

# Méthodes de recherche quantitatives



<http://odiv.free.fr>

## Plan de la présentation

- 1. Le recueil de données
- 2. L'inférence statistique
- 3. Les principales méthodes descriptives
- 4. Les principales méthodes explicatives

Octobre 2006

2



<http://odiv.free.fr>

## 1. Le recueil des données

Octobre 2006

3



<http://odiv.free.fr>

Il ne s'agit pas ici de présenter toutes les démarches mais simplement de les évoquer pour saisir les enjeux  
La distinction entre variables qualitatives et quantitatives ne sera faite que par la suite.

Octobre 2006

4



<http://odiv.free.fr>

## 4 éléments principaux

- Nature des données collectées
- Mode de collecte des données
- Nature du terrain d'observation et de l'échantillon
- Sources de données

Octobre 2006

5



<http://odiv.free.fr>

## Nature des données collectées (1/2)

Il convient d'identifier les informations nécessaires pour répondre à la question de recherche  
La connaissance préalable du domaine de recherche est indispensable (évident pour le questionnaire plus délicat pour les études de cas)

Octobre 2006

6



<http://odiv.free.fr>

## Nature des données collectées (2/2)

- Primaires: recueillies pour répondre à un besoin spécifique
- Secondaires: utilisation de données existantes

Avantages/inconvénients

Octobre 2006

7



## Le mode de collecte des données

Il doit permettre de saisir toutes les informations pertinentes pour répondre à la question

Questionnaire fermé, entretien ouvert, entretien semi-directif, observation, expérimentation...

Il est possible de retenir plusieurs approches

Octobre 2006

8



## Nature du terrain d'observation

Il doit être cohérent avec la problématique

Exemple: processus d'achat d'une automobile étudié auprès d'étudiants

Octobre 2006

9



## Sources de données

Déterminer la taille et la composition de l'échantillon

Octobre 2006

10



## Recueil des données

Questions à se poser. Pour la plupart, elles se recoupent.

Octobre 2006

11



## Questions sur le recueil de données (1/5)

Nature des données:

- Quelles sont les informations dont j'ai besoin pour répondre à ma question de recherche?
- Les données collectées sont-elles adaptées à la méthode d'analyse retenue?

Octobre 2006

12



## Questions sur le recueil de données (2/5)

Mode de collecte des données:

- Le mode de collecte des données est-il adapté à la question de recherche?
- Le mode de collecte des données est-il adapté aux traitements envisagés?

Octobre 2006

13



## Questions sur le recueil de données (3/5)

Nature du terrain et de l'échantillon:

- Le terrain permet-il de répondre à la question?
- La taille de l'échantillon est-elle suffisante pour appliquer la méthode de traitement choisie?
- La composition de l'échantillon pose-t-elle des problèmes en termes de validité de la recherche?

Octobre 2006

14



## Questions sur le recueil de données (4/5)

Sources de données:

- Aptitudes des interlocuteurs?
- Autres interlocuteurs?
- Mes interlocuteurs sont-ils les meilleurs?
- Est-il intéressant d'interroger d'autres interlocuteurs?
- Existe-t-il d'autres sources (le principe de triangulation)?

Octobre 2006

15



## Questions sur le recueil de données (5/5)

Faisabilité:

- Coût et durée acceptables?
- Possibilité de sous-traitance?
- Mon terrain est-il accessible?
- Le mode de recueil des données est-il acceptable pour le terrain et les sujets d'observation (forme, durée, anonymat, confidentialité...)

Octobre 2006

16



## Définition de la mesure

Règle d'attribution de nombres à des caractéristiques des objets

Un outil de collecte (de mesure) privilégié des informations quantitatives: le questionnaire.

Octobre 2006

17



Octobre 2006

18



## Les critères de qualité d'une mesure

- la fiabilité (ou fidélité)
- la validité
- la sensibilité

Octobre 2006

19



## La fiabilité

L'accroissement de la fiabilité d'un instrument de mesure passe par la réduction de la partie aléatoire de l'erreur de mesure ; ainsi, en mesurant plusieurs fois le même phénomène avec le même instrument, les résultats doivent être aussi proches que possibles.

Octobre 2006

20



## La validité

« Mesure-t-on ce qu'on cherche à mesurer? »

Elle prend plusieurs formes :

- la validité de contenu
- la validité de trait
- la validité nomologique

Octobre 2006

21



## La validité de contenu

La validité de contenu ou faciale se réfère au degré de correspondance entre les concepts et leurs indicateurs.

Elle est atteinte par les réflexions théoriques.

Octobre 2006

22



## La validité de trait (1/2)

La validité de trait ou de construit permet de s'assurer que les indicateurs des construits fournissent une bonne représentation du phénomène étudié. Il faut préciser que cette démarche est une démarche de réfutation. Il n'est pas possible de démontrer la validité absolue mais seulement la non-validité des résultats

Octobre 2006

23



## La validité de trait (2/2)

Deux vérifications possibles:

- La validité convergente est établie en montrant qu'en mesurant le même trait avec différents outils, les résultats sont similaires donc fortement corrélés. Cette dimension de la validité est proche de la fiabilité.
- En revanche, si les indicateurs sont supposés mesurer des phénomènes différents, ils doivent être faiblement corrélés entre eux car ils doivent permettre de discriminer les phénomènes entre eux ; c'est la validité discriminante.

Octobre 2006

24



## La validité nomologique

La validité nomologique ou prédictive concerne la liaison entre les concepts ; il s'agit de savoir si les relations entre les mesures d'un concept et celles d'autres concepts sont ou non en conformité avec les prédictions issues de la théorie

Octobre 2006

25



http://odiv.free.fr

## La sensibilité

La sensibilité désigne simplement la capacité d'un instrument à enregistrer des variations fines d'un phénomène. Cette caractéristique est très rarement étudiée

Octobre 2006

26



http://odiv.free.fr

## Illustration

	P3	P5	P7	P10	AMENGP	AMCOHT	ANCB
P5	-0,0112						
P7	0,2793 *	0,2475 *					
P10	0,2843 *	0,2281	0,5079 **				
AMENGP	0,3994 **	0,0200	-0,0565	-0,0542			
AMCOHT	0,0838	0,6647 **	0,1829	0,1511	0,1913		
ANCB	-0,0189	0,0270	0,3264 **	-0,2804 *	-0,0874	0,1202	
ANCDD	0,2318 *	0,0803	0,3011 **	-0,4974 **	0,2242	0,1775	0,3974 **

Triangles hétérotraits-monométhode  
Diagonale monotrait-multiméthodes  
Triangles hétérotraits-hétérométhodes

Octobre 2006

27



http://odiv.free.fr

## La démarche idéale (le paradigme de Churchill)

- Définition du domaine conceptuel (revue de littérature, avis d'experts...)
- Phase exploratoire (entretiens et pré-test)
- Phase confirmatoire

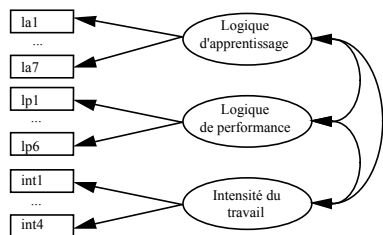
Octobre 2006

28



http://odiv.free.fr

## Exemple d'analyse confirmatoire



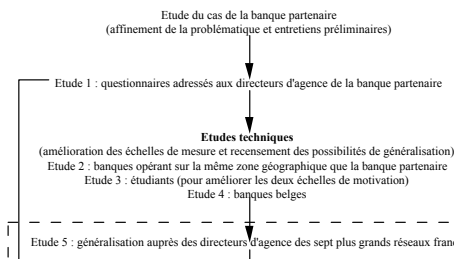
Octobre 2006

29



http://odiv.free.fr

## Une application



Octobre 2006

30



http://odiv.free.fr

## 2. L'inférence statistique

Octobre 2006

31



## Deux grandes catégories de variables

- Les variables qualitatives
- Les variables quantitatives

Octobre 2006

32



## Une hiérarchisation des variables

Type de variables	Exemples
Quantitatives	Revenu
Ordinales	Ordre de préférence
Qualitatives/nominales	CSP



Octobre 2006

33



## Décrire une variable qualitative

Un indicateur: les fréquences

Octobre 2006

34



## Décrire une variable quantitative

Des indicateurs multiples:

- La moyenne
- L'écart type – la variance
- Les quartiles – la médiane – les fractiles
- Le minimum – le maximum – l'étendue
- Les coefficients d'asymétrie et d'aplatissement
- les fréquences – le mode
- ...

Octobre 2006

35



## La théorie de l'échantillonnage

Octobre 2006

36



## Définition

Le terme de population a une signification particulière en statistique : c'est l'ensemble total d'objets ou d'individus à étudier à partir duquel est extrait un échantillon.

Octobre 2006

37



## L'inférence statistique

L'objectif est de tirer des conclusions sur les caractéristiques de la population à partir des observations faites sur l'échantillon

Octobre 2006

38



## Notations

	moyenne	écart type	proportion	effectif
Population	$\mu$	$\sigma$	$\pi$	-
Echantillon	$\bar{x}$	s ou s'	p	n

Octobre 2006

39



Pour illustrer les fondements de la théorie de l'échantillonnage, deux cas seront évoqués:

- La comparaison d'une moyenne à un standard
- Une comparaison de proportions entre plusieurs groupes

Octobre 2006

40



## Application 1 Le test de comparaison d'une moyenne à un standard

Le prix du kilo de chocolat était en moyenne équivalent à 12,50 euros en novembre 2001.

Pour évaluer l'impact du passage à l'euro, un échantillon de 10 prix est observé en mars 2002. Il fait apparaître une moyenne de 12,15 euros avec un écart type de 1,5 euros.

Avec un risque alpha de 5% peut-on considérer que le prix du chocolat a baissé?

Octobre 2006

41



## Propriété utilisée

$$\bar{x} \rightarrow St_{v=n-1} \left( \mu; \frac{s'}{\sqrt{n}} = \frac{s}{\sqrt{n-1}} \right)$$

Octobre 2006

42



## Reformulation de l'énoncé

$$n = 10$$

$$s = 1,5$$

$$\alpha = 5\%$$

$$\bar{x} = 12,15$$

Octobre 2006

43



<http://odiv.free.fr>

## Les hypothèses

$$H_0 : \mu = 12,5$$

$$H_1 : \mu < 12,5$$

Octobre 2006

44



<http://odiv.free.fr>

Si  $H_0$  est vraie

$$H_0 : \mu = 12,5$$

$$\bar{x} \rightarrow St_{\nu=9} \left( 12,5; \frac{s}{\sqrt{n-1}} = \frac{1,5}{\sqrt{9}} = 0,5 \right)$$

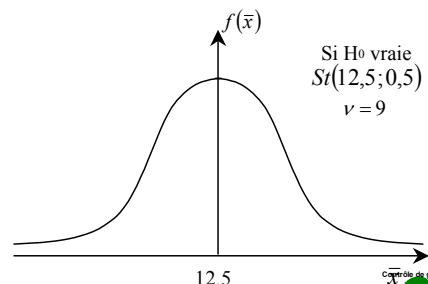
Octobre 2006

45



<http://odiv.free.fr>

## Représentation graphique



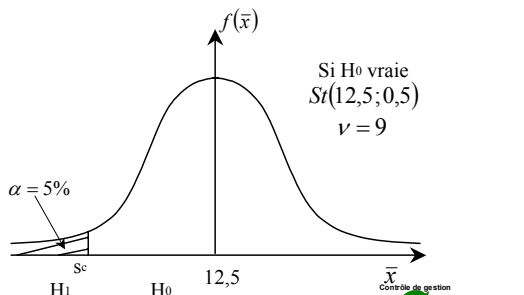
Octobre 2006

46



<http://odiv.free.fr>

## Représentation de la règle de décision



Octobre 2006

47

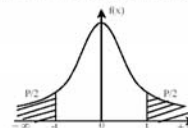


<http://odiv.free.fr>

## Table de Student

### Loi de Student

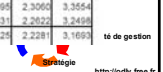
Valeurs de t ayant la probabilité P d'être dépassées en valeur absolue.



P/2	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	1%
1	0,1554	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3137	12,7062	63,6555
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3902	1,8856	2,9202	4,3027	9,9250
3	0,1309	0,2707	0,4242	0,5844	0,7649	0,9789	1,2498	1,6377	2,3534	3,1824	5,8408
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7765	4,6041
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9169	1,1558	1,4759	2,0150	2,5706	4,0321
6	0,1311	0,2648	0,4043	0,5534	0,7175	0,9057	1,1342	1,4398	1,9432	2,4469	3,7074
7	0,1303	0,2632	0,4016	0,5491	0,7111	0,8966	1,1162	1,4146	1,8948	2,3646	3,4995
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8869	1,1081	1,3968	1,8595	2,3062	3,3554
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8824	1,0997	1,3830	1,8331	2,2622	3,2498
10	0,1289	0,2602	0,3966	0,5415	0,6988	0,8791	1,0931	1,3722	1,8125	2,2281	3,1693

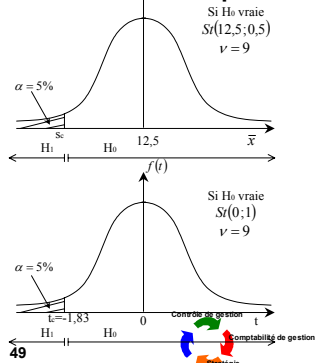
Octobre 2006

48



<http://odiv.free.fr>

## Détermination du seuil critique



Octobre 2006

## Calcul du seuil critique

$$\frac{s_c - 12,5}{0,5} = -1,83$$

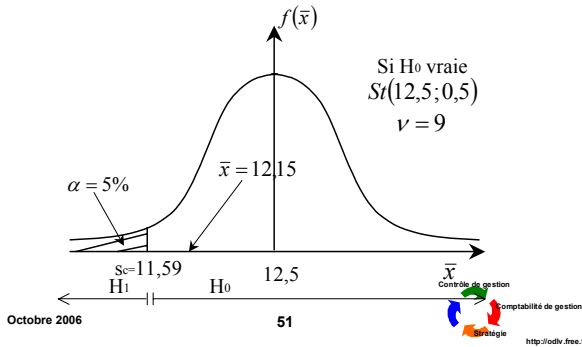
$$\Rightarrow s_c = -1,83 \times 0,5 + 12,5 = 11,59$$

Octobre 2006

50

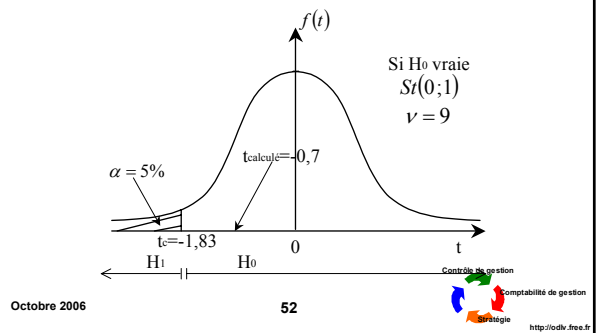


## Représentation graphique finale



Octobre 2006

## Autre représentation graphique



Octobre 2006

Malgré une baisse sensible du prix observé sur l'échantillon, pour quelle raison considère-t-on que le prix du chocolat n'a pas baissé?

Octobre 2006

53



Faisons varier les différents paramètres pour évaluer leur impact sur la décision.

Octobre 2006

54



## Situations à comparer

$\bar{x}$	s	n	t	Hyp acceptée
12,15	1,5	10	-0,70	H <sub>0</sub>
12,15	1,5	<b>200</b>		
12,15	<b>0,3</b>	10		
<b>11,05</b>	1,5	10		

Octobre 2006

55



<http://odiv.free.fr>

## Décisions

$\bar{x}$	s	n	t	Hyp acceptée
12,15	1,5	10	-0,70	H <sub>0</sub>
12,15	1,5	<b>200</b>	-3,29	H <sub>1</sub>
12,15	<b>0,3</b>	10	-3,50	H <sub>1</sub>
<b>11,05</b>	1,5	10	-2,90	H <sub>1</sub>

Octobre 2006

56



<http://odiv.free.fr>

## Application 2 le test du chi2

Les effets d'un somnifère sont testés en l'administrant à quelques patients.

Un placebo est administré à groupe témoin.

Octobre 2006

57



<http://odiv.free.fr>

## Les observations

Observations	Bien dormi	Mal dormi	Total
Somnifère	53	28	81
Placebo	40	32	72
Total	93	60	153

Octobre 2006

58



<http://odiv.free.fr>

## Les hypothèses testées

- H<sub>0</sub> : les variables sont indépendantes
- H<sub>1</sub> : les variables sont dépendantes

Octobre 2006

59



<http://odiv.free.fr>

## Si l'hypothèse H<sub>0</sub> est vraie

La proportion de personne ayant bien dormi (ou mal dormi) est la même dans nos deux groupes.

$$93/153=60,78\%$$

Octobre 2006

60



<http://odiv.free.fr>

# Le tableau théorique

Observations	Bien dormi	Mal dormi	Total
Somnifère	53	28	81
Placebo	40	32	72
Total	93	60	153

Théorique	Bien dormi	Mal dormi	Total
Somnifère	49,24	31,76	81
Placebo	43,76	28,24	72
Total	93	60	153



Pour évaluer si ces deux tableaux sont identiques, nous allons calculer la distance qui les sépare.



# Métrique utilisée

$$\sum_{i=1}^l \sum_{j=1}^c \frac{(n_{ij} - n_{ij}')^2}{n_{ij}'} \rightarrow \chi^2 \text{ avec } \nu = (l-1) \times (c-1)$$



# Le calcul de la distance

Observations	Bien dormi	Mal dormi	Total
Somnifère	53	28	81
Placebo	40	32	72
Total	93	60	153

Théorique	Bien dormi	Mal dormi	Total
Somnifère	49,24	31,76	81
Placebo	43,76	28,24	72
Total	93	60	153

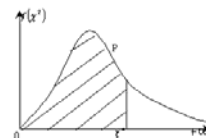
Distances	Bien dormi	Mal dormi	Total
Somnifère	0,29	0,45	0,73
Placebo	0,32	0,50	0,83
Total	0,61	0,95	1,56



Les tables statistiques vont nous permettre d'évaluer l'importance de cette distance.

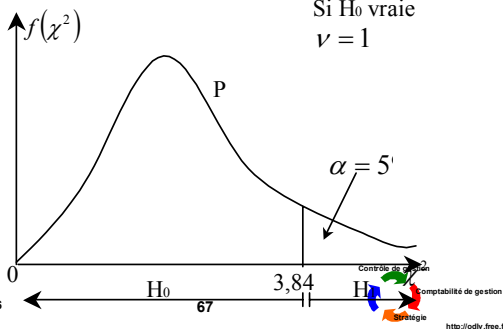
# Loi du $\chi^2$

Valeur de  $\chi^2$  ayant la probabilité P d'être dépassée.



taille P	0,5%	1,0%	2,5%	5,0%	10,0%	50,0%	90,0%	95,0%	97,5%	99,0%	99,5%
1	0,004	0,008	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,821	1,145	1,610	4,351	9,236	11,070	12,832	15,086	16,750
6	0,676	0,922	1,237	1,835	2,204	5,348	10,845	12,592	14,449	16,912	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,160	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955





Que se passe-t-il si les effectifs observés sont 10 fois plus importants?



## Observations

Observations	Bien dormi	Mal dormi	Total
Somnifère	530	280	810
Placebo	400	320	720
Total	930	600	1530



## Calcul de la distance

Observations	Bien dormi	Mal dormi	Total
Somnifère	530	280	810
Placebo	400	320	720
Total	930	600	1530

Théorique	Bien dormi	Mal dormi	Total
Somnifère	492,35	317,65	810
Placebo	437,65	282,35	720
Total	930	600	1530

Distances	Bien dormi	Mal dormi	Total
Somnifère	2,88	4,46	7,34
Placebo	3,24	5,02	8,26
Total	6,12	9,48	15,60



L'hypothèse  $H_1$  est acceptée.



## Tableau de synthèse des tests

Type de variable	2 échantillons	Plus de 2 échantillons	Mesures d'association
Nominale	Test du Chi-deux		Coefficient Phi
Ordinale	Test de Mann-Witney	Test de Kruskal-Wallis	Coefficient de corrélation des rangs
Métrique	Test de Student	Analyse de variance	Coefficient de corrélation linéaire



### 3. Les principales méthodes descriptives

- Les méthodes les plus simples (méthodes univariées)
- Les méthodes d'analyse factorielle
- Les méthodes de classification

Octobre 2006

73



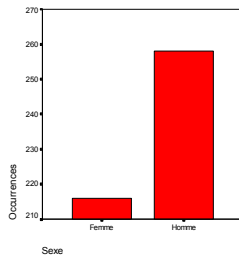
### Les méthodes les plus simples

Octobre 2006

74



### Diagramme en bâtons, secteurs...

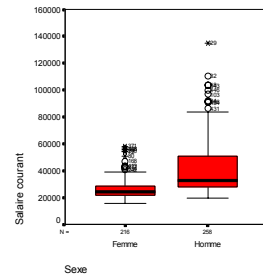


Octobre 2006

75

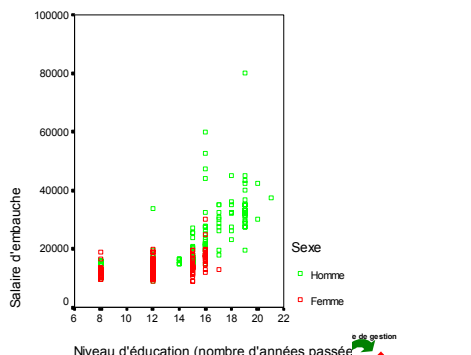


### Les boîtes à moustaches



Octobre 2006

76



Octobre 2006

77



### Les méthodes d'analyse factorielle

Il convient de distinguer les méthodes permettant de traiter les variables quantitatives (ACP) de celles permettant de traiter les variables qualitatives (AFC/AFCM).

Octobre 2006

78



## Exemple d'ACP (1/2)

Matrice des composantes après rotation<sup>a</sup>

	Composante		
	1	2	3
vendeurs agréables	.894		
vendeurs attentifs	.887		
personnel souriant	.884		
vendeurs compétents	.867		
vendeurs disponibles	.777		
vendeurs poussent à la vente	-.509		
prix attractifs		.893	
réachalandages fréquents		.891	
beaucoup de produits nouveaux		.887	
large gamme de produits		.707	
des coloris très tendance		.540	
présentation agréable			.884
magasin bien situé			.871
la musique est agréable			.802
vitrine attrayantes			.734
les produits se trouvent facilement			.599

Méthode d'extraction : Analyse en composantes principales.  
Méthode de rotation : Varimax avec normalisation de Kaiser.  
<sup>a</sup> La rotation a convergé en 4 itérations.

Octobre 2006



http://odiv.free.fr

## Exemple d'ACP (2/2)

- Axe 1 : Qualité des vendeurs
- Axe 2 : Qualité de « l'offre »
- Axe 3 : Qualité de l'environnement

Octobre 2006

80



http://odiv.free.fr

## Une application

La validation du modèle AMI

- (A) Orienter les **A**ctions et les comportements des acteurs
- (M) **M**odéliser les relations entre les ressources et les finalités
- (I) **I**nterconnecter la stratégie et le quotidien

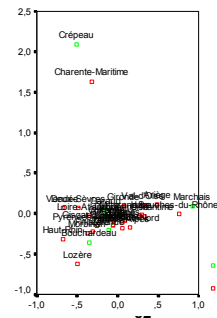
Octobre 2006

81



http://odiv.free.fr

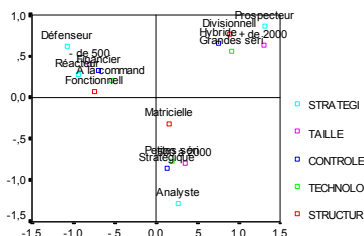
## Exemple d'AFC



Octobre 2006

http://odiv.free.fr

## Exemple d'AFCM (1/2)

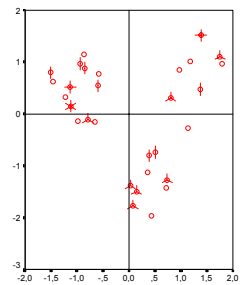


Octobre 2006



http://odiv.free.fr

## Exemple d'AFCM (2/2)



Octobre 2006

84



http://odiv.free.fr

# Les méthodes de classification

Octobre 2006

85



# Les méthodes de classification

Ces méthodes ont pour objectif de constituer des groupes homogènes (généralement d'observations mais parfois de variables).

On distingue les méthodes hiérarchiques de méthodes non hiérarchiques.

Octobre 2006

86



## Méthodes hiérarchique / non hiérarchiques (1/2)

Les méthodes hiérarchiques présentent l'intérêt de ne pas avoir à préciser a priori le nombre de groupes à constituer mais elle ne permettent pas de traiter un grand nombre d'observation (quelques centaines)

Octobre 2006

87



## Méthodes hiérarchique / non hiérarchiques (2/2)

Les méthodes non hiérarchiques présentent l'intérêt de pouvoir traiter un grand nombre d'observation (des millions) mais elles présentent l'inconvénient de devoir fixer a priori le nombre de groupes à constituer.

Octobre 2006

88



## Le principe des méthodes non hiérarchiques

Il est nécessaire de préciser a priori le nombre de groupes à constituer.

Octobre 2006

89



## Exemple

	X	Y
O1	1,0	2,0
O2	1,3	2,1
O3	-0,5	-2,0
O4	-1,0	-2,3
O5	0,0	1,0
O6	1,5	-1,4
O7	-0,9	-1,7
O8	2,0	-1,0
O9	-1,0	0,5
O10	0,9	1,6
O11	-0,8	-1,5
O12	1,4	-2,1
O13	0,4	1,3
O14	-0,6	-2,7
O15	1,6	-1,4
O16	0,7	2,2
O17	-1,0	0,5
O18	0,5	1,3
O19	-0,8	-1,6
O20	1,9	-2,0
O21	0,1	1,1
O22	-1,0	-2,4
O23	1,8	-1,7
O24	1,7	-2,0
O25	0,9	2,3
O26	0,6	0,7
O27	0,7	1,5
O28	2,2	-1,0
O29	2,0	-0,5
O30	1,5	-1,7
O31	-0,4	-1,7
O32	-0,5	-1,7

Octobre 2006

90



32 observations sont décrites par deux variables.

Nous allons constituer 3 groupes.

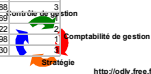
Pour cela, trois points sont sélectionnés au hasard : les trois premiers (O1, O2 et O3).

Les distances entre chaque point et ces trois points sont calculées.



## Le calcul des distances

obs	X	Y	O1	O2	O3	dist min	G
O1	1	2	0,00	0,32	4,27	0,00	1
O2	1,5	2,1	0,32	0,00	4,48	0,00	2
O3	-0,5	-2	4,27	4,48	0,00	0,00	3
O4	-1	-2,3	4,74	4,96	0,58	0,58	3
O5	0	1	1,41	1,70	3,04	1,41	1
O6	1,5	-1,4	3,44	3,51	2,09	2,09	3
O7	-0,9	-1,7	4,18	4,39	0,50	0,50	3
O8	2	-1	3,18	3,18	2,69	2,69	3
O9	-1	0,5	2,50	2,80	2,55	2,50	1
O10	0,9	1,6	0,41	0,64	3,86	0,41	1
O11	-0,8	-1,5	3,84	4,17	0,50	0,50	3
O12	1,4	-2,1	4,12	4,20	1,90	1,90	3
O13	0,4	1,3	0,92	1,20	3,42	0,92	1
O14	-0,5	-2,7	4,88	5,16	0,71	0,71	3
O15	1,6	-1,4	3,45	3,51	2,18	2,18	3
O16	0,7	2,2	0,36	0,61	4,37	0,36	1
O17	-1	0,5	2,50	2,80	2,55	2,50	1
O18	0,5	1,3	0,86	1,13	3,45	0,86	1
O19	-0,8	-1,6	4,02	4,25	0,50	0,50	3
O20	1,9	-2	4,10	4,14	2,40	2,40	3
O21	0,1	1,1	1,27	1,56	3,16	1,27	1
O22	-1	-2,4	4,83	5,05	0,64	0,64	3
O23	1,8	-1,7	3,79	3,83	2,32	2,32	3
O24	1,7	-2	4,06	4,12	2,20	2,20	3
O25	0,9	2,3	0,32	0,45	4,52	0,32	1
O26	0,5	0,7	1,39	1,61	2,88	1,39	1
O27	0,7	1,5	0,58	0,85	3,70	0,58	1
O28	2,2	-1	3,23	3,23	2,88	2,88	3
O29	2	-0,5	2,69	2,69	2,92	2,69	3
O30	1,5	2	0,50	0,22	4,47	0,22	2
O31	-0,4	0,6	1,98	2,27	2,60	1,98	1
O32	-0,5	-1,7	3,97	4,20	0,30	0,30	3



## Le cas de l'observation 10

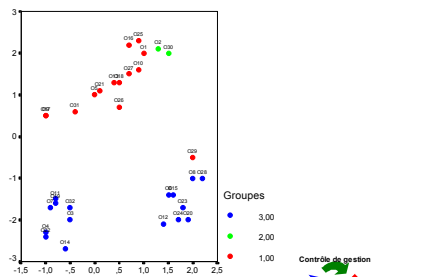
Les distances entre cette observation et les points O1, O2 et O3 sont respectivement de 0,41, 0,64 et 3,86.

Cette observation est donc rattachée au point O1.

Ce calcul est fait pour tous les points.



## Étape 1



Les centres de gravités des trois groupes sont calculés avant de recalculer les distances pour réaffecter les points à ces centres de gravité.

## Le calcul des centres de gravité

CG1	0,38	1,15
CG2	1,40	2,05
CG3	0,60	-1,36



## Étape 2

les distances sont recalculées par rapport aux centres de gravité.

Octobre 2006

97

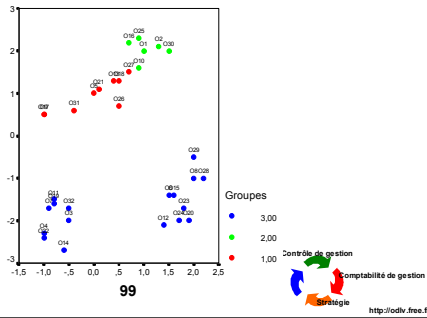


obs	X	Y	CG1	CG2	CG3	dist min	G
O1	1	2	1,05	0,40	3,38	0,40	2
O5	0	1	0,41	1,75	2,43	0,41	1
O9	-1	0,5	1,52	2,86	2,45	1,52	1
O10	0,9	1,6	0,69	0,67	2,97	0,67	2
O13	0,4	1,3	0,15	1,25	2,66	0,15	1
O16	0,7	2,2	1,10	0,72	3,56	0,72	2
O17	-1	0,5	1,52	2,86	2,45	1,52	1
O18	0,5	1,3	0,19	1,17	2,66	0,19	1
O21	0,1	1,1	0,28	1,61	2,51	0,28	1
O25	0,9	2,3	1,28	0,56	3,67	0,56	2
O26	0,5	0,7	0,47	1,62	2,06	0,47	1
O27	0,7	1,5	0,48	0,89	2,86	0,48	1
O29	2	-0,5	2,31	2,62	1,64	1,64	3
O31	-0,4	0,6	0,95	2,31	2,20	0,95	1
O2	1,3	2,1	1,32	0,11	3,53	0,11	2
O30	1,5	2	1,41	0,11	3,47	0,11	2
O3	-0,5	-2	3,27	4,47	1,27	1,27	3
O4	-1	-2,3	3,72	4,97	1,86	1,86	3
O6	1,5	-1,4	2,79	3,45	0,90	0,90	3
O7	-0,9	-1,7	3,12	4,40	1,54	1,54	3
O8	2	-1	2,69	3,11	1,44	1,44	3
O11	-0,8	-1,5	2,90	4,18	1,41	1,41	3
O12	1,4	-2,1	3,41	4,15	1,09	1,09	3
O14	-0,6	-2,7	3,97	5,15	1,80	1,80	3
O15	1,6	-1,4	2,83	3,46	1,00	1,00	3
O19	-0,8	-1,6	2,99	4,26	1,42	1,42	3
O20	1,9	-2	3,50	4,08	1,45	1,45	3
O22	-1	-2,4	3,81	5,06	1,91	1,91	3
O23	1,8	-1,7	3,18	3,77	1,25	1,25	3
O24	1,7	-2	3,42	4,06	1,27	1,27	3
O28	2,2	-1	2,82	3,15	1,64	1,64	3
O29	-0,5	-1,7	3,98	4,20	1,15	1,15	3

Octobre 2006

98

## Étape 2



Octobre 2006

99

## Étape 3

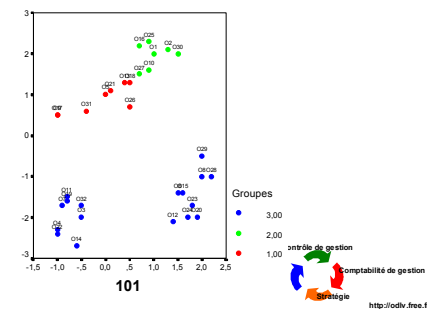
Les centres de gravité des groupes qui viennent d'être constitués sont recalculés et les points réaffectés jusqu'à ce que la solution converge (plus de changement d'une étape à l'autre).

Octobre 2006

100



## Étape 3



Octobre 2006

101

## Étape 4

Octobre 2006

102



On remarque que les résultats obtenus aux étapes 3 et 4 sont identiques. La démarche s'arrête.



La solution obtenue peut surprendre par rapport à la représentation graphique. En effet, la solution obtenue dépend des points sélectionnés au départ.

C'est la raison pour laquelle les procédures de calcul prévoient la possibilité de préciser les coordonnées des premiers centres de groupes.



## Le principe des méthodes hiérarchiques

L'intérêt de ces méthodes est qu'il n'est pas nécessaire de préciser a priori le nombre de groupes à constituer.



## Exemple

obs	X	Y
O1	1,2	-2,0
O2	1,3	2,1
O7	-0,9	-1,7
O9	2,0	-1,5
O11	-0,8	-1,5
O13	0,4	1,3
O14	-0,6	-2,7
O20	1,9	-2,0
O21	0,1	1,1
O22	-1,0	-2,4
O23	1,8	-1,7
O26	0,5	0,7
O30	1,5	2,0
O32	-0,5	-1,7



La première étape consiste à calculer la matrice des distances entre les observations. C'est la raison pour laquelle cette méthode n'est applicable que sur des fichiers comprenant peu d'observations.



## La matrice des distances entre observations

Matrice de proximité

Carré de la distance Euclidienne

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,00	16,820	4,900	,890	4,250	11,530	3,730	,460	10,820	9,000	4,90	7,780	16,090	2,980
2	16,820	0,00	19,290	13,450	17,370	1,450	26,950	17,170	2,440	25,540	14,69	2,600	0,50	17,680
3	4,900	19,290	0,00	8,450	0,50	10,690	1,090	7,930	8,840	5,00	7,290	7,720	19,450	,160
4	,890	13,450	8,450	,000	7,840	10,400	8,200	,260	10,370	9,810	,080	7,090	12,500	6,290
5	4,250	17,370	0,50	7,840	,000	9,280	1,480	7,540	7,570	8,90	6,900	6,530	17,540	,130
6	11,530	1,450	10,690	10,400	9,280	,000	17,000	13,140	,130	15,650	10,36	370	1,700	9,810
7	3,730	26,950	1,090	8,200	1,480	17,000	,000	6,740	14,930	,250	6,760	12,770	26,500	1,010
8	,460	17,170	7,930	,260	7,540	13,140	6,740	,000	12,850	8,570	,100	9,250	16,160	5,850
9	10,820	2,440	8,840	10,370	7,570	,130	14,930	12,850	,000	13,460	10,73	,320	2,770	8,200
10	9,000	25,540	5,00	9,810	,080	15,650	,250	8,570	13,460	,000	8,330	11,880	25,610	,740
11	,490	14,690	7,290	,080	6,900	10,960	6,760	,100	10,730	8,330	,000	7,450	13,780	5,290
12	7,780	2,600	7,720	7,090	6,530	,370	12,770	9,250	,320	11,880	7,450	,000	2,690	6,760
13	16,090	,050	19,450	12,500	17,540	1,700	26,500	16,160	2,770	25,610	13,78	2,690	,000	17,690
14	2,980	17,680	,160	6,290	,130	9,810	1,010	5,850	8,200	,740	5,290	6,760	17,690	,000

Ceci est une matrice de dissimilarités



Cette matrice permet d'identifier les observations qui sont les plus proches (observations 3 et 5).

Ces deux observations sont agrégées et remplacées par leur centre de gravité.

La matrice des distances sont recalculées entre ce centre de gravité et les autres observations.

Octobre 2006

109



<http://odiv.free.fr>

La démarche est poursuivie selon le même principe jusqu'à ce que toutes les observations soient agrégées en un seul groupe.

Octobre 2006

110



<http://odiv.free.fr>

## Les 13 étapes de la démarche

Chaîne des agrégations

Etape	Regroupement de classes		Coefficients	Etape d'apparition de la classe		Etape suivante
	Classe 1	Classe 2		Classe 1	Classe 2	
1	3	5	.025	0	0	5
2	2	13	.050	0	0	11
3	4	11	.090	0	0	6
4	6	9	.155	0	0	8
5	3	14	.243	1	0	10
6	4	8	.350	3	0	9
7	7	10	.475	0	0	10
8	6	12	.683	4	0	11
9	1	4	1.104	0	6	12
10	3	7	2.118	5	7	12
11	2	6	4.724	2	8	13
12	1	3	16.535	9	10	13
13	1	2	55.186	12	11	0

Octobre 2006

111



<http://odiv.free.fr>

## Interprétation

- Lors de la première étape, les observations 3 et 5 sont agrégées.
- Lors de l'étape 5, l'observation 14 est ajoutée au groupe.
- A l'étape 10, c'est l'observation 7 qui est ajoutée.

Octobre 2006

112



<http://odiv.free.fr>

Ces étapes permettent de tracer un arbre hiérarchique ou dendrogramme.

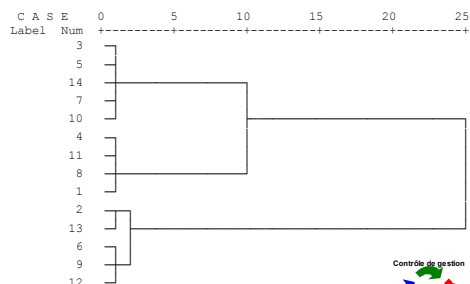
Octobre 2006

113



<http://odiv.free.fr>

## Le dendrogramme



Octobre 2006

114



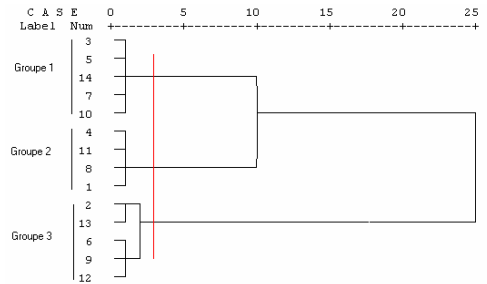
<http://odiv.free.fr>

Les branches de l'arbre (trais horizontaux) représentent les distances qui existent entre les observations.

Les observations 3 et 5 sont proches. Les observations 14 et 8 sont plus éloignées...



## Une solution : 3 groupes homogènes



La solution choisie (3 groupes) permet de remplir des objectifs contradictoires :

- Constituer peu de groupes
- Constituer des groupes homogènes



## Remarque

Les présentations des méthodes hiérarchiques et non hiérarchiques qui viennent d'être faites permettent de comprendre leurs principes.

Il faut toutefois noter la diversité des méthodes existantes principalement :

- choix des métriques c'est-à-dire manière de calculer les distance
- les méthodes d'agrégation (méthodes hiérarchiques)



## Précautions d'utilisation

- La standardisation des variables (les unités de mesure des variables ont un impact direct sur les distances calculées)
- Le nombre de variables (si le nombre de variables par dimensions n'est pas équilibré, les dimensions sur-représentées seront prépondérantes dans la constitution des groupes)



## Une solution

Pour surmonter ces deux obstacles, il est possible de faire une classification sur les résultats d'une analyse factorielle (ACP, AFC ou AFCM).



## 4. Les principales méthodes explicatives

- Les méthodes de régression
- Les méthodes dérivées de la régression
- Les modèles causaux
- Les modèles multi-niveaux

Octobre 2006

121



## Remarque : la causalité est une notion complexe

- Un phénomène peut avoir une ou plusieurs causes
- Un phénomène A est relié de façon causale à un autre B si la présence de B rend A plus probable
- Une relation de causalité ne peut jamais être prouvée mais seulement inférée
- Idéalement, il faudrait montrer que la présence du phénomène B intervient systématiquement avant celle de A (diffère de la corrélation).

Octobre 2006

122



## Les méthodes de régression

Octobre 2006

123



## Exemple

Quelles sont les variables explicatives du chiffre d'affaires d'un point de vente?

Octobre 2006

124



- Surface
- Qualité de l'emplacement
- Effectif
- Budget publicitaire
- Concurrence
- ...

Octobre 2006

125



## Formalisation

$$CA_i = \beta_0 + \beta_1 Surface_i + \beta_2 Publicité_i + \beta_3 Concurrence_i + \varepsilon_i$$

Octobre 2006

126



# Les méthodes dérivées des méthodes de régression

Avec ces méthodes, les variables expliquées sont particulières:

- Variable dichotomique (0 ou 1)
- Proportion

Octobre 2006

127



<http://odiv.free.fr>

# Applications

- Explication du choix d'une méthode comptable
- Explication de la défaillance des entreprises
- ...
- Explication du taux de retour à un mailing
- Explication de l'efficacité d'un insecticide
- ...

Octobre 2006

128



<http://odiv.free.fr>

# Les modèles causaux

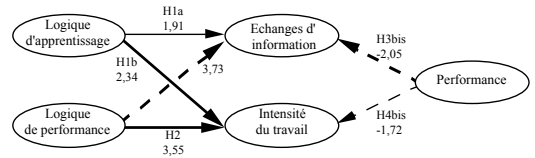
Octobre 2006

129



<http://odiv.free.fr>

# Exemple



Octobre 2006

130



<http://odiv.free.fr>

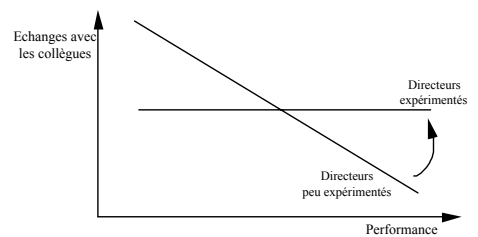
# Les modèles multi-niveaux

Octobre 2006

131



<http://odiv.free.fr>



Octobre 2006

132



<http://odiv.free.fr>

## Principales références

- Chalmers A. (1987), *Qu'est-ce que la science ?*, La Découverte.
- Evrard Y., Pras B., Roux E. (1993), *Market – Etudes et recherches en marketing*, Nathan.
- Thiétard R.-A. et coll. (1999), *Méthodes de recherche en management*, Dunod.

Octobre 2006

133



Les méthodes évoquées sont décrites dans ces ouvrages.

Des supports de cours sont disponibles.

Octobre 2006

134



## Forums pour centraliser l'information

Thèmes abordés :

- Évaluation des coûts
- ERP
- Management international
- Contrôle organisationnel
- Normes comptables IAS/IFRS (Hubert Tondeur)

Octobre 2006

135

