100% ACTUAIRES & DATA SCIENCE
Jeudi 17 novembre 2022

Institut des ACTUAIRES

Encadrement de l'IA et renouveau des questions de discrimination dans les travaux actuariels

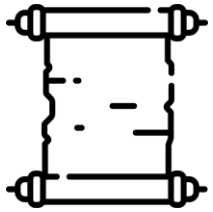
Antoine Chancel
SCOR

Fabien Faivre
Macif

Antoine Ly
SCOR

Marguerite Saucé
SCOR

Introduction



Sujet ancien

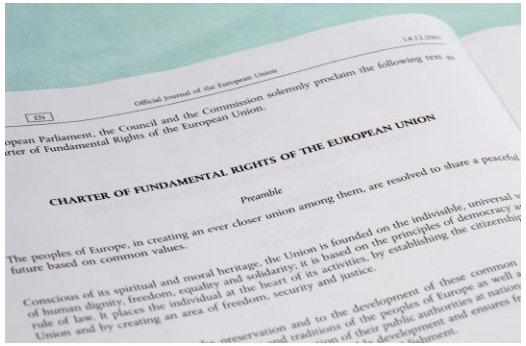
Mythe actuaire et vision objective du risque remis en cause

Renouveau ML

Subjectivité partout : données process narratif du risque construction modèles

Attention des régulateurs sur activités cœur de métier, notamment enjeux discrimination et équité

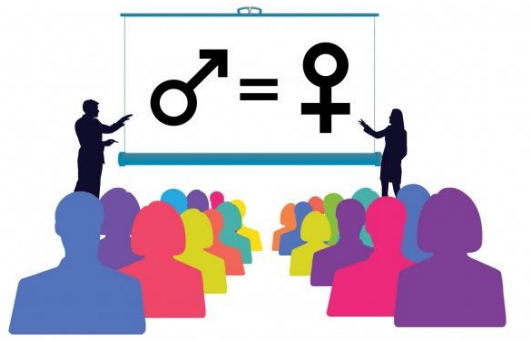
PARTIE 1
Une réglementation en pleine évolution



2000 — Charte des droits fondamentaux de l'UE

Article 21 - Non-discrimination

- 1. Est interdite toute discrimination fondée notamment sur le sexe, la race, la couleur, les origines ethniques ou sociales, les caractéristiques génétiques, la langue, la religion ou les convictions, les opinions politiques ou toute autre opinion, l'appartenance à une minorité nationale, la fortune, la naissance, un handicap, l'âge ou l'orientation sexuelle.



2002

Directive 2002/73/CE égalité de traitement entre hommes et femmes en ce qui concerne l'accès à l'emploi, à la formation et à la promotion professionnelles, et les conditions de travail

Article 2

- **Discrimination directe** : la situation dans laquelle une personne est traitée de manière moins favorable en raison de son sexe qu'une autre ne l'est, ne l'a été ou ne le serait dans une situation comparable.

- **Discrimination indirecte** : la situation dans laquelle une disposition, un critère ou une pratique apparemment neutre désavantagerait particulièrement des personnes d'un sexe par rapport à des personnes de l'autre sexe, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour parvenir à ce but soient appropriés et nécessaires.

100% ACTUAIRES & DATA SCIENCE
Jeudi 17 novembre 2022

Lecture subjective et non exhaustive

Institut des ACTUAIRES

2012 **CJUE : "Gender Insurance Directive"**

« La prise en compte du sexe de l'assuré en tant que facteur de risque dans les contrats d'assurance constitue une discrimination »

FRA EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

2018 **FRA : #BigData: Discrimination in data-supported decision making**

« The use of algorithms in making decisions and building automated processes may have a significant impact on people's lives »

https://fra.europa.eu/fr/publication/2018/bigdata-discrimination-data-supported-decision-making#publication-tab-11

2019 **AI HLEG : Ethics Guidelines for Trustworthy AI**

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

The concept of Trustworthy AI was introduced by the High-Level Expert Group on AI (AI HLEG) in the Ethics Guidelines for Trustworthy Artificial Intelligence (AI) and is based on seven key requirements:

1. Human Agency and Oversight;
2. Technical Robustness and Safety;
3. Privacy and Data Governance;
4. Transparency;
5. Diversity, Non-discrimination and Fairness;
6. Environmental and Societal well-being; and
7. Accountability.

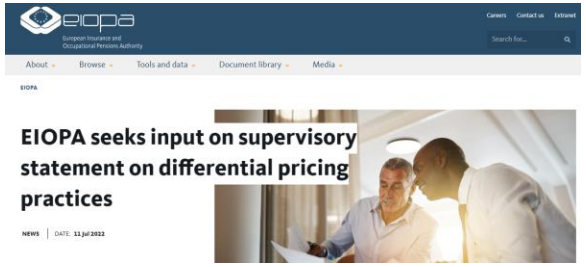
Lecture subjective et non exhaustive

2021

EIOPA :

AI governance principles: towards ethical and trustworthy AI in the European insurance sector

https://www.eiopa.europa.eu/sites/default/files/publications/reports/eiopa-ai-governance-principles-june-2021.pdf



2022

EIOPA :

Consultation sur les pratiques de tarification différenciées en non-vie

https://www.eiopa.europa.eu/media/news/eiopa-seeks-input-supervisory-statement-differential-pricing-practices_en

Lecture subjective et non exhaustive

2021 → 2025 ?

AI Act

EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

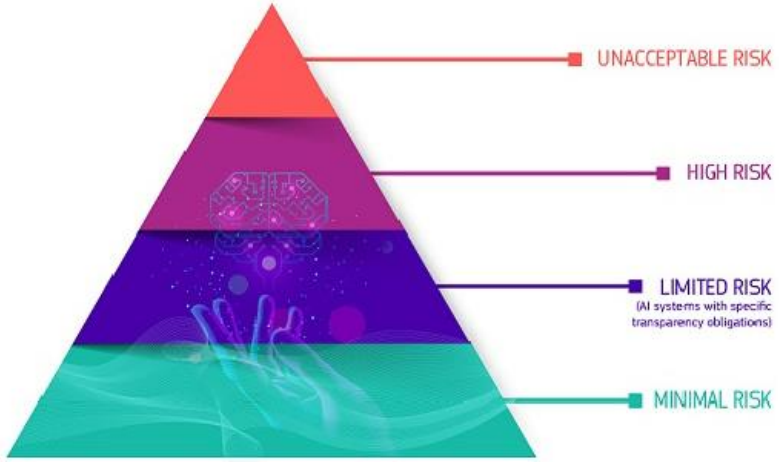
2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}



4 niveaux de risque

MAIS contrairement à S2

Évaluation des risques exogène

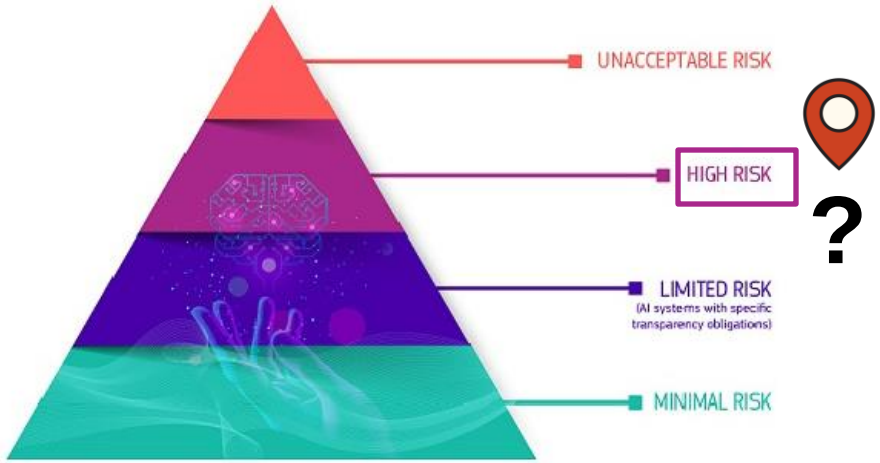
Liste définie par l'Annexe III en fonction des domaines d'application

Commission amende liste

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

2021 → 2025 ?

AI Act



Annexe III : "5(e). AI systems intended to be used for **risk assessment** in relation to natural persons **and pricing** in the case of **life and health insurance** with the exception of AI systems put into service by providers that are micro and small-sized enterprises as defined in the Annex of Commission Recommendation 2003/361/EC for their own use."

21/04/2021 — **PAS** d'assurance

29/11/2021 — **TOUTE** l'assurance (tarification, souscription, gestion sinistre)

15/07/2022 — **PAS** d'assurance

19/10/2022 — **SANTE / VIE** Tarification et évaluation des risques

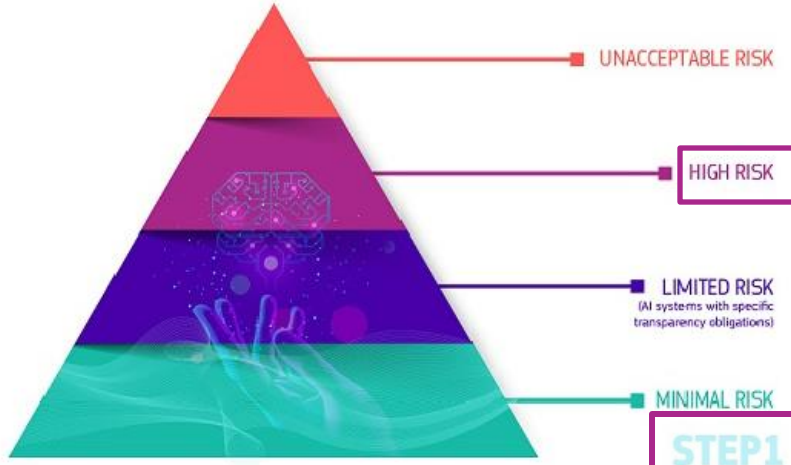
03/11/2022 — idem

https://artificialintelligenceact.eu/wp-content/uploads/2022/11/AIA-CZ-Draft-for-Coreper-3-Nov-22.pdf

Lecture subjective et non exhaustive

2021 → 2025 ?

AI Act



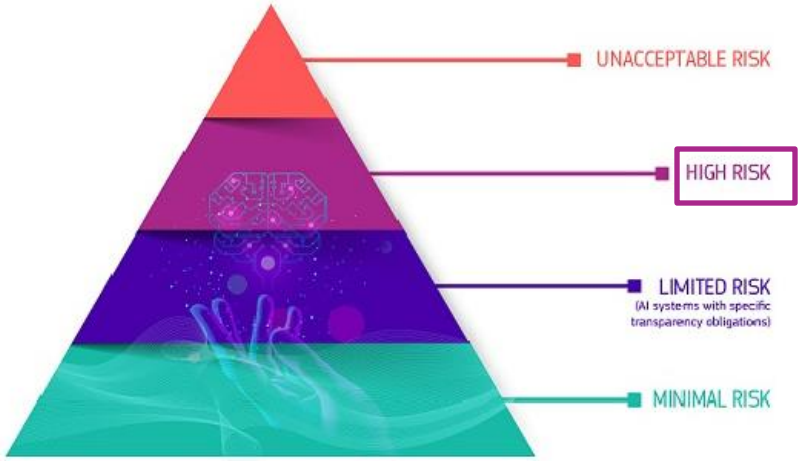
=CNIL.?

https://www.conseil-etat.fr/actualites/s-engager-dans-l-intelligence-artificielle-pour-un-meilleur-service-public

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

2021 → 2025 ?

AI Act



Equité :

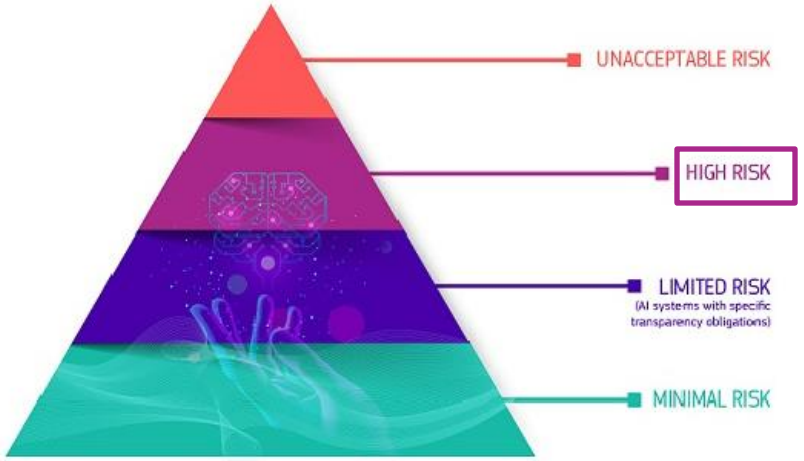
Obligation de respect de la Charte des Droits Fondamentaux de l'UE

Article 10.5 : possibilité de constituer des **bases de données personnelles sensibles** permettant par exemple des statistiques ethniques. Cela autorise la mesure directe des biais statistiques, sources potentielles de discrimination.

https://www.math.univ-toulouse.fr/~besse/pub/StatSociete012022-AI-act-stat.pdf

2021 → 2025 ?

AI Act



Equité :

Obligation de respect de la Charte des Droits Fondamentaux de l'UE (article 20)

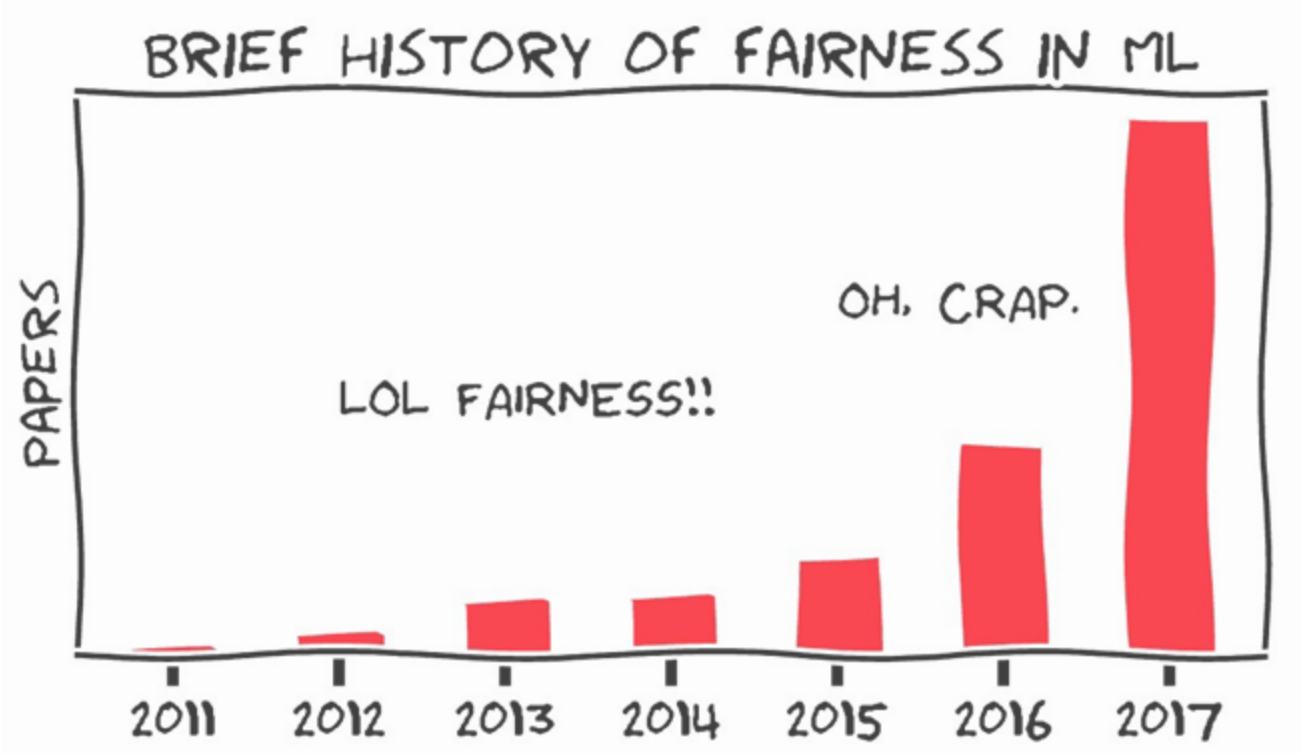
MAIS

- Choix des métriques de biais laissées à l'initiative du concepteur
- Dizaines de définitions de biais statistiques pouvant être à l'origine de sources de discrimination. Lesquelles considérer en priorité ?

La recherche d'un biais systémique ou de société est requise dans l'analyse préalable des données (art. 10, 2. (f)), ainsi que l'obligation de détailler les performances (précision) par groupe ou sous-groupe d'un système d'IA (art. 13, 3., (b) iv). Ceci permet de prendre en compte **certains** types de biais.

https://www.math.univ-toulouse.fr/~besse/pub/StatSociete012022-AI-act-stat.pdf

PARTIE 2
Multiplicité des définitions d'équité



Discrimination

Directe

Différence de traitement basée sur des caractéristiques protégées

$$\mathbb{E}[Y|X,S]$$

Indirecte

Différence de traitement sans utilisation des caractéristiques protégées

$$\mathbb{E}[Y|X]$$

MAIS

$$X \not\perp\!\!\!\perp S$$

Proxys : inférence (volume données, complexité algorithmes)

Discrimination

Directe

Différence de traitement basée sur des caractéristiques protégées

$$\mathbb{E}[Y|X,S]$$

Indirecte

Différence de traitement sans utilisation des caractéristiques protégées

$$\mathbb{E}[Y|X]$$

MAIS

$$X \not\perp\!\!\!\perp S$$

Proxys : inférence (volume données, complexité algorithmes)



Définies par la loi (selon juridiction, secteur d'activité)

Réglementation actuelle :

- ne pas utiliser les variables protégées
- éviter certains proxys

Pas de règles pour évaluer la discrimination indirecte en assurance

Premiers efforts de définition :
US, Disparate Impact

Equité actuarielle

Équité face au risque ⇒ Classes de risque homogènes ⇒ Segmentation et mutualisation

→ Critères socialement inacceptables

→ Corriger quelque chose qui n'est pas dans les données

Eviter discrimination directe : variables sensibles non collectées

Impossible de détecter si discrimination → AI Act prévoit collecte → Acceptation ? Sincérité réponses ?

Equité statistique

S variable protégée
Classification : Y classe réelle et Ŷ classe estimée (binaires)

		Classe estimée	
		Positive	Négative
Classe réelle	Positive	TP	FN
	Négative	FP	TN

Matrice de confusion

Equité de groupe

Parité statistique : $\hat{Y} \perp\!\!\!\perp S$ ⇔ mêmes taux d'acceptation (AR) pour tous les groupes

$$\mathrm{AR} = \frac{\mathrm{TP} + \mathrm{FP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Egalité des chances: $\hat{Y} \perp\!\!\!\perp S|Y$ ⇔ mêmes taux de vrais (TPR) et faux positifs (FPR) pour tous les groupes

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

Egalité des opportunités : mêmes taux de vrais positifs (TPR) pour tous les groupes

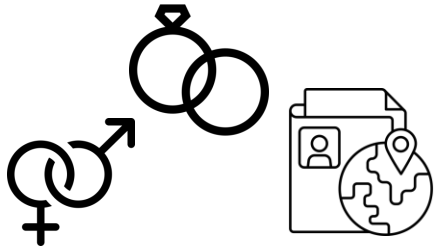
Equité individuelle

Individus similaires traités de manière similaire
Proximité entre individus : définition d'une distance

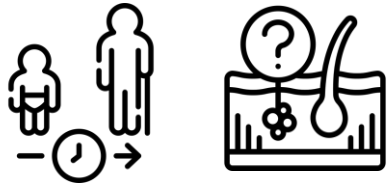
Définitions pas toutes compatibles

PARTIE 3
Impact et mitigation : un exemple en acceptation du risque

SEER : mélanome cutané, US



Variables sensibles



Variables non sensibles

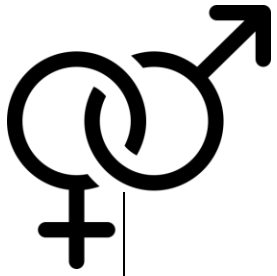
Taux différents selon sexe, origine, état civil

Problématique :
Modéliser des taux de mortalité "justes"

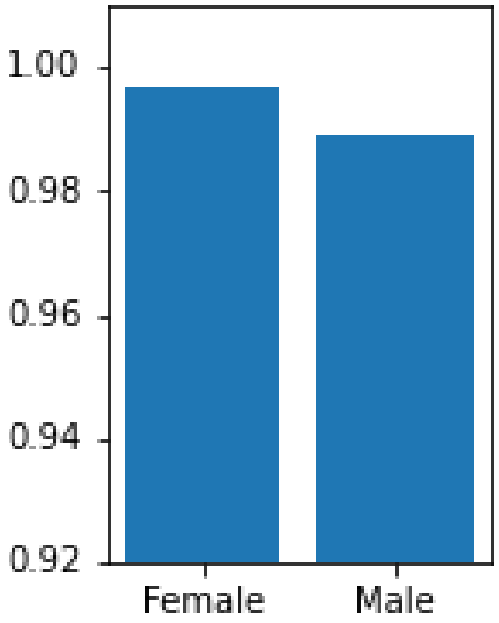
Choix d'une définition : parité statistique

Modèle simple et interprétable : régression logistique

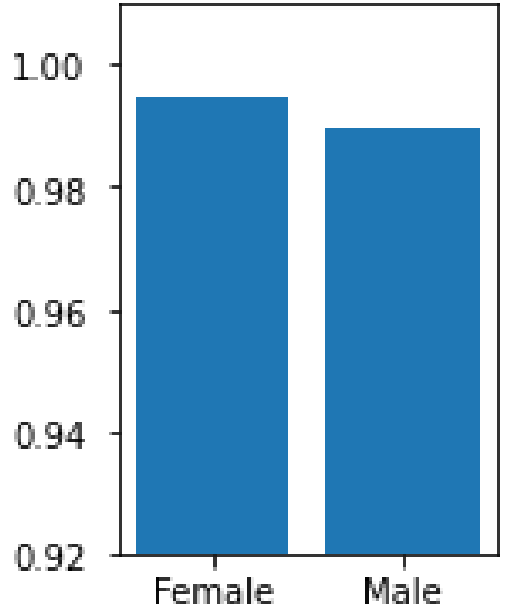
Taux d'acceptation par sexe



Modèle **avec** variables sensibles

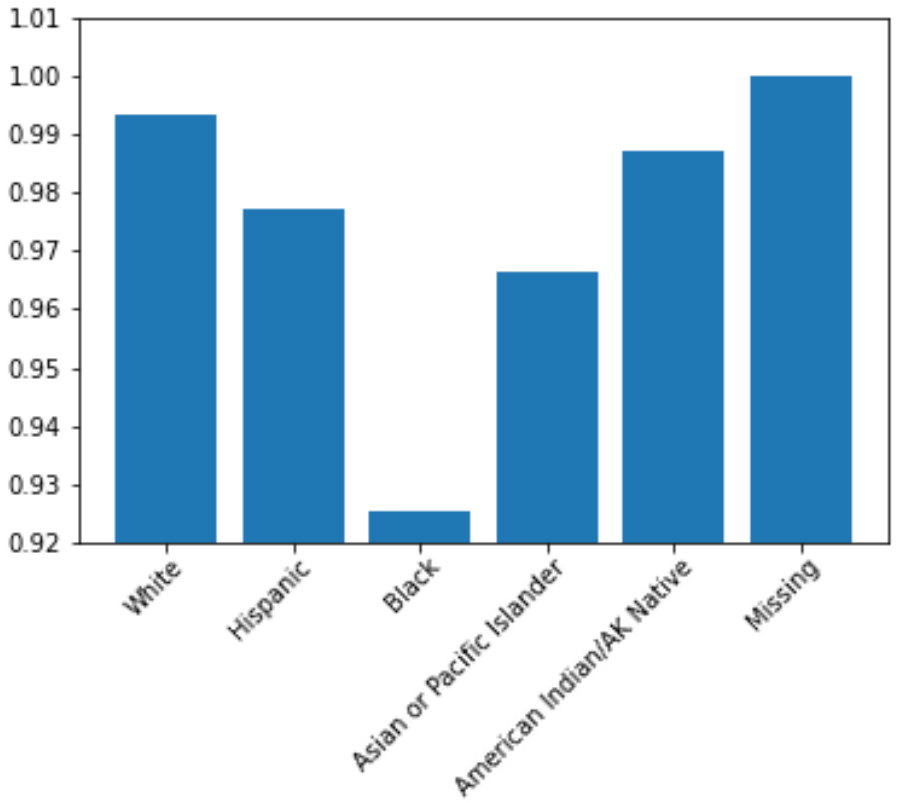


Modèle **sans** variables sensibles

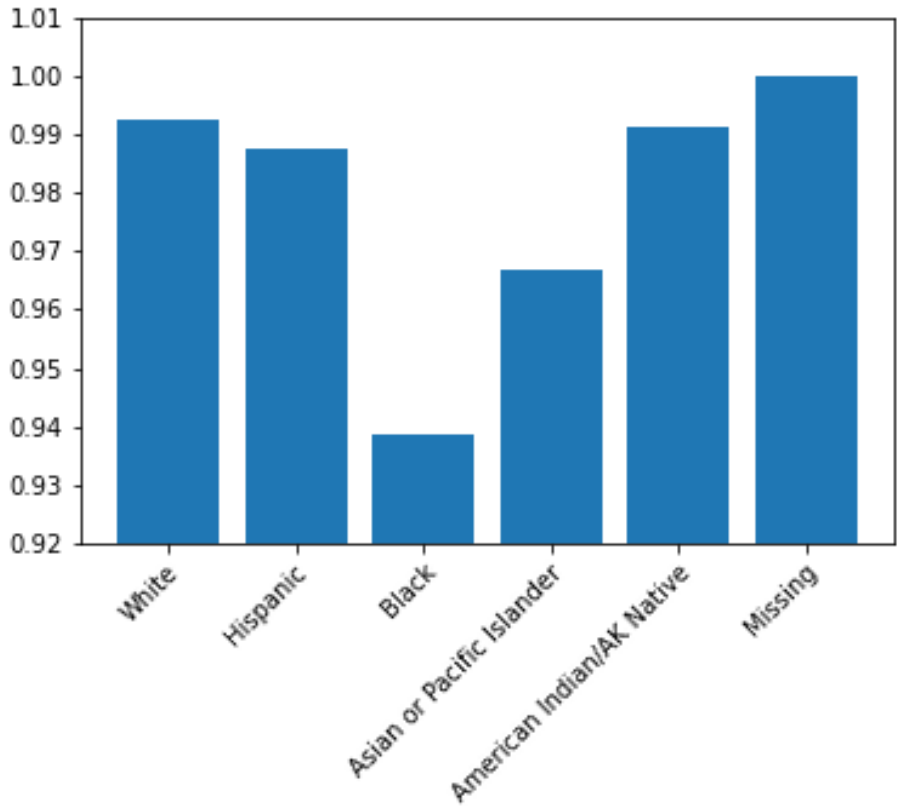


Taux d'acceptation par origine

Modèle **avec** variables sensibles

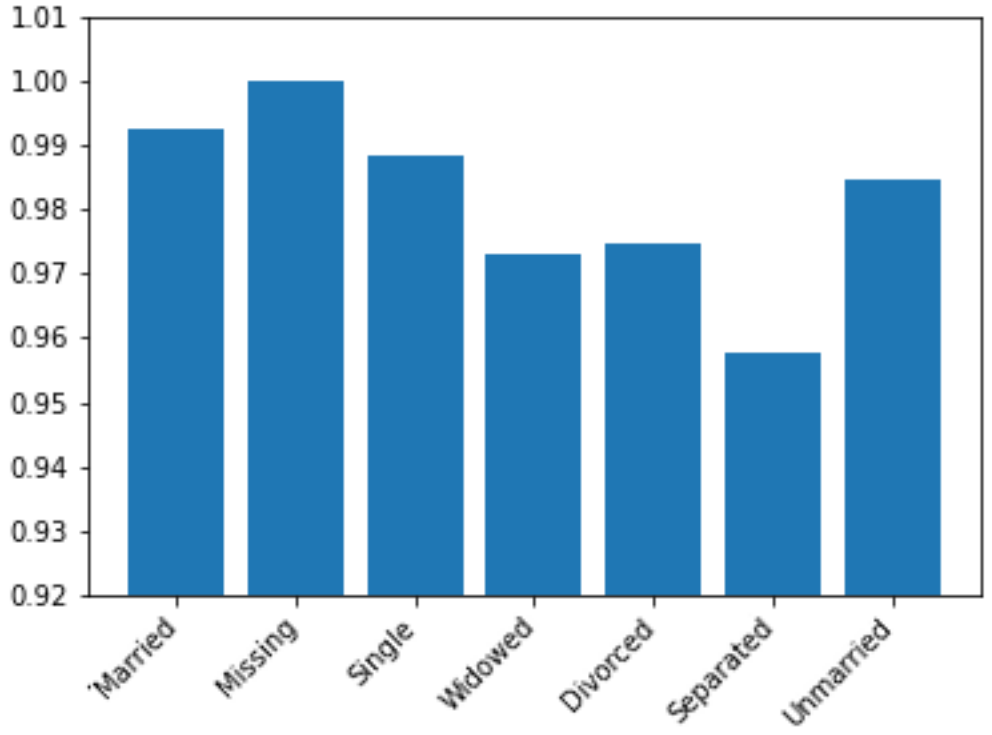


Modèle **sans** variables sensibles

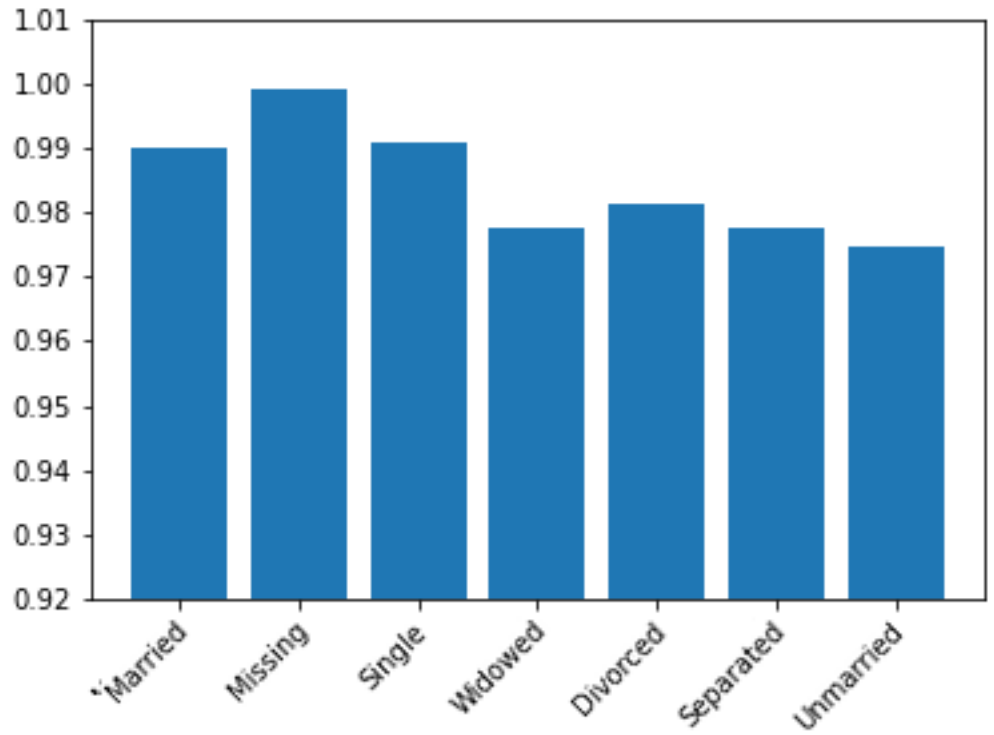


Taux d'acceptation par état civil

Modèle **avec** variables sensibles



Modèle **sans** variables sensibles



Modèle **avec** variables sensibles

↓

Discrimination directe

Modèle **sans** variables sensibles

↓

Discrimination indirecte

Résultats (équité) selon la structure de dépendance

Pas de mesure sans variables sensibles : collecte nécessaire

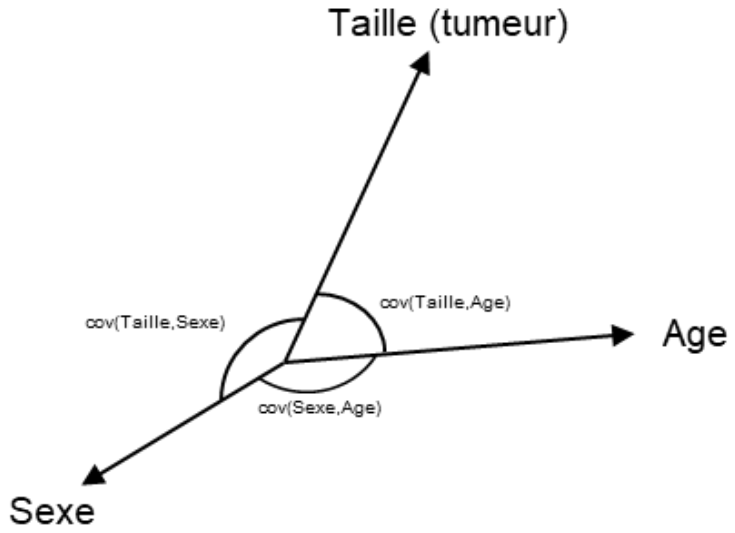
Comment éviter la discrimination indirecte ?

Réduire l'iniquité

Parité statistique pour l'équité : $\hat{Y} \perp\!\!\!\perp S$ ⟶ Indépendance ≈ corrélation (ordre 1)

Variables : vecteurs dans l'espace des variables centrées et de variance finie

Covariance = produit scalaire



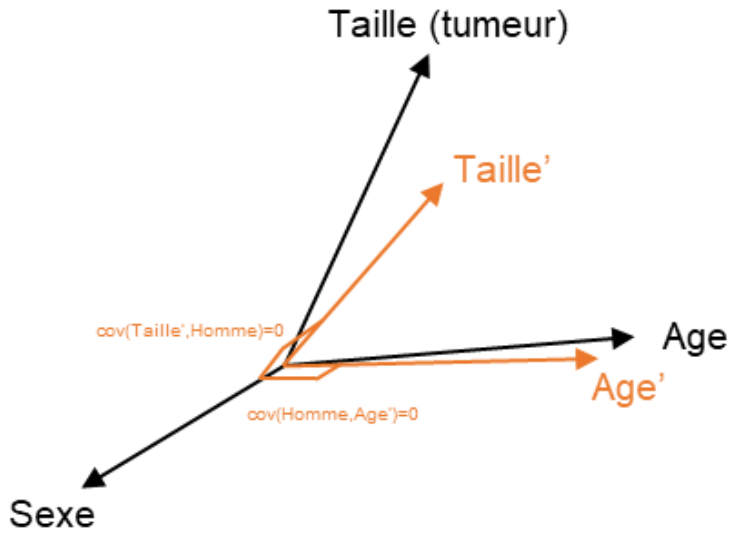
Absence de corrélation ⇔ vecteurs orthogonaux

Réduire l'iniquité

Parité statistique pour l'équité : $\hat{Y} \perp\!\!\!\perp S$ ⟶ Indépendance ≈ corrélation (ordre 1)

Variables : vecteurs dans l'espace des variables centrées et de variance finie

Covariance = produit scalaire



But : transformer les vecteurs non sensibles tels que Taille' ⊥ Sexe et Age' ⊥ Sexe

Utiliser variables transformées décorrélées dans modèle

Absence de corrélation ⇔ vecteurs orthogonaux

Zoom sur la méthode du changement de base :

1 → s : sensibles
s+1 → n : non sensibles

- Les vecteurs sensibles $u_1, \ldots, u_s$ ne changent pas
- Les vecteurs non sensibles transformés sont orthogonaux aux vecteurs sensibles

Pour chaque nouveau vecteur u_k' :

$u_k' \perp u_1, \ldots, u_k' \perp u_s$ ⟶ s équations, s+1 inconnues ⟶ Infinité de solutions

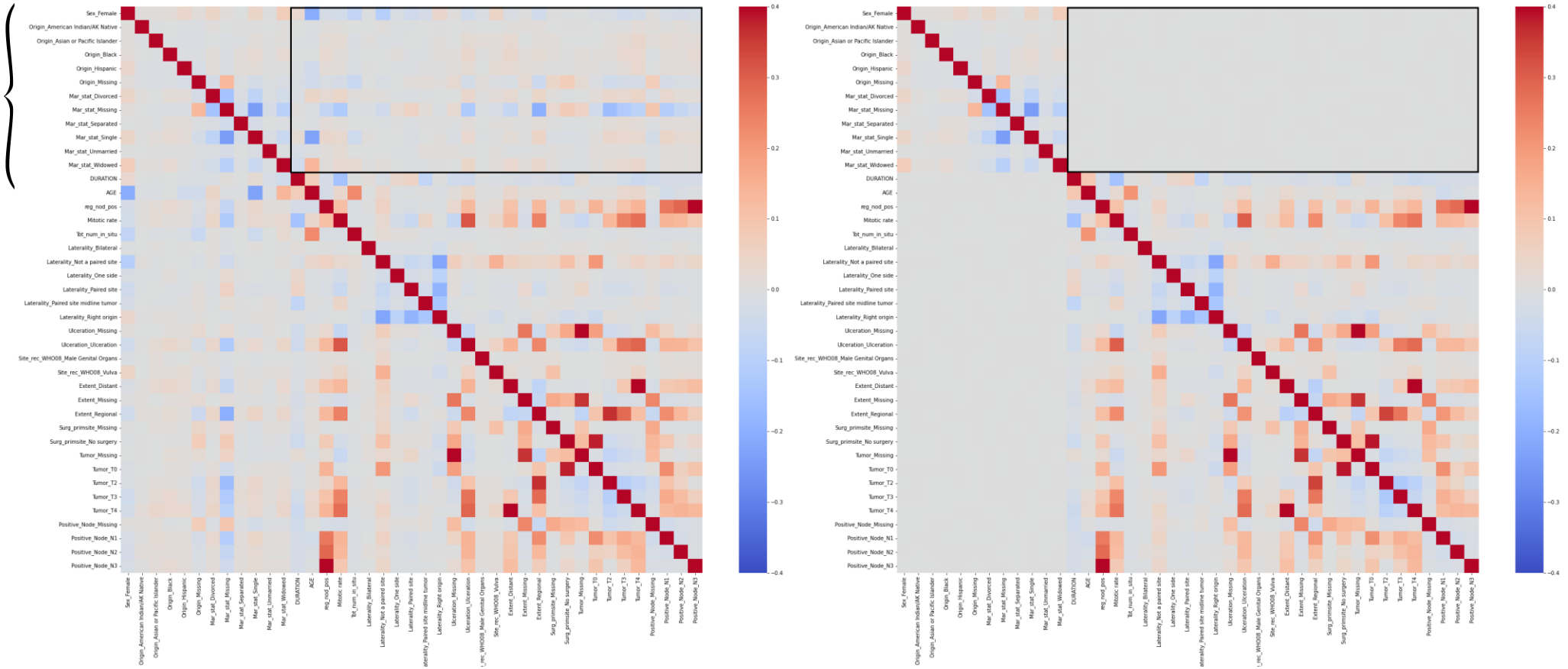
Idée : **vecteur transformé proche de celui d'origine** : min $d(u_k', u_k)$

↳ distance définie par le produit scalaire : $d(u,v) = < u-v, u-v >$

Nouveau vecteur = combinaison linéaire ancien et sensibles

⇒ Pour chaque vecteur transformé : s+1 équations linéaires à s+1 inconnues

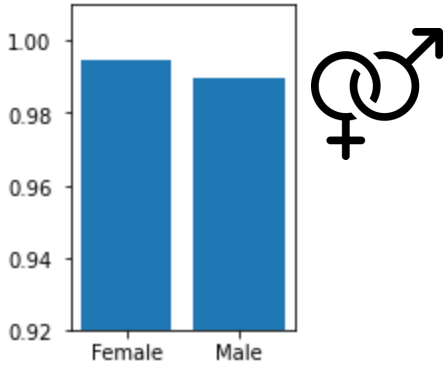
Sensibles

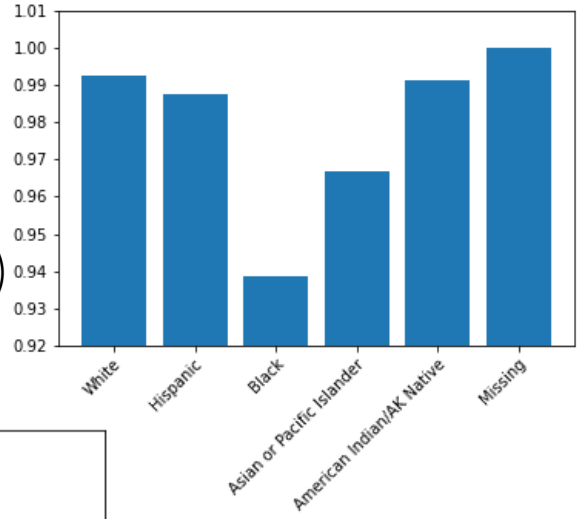


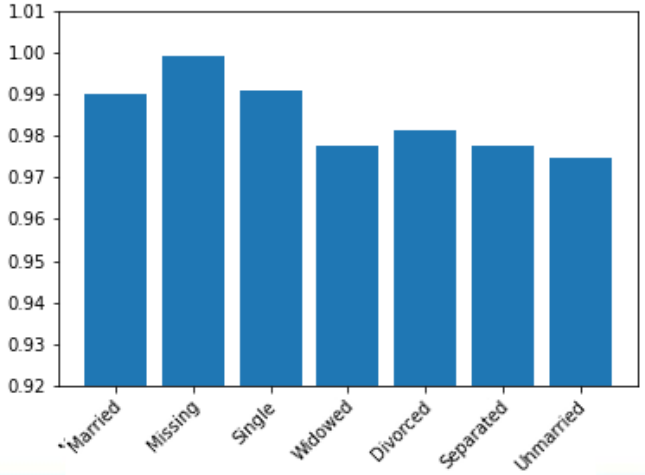
Variables d'origine

Variables transformées

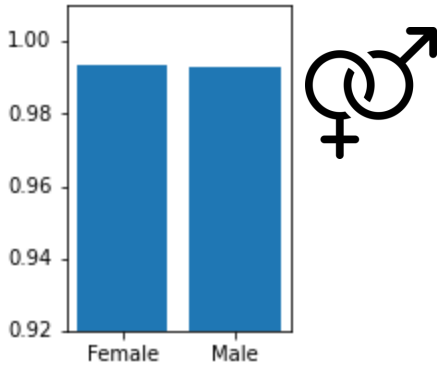
Modèle **sans** variables sensibles

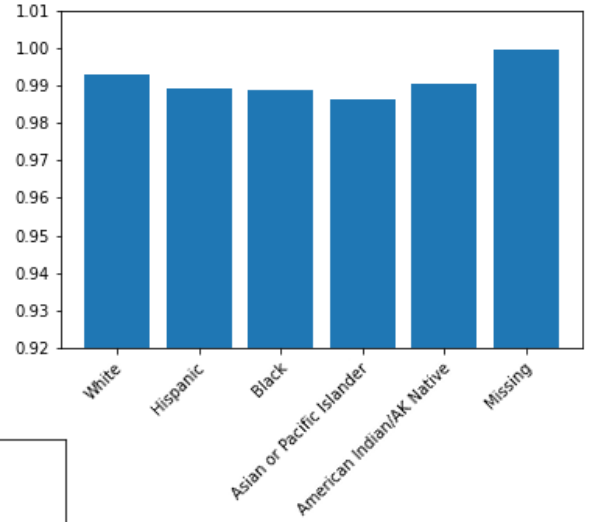


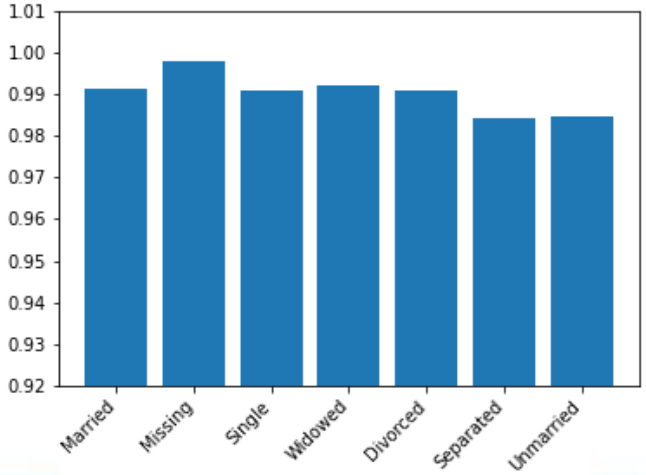




Modèle avec variables **transformées**





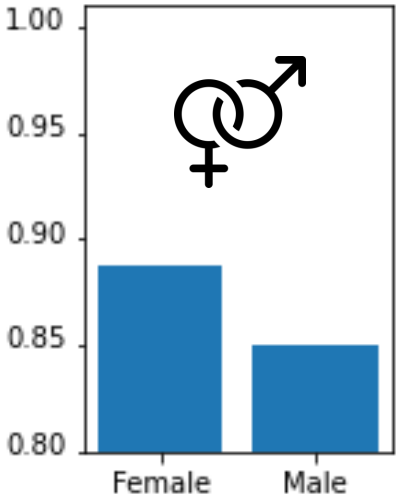


Différences restantes : dépendance non linéaire

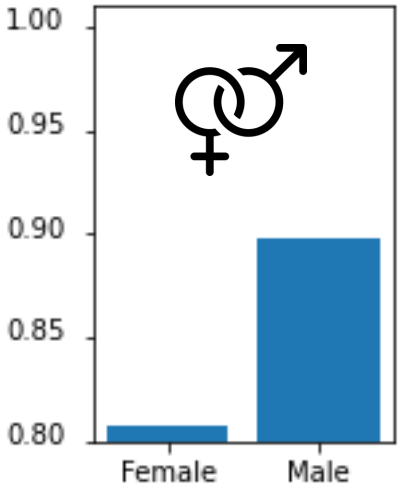
En regardant une autre définition d'équité

Egalité des chances : mêmes taux de vrais et de faux positifs

FPR sans variables sensibles



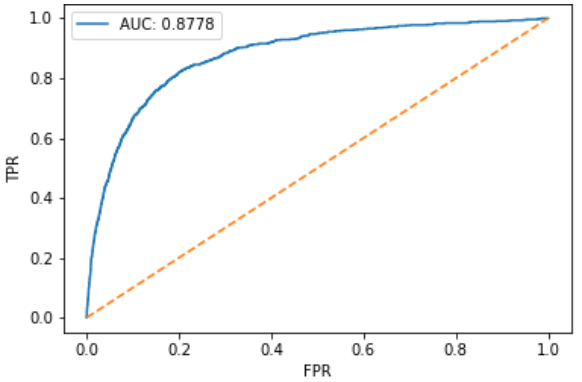
FPR avec variables transformées

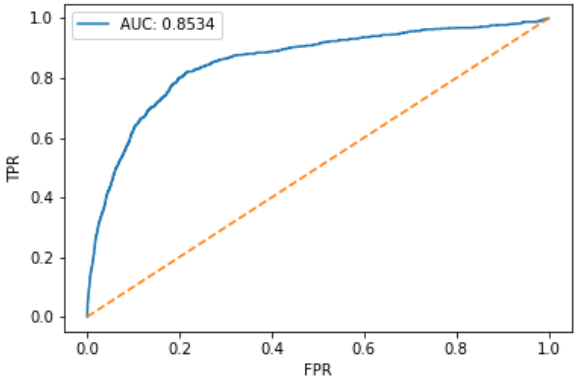


Incompatibilité des définitions d'équité

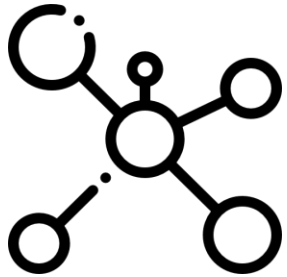
Et la performance ?

Légère baisse de l'AUC



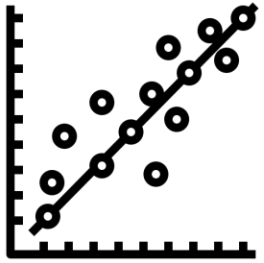


Compromis équité-performance



Modèles complexes peuvent détecter des dépendances non linéaires

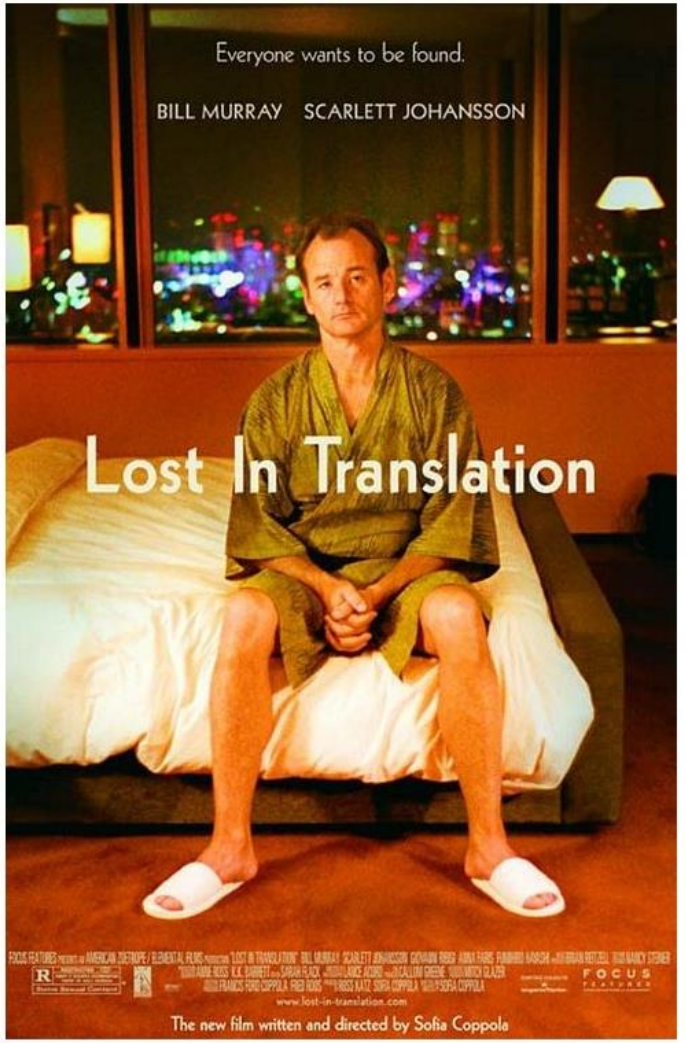
Aller au-delà de la corrélation



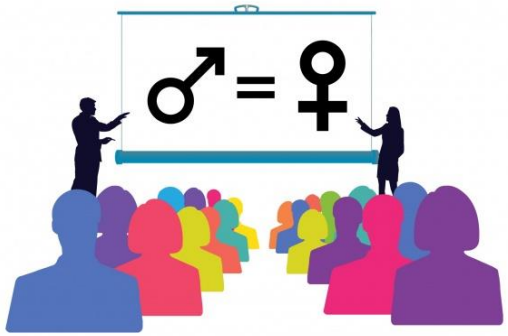
Quels critères pour la régression ?

Quelle est la bonne métrique ?

PARTIE 4
Ouverture sur les approches causales



Tarif et causalité : retour sur la Gender Directive



2.3. L'utilisation d'autres facteurs d'évaluation des risques

2.3.1. *Facteurs corrélés au sexe: la question de la discrimination indirecte*

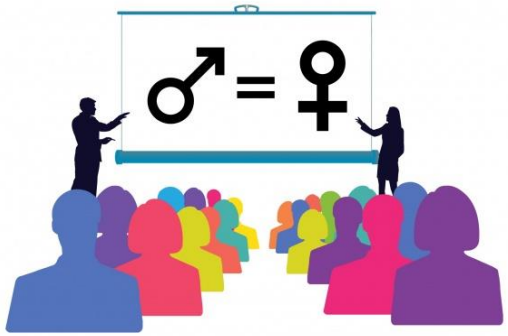
16. L'arrêt Test-Achats se borne à examiner l'utilisation du sexe comme facteur d'évaluation des risques, sans aborder la recevabilité d'autres facteurs utilisés par les assureurs. Toutefois, conformément à l'article Article 2, point b), de la directive, il y a discrimination indirecte lorsqu'un facteur de risque apparemment neutre désavantage en particulier les personnes d'un sexe. Contrairement à la discrimination directe, la discrimination indirecte peut être justifiée si le but est légitime et si les moyens d'y parvenir sont appropriés et nécessaires.

17. L'utilisation de facteurs de risque susceptibles d'être corrélés au sexe reste par conséquent possible, dès lors qu'il s'agit bel et bien de facteurs de risque réels [3].

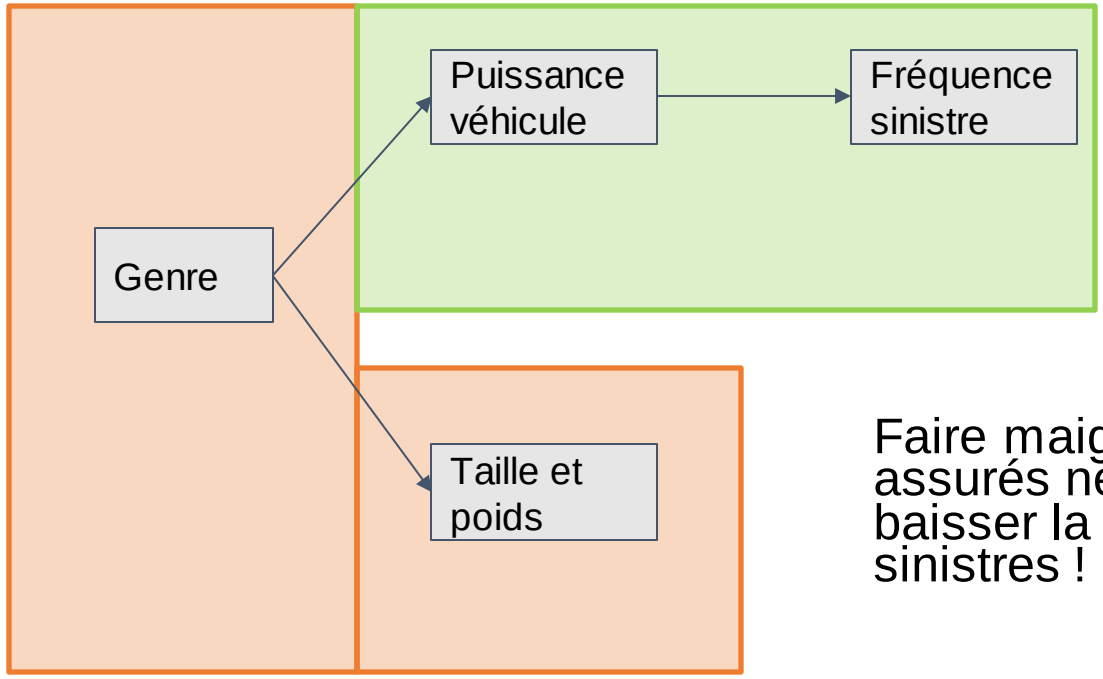
[3] Par exemple, une différenciation des prix fondée sur la taille du moteur de la voiture dans le domaine de l'assurance-automobile doit rester possible, même si statistiquement les hommes conduisent des voitures au moteur plus puissant. Tel ne serait pas le cas d'une différenciation, en matière d'assurance-automobile, fondée sur la taille ou le poids d'une personne.

https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:011:0001:0011:fr:PDF

Tarif et causalité : retour sur la Gender Directive



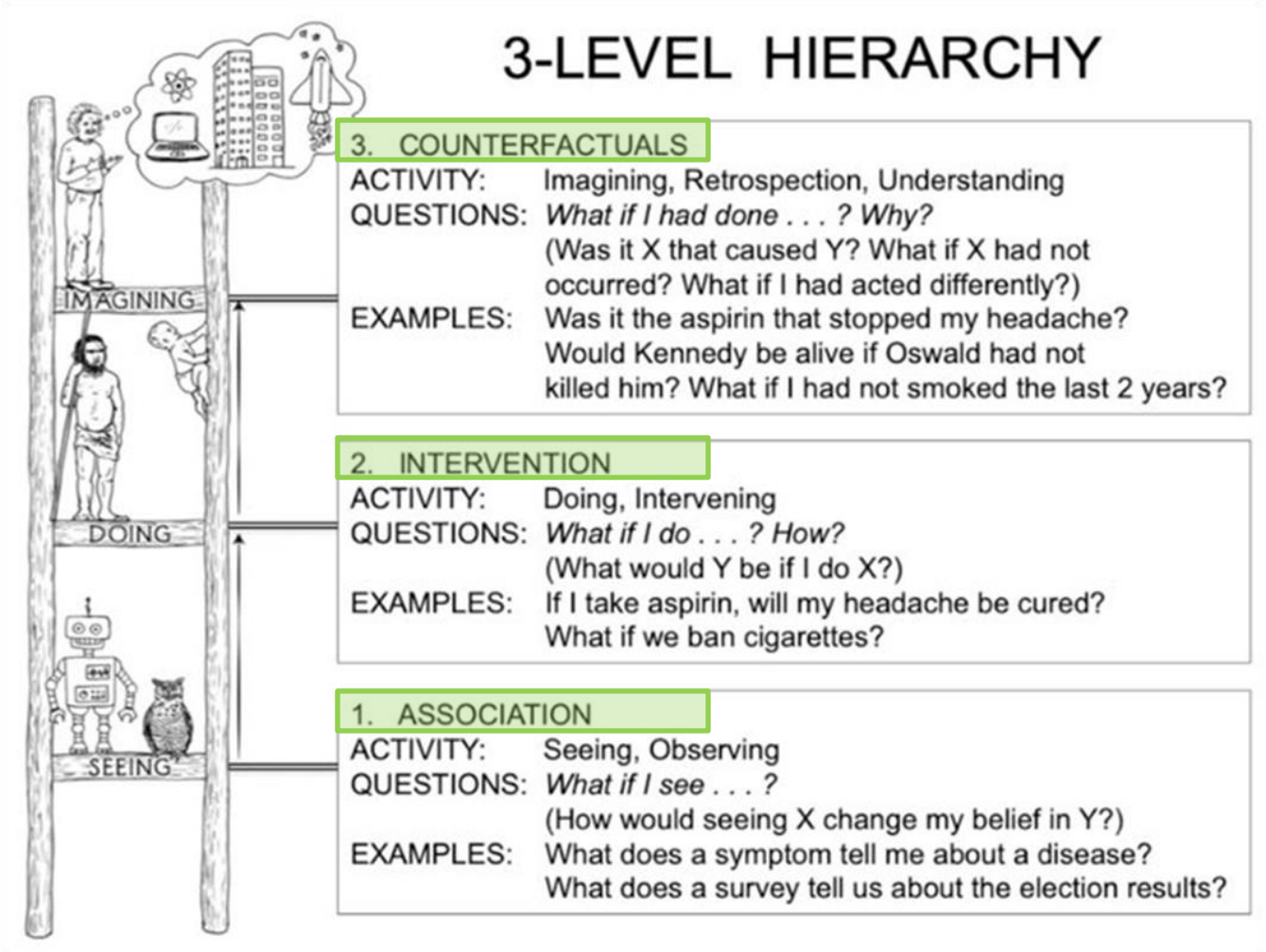
Par exemple, une différenciation des prix fondée sur la taille du moteur de la voiture dans le domaine de l'assurance-automobile doit rester possible, même si statistiquement les hommes conduisent des voitures au moteur plus puissant. Tel ne serait pas le cas d'une différenciation, en matière d'assurance-automobile, fondée sur la taille ou le poids d'une personne.



Faire maigrir vos assurés ne fera pas baisser la fréquence des sinistres !

Les graphes causaux permettent d'expliquer les hypothèses au-delà des données

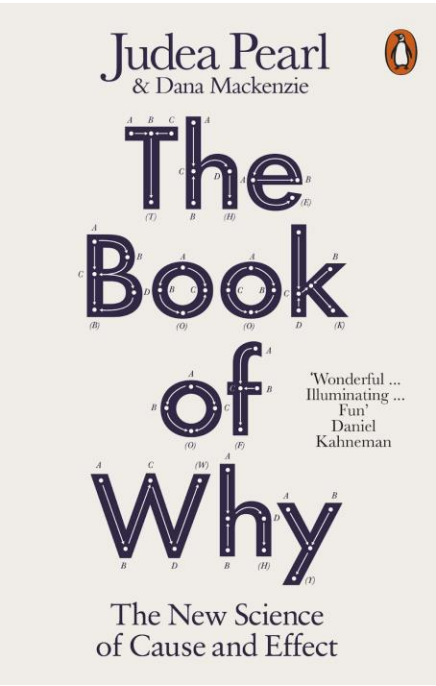
Il existe un langage pour exprimer la causalité !



3. Counterfactuals
$P(y_x|x', y')$

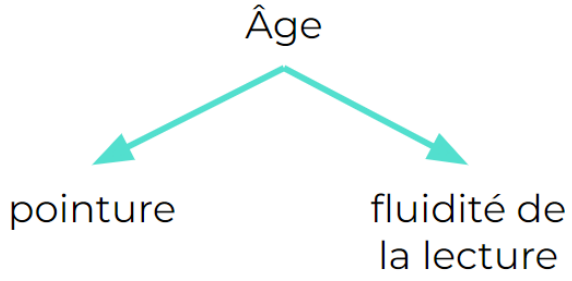
2. Intervention
$P(y|do(x)\ z)$

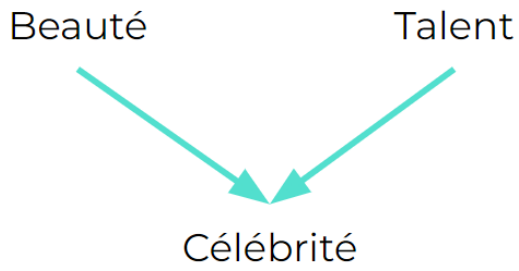
1. Association
$P(y|x)$



http://bayes.cs.ucla.edu/jp_home.html

Il existe un langage pour exprimer la causalité : graphes causaux





$alarme \perp feu \mid fumée$

$pointure \not\perp fluence$

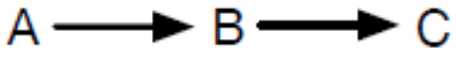
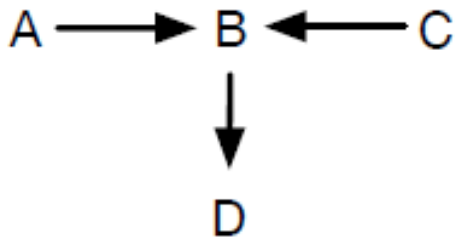
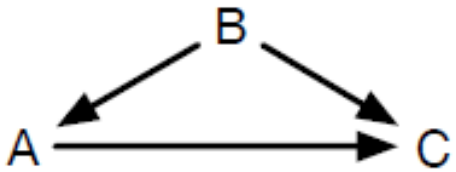
$pointure \perp fluence \mid âge$

$beauté \perp talent$

$beauté \not\perp talent \mid célébrité$

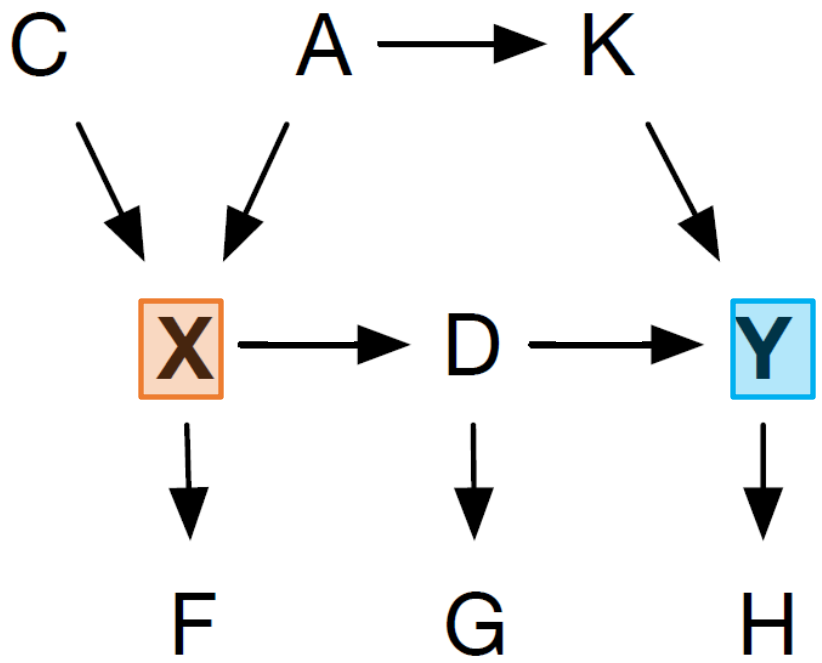
The Book of Why

En quoi est-il important d'expliciter le graphe causal ?

Graph:	Regression:	Implications:
	$\mathbb{E}[C\|\cancel{A}, B]$	$A \perp\!\!\!\perp C\|B$
	$\mathbb{E}[C\|\cancel{A}]$	$A \perp\!\!\!\perp C$
	$\mathbb{E}[C\|A, D]$	$A \not\perp\!\!\!\perp C\|D$
	$\mathbb{E}[C\|A, B]$	$A \perp\!\!\!\perp B$

En quoi est-il important d'expliciter le graphe causal ?

Processus de génération des données (supposé inconnu)



Variable cible

Variable dont on pense a priori qu'elle peut avoir un impact

Quelle importance des variables explicatives sur E[Y |…] ?

Régression linéaire (LR) → Coefficients

Forêt aléatoire (RF) → Importance par permutation

Forêt aléatoire (RF) → Valeurs de Shapley

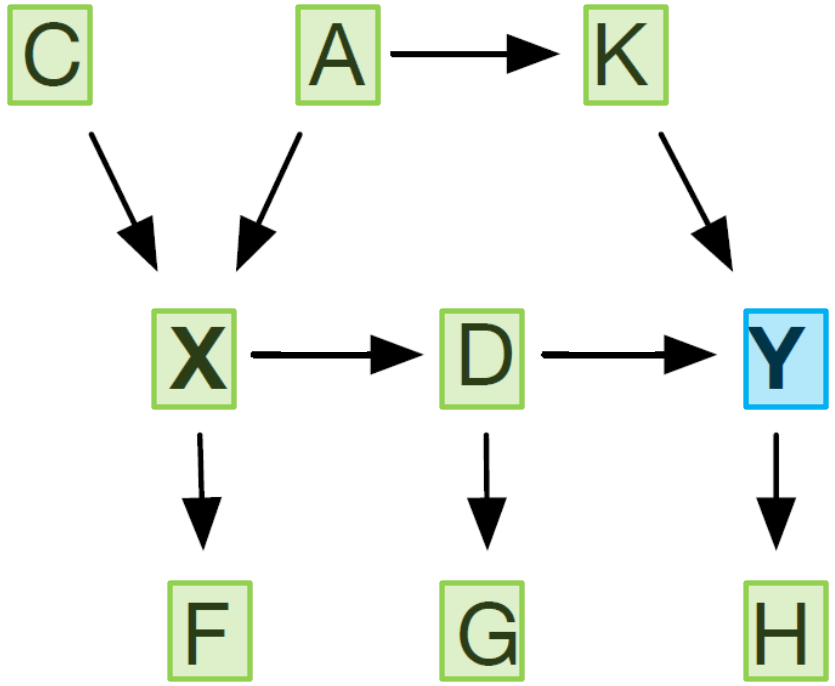
Réseau de neurones (NN) → Valeurs de Shapley

Tri à plat → Corrélation avec Y

Matthew J. Vowels, 2022, https://arxiv.org/abs/2202.09875

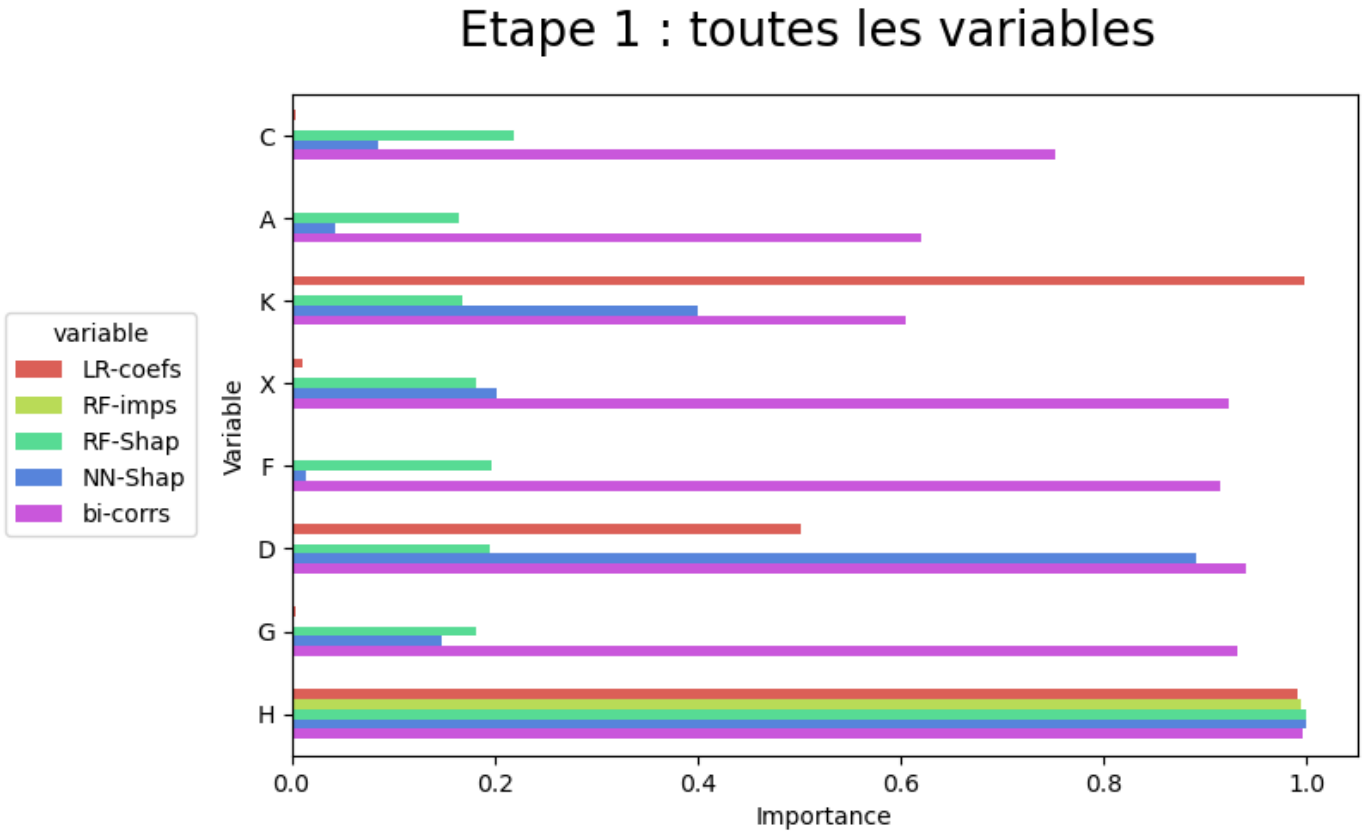
En quoi est-il important d'expliciter le graphe causal ?

Processus de génération des données (supposé inconnu)



Variable cible Variable incluse dans la régression

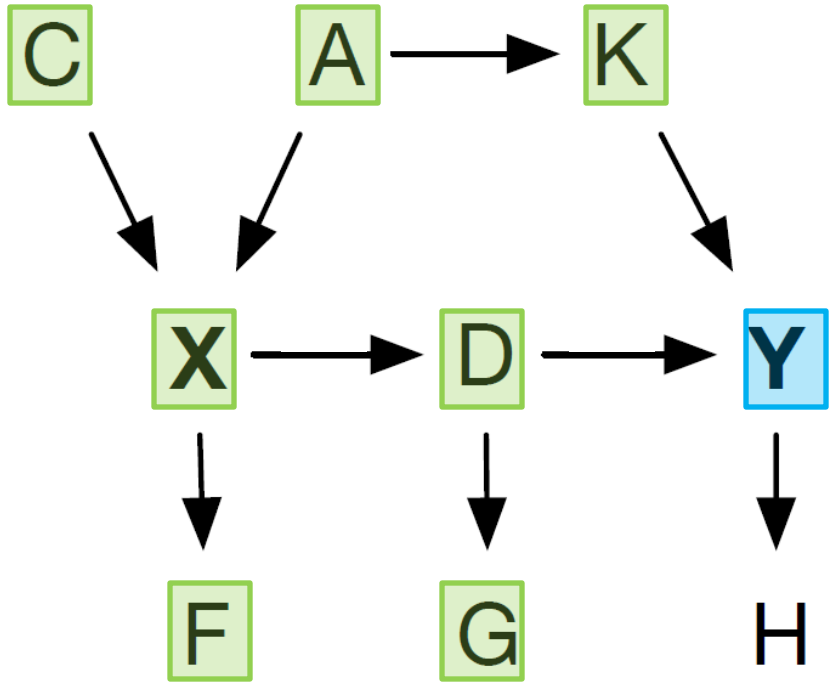
Quelle importance des variables explicatives sur E[Y |…] ?



Matthew J. Vowels, 2022, https://arxiv.org/abs/2202.09875

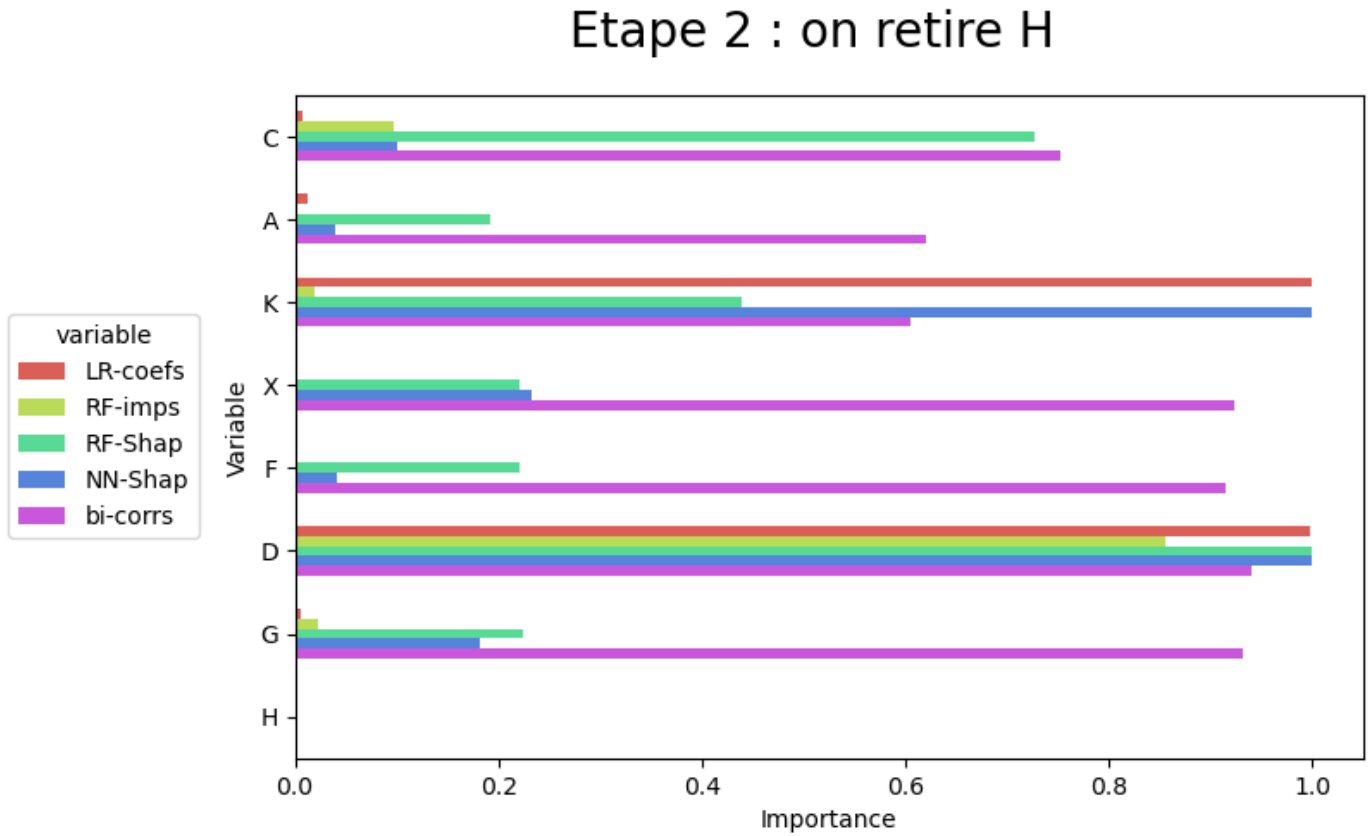
En quoi est-il important d'expliciter le graphe causal ?

Processus de génération des données (supposé inconnu)



Variable cible Variable incluse dans la régression

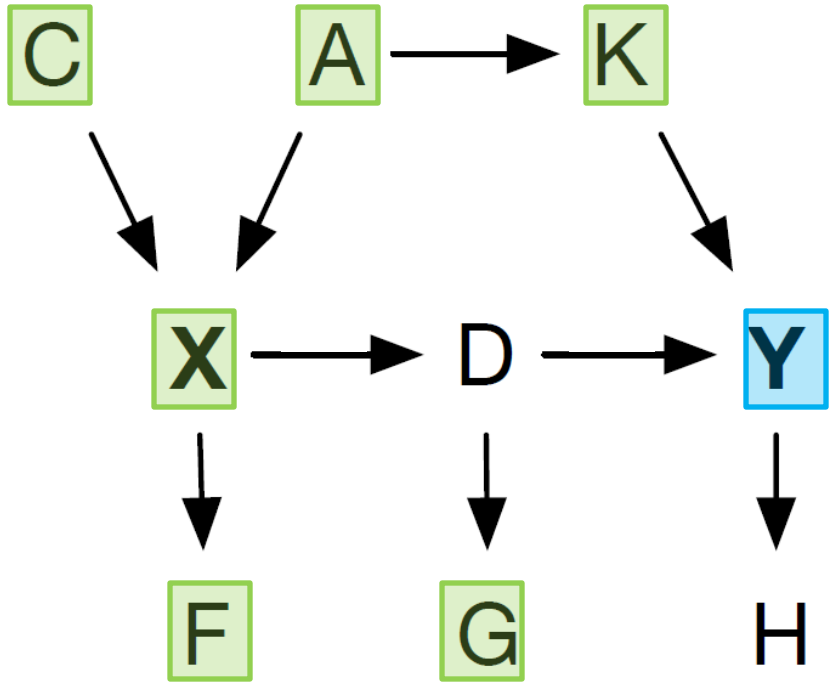
Quelle importance des variables explicatives sur E[Y |…] ?



Matthew J. Vowels, 2022, https://arxiv.org/abs/2202.09875

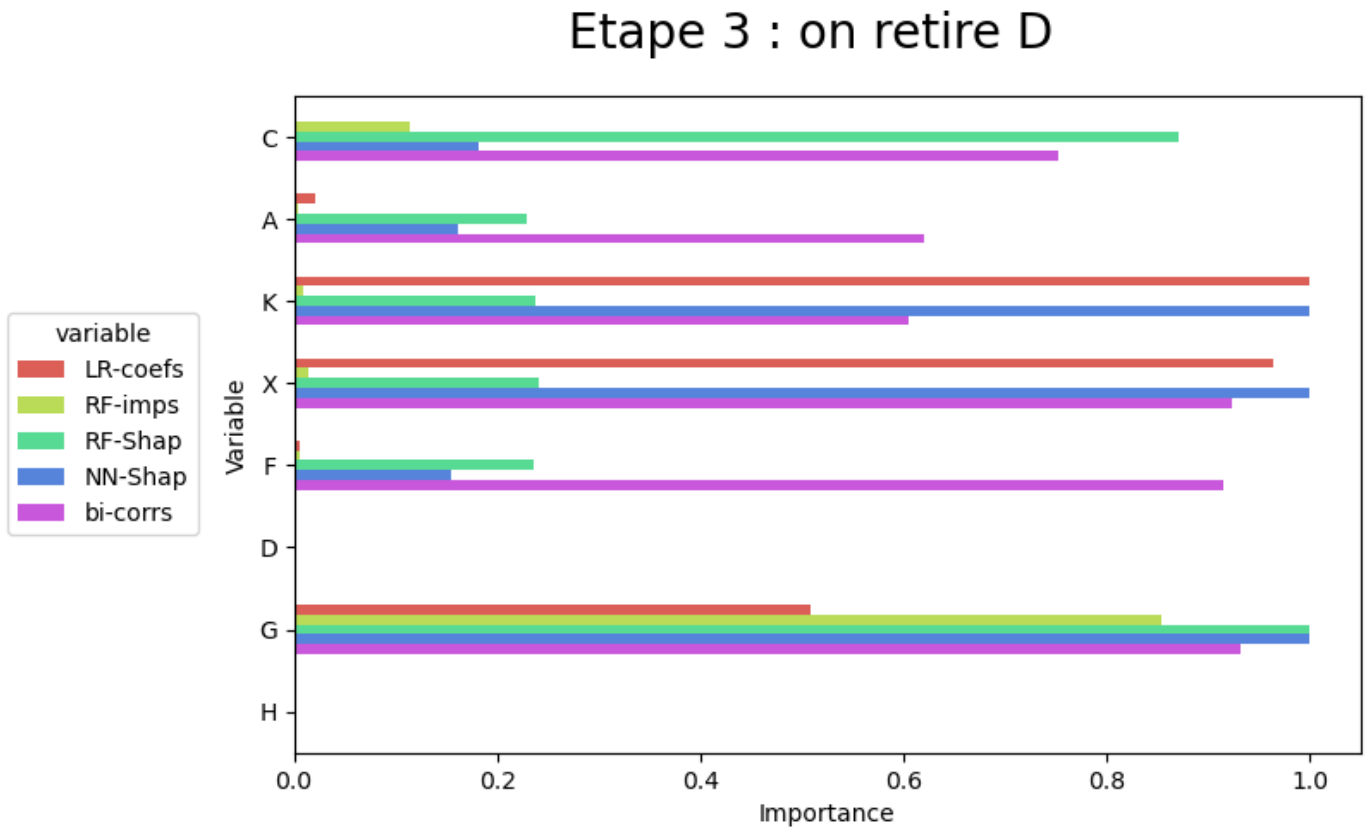
En quoi est-il important d'expliciter le graphe causal ?

Processus de génération des données (supposé inconnu)



Variable cible Variable incluse dans la régression

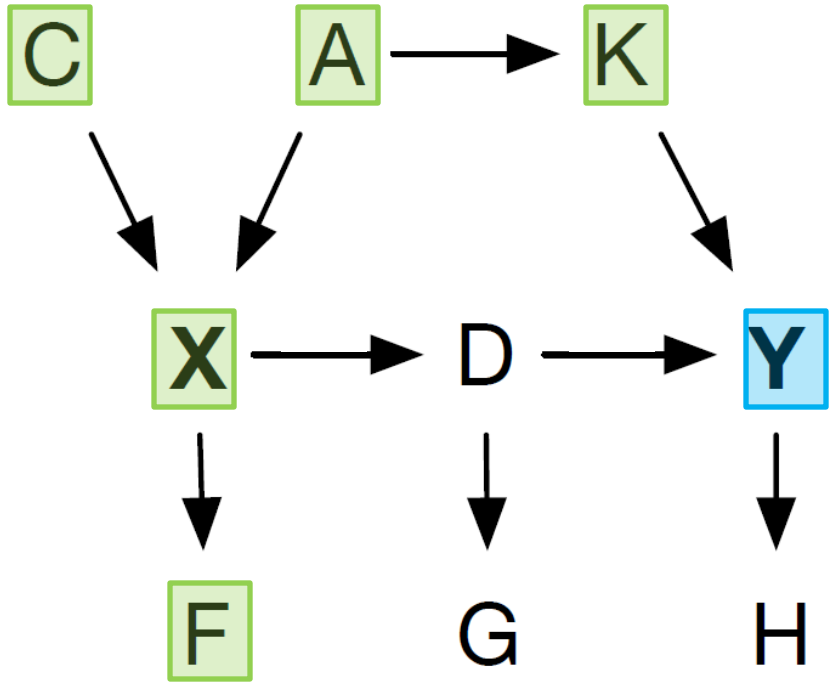
Quelle importance des variables explicatives sur E[Y |…] ?



Matthew J. Vowels, 2022, https://arxiv.org/abs/2202.09875

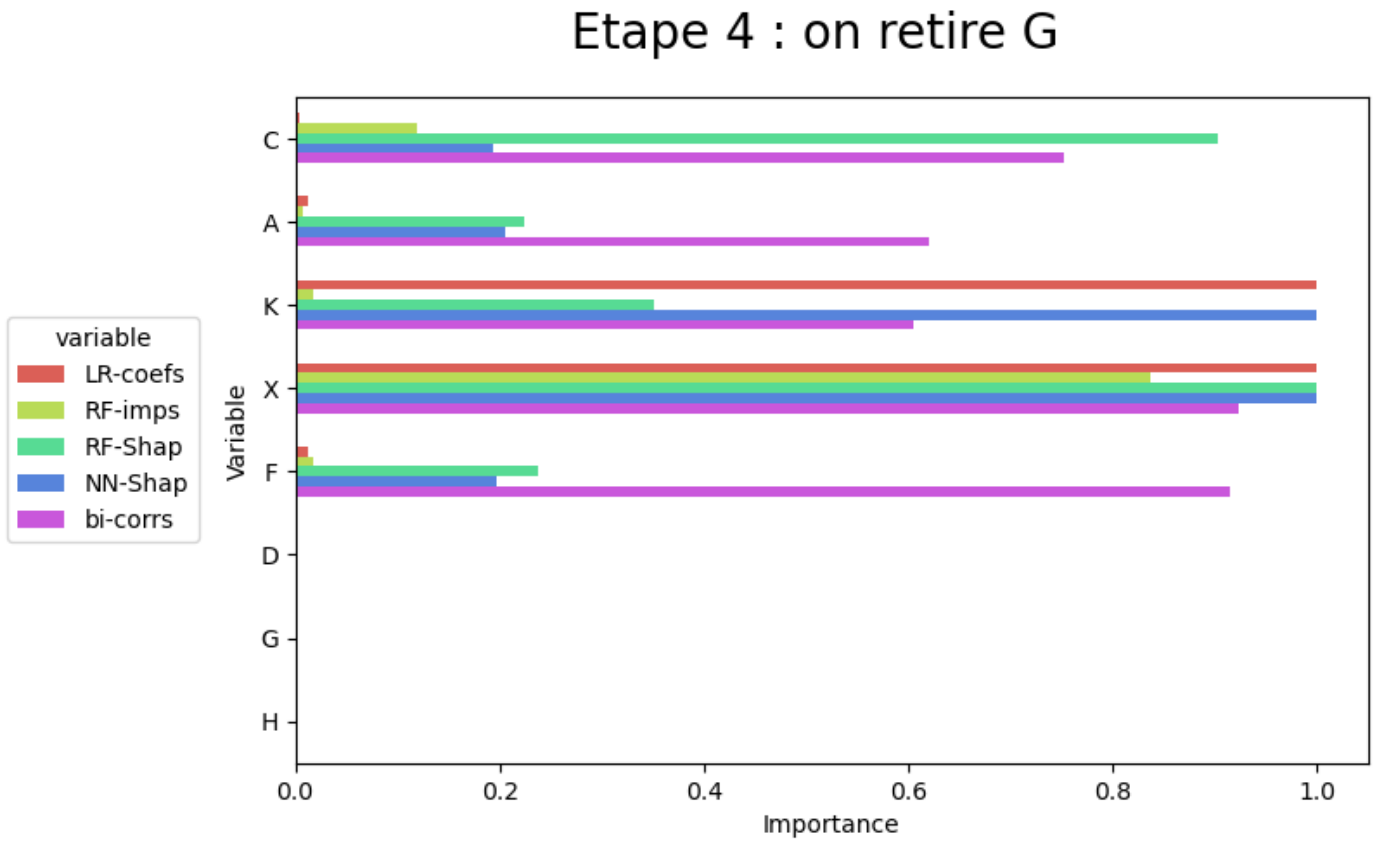
En quoi est-il important d'expliciter le graphe causal ?

Processus de génération des données (supposé inconnu)



Variable cible Variable incluse dans la régression

Quelle importance des variables explicatives sur E[Y |…] ?



Matthew J. Vowels, 2022, https://arxiv.org/abs/2202.09875

Comment trouver un graphe causal compatible avec les données ?

Un vrai sujet de recherche avec de nombreuses pistes

D'ya like DAGs? A Survey on Structure Learning and Causal Discovery

MATTHEW J. VOWELS*, NECATI CIHAN CAMGOZ, and RICHARD BOWDEN, CVSSP, University of Surrey, U.K.

Causal reasoning is a crucial part of science and human intelligence. In order to discover causal relationships from data, we need structure discovery methods. We provide a review of background theory and a survey of methods for structure discovery. We primarily focus on modern methods which leverage continuous optimization, and provide reference to further resources such as benchmark datasets and software packages. Finally, we discuss the assumptive leap required to take us from structure to causality.

1 INTRODUCTION

Causal understanding has been described as 'part of the bedrock of intelligence' [145], and is one of the fundamental goals of science [11, 70, 183, 241–243]. It is important for a broad range of applications, including policy making [136], medical imaging [30], advertisement [22], the development of medical treatments [189], the evaluation of evidence within legal frameworks [183, 218], social science [82, 96, 246], biology [235], and many others. It is also a burgeoning topic in machine learning and artificial intelligence [17, 66, 76, 144, 210, 247, 255], where it has been argued that a consideration for causality is crucial for reasoning about the world. In order to discover causal relations, and thereby gain causal understanding, one may perform interventions and manipulations as part of a randomized experiment. These experiments may not only allow researchers or agents to identify causal relationships, but also to estimate the magnitude of these relationships.

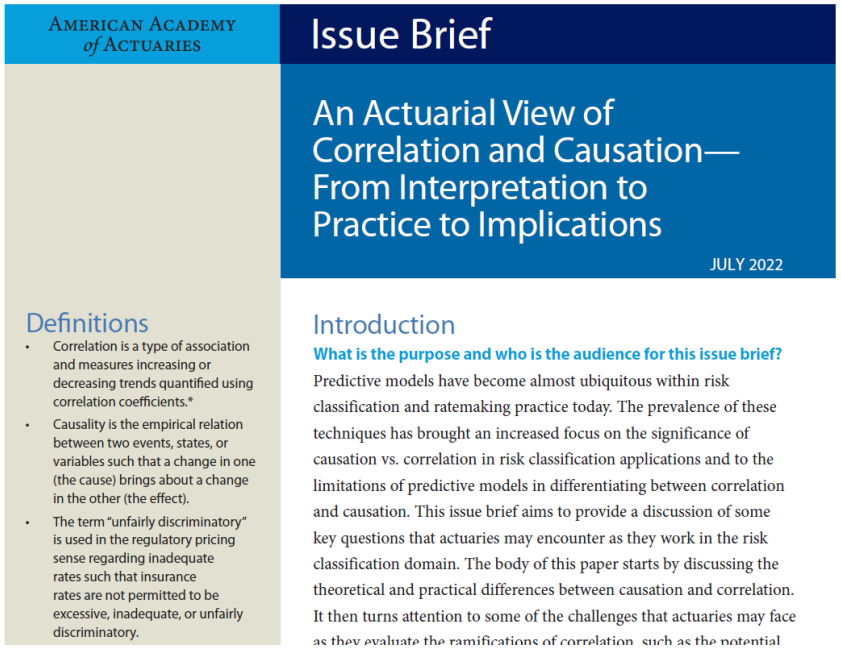
Unfortunately, in many cases, it may not be possible to undertake such experiments due to prohibitive cost, ethical concerns, or impracticality. For example, to understand the impact of smoking, it would be necessary to force different individuals to smoke or not-smoke. Researchers are therefore often left with non-experimental, observational data. In the absence of intervention and manipulation, observational data leave researchers facing a number of challenges: Firstly, observational datasets may not contain all relevant variables - there may exist unobserved/hidden/latent factors (this is sometimes referred to as the third variable problem). Secondly, observational data may exhibit selection bias - for example, younger patients may in general prefer to opt for surgery, whereas older patients may prefer medication. Thirdly, the causal relationships underlying these data may not be known *a priori* - for example, are genetic factors independent causes of a particular outcome, or do they mediate or moderate an outcome? These three challenges affect the discovery and estimation of causal relationships.

To address these challenges, researchers in the fields of statistics and machine learning have developed numerous methods for uncovering causal relations (causal discovery) and estimating the magnitude of these effects (causal inference) from observational data, or from a mixture of

*Corresponding author.

Authors' address: Matthew J. Vowels, m.j.vowels@surrey.ac.uk; Necati Cihan Camgoz, n.camgoz@surrey.ac.uk; Richard Bowden, r.bowden@surrey.ac.uk, CVSSP, University of Surrey, Guildford, Surrey, GU2 7XH, U.K.

Mais une prise de conscience et un attrait de plus en plus marqué dans la communauté actuarielle



AMERICAN ACADEMY of ACTUARIES

Issue Brief

An Actuarial View of Correlation and Causation—From Interpretation to Practice to Implications

JULY 2022

Definitions

- Correlation is a type of association and measures increasing or decreasing trends quantified using correlation coefficients.*
- Causality is the empirical relation between two events, states, or variables such that a change in one (the cause) brings about a change in the other (the effect).
- The term "unfairly discriminatory" is used in the regulatory pricing sense regarding inadequate rates such that insurance rates are not permitted to be excessive, inadequate, or unfairly discriminatory.

Introduction

What is the purpose and who is the audience for this issue brief?

Predictive models have become almost ubiquitous within risk classification and ratemaking practice today. The prevalence of these techniques has brought an increased focus on the significance of causation vs. correlation in risk classification applications and to the limitations of predictive models in differentiating between correlation and causation. This issue brief aims to provide a discussion of some key questions that actuaries may encounter as they work in the risk classification domain. The body of this paper starts by discussing the theoretical and practical differences between causation and correlation. It then turns attention to some of the challenges that actuaries may face as they evaluate the ramifications of correlation, such as the potential

https://www.actuary.org/sites/default/files/2022-07/Correlation.IB_.6.22_final.pdf

Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. (2022). DISCRIMINATION-FREE INSURANCE PRICING. ASTIN Bulletin, 52(1), 55-89. https://doi:10.1017/asb.2021.23

https://arxiv.org/abs/2103.02582

Conclusion

Question ouverte

Académiquement

Réglementairement

Sanction immédiate : préapprobation sur dossier par organisme externe

Explicitation hypothèses conscientes / inconscientes : biais de sélection, narratif autour du risque, construction des modèles

Mouvement mondial

Besoin pour les actuaires de se saisir du sujet

Articulation plus forte avec la direction, sujet plus uniquement technique