

Rédigé par : l'équipe enseignante

Durée : 1h30

A l'intention de : Elèves d'ING1-GI

Document ou matériel autorisés : Calculatrice uniquement

Les questions QCM peuvent être à réponses multiples et sont à points négatifs

1. Exercice 1 : indicateurs numériques

Question 1.1.

On donne la série statistique suivante : 5, 6, 8, 2, 1, 9, 5, 11, 5, 10, 3

La médiane

- =9
 =3
 =5
 =8
 ∈ [3,5]
 ∈ [8,9]
 aucune réponse

Le 1^{er} quartile

- =9
 =3
 =5
 =8
 ∈ [3,5]
 ∈ [8,9]
 aucune réponse

Le 3^{ème} quartile

- =9
 =3
 =5
 =8
 ∈ [3,5]
 ∈ [8,9]
 aucune réponse

Question 1.2.

Soit la répartition suivante :

x_i	Effectif	Fréquence	Fréquence cumulée
[3 ; 4[10	0,024	0,024
[4 ; 6[30	0,061	0,085
[6 ; 8[120	0,246	0,331
[8 ; 9[208	0,431	0,762
[9 ; 10[??	0,184	0,946
[10 ; 12[22	0,051	~1

Sachant que la moyenne calculée sur ce tableau vaut approximativement 8,2, quelle est la valeur de l'effectif non renseigné ?

- =50
 =120
 =90
 =119,7
 =126,8
 aucune réponse

Le 1^{er} quartile est égal à

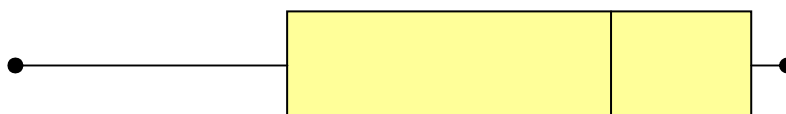
- =7,34
 =8,39
 =120
 =10,2
 =210
 aucune réponse

La médiane est égale à

- =7,34
 =8,39
 =120
 =240
 =50%
 aucune réponse

Question 1.3.

Soit la répartition d'une variable quantitative résumée par la boîte de Tuckey suivante



- La moyenne et la médiane sont à peu près égales
- La moyenne est plus grande que la médiane
- La moyenne est plus petite que la médiane
- L'écart-type et la moyenne sont à peu près égaux
- La moyenne est plus petite que l'écart-type
- La moyenne est plus grande que l'écart-type
- On ne peut rien dire sur l'ordre des valeurs de la médiane et de la moyenne
- On ne peut rien dire sur l'ordre des valeurs de l'écart-type et de la moyenne

2. Exercice 2 : croisement de deux variables

Question 2.1.

Soient $\{x_i\}_{i=1,\dots,n}$ et $\{y_i\}_{i=1,\dots,n}$ deux séries numériques observées sur les mêmes individus.

Si le coefficient de corrélation est égal à 1 alors les droites de régression et y en x et de x en y sont

- Confondues Parallèles Perpendiculaires

Le coefficient de détermination est toujours supérieur au coefficient de corrélation

- Vrai Faux

Si le coefficient de corrélation est nul alors

- x est non corrélés à y x et y sont indépendantes
 y est non corrélés à x x et y sont liés fonctionnellement

Question 2.2.

Dans une population Ω de taille n, on observe deux variables :

- une qualitative, $x=\{x_k\}_{k=1,\dots,n}$, à p modalités notées, m_1,\dots,m_p
- une quantitative continue $y=\{y_k\}_{k=1,\dots,n}$ de moyenne \bar{y} et de variance s_y^2 .

On suppose que les modalités de la série x définissent des sous-populations, $\Omega = \Omega_1 \cup \dots \cup \Omega_p$ où $\Omega_i \cap \Omega_j = \emptyset$, de tailles respectives n_1,\dots,n_p .

On peut alors considérer les restrictions de la caractéristiques y sur chacune des sous-populations et calculer les indicateurs numériques usuels pour chaque modalités de x,

- moyennes : \bar{y}_i , $i=1,\dots,p$
- variances : s_i^2 , $i=1,\dots,p$

- 1) Rappelez la formule de la décomposition de la variance
- 2) Expliquez l'objectif d'un algorithme de clustering.
- 3) Peut-on utiliser le rapport de corrélation pour comparer deux partitions obtenues par une méthode de clustering ?
- 4) Que devient le rapport de corrélation si le nombre de clusters est égal aux nombres d'observations ?

Question 2.3.

Considérons le cas de deux variables qualitatives résumées dans un tableau de contingence.

- 1) Quel indicateur numérique permet de mesurer le lien entre les deux variables. Expliquez ce qu'il mesure.

ING1-GI : DATA EXPLORATION - RATRAPAGE 2016-2017

- 2) On étudie la consommation de viande en fonction de la situation familiale sur un échantillon de 226 personnes.
- (a) Faire une phrase pour expliquer chaque case grisée
- (b) A partir des profils lignes et/ou colonnes, étudiez le lien entre ces variables

Consommation viande

<i>Situation familiale</i>	3 fois	6 fois et plus	1 fois et moins	Tot
Couple	16,4%	32,7%	3,5%	52,7%
Famille	1,8%	2,2%	0,4%	4,4%
Seul	18,6%	21,7%	2,7%	42,9%
Tot	36,7%	56,6%	6,6%	100%

Tableau des fréquences

Consommation viande

<i>Situation familiale</i>	3 fois	6 fois et plus	1 fois et moins
Couple	31,1%	62,2%	6,7%
Famille	40,0%	50,0%	10,0%
Seul	43,3%	50,5%	6,2%

Profils lignes

Consommation viande

<i>Situation familiale</i>	3 fois	6 fois et plus	1 fois et moins
Couple	44,6%	57,8%	53,3%
Famille	4,8%	3,9%	6,7%
Seul	50,6%	38,3%	40,0%

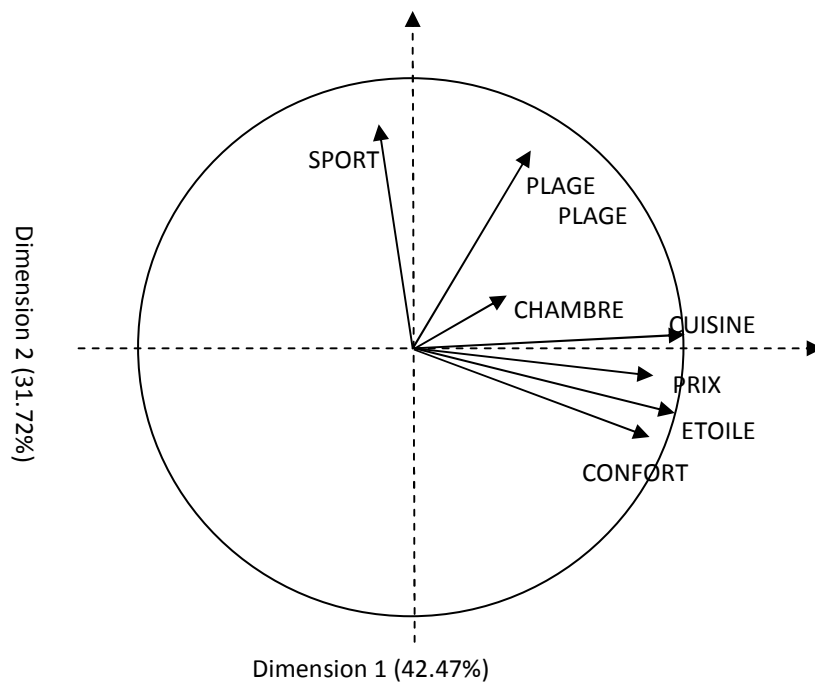
Profils colonnes

3. Exercice 3 : Analyse en composantes principales

Question 3.1

Dans cet exercice, on étudie un ensemble d'hôtels au travers 7 variables quantitatives : PLAGE, CONFORT, SPORT, CUISINE, ETOILE, PRIX, CHAMBRE. Pour cela, on effectue une ACP. Le graphique des variables se trouve ci-dessous.

- 1) Quelles sont les variables qui contribuent à la construction de l'axe 1? A l'axe 2?
- 2) Quelles sont les variables qui ne peuvent être interprétées sur ce graphique? Justifiez votre réponse.
- 3) Citez deux variables qui sont peu corrélées entre elles? Justifiez votre réponse.
- 4) Citez une variable dont le coefficient de corrélation avec la première composante principale est proche de 1? Justifiez votre réponse.
- 5) Combien vaut le coefficient de corrélation entre les deux composantes principales? Justifiez votre réponse.



Question 3.2.

On considère 3 variables X_1 , X_2 , et X_3 telles que les coefficients de corrélation sont

$$r(X_1, X_2) = 0,1$$

$$r(X_1, X_3) = 0,8$$

$$r(X_2, X_3) = -0,2$$

Quel(s) graphique(s) des variables résultant d'une ACP vous semble possible.

- | | | | | | |
|--------------------------|--|--------------------------|--|--------------------------|--|
| <input type="checkbox"/> | | <input type="checkbox"/> | | <input type="checkbox"/> | |
| <input type="checkbox"/> | | <input type="checkbox"/> | | <input type="checkbox"/> | |