

Objectif

- Acquérir des bases en pratiques d'enquêtes quantitatives, statistiques descriptives et explicatives, et
- Maîtriser la pratique de l'analyse de données quantitatives, à partir de cas et d'exemple de traitement de données et de mise en forme de résultats d'études.

Bibliographie

Malhotra N.K., Décaudin J-M., Bouguerra A., Bories D. (2014), Etudes marketing avec SPSS, 6ème édition, Pearson Education

Giannelloni J.L., Verette E. (2019), Études de marché, Vuibert.

RAPPELS

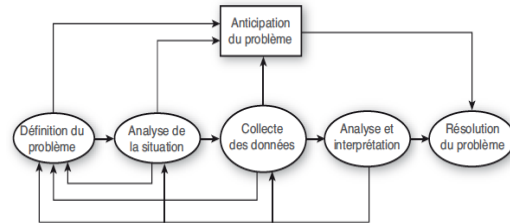
Les études marketing

- Les études et recherche marketing ont pour but d'aider le responsable marketing à résoudre un problème spécifique, à contrôler ses performances, à planifier les décisions (Evrard, Desmet, Pras et Roux, 2009)
- Prendre une meilleure décision que celle qui aurait été prise en sollicitant le hasard
- Leur objectif est de lier l'entreprise à son environnement en :
 - développant des instruments de mesure,
 - collectant et en analysant des données,
 - et en communiquant les résultats et leur interprétation



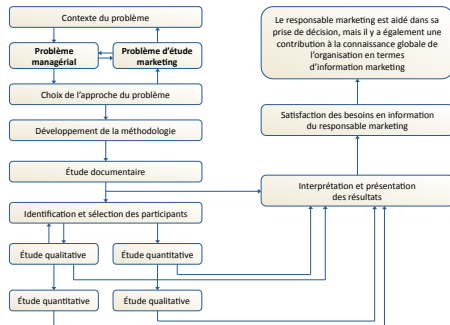
ISFIA

La démarche d'étude



ISFIA

Démarche générale



D'après Naresh K. Malotra & David F. Birks, Marketing Research

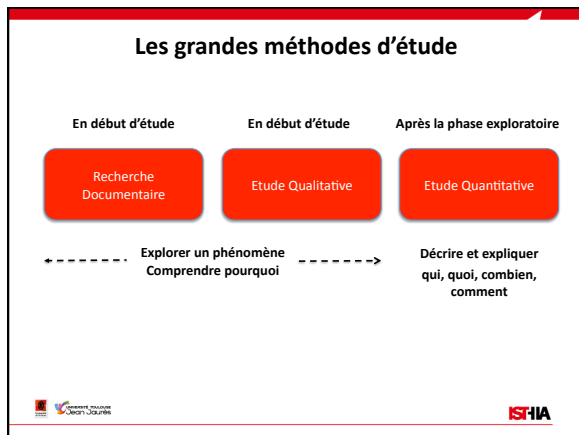
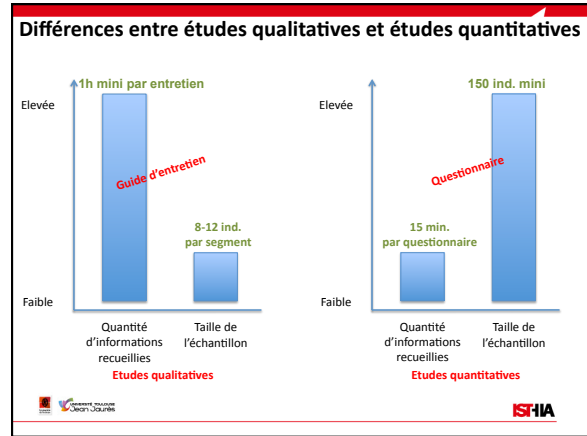
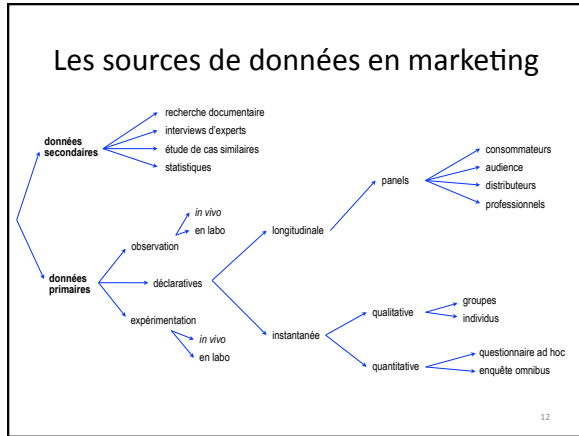
Du problème managérial à la technique d'étude

Problème managérial	Objectifs d'étude	Techniques d'étude
Existe-t-il un marché potentiel pour un nouveau produit ?	<ul style="list-style-type: none"> – Tester les réactions des consommateurs à l'idée – Tester la composition du produit – Estimer le taux d'essai et de réachat 	<ul style="list-style-type: none"> – Test de concept – Test de formule – Marché-test simulé – Marché-témoin
	– Connaître les concurrents	<ul style="list-style-type: none"> – Analyse de la concurrence – Panels
	– Connaître les attentes des consommateurs	<ul style="list-style-type: none"> – Identifier les bénéfices recherchés par les consommateurs – Étude de segmentation
	– Détecter les forces et faiblesses de la marque	<ul style="list-style-type: none"> – Étude du capital marque – Analyse des images de marque des concurrents
	– Déterminer un prix de vente	<ul style="list-style-type: none"> – Test de prix psychologiques – Analyse conjointe



adapté de Vernette, 2000.

ISFIA



Le data mining

- Ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases de données informatiques (souvent de grande taille), de façon automatique ou semi-automatique, en vue de détecter des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données.
- L'art d'extraire des informations, voire des connaissances à partir de données.

Source : adapté de Tufféry, 2005.

ISFIA

Le data mining

- Le data mining descriptif :
 - Il vise à mettre en évidence des informations présentes mais cachées par le volume des données
 - Ex : Classifications automatiques d'individus et des recherches d'associations de produits
- Le data mining prédictif :
 - Il vise à extrapoler de nouvelles informations à partir des informations présentes, ces nouvelles informations pouvant prendre la forme de classements ou scorings (sélection de clients selon certains critères), ou de prédictions comme l'appétence pour un produit (probabilité d'achat futur) ou le risque d'attrition (probabilité de départ à la concurrence).

Source : adapté de Tufféry, 2005.



L'échantillon



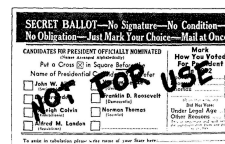
Exemple historique

- Tout commence aux États-Unis, lorsque Franklin D. Roosevelt se représente contre Alf Landon aux élections de 1936
- Deux hommes s'affrontent pour pronostiquer le résultat de ces élections :
 - Codely, rédacteur en chef du Literary Digest
 - Gallup, créateur de l'institut éponyme



Literacy Digest

- Méthode utilisée :
 - Création d'un listing de 10 millions de noms :
 - Annuaire des États-Unis
 - Liste des abonnés au magazine
 - Liste de propriétaires de voitures
 - Envoi d'un mailing
 - Retour de réponse au magazine :
 - Taux de retour : 25%



➔ Résultat du sondage :

- Landon : 57%
- Roosevelt : 43%



Gallup


- L'institut fondé par George H. Gallup se basa sur un échantillon limité à 5 000 sondés en reconstituant une population en miniature




ISHIA

Le résultat

Nominee	Franklin D. Roosevelt	Aif Landon
Party	Democratic	Republican
Home state	New York	Kansas
Running mate	John Nance Garner	Frank Knox
Electoral vote	523	8
States carried	46	2
Popular vote	27,762,648	18,681,882
Percentage	60.9%	36.5%



- Les biais de Literary Digest :
 - Le biais de sélection
 - Le biais de non réponse

ISHIA

La naissance des instituts de sondage

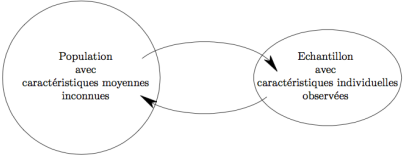
- La victoire de Gallup marque la naissance des instituts de sondage :
 - Gallup est le père de l'échantillon représentatif par quotas, le premier à avoir eu l'idée de reconstituer une population en miniature
- Deux ans après, les sondages sont importés en France par Jean Stoetzel, philosophe et sociologue, créateur en 1938 de l'Institut français d'opinion publique (IFOP).
- La méthode Gallup est construite autour de deux étapes :
 - la définition de la population à étudier et la sélection de l'échantillon

GALLUP

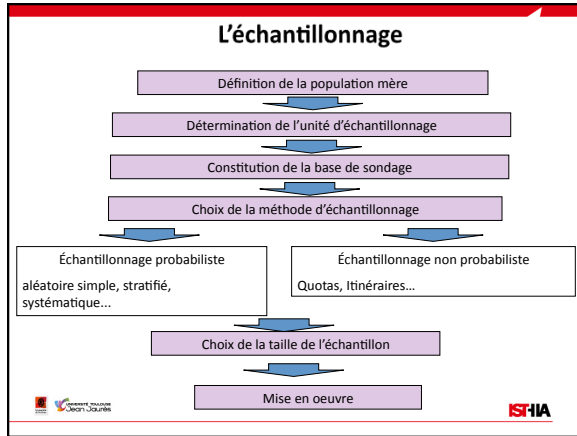
ISHIA

L'échantillonnage

- Après traitement, les résultats obtenus auprès de l'échantillon sont extrapolés à la population étudiée
- Pour assurer la fiabilité de cette extrapolation :
 - L'échantillon doit être représentatif



ISHIA



L'échantillonnage : La population mère

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : La population mère

- **Population mère (population de référence) :**
 - La population sur laquelle porte l'étude et au sein de laquelle est prélevé l'échantillon
 - Exemples :
 - Pour un sondage électoral la population mère est donc constituée des électeurs.
 - Pour une étude de satisfaction, la population mère est constituée des clients de l'enseigne de distribution
 - Ensemble des individus ou objets possédant les informations désirées pour répondre aux objectifs de l'étude
 - Ensemble de la population intéressée, a priori par le produit, la gamme de produit ou l'entreprise ou plus généralement par l'objet du sondage
- **Attention :**
 - S'il y a des erreurs dans la définition de la population-mère, elles se répercutent dans l'échantillon :
 - le résultat est alors biaisé

IST-HIA

L'échantillonnage : L'unité d'échantillonnage

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : l'unité d'échantillonnage

- **Représentation concrète des éléments de la population mère**
- **Sous ensemble de la population de base qui est interrogé après sélection lors d'une enquête**
 - Exemples :
 - Un ménage sera représenté par le chef de famille ou la personne intéressée par les achats
 - Une entreprise sera représentée par le directeur de la communication pour une étude sur l'évolution du média télévision
 - Un distributeur sera représenté par un directeur de magasin ou un chef de rayon



La base de sondage

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

La base de sondage

- Elle consiste en une **liste à partir de laquelle la sélection des éléments va être opérée**
- Pour les produits s'adressant à des niches l'utilisation de bases de données spécifiques et coûteuses est nécessaire
- 2 solutions :
 - Constituer sa base de sondage
 - Louer une base de sondage
 - Acheter une base de sondage
- Sources :
 - Annuaire téléphonique
 - Annuaire d'entreprise
 - Plan d'une ville
 - Fichier loué à FT ou EDF...



La méthode d'échantillonnage

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

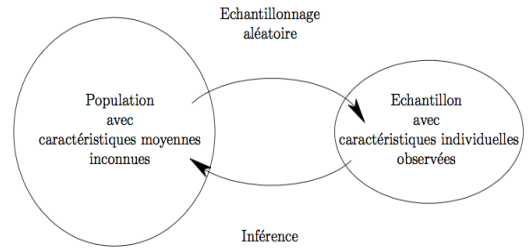
L'échantillonnage : la méthode

Méthodes aléatoires/probabilistes

Échantillon aléatoire simple
 Échantillon aléatoire systématique
 Échantillon aléatoire stratifié
 Échantillon aléatoire en grappes

Méthodes non aléatoires/empiriques

Échantillon en quotas
 Échantillon de volontariat
 Échantillon en « boule de neige »...

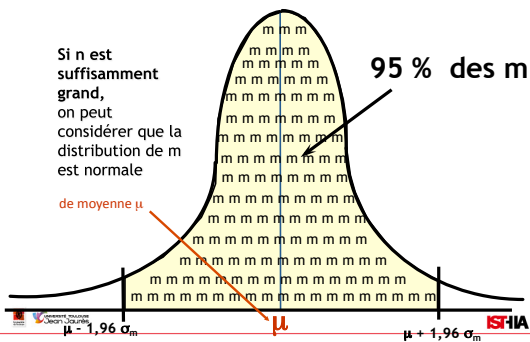


Distribution d'échantillonnage de la moyenne

Si n est suffisamment grand, on peut considérer que la distribution de m est normale

de moyenne μ

95 % des m



Méthodes d'échantillonnage probabilistes



Méthode d'échantillonnage probabiliste

- Définition
 - Les échantillons probabilistes ou aléatoires sont :
 - **constitués par tirage au sort dans la population mère**
 - pour laquelle on dispose de la liste complète de toutes les unités de sondage qui la composent (individus, familles, entreprises, etc.).
- On distingue 4 méthodes :
 - Échantillonnage aléatoire simple
 - Échantillonnage aléatoire systématique
 - Échantillonnage stratifié
 - Échantillonnage en grappes et à plusieurs degrés



IST-HIA

Echantillonnage aléatoire simple

- Principe
 - **Chaque membre d'une population a une chance égale d'être inclus à l'intérieur de l'échantillon**
 - Chaque combinaison de membres de la population a aussi une chance égale de composer l'échantillon
- Mode d'administration
 - Vous devez dresser une liste de toutes les unités incluses dans la population observée pour sélectionner un échantillon aléatoire simple
 - Un échantillonnage aléatoire simple peut s'effectuer avec ou sans remise
- Avantages :
 - Facile à mettre en œuvre
- Inconvénients :
 - Risque de non-représentativité, le coût



IST-HIA

Méthodes d'échantillonnage non probabilistes

IST-HIA
Institut Supérieur du Tourisme de l'Hôtellerie et de l'Alimentation

Méthode d'échantillonnage non-probabiliste

- Définition
 - La méthode d'échantillonnage non-probabiliste est **utilisée lorsqu'il n'est pas possible de constituer une liste exhaustive** de toutes les unités du sondage
 - Dans le cas de l'échantillonnage probabiliste, chaque unité a une chance d'être sélectionnée
 - Ce n'est plus vrai dans le cas de l'échantillonnage non-probabiliste
 - On se fixe alors comme règle que l'échantillon retenu doit avoir la même composition que la population mère par rapport à une ou plusieurs caractéristiques



IST-HIA

Echantillonnage par quotas

- **Principe :**
 - Il s'effectue jusqu'à ce qu'un nombre précis d'unités (de quotas) pour diverses sous-populations ait été sélectionné
- **Mode d'administration :**
 - Les quotas peuvent être fondés sur des proportions de la population. (par exemple 50% d'hommes et 50% de femmes)
 - Ne retenir qu'un nombre restreint de quotas. Au delà de 2 ou 3 quotas, on complique la tâche des enquêteurs
- **Avantages :**
 - L'échantillonnage par quotas est généralement moins coûteux que l'échantillonnage aléatoire
 - Il est également facile à administrer
- **Inconvénients :**
 - Certaines unités peuvent n'avoir aucune chance d'être sélectionnées



IST-HIA

Echantillonnage par quotas

- Si l'on part d'une population telle que décrite dans le tableau ci-dessous :

	Sexe	Age	Pays		
Masculin	600	Moins de 25 ans	France	250	
Féminin	400	Entre 25 et 60 ans	Espagne	200	
		Plus de 60 ans	Italie	250	
			Suisse	150	
			Belgique	150	
Total	1000	Total	1000	Total	1000

- Et que nous décidons un taux de sondage de 1/5 (20%), nous interrogerons donc 200 personnes avec la répartition suivante :

	Sexe	Age	Pays		
Masculin	120	Moins de 25 ans	France	50,00	
Féminin	80	Entre 25 et 60 ans	Espagne	40,00	
		Plus de 60 ans	Italie	50,00	
			Suisse	30,00	
			Belgique	30,00	
Total	200	Total	200	Total	200

- Le choix des individus au sein de ces échantillons se réalise de manière aléatoire



IST-HIA





TOURISME
Tourism



HÔTELLERIE
Hospitality



ALIMENTATION
Food Studies

Choix de la méthode d'échantillonnage




L'échantillonnage : choix de la méthode

- **Les objectifs de l'enquête**
 - Répétée dans le temps
 - Portée de l'enquête
- **Nature de la population étudiée**
 - Populations rares
 - Existence d'une base de sondage
- **La possibilité de calculer l'erreur d'échantillonnage :**
 - Préférence aux méthodes probabilistes
- **Le budget disponible**
- **Le mode d'administration du questionnaire**



IST-HIA

**L'échantillonnage :
La taille de
l'échantillon**

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- La représentativité de l'échantillon dépend de :
 - la méthode d'échantillonnage adoptée
- La précision des résultats obtenus dépend de :
 - la taille de l'échantillon

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- Dans le domaine probabiliste :
 - La taille de la population mère N n'a aucune incidence sur la taille de l'échantillon n
- La détermination de la taille de l'échantillon sera différente selon que nous sommes :
 - Dans un sondage non exhaustif ($n/N < 1/7$)
 - Petite taille d'échantillon par rapport à la taille de la population mère
 - Dans un sondage exhaustif ($n/N \geq 1/7$)
 - Grande taille d'échantillon par rapport à la taille de la population mère

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

**L'échantillonnage :
La taille de
l'échantillon**

Dans un sondage non
exhaustif

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- La taille de l'échantillon dans un sondage non exhaustif :
 - Si $n/N < 1/7$; $n \geq 30$; méthode d'échantillonnage probabiliste
 - Taille minimale : $n = \frac{t^2 pq}{e^2}$
 - t = coefficient dont la valeur dépend du niveau de confiance (table de la loi normale)
 - En général : niveau de confiance de 95%
 - 95 chances sur 100 pour que l'estimation soit comprise entre $p-e$ et $p+e$
 - p = fréquence observée dans l'échantillon
 - $q = 1-p$
 - e = marge d'erreur que l'on peut tolérer dans l'estimation



L'échantillonnage : taille de l'échantillon

- La taille de l'échantillon dans un sondage non exhaustif (suite) :
 - Pour les valeurs de p et de q : 2 cas
 - Une **pré-enquête** a été réalisée sur un échantillon de taille au moins égale à 30 pour estimer p
 - Ex : enquête auprès de 40 clients -> taux de satisfaction de 70% -> seuil de confiance de 95%
 - $n = \frac{1,96^2 * 0,7 * 0,3}{0,05^2} = 323$
 - Pas d'enquête préalable :
 - $p=1/2$; $q=1/2$
 - $n = 385$



L'échantillonnage : taille de l'échantillon

- La taille de l'échantillon dans un sondage non exhaustif (suite) :

Taille d'échantillon pour $p = 0,5$ et $t = 2$

Marge d'erreur	Taille minimale de n
0,02	2500
0,03	1111
0,04	625
0,05	400

En multipliant par 4 la taille de l'échantillon (ex. : 625 à 2500)



La marge d'erreur est seulement 2 fois plus petite (ex. : de 4 à 2%)



L'échantillonnage : le calcul de l'erreur

- La précision d'une estimation effectuée lors d'un sondage non exhaustif :
 - La marge d'erreur d'une estimation varie en fonction de la taille de l'échantillon mais moins qu'on ne le pense
 - $e = \pm t \times \sqrt{\frac{pq}{n}}$
 - La précision d'une estimation varie en fonction de :
 - La racine carrée de la taille



L'échantillonnage : taille de l'échantillon

Proportion observée f ₀ n (taille d'échantillon)	5 %		8 %		10 %		15 %		20 %		25 %		30 %		35 %		40 %		50 %	
	95 %	92 %	90 %	85 %	80 %	75 %	70 %	65 %	60 %	55 %	50 %	45 %	40 %	35 %	30 %	25 %	20 %	15 %	10 %	5 %
100							8	8,6	9,2	9,6	9,8	10								
150							5,7	6,4	6,9	7,3	7,6	7,8	8							
200				4,3	5,1	5,7	6,1	6,5	6,8	6,9	7,1									
250	2,8	3,4	3,8	4,5	5	5,4	5,8	6	6,2	6,3										
300	2,5	3,1	3,5	4,2	4,6	5	5,3	5,6	5,7	5,8										
350	2,3	2,9	3,2	3,8	4,2	4,6	4,9	5,1	5,2	5,3										
400	2,2	2,7	3	3,6	4	4,3	4,6	4,8	4,9	5										
500	2	2,4	2,7	3,2	3,6	3,9	4,1	4,3	4,4	4,5										
600	1,8	2,2	2,4	3	3,3	3,5	3,8	3,9	4	4,1										
700	1,7	2,1	2,3	2,7	3	3,3	3,5	3,5	3,7	3,8										
800	1,5	1,9	2,1	2,5	2,8	3	3,2	3,3	3,4	3,5										
900	1,5	1,8	2	2,4	2,7	2,9	3	3,1	3,2	3,3										
1 000	1,4	1,7	1,8	2,3	2,5	2,7	2,9	3	3	3,1										
1 500	1,2	1,4	1,5	1,9	2,1	2,3	2,4	2,5	2,6	2,6										
2 000	1	1,2	1,3	1,6	1,8	2	2,1	2,2	2,2	2,3										
3 000	0,8	1	1,1	1,3	1,4	1,5	1,6	1,7	1,8	1,8										
5 000	0,6	0,8	0,8	1	1,1	1,2	1,3	1,4	1,4	1,4										
10 000	0,4	0,5	0,6	0,7	0,8	0,9	0,9	1	1	1										

PRÉCISION DE L'ESTIMATION D'UNE PROPORTION
CALCULÉE À PARTIR D'UN ÉCHANTILLON

IST-HIA

Lire les résultats d'une enquête

Exemple :
Baromètre de consommation
et de perception des produits
biologiques en France



UNIVERSITÉ SCOLAIRE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
et de l'Hôtellerie et de l'Alimentation

Méthodologie

- L'étude menée en 2009 est la 7^{ème} vague du baromètre.
- Etude quantitative, réalisée en face à face à domicile, du 26 au 30 octobre 2009.
- Recueil des données effectué à l'aide du système WAPI (Wireless Assisted Personal Interview).
- Echantillon de 1015 personnes, représentatif de la population française (en terme de sexe, âge - 15 ans et plus -, catégorie socio-professionnelle du chef de famille, région et taille d'agglomération).
- Questionnaire d'une vingtaine de minutes.

Une partie commune posée
à tous les interviewés
base : population française
questionnaire de 10 minutes

Une partie posée uniquement aux
consommateurs-acheteurs*
base : consommateur de produits bio
questionnaire de 10 minutes

- Tests de significativité à partir de 95%
- * Question utilisée pour distinguer les consommateurs et acheteurs de produits biologiques : « Consommez-vous des produits biologiques ? »
(SI « au moins une fois par mois » - consommateur - SI au moins 1 achat au cours des 4 dernières semaines)

UNIVERSITÉ SCOLAIRE
Jean Jaurès

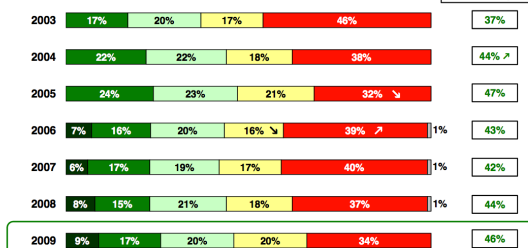
CSA i'bio

Consommation de produits biologiques

Consommez-vous des produits biologiques ?

Base : 1015

Consommateurs au
moins une fois par mois



■ Tous les jours ■ Au moins une fois par semaine ■ Environ une fois par mois ■ Moins d'une fois par mois ■ Jamais ■ Group

Evolution significative entre les vagues



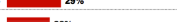
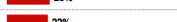
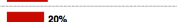




UNIVERSITÉ SCOLAIRE
Jean Jaurès

IST-HIA

Raisons de non-achat de produits biologiques

Je vais vous offrir un certain nombre de propositions. Pour chacune d'entre elles, pouvez-vous me dire si, oui ou non, c'est une raison pour laquelle vous n'avez pas acheté récemment ou que vous n'achetez jamais de produits biologiques ?

Base : n'a pas acheté de produits biologiques ces 4 dernières semaines ou n'en achète jamais = 80%

	2008 b=421	2007 b=504	2006 b=403	2005 b=469	2004 b=573
prix des produits biologiques trop élevé 	79%	75%	78%	78%	74%
pas le réflexe d'en acheter 	64%	68%	69%	65%	59%
manque d'information sur les spécificités 	29%	31%	34% ↘	41% ↗	38%
pas intéressé par l'agriculture 	23%	22%	23%	22%	17%
pas confiance dans les produits biologiques du commerce 	22%	22%	21% ↘	28%	26%
ne les trouve pas facilement dans le magasin habituel 	20%	20%	/	/	/
qualité pas satisfaisante 	16%	14%	14% ↘	19%	20%
pas intéressé par l'alimentation en général 	15%	16%	16% ↘	23%	11%
magasin habituel n'en vend pas 	11%	12%	13% ↘	19%	18%

↗ Evolution significative entre les vagues



IST-HIA



TOURISME
Tourism



HÔTELLERIE
Hospitality



ALIMENTATION
Food Studies

L'échantillonnage : La taille de l'échantillon

Dans un sondage exhaustif



IST-HIA
Institut Supérieur du Tourisme
et de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- La taille de l'échantillon dans un sondage exhaustif ($n/N \geq \frac{1}{17}$) :
 - Dans ce cas, la taille de la population (petite p/r à celle de l'échantillon) aura une incidence sur la taille minimale de l'échantillon car on se rapproche d'un recensement
 - Ex. : enquête auprès d'entreprises de transport routier du Sud Ouest
 - 620 entreprises ; on se contente d'une estimation à 3% près
 - Pour un sondage non exhaustif, il faudrait interroger 1000 pers...?
 - Méthode :
 - On calcule la taille de l'échantillon comme dans le cas d'un échantillon non exhaustif (ex. : $p=q=1/2$; $t=2$; $e=3\%$) $\rightarrow n = 1000$ mais $N = 620$
 - On calcule la taille minimale d'un sondage exhaustif :

$$n' = \frac{nN}{n + N} = \frac{1000 \cdot 620}{1000 + 620} = 383 \text{ chefs d'entreprises}$$



IST-HIA

L'échantillonnage : le calcul de l'erreur

- Dans le cas des méthodes empiriques :
 - Il faut tenir compte du biais d'échantillonnage
 - Appliquer la correction de Kish au calcul de e
- $$e = \pm t \times \sqrt{\frac{pq}{n}} \times 1,4$$
- Exemple :
 - Échantillon par quotas de 650 pers. ;
 - Seuil de confiance de 95%
 - \rightarrow marge d'erreur de 5,6% au lieu de 4%



IST-HIA

Le choix de la taille d'échantillon

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- La détermination de la taille d'échantillon dépend de deux facteurs antagonistes :
 - La précision souhaitée
 - Le budget disponible

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

L'échantillonnage : taille de l'échantillon

- Plus la taille est importante et plus les résultats sont précis.
- Plus la taille est importante et plus le coût est élevé
- Dans l'idéal, la taille de l'échantillon doit être celle qui permet d'atteindre le meilleur équilibre entre :
 - Le risque de commettre des erreurs d'échantillonnage
 - Le coût de l'échantillonnage

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

Les études quantitatives

Le questionnaire

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

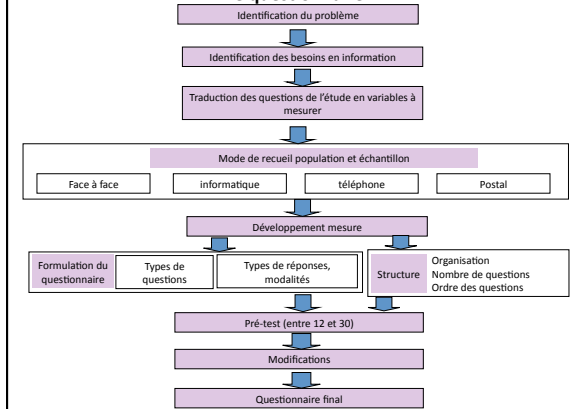
Le questionnaire

- Qu'est-ce qu'un questionnaire ?
 - « Un ensemble formalisé de questions destiné à obtenir l'information des répondants »
- Objectifs :
 - Traduire le besoin d'information en questions auxquelles le répondant peut et veut répondre
 - Encourager et motiver le répondant à coopérer et à s'impliquer dans l'enquête
 - Minimiser l'erreur dans les réponses



ISFIA

Le questionnaire



Le questionnaire : la rédaction

- Lister toutes les informations à rechercher :
 - Sur les besoins de la clientèle,
 - les caractéristiques idéales du produit,
 - les informations sur le prix auquel les interviewés accepteraient de payer le produit ou service,
 - sur les facteurs qui pourraient conduire les clients à venir acheter dans telle entreprise plutôt qu'une autre



ISFIA

Le questionnaire : la dynamique du questionnaire

- Le questionnaire n'est pas une simple série de questions
- C'est un instrument de mesure à construire :
 - de façon logique et cohérente
 - en veillant à mettre à l'aise et à intéresser
- La progressivité des questions : technique de l'entonnoir
 - Commencer par des questions qui ne demandent aucun effort de réflexion
 - Prévoir quelques grandes parties et ménager des transitions pour éveiller l'intérêt : « Nous allons commencer par... Maintenant nous allons parler de... Nous allons aborder... »



ISFIA

Le questionnaire : l'organisation

L'introduction

- Présenter les buts de l'étude (de façon vague toutefois)
- Remercier le répondant
- Préciser qu'il n'y a ni bonnes ni mauvaises réponses
- Préciser que les réponses sont anonymes

Le corps du questionnaire

- Commencer par des questions faciles
- Aller du plus général au particulier (méthode de l'entonnoir)
- Regrouper les questions par thèmes
- Introduire des questions filtres: **ne pas transformer le questionnaire en labyrinthe**

La fiche signalétique - En fin de questionnaire



Introduction questionnaire - Exemple

- Bonjour,

Vous avez été choisi pour composer un échantillon réduit mais représentatif des clients de X

Ce questionnaire porte sur la qualité de votre relation avec X

Certaines questions peuvent vous paraître un peu étranges ou encore répétitives mais veuillez essayer d'y répondre aussi spontanément que possible

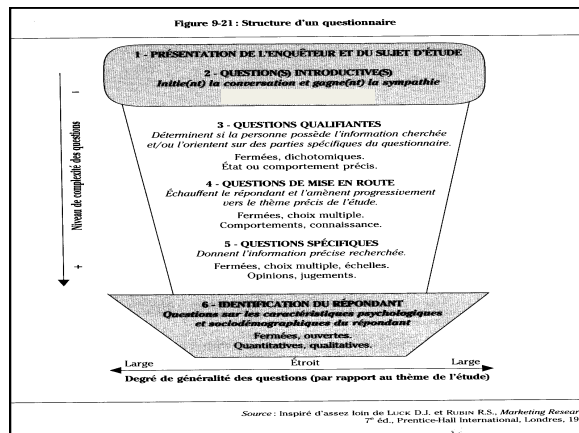
Il n'y a ni bonne, ni mauvaise réponse, seul votre avis compte

De plus, vos réponses resteront anonymes

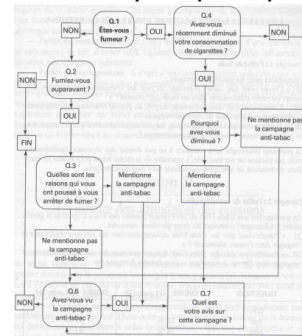
Vous pouvez donc répondre en toute franchise



Figure 9-21: Structure d'un questionnaire



Le questionnaire : la dynamique du questionnaire



Le questionnaire : le pré-test

Toujours pré-tester un questionnaire

Le questionnaire

Les biais à éviter

UNIVERSITÉ POLYTECHNIQUE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

Le questionnaire : les biais à éviter

- Ne pas introduire des mots, des expressions qui orienteront les réponses des interlocuteurs
- Les questions posées doivent être claires, précises et compréhensibles
- Les questions ne doivent pas demander trop d'efforts trop longs et trop soutenus

Le questionnaire : les biais à éviter

- Les effets d'interaction entre les questions
 - L'effet de halo
 - L'effet de contamination
 - L'effet de recoupement
- Les biais de réponse
 - La réticence à répondre
 - Les questions combinées
 - Les questions longues
 - Les questions vagues
 - Les questions techniques
 - Les biais de suggestion
 - Les biais de prestige
 - Les biais de défense

Le questionnaire : les biais à éviter

- **L'effet de halo :**

- La personne peut avoir tendance à répondre de la même manière
- Elle aligne l'opinion sur une opinion moyenne

Pour vous, le développement durable, c'est un thème...

	Pas du tout d'accord				Inconnu				Tout à fait d'accord			
8. important.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. intéressant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. passionnant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. étrange.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. fascinant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. nécessaire.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. qui a de la signification pour vous.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. ... vous concerne.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. ... a de la valeur pour vous.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. ... vous implique.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Les réponses données risquent de ne pas être discriminantes
- Par paresse ou lassitude, le répondant va répondre de la même façon en fonction de son opinion générale sur l'évolution des prix
 - **Solution : changer le sens des échelles**



Le questionnaire : les biais à éviter

- **L'effet de contamination :**

- Influence directe d'une question sur les suivantes
- Le fait d'avoir déjà répondu à une question peut influencer les réponses aux questions suivantes
- Ex : Questionnaires de satisfaction client
 - Question concernant la satisfaction globale du client en début ou fin de questionnaire ?



Le questionnaire : les biais à éviter

- **L'effet de recouvrement :**

- Contamination que l'on peut qualifier de « maîtrisée » (volontaire de la part du chargé d'étude)
- On pose la même question sous plusieurs formes à des stades différents du questionnaire
- Vérifier si les réponses sont cohérentes



Le questionnaire : les biais à éviter

- **Les effets d'interaction entre les questions**

- L'effet de halo
- L'effet de contamination
- L'effet de recouvrement

- **Les biais de réponse**


- La réticence à répondre
- Les questions combinées
- Les questions longues
- Les questions vagues
- Les questions techniques
- Les biais de suggestion
- Les biais de prestige
- Les biais de défense



Le questionnaire : les biais à éviter

- **La réticence à répondre : « Erreur systématique d'évaluation de la réponse »**

1. Est-ce que le répondant est informé ?
2. Usage, familiarité, expérience passée ?
3. Le répondant peut-il se souvenir ?





X

✓

Combien de cannettes de jus de fruit avez-vous consommé ces dernières semaines ?

A quelle fréquence consommez-vous une cannette de jus ?


1. moins d'une fois par semaine
2. de 1 à 3 fois par semaine
3. de 4 à 6 fois par semaine
4. plus de 6 fois par semaine

Le questionnaire : les biais à éviter


- **Question combinée :**
 - Le répondant ne répond pas à la bonne question
 - Une ou plusieurs questions



"Pensez-vous que Coca-Cola est une boisson douce qui rafraîchit ?"



Question double ! Préférez deux questions :

"Pensez-vous que Coca-Cola est une boisson douce ?" ET
"Pensez-vous que Coca-Cola est une boisson rafraîchissante ?"





Le questionnaire : les biais à éviter

- **Questions longues**
 - « Pensez-vous que les concessionnaires automobiles sont des commerçants indépendants comme les marchands de meubles qui sont propriétaires de leur magasin ou que ce sont des employés des firmes automobiles ? »
- **Questions vagues**
 - « Ecoutez-vous régulièrement la radio ? »
- **Questions techniques**
 - « Connaissez-vous la lessive atomisée ? »
 - « Pensez-vous que le processus de déshydratation des potages diminue leur contenu en calories ? »

X

X

X






Le questionnaire : les biais à éviter

- **Les biais de suggestion : influence de l'enquêteur**
 - Devrions-nous acheter des voitures importées puisque cela crée du chômage en France ?
 - Oui
 - non
 - Sans avis
- **Les biais de prestige**
 - Le répondant peut tenter de masquer une part de la vérité ou de mettre en exergue les déterminants valorisants
 - « Diriez-vous que votre entreprise a une bonne réputation ? »

X

X

Le questionnaire : les biais à éviter

• Les biais de défense

- « Quel est votre âge » ➔ demander la date
- « Quels sont vos revenus » ➔ proposer des classes



Le questionnaire : les biais à éviter

- S'assurer que les questions ne sont pas biaisées (incluant la réponse).
- Poser des questions simples :
 - une question à la fois.
- Poser des questions précises.
- Éviter le jargon et les termes familiers ou techniques.
- Éviter les mots compliqués.
- Éviter les termes vagues ou ambigus (« souvent »).
- Éviter les structures de questions trop complexes (double négation).
- Proposer des plages de réponses pour les sujets sensibles.
- S'assurer que les réponses proposées sont mutuellement exclusives :
 - a. 1 heure et moins
 - b. 1-3 heures
 - c. 3 heures et plus➔ Remplacer cette échelle par :
 - a. moins d'1 heure
 - b. de 1 heure à 3 heures (inclus)
 - c. plus de 3 heures
- Proposer une réponse « autres ».

Le questionnaire
Le format des questions

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ POLYTECHNIQUE
JEAN JAURÈS

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Alimentation

Le questionnaire : Le format des questions

- Les questions ouvertes
- Les questions fermées
 - Questions dichotomiques
 - Questions à choix multiples et à réponse unique
 - Questions à choix multiples et à réponses multiples
- Les questions à catégories spécifiques
 - L'échelle de Likert
 - L'échelle d'Osgood (différenciateur sémantique)
 - Les échelles Icones

Le questionnaire : Le format des questions

• Les questions ouvertes

- On distingue deux types de question ouverte :
 - celle qui amène une réponse sous la forme de chiffres
 - celle qui amène une réponse sous la forme d'un discours

Combien d'enfants à charge avez-vous ? _____

Quelles marques de yaourts connaissez-vous ?



ISFIA

Le questionnaire : Le format des questions

• Les questions fermées

– Questions dichotomiques

Sexe : H F

Consommez-vous des produits Danone ? Oui Non

– Questions à choix multiples et à réponse unique

Quelle est votre situation familiale ?

Célibataire

Union libre

Marié

Divorcé, séparé, veuf



ISFIA

Le questionnaire : Le format des questions

• Les questions fermées

– Questions à choix multiples et à réponses multiples

A quel(s) repas consommez-vous le plus fréquemment des produits d'agriculture biologique ?

Au petit déjeuner

Au déjeuner

Au dîner

Plutôt entre les repas



ISFIA

Le questionnaire : Le format des questions

• Les échelles de rang (ou ordonnées)

- On demande à l'enquêté de hiérarchiser ses réponses

- Classez par ordre de préférences vos marques de fromages préférées ?



ISFIA

Le questionnaire : Le format des questions

- Les questions à catégories spécifiques
 - L'échelle de Likert
 - Attribuez une note de 1 à 5 à chacune des affirmations suivantes
 - La modalité centrale est essentielle (sinon mesure ordinale)**

Pour moi, protéger l'environnement consiste à :

	Pas du tout d'accord	Plutôt pas d'accord	Ni d'accord, Ni pas d'accord	Plutôt d'accord	Tout à fait d'accord
Utiliser des lessives sans phosphate	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Trier mes déchets chez moi	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

jamais – occasionnellement – souvent - presque toujours - toujours

Le questionnaire : Le format des questions

- L'échelle d'Osgood (différenciateur sémantique)
 - Positionner à deux extrémités d'une graduation deux antonymes
 - Ceux-ci peuvent être remplacés par les formulations plus complètes
 - Ex : « qui appartient au passé » - « résolution moderne »

Pour moi, personnellement, l'entreprise X est :

	+2	+1	0	-1	-2	
Agréable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Désagréable
Sympathique	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Antipathique

Le questionnaire : Le format des questions

- Les échelles Icones

Quel est votre satisfaction à propose de la marque de viande X ?

1

2

3

4

5

Nom	Description	Exemple
A. Questions fermées		
Dichotomique	Deux réponses possibles.	Longue vous avez organisé ce voyage avec vous contacté la compagnie aérienne SAS ? Oui Non
Choix multiple	Trois réponses possibles ou davantage.	Vous avez effectué ce voyage avec : <input type="checkbox"/> aucune autre personne <input type="checkbox"/> votre conjoint <input type="checkbox"/> vos enfants <input type="checkbox"/> un ou plusieurs collègues <input type="checkbox"/> un service organisé <input type="checkbox"/> autres
Echelle de Likert	Une phrase par rapport à laquelle le répondant exprime son degré d'accord ou de désaccord.	Les petites compagnies aériennes offrent un général un meilleur service que les grandes compagnies aériennes. <input type="checkbox"/> Pas du tout d'accord <input type="checkbox"/> Plutôt pas d'accord <input type="checkbox"/> Ni d'accord ni pas d'accord <input type="checkbox"/> Plutôt d'accord <input type="checkbox"/> Tout à fait d'accord
Echelle sémantique différentielle	Une échelle contenant deux mots opposés. Le répondant indique le point correspondant à ses opinions.	Selon vous, SAS est-elle une compagnie aérienne de grande ou de petite taille ? Garde ————— Piste
Echelle d'importance	Une échelle mesurant l'importance d'un critère.	Selon vous, est-il important de pouvoir choisir le vin servi avec le plat ou repas ? <input type="checkbox"/> Pas du tout important <input type="checkbox"/> Peu important <input type="checkbox"/> Moyennement important <input type="checkbox"/> Assez important <input type="checkbox"/> Très important
Echelle de jugement	Une échelle évaluant un produit ou un service sur un critère.	Selon vous, les repas servis à bord des lignes SAS sont-ils de bonne ou mauvaise qualité ? <input type="checkbox"/> Très mauvaise <input type="checkbox"/> Plutôt mauvaise <input type="checkbox"/> Ni bonne ni mauvaise <input type="checkbox"/> Assez bonne <input type="checkbox"/> Très bonne
Echelle d'intention d'achat	Une échelle demandant l'intention d'achat du répondant.	Si l'on vous proposait un service d'appel téléphonique à bord d'un avion, savez-vous l'acheter ? <input type="checkbox"/> Certainement pas <input type="checkbox"/> Probablement pas <input type="checkbox"/> Un peu pas <input type="checkbox"/> Probablement <input type="checkbox"/> Certainement

Le questionnaire
Les types de mesures
&
Les traitements associés

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

Le questionnaire : les types de mesures

- Qu'est ce que mesurer ?
 - Établir une relation entre l'objet étudié et un symbole le représentant
 - Les symboles utilisés seront le plus souvent des chiffres
 - Il s'agit de mettre en correspondance l'univers réel sur lequel porte l'étude avec un système symbolique représenté par des chiffres
 - La mesure peut être définie comme « les règles d'attribution de nombres à des caractéristiques des objets »
 - L'opérationnalisation : correspondance entre niveau théorique et niveau empirique

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA

Le questionnaire : les types de mesures

K catégories

Ordonnées

Non → **Nominales** Catégories que l'on nomme avec un nom. Exemple : homme ou femme, le nom de la voiture, une couleur

Oui → Intervalles égaux entre catégories

Non → **Ordinales** Les variables ordinales sont des catégories qui sont naturellement ordonnées. Un même intervalle de "1" n'a pas toujours la même signification. Ex : Le classement à une course. Arriver 1er ou 2ème à une course n'a pas la même signification qu'arriver 1526 ou 1527ème.

Oui → Y-a-t-il un zéro naturel ?

Non → **Intervalle** L'intervalle entre deux catégories à toujours la même valeur, la même signification. La différence entre 10°C et 11°C est identique à celle entre 20°C et 21°C. Le zéro ne correspond pas l'absence de l'élément. A 0°C, il y a toujours une température.

Oui → **Mesures Métriques**

Ratio Variables d'intervalles avec un zéro naturel. Exemple pour la durée d'un test, à 0, il n'y a pas de temps.

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA

Le questionnaire : les types de mesures

K catégories

Ordonnées

Non → **Nominale** Statistiques descriptives possibles

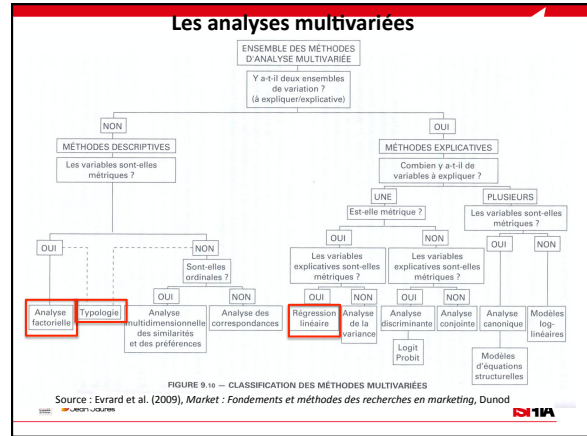
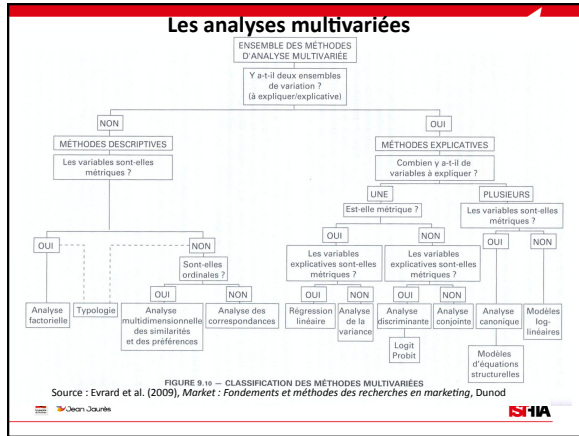
Oui → Intervalles égaux entre catégories

Non → **Ordinale** Compter => Calcul d'effectifs

Oui → **Métrique** Compter => Calcul d'effectifs
Moyenne, Variance, Ecart-type => sous SPSS :
« Descriptives »

UNIVERSITÉ TOULOUSE
Jean Jaurès

IST-HIA





TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

Stratégies d'analyse

IST-HIA
Institut Supérieur du Tourisme et de l'Alimentation

Stratégies d'analyse

- 1 **Univariée** : Tris à plat/description de l'échantillon/recodage
- 2 **Bivariée** : Tris croisés/mesures d'association
- 3 **Multivariée** : Descriptive : par groupes de variables
Explicative : entre groupes

IST-HIA

Stratégies d'analyse

- L'analyse univariée :
 - Examiner la distribution des modalités de réponse pour une variable.
 - Dans le cas d'une variable qualitative, par exemple, il s'agit d'un tri à plat, c'est-à-dire le dénombrement des observations correspondant à chaque modalité de la variable.
- L'analyse bivariée :
 - Etudier les relations entre deux variables.
 - Dans le cas de variables qualitatives, il s'agit d'un tableau croisé dénombrant les nombres d'observations correspondant à chaque combinaison possible des deux variables.
 - Consiste à mesurer la force de l'association entre deux variables (grâce par exemple au coefficient de corrélation pour des variables métriques).



Stratégies d'analyse

- L'analyse multivariée :
 - Permet de dépasser les techniques précédentes.
 - Permet le traitement simultané de plus de deux variables. L'ambition du chargé d'étude qui utilise des analyses multivariées peut se situer à deux niveaux :
 - décrire les données à sa disposition ou
 - expliquer des phénomènes grâce aux données à sa disposition.
 - On peut chercher à expliquer soit des différences (par exemple, entre les hommes et les femmes, les clients et les non-clients d'un magasin), soit des relations entre variables.



Méthodes multivariées à vocation descriptive

	Variables		
	Nominale	Ordinale	Métrique
Variable	Analyse factorielle des correspondances		Analyse factorielle
Individu	Typologies		

Méthodes explicatives

Une variable dépendante	Variables indépendantes		
	Nominale	Ordinale	Métrique
Nominale	Logit		Régression logistique Analyse discriminante
Ordinale	Analyse conjointe	Logit ordonné	
Métrique	Analyse de variance		Régression linéaire multiple
Plusieurs variables dépendantes	Nominale	Ordinale	Métrique
	Nominale		
	Métrique		
	Modèles log-linéaires		
	MANOVA		Équations structurelles



Décrire une variable quantitative

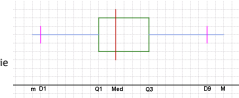
- Plusieurs indicateurs permettent de décrire une variable quantitative :
 - Les indicateurs de **tendance centrale** :
 - moyenne, médiane, mode.
 - Les indicateurs de **dispersion** :
 - étendue, variance, écart type, coefficient de variation.
 - Les indicateurs de **forme de la distribution** :
 - asymétrie, aplatissement.
 - Des **représentations graphiques** :
 - Ex : histogrammes ou boîtes à moustaches → description simple des variables quantitatives.



- On mesure la symétrie et la forme de la distribution par l'asymétrie et l'aplatissement.
- Le coefficient de symétrie (skewness) :
 - Mesure l'asymétrie d'une distribution.
 - Une distribution normale est symétrique → valeur de skewness de 0.
- Le coefficient d'aplatissement (kurtosis) :
 - Permet de mesurer le relief ou la platitude d'une courbe issue d'une distribution de fréquences.
 - Distribution normale → coefficient de kurtosis est de 0.

Intervalle de Tukey (boite à moustache)

- **Objectif :**
 - La boîte à moustaches (Box Plot) est un moyen rapide de figurer le profil essentiel d'une série statistique quantitative
 - Elle a été inventée en 1977 par John Tukey, mais peut faire l'objet de certains aménagements selon les utilisateurs.
- **Caractéristiques de la « boîte » :**
 - Longueur :
 - du 1^{er} quartile au 3^{ème} quartile
 - Coupe :
 - un trait vertical à hauteur de la médiane
- **Extrémités :**
 - Deux traits horizontaux
 - un va du 1^{er} quartile à la valeur minimale de la série
 - l'autre du 3^{ème} quartile à la valeur maximale
 - Sur ces deux moustaches :
 - on représente également en général les valeurs du 1^{er} et du dernier décile par des traits verticaux.



TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

La dispersion

Dispersion : Ecart-type (moment 2)

- La variance et l'écart-type mesurent eux la dispersion des variables autour de la moyenne

- **Variance (V) :**
 - moyenne des carrés des écarts à la moyenne

$$V(X) = \sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

- **Ecart-type :**
 - observation «j» de la variable X ; μ : moyenne de la variable X ; n : nombre d'observations;
 - racine carrée de la variance

$$\sigma(X) = \sqrt{V(X)}$$

Loi normale

- Dans une distribution normale :
 - moyenne arithmétique = médiane = mode
- L'évasement de la courbe correspond à l'écart-type.

Source : <http://biblioxtm.ugr.ac.ca/stat/Fichesstat/Distribution/normale.htm>

ISFIA

Symétrie d'une distribution (moment 3 standardisé)

- **Symétrie (skewness) : degré de symétrie dans la distribution à gauche et à droite de la moyenne**

- **Interprétation**
 - $SY > 0$: longue queue vers la droite (courbe A)
 - $SY = 0$; la distribution est symétrique p/r à la valeur centrale
 - $SY < 0$; longue queue vers la gauche (courbe B)
- **Valeur cible :**
 - Valeur cible = 0
 - Acceptable si entre [-1;+1] ou même [-2;+2]

ISFIA

Aplatissement d'une distribution (moment 4 standardisé)

- **Aplatissement (Kurtosis): degré de concentration des valeurs autour de la moyenne**

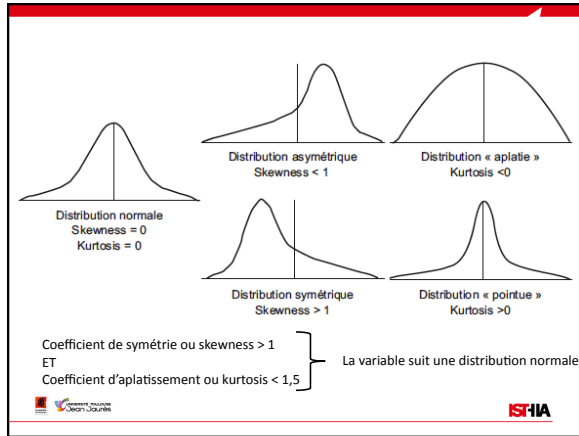
- **Interprétation**
 - $Ku > 0$; la distribution se situe en dessous d'une loi normale (B)
 - $Ku = 0$; la distribution est ajustée en amplitude à une loi normale (A)
 - $Ku < 0$; la distribution se situe au dessus d'une loi normale (C)
- **Valeur cible**
 - Valeur cible d'une loi normale = 0 (mésokurtique A)
 - Valeur < 0 -> Aplatissement négatif (leptokurtique B)
 - « courbe pointue, queues plus longues »
 - Valeur > 0 -> Aplatissement positif (platykurtique C)
 - « courbe plate, queues plus courtes que Normale »

ISFIA

Symétrie et Aplatissement

Valeur des coefficients		Conclusions sur la normalité des données
Coefficient d'asymétrie (Skewness)	Coefficient d'aplatissement (Kurtosis)	
< 1	$< 1,5$	Les données peuvent être considérées comme normales
< 2	< 7	Les données peuvent être considérées comme modérément normales
< 3	< 21	Les données peuvent être considérées comme non-normales

ISFIA



Les analyses bivariées

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ SOLAIRE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

Les analyses bivariées

- Les tests d'association :
 - échelles nominales :
 - Le χ^2 :
 - comparaison d'une variable théorique à une variable observée. (ex : csp et choix des marques)
 - échelle ordinale :
 - Kologorov Smirnov, Coefficients de corrélation de rangs de Spearman....
 - Vérifier que l'ordre n'est pas dû au hasard.
 - échelles métriques :
 - Anova
 - Régressions simples

UNIVERSITÉ SOLAIRE
Jean Jaurès

IST-HIA

Le test du Khi-deux

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

UNIVERSITÉ SOLAIRE
Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme
de l'Hôtellerie et de l'Administration

Le test du Khi-deux

- Le test du khi-deux (χ^2) cherche à tester si deux variables qualitatives (nominales ou ordinales) sont significativement associées.
- En réalité, c'est l'indépendance des variables qualitatives, présentées dans un tableau croisé, qui est testée.
- On cherche à vérifier si l'association des deux variables est suffisamment forte pour que l'hypothèse de leur indépendance puisse être rejetée.
- Le principe :
 - Comparer la distribution observée (O_{ij}), c'est-à-dire les effectifs que l'on peut lire dans le tableau croisé, à une distribution théorique (T_{ij}) qui correspond à l'hypothèse selon laquelle les deux variables sont indépendantes.



ISFIA

Le test du Khi-deux

- le test du Khi2 calcule la différence entre :
 - Les effectifs observés dans l'échantillon et
 - Les effectifs qui auraient du être observés s'il n'y avait pas de lien entre les deux var. qualitatives (effectifs théoriques)
- Si cette différence est grande :
 - on rejette l'indépendance entre les deux variables



ISFIA

Le test du Khi-deux

- **Formule du test :**
$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - T_{ij})^2}{T_{ij}}$$
 - i = numéro de la ligne ;
 - j = numéro de la colonne ;
 - l = nombre de lignes, c'est-à-dire le nombre de modalités de la variable présentée en lignes ;
 - c = nombre de colonnes, c'est-à-dire le nombre de modalités de la variable présentée en colonnes
- Le test du khi deux suit une loi à $(l-1)(c-1)$ ddl pour un seuil de risque α
- Règles de décision :
 - La valeur du Khi deux calculée est comparée à la valeur critique lue dans la table du Khi deux
 - Si $\text{Khi2}_{\text{calculé}} \leq \text{Khi2}_{\text{table}}$: on accepte H_0
 - Si $\text{Khi2}_{\text{calculé}} > \text{Khi2}_{\text{table}}$: on rejette H_0



ISFIA

Le test du Khi 2 (exemple)

		Destination Choisie			
		Baléares	Rocheuses	Egypte	Total
Niveau d'études	Effectif Obs.				
	Effectif Théo.	300 (270)	50 (112,5)	100 (67,5)	450
Primaire	Effectif Obs.	250	80	20	
	Effectif Théo.	(210)	(87,5)	(52,5)	350
Secondaire	Effectif Obs.	50	120	30	
	Effectif Théo.	(120)	(50)	(30)	200
Supérieur					
Total		600	250	150	1000

Exemple de calcul des Effectifs Théoriques = $(450 * 600) / 1000 = 270$



ISFIA

- Valeur calculée du Khi 2
 $(300-270)/270 + (50 - 112,5)/112,5 + \dots + (30-30)/30 = 220,91$
- Nombre de degré de liberté
 $(l-1)(c-1) = (3-1)(3-1) = 4$
- Valeur du Khi 2 donnée par la table
 - Pour 4 ddl et un seuil de confiance de 95% -> Khi 2 = 9,49
- Comparaison du Khi-deux calculé et du Khi-deux Théorique
 - Khi-deux calculé (220,91) > Khi-deux Théorique (9,49) -> on rejette H0**
- Conclusion
 - Le niveau d'étude a donc une influence sur le choix de la destination touristique. Il faudra donc orienter les actions marketing (contenu des brochures, réservation internet...)

ISFIA

Le test du Khi-deux : en résumé

- Les hypothèses du test :
 - H0 -> effectifs observés = effectifs théoriques
 - H0 -> effectifs observés ≠ effectifs théoriques
- On calcule pour chacune des cases du tableau :
 - Les effectifs théoriques qui auraient été observés si H0 est vraie (indépendance)
- Si H0 est vraie :
 - Les effectifs observés dans l'échantillon et les effectifs théoriques doivent être égaux (aux erreurs d'échantillonnage et d'observation près)
- Si H0 est fautive :
 - Les effectifs sont différents

ISFIA

Table du Khi deux

Tableau du Khi-deux (χ²)

pour	χ²(0,99)	χ²(0,95)	χ²(0,90)	χ²(0,85)	χ²(0,80)	χ²(0,75)	χ²(0,70)	χ²(0,65)	χ²(0,60)
1	0,0158	0,455	1,074	1,642	2,708	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,354	2,366	3,859	4,962	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,968	13,277	18,467
5	1,510	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,352
8	3,490	7,344	9,524	11,030	13,562	15,507	18,356	20,090	26,125
9	4,168	8,343	10,658	12,242	14,884	16,919	19,679	21,668	27,877
10	4,865	9,342	11,781	13,442	16,357	18,307	21,029	23,029	29,588
11	5,578	10,341	12,899	14,631	17,775	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	19,249	21,226	24,454	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,673	29,141	36,123
15	8,547	14,338	17,222	19,311	22,307	24,986	28,039	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,088	16,338	19,611	21,615	24,789	27,587	30,969	33,409	40,790
18	10,865	17,338	20,801	22,760	25,989	28,889	32,346	34,805	42,312
19	11,651	18,338	21,989	23,900	27,204	30,144	33,667	36,191	43,820
20	12,445	19,337	22,779	25,038	28,452	31,410	35,020	37,566	45,315
21	13,240	20,337	23,668	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,639	27,301	30,813	33,924	37,668	40,290	48,268
23	14,848	22,337	25,618	28,429	32,027	35,172	38,968	41,638	49,728
24	15,659	23,337	26,696	29,553	33,198	36,416	40,270	42,980	51,178
25	16,473	24,337	27,772	30,675	34,382	37,652	41,566	44,314	52,610
26	17,292	25,336	28,846	31,795	35,593	38,885	42,856	45,642	54,022
27	18,114	26,336	30,019	32,912	36,841	40,113	44,140	46,963	55,418
28	18,939	27,336	31,291	34,027	37,916	41,337	45,419	48,278	56,800
29	19,768	28,336	32,661	35,139	39,087	42,557	46,693	49,588	58,162
30	20,599	29,336	33,930	36,250	40,256	43,773	47,962	50,892	59,503

Quand le nombre de degrés de liberté est élevé, χ^2 est à peu près distribué normalement autour de $\sqrt{2(d.d.f.-1)}$ avec une variance égale à 1.

ISFIA

Test du Khi-deux

Pour obtenir un tableau croisé et faire le test d'indépendance du χ^2 , sélectionner **Analyse** → **Statistiques descriptives** → **Tableaux croisés**

Source : Denis Darpy

ISFIA

Test du Chi-deux

Lecture d'un tableau croisé :

Tableau croisé domicile * motif de déplacement

domicile		motif de déplacement		Total
		personnel	professionnel	
Paris	Effectif	176	259	435
	% dans domicile	40,5%	59,5%	100,0%
	% dans motif de déplacement	27,5%	19,6%	22,1%
Banlieue	Effectif	160	339	519
	% dans domicile	34,7%	65,3%	100,0%
	% dans motif de déplacement	28,1%	26,8%	26,4%
Province	Effectif	199	602	801
	% dans domicile	24,8%	75,2%	100,0%
	% dans motif de déplacement	31,0%	45,3%	40,8%
Etranger	Effectif	86	124	210
	% dans domicile	41,0%	59,0%	100,0%
	% dans motif de déplacement	13,4%	9,4%	10,7%
Total	Effectif	641	1324	1965
	% dans domicile	32,6%	67,4%	100,0%
	% dans motif de déplacement	100,0%	100,0%	100,0%

Nombre d'individus se déplaçant pour des raisons personnelles et habitant Paris = 176/435

Pourcentage des individus se déplaçant pour des raisons personnelles et habitant Paris, parmi tous les individus habitant Paris = 176/641

Pourcentage des individus se déplaçant pour des raisons personnelles et habitant Paris parmi tous les individus se déplaçant pour des raisons personnelles = 176/641

Pourcentage des individus se déplaçant pour des raisons personnelles et habitant Paris parmi tous les individus de l'échantillon = 176/1965

% Total

% Colonie

% Ligne

Attention : l'effectif théorique de chaque cellule du tableau croisé doit être supérieur à 5. Dans le cas contraire, il peut être nécessaire de procéder à des regroupements de modalités

Test du Khi-deux

Le test du χ^2 permet de vérifier si une relation entre deux variables (non métriques) existe dans la population.

Le test du χ^2 permet donc d'accepter ou de rejeter l'hypothèse H_0 "il n'y a pas de relation entre les deux variables dans la population dont est issu l'échantillon".

Ici, la valeur du χ^2 est relativement grande, ce qui lui permet d'être supérieure à la valeur critique correspondant au seuil de signification statistique de 0,05. Ainsi, nous avons une signification de 0,000 ce qui nous permet de rejeter H_0 et de conclure qu'il existe bien une relation entre les deux variables dans la population

	Chi-deux	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	41,837 ^a	3	,000
Rapport de vraisemblance	42,243	3	,000
Association linéaire par linéaire	9,732	1	,002
Nombre d'observations valides	1965		

^a 0 cellules (.0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 68,50.

Remarque : dans la lecture du tableau du Khi-deux, il est préférable de se référer au seuil de signification statistique qui est toujours le même (0,05) plutôt qu'à la valeur du χ^2 qui varie selon le nombre de degrés de liberté.

Source : Denis Darpy

TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

Le coefficient de corrélation

IST-HIA
Institut Supérieur du Tourisme de l'Hôtellerie et de l'Alimentation

Le calcul du coefficient de corrélation

Objectif :
vérifier la qualité de l'ajustement linéaire et l'intensité de la relation entre 2 variables métriques

- Formule :
$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \Leftrightarrow r = \frac{(\sum_{i=1}^n x_i y_i) - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2) - n \bar{x}^2} \sqrt{(\sum_{i=1}^n y_i^2) - n \bar{y}^2}}$$
- r est toujours compris entre -1 et 1
 - Si r = 1 très forte corrélation (lien linéaire) et variation dans le même sens
 - Si r = 0 pas de lien entre les variables
 - Si r = -1 très forte corrélation et variation de sens opposés

138

TOURISME
Tourism

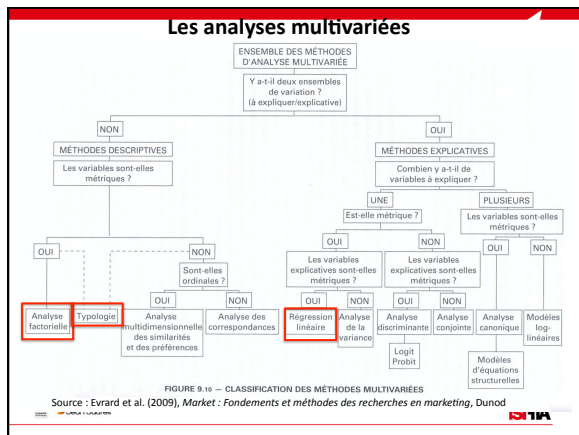
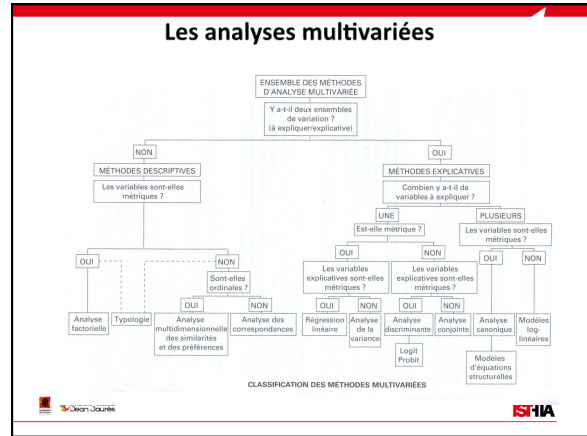
HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

Les analyses multivariées

UNIVERSITÉ TOULOUSE Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme de l'Hôtellerie et de l'Alimentation



TOURISME
Tourism

HÔTELLERIE
Hospitality

ALIMENTATION
Food Studies

Présentation des résultats

UNIVERSITÉ TOULOUSE Jean Jaurès

IST-HIA
Institut Supérieur du Tourisme de l'Hôtellerie et de l'Alimentation

La présentation des résultats

Elle se fait sous la forme d'un rapport structuré comprenant :

1. Problématique et objectifs de l'étude
2. Méthodologie
3. Résultats détaillés présentés sous forme de graphiques, de tableaux accompagnés de commentaires
4. Synthèse et recommandations principales
5. Conclusion