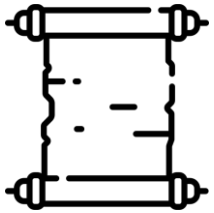
IA et éthique en assurance : une nouvelle solution pour atténuer la discrimination par proxy dans la modélisation du risque

Marguerite Saucé
ENSAE
28 juin 2023

Introduction



Sujet ancien, notions complexes et philosophiques

Modèles dénoncés comme discriminants dans de nombreux secteurs

Attention croissante des régulateurs et du public sur les sujets de discrimination et d'équité

4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

Sommaire

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination

1. Concepts clés et réglementation sur la discrimination
1.1 L'assurance vie

Segmentation et mutualisation

Les pertes sont la responsabilité collective du pool

Classes homogènes : adapter les primes aux profils de risque et éviter l'antisélection

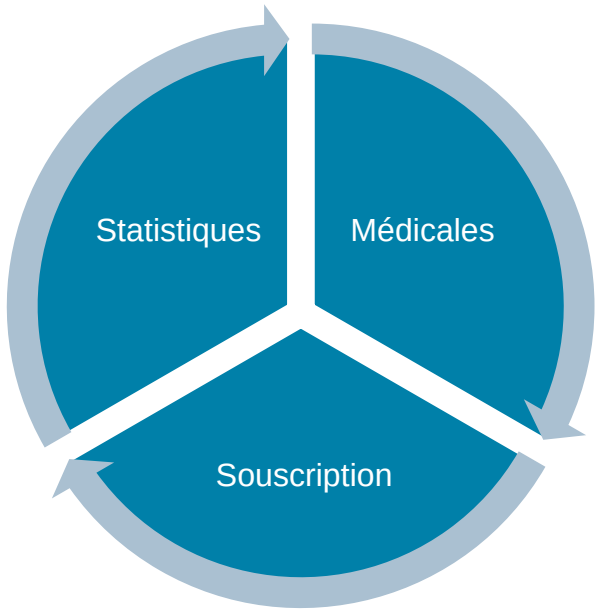
Equité actuarielle

Les assurés contribuent proportionnellement à leur risque

Mais comment déterminer le risque réel ?

Modélisation du risque

Trois types de contraintes

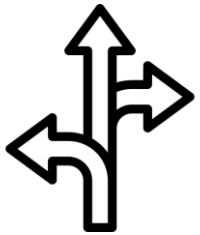


Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.2 Equité et biais

La nécessité de pratiques équitables

Réputation et régulation



Une multitude de points de vue

Légal, éthique, statistique

Réglementation actuelle : éviter l'utilisation de certains critères

→ Comment vérifier s'il y a discrimination ?

Les types de biais

I – les classes ne reflètent pas le risque

II – les classes reflètent une corrélation non causale avec le risque

III – les classes reflètent une réalité statistique causale mais éthiquement inacceptable

Institut
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.2 Equité et biais dans les données

Directe

Les variables protégées sont explicitement utilisées pour prendre la décision

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

Indirecte

Le traitement est apparemment neutre (basé sur des variables non protégées) mais les groupes sont traités différemment

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

MAIS

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

Proxys : inférence

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.2 Equité et biais dans les données

Directe

Les variables protégées sont explicitement utilisées pour prendre la décision

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

Indirecte

Le traitement est apparemment neutre (basé sur des variables non protégées) mais les groupes sont traités différemment

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

MAIS

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

Quelles sont les variables protégées ?

Proxys : inférence

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.2 Equité et biais dans les données

Quelles sont les variables protégées ?



Définies par la loi : dépendent de la juridiction, du secteur d'activité…

Réglementation actuelle :
- Interdiction d'utiliser les variables protégées
- Eviter d'utiliser certains proxys

Pas de règles pour mesurer la discrimination indirecte en assurance

Quelques premiers efforts de définition aux Etats-Unis avec le *Disparate Impact*

Institut
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.3 L'expression de l'équité et du biais dans les données

Variables sensibles

En France : grossesse, information génétique

Aux Etats-Unis : seulement 9 Etats interdisent l'utilisation de l'origine nationale

Variables proxy

Nome → sexe et origine nationale

Adresse → origine nationale (surtout aux Etats-Unis)

Métier → sexe

De très nombreuses variables peuvent servir de proxy

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.4 Comment mesurer l'équité

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

Matrice de confusion

<u>Cadre défini</u>
S variable protégée
Classification binaire : Y classe réelle et Ŷ classe estimée avec un modèle linéaire

<u>Equité de groupe</u>

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

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

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

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

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

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

<u>Equité individuelle</u>

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

Définitions pas toutes compatibles

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Concepts clés et réglementation sur la discrimination
1.4 Comment mesurer l'équité

Institut des ACTUAIRES

Cadre défini
S variable protégée
Classification binaire : Y classe réelle et Ŷ classe estimée avec un modèle linéaire

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

Matrice de confusion

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

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

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

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

But : trouver des méthodes permettant d'atténuer la discrimination

1) Données simulées : contrôle des relations entre variables
2) Cas pratique avec des données réelles

Institut
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

2. Données simulées

2. Données simulées
2.1 Le processus de simulation

X : variables non sensibles (vecteur Gaussien)

A and B : variables sensibles (Bernoulli)

Y : variable d'intérêt (Bernoulli)

Corrélées entre elles de manière contrôlée
→ copule gaussienne

Processus de simulation

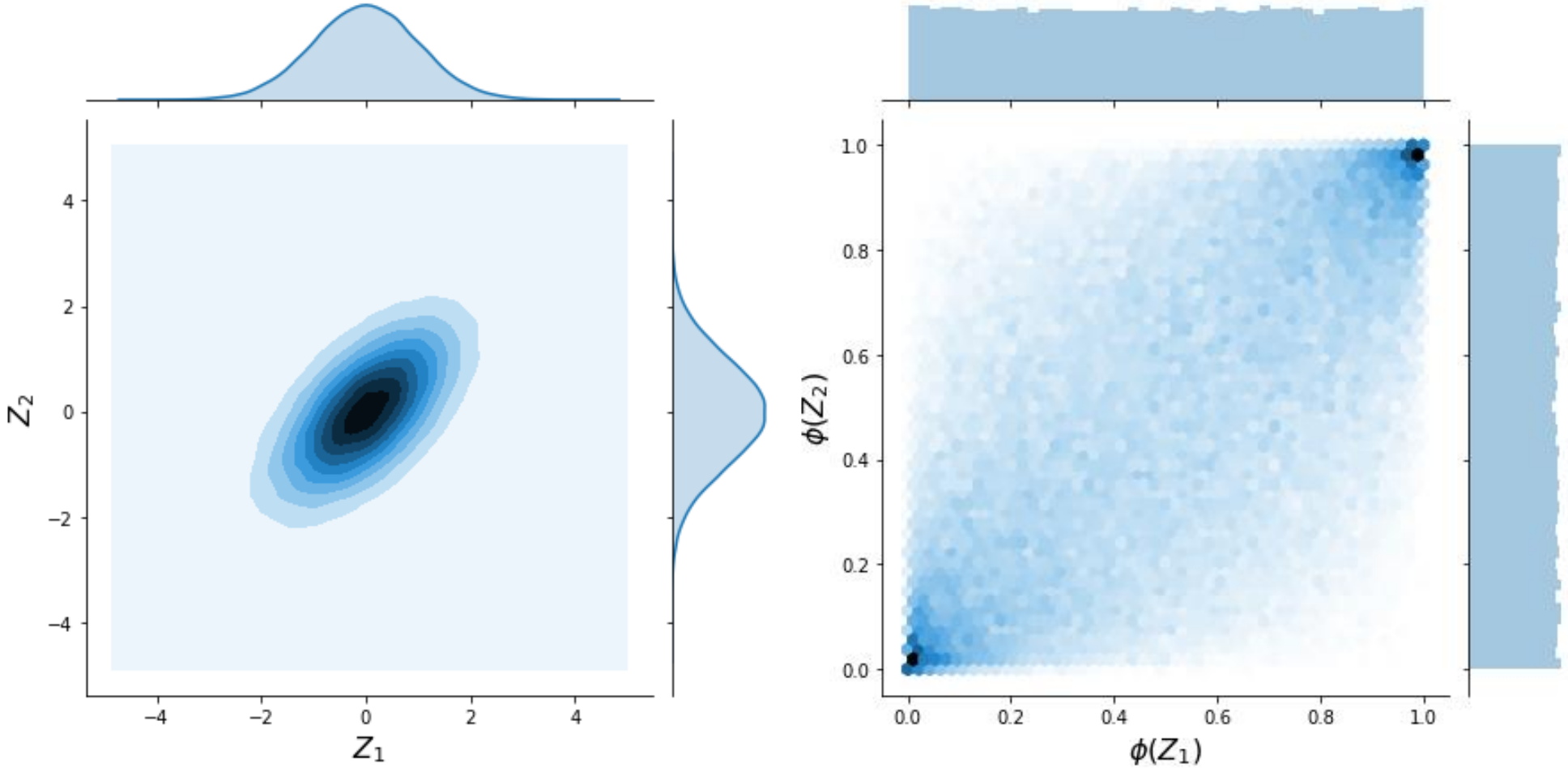
1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$

Institut
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

2. Données simulées
2.1 Le processus de simulation

Processus de simulation

1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$



$$R_Z = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$\rho = 0.6$$

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

2. Données simulées
2.1 Le processus de simulation

Processus de simulation

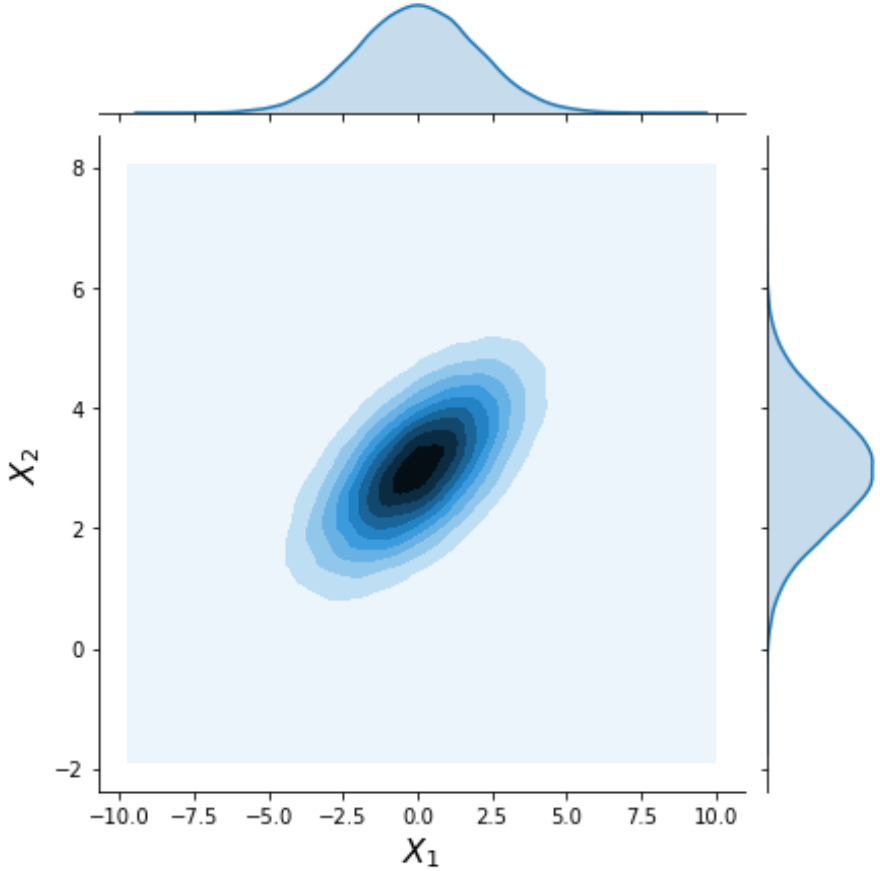
1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$

$$R_Z = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

$$corr(X_1, X_2) = \rho$$



Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$

2. Données simulées

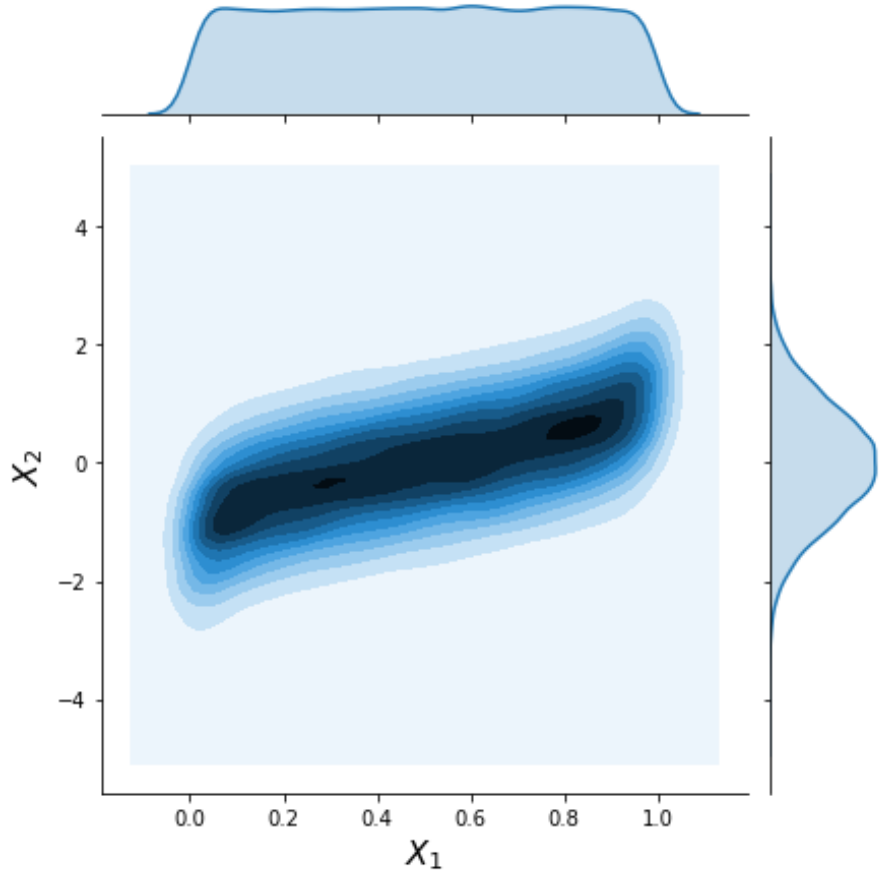
2.1 Le processus de simulation

$$R_Z = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$X_1 \sim U([a, b])$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

$$corr(X_1, X_2) = \sqrt{\frac{3}{\pi}}\,\boldsymbol{\rho}$$



Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

2. Données simulées
2.1 Le processus de simulation

Processus de simulation

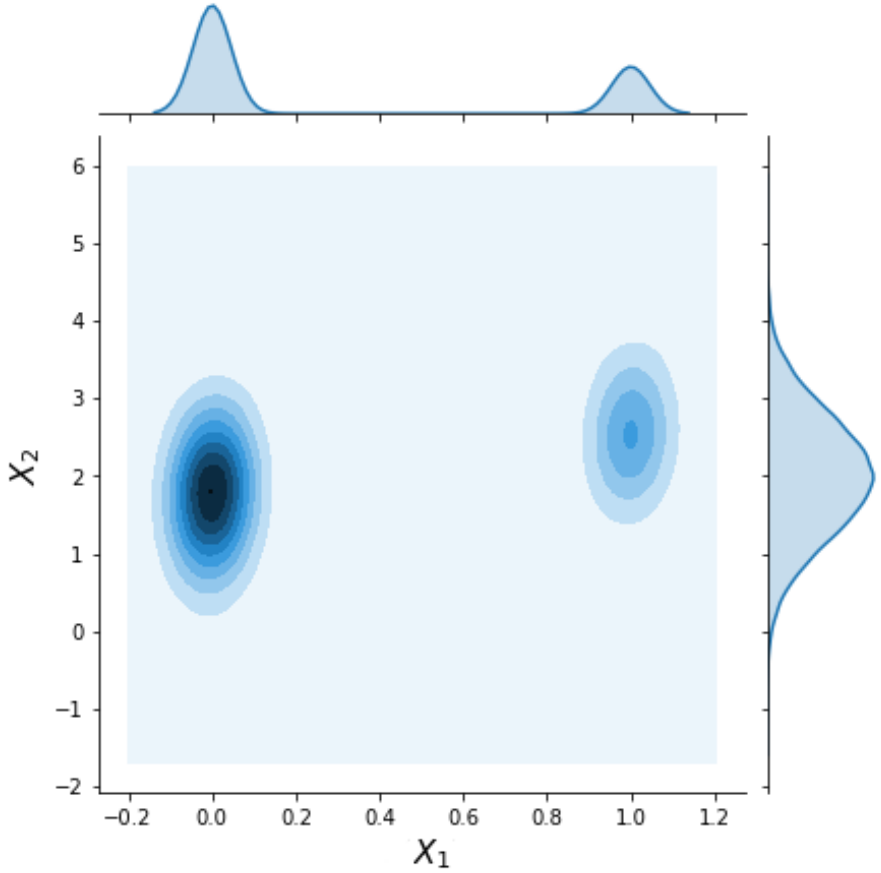
1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$

$$R_Z = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$X_1 \sim B(p)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

$$corr(X_1, X_2) = \frac{\rho e^{-\frac{(\Phi^{-1}(1-p))^2}{2}}}{\sqrt{(1-p)p2\pi}}$$



Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

1. Simuler $Z \sim N_n(0, R_Z)$
2. Calculer $U = (\Phi(Z_1), \dots, \Phi(Z_n))$
3. Calculer $(F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$

2. Données simulées
2.1 Le processus de simulation

Pour les données finales

On pose

$$X^{(i)} \sim N(\mu_i, \sigma_i^2)$$
$$A \sim B(p_a)$$
$$B \sim B(p_b)$$
$$Y \sim B(p_y)$$

$$R_{Z,ij} = \begin{cases} 1 \; si \; i = j \\ corr(Z_i, Z_j) \; sinon \end{cases}$$

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

2. Données simulées
2.2 Statistiques descriptives

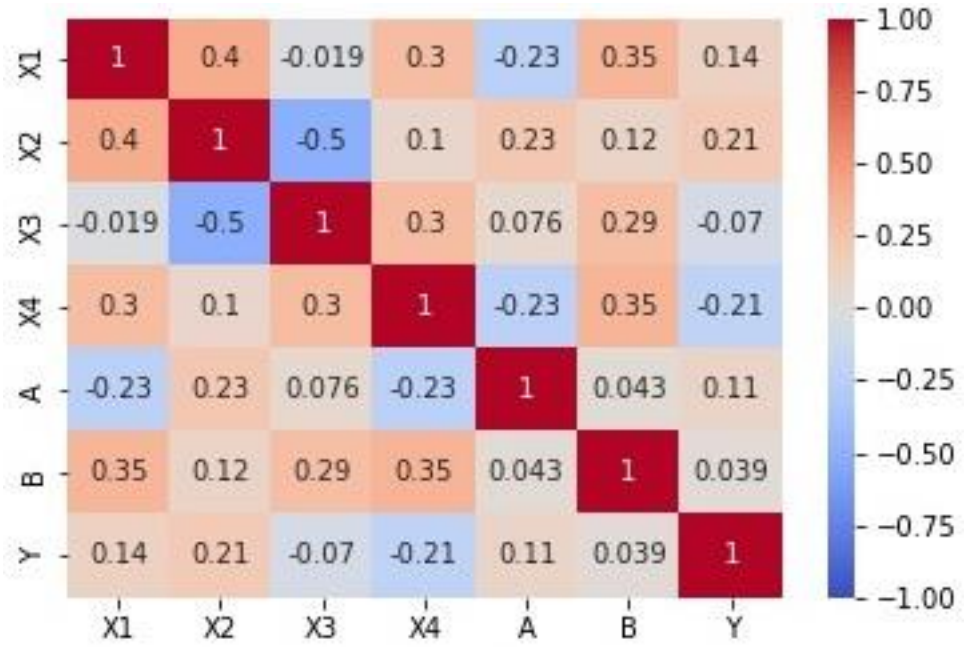
100 simulations → intervalles de confiance sur les résultats

Pour les données finales

On pose

$$X^{(i)} \sim N(\mu_i, \sigma_i^2)$$
$$A \sim B(p_a)$$
$$B \sim B(p_b)$$
$$Y \sim B(p_y)$$

$$R_{Z,ij} = \begin{cases} 1 \; si \; i = j \\ corr(Z_i, Z_j) \; sinon \end{cases}$$



Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

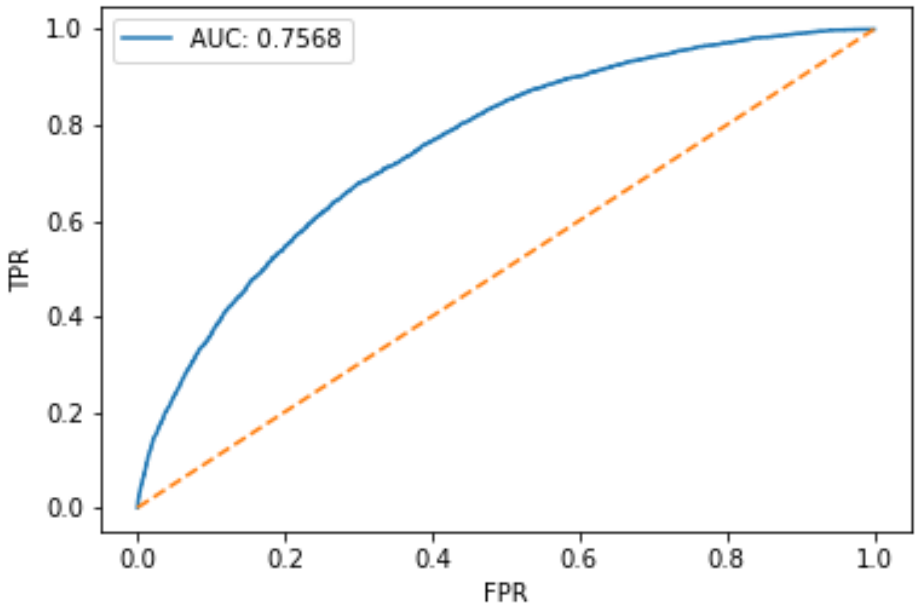
3. Atténuation de la discrimination sur les données simulées

3. Atténuation de la discrimination sur les données simulées

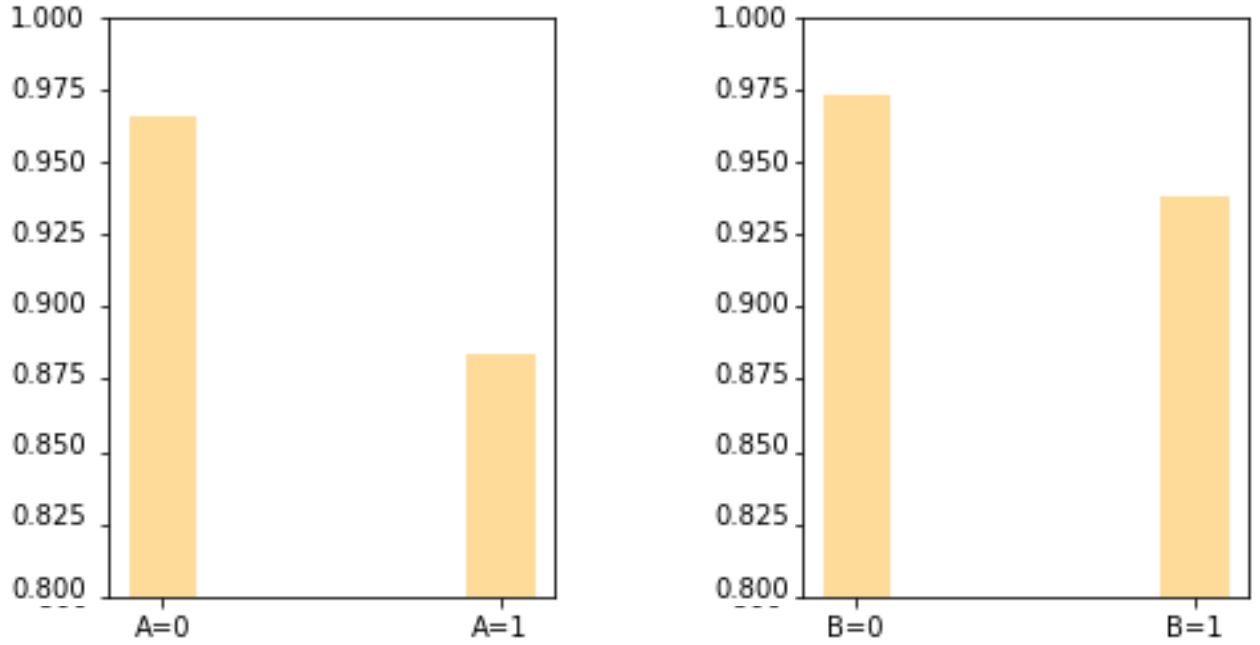
3.1 Modèle de régression logistique sans pré-traitement

Performance

(%)	With all variables
Accuracy	81.05 ± 0.05



Equité : taux d'acceptation



Selon toutes les définitions d'équité,
les groupes A=1 et B=1 sont désavantagés par le modèle.

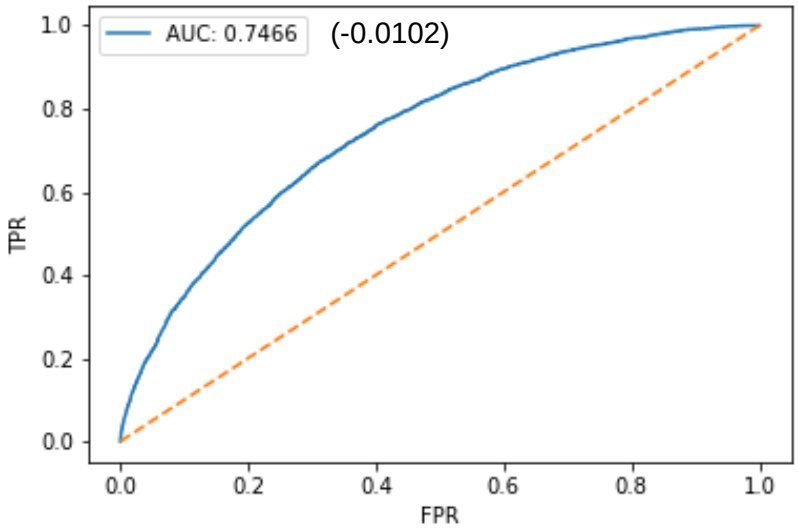
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

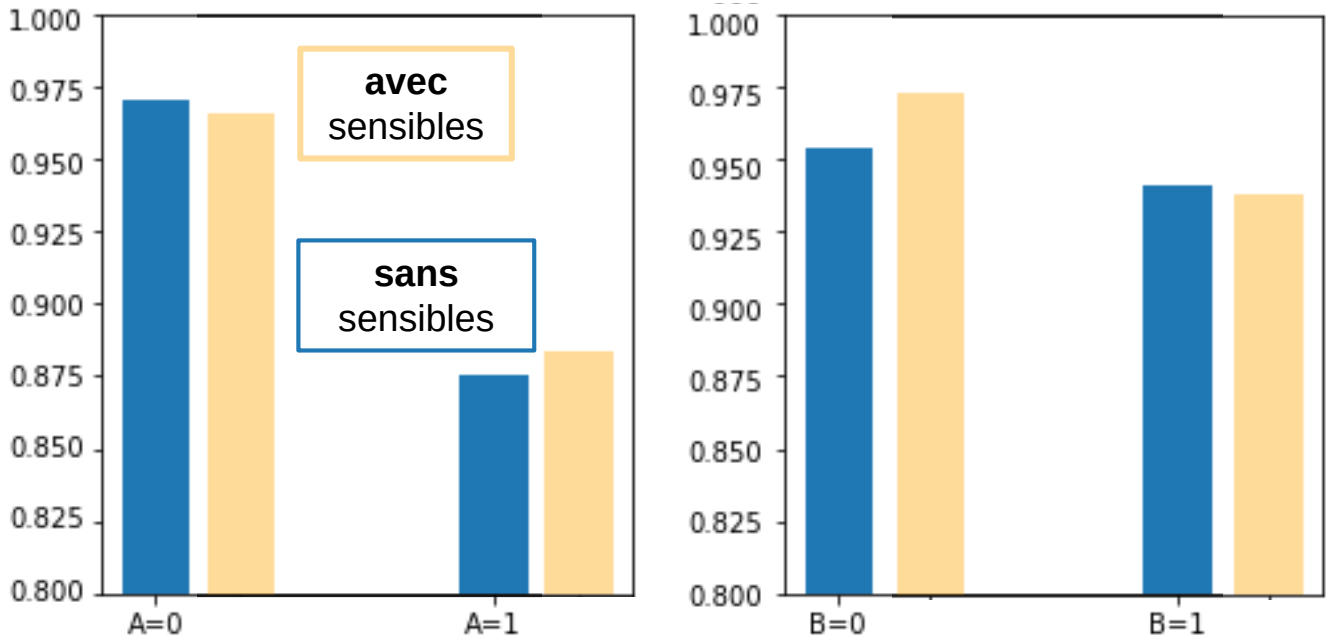
3.2 Sans les variables protégées : pas de discrimination directe

Performance

(%)	With all variables	Without protected variables
Accuracy	81.05 ± 0.05	81.04 ± 0.05



Equité : taux d'acceptation



Les groupes A=1 et B=1 sont encore désavantagés.

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

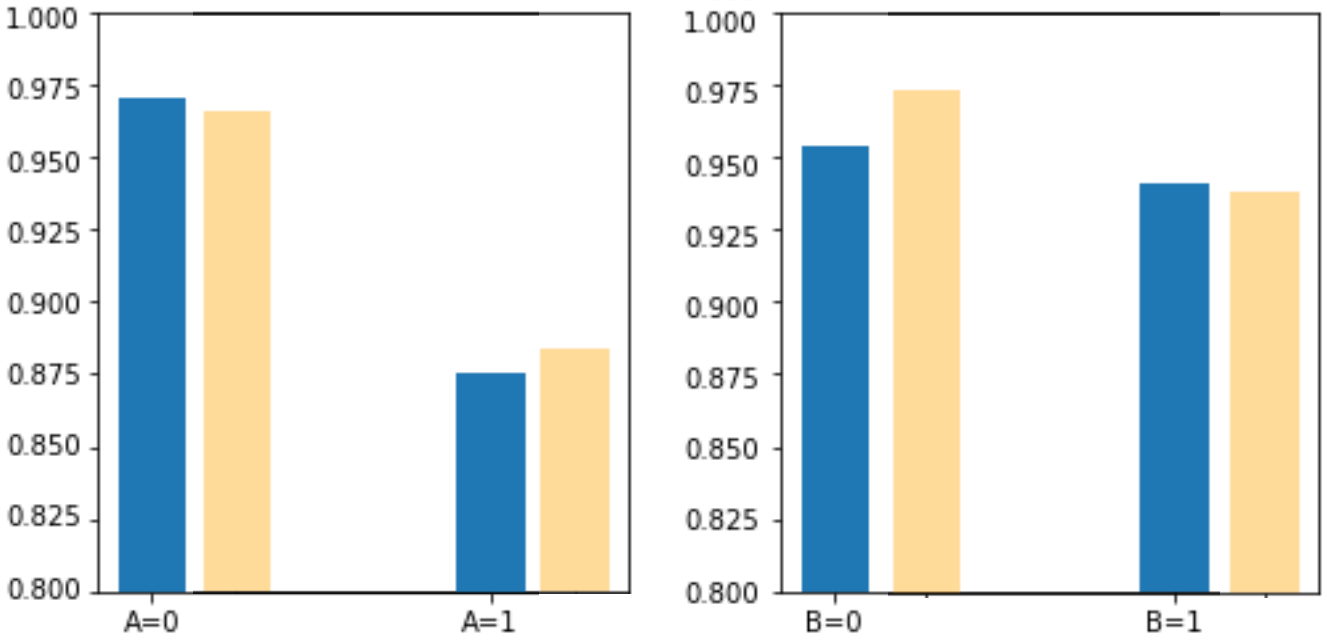
3. Atténuation de la discrimination sur les données simulées

3.2 Sans les variables protégées : pas de discrimination directe

Modèle **avec** variables sensibles

↓

Discrimination directe



Modèle **sans** variables sensibles

↓

Discrimination indirecte

Pas de mesure sans accès aux variables sensibles : leur collecte est indispensable.

Les résultats (équité) dépendent de la structure de dépendance.

Comment éviter la discrimination indirecte ?

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

3.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

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

↓

Indépendance ≈ absence de corrélation (au 1er ordre)

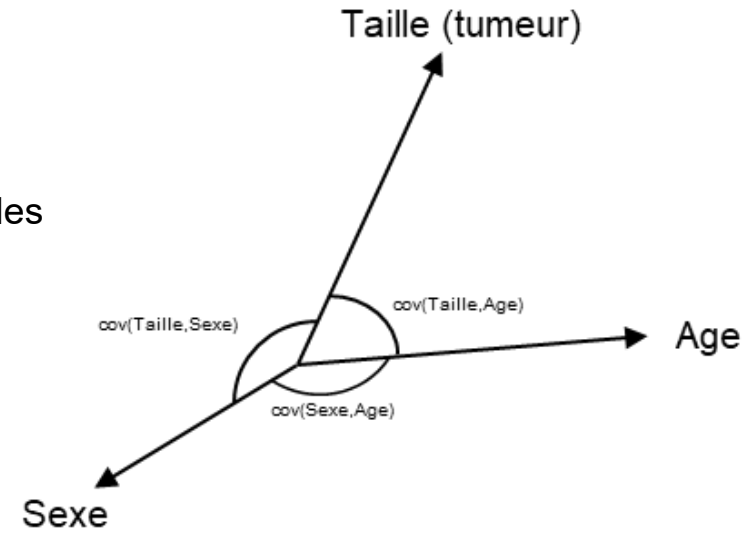
		Classe estimée	
		Positive	Négative
Classe réelle	Positive	TP	FN
	Négative	FP	TN

Matrice de confusion

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

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

Covariance = produit scalaire



Absence de corrélation ⇔ vecteurs orthogonaux

Institut
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

3.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

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

↓

Indépendance ≈ absence de corrélation (au 1er ordre)

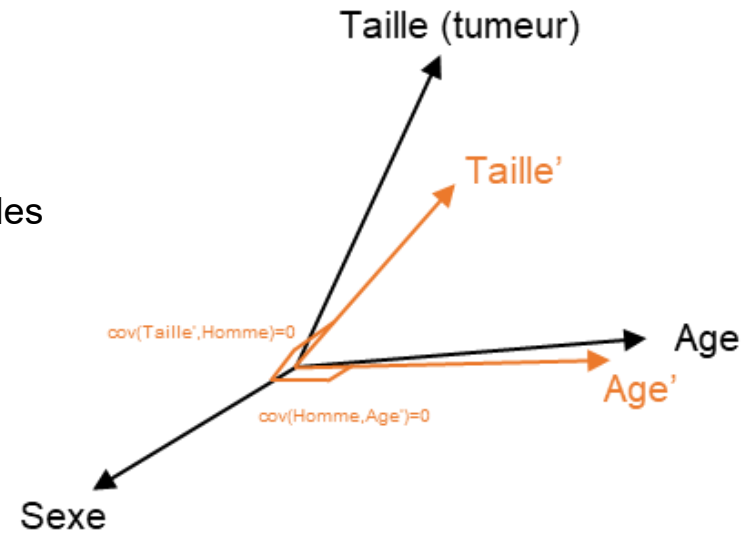
		Classe estimée	
		Positive	Négative
Classe réelle	Positive	TP	FN
	Négative	FP	TN

Matrice de confusion

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

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

Covariance = produit scalaire



Absence de corrélation ⇔ vecteurs orthogonaux

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

↓

Utiliser les variables transformées décorrélées dans le modèle

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

3.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

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 : $d(u,v) = \langle u-v, u-v \rangle$

Nouveau vecteur = combinaison linéaire ancien et sensibles

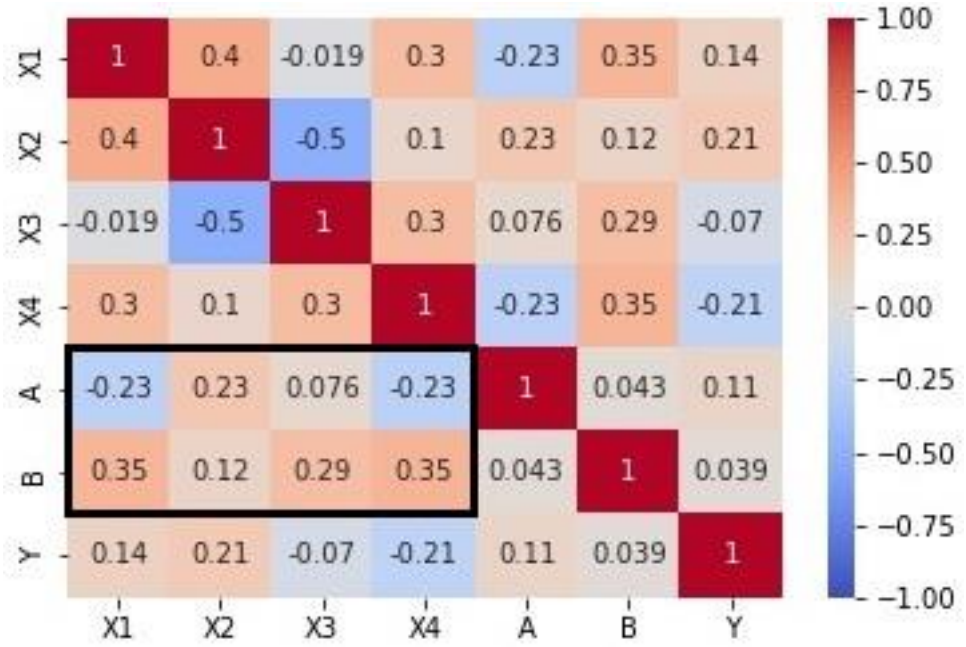
⇒ Pour chaque vecteur transformé : s+1 équations linéaires à s+1 inconnues

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

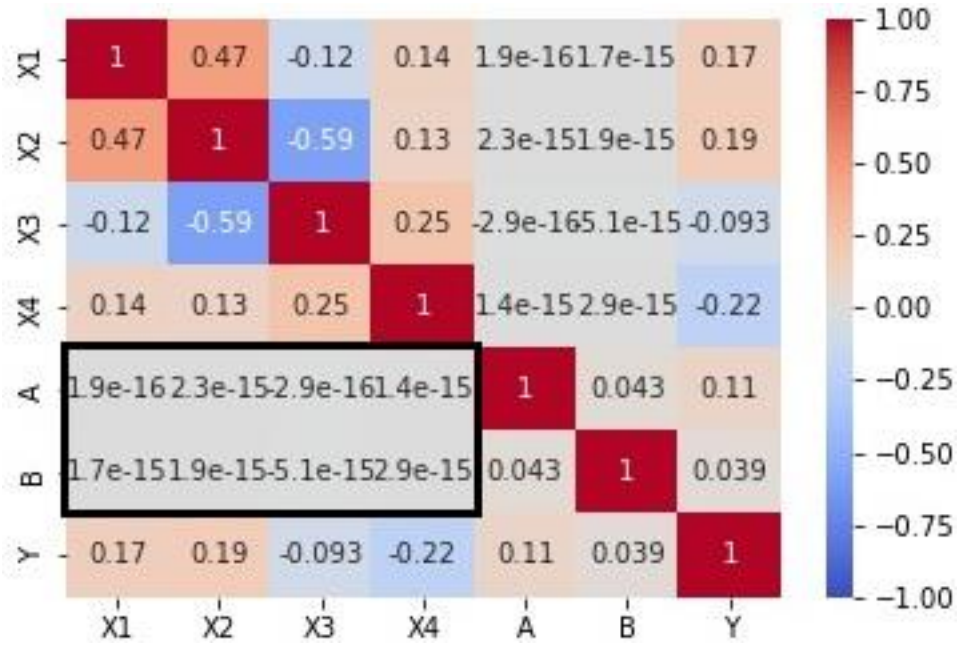
3. Atténuation de la discrimination sur les données simulées

3.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

Avant transformation



Après transformation



Les nouveaux vecteurs non sensibles sont maintenant décorrélés des vecteurs sensibles !

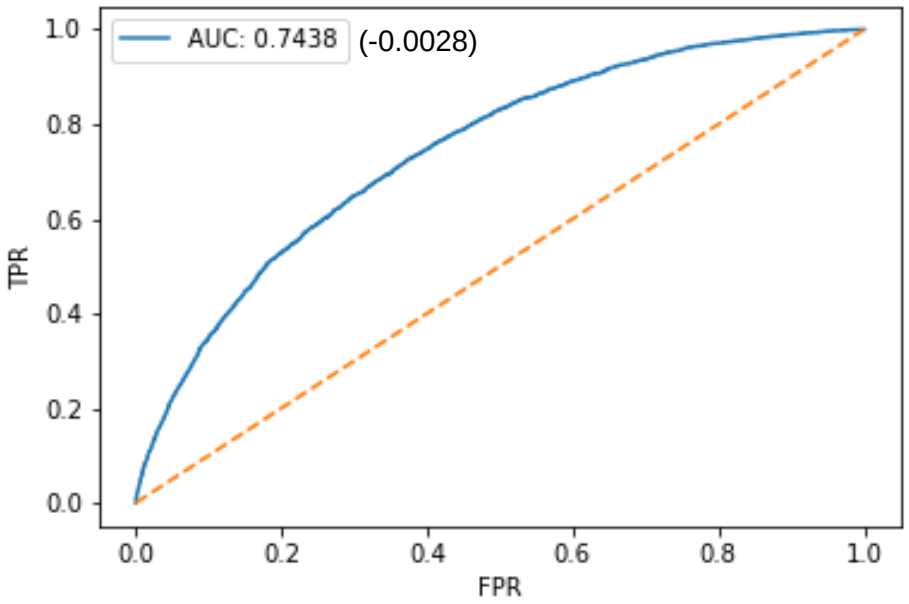
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

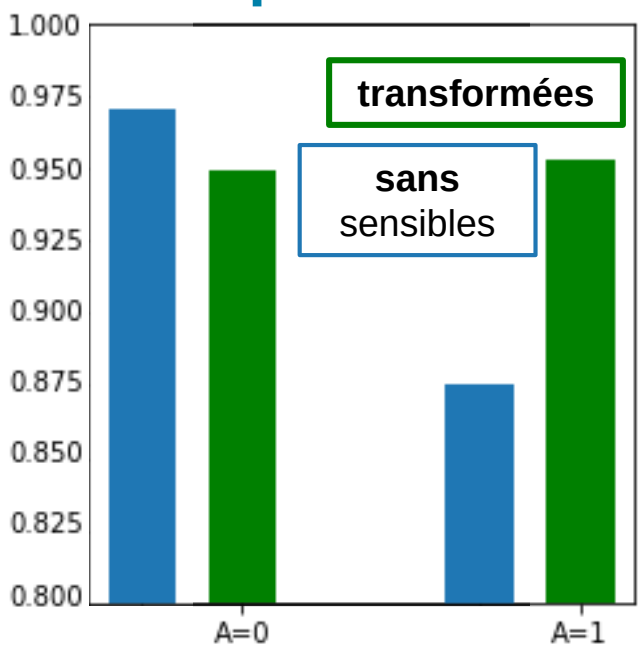
3.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

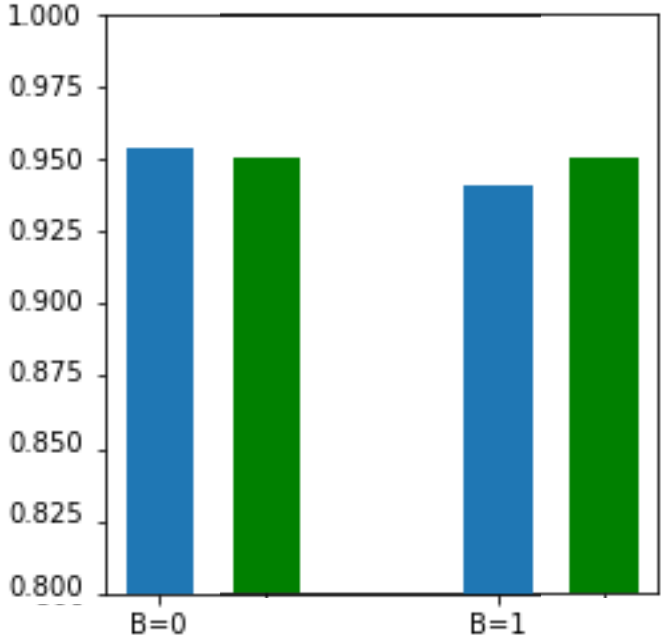
Performance

(%)	Without protected variables	With transformed variables
Accuracy	81.04 ± 0.05	80.81 ± 0.05



Equité





Les taux d'acceptation sont maintenant très proches pour tous les groupes : on a approché la parité statistique.

4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

3. Atténuation de la discrimination sur les données simulées

3.4 Conclusion sur les méthodes

On a approché la parité statistique

Corrélation : dépendance à l'ordre 1 uniquement → il pourrait y avoir des dépendances non-linéaires entre variables

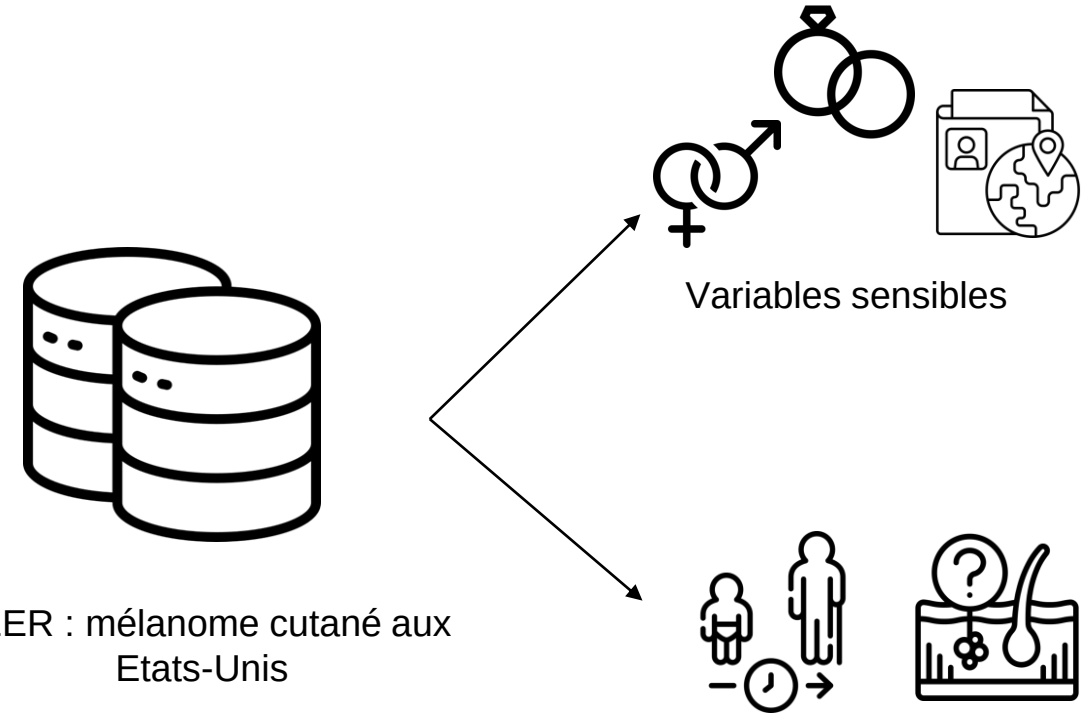
Performance : légère baisse de précision et d'AUC, car nous avons une information transformée et moins complète.

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.1 Présentation de la base de données



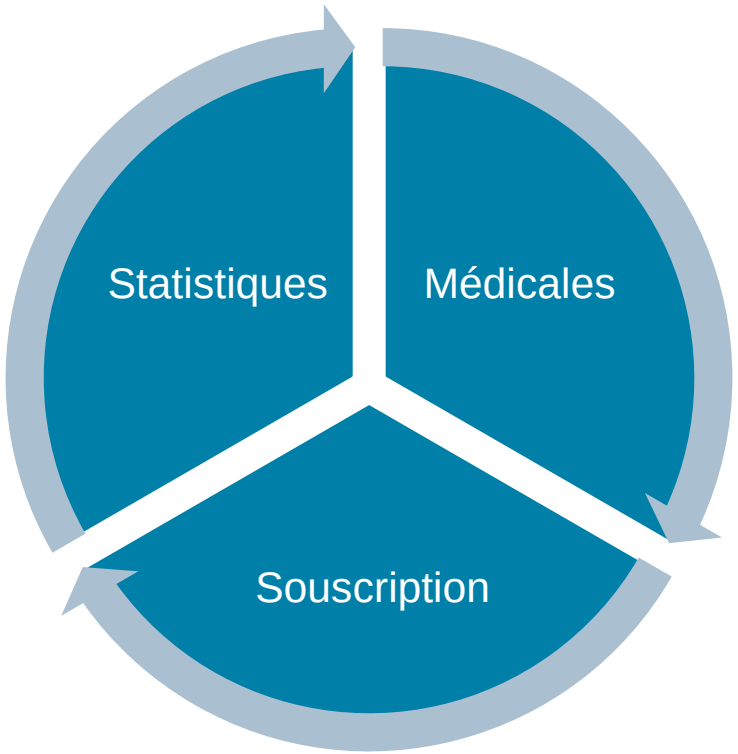
SEER : mélanome cutané aux Etats-Unis

Comment modéliser des taux de mortalité équitables ?

4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.1 Présentation de la base de données



Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.2 Spécificité de l'analyse de survie

Observation des données partielle à cause des phénomènes de censure et troncature → si ignorés, sous-estimation de la probabilité d'occurrence de l'événement d'intérêt

Modèles spécifiques à la survie

Modification de la structure des données pour utiliser des modèles standards avec l'exposition en poids

i	Age_dx	Year_dx	Survival (months)	Death_melanoma
1	25	2002	25	0
2	37	2004	4	1
3	56	2010	58	1

i	Survival (years) S_j	Max_duration (years)	j	Duration$_j$ (years)	Age	Year	Death_melanoma_j $d_{i,j}$	Exposure $e_{i,j}$
1	25/12=2.08	3	1	0	25	2002	0	1
			2	1	26	2003	0	1
			3	2	27	2004	0	0.08
2	4/12=0.33	1	1	0	37	2004	1	1
3	58/12=4.83	5	1	0	56	2010	0	1
			2	1	57	2011	0	1
			3	2	58	2012	0	1
			4	3	59	2013	0	1
			5	4	60	2014	1	1

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.2 Spécificité de l'analyse de survie

Taux de mortalité à 5 ans avec les données observées : un premier aperçu de l'influence de certaines variables

$$\hat{q}_5 = \frac{\sum_{i=1}^{l_8} d_{i,5}}{\sum_{i=1}^{l_5} e_{i,5}}$$

Décès dans les 5 ans

Exposition individuelle initiale

	Melanoma of the skin	All causes
M0	5.28%	
M1		83.00%

Five-year mortality rate by presence of metastasis

		Melanoma of the skin	All causes
M0	Missing	5.36%	
	I	1.44%	
	II	14.79%	
	III	29.87%	
M1	IV		83.00%

Five-year mortality rate by stage

Les taux sont cohérents avec la littérature médicale

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.2 Spécificité de l'analyse de survie

Taux de mortalité à 5 ans avec les données observées : un premier aperçu de l'influence de certaines variables

		Melanoma of the skin	All causes
M0	Women	3.82%	
	Men	6.38%	
M1	Women		81.09%
	Men		83.92%

Five-year mortality rate by gender

		Melanoma of the skin	All causes
M0	Hispanic	7.34%	
	Missing	0.07%	
	American Indian/AK Native	9.86%	
	Asian or Pacific Islander	11.73%	
	Black	18.82%	
	White	5.21%	
M1	Hispanic		86.53%
	Missing		35.71%
	American Indian/AK Native		88.45%
	Asian or Pacific Islander		93.85%
	Black		81.85%
	White		82.71%

Five-year mortality rates by origin

		Melanoma of the skin	All causes
M0	Divorced	9.71%	
	Married	5.86%	
	Missing	1.75%	
	Separated	8.56%	
	Single	6.27%	
	Unmarried	7.71%	
	Widowed	11.17%	
M1	Divorced		85.15%
	Married		81.61%
	Missing		77.78%
	Separated		80.51%
	Single		82.94%
	Unmarried		74.38%
	Widowed		90.75%

Five-year mortality rates by marital status

Taux différents selon le sexe, l'origine nationale, l'état civil

Attention : la distribution des autres variables n'est pas la même dans toutes les catégories

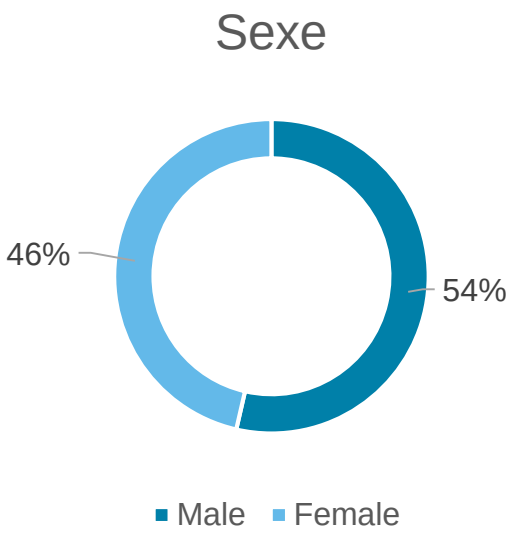
Intuition : un modèle utilisant directement ces données peut être discriminant envers certains groupes

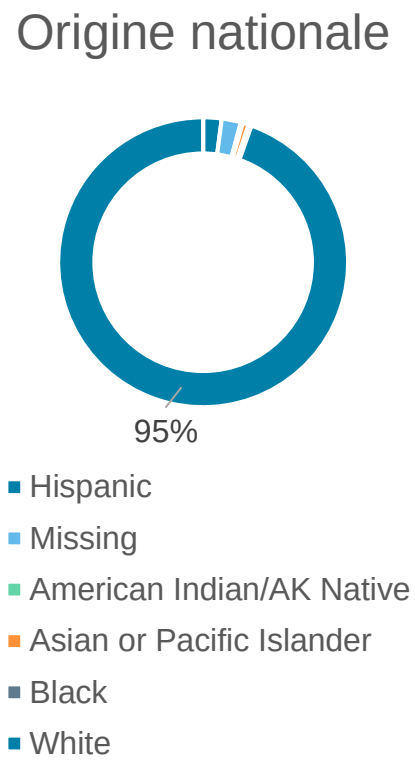
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

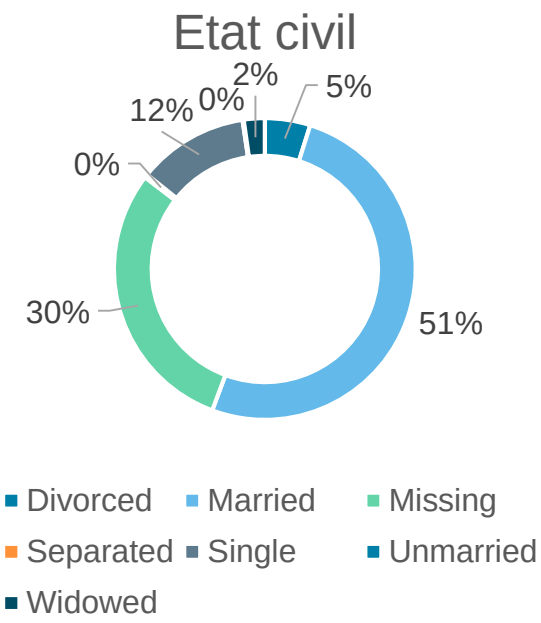
4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.3 Statistiques descriptives

Déséquilibre du nombre d'observations pour les variables sensibles





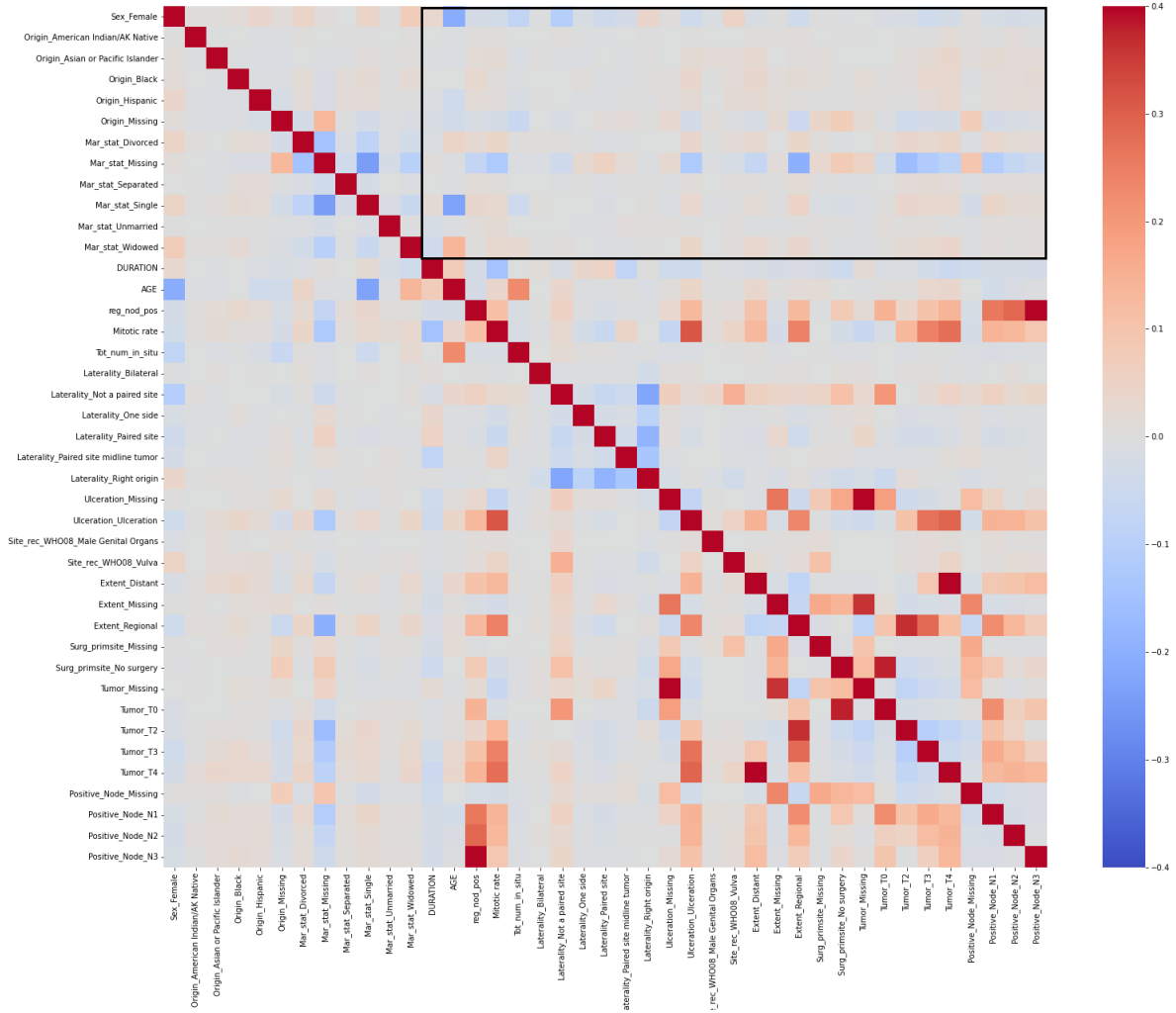


Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

4. Cas pratique : la mortalité des individus atteints du mélanome cutané

4.3 Statistiques descriptives

Présence de corrélations entre variables sensibles et non sensibles qui peuvent servir de proxy



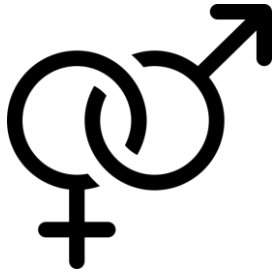
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

5. Atténuation de la discrimination sur les données réelles

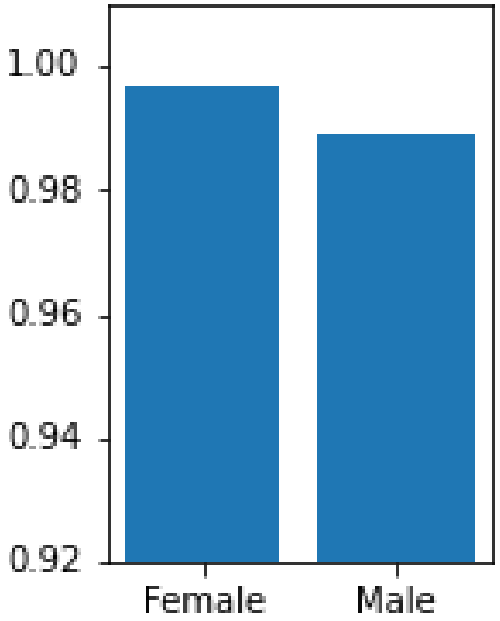
5. Atténuation de la discrimination sur les données réelles

5.1 Modèle de régression logistique avec et sans variables sensibles

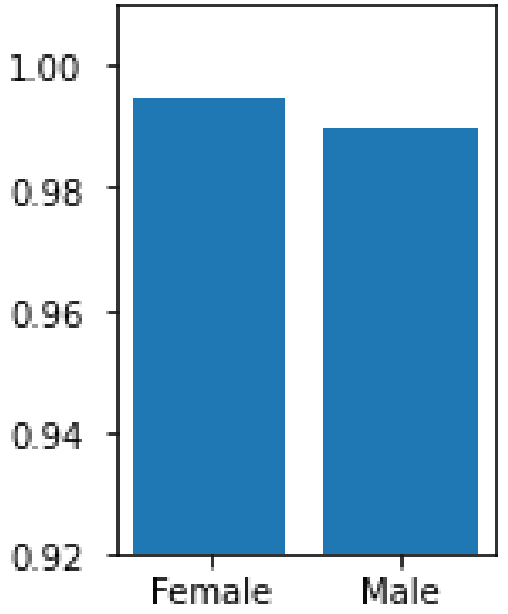
Taux d'acceptation par sexe



Modèle **avec** variables sensibles



Modèle **sans** variables sensibles



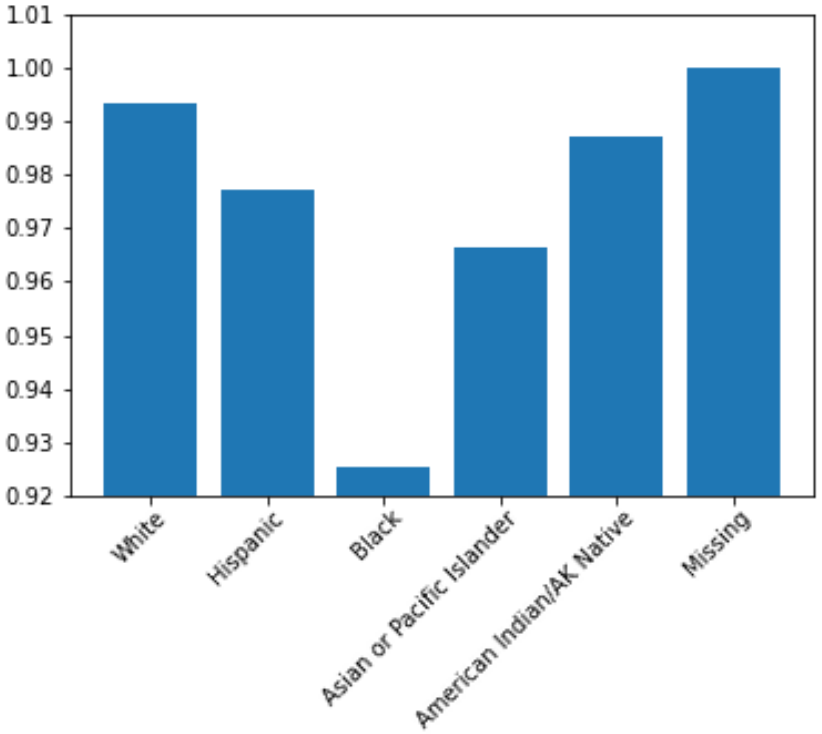
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

5. Atténuation de la discrimination sur les données réelles

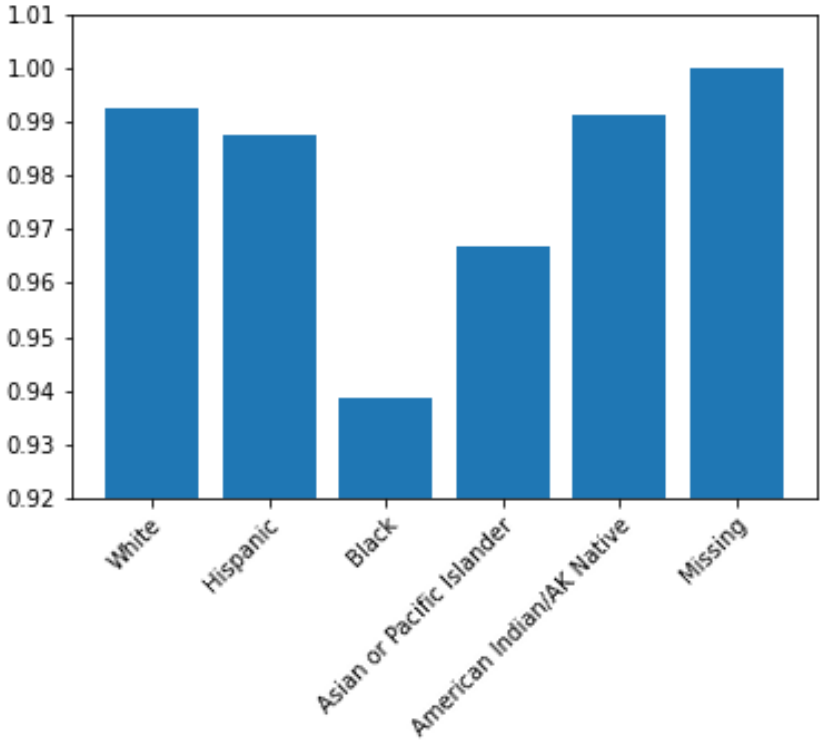
5.1 Modèle de régression logistique avec et sans variables sensibles

Taux d'acceptation par origine nationale

Modèle **avec** variables sensibles



Modèle **sans** variables sensibles

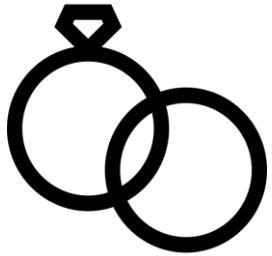


Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

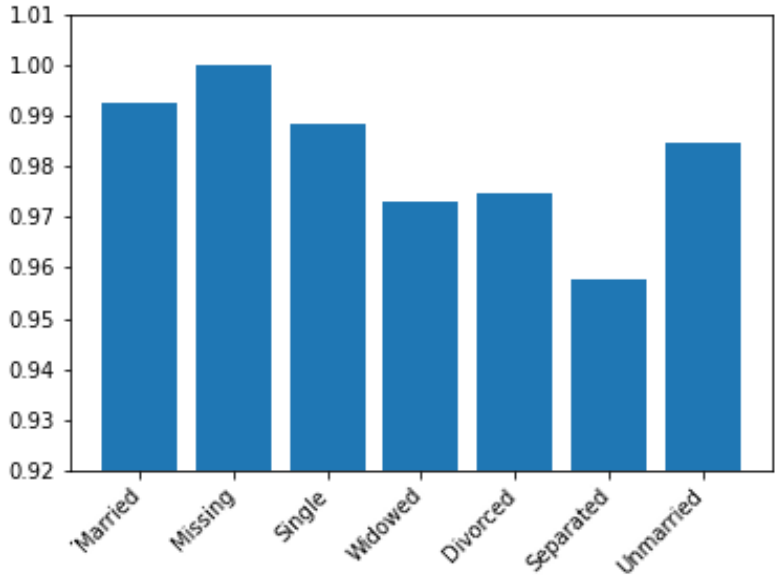
5. Atténuation de la discrimination sur les données réelles

5.1 Modèle de régression logistique avec et sans variables sensibles

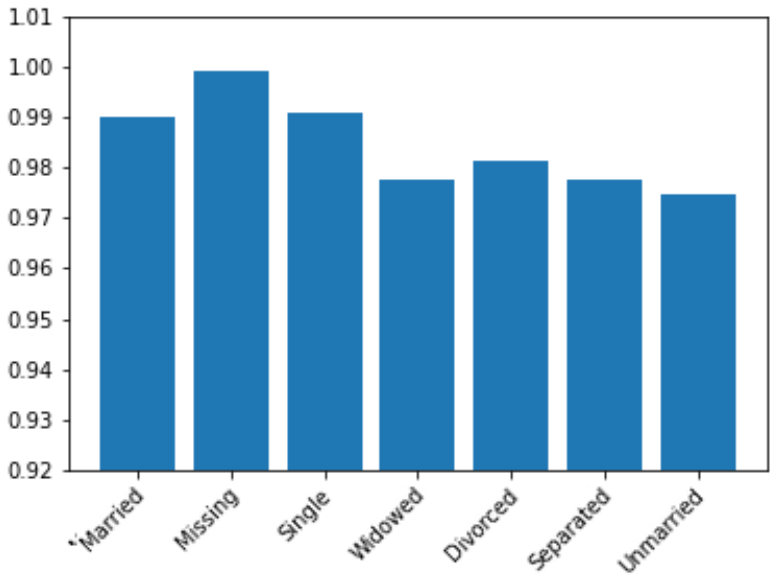
Taux d'acceptation par état civil



Modèle **avec** variables sensibles



Modèle **sans** variables sensibles

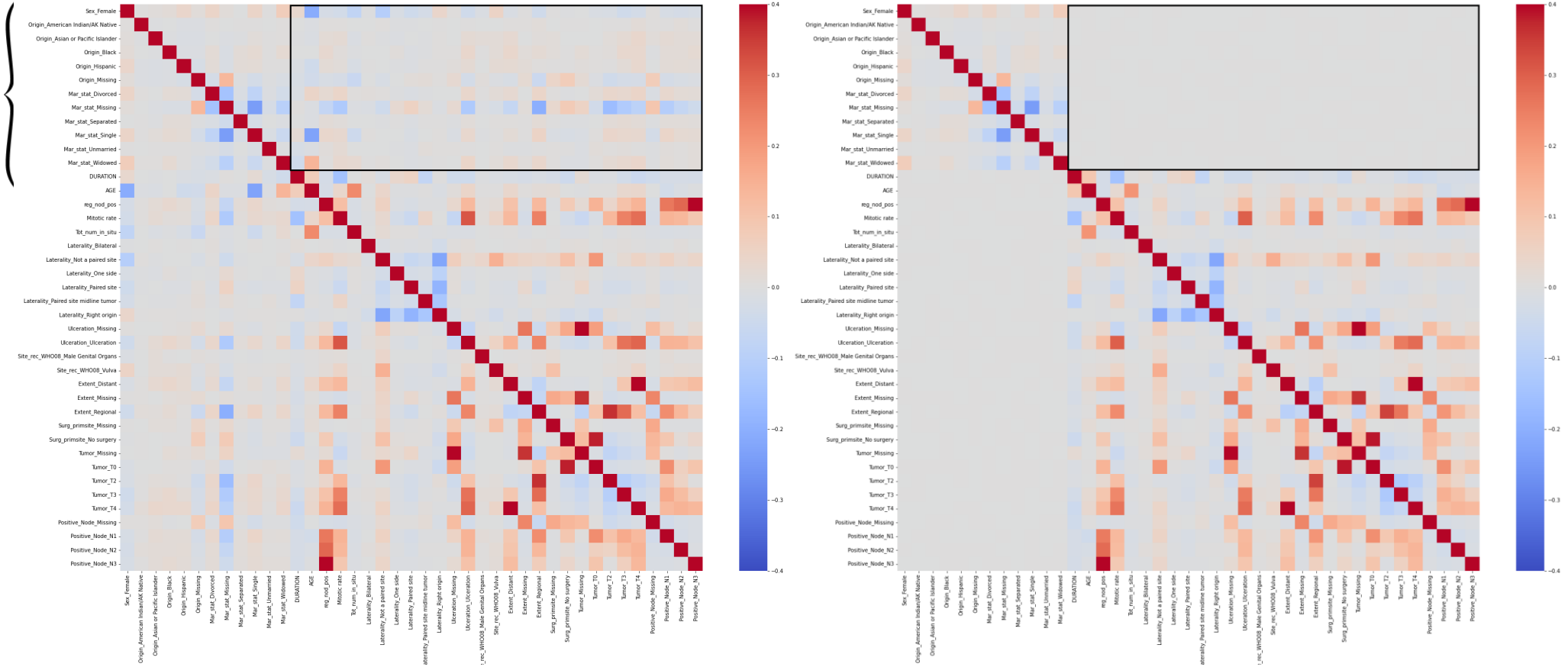


Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

5. Atténuation de la discrimination sur les données réelles

5.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

Sensibles



Variables d'origine

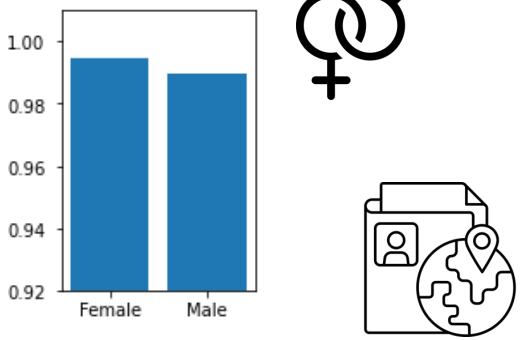
Variables transformées

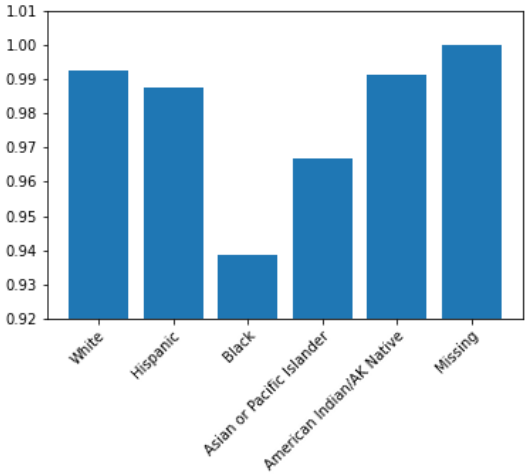
Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

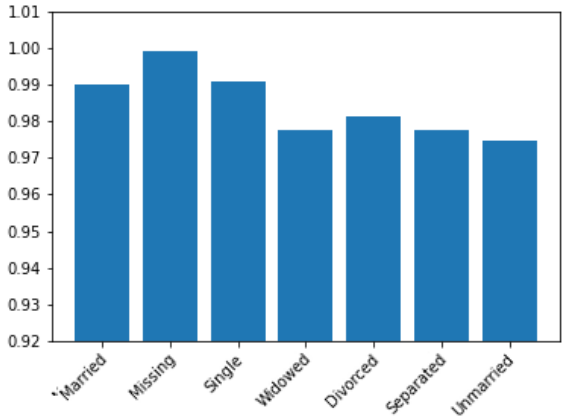
5. Atténuation de la discrimination sur les données réelles

5.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

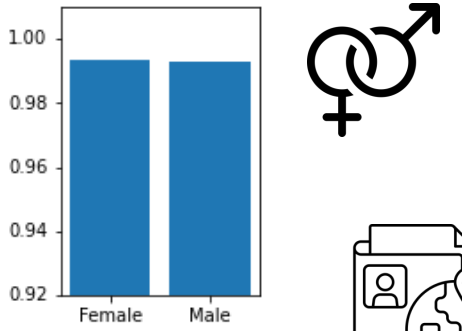
Modèle **sans** variables sensibles

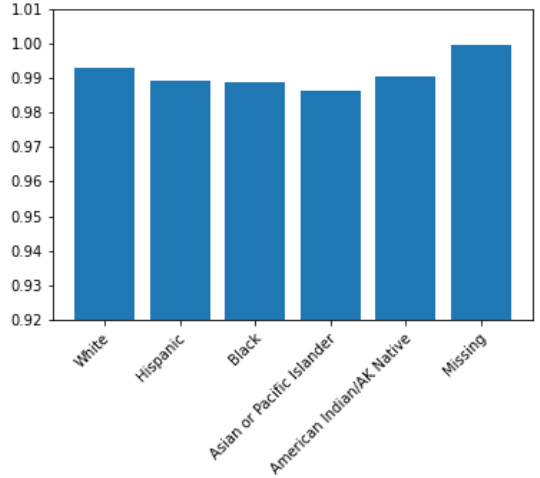


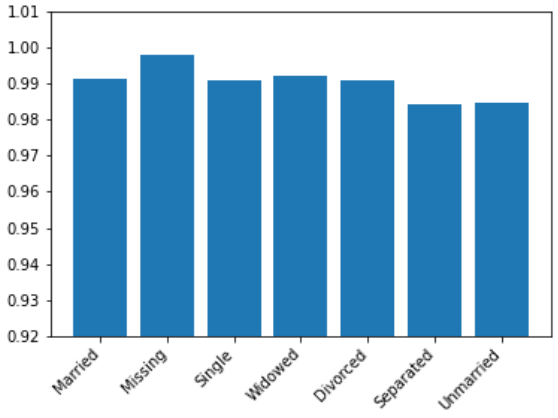




Modèle avec variables **transformées**







Ecarts restants : dépendances non linéaires

INSTITUT
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

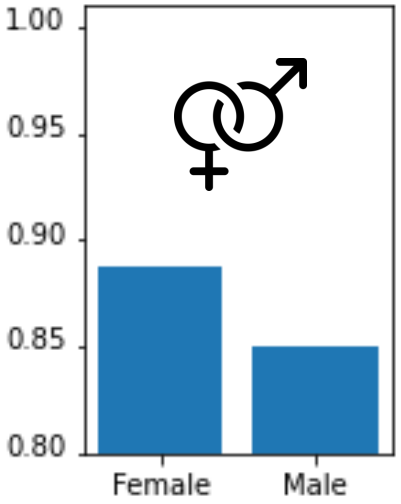
5. Atténuation de la discrimination sur les données réelles

5.3 Transformation des variables non sensibles pour atténuer la discrimination indirecte

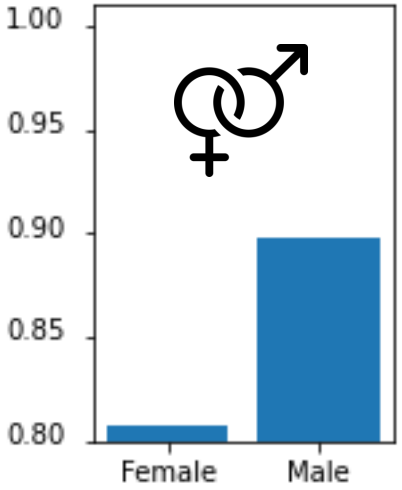
En regardant une autre définition d'équité

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

FPR **sans** variables sensibles



FPR avec variables **transformées**

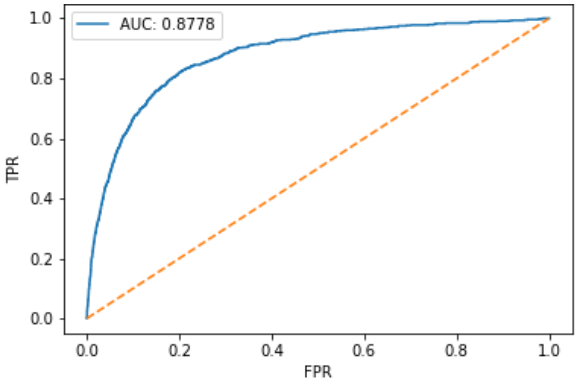


Incompatibilité des définitions d'équité

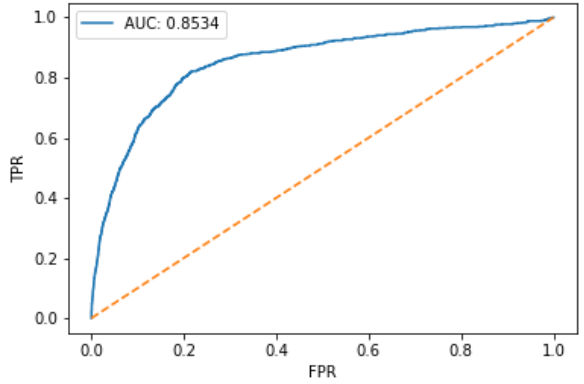
Et la performance ?

Légère baisse de l'AUC

sans variables sensibles



variables **transformées**



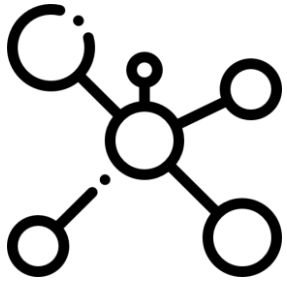
Compromis équité-performance

Institut des Actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com

Conclusion

Conclusion

La méthode est un **succès** : approximation de la parité statistique
→ atténuation de la discrimination indirecte selon cette définition



Des modèles complexes peuvent facilement détecter des relations non linéaires entre variables → aller **au-delà de la décorrélation**

Hypothèse d'un cadre de classification binaire → **étendre** à d'autres cas (régression)

Problématique encore largement ouverte académiquement et réglementairement

Bonne définition d'équité ?

Réglementation en plein essor
(ex AI Act en Europe)

4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com