

Licence d'Informatique 2

Analyse de Données Utilisateur (C5-160412)

TD 1 – Description de Données

Carl FRÉLICOT – Dpt Info / Lab MIA

1. Des Données

Un tableau de données croise des *individus* (ou *observations*, *exemples*, *instances*) et des *variables* (ou *attributs*, *primitives*). Le *type* d'une donnée est généralement celui de la variable observée. En analyse de données, soit la variable a un rôle descriptif (*feature*), soit on cherche à l'expliquer (*target*).

- Quels *types* de données/variables connaissez-vous ?
Nommez-les et donnez un exemple pour chaque type.
- Qu'est-ce qu'une variable binaire ?
- Savez-vous ce qu'est une *méta-donnée* ? Donnez un exemple.
- Connaissez-vous la différence entre des données *structurées* et des données *non structurées* ?

N°	Poids	Nombre de frères et sœurs	Situation familiale	Force	Nationalité	Sexe	Date	Heure
1	41,5	1	Célibataire	Faible	Indien	Homme	31/10/1982	10:12
2	33,4	1	Célibataire	Normal	Suisse	Femme	31/10/1983	15:14
3	37,5	2	Célibataire	Fort	Sénégalais	Femme	31/12/1983	14:48
4	33,5	1	Divorcé	Faible	Australien	Femme	01/01/1984	20:15
5	39,7	2	Divorcé	Normal	Birman	Homme	02/01/1984	06:02
6	30,8	1	Marié	Fort	Belge	Homme	31/10/2010	05:45
7	37,4	3	Marié	Très fort	Portugais	Femme	21/10/1983	14:47
8	38,2	1	Marié	Invincible	Brésilien	Homme	02/10/1956	21:14
9	43	3	Veuf	Fort	Russe	Femme	14/09/1929	14:54
10	38,5	2	Veuf	Normal	Serbe	Homme	27/08/1902	02:12

Exercice 1

Identifiez les types de variables dans le tableau ci contre.

2. Description de Données

[Stat. Desc. Unidimensionnelle]

Les outils de ce qu'on appelle la *statistique descriptive* constituent la base de l'analyse de données en ce sens qu'ils permettent de représenter et fournir des résumés synthétiques (numériques/graphiques) des séries de valeurs observées, outils adaptés à leur type. Il s'agit d'une démarche dite *exploratoire*.

- Connaissez-vous des descripteurs numériques ? Lequels, et que sont-ils censés représenter ?
- Savez-vous ce qu'on entend par *une statistique* ? Et par *une statistique d'ordre* ?
- Connaissez-vous les représentations graphiques suivantes : diagramme en bâtons (*bargraph*), diagramme en secteurs (*piechart*), histogramme (*histogram*), boîte à moustaches (*boxplot*) ?
Sont-elles spécifiques à un type de données ?
- Savez-vous ce que signifie *normaliser* des données, les *centrer*, les *réduire* ? Et à quoi cela sert ?

Exercice 2

Le *Thermal Design Power* (en Watts) de trois types de microprocesseurs *Untel* (A, B, C) a été mesuré :

	Σ										Σ^2	
type A	50,1	58,1	58,4	59,4	64,2	64,9	65,4	67,8	68,1	73,3	1 394	98 491, 86
	74,7	74,9	75,1	75,4	76	76	76,6	76,7	77,6	81,3		
type B	58	61,8	63,7	64	64,4	64,9	65,3	65,8	66,8	66,9	1 334, 4	89 249, 62
	67,5	67,5	67,8	68,5	68,7	69,1	69,5	70,4	71,8	72		
type C	58,6	61,8	62,8	63	63,1	63,3	63,7	64,2	64,4	65,3	1 376, 6	95 540, 62
	69,3	70,9	73,7	73,8	73,9	74,4	76,3	76,9	78,4	78,8		

2-1) Pour le type A, calculez : (puis B, puis C)

- la moyenne arithmétique \bar{x} et la médiane Q_2 ,
- les deux autres quartiles et l'écart-type s .

2-2) Complétez la figure ci-contre.

2-3) À chaque processeur testé, on associe l'étiquette "faible", "fort" et "moyen" selon que son *TDP* est inférieur à 65 W, supérieur à 70 W ou entre les deux.

- complétez le tableau ci-dessous, puis
- dessinez le diagramme en secteurs correspondant.

TDP	faible	moyen	fort	Σ
effectif				60

