



UNIVERSITÉ ABDELMALEK ESSAËDI جامعة عبد الملك السعدي

المدرسة الوطنية للتجارة و التسيير - طنجة

ⵜⴰⵎⴰⵏⵜ ⵜⴰⵏⴳⴷⴰⵢⵜ ⵜⴰⵎⴰⵏⴳⴷⴰⵢⵜ ⵜⴰⵏⴳⴷⴰⵢⵜ ⵜⴰⵏⴳⴷⴰⵢⵜ ⵜⴰⵏⴳⴷⴰⵢⵜ

Ecole Nationale de Commerce et de Gestion - Tanger



Module : Analyse Des Données

(Semestre 5) – Automne 2023

Zineb El Andaloussi
z.elandaloussi@uae.ac.ma

Chapitre préliminaire

Introduction à l'Analyse Des Données (ADD)

I / ANALYSTE DE DONNEES

DE QUOI PARLE-T-ON?



Le Métier du Data Analyst

Rattachement : Le Data Analyst est généralement rattaché à la **direction des systèmes d'information (DSI)** d'une entreprise.

Missions : Le Data Analyste a pour missions

- ✓ **L'utilisation** de différentes technologies pour extraire les données de sources primaires et secondaires. Ces données peuvent être liées aux clients, aux marchés, aux produits, aux concurrents ou à l'environnement en général...
- ✓ **Recueillir**, analyser et interpréter les données pour en dégager des observations business utiles pour le diagnostic et la prédiction ;
- ✓ La **réorganisation** des données dans un format lisible ;
- ✓ **L'évaluation** de la qualité et la fiabilité des données collectées ;
- ✓ Le **développement** et le maintien de bases de données ;
- ✓ **L'analyse** des tendances locales, nationales et mondiales qui ont un impact sur l'entreprise et son secteur d'activité ;
- ✓ **Le travail** en équipe avec les managers pour orienter les prises de décision du Management et d'améliorer ainsi les performances et la croissance de l'entreprise ;
- ✓ **La préparation** de rapports d'analyse finaux, ce qui permet aux PP de prendre des décisions importantes sur la base de divers faits et tendances.

Le Métier du Data Analyst



Le Métier du Data Analyst

Formation et profil : Le Data Analyst est généralement issu d'une formation Bac+5 universitaire ou d'une grande école. Les profils qui ont suivi un cursus en mathématiques, statistiques, économie, marketing ou en informatique sont parmi les plus privilégiés par les entreprises. Le DA doit faire preuve de:

- ✓ Bonne Maîtrise des techniques d'analyse des données et des méthodologies statistiques
- ✓ Maîtrise de divers outils et logiciels (Excel, La connaissance de logiciels de visualisation de données tels que Tableau et Web Analytics, IBM SPSS, Nvivo, .. etc.),
- ✓ Une appétence pour les langages de programmation (ex. Python, Matlab, R, SQL...).
- ✓ Aisance communicationnelle, relationnelle et rédactionnelle
- ✓ Passion pour les chiffres.
- ✓ Un Anglais courant est très apprécié pour ce type de poste.

Quelle évolution pour le Métier de DA?

A l'heure du big data les perspectives d'évolution du poste sont nombreuses:

- **Définition des big Data** : Vastes ensembles de données collectées par les entreprises, caractérisées par le volume, la variété et la vélocité avec laquelle elles sont générées, collectées et traitées. Cela les différencie des mégadonnées des données traditionnelles.
- **Provenance**: A titre d'exemple, l'activité des internautes génère une myriade de données. Les clics, les applications mobiles et les réseaux sociaux capturent de nombreuses informations. Ainsi, Facebook ingère chaque jour plus de 500 terabytes de nouvelles données dans ses bases de données.
- **Importance** : Les Data-Driven Companies, peuvent analyser les Big Data pour découvrir de précieuses informations sur les besoins et **les attentes de ses clients** afin de créer de nouveaux produits ou des campagnes marketing ciblées et d'accroître ainsi la fidélité client (ou d'augmenter le taux de conversion)

Quelle évolution pour le Métier de DA?

- A l'heure du big data les perspectives d'évolution du poste de DA sont nombreuses, notamment dans les secteurs des services financiers, des nouvelles technologies, du conseil, ou de l'industrie
- Selon les entreprises, le DA pourra s'orienter naturellement vers des postes de **Lead Data Analyst** ou **Data Scientist**, en **Business Intelligence** ou vers des postes plus managériaux tels que **Responsable Data** ou **Chief Data Officer**.
- En termes de rémunération, un Data Analyst junior en Europe pourra démarrer sa carrière avec un salaire annuel de 3800 €.

II / LES DONNEES

Définition et typologie

Qu'est-ce que les données ?



- Les données consistent en des observations ou des mesures collectées pour répondre à une problématique précise, pour résoudre un problème particulier et pour mieux comprendre un phénomène.
- Les données sont des sources d'information au sujet d'une population spécifique ; elles peuvent donc se référer à différentes unités (un individu, un groupe d'individus, des objets, etc.)

Typologie des données selon la méthode de recherche utilisée pour les recueillir

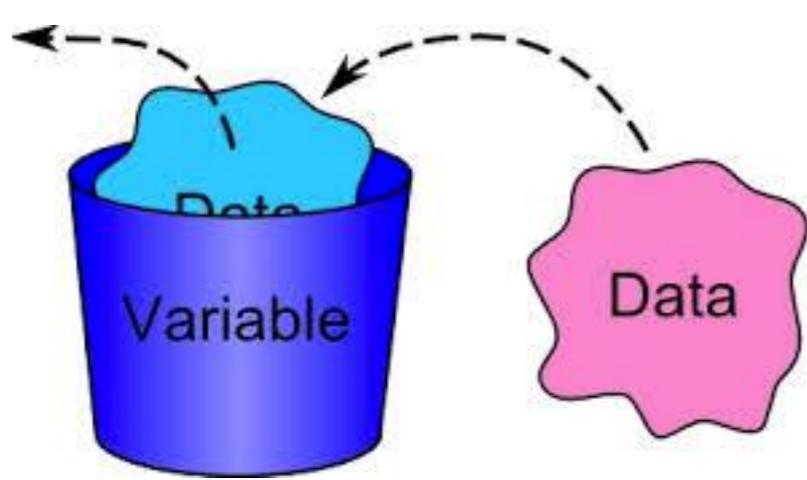
	Qualitatif	Quantitatif
But	Décrire Comprendre en profondeur des problèmes, des besoins et des Motivations, des freins..	Quantifier Confirmer ou tester de manière chiffrée des théories. des suppositions Prédire, Expliquer, Décrire
Répond à la question	Pourquoi ?	Quoi ? Combien ?
S'appuie sur	Mots et Expressions	Nombres et Statistiques
Questions	Ouvertes	Fermées
Résultats exprimés en	Mots et Verbatims	Graphiques et Pourcentage
Nombre de répondants	Petit nombre	> 100 personnes 200 personnes conseillé
Outils de collecte des données	Ex. Entretiens, focus groups, observation...	Ex. Questionnaire

Typologie des données selon la source

Données Primaires	Données secondaires
<ul style="list-style-type: none">• Elles sont collectées sur le terrain et sont spécialement produites en vue de réaliser une étude, c'est-à-dire pour répondre à une problématique spécifique.• Elles sont obtenues directement en utilisant des méthodes adaptées (qualitatives, quantitatives ou mixtes).	<ul style="list-style-type: none">• Ce sont des données de seconde main. Elles ont déjà été recueillies, traitées, interprétées par d'autres analystes.• Elles constituent une information déjà existante et accessible.• Elles peuvent provenir de différentes sources (rapports internes, rapports produits par des instituts de recherche, des organismes de statistiques, ...)

III / LES VARIABLES

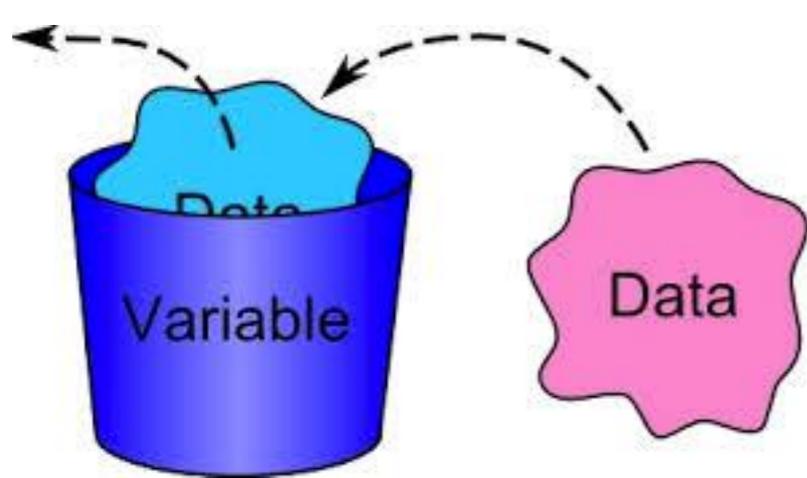
Définition et typologie



Les variables

Notion de variable

- Une **unité statistique** est une unité d'observation ou de mesure pour laquelle des données sont recueillies. Elle peut s'agir d'une personne, d'une entreprise, etc.
- Sur cette unité, on mesure un caractère (un trait). Il s'agit d'un critère pertinent au regard de l'étude et qui permet de décrire l'individu (ex. pour une entreprise, on peut mesurer les caractères chiffre d'affaires, année de création, ventes à l'export, effectif des employés, secteur d'activité, etc.)
- Les différentes valeurs de la variable sont appelées **modalités**.
- L'ensemble des modalités de la variable est appelé le **domaine de la variable** (ex. le domaine de la variable *Sexe* est {M, F})
- Différentes valeurs sont alors possibles pour le caractère statistique, d'où le mot « variable statistique ».
- Au plan strictement sémantique, le terme « variable » suppose qu'une réponse à une question donnée peut varier d'un individu à un autre. Donc : Si le caractère mesuré peut prendre différentes valeurs, on dit alors que **ce caractère est une variable**.
- *Une variable est donc « un critère par lequel on classe des individus dans des catégories ».*



Les variables

Notion de variable

Exemples de populations, individus et variables statistiques

Population	Individu	Variables statistiques
Ensemble des abonnés d'un opérateur de téléphonie	Un abonné	Âge, sexe, services possédés, montant de la facture mensuelle, etc.
Entreprises de BTP	Une entreprise	Nombre de salariés, chiffre d'affaires, délais de paiement, etc.
Usagers d'un restaurant d'entreprise	Un usager	Emploi, âge, sexe, fréquentation, dépense, appréciation, etc.
Ensemble des pièces produites par une chaîne de production	Une pièce	Diamètre, poids, robustesse, etc.

INDEPENDENT
VARIABLE

Influences

DEPENDENT
VARIABLE



Les variables

Typologie des variables

Classification des variables selon la chaîne causale

La variable indépendante : représente habituellement un facteur causal, un élément qui semble déterminant dans l'explication d'un phénomène. Dans certains cas, la variable indépendante est celle qui survient en premier lieu, dans le temps ; il y a donc une certaine antériorité.

La variable dépendante : est celle qui est influencée, celle dont on observe les variations. Il s'agit de « Toute grandeur dont le niveau peut être connu lorsqu'on connaît le niveau atteint par une autre grandeur dite variable indépendante ». On dit alors que la variable dépendante est fonction de la variable indépendante. Les variables dépendantes sont les variables manipulées, expliquées par les autres variables ; elles représentent ce que l'on cherche à savoir.

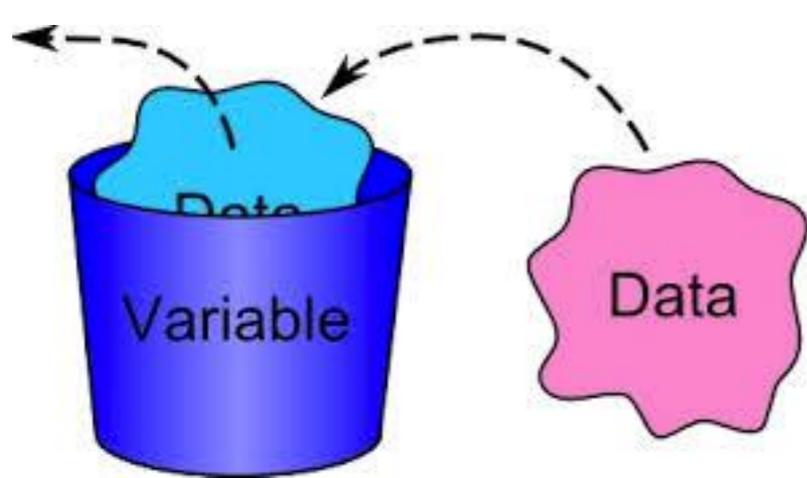
Variable indépendante

Les opinions des répondants
face au prix et à la fiabilité
d'une voiture



Variable dépendante

Le choix d'une marque
de voiture



Les variables

Typologie des variables

La classification Variable qualitative / quantitative

-**La variable est dite qualitative** quand ses modalités sont exprimées sous forme de « catégories ». Les variables qualitatives (ou catégorielles) n'ont pas donc de valeur numérique. Leurs modalités s'expriment sous forme d'énoncés qualitatifs, de façon littérale (les opérations arithmétiques telles que moyenne, somme, etc. n'ont pas de sens)

Illustration

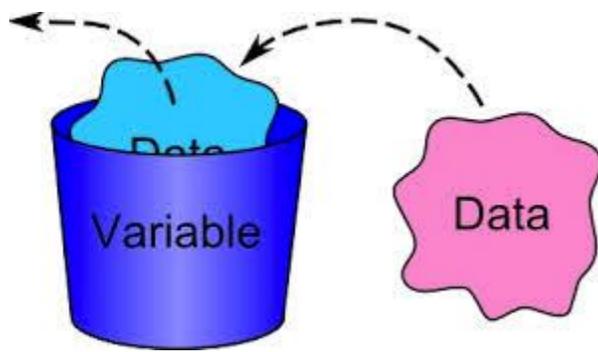
Question 1 : Êtes-vous ? : Femme Homme

Question 2 : Que prévoyez-vous pour les douze prochains mois ? Selon vous, en comparaison avec l'évolution des douze derniers mois, les prix à la consommation vont-ils :

- Augmenter plus fortement
- Augmenter dans la même proportion
- Augmenter moins fortement
- Rester pratiquement inchangés
- Diminuer

Question 3: Êtes-vous satisfait des services de nos livreurs GLOVO ? |

- Pas du tout
- Peu satisfait
- Satisfait
- Assez satisfait
- Très satisfait



Les variables

Typologie des variables

La classification Variable qualitative / quantitative

Une variable est dite quantitative quand elle est associée à un caractère mesurable, quand toutes ses modalités sont numériques c'est-à-dire des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

Illustration

Question 1: *Donnez un note de 0 à 5 exprimant votre satisfaction vis-à-vis de nos programmes d'enseignement*

0 1 2 3 4 5

Question 2: *Veillez indiquer l'âge de votre premier enfant :*

inférieur à 7 ans entre 7 et 14 ans entre 14 et 21 ans plus de 21 ans



Les variables

Variables statistiques et échelles de mesure

Une **échelle de mesure** est un ensemble de gradations permettant d'attribuer une valeur attendue à une question. L'utilisation d'une échelle de mesure pour obtenir une réponse à une question oriente le **répondant** sur les nuances souhaitées à sa réponse et permet d'avoir une information fiable et facile à traiter.

Variables	Structure
Qualitatives	Nominale
	Ordinale
Quantitatives	Intervalle
	Rapport



Les variables

Variables statistiques et échelles de mesure

Les variables qualitatives peuvent être mesurées par des échelles nominales ou ordinales

Les variables qualitatives mesurées sur une échelle Nominale

NOM

L'échelle nominale

- Avec les échelles nominales on cherche, tout simplement, à donner un nom (nom = nominale) et non une valeur numérique aux phénomènes que l'on observe.
- Les modalités de la variable sont de simples catégories non hiérarchisées. Il n'y a donc pas de modalité dont la ou les propriétés puissent être considérées comme inférieures ou supérieures à celles d'une autre modalité.
- La seule opération mathématique possible avec cette échelle est de compter les effectifs dans chacune des modalités et de calculer les fréquences. Il serait impensable de faire par exemple une moyenne à partir de ces différentes modalités.



Les variables

Variables statistiques et échelles de mesure

Les variables qualitatives mesurées sur une échelle Nominale

Question 1: *Seriez-vous favorable à la reprise des cours en mode présentiel ?* (Variable nominale dichotomique)

Oui Non

Question 2: *Quels outils utilisez-vous dans le cadre de votre enseignement ?* (Variable nominale polychotomique)

Echange de documents Evaluations en ligne des étudiants Outils de travail collaboratif (Wiki, etc.) Outil de communication avec les étudiants (courriels, forums, etc.) Sites web, blog Création de contenus pédagogiques (diapos, tableur, enregistrements, etc.) Présentation de contenus accessibles sur Internet (Podcasts, vidéos, etc.) Aucun Autres

Question 3: *Si vous avez du travail, quelle est votre catégorie professionnelle?* (Variable nominale polychotomique)

Employé ou cadre ouvrier spécialisé ouvrier indépendant chef d'entreprise agriculteur autres



Les variables

Variables statistiques et échelles de mesure

Les variables qualitatives mesurées sur une échelle Ordinale



▪ *L'échelle ordinale*

- L'échelle ordinale est similaire à l'échelle nominale exceptée qu'elle permet d'établir une relation d'ordre entre les éléments d'un ensemble. Les modalités sont ordonnées ou classées par ordre de grandeur, sans toutefois être capable d'évaluer de façon quantitative la distance qui les sépare.
- Avec cette échelle de mesure, on peut calculer des fréquences, mais aussi des moyennes et d'autres indicateurs statistiques.



Les variables

Variables statistiques et échelles de mesure

Les variables qualitatives mesurées sur une échelle Ordinale

Illustration

Question 1: Quelle est la note que vous accordez à notre livreur ?



A B C

Question 2: Etes-vous satisfait de la qualité des cours à distance ?

Pas du tout satisfait Insatisfait Ni satisfait ni insatisfait Satisfait Très satisfait

Question 3: Votre enfant aîné est-il ?

Ecolier Collégien Lycéen Universitaire

NB: la distance entre pas du tout satisfait et insatisfait ne peut pas être évaluée de manière quantitative.



Les variables

Variables statistiques et échelles de mesure

Les variables qualitatives mesurées sur une échelle Ordinale

Exemples de variables qualitatives et de modalités

Variable	Modalités possibles	Variable qualitative nominale/ordinale
Catégorie socioprofessionnelle	Ouvrier qualifié, agent de maîtrise, ingénieur, etc.	Nominale
Sexe	Femme, homme	Nominale
Jugement	Très insatisfait, insatisfait, satisfait, très satisfait	Ordinale

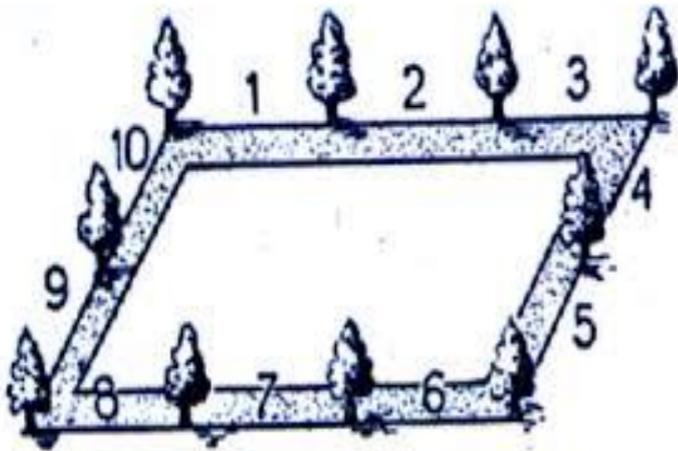


Les variables

Variables statistiques et échelles de mesure

Les variables quantitatives peuvent être mesurées par des échelles d'intervalles ou de rapports

Les variables quantitatives mesurées sur une échelle d'intervalles



▪ *L'échelle d'intervalles*

-L'échelle d'intervalles définit numériquement les intervalles entre les données.

-**Principe:** Deux intervalles prélevés renvoient toujours à la même quantité: l'intervalle entre deux catégories à toujours la même valeur, la même signification.

-On utilise cette échelle lorsque la comparaison d'intervalles est possible.

-A noter que le point zéro est fixé arbitrairement et ne correspond en rien à l'absence de phénomène.



Les variables

Variables statistiques et échelles de mesure

Les variables quantitatives peuvent être mesurées par des échelles d'intervalles ou de rapports

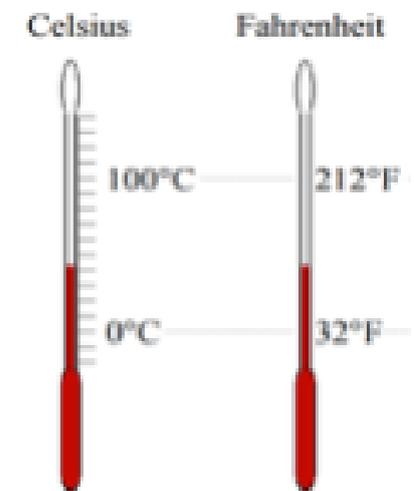
Les variables quantitatives mesurées sur une échelle d'intervalles

Illustration

Question : Quelle est la température qui règne dans cette région du Canada ?

- $[-20^{\circ} -10^{\circ}]$ $[-10^{\circ} 0^{\circ}]$ $[0^{\circ} 10^{\circ}]$ $[10^{\circ} 20^{\circ}]$ $[20^{\circ} 30^{\circ}]$

Un exemple typique est la température mesurée en degrés Celsius. Nous pouvons dire qu'une température de 20 degrés est plus élevée qu'une température de 10 degrés, et qu'une augmentation de 10 à 30 degrés est deux fois plus importante qu'une augmentation de 10 à 20 degrés. A 0°C , il y a toujours une température. C'est en fait « la température à laquelle l'eau gèle ».





Les variables

Variables statistiques et échelles de mesure

Les variables quantitatives peuvent être mesurées par des échelles d'intervalles ou de rappports

Les variables quantitatives mesurées sur une échelle de rapports



▪ *L'échelle rapports*

- L'échelle de rapport est la plus achevée des mesures.
- L'échelle de rapport implique que la distance entre deux unités est la même tout au long de l'échelle (tout comme dans l'échelle d'intervalle) et que le zéro existe et a un sens. Le zéro y est universel et signifie absence du phénomène ou mesure nulle.



Les variables

Variables statistiques et échelles de mesure

Les variables quantitatives peuvent être mesurées par des échelles d'intervalles ou de rappports

Les variables quantitatives mesurées sur une échelle de rapports

Illustration

Question: Quel est votre Revenu mensuel?

[0-2000[[2000-4000[[4000-6000[[6000-8000[8000 ou plus

Synthèse

LES ÉCHELLES DE MESURE ET LES PROPRIÉTÉS DES NOMBRES

<i>Échelles</i>	<i>Propriétés</i>			
	Classement	Ordre	Distance	Zéro absolu
Nominale	Oui	Non	Non	Non
Ordinale	Oui	Oui	Non	Non
Intervalles	Oui	Oui	Oui	Non
Rapport	Oui	Oui	Oui	Oui

Chapitre 1

**Analyse Des Données (ADD) issues des enquêtes
quantitatives par questionnaire**

Les enquêtes quantitatives par questionnaire

- 1. But des méthodes quantitative par questionnaire et Définition du questionnaire**
- 2. L'analyse des données issues d'enquêtes par questionnaires**

NB: Dans ce cours, nous allons:

- Privilégier les outils statistiques les plus couramment utilisés, au détriment des techniques habituellement enseignées dans les cours et manuels de statistique
- Favoriser l'exposé des principes, des logiques, des intérêts et des limites des outils statistiques, au détriment des aspects **mathématiques**

1. But des méthodes quantitative par questionnaire et Définition du questionnaire

-Le but des méthodes quantitatives de collecte de données

- *Quantifier et / ou mesurer les phénomènes étudiés*
- *Procéder au traitement statistique des informations collectées pour étudier des relations potentielles entre variables*

-Les enquêtes par questionnaire sont la principale méthode quantitative de collecte de données

- *Il s'agit d'interroger, à l'aide d'un questionnaire un échantillon d'individus le plus représentatif possible de l'ensemble de la population étudiée afin de collecter des données concernant un domaine particulier, telles que les opinions sur un sujet, les raisons d'un comportement, ou n'importe.*
- *Le questionnaire est document écrit qui permet de réunir un certain nombre d'informations sur des personnes, des entreprises, etc. Il comporte des questions standards et, sauf exceptions, les possibilités de réponses sont limitées par une liste exhaustive.*

2. L'analyse des données issues d'enquêtes par questionnaires

Conception du questionnaire utilisé dans cette partie du cours

Alpha Négoce a mis en place une action de formation en anglais destinée aux membres du département commercial (30 salariés). Chaque salarié a la possibilité de bénéficier de cours individuels en face à face avec l'un des deux formateurs (Jane ou John) désignés par le centre « *Speak English Center* ». Au Bout du 3 ème mois de formation, l'entreprise vous charge de conduire une enquête – satisfaction auprès des 30 participants à la formation.

Conception du questionnaire utilisé dans cette partie du cours

Questionnaire de satisfaction destiné aux salariés	
Q1. Quel est votre sexe?	Homme Femme
Q2. Quel est votre âge? ans
Q3. Qui est votre formateur ?	John Jane
Q4. Quel est le nombre total de séances que vous avez suivies? Séances
Q5. Pouvez-vous donner une note de 1 à 7 pour évaluer votre de satisfaction vis-à-vis du formateur?	
1 2 3 4 5 6 7	

Administration du questionnaire et constitution du fichier des données

- Dans le fichier des données, **les individus sont présentés en lignes** et les **variables sont en colonnes**.
- Chaque cellule présente la valeur que prend une variable pour un individu donné.

	Sexe	Age	Form	Nbr_Seanc	Note_satisf
1	M	21	John	8	5
2	M	22	John	8	4
3	M	25	John	9	7
4	F	36	Jane	6	2
5	F	41	Jane	4	1
6	F	28	Jane	5	3
7	M	26	John	12	5
8	F	38	Jane	7	3
9	M	35	John	10	5
10	M	24	John	11	6
11	F	41	Jane	6	4
12	F	34	Jane	9	5
13	M	32	John	9	4
14	F	38	Jane	5	1
15	M	22	John	10	5

15	M	33	John	12	5
16	F	42	Jane	7	5
17	M	31	John	5	4
18	M	33	John	8	6
19	M	40	John	8	6
20	F	47	Jane	9	7
21	F	27	Jane	6	4
22	F	44	Jane	6	3
23	M	43	John	3	5
24	F	33	Jane	7	3
25	M	45	John	10	2
26	F	36	Jane	7	5
27	M	49	John	8	4
28	F	39	Jane	6	4
29	F	35	Jane	7	4
30	F	38	Jane	3	2

Saisie des données sur le logiciel IBM SPSS

- **Présentation du logiciel SPSS**
- **Exécution de SPSS**
- **Saisie des variables et des observations (l'option d'importation de fichiers Excel est aussi envisageable)-----Voir tutoriel**

L'analyse statistiques des données du questionnaire sur SPSS

Selon le nombre de variables impliquées dans l'analyse, on peut effectuer des analyses univariées, des analyses bivariées et des analyses multivariées

Analyse univariée	L'étude des fréquences, des mesures de tendance centrale, de dispersion et de forme
Analyse bivariée	A partir des tableaux croisés, analyse de variance, la corrélation, la régression simple
Analyse multivariée	L'analyse des correspondances, analyse en composantes principales, analyse discriminante et régression multiple

L'analyse univariée



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Analyse univariée

Appelées également **tri à pla**, la **SU** a pour principales fonctions de :

- **R**ésumer les renseignements obtenus sur chaque variable dans un tableau simple ou dans une figure - Dénombrer les réponses obtenues - Compter l'apparition des différentes modalités d'une seule variable.
- **S**e familiariser avec les données (« *Getting a feel for the data* »)
- **S'**assurer de la représentativité de l'échantillon en comparer les profils des variables d'identification (réponses aux questions signalétiques) des individus de l'échantillon aux profils de ces mêmes variables dans la population étudiée
- **É**liminer du dépouillement des questions inintéressantes parce qu'elles sont restées la plupart du temps sans réponse, ou que les enquêtés y ont presque tous répondu de la même façon.

Statistique univariée (SU)

Statistique univariée appliquée aux variables qualitatives

Dans le cas des variables **qualitatives**, l'information est résumée dans :

- **Des tableaux**: des tableaux de distributions qui comportent, dans une première colonne les modalités et dans une seconde les effectifs.

Distribution de fréquences		
Fréquence absolue	Fréquence relative	Fréquence cumulée
$\sum_{i=1}^k n_i = n$	$f_i = \frac{n_i}{n}$	$F_1 = f_1$ $F_2 = f_1 + f_2$ $F_3 = f_1 + f_2 + f_3$: : $F_k = f_1 + \dots + f_k = 1$

n_i : nombre d'éléments ayant pour une modalité / valeur d'une variable X et n : nombre total d'éléments

Note : Pourcentage = fréquence * 100

- **Des représentations graphiques** qui se font par le biais de **diagrammes en secteurs** ou en barres.

LET'S
PLAY

Statistique univariée (SU)

Statistique univariée appliquée aux variables qualitatives

Statistiques

Sexe

N	Valide	30
	Manquant	0

Sexe

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Homme	14	46,7	46,7	46,7
	Femme	16	53,3	53,3	100,0
Total		30	100,0	100,0	

Statistiques

Nom du formateur

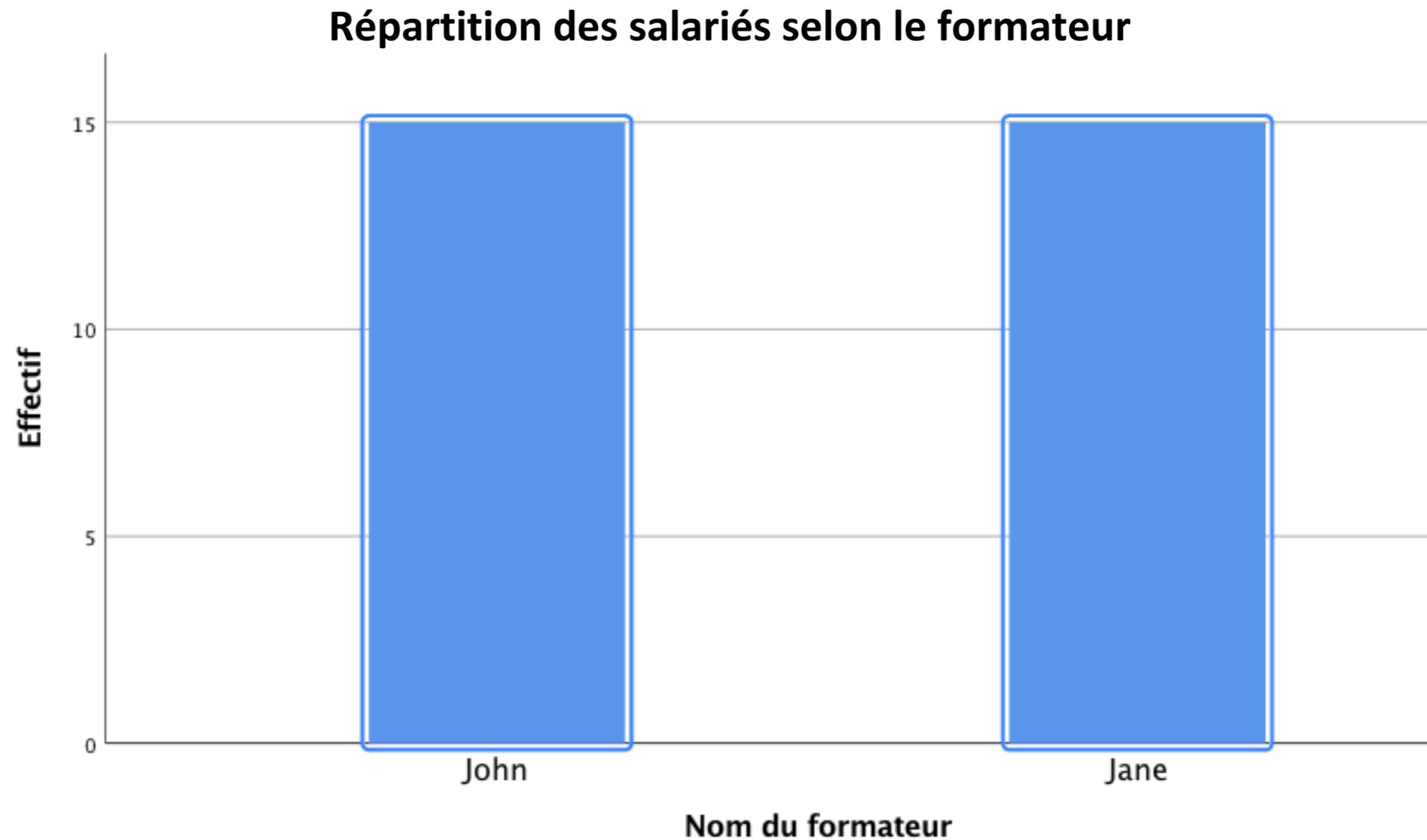
N	Valide	30
	Manquant	0

Nom du formateur

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	John	15	50,0	50,0	50,0
	Jane	15	50,0	50,0	100,0
Total		30	100,0	100,0	

Statistique univariée (SU)

Statistique univariée appliquée aux variables qualitatives



Statistique univariée (SU)

Statistique univariée appliquée aux variables quantitatives

Dans le cas des variables quantitatives, l'information est résumée dans :

- Des tableaux de distribution** qui comportent, dans une première colonne les modalités et dans une seconde les effectifs
- Des graphiques** : des diagrammes en bâtons pour les variables discrètes (les modalités sont des nombres entiers) et des histogrammes pour les variables continues (qui sont susceptibles de prendre n'importe quelle valeur dans un intervalle donné).
- Plusieurs indicateurs statistiques**
 - 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.
 - etc.

LET'S
PLAY

Statistique univariée (SU)

Statistique univariée appliquée aux variables quantitatives : *Tableau de distribution*

Tableau de distribution pour la variable « note de satisfaction »

Note de satisfaction					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	1	2	6,7	6,7	6,7
	2	3	10,0	10,0	16,7
	3	4	13,3	13,3	30,0
	4	8	26,7	26,7	56,7
	5	8	26,7	26,7	83,3
	6	3	10,0	10,0	93,3
	7	2	6,7	6,7	100,0
	Total		30	100,0	100,0

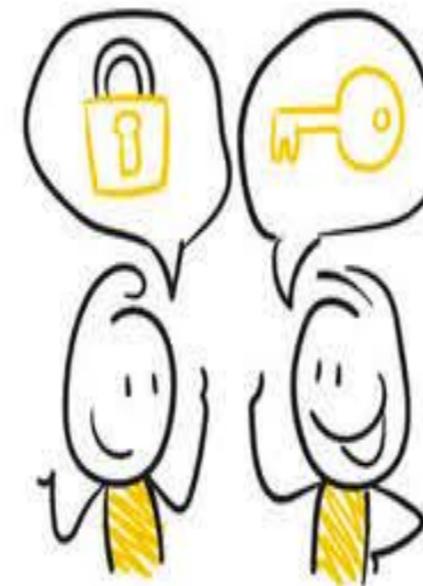
Statistique univariée (SU)

Statistique univariée appliquée aux variables quantitatives : *Tableau de distribution*

		Age			
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	21	1	3,3	3,3	3,3
	22	1	3,3	3,3	6,7
	24	1	3,3	3,3	10,0
	25	1	3,3	3,3	13,3
	26	1	3,3	3,3	16,7
	27	1	3,3	3,3	20,0
	28	1	3,3	3,3	23,3
	31	1	3,3	3,3	26,7
	32	1	3,3	3,3	30,0
	33	3	10,0	10,0	40,0
	34	1	3,3	3,3	43,3
	35	2	6,7	6,7	50,0
	36	2	6,7	6,7	56,7
	38	3	10,0	10,0	66,7
	39	1	3,3	3,3	70,0
	40	1	3,3	3,3	73,3
	41	2	6,7	6,7	80,0
	42	1	3,3	3,3	83,3
	43	1	3,3	3,3	86,7
	44	1	3,3	3,3	90,0
45	1	3,3	3,3	93,3	
47	1	3,3	3,3	96,7	
49	1	3,3	3,3	100,0	
Total		30	100,0	100,0	



Tableau très longs?



Recorder les variables

LET'S
PLAY

Statistique univariée (SU)

Statistique univariée appliquée aux variables quantitatives : *Tableau de distribution*

Statistiques					
âge des salariés					
N	Valide	30			
	Manquant	0			

âge des salariés					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	entre 21 et 25	4	13,3	13,3	13,3
	entre 26 et 30	3	10,0	10,0	23,3
	entre 31 et 35	8	26,7	26,7	50,0
	entre 36 et 40	7	23,3	23,3	73,3
	entre 41 et 45	6	20,0	20,0	93,3
	entre 46 et 50	2	6,7	6,7	100,0
	Total	30	100,0	100,0	

Statistique univariée (SU)

Statistique univariée appliquée aux variables quantitatives: *Indicateurs de tendance centrale*

La moyenne arithmétique

- Somme des valeurs de toutes les observations divisée par l'effectif.
- La moyenne révèle la tendance centrale en ce sens que les réponses se trouvent réparties de part et d'autre de la moyenne.

Si on note par : x_i : la $i^{\text{ème}}$ observation, la moyenne arithmétique est définie par :

$$\bar{X} = \frac{1}{n} \sum_i n_i x_i = \sum_i f_i x_i \text{ avec } n = \sum_i n_i.$$

Salarié	1	2	3	4	5	6	7	8	9	10	11	Sum
Age (xi)	46	23	34	25	28	31	23	40	36	45	24	355

Moyenne = somme des xi / n = 355/11= 32 ans

NOTA BENE! *La moyenne a un gros défaut: elle est sensible aux valeurs extrêmes.*

Statistique univariée (SU)

Statistique univariée appliquée aux variables métriques (quantitatives) : *Indicateurs de tendance centrale*

Le mode

Le mode est la valeur la plus fréquente dans un échantillon. C'est la valeur de la variable qui se répète le plus de fois.

Salarié	1	2	3	4	5	6	7	8	9	10	11	Sum
Age (xi)	46	23	34	25	28	31	23	40	36	45	24	355

Mode = ... ans

LET'S
PLAY

Statistique univariée (SU)

Statistique univariée appliquée aux variables métriques (quantitatives) : *Indicateurs de tendance centrale*

		Statistiques		
		Age	Nombre de séances	Note de satisfaction
N	Valide	30	30	30
	Manquant	0	0	0
Moyenne		35,20	7,37	4,13
Médiane		35,50	7,00	4,00
Mode		33 ^a	6 ^a	4 ^a

a. Présence de plusieurs modes. La plus petite valeur est affichée.

L'analyse statistique Bivariée

- Analyse de la relation entre deux variables qualitatives
- Analyse de la relation entre deux variables quantitatives

Analyse de la relation entre deux variables qualitatives



Analyse des données de
l'enquête par questionnaire

L'analyse des données du
questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

LES TESTS D'HYPOTHÈSES

- Pour mesurer l'indépendance des variables, il faut utiliser des tests d'hypothèses.
- Ces tests nous permettent de déterminer si les relations entre deux variables données sont dues au hasard ou sont réellement **significatives**.
- Dans l'analyse bivariée, les tests statistiques à effectuer varient selon la nature des variables. Pour mesurer la relation entre deux variables qualitatives. On utilise souvent le **test de Chi-deux**



Analyse des données de
l'enquête par questionnaire

L'analyse des données du
questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

LES TESTS D'HYPOTHÈSES

- Le Chi-carré est un test d'indépendance: Ainsi, il y a deux hypothèses
 - L'hypothèse H_0 (nulle) présume qu'il n'y a pas de relation entre les variables (les deux variables sont indépendantes).
 - L'hypothèse H_1 : au contraire, affirme qu'il y a une relation entre les deux variables.



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

TABLEAU DE CONTINGENCE

LA PUBLICITÉ ET L'ACHAT D'UN FORFAIT-VOYAGE

		Achat du forfait-voyage		
		Oui	Non	Total
A vu la publicité dans le dernier mois	Oui	80	120	200
	Non	15	85	100
Total		95	205	300

Nous voyons dans ce tableau que 40 % (80/200) de ceux qui ont vu la publicité dans le dernier mois ont acheté le forfait-voyage.

Le test du khi-carré nous permettra d'affirmer si la publicité influence l'achat d'un forfait-voyage ou l'inverse.



Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 1: Formulation des hypothèses

Dans notre cas, nous pouvons formuler deux hypothèses (test d'hypothèses):

1. l'hypothèse H_0 : la publicité n'a aucune influence sur l'achat du forfait-voyage

2. l'hypothèse H_1 : la publicité a une influence sur l'achat du forfait voyage.

Les résultats des tests de Khi deux permettront de trancher entre ces deux hypothèses et d'en tirer les conséquences pour l'interprétation des données.



Analyse des données de
l'enquête par questionnaire

L'analyse des données du
questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 2: Déterminer le seuil de signification (α)

α indique la probabilité que le test nous révèle qu'il existe un lien entre les deux variables alors que dans les faits ce lien n'existe pas.

Par exemple, on peut accepter un risque de 5% donc $\alpha=5\%$



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 3: Produire le tableau de contingence et calculer les effectifs théoriques

Le test du khi-carré fonctionne essentiellement dans la comparaison entre un effectif observé et un effectif théorique.

L'effectif théorique correspond à l'effectif qu'on aurait **s'il n'y avait pas de lien entre les variables.**

La formule générale est la suivante:

$$\chi^2 = \sum \sum \frac{(F_o - F_{th})^2}{F_{th}}$$

Où:

F_o : la fréquence observée; ce sont les résultats obtenus sur le terrain.

F_{th} : la fréquence théorique; elle est obtenue par le calcul: (Total de la colonne \times Total de la rangée) / Grand total.

Fréquence théorique =

(Total de la colonne) * (Total de la ligne) / Grand total

Ex. 63,3 = 95 * 200 / 300



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

TABLEAU DE CONTINGENCE

<i>LA PUBLICITÉ ET L'ACHAT D'UN FORFAIT-VOYAGE</i>			
<i>A vu la publicité dans le dernier mois</i>	<i>Achat du forfait-voyage</i>		<i>Total</i>
	<i>Oui</i>	<i>Non</i>	
<i>Oui</i>	80	120	200
<i>Non</i>	15	85	100
<i>Total</i>	95	205	300

Fréquence théorique de $63,3 = 95 * 200 / 300$



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

<i>Fréquence observée</i> (F_o)	<i>Fréquence théorique</i> (F_{th})	$F_o - F_{th}$	$(F_o - F_{th})^2$	$\frac{(F_o - F_{th})^2}{F_{th}}$
80	63,3	16,7	278,9	$\frac{278,9}{63,3} = 4,4$
120	136,6	-16,6	275,6	$\frac{275,6}{136,6} = 2,0$
15	31,6	-16,6	275,6	$\frac{275,6}{31,6} = 8,7$
85	68,3	16,7	278,9	$\frac{278,9}{68,3} = 4,1$
KHI-CARRÉ				19,2



Analyse des données de
l'enquête par questionnaire

L'analyse des données du
questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 4: Vérifier les conditions d'applicabilité du test

- l'échantillon ≥ 30 (OUI)
- Tous les effectifs théoriques ≥ 5 (OUI)

Étape 5: Nous calculons le nombre de degrés de liberté

$$v = (\text{Nombre de lignes} - 1) (\text{Nombre de colonnes} - 1) = (2 - 1) * (2 - 1) = 1$$



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 4: Consulter la table du khi-carré: celle-ci nous donne les valeurs critiques en fonction du seuil de signification et du nombre de degré de liberté. Avec un nombre de degrés de liberté égale à 1 et un seuil de signification égale à 5%, la valeur critique de khi carré sera égale à 3,84

$v \backslash \alpha$	10%	5%	2,5%	1%	0,5%
1	2,71	3,84	5,02	6,54	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,82	9,35	11,35	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,65	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,54	20,09	21,96
9	14,68	16,92	19,02	21,67	23,59
10	15,99	18,31	20,48	23,21	25,19
11	17,28	19,68	21,92	24,73	26,76
12	18,55	21,03	23,34	26,22	28,30



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

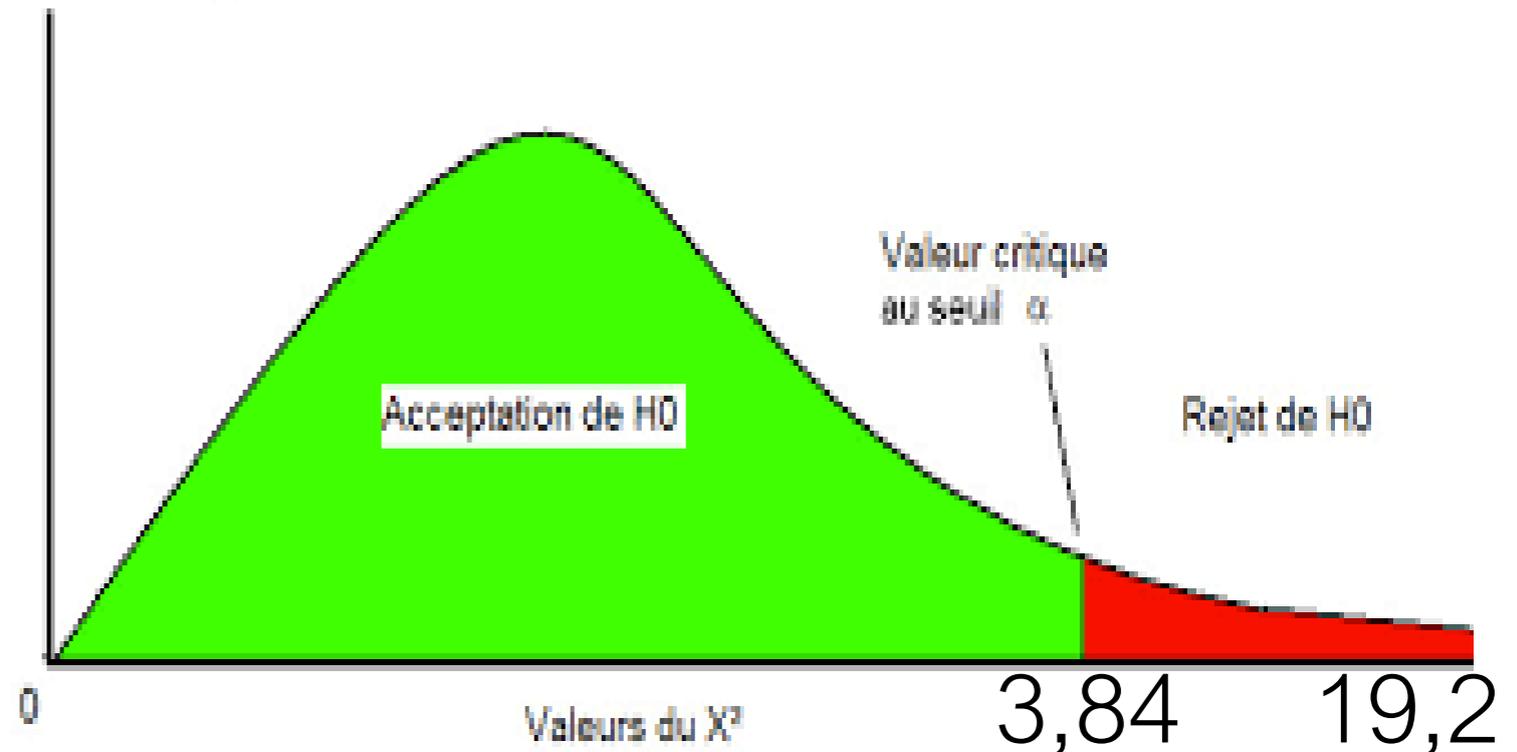
Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Étape 5: Décision

- **Khi-carré > seuil critique** : rejeter l'hypothèse H_0 et par conséquent accepter l'hypothèse H_1 = affirmer qu'il y a un lien entre les variables
- **Khi-carré < seuil critique** : accepter l'hypothèse H_0 et par conséquent rejeter l'hypothèse H_1 = absence de lien entre les variables

Dans notre cas:

- Khi-carré = 19,2 > 3,84 (seuil critique figurant dans la table de Khi-deux)
- Le Khi-carré se trouve dans dans la zone de rejet de H_0 (accepter H_1)



Interprétation: Au seuil de signification de 5%, on doit admettre qu'il existe un lien entre le fait de voir la publicité et d'acheter le pack voyage (accepter H_1)



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

OUTPUT dans SPSS VIEWER

• *La valeur de khi carré est égale à 8,571 avec 1 degré de liberté. avec $8,571 > 3,84$. cette valeur se trouve dans dans la zone de rejet de H_0*

Conclusion: Au seuil de signification de 5%, on peut affirmer qu'il existe un lien entre le sexe du participant et le formateur

NB: La règle de décision du test sur SPSS se base sur un **niveau de signification statistique (valeur p)**.

- Si $p < 0,05$ (seuil de signification) : on rejette H_0 et on confirme H_1 (un lien existe entre les deux variables)
- Si $p > 0,05$ (seuil de signification) : On accepte H_0 et on rejette H_1 (absence de lien entre les deux variables)



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Test d'indépendance de Khi-deux*

Tableau croisé Sexe * Nom du formateur

			Nom du formateur		
			John	Jane	Total
Sexe	Homme	Effectif	11	3	14
		Effectif théorique	7,0	7,0	14,0
		Résidu	4,0	-4,0	
	Femme	Effectif	4	12	16
		Effectif théorique	8,0	8,0	16,0
		Résidu	-4,0	4,0	
Total	Effectif		15	15	30
	Effectif théorique		15,0	15,0	30,0

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	8,571 ^a	1	,003



Analyse des données de
l'enquête par questionnaire

L'analyse des données du
questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Mesures d'association*

*Mesurer la force de la relation entre deux variables qualitatives :
les mesures d'association*

Si on affirme qu'une relation existe entre deux variables, on peut donc déterminer la force de ce lien en utilisant

- le coefficients Phi
- le V de Cramer
- le coefficients de contingence



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables qualitatives : *Mesures d'association*

Le Phi

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Tableaux 2*2

Le V de Cramer

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Tableaux dépassant 2*2

Valeur	Force du lien statistique
0	Absence de relation
Entre 0,05 et 0,10	Très faible
Entre 0,10 et 0,20	Faible
Entre 0,20 et 0,40	Modérée
Entre 0,40 et 0,80	Forte
Entre 0.80 et 1	Louche (Colinéarité)

Le coefficient de contingence de Pearson (C)

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

grands tableaux / grands échantillons

- parfaite si la valeur de C = 1
- très forte si C se situe entre 0,8 et 1
- forte si C se situe entre 0,5 et 0,8.
- d'intensité moyenne si C se situe entre 0,2 et 0,5.
- faible si C se situe entre 0 et 0.2.
- nulle si C = 0

Analyse de la relation
entre deux variables
quantitatives



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

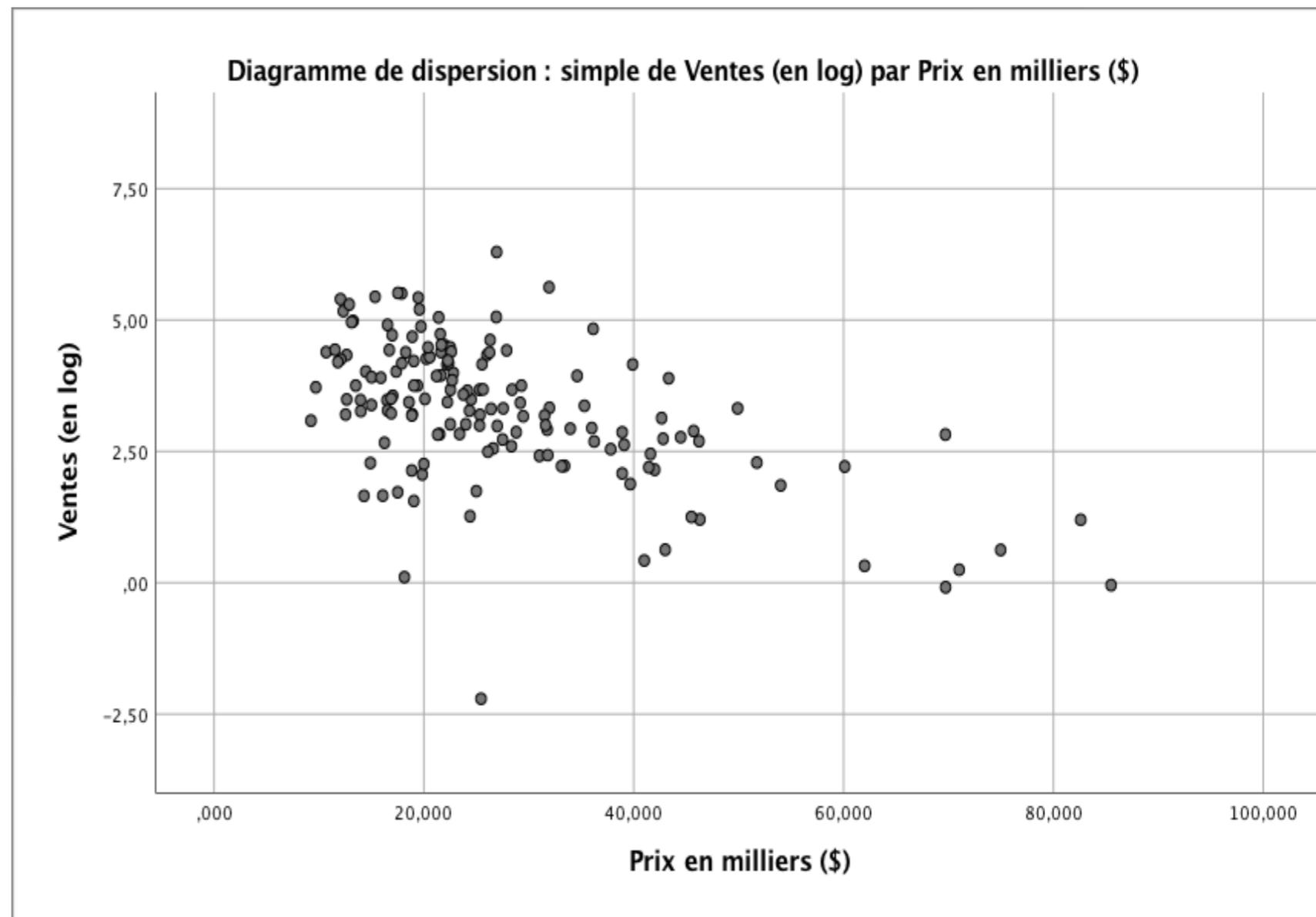
Statistique BIVARIÉE appliquée à deux variables quantitatives : *La représentation graphique*

- Diagramme de dispersion
- Nuage de points

Exemple CARSALES

L'évaluation du lien à partir du diagramme de dispersion :

Ce nuage de points a une forme ovoïde qui tend à former une ligne droite, cela indique une relation de type linéaire entre deux variables quantitatives et on parle de **corrélacion linéaire**





Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables **quantitatives** : *Mesure de la relation*

Le coefficient de corrélation linéaire de Pearson

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \times \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables quantitatives : *Mesure de la relation*

Exemple prix et ventes d'un échantillon de 10 voitures

prix x	ventes y	X= x- moyenne des x	Y= y- moyenne des y	XY	X au carré	Y au carré
42,000	8,588	14,552	-13,955	-203,07	211,76	194,73
23,990	20,397	-3,458	-2,146	7,42	11,96	4,60
33,950	18,780	6,502	-3,763	-24,47	42,28	14,16
62,000	1,380	34,552	-21,163	-731,21	1193,84	447,86
26,990	19,747	-0,458	-2,796	1,28	0,21	7,82
33,400	9,231	5,952	-13,312	-79,23	35,43	177,20
38,900	17,527	11,452	-5,016	-57,44	131,15	25,16
21,975	91,561	-5,473	69,018	-377,74	29,95	4763,53
25,300	39,350	-2,148	16,807	-36,10	4,61	282,49
31,965	27,851	4,517	5,308	23,98	20,40	28,18
274,480	225,427			-1476,58	1681,59	5945,72
				Racine carré	41,01	77,11



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables quantitatives : *Mesure de la relation*

$$\text{Moyenne des } x_i = \bar{X} = (\sum x_i) / N = 274,480 / 10 = 27,44$$

$$\text{Moyenne des } y_i = \bar{Y} = (\sum y_i) / N = 225,427 / 10 = 22,54$$

$$r = -1476,58 / (41,01 * 77,11) = -0,47$$



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables quantitatives : *Mesure de la relation*

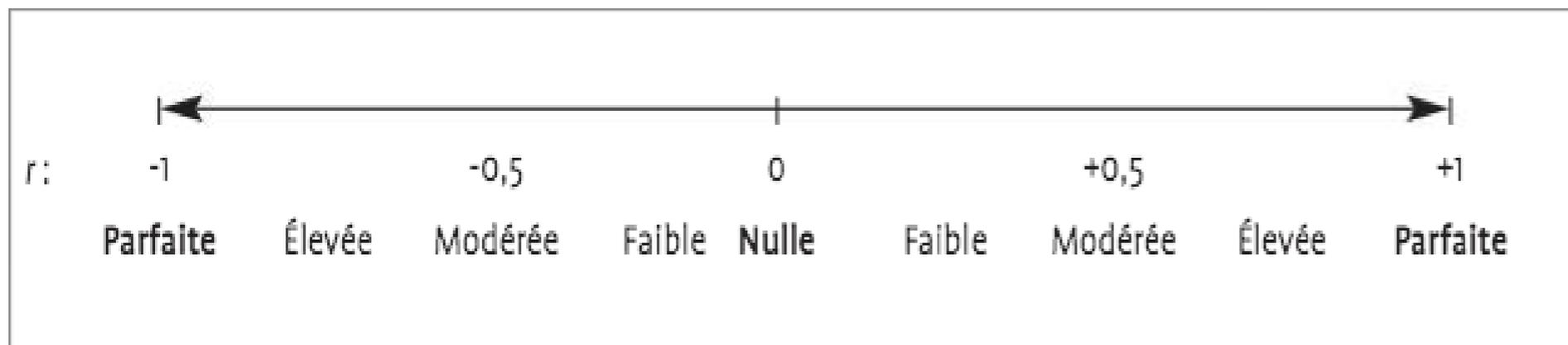
Interprétation du coefficient de corrélation linéaire de Pearson

Sens de la corrélation

- Si r est NÉGATIF : alors les deux variables varient en sens contraire
- Si r est POSITIF : alors les deux variables varient dans le même sens
- Si $r=0$ pas de corrélation

Force de la corrélation

Les valeurs de r sont toujours comprises entre -1 et 1.



Dans l'exemple: un **R négatif** indique en conformité avec la figure, une corrélation négative entre les deux variables: cela signifie que les deux variables varient en sens contraire (elles entretiennent une relation linéaire négative): une augmentation du prix est associée à une diminution des ventes. Une valeur de **-0,47** indique que cette corrélation est de faible à modérée



Analyse des données de l'enquête par questionnaire

L'analyse des données du questionnaire sur SPSS

Statistique BIVARIÉE

Statistique BIVARIÉE appliquée à deux variables quantitatives : *Mesure de la relation*

Exemple CARSALES (corrélation prix et ventes des voitures)

		Corrélations	
		Prix en milliers (\$)	Ventes (en log)
Prix en milliers (\$)	Corrélation de Pearson	1	-,553**
	Sig. (bilatérale)		,000
	N	155	155
Ventes (en log)	Corrélation de Pearson	-,553**	1
	Sig. (bilatérale)	,000	
	N	155	157

** . La corrélation est significative au niveau 0.01 (bilatéral).

FIN