**📐 1. Statistique descriptive**

Résumer, visualiser, comprendre les données.

* **1. Types de variables en statistique**
* Avant de calculer quoi que ce soit, il faut **identifier le type de variable**, car c’est cela qui **détermine** les outils statistiques à utiliser.

| **Type de variable** | **Sous-type** | **Exemples** |
| --- | --- | --- |
| **Qualitative (catégorielle)** | Nominale | Sexe, région, profession |
|  | Ordinale | Niveau d’éducation (primaire, secondaire…), satisfaction |
| **Quantitative (numérique)** | Discrète | Nombre d’enfants, consultations |
|  | Continue | Taille, poids, âge, revenu |

* **📊 2. Quand calculer les indicateurs, selon le type de variable**

| **Indicateur** | **À utiliser avec...** | **Quand ? (Objectif)** |
| --- | --- | --- |
| **Mode** | Qualitative ou quantitative | Trouver la valeur la plus fréquente |
| **Médiane** | Quantitative, ou qualitative ordinale | Quand il y a des valeurs extrêmes ou une distribution asymétrique |
| **Moyenne** | Quantitative continue ou discrète | Quand les données sont symétriques et sans valeurs extrêmes |
| **Étendue** | Quantitative | Pour connaître la variation brute (max – min) |
| **Variance / Écart-type** | Quantitative | Pour mesurer la dispersion autour de la moyenne |
| **Coefficient de variation (CV)** | Quantitative | Pour comparer la dispersion entre deux séries de données |
| **Fréquences (%)** | Tous types | Pour résumer la répartition d’une variable |

* **🎯 Exemple pratique**
* **Variable : Âge des femmes enceintes** (quantitative continue)  
  ✔️ On peut calculer : moyenne, médiane, écart-type, étendue
* **Variable : Région de résidence** (qualitative nominale)  
  ✔️ On peut calculer : mode, fréquence, pourcentage
* **Variable : Niveau d’éducation** (qualitative ordinale)  
  ✔️ On peut calculer : médiane, mode, fréquence
* Variance, écart-type, coefficient de variation
* Diagrammes (boîtes à moustaches, histogrammes, barres, camemberts)
* Tableaux croisés et distributions

**📊 1. Diagrammes (Représentations graphiques)**

Les diagrammes permettent de **visualiser rapidement** la structure d'une variable ou d'une relation entre plusieurs variables.  
Le type de graphique dépend du **type de variable**.

**🧩 A. Pour une variable qualitative**

| **Diagramme** | **À utiliser quand…** | **Exemple** |
| --- | --- | --- |
| **Diagramme en barres** | Variable nominale ou ordinale (sexe, région) | Répartition des accouchements par région |
| **Diagramme circulaire (camembert)** | Répartition en pourcentage | % de femmes par niveau d’éducation |

🔎 Remarque : **le diagramme en barres est préféré** au camembert pour une meilleure lisibilité.

**📏 B. Pour une variable quantitative**

| **Diagramme** | **À utiliser pour…** | **Exemple** |
| --- | --- | --- |
| **Histogramme** | Répartition des effectifs par classe | Répartition des âges des femmes |
| **Polygone de fréquence** | Pour visualiser les tendances de distribution | Évolution du nombre de visites prénatales |
| **Boîte à moustaches (boxplot)** | Voir la dispersion, les médianes et les valeurs extrêmes | Comparaison de l’âge par région |

**📌 Exemple simple : âge des femmes (quantitative continue)**

* **Histogramme** : montre si l’âge est concentré autour d’une valeur.
* **Boxplot** : montre si certaines femmes sont beaucoup plus jeunes ou plus âgées que la moyenne.

**🧮 2. Tableaux croisés (ou tableaux de contingence)**

**📚 Définition :**

Un **tableau croisé** permet de croiser **deux variables qualitatives** ou une qualitative avec une quantitative discrète.  
Il montre comment les modalités d'une variable **se répartissent en fonction** d'une autre.

**🧱 Structure :**

|  | **Variable B : Modalités →** | **TOTAL** |
| --- | --- | --- |
| **Variable A** | B1 | B2 |
| A1 | 10 | 5 |
| A2 | 4 | 6 |
| TOTAL | 14 | 11 |

On peut ensuite **ajouter des pourcentages** en ligne ou en colonne.

**🎯 Quand l’utiliser :**

* Croiser le **niveau d’éducation** avec le **type d’accouchement** (assisté ou non)
* Croiser **lieu de résidence** avec **nombre de visites prénatales**
* Croiser **âge** (en classe) avec **recours aux soins**

**💡 Avantage :**

Le tableau croisé permet de **repérer rapidement une association** (ou non) entre deux variables.

**📌 Résumé : Quand utiliser quoi ?**

| **Type de données** | **Représentation recommandée** |
| --- | --- |
| 1 variable qualitative | Barres, camembert |
| 1 variable quantitative | Histogramme, boxplot |
| 2 variables qualitatives | Tableau croisé + barres/groupées |
| 1 qualitative + 1 quantitative | Boxplot par groupe, moyenne par modalité |

🎯 **Objectif** : savoir bien **résumer une base de données** avant d’aller plus loin.

**🧪 2. Statistique inférentielle**

Tirer des conclusions sur une population à partir d’un échantillon.

* Notion de population/échantillon
* Estimation (ponctuelle et par intervalle de confiance)
* Tests d’hypothèse : Student, chi², ANOVA, etc.
* P-value et erreurs de type I/II

🎯 **Objectif** : savoir **prouver ou réfuter** une hypothèse avec rigueur.

**2. Concepts de base à comprendre**

| **Concept** | **Explication simple** |
| --- | --- |
| **Population** | Ensemble total des individus (ex. toutes les femmes enceintes au Burkina Faso) |
| **Échantillon** | Sous-ensemble choisi pour l’étude |
| **Paramètre** | Valeur vraie dans la population (souvent inconnue) |
| **Estimateur** | Valeur calculée à partir de l’échantillon (ex. moyenne, proportion) |
| **Erreur d’échantillonnage** | Différence entre l’estimateur et le paramètre |
| **Intervalle de confiance** | Intervalle dans lequel on estime que le paramètre se trouve avec un certain niveau de confiance (souvent 95 %) |
| **Test d’hypothèse** | Méthode pour décider si une différence observée est statistiquement significative |

**📊 3. Principaux outils de l’inférence**

| **Outil** | **Quand l’utiliser** |
| --- | --- |
| **Estimation d’une moyenne** | Âge moyen, revenu moyen… |
| **Estimation d’une proportion** | % de femmes ayant accouché dans un centre |
| **Test de Student (t-test)** | Comparer deux moyennes |
| **Test du chi²** | Vérifier l’association entre deux variables qualitatives |
| **Test de proportion** | Comparer deux proportions |
| **Intervalle de confiance** | Délimiter une zone plausible pour un paramètre |
| **Régression linéaire/logistique** | Prédire une variable en fonction d’autres |

**🧠 Exemple simple**

**Question** : Le taux d’accouchement assisté est-il plus élevé chez les femmes instruites que chez les femmes non instruites ?

* **Population** : femmes enceintes du Burkina Faso
* **Échantillon** : enquête nationale (EDS, MICS, etc.)
* **Variable 1** : niveau d’instruction
* **Variable 2** : recours à l’accouchement assisté (oui/non)
* **Test à utiliser** : **test du chi²** (2 variables qualitatives)

**🚨 Attention : les limites**

* Toute **inférence comporte une incertitude**
* Il faut **respecter les conditions d’application** des tests (normalité, indépendance…)
* Ne pas **confondre** signification statistique et importance pratique

**📊 3. Analyse de régression**

Étudier les relations entre variables.

* Régression linéaire simple et multiple
* Régression logistique (binaire/multinomiale)
* Interprétation des coefficients
* Résidus, multicolinéarité, R², AIC/BIC

**1. Qu’est-ce que la régression ?**

La régression est une méthode statistique qui permet de :

| **But** | **Exemple (mémoire)** |
| --- | --- |
| **Modéliser une relation** entre une **variable dépendante** (à expliquer) et une ou plusieurs **variables indépendantes** (explicatives) | Expliquer l'accouchement assisté par l'âge, le niveau d'instruction, la région… |
| **Prédire** la valeur d’une variable | Prédire la probabilité d’avoir un accouchement assisté |
| **Mesurer l’effet** d’une variable en **contrôlant les autres** | Quel est l’effet net du niveau d’éducation sur le recours aux soins ? |

**🧩 2. Types de régression**

| **Type de régression** | **Quand l’utiliser ?** | **Variable dépendante (Y)** |
| --- | --- | --- |
| **Régression linéaire simple** | 1 variable explicative | Quantitative (ex. âge) |
| **Régression linéaire multiple** | Plusieurs variables explicatives | Quantitative |
| **Régression logistique binaire** | Pour prédire une variable **binaire** (oui/non) | Qualitative (0/1) |
| **Régression logistique multinomiale** | Variable dépendante avec **plus de 2 catégories** (sans ordre) | Qualitative nominale |
| **Régression ordinale** | Variable dépendante ordonnée | Qualitative ordinale |

**🎯 Exemples concrets :**

1. **Régression linéaire**  
    Y = âge de la femme  
    X = nombre de CPN  
    → *Plus une femme a de CPN, plus son âge augmente-t-il ?*
2. **Régression logistique binaire**  
    Y = accouchement assisté (1 = oui, 0 = non)  
    X = niveau d’instruction, lieu de résidence, parité  
    → *Quels facteurs influencent le recours à un accouchement assisté ?*

**⚙️ 3. Interprétation des résultats**

**➤ Pour une régression linéaire :**

* Coefficient positif : l’augmentation de X entraîne une **augmentation de Y**
* Coefficient négatif : l’augmentation de X entraîne une **diminution de Y**
* R² = pourcentage de la variance de Y expliqué par les X

**➤ Pour une régression logistique :**

* On interprète les **odds ratios (OR)** :
  + OR > 1 : la variable augmente la probabilité de l'événement
  + OR < 1 : la variable diminue cette probabilité
* Ex. : OR = 2,3 pour l’instruction secondaire → une femme instruite a **2,3 fois plus de chances** d’accoucher assistée

**⚠️ Conditions importantes**

Avant d’interpréter une régression, vérifier :

* **Linéarité** (pour les modèles linéaires)
* **Indépendance** des observations
* **Pas de multicolinéarité** (variables explicatives fortement corrélées entre elles)
* **Bonne spécification du modèle**

**📦 Résumé : à quoi sert chaque régression ?**

| **Objectif** | **Type de régression à utiliser** |
| --- | --- |
| Expliquer une variable **quantitative** | Linéaire (simple ou multiple) |
| Expliquer une variable **binaire** | Logistique binaire |
| Expliquer une variable **nominale** | Logistique multinomiale |
| Expliquer une variable **ordinale** | Logistique ordinale |

🎯 **Objectif** : savoir **modéliser une variable cible à partir de prédicteurs**.

**🎯 4. Échantillonnage et plan d'enquête**

Obtenir des données représentatives et fiables.

* Techniques : aléatoire simple, stratifié, en grappes
* Calcul de taille d’échantillon
* Biais d’échantillonnage, poids et redressement

🎯 **Objectif** : **concevoir une enquête** ou une étude statistique solide.

**📘 5. Probabilités**

La base logique de la statistique.

* Événements, espaces, probabilités conditionnelles
* Lois : binomiale, normale, Poisson, exponentielle…
* Espérance, variance, loi des grands nombres, théorème central limite

**📚 1. Qu’est-ce qu’une probabilité ?**

La **probabilité** est un **nombre compris entre 0 et 1** (ou entre 0 % et 100 %) qui mesure **la chance qu’un événement se produise**.

| **Probabilité** | **Interprétation** |
| --- | --- |
| 0 | Impossible |
| 1 | Certain |
| 0,5 | Équiprobable (ex. pile ou face) |

**🧩 2. Vocabulaire de base**

| **Terme** | **Définition** | **Exemple** |
| --- | --- | --- |
| **Expérience aléatoire** | On ne peut pas prédire le résultat | Tirer une carte au hasard |
| **Événement** | Résultat ou ensemble de résultats | Obtenir une carte rouge |
| **Univers** | Ensemble de tous les résultats possibles | Toutes les cartes du jeu |

**🔢 3. Calcul de la probabilité**

Si tous les cas sont **équiprobables** :

P(eˊveˊnement)=nombre de cas favorablesnombre de cas possiblesP(\text{événement}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}P(eˊveˊnement)=nombre de cas possiblesnombre de cas favorables​

**Exemple :**

* Univers = {1, 2, 3, 4, 5, 6}
* P(obtenir un nombre pair) = 3/6 = 0,5

**⚙️ 4. Règles fondamentales**

**➤ A. Règle de l’union (ou OU)**

Si deux événements A et B sont **incompatibles** (ne peuvent pas arriver en même temps) :

P(A ou B)=P(A)+P(B)P(A \text{ ou } B) = P(A) + P(B)P(A ou B)=P(A)+P(B)

**➤ B. Règle de l’intersection (et)**

Si A et B sont **indépendants** (l’un n’influence pas l’autre) :

P(A et B)=P(A)×P(B)P(A \text{ et } B) = P(A) \times P(B)P(A et B)=P(A)×P(B)

**🧠 5. Probabilité conditionnelle**

C’est la probabilité que **B arrive sachant que A est déjà arrivé** :

P(B∣A)=P(A∩B)P(A)P(B | A) = \frac{P(A \cap B)}{P(A)}P(B∣A)=P(A)P(A∩B)​

**Exemple santé :**

* P(accouchement assisté | femme instruite) = probabilité qu’une femme instruite ait accouché avec assistance.

**🧮 6. Loi des probabilités :**

| **Loi** | **Quand l’utiliser** | **Exemple** |
| --- | --- | --- |
| **Loi uniforme** | Cas équiprobables | Tirage au sort |
| **Loi binomiale** | Nombre de succès dans plusieurs essais | Nombre de femmes ayant eu ≥4 CPN |
| **Loi normale** | Phénomènes naturels (taille, poids, âge) | Distribution des âges des femmes enceintes |

**📌 Application à ton mémoire (santé)**

| **Situation** | **Type de probabilité** |
| --- | --- |
| Quelle est la chance qu'une femme ait au moins 4 CPN ? | Proportion ou estimation |
| Quelle est la probabilité qu’une femme instruite accouche dans un centre de santé ? | Probabilité conditionnelle |
| Quelle est la chance d’avoir ≥3 femmes sur 5 ayant eu un accouchement assisté ? | Loi binomiale |

**✅ Résumé**

| **Élément** | **Ce qu’il faut retenir** |
| --- | --- |
| Probabilité = mesure d’incertitude | Entre 0 et 1 |
| Probabilité conditionnelle | Ce qu’on cherche *sachant* une condition |
| Probabilités combinées | Avec des règles (addition, multiplication) |
| Utilisation | Base des tests, de la régression, de l’inférence |

🎯 **Objectif** : comprendre **les distributions derrière les méthodes** statistiques.

**📦 6. Analyse multivariée**

Travailler avec plusieurs variables simultanément.

* ACP, AFC, ACM
* Classification (hiérarchique, K-means)
* Analyse discriminante, clustering
* Corrélations multiples

**📚 1. Définition**

L’**analyse multivariée** désigne l’ensemble des méthodes statistiques utilisées pour **analyser simultanément plusieurs variables** (généralement plus de deux), dans le but de :

| **Objectif** | **Exemple** |
| --- | --- |
| Résumer l'information | Résumer les facteurs socio-démographiques |
| Détecter des relations | Voir si l’éducation, l’âge et la résidence influencent le recours aux soins |
| Classer ou regrouper des individus | Regrouper les femmes selon leur profil d’utilisation des soins |
| Expliquer une variable cible à partir de plusieurs variables | Expliquer l’accouchement assisté par les caractéristiques des femmes |

**🧩 2. Quand l’utiliser ?**

Tu utilises l’analyse multivariée quand :

* Tu as **plusieurs variables explicatives** (X1, X2, X3…)
* Et **une ou plusieurs variables à expliquer** (Y)

➡️ Exemple :  
**Y = recours à l’accouchement assisté**  
**X = âge, instruction, région, parité, nombre de CPN**

**🧪 3. Principales méthodes d’analyse multivariée**

Voici les grandes familles avec leur utilité :

| **Méthode** | **Objectif** | **Variable dépendante** |
| --- | --- | --- |
| **Régression linéaire multiple** | Prédire une variable quantitative | Quantitative |
| **Régression logistique multiple** | Prédire une variable binaire (oui/non) | Binaire (0/1) |
| **Analyse en composantes principales (ACP)** | Réduire les dimensions, synthèse | Aucune (exploratoire) |
| **Analyse factorielle des correspondances (AFC)** | Explorer les relations entre variables qualitatives | Aucune (exploratoire) |
| **Analyse discriminante** | Classer les individus selon des groupes | Qualitative |
| **Analyse de clusters (classification)** | Regrouper des individus similaires | Aucune (non supervisée) |
| **Régression multinomiale / ordinale** | Variable qualitative à plusieurs modalités | Nominale / Ordinale |

**💡 4. Exemple concret lié à ton mémoire**

**🎯 Objectif :**

Identifier les **facteurs associés au recours à l’accouchement assisté**

| **Méthode** | **Pourquoi c’est adapté** |
| --- | --- |
| **Régression logistique multiple** | Parce que la variable dépendante est binaire (assisté = oui/non) |
| **ACP ou AFC** | Pour explorer les profils de femmes (âge, éducation, résidence, etc.) |
| **Analyse de cluster** | Pour créer des profils-types de femmes (ex. : "femme urbaine instruite", "femme rurale non scolarisée", etc.) |

**📌 5. Étapes d’une analyse multivariée**

1. **Préparation des données**
   * Vérification des valeurs manquantes
   * Recodage des variables
   * Centrage et réduction (si nécessaire)
2. **Choix de la méthode**
   * Selon le type de variable dépendante
3. **Lancement de l’analyse avec logiciel**
   * Excel (limité), SPSS, Stata, R, Python…
4. **Interprétation des résultats**
   * Coefficients, p-values, odds ratios, inertie, axes, regroupements
5. **Validation du modèle**
   * Significativité, ajustement, colinéarité, tests d’ajustement

**📊 6. Résumé**

| **Ce que tu veux faire** | **Méthode** |
| --- | --- |
| Expliquer un phénomène binaire (oui/non) | Régression logistique |
| Explorer la structure de plusieurs variables quantitatives | ACP |
| Étudier des associations entre variables qualitatives | AFC |
| Créer des groupes d’individus similaires | Clustering |
| Comparer des profils | Analyse discriminante |

**✅ Et après ?**

Souhaites-tu :

* Un **exemple complet pas à pas** avec des données fictives ?
* Commencer par la **régression logistique multiple** (la plus utilisée en santé publique) ?
* Un **fichier Excel ou SPSS** pour tester ensemble une analyse ?

Je peux aussi t'aider à appliquer ça directement sur des données réelles comme celles de l’**EDS** ou du **RGPH**, si tu en as.

🎯 **Objectif** : **explorer, réduire, segmenter** des données complexes.

**💻 7. Compétences logicielles**

Passer de la théorie à l’analyse réelle.

* **Excel** : traitement de base
* **SPSS / Stata** : statistiques sociales et santé
* **R / Python** : analyses avancées, visualisation, automatisation
* **Power BI / Tableau** : tableaux de bord, data viz

🎯 **Objectif** : savoir **manipuler, analyser et présenter** les données.

**💼 8. Éthique et communication des résultats**

Être un(e) statisticien(ne) responsable et utile.

* Anonymisation, consentement, biais
* Clarté dans l’interprétation des résultats
* Adapter les résultats au public (décideurs, chercheurs, non-experts)

🎯 **Objectif** : **diffuser des résultats fiables et utiles** à la prise de décision.

Lorsque nous recevons une base de données brute issue d’une enquête ou d’un recueil de terrain, quels sont les premiers réflexes qu’un statisticien doit avoir ? Comment passer de ces données brutes à des indicateurs statistiques clairs et pertinents pour l’analyse ?

**Comprendre la source et le contexte**

* Quelle est la **finalité de l’enquête** ?
* Quels types de **variables** sont collectées (quantitatives, qualitatives, dates, etc.) ?
* Quelle est l’**unité statistique** (femme, ménage, accouchement…) ?

**🧹 2. Nettoyer la base**

| **Tâche** | **Pourquoi** |
| --- | --- |
| Supprimer les doublons | Éviter le biais |
| Identifier les valeurs manquantes | Pour décider : supprimer ou imputer |
| Vérifier les formats de variables | Nombres, dates, codes texte... |
| Vérifier les incohérences | Ex. : une femme ayant 0 grossesse mais ayant accouché |

**🗂 3. Recoder / transformer certaines variables**

* Transformer les **textes** en **codes** (ex. : "Oui"/"Non" → 1/0)
* Regrouper des modalités (ex. âge en tranches : 15–19, 20–34, 35–49)
* Créer des **variables dérivées** (ex. : nombre total de CPN ≥ 4 → indicateur "CPN complète")

**🧮 II. Construire ou reformuler les indicateurs statistiques**

Un **indicateur** est une **variable synthétique** qui permet de mesurer une situation de manière claire et comparable.

**📌 1. Définir l’objectif de l’indicateur**

Ex. : mesurer le **recours aux soins prénatals complets** (au moins 4 visites)

**🧪 2. Identifier les variables de base nécessaires**

Ex. : Variable nb\_cpn = nombre de consultations prénatales déclarées

**🔁 3. Créer l’indicateur par transformation ou recodage**

| **Exemple d’indicateur** | **Construction** |
| --- | --- |
| **CPN complète** | CPN\_4 = 1 si nb\_cpn ≥ 4, sinon 0 |
| **Accouchement assisté** | acc\_assiste = 1 si accouchement dans une structure de santé ou assisté par personnel qualifié |
| **Inégalité de recours** | Comparaison entre milieux urbain/rural ou entre niveaux d’instruction |

**📊 4. Calculer les statistiques descriptives**

* **Taux ou proportions** (ex. : % de femmes ayant eu ≥4 CPN)
* **Ratios** (ex. : accouchement assisté / total des accouchements)
* **Croisements** par région, par tranche d’âge, par niveau d’éducation…

**1. Analyse descriptive (statistique de base)**

But : décrire les caractéristiques générales de la population.

**📊 Pour les variables quantitatives :**

* revenu\_mensuel, nombre\_enfants, consultation\_prenatale
* **Moyenne**, **médiane**, **écart-type**, **min**, **max**, **quartiles**
* Histogrammes, boxplots

**📋 Pour les variables qualitatives :**

* gender, niveau\_instruction, statu\_matrimonial, zone\_residense, lieu\_accouchement, autonomie\_decisionnelle, etc.
* **Tableaux de fréquences** (valeurs absolues et pourcentages)
* Graphiques en barres ou camemberts

**✅ 2. Analyse bivariée (relations entre deux variables)**

**📌 a. Analyse qualitative vs qualitative**

* Exemples :
  + niveau\_instruction vs autonomie\_decisionnelle
  + zone\_residense vs lieu\_accouchement
  + assurance\_santé vs consultation\_prenatale (regroupée en classes)
* **Test du Chi² d’indépendance**
* **Tableaux croisés** + **graphes empilés** ou **groupés**

**📌 b. Quantitative vs qualitative**

* Exemples :
  + revenu\_mensuel selon autonomie\_decisionnelle
  + nombre\_enfants selon niveau\_instruction
* **Boxplots** ou **violin plots**
* **Test de comparaison de moyennes** (ANOVA, t-test)

**📌 c. Quantitative vs quantitative**

* revenu\_mensuel vs nombre\_enfants ou consultation\_prenatale
* **Nuages de points (scatterplots)**
* **Corrélation (Pearson, Spearman)**

**✅ 3. Analyse avancée (modélisation, régression)**

**📌 a. Régression logistique**

* Si tu veux prédire une variable binaire (ex. : autonomie\_decisionnelle = élevée vs faible)
* Variables explicatives : niveau\_instruction, zone\_residense, assurance\_santé, etc.

**📌 b. Régression linéaire**

* Pour prédire un score ou une valeur continue (ex. : consultation\_prenatale)

**📌 c. Analyse de correspondances multiples (ACM)**

* Pour explorer les **relations entre plusieurs variables qualitatives**
* Très utile si tu veux faire une **typologie** des femmes

**📌 d. Clustering / segmentation**

* Segmenter les femmes selon leurs caractéristiques (instruction, autonomie, revenus, etc.)

**Compétences techniques/statistiques**

1. **Analyse exploratoire des données (EDA)**
   * Nettoyage, traitement, et visualisation de données
   * Résumés statistiques (moyenne, médiane, écart-type, etc.)
   * Détection des valeurs aberrantes et traitement des données manquantes
2. **Méthodes statistiques fondamentales**
   * Régressions (linéaire, logistique…)
   * Tests d’hypothèses (t-test, chi², ANOVA…)
   * Intervalles de confiance, distributions, etc.
3. **Modélisation avancée (si applicable)**
   * Séries temporelles
   * Analyse multivariée
   * Méthodes bayésiennes
   * Machine learning (selon ton stage)

**💻 Compétences informatiques / outils**

1. **Logiciels et langages**
   * **R** ou **Python** (essentiel pour le traitement et la modélisation)
   * **Excel** (très utilisé en entreprise pour des analyses simples)
   * **SAS**, **SPSS**, **Stata** (si utilisés dans ton secteur)
2. **Manipulation de données**
   * Pandas (Python), dplyr / tidyr (R)
   * SQL pour interroger des bases de données
3. **Visualisation de données**
   * ggplot2 (R), matplotlib / seaborn / plotly (Python)
   * Tableau, Power BI (pour les tableaux de bord dynamiques)

**📈 Compétences en communication de données**

* **Savoir interpréter et vulgariser les résultats** (expliquer à un non-statisticien)
* **Créer des rapports clairs et visuellement attrayants**
* Présenter des résultats lors de réunions ou dans des slides PowerPoint

**🤝 Compétences professionnelles générales**

* **Rigueur et souci du détail**
* **Capacité à travailler en équipe** (surtout avec des non-statisticiens)
* **Autonomie et capacité d’auto-apprentissage**
* **Respect des délais et organisation**

**✅ Exemple de choses concrètes à montrer pendant ton stage**

* Ton **code propre et bien commenté**
* Des **analyses rigoureuses et reproductibles**
* Des **rapports ou présentations claires**
* Des propositions d’amélioration ou d’approfondissement d’analyses existantes
* Une **bonne compréhension du contexte métier** de tes analyses