

Séance n°4/4
15 février 2019

Jean-Charles Rigaux Méthodologie DEA Doctorat 15/02/2019 36

Liste des concepts et notions utiles...

Population statistique

Jean-Charles Rigaux Méthodologie DEA Doctorat 15/02/2019 37

Échantillon

Comment a-t-il été obtenu ?

SONDAGE

Méthodes aléatoires

Méthodes empiriques

Tirage **AVEC** remise

Tirage **SANS** remise

Jean-Charles Rigaux Méthodologie DEA Doctorat 15/02/2019 38

Échantillon représentatif

Échantillon obtenu par une méthode aléatoire

Jean Claude Riguer Méthodologie DEA Doctorat 15/02/2019 59

Traitement des données assisté par informatique



SPAD
CHIC
SPSS
Excel

15/02/2019 60

Mises en ordre des données

Jean Claude Riguer Méthodologie DEA Doctorat 15/02/2019 61

A screenshot of a Microsoft Excel spreadsheet. The spreadsheet has columns labeled 'ident.', 'nom', 'V01', 'V02', 'V03', and 'V04'. The 'ident.' column contains values 'I001', 'I002', 'I003', and 'I005'. To the right of the spreadsheet, a stick figure is holding a scale. The slide footer contains the text 'Jean-Claude Rigoir - Méthodologie - DEA 3 - Dextera' and the date '15/02/2019'.

... pour réaliser les opérations...

The slide features a stick figure holding a magnifying glass over a large orange oval. The footer includes 'Jean-Claude Rigoir - Méthodologie - DEA 3 - Dextera' and '15/02/2019'.

Analyse uni-variée

Étude pas à pas de chacune des variables :

Quantitatives

Qualitatives

The slide shows a stick figure holding a hammer. The footer contains 'Jean-Claude Rigoir - Méthodologie - DEA 3 - Dextera' and '15/02/2019'.



Mise en tableaux
Mise en graphiques

Jean-Charles Riquier Méthodologie DEIA D'Informatique 15/02/2019 65

Pour DECRIRE...
Statistique descriptive...

Jean-Charles Riquier Méthodologie DEIA D'Informatique 15/02/2019 66

Vitesse en km/h	0	1	2	3	4	5	6	7	8	Total
Effectif	3	7	20	30	87	50	12	3	0	212
Fréquences	1,4%	3,3%	9,4%	14,1%	40,6%	23,6%	5,7%	1,4%	0%	100%
Fréquences relatives	0,033	0,067	0,200	0,286	0,387	0,273	0,057	0,014	0,000	1,000

Quelle caractéristique qualifie l'ensemble des données ci-dessus ? Répondre à cette question en utilisant les termes de la statistique descriptive.

Diagramme en bâtons de la vitesse en km/h





Diagramme circulaire



Dans le tableau ci-dessus, nous présentons le détail des algorithmes de calcul de la moyenne et de la variance de la variable Vitesse en km/h. Répondre à cette question en 200.

Jean-Charles Riquier Méthodologie 15/02/2019 67

Tableau statistique des distributions d'effectifs et de fréquences d'une variable quantitative discrète

Valeurs de V05	0	1	2	3	4	5	6	7	8	Effectif total
Effectifs	3	7	29	56	87	53	12	3	0	250
Fréquences	0,012	0,028	0,116	0,224	0,348	0,212	0,048	0,012	0	1
Fréquences en pourcentage	1,20%	2,80%	11,60%	22,40%	34,80%	21,20%	4,80%	1,20%	0,00%	100%

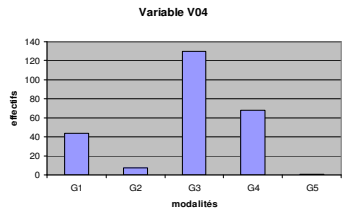
Tableaux statistiques des distributions d'effectif et de fréquence d'une variable quantitative continue

V02	[20; 25]	[25; 30]	[30; 35]	[35; 40]	[40; 45]	[45; 50]	[50; 55]	[55; 60]
effectif	12	42	47	42	50	45	9	3
fréquence	0,048	0,168	0,188	0,168	0,200	0,180	0,036	0,012

Tableaux statistiques des distributions d'effectif et de fréquence d'une variable qualitative

V04	G1	G2	G3	G4	G5
effectif	44	7	130	68	1
fréquence	0.176	0.028	0.520	0.272	0.004

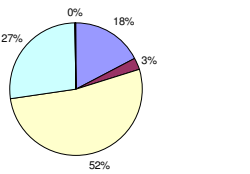
Diagramme en bâtons des distributions d'effectif et de fréquence



Jean-Claude Rigolet

15/02/2019 71

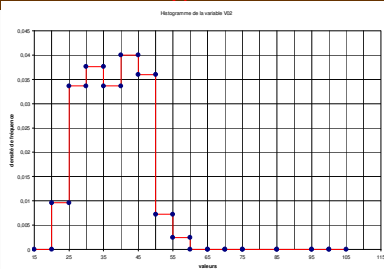
Diagramme circulaire des distributions d'effectif et de fréquence



Jean-Claude Rigolet

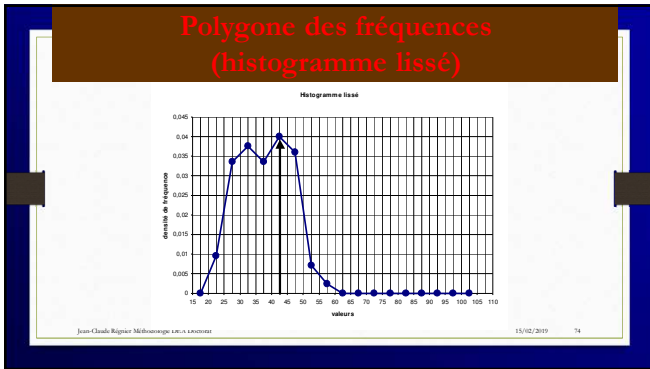
15/02/2019 72

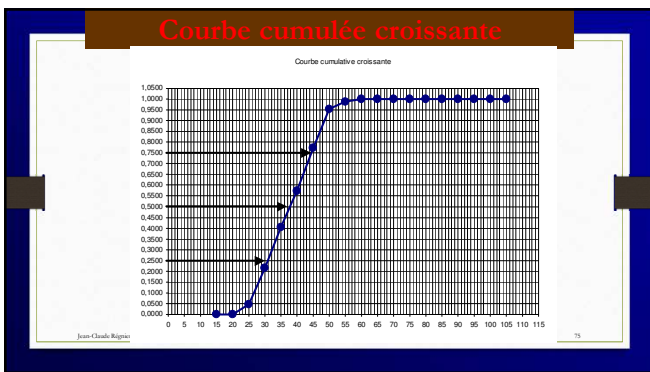
Histogramme



Jean-Claude Rigolet Méthodologie DREA Diverstet

15/02/2019 73





Analyse bi-variée

Étude conjointe de 2 variables :

- Quantitative X Quantitative
- Qualitative X Qualitative
- Qualitative X Quantitative
- Qualitative X Textuelle

Jean-Claude Riquier Méthodologie I.N.S.A. Doctorat 15/02/2019 76

Croisement de deux variables qualitatives
Tableau de contingence

TQ4A	f1	f2	m	F1	F2	
L1	15	14	13	23	25	90
L2	21	25	13	30	30	119
L3	13	12	7	5	4	41
	49	51	33	58	59	250

Jean Claude Régnier Méthodologie DEA Docteur 15/02/2019 77

Croisement de deux variables qualitatives
Tableau des fréquences

TQ4B	f1	f2	m	F1	F2	
L1	0,06	0,056	0,052	0,092	0,1	0,36
L2	0,084	0,1	0,052	0,12	0,12	0,476
L3	0,052	0,048	0,028	0,02	0,016	0,164
	0,196	0,204	0,132	0,232	0,236	1

Jean Claude Régnier Méthodologie DEA Docteur 15/02/2019 78

Croisement de deux variables qualitatives
**Tableau des fréquences conditionnelles :
 tableau des profils ligne**

TQ4C	f1	f2	m	F1	F2	
L1	0,167	0,156	0,144	0,256	0,278	1
L2	0,176	0,21	0,109	0,252	0,252	1
L3	0,317	0,293	0,171	0,122	0,098	1
	0,196	0,204	0,132	0,232	0,236	1

Jean Claude Régnier Méthodologie DEA Docteur 15/02/2019 79

Analyse multi-variée

Étude conjointe de 3 ou plus variables :

- Quantitatives**
Analyse factorielle en composantes principales
- Qualitatives**
Analyse factorielle des correspondances multiples
- Qualitatives X Quantitatives**
Analyse factorielle des tableaux mixtes

Jean Claude Rigier Méthodologie DEA Doctorat 15/02/2019 80

Tableur

The image shows two overlapping windows. The background window is Microsoft Excel 365, displaying a spreadsheet with columns A, B, and C, and rows 1 through 5. The foreground window is the 'À propos de OpenOffice.org' (About OpenOffice.org) dialog box, which includes the OpenOffice.org logo and the following text: 'OpenOffice.org 3.3.0 (000330m2) (Build#9567) Copyright © 2000, 2010 Oracle et/ou ses filiales. Tous droits réservés. Ce produit a été créé par Oracle, à partir d'OpenOffice.org. OpenOffice.org remercie tous les membres de la communauté et, en particulier, ceux mentionnés sur le site http://www.openoffice.org/welcome/credits.html.'

Jean Claude Rigier Méthodologie DEA Doctorat

COHERIS ANALYTICS SPAD

Version 9.1.22

Copyright 1982-2017 COHERIS
www.coheris.com

Java 1.8.0_201 64 bits Windows 64 bits

Jean Claude Rigier Méthodologie DEA Doctorat 15/02/2019 82

Nature des variables SPAD 9.1

Classement des Variables statistiques

The diagram shows a hierarchy where 'Variables Statistiques' branches into 'continues' and 'nominales'. A separate box labeled 'Variables textuelles' is also shown, indicating its relationship to the overall classification.

continues nominales

Variables textuelles

Jean Claude Riquier Méthodologie DCA Doctorat 15/02/2019 83

Nature des variables SPAD 9.1

de lignes à lire 10000

Nouveau nom	Stockage	Rôle
Individus	Chaîne	id Identifiant
Groupe	Chaîne	Nominale
V01_Sexe (N)	Chaîne	Nominale
AGE (C)	Réel	Continue
V03_Vous êtes-vous déjà trouvé...	Chaîne	Nominale
V04_Quand des connaissances e...	Chaîne	Nominale
V09a (T mot1)Aujourd'hui quév...	Chaîne	Textuelle
V09b (T mot2)Aujourd'hui quév...	Chaîne	Textuelle
V09c (T mot3)Aujourd'hui quév...	Chaîne	Textuelle
V09d (T phrase)Aujourd'hui qué...	Chaîne	Textuelle
[V10] Quel niveau d'intérêt perso...	Entier	Continue
[V11] Quel niveau d'utilité attrib...	Entier	Continue
[V17] Depuis la première classe ...	Chaîne	Nominale

Jean Claude Riquier 15/02/2019 84

Nature des variables SPAD 9.1

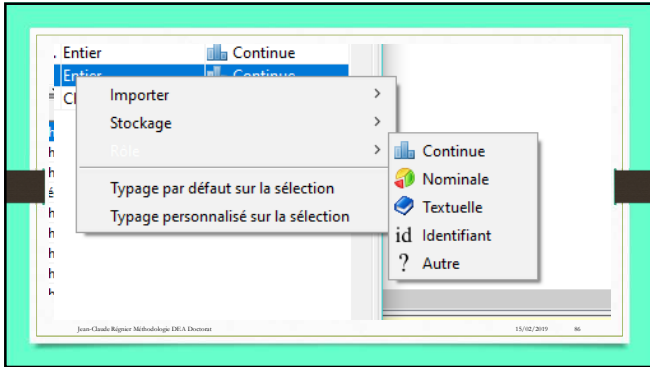
The screenshot shows a context menu with the following options:

- Importer
- Stockage
- Rôle
- Typage par défaut sur la sélection
- Typage personnalisé sur la sélection

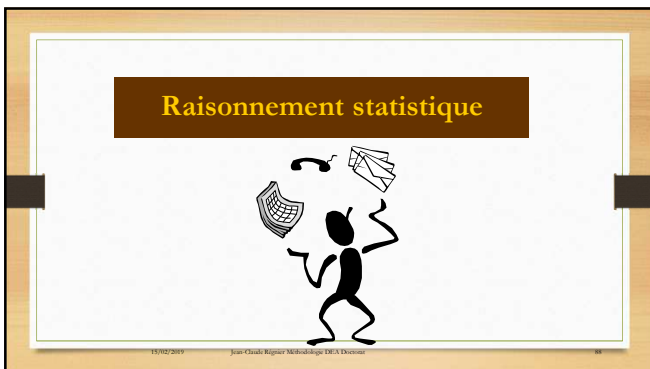
A sub-menu is open for 'Typage personnalisé sur la sélection', showing the following options:

- Chaîne
- Réel
- Entier
- Date
- Autre

Jean Claude Riquier







Estimer

Tester une hypothèse

Jean-Claude Régnier Méthodologie DEA Docteur 15/02/2019 49

Déterminisme et Aléatoire

15/02/2019 Jean-Claude Régnier Méthodologie DEA Docteur 50

Formuler des hypothèses alternatives

[H₀] les dés sont bien équilibrés *contre* **[H₁]** un dé, au moins, est truqué

[H₀] filles et garçons parviennent à des performances identiques en mathématiques *contre* **[H₁]** il existe une différence

Jean-Claude Régnier Méthodologie DEA Docteur 15/02/2019 51

Base du raisonnement statistique

		« Situation de nature » inconnue	
		Ho vraie H1 fausse	H1 vraie Ho fausse
Décision du chercheur	Rejeter Ho Conserver H1	Erreur de première espèce α	Correcte
	Conserver Ho Rejeter H1	Correcte	Erreur de seconde espèce β

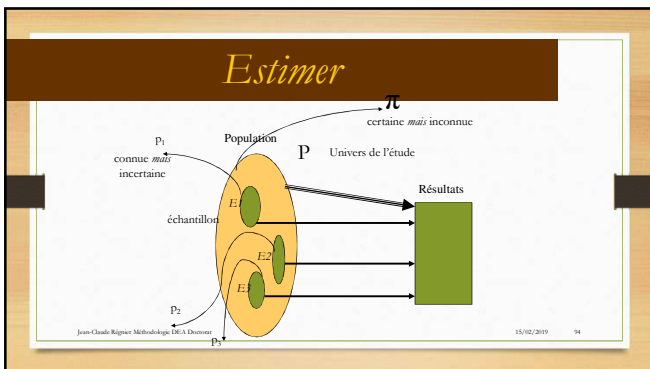
Jean-G. 12

Base du raisonnement statistique

		« Situation de nature » inconnue	
		Ho vraie H1 fausse	H1 vraie Ho fausse
Décision du chercheur	Rejeter Ho Conserver H1	Prob{Rejeter Ho sachant Ho vraie}= α	Prob{Conserver H1 sachant H1 vraie}= $1-\beta$
	Conserver Ho Rejeter H1	Prob{Conserver Ho sachant Ho vraie}= $1-\alpha$	Prob{Rejeter H1 sachant H1 vraie}= β


Jean-Claude Rippen M

Sommaire



Édition de l'arbre des méthodes

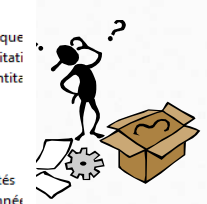
- Méthodes
 - Imports / Exports de données
 - Data Management
 - Statistiques descriptives et tests
 - Analyses factorielles
 - Classifications - Typologies
 - Modélisation statistique
 - Algorithmes de Data Mining
 - Séries chronologiques
 - Text Mining
 - Modèles structurels
 - Langages de script
 - Industrialisation - Archivage
 - Process



Jean-Claude Riquier Méthodologie DEA Doctorat 15/02/2019 95

Statistiques descriptives et tests


- Statistiques de base
 - Générateur automatique de graphique
 - Caractérisation d'une variable qualitative
 - Caractérisation d'une variable quantitative
 - Tableaux croisés
- Tests statistiques
 - Matrice Importance - Satisfaction
 - Analyse bivariée
 - Marquage sémantique des modalités
 - Construction d'une matrice de données
 - Sélection des données d'une matrice p
 - Statistiques sur questions multiples



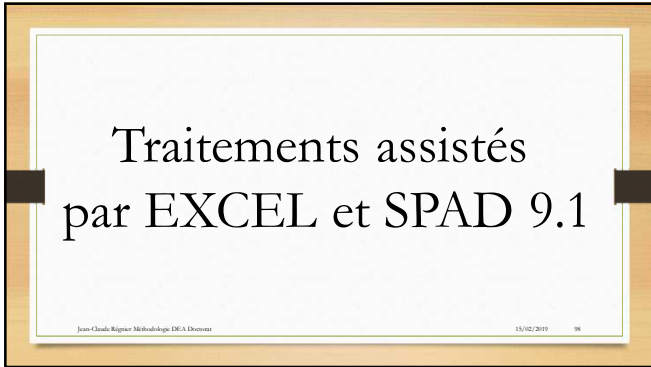
Jean-Claude Riquier Méthodologie DEA Doctorat 15/02/2019 96

Text Mining

- Détection des langues
- Construction du vocabulaire
- Nuage de mots
- Lemmatiseur semi-automatique
- Édition des contextes des mots
- Détection de segments répétés
- Création de segments répétés
- Vocabulaire caractéristique de groupe:
- Création d'un tableau lexical
- AFC sur tableau lexical
- Création d'une table mots/segments/
- Archivage des textes transformés



Jean-Claude Riquier Méthodologie DEA Doctorat 15/02/2019 97



Traitements assistés
par EXCEL et SPAD 9.1

Jean-Claude Rigier Méthodologie DEIA Dextera 15/02/2019 18
