

# Étude cas-témoin sur l'infarctus du myocarde

Biostatistiques

**LANCON Cindy, FRISON Victor**

Encadrant : Davit Varron

2026-01-15

# Plan

## 1 Introduction

### ► Introduction

### ► Regression logistique simple / univarié

### ► Regression multivariée logistique multiple

### ► Conclusion

# Introduction

## 1 Introduction

### Design et Population :

- **Type** : Étude multicentrique sur l'infarctus du myocarde aigu.
- **Echantillon global** :  $N = 520$  patients.

### Constitution des groupes :

- **260 Cas** : Patients hospitalisés pour un 1<sup>er</sup> infarctus confirmé.
- **260 Témoins** : Patients hospitalisés (mêmes centres) pour pathologies non cardio-vasculaires, sans antécédents coronariens.

### Variables et Modélisation

#### Variable réponse ( $Y$ ) :

Statut du patient

**Modalité de référence : Témoin**

*Objectif : Modéliser la probabilité d'être Cas.*

#### Facteurs de risque étudiés ( $X$ ) :

*Cliniques* : Âge, Sexe, IMC, HTA, Diabète, Cholestérol.

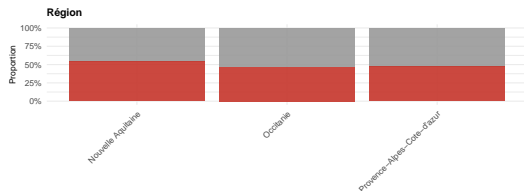
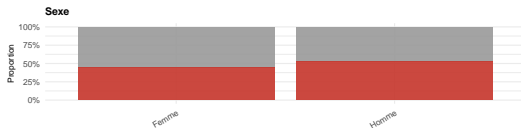
*Mode de vie* : Activité physique, Alcool, Région.

*Tabagisme* : Jamais / Ancien / Actuel.

**Réf Tabac : Jamais fumeur.**

# Analyse exploratoire et Statistiques Descriptives

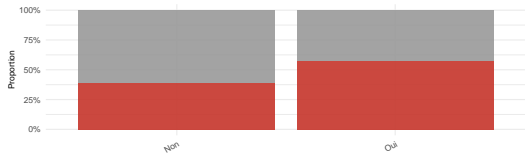
## 1 Introduction



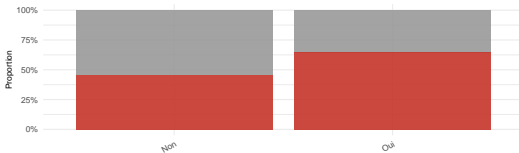
# Analyse exploratoire et Statistiques Descriptives

## 1 Introduction

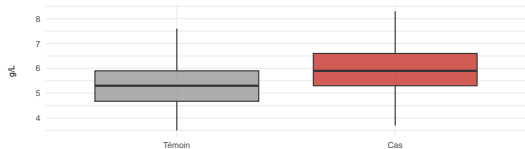
**Hypertension**



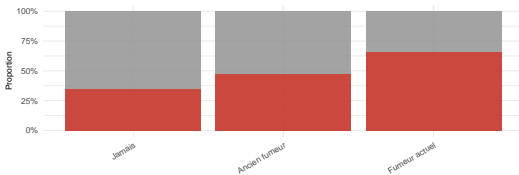
**Diabète**



**Cholestérol Total**

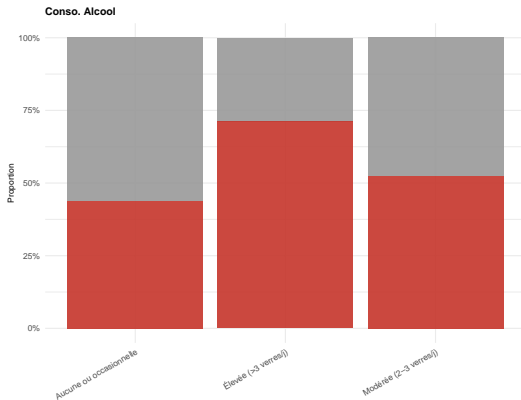
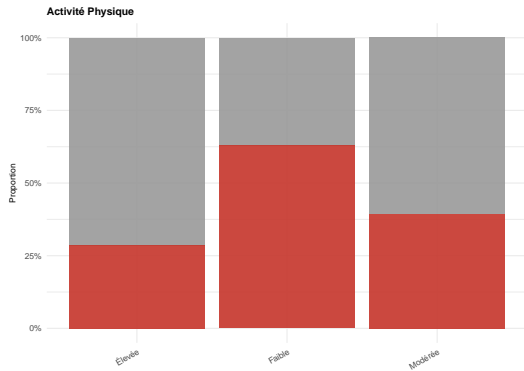


**Tabagisme**



# Analyse exploratoire et Statistiques Descriptives

## 1 Introduction



# Plan

## 2 Regression logistique simple / univarié

► Introduction

► Regression logistique simple / univarié

► Regression multivariée logistique multiple

► Conclusion

# Variable Âge

## 2 Regression logistique simple / univarié

### Résultats du modèle logistique :

- Coefficient estimé ( $\hat{\beta}_{age}$ ) : 0.083
- Odds Ratio (OR) : 1.09
- Intervalle de Confiance à 95% : [1.06 ; 1.11]
- Significativité (p-value) :  $p < 0.001$

**Analyse de l'Odds Ratio :** Pour chaque année d'âge supplémentaire, l'odds de faire un infarctus est multiplié par environ **1.09**. Concrètement, le risque augmente de **6% à 11%** par an.

### Conclusion

L'intervalle de confiance ne contient pas la valeur 1 et la p-value est extrêmement faible. **L'âge est donc un facteur de risque significativement associé à la survenue de l'infarctus.**



### Résultats du modèle logistique :

- Coefficient estimé ( $\hat{\beta}_{imc}$ ) : 0.148
- Odds Ratio (OR) : 1.16
- Intervalle de Confiance à 95% : [1.10 ; 1.22]
- Significativité (p-value) :  $p < 0.001$

**Analyse de l'Odds Ratio** : Pour chaque unité d'IMC supplémentaire, l'odds d'être malade est multiplié par environ **1.16**. Concrètement, le risque augmente de **10% à 22%** par unité.

### Conclusion

L'intervalle ne contient pas 1 et la p-value est extrêmement faible ( $6.7 \times 10^{-10}$ ). **L'IMC est donc significativement associé à la survenue de l'infarctus.**

### Résultats du modèle logistique :

- Coefficient estimé ( $\hat{\beta}_{chol}$ ) : 0.785
- Odds Ratio (OR) : 2.19
- Intervalle de Confiance à 95% : [1.79 ; 2.71]
- Significativité (p-value) :  $p < 0.001$

**Analyse de l'Odds Ratio :** Pour chaque unité de cholestérol supplémentaire, l'odds est multiplié par environ **2.19**. Le risque augmente considérablement, de **79% à 171%** par unité.

### Conclusion

L'intervalle est nettement supérieur à 1. **Le cholestérol total est un facteur de risque majeur et hautement significatif.**

# Variable Sexe

## 2 Regression logistique simple / univarié

### Résultats du modèle logistique (Réf : Femme) :

- Coefficient estimé ( $\hat{\beta}_{Homme}$ ) : 0.307
- Intervalle de Confiance de l'OR à 95% : [0.95 ; 1.95]
- Significativité (p-value) :  $> 0.05$  (Non significatif)

**Analyse :** L'intervalle de confiance de l'Odds Ratio contient la valeur 1 ([0.95; 1.95]). Cela signifie qu'il n'y a pas de différence de risque statistiquement prouvée entre les hommes et les femmes dans cet échantillon.

### Conclusion

La p-value est supérieure au seuil de 5% et l'IC contient 1. **Le sexe n'est pas statistiquement associé à la survenue de l'infarctus.**

# Variable Tabagisme

## 2 Regression logistique simple / univarié

### Résultats par rapport aux non-fumeurs (Réf : Jamais) :

- **Ancien Fumeur :**
  - OR IC 95% : [1.07 ; 2.61] ( $p = 0.023$ )
- **Fumeur Actuel :**
  - OR IC 95% : [2.34 ; 5.56] ( $p < 0.001$ )

### Analyse des risques :

- Être **ancien fumeur** augmente le risque (OR min : 1.07).
- Être **fumeur actuel** l'augmente drastiquement (OR entre 2.34 et 5.56).

### Conclusion

Le test global est très significatif ( $p < 0.001$ ). **Le tabagisme est fortement associé à la maladie, avec un effet dose-réponse (Actuel > Ancien > Jamais).**

# Variable Hypertension

## 2 Regression logistique simple / univarié

### Résultats du modèle logistique (Réf : Non) :

- Coefficient estimé ( $\hat{\beta}_{HTA}$ ) : 0.75
- Intervalle de Confiance de l'OR à 95% : [1.47 ; 3.03]
- Significativité (p-value) :  $p < 0.001$  ( $3.6 \times 10^{-5}$ )

**Analyse de l'Odds Ratio** : Le fait d'être hypertendu multiplie le risque (l'odds) d'être un cas par au moins **1.47** et au plus **3.03** par rapport aux sujets sains.

### Conclusion

L'intervalle de confiance est nettement supérieur à 1. L'hypertension artérielle est un facteur de risque majeur statistiquement associé à l'infarctus.

# Variable Diabète

## 2 Regression logistique simple / univarié

### Résultats du modèle logistique (Réf : Non) :

- Coefficient estimé ( $\hat{\beta}_{diabète}$ ) : 0.77
- Odds Ratio (OR) : 2.16 ( $e^{0.77}$ )
- Intervalle de Confiance à 95% : [1.42 ; 3.34]
- Significativité (p-value) :  $p < 0.001$  ( $3.2 \times 10^{-4}$ )

**Analyse de l'Odds Ratio** : Être diabétique multiplie l'odds d'être un cas par au moins **1.42** et au plus **3.34** par rapport à un sujet non diabétique.

### Conclusion

La p-value est très faible et l'intervalle exclut 1. Le **diabète est statistiquement associé à la contraction de la maladie.**

### Résultats par rapport à une activité Élevée (Réf) :

- **Activité Faible :**
  - OR IC 95% : [2.48 ; 7.51] ( $p < 0.001$ )
- **Activité Modérée :**
  - OR IC 95% : [0.91 ; 2.93] ( $p = 0.104$ , Non Sign.)

**Analyse des risques :** Une activité physique **faible** augmente drastiquement le risque ( $OR > 2.48$ ). En revanche, la différence entre activité modérée et élevée n'est pas significative.

### Conclusion

Le test global est significatif ( $p < 0.001$ ). **La sédentarité (activité faible) est un facteur de risque majeur par rapport à une activité sportive régulière.**

# Variable Alcool

## 2 Regression logistique simple / univarié

### Résultats par rapport à Aucune/Occas. (Réf) :

- **Consommation Élevée :**
  - OR IC 95% : [1.80 ; 5.79] ( $p < 0.001$ )
- **Consommation Modérée :**
  - OR IC 95% : [0.96 ; 2.08] ( $p = 0.08$ , Non Sign.)

**Analyse des risques :** Une consommation **élevée** multiplie le risque par environ 3 (OR estimé à 3.15). Une consommation modérée ne montre pas de sur-risque significatif.

### Conclusion

Le test global est significatif ( $p < 0.001$ ). **L'abus d'alcool (consommation élevée) est fortement associé à la maladie.**



# Plan

## 3 Regression multivariée logistique multiple

- ▶ Introduction
- ▶ Regression logistique simple / univarié
- ▶ Regression multivariée logistique multiple
- ▶ Conclusion

# Création du modèle

## 3 Regression multivariée logistique multiple

### 1. Objectif

- Ajuster l'effet de chaque facteur en contrôlant les **facteurs de confusion**.
- Obtenir un modèle "nettoyé" des variables non pertinentes.

### 2. Méthode de sélection

- **Type** : Descendante (*Backward*).
- **Critère** : Minimisation de l'**AIC** (Akaike Information Criterion).
- *Principe* : On part du modèle complet et on retire itérativement les variables qui n'apportent pas d'information significative.

### Implémentation R :

```
# Modèle complet (toutes variables)
mod_full <- glm(statut_cas_temoin ~ .,
                data = data,
                family = binomial)

# Sélection Stepwise (AIC)
# direction = "backward"
mod_final <- stepAIC(mod_full,
                    trace = FALSE)

# Comparaison des modèles
AIC(mod_full, mod_final)
```

# Résultat de la sélection

## 3 Regression multivariée logistique multiple

### Résultats de la sélection

La procédure améliore la qualité du modèle (baisse de l'AIC de **526.83** à **525.09**).

**Variables exclues (Non Sign.) :**

*Sexe*

*Région*

**Variables conservées :** Âge, Tabagisme, IMC, HTA, Diabète, Cholestérol, Activité, Alcool.

# Comparaison modèle univarié et multivarié

## 3 Regression multivariée logistique multiple

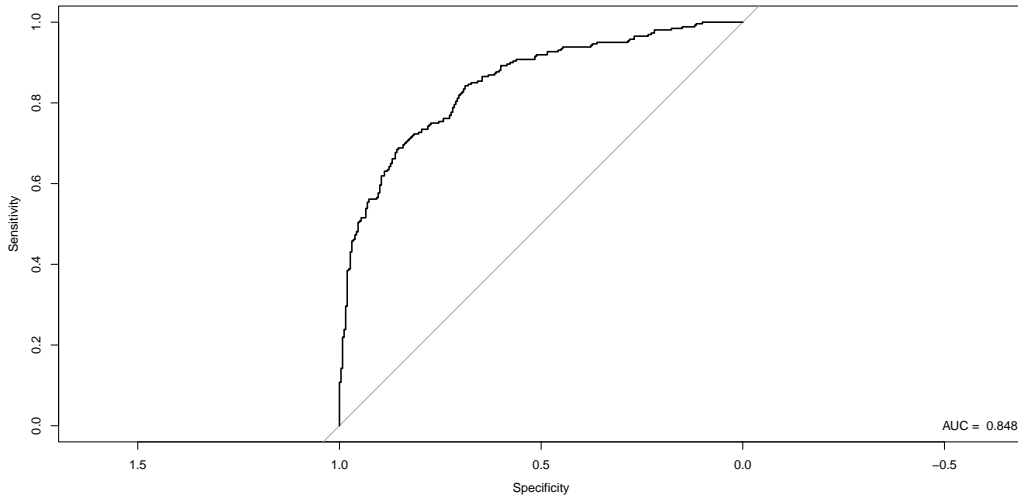
Table 1: Comparaison des odds ratios bruts (univariés) et ajustés (multivariés)

Variable	Analyse univariée			Analyse multivariée		
	OR	IC 95%	p	OR	IC 95%	p
Âge (par année)	1.09	[1.06 ; 1.11]	<0.001	1.08	[1.05 ; 1.11]	<0.001
Tabagisme - Ancien fumeur	1.67	[1.07 ; 2.61]	0.024	1.50	[0.88 ; 2.59]	0.140
Tabagisme - Fumeur actuel	3.59	[2.34 ; 5.56]	<0.001	2.70	[1.62 ; 4.57]	<0.001
IMC (par kg/m <sup>2</sup> )	1.16	[1.10 ; 1.22]	<0.001	1.17	[1.10 ; 1.25]	<0.001
Hypertension - Oui	2.11	[1.47 ; 3.03]	<0.001	1.98	[1.27 ; 3.10]	0.003
Diabète - Oui	2.16	[1.42 ; 3.34]	<0.001	2.19	[1.29 ; 3.75]	0.004
Cholestérol total (par mmol/L)	2.19	[1.79 ; 2.71]	<0.001	2.18	[1.71 ; 2.82]	<0.001
Activité physique - Faible	4.26	[2.48 ; 7.51]	<0.001	2.85	[1.50 ; 5.56]	0.002
Activité physique - Modérée	1.62	[0.91 ; 2.93]	0.104	1.38	[0.70 ; 2.79]	0.356
Consommation alcool - Élevée	3.16	[1.80 ; 5.79]	<0.001	2.88	[1.46 ; 5.86]	0.003
Consommation alcool - Modérée	1.42	[0.96 ; 2.08]	0.080	1.60	[0.98 ; 2.63]	0.061

# Performance du modèle

## 3 Regression multivariée logistique multiple

Courbe ROC du modèle final



# Interactions dans le modèle

## 3 Regression multivariée logistique multiple

**Le concept :** Il est possible que l'effet d'un facteur de risque dépende du niveau d'une autre variable (effet multiplicatif et non plus additif).

**Nous avons testé 5 interactions biologiquement plausibles :**

1. **Tabagisme**  $\times$  **Sexe** : L'effet du tabac est-il différent chez les femmes ?
2. **Tabagisme**  $\times$  **Âge** : L'impact augmente-t-il avec l'âge ?
3. **IMC**  $\times$  **Activité physique** : Le sport compense-t-il le surpoids ?
4. **Diabète**  $\times$  **Hypertension** : Y a-t-il un effet synergique ("cocktail explosif") ?
5. **Cholestérol**  $\times$  **Âge** : L'effet du cholestérol varie-t-il selon l'âge ?

# Interactions dans le modèle

## 3 Regression multivariée logistique multiple

Interaction.Testée	P.value	Significatif
Tabac * Sexe	0.30321	Non
Tabac * Âge	0.60534	Non
IMC * Activité	0.41417	Non
Diabète * HTA	0.73273	Non
Cholestérol * Âge	0.77545	Non

### Validation du modèle additif

Aucune interaction n'est significative ( $p > 0.05$ ). Cela suggère que les effets des facteurs de risque sont additifs et indépendants les uns des autres dans notre échantillon.

**Le modèle final sans interaction est donc conservé.**

# Plan

## 4 Conclusion

- ▶ Introduction
- ▶ Regression logistique simple / univarié
- ▶ Regression multivariée logistique multiple
- ▶ Conclusion



# Synthèse des Résultats

## 4 Conclusion

**1. Facteurs de risque indépendants confirmés :** L'analyse multivariée a permis d'isoler les facteurs ayant un impact direct, indépendamment des facteurs de confusion :

- **Facteurs Cliniques :**

- Âge, IMC élevé.
- Hypertension, Diabète.
- Cholestérol élevé.

- **Mode de vie :**

- Tabagisme **actif**.
- Activité physique faible.
- Consommation d'alcool élevée.

**2. Facteurs non conservés :** Le statut d'*ancien fumeur*, le *sexe* et la *région* ne sont plus significatifs après ajustement.

### Qualité du modèle

Le modèle final est performant (**AUC = 0.85**) et valide l'hypothèse d'additivité des risques (aucune interaction significative).

# Conclusion et perspectives

## 4 Conclusion

**Conclusion principale** : Nos résultats soulignent l'importance cruciale des **facteurs de risque modifiables** (Tabac, Sédentarité, Alcool) dans la prévention de l'infarctus.

### Limites de l'étude

- **Causalité** : Étude cas-témoin (observationnelle) → montre des associations, pas des causalités.
- **Variables manquantes** : Pas de données sur l'alimentation ou les antécédents familiaux.

### Perspectives

Si des données de suivi temporel étaient disponibles :

- Réalisation d'une **Analyse de Survie** (Modèle de Cox).
- Objectif : Étudier le *temps* jusqu'à la survenue de l'événement.