

Séance 4 : Relier des variables entre elles

De la description à l'explication

Mattéo Lanoë

Collège universitaire de Sciences Po, Campus de Poitiers

Printemps 2026

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.
- ▶ Tendances centrale (Moyenne, Médiane, Mode).

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.
- ▶ Tendance centrale (Moyenne, Médiane, Mode).
- ▶ Dispersion (Écart-type, Étendue).

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.
- ▶ Tendance centrale (Moyenne, Médiane, Mode).
- ▶ Dispersion (Écart-type, Étendue).

Aujourd'hui (Séance 4) :

- ▶ Relier **deux** variables entre elles.

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.
- ▶ Tendance centrale (Moyenne, Médiane, Mode).
- ▶ Dispersion (Écart-type, Étendue).

Aujourd'hui (Séance 4) :

- ▶ Relier **deux** variables entre elles.
- ▶ Passer de la description à l'**explication**.

Introduction : Où en sommes-nous ?

La semaine dernière (Séance 3) :

- ▶ Décrire **une** variable seule.
- ▶ Tendance centrale (Moyenne, Médiane, Mode).
- ▶ Dispersion (Écart-type, Étendue).

Aujourd'hui (Séance 4) :

- ▶ Relier **deux** variables entre elles.
- ▶ Passer de la description à l'**explication**.
- ▶ Choisir le bon outil selon la nature des variables.

1. Pourquoi relier des variables ?

L'objectif est d'examiner le lien entre une chose qu'on observe et une chose qui l'explique.

1. Pourquoi relier des variables ?

L'objectif est d'examiner le lien entre une chose qu'on observe et une chose qui l'explique.

Vocabulaire essentiel

1. **Variable Dépendante, Y (À expliquer)** : C'est le comportement social étudié.
Ex : Le vote, le salaire, la survie dans le Titanic.

1. Pourquoi relier des variables ?

L'objectif est d'examiner le lien entre une chose qu'on observe et une chose qui l'explique.

Vocabulaire essentiel

1. **Variable Dépendante, Y (À expliquer)** : C'est le comportement social étudié.
Ex : Le vote, le salaire, la survie dans le Titanic.
2. **Variable Indépendante, X (Explicative)** : C'est la cause supposée (souvent socio-démographique).
Ex : L'âge, le genre, la classe sociale.

1. Pourquoi relier des variables ?

L'objectif est d'examiner le lien entre une chose qu'on observe et une chose qui l'explique.

Vocabulaire essentiel

1. **Variable Dépendante, Y (À expliquer)** : C'est le comportement social étudié.
Ex : Le vote, le salaire, la survie dans le Titanic.
2. **Variable Indépendante, X (Explicative)** : C'est la cause supposée (souvent socio-démographique).
Ex : L'âge, le genre, la classe sociale.

1. Pourquoi relier des variables ?

L'objectif est d'examiner le lien entre une chose qu'on observe et une chose qui l'explique.

Vocabulaire essentiel

1. **Variable Dépendante, Y (À expliquer)** : C'est le comportement social étudié.
Ex : Le vote, le salaire, la survie dans le Titanic.
2. **Variable Indépendante, X (Explicative)** : C'est la cause supposée (souvent socio-démographique).
Ex : L'âge, le genre, la classe sociale.

Exemple : Si je veux expliquer le vote (Dépendante), je vais regarder l'âge ou la catégorie sociale (Indépendantes).

1.1 Le sens de la relation : Qui explique qui ?

Avant de lancer un calcul, il faut poser une hypothèse sociologique. On distingue deux types de variables selon leur rôle :

1.1 Le sens de la relation : Qui explique qui ?

Avant de lancer un calcul, il faut poser une hypothèse sociologique. On distingue deux types de variables selon leur rôle :

1. La Variable DÉPENDANTE, Y

(Celle qu'on veut expliquer)

C'est le phénomène, le comportement ou l'opinion qui varie.

→ C'est l'**EFFET**.

Ex : Le vote, le salaire, la réussite au bac, la survie.

1.1 Le sens de la relation : Qui explique qui ?

Avant de lancer un calcul, il faut poser une hypothèse sociologique. On distingue deux types de variables selon leur rôle :

1. La Variable DÉPENDANTE, Y

(Celle qu'on veut expliquer)

C'est le phénomène, le comportement ou l'opinion qui varie.

→ C'est l'**EFFET**.

Ex : Le vote, le salaire, la réussite au bac, la survie.

2. La Variable INDÉPENDANTE, X

(Celle qui explique)

C'est la caractéristique (souvent stable) qui influence le comportement.

→ C'est la **CAUSE**.

Ex : L'âge, le sexe, le diplôme, l'origine sociale.

1.2 Comment les distinguer? (Astuce)

Posez-vous la question : **"Est-ce que A peut changer B, ou l'inverse?"**

1.2 Comment les distinguer ? (Astuce)

Posez-vous la question : "**Est-ce que A peut changer B, ou l'inverse ?**"

Exemple 1 : Le Vote et l'Âge

- ▶ Est-ce que voter à droite fait vieillir ? **Non.**
- ▶ Est-ce que vieillir augmente les chances de voter à droite ? **Oui (souvent).**
- ▶ **Donc** : Âge = Indépendante / Vote = Dépendante.

1.2 Comment les distinguer ? (Astuce)

Posez-vous la question : "**Est-ce que A peut changer B, ou l'inverse ?**"

Exemple 1 : Le Vote et l'Âge

- ▶ Est-ce que voter à droite fait vieillir ? **Non.**
- ▶ Est-ce que vieillir augmente les chances de voter à droite ? **Oui (souvent).**
- ▶ **Donc** : Âge = Indépendante / Vote = Dépendante.

Exemple 2 : Le Salaire et le Diplôme

- ▶ C'est le diplôme qui permet d'obtenir un salaire (et rarement l'inverse).
- ▶ **Donc** : Diplôme = Indépendante / Salaire = Dépendante.

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.
- ▶ On peut dire "plus que" ou "moins que".

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.
- ▶ On peut dire "plus que" ou "moins que".
- ▶ *Exemples :*

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.
- ▶ On peut dire "plus que" ou "moins que".
- ▶ *Exemples :*
 - ▶ Mention au Bac ($AB < B < TB$)

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.
- ▶ On peut dire "plus que" ou "moins que".
- ▶ *Exemples :*
 - ▶ Mention au Bac (AB < B < TB)
 - ▶ Satisfaction (Faible < Moyenne < Forte)

2.1 La nature des variables : (A) Catégorielles

On les appelle variables **QUALITATIVES** (ou nominales).
Elles servent à **classer** les individus dans des boîtes.

1. Nominales (Des Noms)

- ▶ Ce sont de simples étiquettes.
- ▶ Il n'y a **aucun ordre** logique.
- ▶ *Exemples :*
 - ▶ Sexe (H / F)
 - ▶ Couleur des yeux (Bleu, Marron...)
 - ▶ Région (Bretagne, PACA...)

2. Ordinales (Un Ordre)

- ▶ Ce sont des catégories, mais elles sont **hiérarchisées**.
- ▶ On peut dire "plus que" ou "moins que".
- ▶ *Exemples :*
 - ▶ Mention au Bac (AB < B < TB)
 - ▶ Satisfaction (Faible < Moyenne < Forte)
 - ▶ Classe sociale (Populaire < Moyenne < Sup.)

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.
- ▶ **Le Numéro de dossard** d'un coureur.

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.
- ▶ **Le Numéro de dossard** d'un coureur.
- ▶ **La variable "Survived" du Titanic** : Codée **0** (Mort) et **1** (Survivant).

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.
- ▶ **Le Numéro de dossard** d'un coureur.
- ▶ **La variable "Survived" du Titanic** : Codée **0** (Mort) et **1** (Survivant).
 - ▶ Si je fais la moyenne, j'obtiens 0,38.

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.
- ▶ **Le Numéro de dossard** d'un coureur.
- ▶ **La variable "Survived" du Titanic** : Codée **0** (Mort) et **1** (Survivant).
 - ▶ Si je fais la moyenne, j'obtiens 0,38.
 - ▶ Est-ce qu'un passager "0,38" existe ? Non.

2.2 Attention au piège des chiffres !

Règle d'or

Ce n'est pas parce qu'une variable contient des **chiffres** qu'elle est **quantitative** !

Certains chiffres sont en réalité des **étiquettes** (des codes).

Les exemples pièges :

- ▶ **Le Code Postal** : 75007. Est-ce une quantité ? Non, c'est une adresse. (Moyenne impossible). → **Qualitative**.
- ▶ **Le Numéro de dossard** d'un coureur.
- ▶ **La variable "Survived" du Titanic** : Codée **0** (Mort) et **1** (Survivant).
 - ▶ Si je fais la moyenne, j'obtiens 0,38.
 - ▶ Est-ce qu'un passager "0,38" existe ? Non.
 - ▶ C'est une catégorie (Mort ou Vivant). → **Qualitative**.

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples* :
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).
- ▶ C'est une mesure physique ou monétaire.

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).
- ▶ C'est une mesure physique ou monétaire.
- ▶ *Exemples :*

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).
- ▶ C'est une mesure physique ou monétaire.
- ▶ *Exemples :*
 - ▶ Salaire (1500,50 €).

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).
- ▶ C'est une mesure physique ou monétaire.
- ▶ *Exemples :*
 - ▶ Salaire (1500,50 €).
 - ▶ Taille (1,75 m).

2.3 La nature des variables : (B) Numériques

On les appelle variables **QUANTITATIVES**.

Ici, le chiffre mesure une vraie grandeur. La moyenne a du sens.

1. Discrètes (Des entiers)

- ▶ On compte par "bonds" (1, 2, 3...).
- ▶ Pas de virgule possible dans la réalité.
- ▶ *Exemples :*
 - ▶ Nombre d'enfants (on n'a pas 2,5 enfants).
 - ▶ Nombre de pièces dans un logement.

2. Continues (Le réel)

- ▶ Une infinité de valeurs possibles (avec virgules).
- ▶ C'est une mesure physique ou monétaire.
- ▶ *Exemples :*
 - ▶ Salaire (1500,50 €).
 - ▶ Taille (1,75 m).
 - ▶ Âge exact (34,5 ans).

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

Le "tri à plat" (une seule variable) ne suffit plus. On croise deux variables pour voir les proportions.

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

Le "tri à plat" (une seule variable) ne suffit plus. On croise deux variables pour voir les proportions.

Règle de lecture

Tout dépend de la position de la **Variable Explicative (La Cause)**.

- ▶ Cause en **COLONNE** → On lit les pourcentages en colonne (comparaison horizontale).

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

Le "tri à plat" (une seule variable) ne suffit plus. On croise deux variables pour voir les proportions.

Règle de lecture

Tout dépend de la position de la **Variable Explicative (La Cause)**.

- ▶ Cause en **COLONNE** → On lit les pourcentages en colonne (comparaison horizontale).
- ▶ Cause en **LIGNE** → On lit les pourcentages en ligne.

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

Le "tri à plat" (une seule variable) ne suffit plus. On croise deux variables pour voir les proportions.

Règle de lecture

Tout dépend de la position de la **Variable Explicative (La Cause)**.

- ▶ Cause en **COLONNE** → On lit les pourcentages en colonne (comparaison horizontale).
- ▶ Cause en **LIGNE** → On lit les pourcentages en ligne.

3. Cas n°1 : Quali × Quali

L'outil : Le Tableau Croisé

Le "tri à plat" (une seule variable) ne suffit plus. On croise deux variables pour voir les proportions.

Règle de lecture

Tout dépend de la position de la **Variable Explicative (La Cause)**.

- ▶ Cause en **COLONNE** → On lit les pourcentages en colonne (comparaison horizontale).
- ▶ Cause en **LIGNE** → On lit les pourcentages en ligne.

Le but est de voir si une proportion change d'une catégorie à l'autre.

3.1 Le Tableau Croisé : Lecture en Colonne

Situation la plus fréquente : La variable explicative (la cause) est en colonne.

Exemple : Le Sexe (Cause) influence-t-il la Survie (Effet) ?

3.1 Le Tableau Croisé : Lecture en Colonne

Situation la plus fréquente : La variable explicative (la cause) est en colonne.

Exemple : Le Sexe (Cause) influence-t-il la Survie (Effet) ?

	Variable Explicative		Ensemble
	Femmes	Hommes	
Survivants	74,2 %	18,9 %	38,4 %
Morts	25,8 %	81,1 %	61,6 %
TOTAL	100 %	100 %	100 %

3.1 Le Tableau Croisé : Lecture en Colonne

Situation la plus fréquente : La variable explicative (la cause) est en colonne.

Exemple : Le Sexe (Cause) influence-t-il la Survie (Effet) ?

	Variable Explicative		Ensemble
	Femmes	Hommes	
Survivants	74,2 %	18,9 %	38,4 %
Morts	25,8 %	81,1 %	61,6 %
TOTAL	100 %	100 %	100 %

Comment lire ?

- ▶ On repère le 100% (en bas).

3.1 Le Tableau Croisé : Lecture en Colonne

Situation la plus fréquente : La variable explicative (la cause) est en colonne.

Exemple : Le Sexe (Cause) influence-t-il la Survie (Effet) ?

	Variable Explicative		Ensemble
	Femmes	Hommes	
Survivants	74,2 %	18,9 %	38,4 %
Morts	25,8 %	81,1 %	61,6 %
TOTAL	100 %	100 %	100 %

Comment lire ?

- ▶ On repère le 100% (en bas).
- ▶ On lit : "**Parmi** les femmes, 74% ont survécu..."

3.1 Le Tableau Croisé : Lecture en Colonne

Situation la plus fréquente : La variable explicative (la cause) est en colonne.

Exemple : Le Sexe (Cause) influence-t-il la Survie (Effet) ?

	Variable Explicative		Ensemble
	Femmes	Hommes	
Survivants	74,2 %	18,9 %	38,4 %
Morts	25,8 %	81,1 %	61,6 %
TOTAL	100 %	100 %	100 %

Comment lire ?

- ▶ On repère le 100% (en bas).
- ▶ On lit : "**Parmi** les femmes, 74% ont survécu..."
- ▶ On compare : "...alors que **parmi** les hommes, seulement 19% ont survécu."

3.2 Le Tableau Croisé : Lecture en Ligne

Parfois, la variable explicative est présentée en **LIGNE**.

Exemple : La Classe Sociale (Cause) influence-t-elle la Survie ?

3.2 Le Tableau Croisé : Lecture en Ligne

Parfois, la variable explicative est présentée en **LIGNE**.

Exemple : La Classe Sociale (Cause) influence-t-elle la Survie ?

Classe (Explicative)	Survivants	Morts	TOTAL
1ère Classe	63,0 %	37,0 %	100 %
2ème Classe	47,3 %	52,7 %	100 %
3ème Classe	24,2 %	75,8 %	100 %
<i>Moyenne à bord</i>	<i>38,4 %</i>	<i>61,6 %</i>	<i>100 %</i>

3.2 Le Tableau Croisé : Lecture en Ligne

Parfois, la variable explicative est présentée en **LIGNE**.

Exemple : La Classe Sociale (Cause) influence-t-elle la Survie ?

Classe (Explicative)	Survivants	Morts	TOTAL
1ère Classe	63,0 %	37,0 %	100 %
2ème Classe	47,3 %	52,7 %	100 %
3ème Classe	24,2 %	75,8 %	100 %
<i>Moyenne à bord</i>	<i>38,4 %</i>	<i>61,6 %</i>	<i>100 %</i>

Comment lire ?

- ▶ Le 100% est à droite (au bout de la ligne).

3.2 Le Tableau Croisé : Lecture en Ligne

Parfois, la variable explicative est présentée en **LIGNE**.

Exemple : La Classe Sociale (Cause) influence-t-elle la Survie ?

Classe (Explicative)	Survivants	Morts	TOTAL
1ère Classe	63,0 %	37,0 %	100 %
2ème Classe	47,3 %	52,7 %	100 %
3ème Classe	24,2 %	75,8 %	100 %
<i>Moyenne à bord</i>	<i>38,4 %</i>	<i>61,6 %</i>	<i>100 %</i>

Comment lire ?

- ▶ Le 100% est à droite (au bout de la ligne).
- ▶ On lit : "**Parmi** les passagers de 1ère classe, 63% ont survécu..."

3.2 Le Tableau Croisé : Lecture en Ligne

Parfois, la variable explicative est présentée en **LIGNE**.

Exemple : La Classe Sociale (Cause) influence-t-elle la Survie ?

Classe (Explicative)	Survivants	Morts	TOTAL
1ère Classe	63,0 %	37,0 %	100 %
2ème Classe	47,3 %	52,7 %	100 %
3ème Classe	24,2 %	75,8 %	100 %
<i>Moyenne à bord</i>	<i>38,4 %</i>	<i>61,6 %</i>	<i>100 %</i>

Comment lire ?

- ▶ Le 100% est à droite (au bout de la ligne).
- ▶ On lit : "**Parmi** les passagers de 1ère classe, 63% ont survécu..."
- ▶ On compare : "...contre seulement 24% chez les passagers de 3ème classe."

4.1 Croisement Quali \times Quanti : Le Principe

Comment savoir si une catégorie (Quali) influence une grandeur chiffrée (Quanti) ?

Exemple : Est-ce que la Classe sociale détermine le Prix du billet ?

4.1 Croisement Quali \times Quanti : Le Principe

Comment savoir si une catégorie (Quali) influence une grandeur chiffrée (Quanti) ?

Exemple : Est-ce que la Classe sociale détermine le Prix du billet ?

La Méthode : "Découper et Mesurer"

On ne regarde plus des pourcentages d'individus.

1. On découpe la population en sous-groupes (Les 1ères, les 2èmes, les 3èmes).

4.1 Croisement Quali \times Quanti : Le Principe

Comment savoir si une catégorie (Quali) influence une grandeur chiffrée (Quanti) ?

Exemple : Est-ce que la Classe sociale détermine le Prix du billet ?

La Méthode : "Découper et Mesurer"

On ne regarde plus des pourcentages d'individus.

1. On découpe la population en sous-groupes (Les 1ères, les 2èmes, les 3èmes).
2. Pour **chaque groupe**, on calcule la Moyenne (le niveau) et l'Écart-type (la diversité).

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ (<i>Fort</i>)
2ème Classe	20 £	13 £ (<i>Fort</i>)
3ème Classe	13 £	11 £ (<i>Fort</i>)

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ (<i>Fort</i>)
2ème Classe	20 £	13 £ (<i>Fort</i>)
3ème Classe	13 £	11 £ (<i>Fort</i>)

Interprétation sociologique :

- ▶ **Le Niveau (Moyenne) :** Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ (<i>Fort</i>)
2ème Classe	20 £	13 £ (<i>Fort</i>)
3ème Classe	13 £	11 £ (<i>Fort</i>)

Interprétation sociologique :

- ▶ **Le Niveau (Moyenne) :** Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type) :**

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ (<i>Fort</i>)
2ème Classe	20 £	13 £ (<i>Fort</i>)
3ème Classe	13 £	11 £ (<i>Fort</i>)

Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type)** :
 - ▶ Les pauvres ont payé des prix allant de 0£ à 70£. La différence est énorme. Si on regarde par quartiles, 75% des passagers de 3ème classe payaient moins de 15£. Les 25% restants ont payé des prix beaucoup plus élevés, allant jusqu'à 70£.

4.2 Exemple : Prix du billet selon la Classe

Voici le tableau de résultats que l'on obtient sur R :

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ (<i>Fort</i>)
2ème Classe	20 £	13 £ (<i>Fort</i>)
3ème Classe	13 £	11 £ (<i>Fort</i>)

Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type)** :
 - ▶ Les pauvres ont payé des prix allant de 0£ à 70£. La différence est énorme. Si on regarde par quartiles, 75% des passagers de 3ème classe payaient moins de 15£. Les 25% restants ont payé des prix beaucoup plus élevés, allant jusqu'à 70£.
 - ▶ Les riches (1ère) sont très différents entre eux (écart fort) : il y a des "petits riches" et des multimilliardaires.

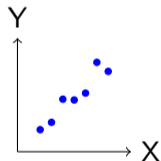
5. Cas n°3 : Quanti \times Quanti (Approche visuelle)

Quand on a deux variables numériques (ex : Âge et Revenu), on utilise d'abord un **Nuage de points**. Par convention, on met **Y** (*variable dépendante, à expliquer*) en **ordonnées**, et **X** (*variable indépendante, qui sert à expliquer*) en **abscisses**.

5. Cas n°3 : Quanti \times Quanti (Approche visuelle)

Quand on a deux variables numériques (ex : Âge et Revenu), on utilise d'abord un **Nuage de points**. Par convention, on met **Y** (*variable dépendante, à expliquer*) en **ordonnées**, et **X** (*variable indépendante, qui sert à expliquer*) en **abscisses**.

Relation Positive

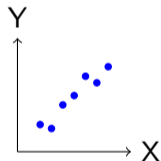


Ça monte ensemble.

5. Cas n°3 : Quanti \times Quanti (Approche visuelle)

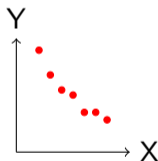
Quand on a deux variables numériques (ex : Âge et Revenu), on utilise d'abord un **Nuage de points**. Par convention, on met **Y** (*variable dépendante, à expliquer*) en **ordonnées**, et **X** (*variable indépendante, qui sert à expliquer*) en **abscisses**.

Relation Positive



Ça monte ensemble.

Relation Négative

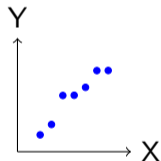


Quand l'un monte, l'autre descend.

5. Cas n°3 : Quanti \times Quanti (Approche visuelle)

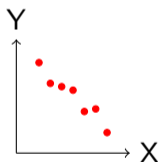
Quand on a deux variables numériques (ex : Âge et Revenu), on utilise d'abord un **Nuage de points**. Par convention, on met **Y** (*variable dépendante, à expliquer*) en **ordonnées**, et **X** (*variable indépendante, qui sert à expliquer*) en **abscisses**.

Relation Positive



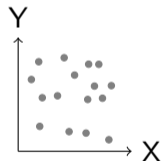
Ça monte ensemble.

Relation Négative



Quand l'un monte, l'autre descend.

Pas de relation



C'est le chaos (nuage dispersé).

5.1 La Covariance : L'intuition

Nous avons vu le nuage de points. Mais comment calculer un "chiffre" qui résume le lien ?

5.1 La Covariance : L'intuition

Nous avons vu le nuage de points. Mais comment calculer un "chiffre" qui résume le lien ?

Le Principe : Le "Produit des Écarts"

Pour chaque individu, on regarde deux choses :

1. Est-il au-dessus ou en dessous de la moyenne en X? (Signe + ou -)

5.1 La Covariance : L'intuition

Nous avons vu le nuage de points. Mais comment calculer un "chiffre" qui résume le lien ?

Le Principe : Le "Produit des Écarts"

Pour chaque individu, on regarde deux choses :

1. Est-il au-dessus ou en dessous de la moyenne en X ? (Signe + ou -)
2. Est-il au-dessus ou en dessous de la moyenne en Y ? (Signe + ou -)

5.1 La Covariance : L'intuition

Nous avons vu le nuage de points. Mais comment calculer un "chiffre" qui résume le lien ?

Le Principe : Le "Produit des Écarts"

Pour chaque individu, on regarde deux choses :

1. Est-il au-dessus ou en dessous de la moyenne en X ? (Signe + ou -)
2. Est-il au-dessus ou en dessous de la moyenne en Y ? (Signe + ou -)

Ensuite, on **multiplie** ces deux résultats.

5.1 La Covariance : L'intuition

Nous avons vu le nuage de points. Mais comment calculer un "chiffre" qui résume le lien ?

Le Principe : Le "Produit des Écarts"

Pour chaque individu, on regarde deux choses :

1. Est-il au-dessus ou en dessous de la moyenne en X? (Signe + ou -)
2. Est-il au-dessus ou en dessous de la moyenne en Y? (Signe + ou -)

Ensuite, on **multiplie** ces deux résultats.

La Formule mathématique (pour info) :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum (x_i - \bar{x}) \times (y_i - \bar{y})$$

5.2 La Covariance : La mécanique des signes

Prenons un exemple concret : Le lien entre **Temps de révision** et **Note**.
Comment la formule réagit-elle aux différents profils d'étudiants ?

5.2 La Covariance : La mécanique des signes

Prenons un exemple concret : Le lien entre **Temps de révision** et **Note**.

Comment la formule réagit-elle aux différents profils d'étudiants ?

Profil	Écart Révision	Écart Note	Produit (Résultat)

5.2 La Covariance : La mécanique des signes

Prenons un exemple concret : Le lien entre **Temps de révision** et **Note**.

Comment la formule réagit-elle aux différents profils d'étudiants ?

Profil	Écart Révision	Écart Note	Produit (Résultat)
Profil A	Il a bossé PLUS (+)	Il a eu MIEUX (+)	$(+) \times (+) = +$

Bilan :

- ▶ Si la somme totale est **positive**, c'est que les profils logiques (A et B) dominent.

5.2 La Covariance : La mécanique des signes

Prenons un exemple concret : Le lien entre **Temps de révision** et **Note**.

Comment la formule réagit-elle aux différents profils d'étudiants ?

Profil	Écart Révision	Écart Note	Produit (Résultat)
Profil A	Il a bossé PLUS (+)	Il a eu MIEUX (+)	$(+) \times (+) = +$
Profil B	Il a bossé MOINS (-)	Il a eu MOINS (-)	$(-) \times (-) = +$

Bilan :

- ▶ Si la somme totale est **positive**, c'est que les profils logiques (A et B) dominent.
- ▶ Le lien général est donc positif.

5.2 La Covariance : La mécanique des signes

Prenons un exemple concret : Le lien entre **Temps de révision** et **Note**.

Comment la formule réagit-elle aux différents profils d'étudiants ?

Profil	Écart Révision	Écart Note	Produit (Résultat)
Profil A	Il a bossé PLUS (+)	Il a eu MIEUX (+)	$(+) \times (+) = +$
Profil B	Il a bossé MOINS (-)	Il a eu MOINS (-)	$(-) \times (-) = +$
Profil C	Il a bossé MOINS (-)	Il a eu MIEUX (+)	$(-) \times (+) = -$

Bilan :

- ▶ Si la somme totale est **positive**, c'est que les profils logiques (A et B) dominent.
- ▶ Le lien général est donc positif.

5.3 De la Covariance à la Corrélacion

La covariance a un défaut majeur : elle dépend des unités (euros, mètres, années...).

Exemple : Une covariance de "5000" est-elle forte ? Impossible à dire.

5.3 De la Covariance à la Corrélacion

La covariance a un défaut majeur : elle dépend des unités (euros, mètres, années...).

Exemple : Une covariance de "5000" est-elle forte ? Impossible à dire.

La Solution : Le Coefficient de Corrélacion (r)

Pour obtenir un score universel, on divise la covariance par la dispersion (les écarts-types).

$$r = \frac{\text{Covariance}(X, Y)}{\sigma_X \times \sigma_Y}$$

5.3 De la Covariance à la Corrélacion

La covariance a un défaut majeur : elle dépend des unités (euros, mètres, années...).

Exemple : Une covariance de "5000" est-elle forte ? Impossible à dire.

La Solution : Le Coefficient de Corrélacion (r)

Pour obtenir un score universel, on divise la covariance par la dispersion (les écarts-types).

$$r = \frac{\text{Covariance}(X, Y)}{\sigma_X \times \sigma_Y}$$

Ce qu'il faut retenir :

- ▶ Le résultat est **toujours** compris entre **-1** et **+1**.

5.3 De la Covariance à la Corrélation

La covariance a un défaut majeur : elle dépend des unités (euros, mètres, années...).

Exemple : Une covariance de "5000" est-elle forte ? Impossible à dire.

La Solution : Le Coefficient de Corrélation (r)

Pour obtenir un score universel, on divise la covariance par la dispersion (les écarts-types).

$$r = \frac{\text{Covariance}(X, Y)}{\sigma_X \times \sigma_Y}$$

Ce qu'il faut retenir :

- ▶ Le résultat est **toujours** compris entre **-1** et **+1**.
- ▶ Il n'a plus d'unité (on peut comparer des choux et des carottes).

5.4 Lire le résultat

Comment interpréter la valeur de r ?

5.4 Lire le résultat

Comment interpréter la valeur de r ?



5.4 Lire le résultat

Comment interpréter la valeur de r ?



Note : Un r de -0.8 est un lien aussi fort qu'un r de $+0.8$. Seul le sens change.