.0000	1.0000	1.0000	0	35.0000	2.0000	3.0000
140.0000	15.0000	1.0000	1.0000	4.0000	230.0000	3.0000	14.0000
100.0000	16.0000	1.0000	1.0000	3.0000	110.0000	3.0000	3.0000
110.0000	21.0000	0.7500	1.0000	0	60.0000	2.0000	3.0000
110.0000	13.0000	1.0000	1.0000	0	25.0000	1.0000	12.0000
100.0000	17.0000	0.6700	1.0000	3.0000	115.0000	3.0000	3.0000
100.0000	17.0000	1.0000	1.0000	3.0000	110.0000	3.0000	3.0000
110.0000	16.0000	0.7500	1.0000	1.0000	60.0000	2.0000	8.0000

]

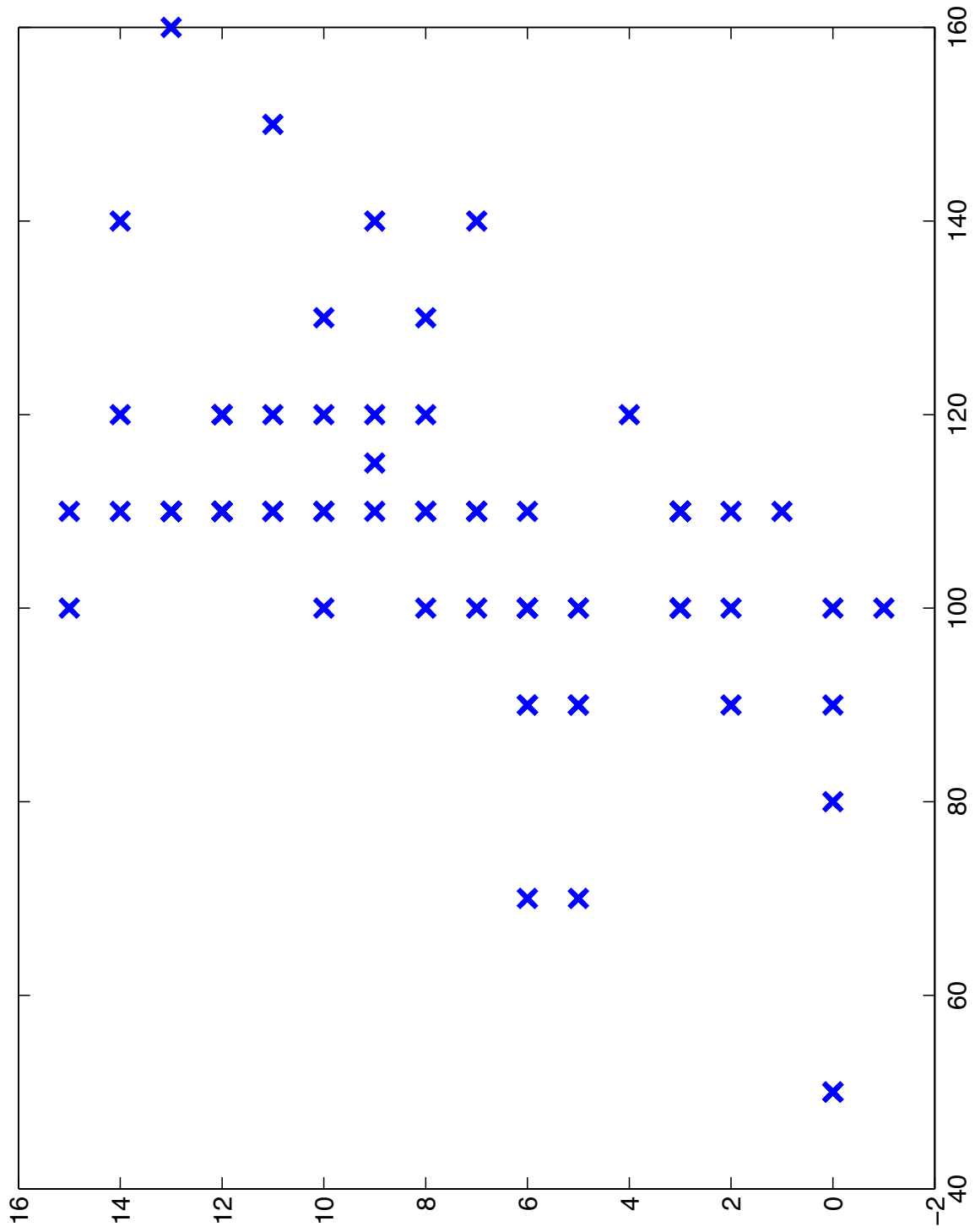


FIGURE 1 – Trouvez la médiane de Tuckey de ce nuage de points.

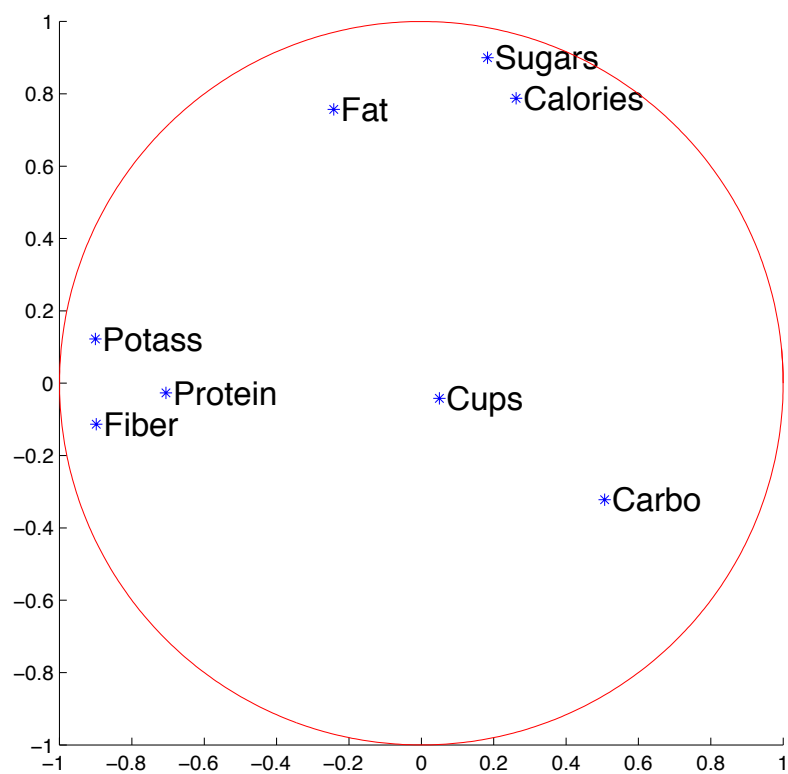


FIGURE 2 – Représentation des variables sur ces deux premières composantes de l'ACP du tableau X des céréales.

Tracez ci-dessous la boîte à moustache de la deuxième variable (représentant la teneur en complexe carbohydrates exprimée en grammes).