

Module 110

Analyser et représenter des données avec des outils

CIE Informaticiens CFC – Journée 1

Flavio Pacifico

CIE 110 : Sommaire

- Qui suis-je ?
- Horaires.
- Présentation des objectifs ICT.
- Approche quantitative ou qualitative ?
- Que veut-on mesurer ?
- Quels sont les objectifs ?
- Quels sont les moyens ?
- Les différents logiciels de statistiques
- Les limites de la quantification

Qui suis-je ?

Activités



Formations

- ✓ CIE 101 – Réaliser et publier un site web – VD
- ✓ CIE 110 – Analyser et représenter des données avec des outils – VD
- ✓ CIE 335 - Réaliser une application pour mobile - VD



Développement

- ✓ Langage Web
 - ✓ CRM en mode web (HTML, CSS, Javascript, Framework JS, PHP)
 - ✓ Base de données SQL
- ✓ C#



Support

- ✓ Support sur les applications développées.

CIE 110 : Déroulement



4.5 jours : cours / participation obligatoire



- ✓ Théorie
- ✓ Exercices
- ✓ Pause matin, midi et après-midi

0.5 jour = examen individuel



- ✓ Avec document

CIE 110 : Objectifs

Analyser et représenter des données avec des outils

Compétence

Lire des données à des fins d'analyse, les valider et établir, à l'aide de métriques, de valeurs limites et d'indicateurs appropriés, des rapports et des représentations graphiques de données de monitoring.

Transmettre des données anonymisées à des tiers.

Objectifs opérationnels

1. Obtenir, à partir des données disponibles, des indicateurs pour l'établissement de rapports.
2. Définir des valeurs limites pour les alertes.
- 3 Élaborer à l'aide d'un outil des représentations graphiques appropriées de données.
- 4 Décrire un concept d'ingénierie de la fiabilité des sites (Site Reliability Engineering [SRE]) incluant l'explication des termes clés, p. ex. indicateur de niveau de service (Service Level Indicator [SLI]), objectif de niveau de service (Service Level Objective [SLO]), etc.
- 5 Analyser des données avec un logiciel approprié.
- 6 Valider les données lues.
- 7 Préparer des données personnelles anonymisées en vue de leur traitement.
- 8 Définir des métriques appropriés pour le monitoring logiciel.
- 9 Extraire des données de logs et les interpréter.

CIE 110 : Introduction

L'analyse de données pour la gestion



L'analyse de données a plusieurs objectifs :

- ✓ Mieux comprendre une situation ou un phénomène.
- ✓ Identifier les déterminants ou facteurs explicatifs de certains comportements.
- ✓ Définir des politiques et des actions plus efficaces en fonction des buts poursuivies.

CIE 110 : Introduction

Avant de commencer



Avant de se lancer dans une étude et dans la mobilisation de données, il convient de se poser un certain nombre de questions qui orienteront à la fois le choix des méthodes, des données à recueillir, des types de résultats et de livrables.

CIE 110 : Introduction

Avant de commencer



Les questions préalables à toute mobilisation de statistiques et de données sont les suivantes :

- ✓ Une approche quantitative est-elle plus pertinente qu'une approche qualitative pour traiter les questions souhaitées ?
- ✓ Que cherche-t-on exactement à mesurer ?
Cette mesure est-elle vraiment possible ?
- ✓ Quels sont les aménagements et réflexions préalables nécessaires ?
- ✓ Quels objectifs cherche-t-on à atteindre avec l'analyse de données ?
- ✓ De quels moyens financiers, humains (compétences, temps disponible) et informatiques (logiciels, données) dispose-t-on ?

CIE 110 : Introduction

Avant de commencer



Nous allons donc voir les points précédents plus en détail étape par étape :

- ✓ Approche quantitative ou qualitative?
- ✓ Que veut-on mesurer ?
- ✓ Quels sont les objectifs ?
- ✓ Quels sont les moyens ?

CIE 110 : Approche quantitative ou qualitative ?



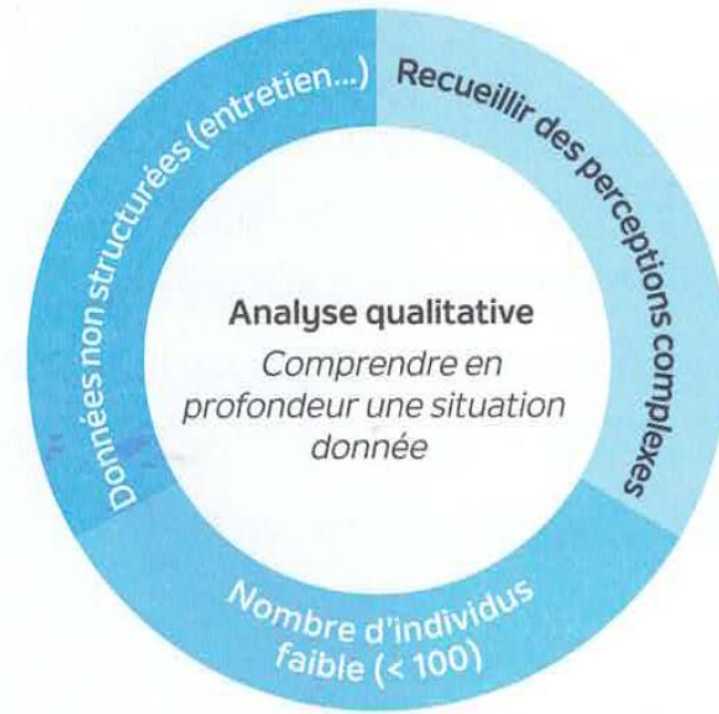
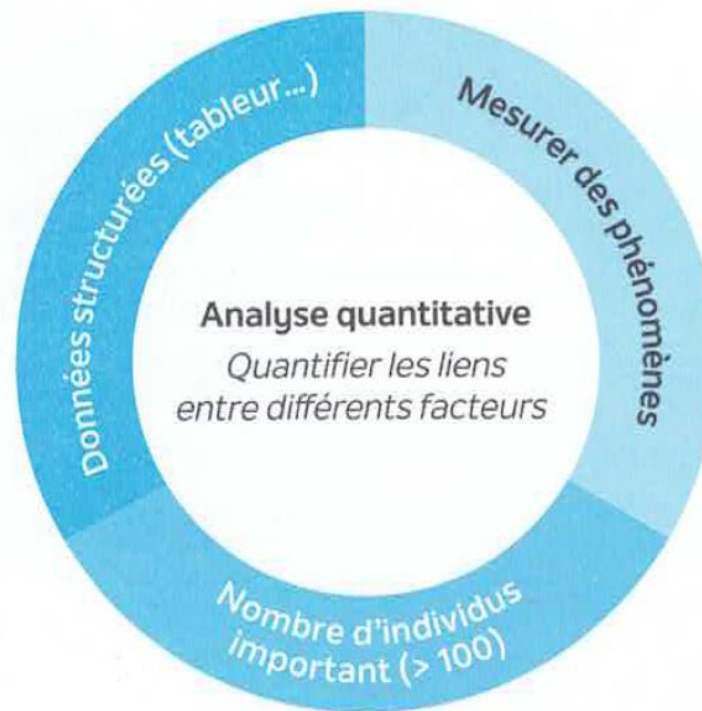
L'approche quantitative correspond à la mobilisation de données généralement structurées sur un nombre important d'individus (en général supérieur à 100). Elle est utile pour mesurer des phénomènes et quantifier des liens entre différents facteurs.

L'approche qualitative correspond à l'analyse de matériau généralement non structuré (texte, discours) sur un faible nombre d'individus (en général inférieur à 100). Elle est utile pour recueillir des perceptions complexes et comprendre en profondeur une situation donnée.

Il est parfois possible de combinées, dans certains cas, les deux méthodes.

CIE 110 : Approche quantitative ou qualitative ?

Le choix entre les deux approches



CIE 110 : Approche quantitative

Pourquoi l'utiliser ?



L'approche quantitative se fonde sur des bases de données généralement structurées (type tableau Excel) et sur un grand nombre d'individus (en général plus de 100 individus).

Elle permet de quantifier des phénomènes et de donner des résultats agrégés, sous forme de moyenne par exemple (moyenne des salaires dans une population), ou de pourcentages (X% des 18-25 ans connaît telle marque).

Elle peut ainsi s'avérer indispensable pour mesurer des variations au sein d'une population (les clients ayant moins de 30 ans achètent plus sur tel site Internet que les clients ayant plus de 50 ans).

CIE 110 : Approche qualitative

Pourquoi l'utiliser ?



L'approche qualitative permet généralement de recueillir des perceptions complexes, de comprendre en profondeur une situation ou un phénomène, ou d'étudier un terme émergent sur lequel on ne dispose d'aucune information ou hypothèse préalables.

Le mode de recueil de l'information (entretiens en face-à-face, observations) est plus adapté pour recueillir les perceptions complexes des individus, ou encore de comprendre les mécanismes relationnels à l'œuvre dans une situation donnée.

CIE 110 : Approche quantitative ou qualitative ?

Pense-bête



- ✓ L'approche quantitative est utile pour quantifier des phénomènes et comparer des populations.
- ✓ L'approche qualitative est utile pour recueillir des perceptions complexes et comprendre des situations locales en profondeur.

CIE 110 : Approche quantitative ou qualitative ?

Quiz



Résultat d'une analyse :

Certains lausannois se plaignent des travaux de rénovations qui perturbent le trafic du métro pendant des mois : « Malheureusement, ça fait 3 mois que le M1 ne s'arrête plus à la station EPFL. Je dois faire un détour de 30 minutes chaque jour. » (Monsieur Dupont).

Approche quantitative ou qualitative?

✓ Qualitative

Pour une analyse qui porterait sur l'efficacité du métro à Lausanne: l'auteur pourrait effectuer une étude qualitative en interrogeant les utilisateurs à travers un entretien, pour ensuite interpréter les résultats.

CIE 110 : Approche quantitative ou qualitative ?

Quiz



Résultat d'une analyse :

D'après notre sondage, 75 % des Lausannois sont insatisfaits. Ils ont répondu NON à la question « Etes-vous satisfait du M1 ? ».

Approche quantitative ou qualitative?

✓ Quantitative

Pour une même analyse sur le métro à Lausanne, l'auteur peut aussi entreprendre une étude quantitative. Il peut faire un sondage auprès des lausannois pour analyser le taux de satisfaction ou d'insatisfaction.

À noter : les résultats chiffrés de satisfaction révélés par une étude quantitative peuvent être complétés par l'analyse d'une étude qualitative qui posera des questions plus précises pour cerner le contentement ou le mécontentement des utilisateurs du métro.

CIE 110 : Approche quantitative ou qualitative ?

Quiz



Vous voulez savoir à quelle fréquence et dans quelles situations les recrues vont à l'armée Suisse.

Approche quantitative ou qualitative?

✓ Quantitative

Une approche quantitative est préférable (questionnaires).

CIE 110 : Approche quantitative ou qualitative ?

Quiz



Vous voulez savoir comment les recrues ont vécu leur début dans l'armée Suisse.
Approche quantitative ou qualitative?

✓ Qualitative

Une étude qualitative est plus appropriée (entretiens individuels).

CIE 110 : Que veut-on mesurer ?

La mobilisation d'une approche quantitative et des statistiques vise généralement à mesurer un objet. Cependant, cet objet peut être de plusieurs natures :



- ✓ Un phénomène.
- ✓ Un lien entre deux phénomènes.
- ✓ L'atteinte d'un ou de plusieurs objectifs.
- ✓ L'évolution d'un phénomène.

La mesure peut aussi concerner des êtres humains ou des objets.

Dans tous les cas, bien définir ce que l'on cherche à mesurer et quelles sont les implications pratiques et éthiques de cette mesure est essentiel.

CIE 110 : Que veut-on mesurer ?

Se poser les bonnes questions



Je cherche à mesurer :

- ✓ Une situation, une chose, un individu, un lien entre deux facteurs, un effet.
- ✓ Une situation à l'instant T, une évolution.
- ✓ Une restitution agrégée, une restitution individuelle.

Je connais les implications :

- ✓ Juridiques : protection des données à caractère personnel.
- ✓ Statistiques: mesure en %, en valeur absolue ou via des outils statistiques.
- ✓ Éthiques : questionnements éthiques sur l'effet de la mesure sur les individus.

CIE 110 : Que veut-on mesurer ?

Pourquoi l'utiliser ?



Les implications du type d'objet mesuré sont à la fois statistiques, juridiques et éthiques. Plusieurs textes encadrent la collecte de données sur des individus. Ainsi, le règlement général de protection des données (RGPD) encadre la collecte et le traitement de données à caractère personnel.

Les mesures qui concernent les êtres humains doivent donner lieu à des questionnements éthiques approfondis. Une mesure n'est généralement pas neutre, au sens où elle peut avoir des effets sur la personne concernée.

Par exemple, mesurer la performance d'un individu à un test donné peut conditionner son avenir professionnel si cette mesure est utilisée comme critère de recrutement.

CIE 110 : Que veut-on mesurer ?

Comment l'utiliser ?

1. Définir le ou les types d'objets mesurés. Voici quelques exemples :

- ✓ Mesurer une situation à un instant T (ex. : l'absentéisme, la fidélisation à une marque).
- ✓ Mesurer des caractéristiques humaines (ex. : la performance, la motivation).
- ✓ Mesurer un objet (ex. : la taille, le prix).
- ✓ Mesurer l'évolution d'une situation, d'un phénomène (ex. : l'évolution de la fidélisation à une marque).
- ✓ Mesurer le lien entre deux phénomènes ou deux variables (ex. : lien entre les caractéristiques socioprofessionnelles et l'absentéisme, ou entre les caractéristiques individuelles et l'achat de tel produit).
- ✓ Mesurer l'atteinte d'objectifs chiffrés (ex. : attente des objectifs de vente, des objectifs de parts de marché).
- ✓ Mesurer les effets d'une politique ou d'une action (ex. : effets d'une publicité sur les intentions d'achat).



CIE 110 : Que veut-on mesurer ?

Comment l'utiliser ?



2. La mesure peut aussi viser un objectif de restitution individuelle (ex. : mesurer la performance d'un individu) ou agrégée (ex. : indiquer la performance moyenne d'un groupe d'individus).

Ces différents types de mesure peuvent se combiner : il est ainsi possible de mesurer l'évolution de la performance moyenne d'un groupe d'individus.

3. Définir les implications statistiques, juridiques et éthiques associées.

CIE 110 : Que veut-on mesurer ?

Pense-bête



- ✓ Cherchez-vous à mesurer une situation à un instant T ou une évolution ?
- ✓ La mesure concerne-t-elle des individus ou des objets ?
- ✓ Visez-vous une restitution agrégée ou individuelle ?

CIE 110 : Quels sont les objectifs ?

La mobilisation d'une approche quantitative peut viser plusieurs objectifs : décrire, comprendre, expliquer, prédire. Ces objectifs correspondent en fait à quatre questions :

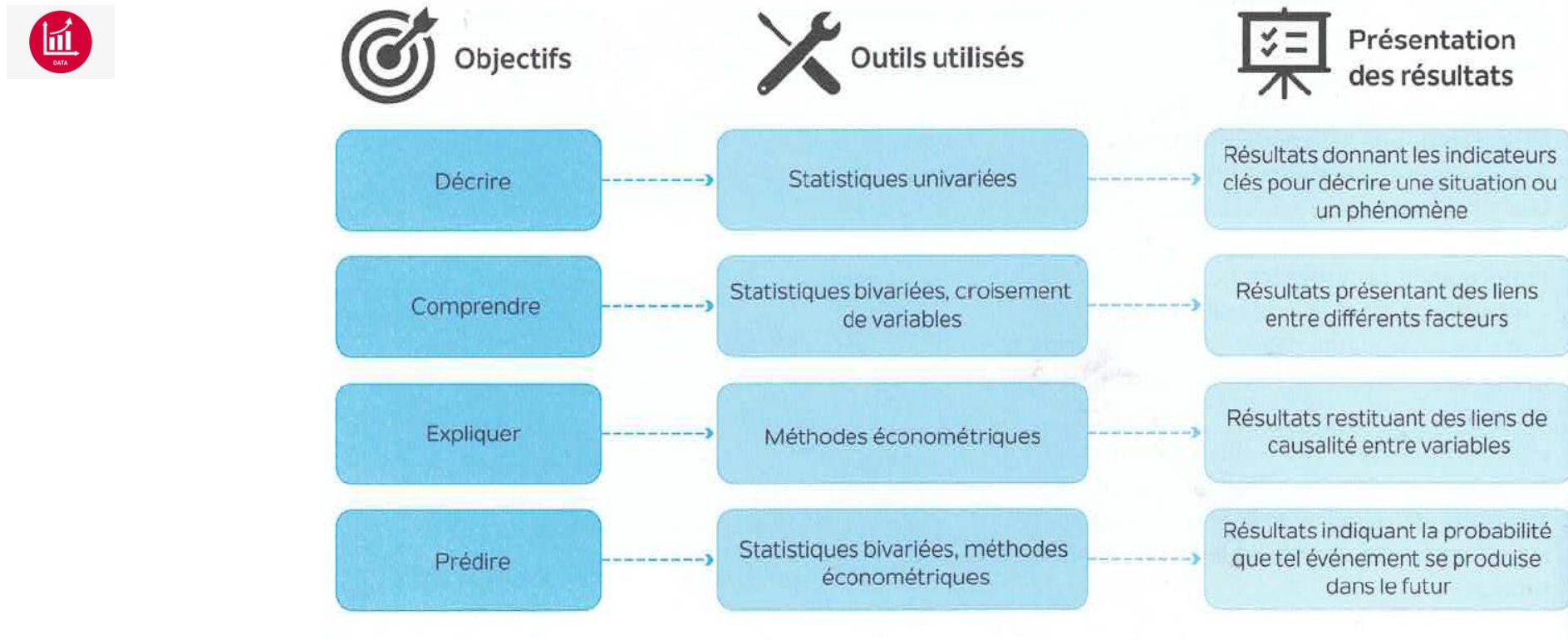


- ✓ Décrire : Quelle est la situation ?
- ✓ Comprendre : Qu'est-ce qui se joue dans cette situation ?
- ✓ Expliquer : Pourquoi la situation est-elle telle qu'elle est ?
- ✓ Prédire : Quelle sera la prochaine situation ?

Il est essentiel de bien définir le ou les objectifs poursuivis avant de commencer à mobiliser les outils statistiques. En effet, selon les objectifs choisis, les méthodologies utilisées ne seront pas les mêmes, ni la présentation des résultats.

CIE 110 : Quels sont les objectifs ?

Comment les objectifs conditionnent la démarche



CIE 110 : Quels sont les objectifs ?

Pourquoi l'utiliser ?



L'usage des statistiques peut viser la description d'un objet, d'une situation ou d'un phénomène. Il s'agit alors de répondre à la question « quelle est la situation ? ».

La compréhension de la situation vise à répondre à la question « qu'est-ce qui se joue dans la situation ? ».

Expliquer la situation revient à répondre à la question « pourquoi la situation est-elle telle qu'elle est ? ».

Enfin, prédire une situation revient à répondre à la question « quelle sera la prochaine situation ? »

CIE 110 : Quels sont les objectifs ?

Comment l'utiliser ?



1. Définir le ou les objectifs recherchés.
2. Définir les outils statistiques adéquats pour atteindre ces objectifs :
 - ✓ Pour décrire la situation : utiliser les statistiques univariées.
 - ✓ Pour comprendre la situation : utiliser les statistiques bivariées et l'étude de liens entre les variables ou encore l'analyse géométrique des données.
 - ✓ Pour expliquer la situation : utiliser les méthodes économétriques.
 - ✓ Pour prédire la situation : utiliser les statistiques bivariées et les méthodes économétriques. Ces outils et méthodes sont à mobiliser de manière différente que pour comprendre ou expliquer.

CIE 110 : Quels sont les objectifs ?

Comment l'utiliser ?



3. Prévoir un mode de restitution adapté :

- ✓ Pour décrire la situation : ex. : les effectifs de l'entreprise sont composés à X% de cadres et X% de non-cadres, le chiffre d'affaires de l'entité s'élève à X CHF, la moyenne d'âge des clients s'élève à X ans.
- ✓ Pour comprendre la situation : ex. : l'engagement au travail des salariés semble lié à la fois à la politique de l'entreprise en matière de qualité de vie au travail et d'équilibre vie privée-vie professionnelle et à des caractéristiques individuelles.
- ✓ Pour expliquer la situation : ex. : à profil identique ou à caractéristiques comparables, les personnes ayant vu telle publicité ont une probabilité d'achat du produit X% supérieure aux personnes n'ayant pas vu la publicité.
- ✓ Pour prédire la situation : ex. : tel client a une probabilité future d'achat de X%, ou encore, les ventes globales s'élèveront à X CHF.

CIE 110 : Quels sont les objectifs ?

Pense-bête



- ✓ Définissez bien l'objectif poursuivi et mobilisez les méthodes adaptées.
- ✓ Présentez vos résultats de façon à atteindre votre objectif, c'est-à-dire en répondant correctement à la question ou aux questions que vous vous posez.

CIE 110 : Quels sont les moyens ?

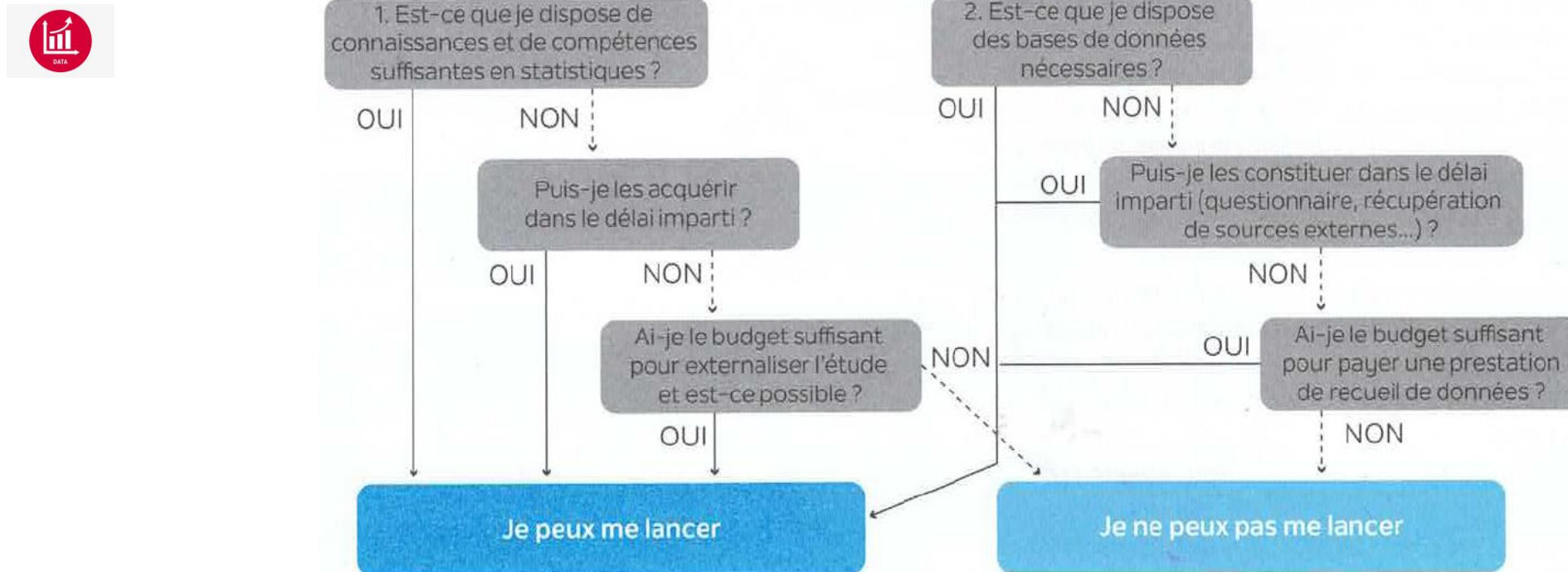
Avant d'entamer une démarche mobilisant des statistiques ou une approche quantitative, il convient de bien définir les moyens dont on dispose sur différents plans:



- ✓ Sur le plan des compétences et connaissances : de quelles connaissances et compétences je dispose sur les statistiques et les logiciels ?
Puis-je acquérir rapidement ces connaissances et compétences ?
- ✓ Sur le plan des données : de quelles bases de données je dispose ?
- ✓ Sur le plan financier : de quel budget je dispose ?

CIE 110 : Quels sont les moyens ?

Les moyens à vérifier avant de commencer



CIE 110 : Quels sont les moyens ?

Pourquoi l'utiliser ?



Les projets d'étude statistique peuvent s'avérer coûteux en temps et en argent. Autant donc s'assurer dès l'amont que les moyens nécessaires pour la bonne réussite du projet sont réunis.

Ces moyens sont les suivants :

- ✓ Compétences et connaissances dans le domaine statistique : maîtrise des notions statistiques de base, maîtrise des logiciels de statistiques, notamment.
- ✓ Bases de données adéquates, permettant de traiter la question posée. Ces données ne doivent pas avoir été agrégées au préalable. Des données agrégées limitent en effet les possibilités de croisement et d'analyse.

En cas d'absence de ces moyens, il peut être possible de rémunérer des prestations externes ou d'acheter des bases de données adaptées, mais cela peut représenter un budget important.

CIE 110 : Quels sont les moyens ?

Comment l'utiliser ?



1. Définir les moyens à disposition :

- ✓ Compétences et connaissances : si l'on ne dispose pas de compétences ou de connaissances statistiques, peut-être peut-on les acquérir ?
- ✓ Bases de données : si l'on ne dispose pas des bases de données adéquates, peut-être peut-on les constituer soi-même ?

2. Identifier les possibilités pour pallier l'absence de ces moyens, le cas échéant.

3. Savoir renoncer à un projet d'étude statistique si les conditions de réussite ne sont pas réunies.

CIE 110 : Quels sont les moyens ?

Pense-bête



- ✓ Évaluez vos compétences et connaissances en statistiques ainsi que votre maîtrise des logiciels statistiques. Si vos compétences ne sont pas suffisantes, prévoyez de faire appel à un prestataire.
- ✓ Cherchez les bases de données dont l'analyse pourrait vous permettre de répondre à la question ou aux questions que vous vous posez. Si vous ne disposez pas des données nécessaires, prévoyez de faire appel à un prestataire.
- ✓ Sachez renoncer si les conditions de réussite du projet d'étude ne sont pas réunies.

CIE 110 : Logiciels de statistiques

Il existe différents logiciels de statistiques, les plus courants étant :

- ✓ Excel
- ✓ R
- ✓ SAS
- ✓ SPSS
- ✓ Stata




Il est essentiel de bien connaître les caractéristiques de chacun avant de choisir lequel utiliser. Pour permettre un choix éclairé, il est possible de lister les caractéristiques sur les points suivants :

- ✓ Coût
- ✓ Ergonomie
- ✓ Facilité d'installation
- ✓ Possibilité de mener des analyses complexes
- ✓ Possibilité d'améliorer facilement l'aspect visuel des graphiques

CIE 110 : Logiciels de statistiques

Tableau comparatif des différents logiciels

	Excel	R	SAS	SPSS	STATA
 Coût	Achat de la licence Microsoft.	Gratuit	Achat de licence nécessaire.		
Facilité d'installation	Très facile.	Très facile sous Windows, plus difficile sous MacOS, notamment pour certains packages.	Très facile		
Ergonomie	Très ergonomique, aucune maîtrise du code nécessaire.	Nécessite une certaine maîtrise du code informatique ; possibilité de télécharger des modules (packages) permettant d'accéder à une interface plus conviviale.	Possibilité de l'utiliser sans le code informatique avec une interface spécifique et souvent plus coûteuse, mais la maîtrise du code reste recommandée pour les usages les plus sophistiqués.		
Analyse complexes	Limitées : possibilité d'acheter un module complémentaire (XLSTAT).	Quasiment illimitées (développement régulier de packages adaptés).	Possibles.		
Graphiques	Très faciles à personnaliser.	Difficiles à personnaliser, surtout en cas de non-maîtrise du code informatique.	Difficiles à personnaliser, surtout en cas de non-maîtrise du code informatique.		

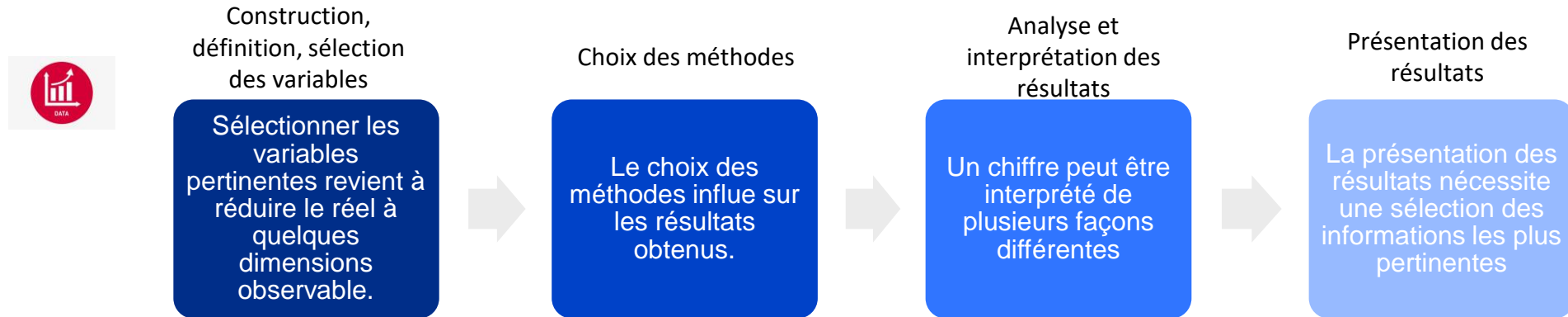
CIE 110 : Les limites de la quantification



Même si le discours ambiant sur l'objectivité des statistiques tend à le faire oublier, les statistiques reposent, tout autant que les études qualitatives, sur des choix humains, et donc par essence subjectifs, ce qui peut introduire des biais. Les choix humains et subjectifs lors des étapes essentielles d'une étude statistique influencent :

- ✓ La construction, la définition et la sélection des variables
- ✓ Le choix des méthodes
- ✓ L'analyse et l'interprétation des résultats
- ✓ La présentation des résultats

CIE 110 : Les limites de la quantification



CIE 110 : Les limites de la quantification

Pourquoi l'utiliser ?



Nous vivons dans une société qui accorde un crédit important aux chiffres, aux statistiques, à tout ce qui est de l'ordre de la quantification.

Les chiffres et les statistiques sont vus comme des garanties d'objectivité et de scientificité. Cependant, ce crédit repose sur le mythe d'une quantification neutre et d'une statistique qui ne serait que le reflet de la réalité.

De nombreux travaux de recherche ont remis en cause ce mythe, en montrant par exemple les différents biais associés à la quantification, ou encore l'importance des choix subjectifs dans les études statistiques.

CIE 110 : Les limites de la quantification

Comment l'utiliser ?



1. À chaque étape de l'analyse, prendre conscience des choix effectués : La construction, la définition et la sélection des variables

- ✓ Construction, définition et sélection des variables : cette étape représente toujours une réduction de la complexité du réel, puisqu'il s'agit de « mettre en données » des phénomènes ou des objets.
De ce fait, cette étape conduit à rendre plus visibles certains éléments et à en masquer d'autres.
- ✓ Choix des méthodes : cette étape est tout sauf neutre. Une analyse « toutes choses égales par ailleurs » ne donnera pas du tout les mêmes résultats qu'une analyse plus descriptive - les résultats peuvent s'avérer apparemment contradictoires.

CIE 110 : Les limites de la quantification

Comment l'utiliser ?



2. Identifier l'influence des choix effectués sur les résultats.

Un même chiffre peut donner lieu à une multiplicité d'interprétations. Par exemple, les interprétations « La majorité des Suisses sont contre cette proposition de loi » et « La moitié des Suisses sont pour cette proposition de loi » peuvent toutes les deux être prononcées à partir du chiffre « 52 % des Suisses sont contre cette proposition de loi ».

Or, ces deux interprétations semblent s'opposer et ne produiront pas du tout le même effet chez l'interlocuteur.

CIE 110 : Les limites de la quantification

Comment l'utiliser ?



3. Questionner l'influence des résultats présentés sur les actions des interlocuteurs.

Cette dernière étape se fonde toujours sur la sélection des résultats les plus pertinents. Les choix opérés ne sont donc pas neutres, puisqu'ils mettent en avant certains points au détriment d'autres.

CIE 110 : Les limites de la quantification

Pense-bête



- ✓ Prenez bien conscience du fait que l'usage de statistiques n'est pas une garantie d'objectivité.
- ✓ Prenez également conscience de l'influence des choix méthodologiques effectués sur les résultats finaux.